CALIBRATION AND EVALUATION OF WRF-HYDRO MODELING SYSTEM FOR EXTREME RUNOFF SIMULATIONS: USE OF HIGH-RESOLUTION SEA SURFACE TEMPERATURE (SST) DATA

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

BY

BERİNA MİNA KILIÇARSLAN

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

FEBRUARY 2022

Approval of the thesis:

CALIBRATION AND EVALUATION OF WRF-HYDRO MODELING SYSTEM FOR EXTREME RUNOFF SIMULATIONS: USE OF HIGH-RESOLUTION SEA SURFACE TEMPERATURE (SST) DATA

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

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences**

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

Prof. Dr. İsmail Yücel
Supervisor, **Civil Engineering, METU**

Examining Committee Members:

Prof. Dr. İsmail Yücel
Civil Engineering, METU

Prof. Dr. Elçin Kentel
Civil Engineering, METU

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

Assoc. Prof. Dr. Koray K. Yılmaz
Civil Engineering, METU

Assoc. Prof. Dr. Marouane Temimi
Civil, Environmental, and Ocean Engineering, Stevens Institute of Technology

Date: 11.02.2022

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

Name Last name : Berina Mina Kılıçarslan

Signature :

ABSTRACT

CALIBRATION AND EVALUATION OF WRF-HYDRO MODELING SYSTEM FOR EXTREME RUNOFF SIMULATIONS: USE OF HIGH-RESOLUTION SEA SURFACE TEMPERATURE (SST) DATA

Kılıçarslan, Berina Mina
Master of Science, Civil Engineering
Supervisor: Prof. Dr. İsmail Yücel

February 2022, 146 pages

This study investigates the impact of the spatio-temporal accuracy of four different sea surface temperature (SST) datasets on the accuracy of the Weather Research and Forecasting (WRF)-Hydro system to simulate hydrological response during two catastrophic flood events over the Eastern Black Sea (EBS) and the Mediterranean (MED) regions of Turkey. Three daily-updated and high spatial resolution external SST products (GHRSST, Medspiration, and NCEP-SST) and one coarse-resolution and time-invariant SST product (ERA5- and GFS-SST for EBS and MED regions, respectively) already embedded in the initial and the boundary conditions datasets of WRF model are used in deriving near-surface atmospheric variables through WRF. Event-based calibration is performed to the WRF-Hydro system using hourly and daily streamflow data in both regions. Two different calibration methods, namely step-wise and automated calibration, are applied for the calibration process. Dynamically Dimensioned Search (DDS) Algorithm is utilized as the automated calibration algorithm to investigate the improvements in hydrograph simulations alternative to the step-wise approach. The uncoupled model simulations for independent SST events are conducted to assess the impact of SST-triggered precipitation on simulated extreme runoff. Despite the fact that manual calibration shows better performance with fewer iteration numbers, results show that the process-based automated calibration approach considering the impact of parameters on the hydrological behavior exhibits a promising performance, particularly in the EBS region. Some localized and temporal differences in the occurrence of the flood events with respect to observations depending on the SST representation are noticeable. SST products represented with higher cross-correlations (GHRSST and Medspiration) revealed significant improvement in flood hydrographs for both regions. The GHRSST dataset shows a substantial improvement in NSE (~70%) and KGE (from 0.06 to 0.3) with respect to the invariable SST (ERA5) in simulated runoffs over the EBS region. Reduction in RMSE up to 20% and an increase in correlation from 0.3 to 0.8 is observed for the same region. The use of both GHRSST and Medspiration SST data characterized with high spatio-temporal correlation resulted in runoff simulations exactly matching the observed runoff peak of 300 m^3/s by reducing the overestimation seen in invariable SST (GFS) in the MED region. In addition, the KGE values are increased from 0.1 to 0.3 for the hydrographs generated with high-resolution SST simulations. Improved precipitation simulation skills of the WRF model with the detailed SST representation show that the hydrographs of GHRSST and Medspiration simulations show better performance compared to the simulated hydrographs by observed precipitation.

Keywords: Calibration, Sea Surface Temperature, WRF, WRF-hydro

ÖZ

AŞIRI AKIM SİMÜLASYONLARI İÇİN WRF-HYDRO MODEL SİSTEMİNİN KALİBRASYONU VE DEĞERLENDİRİLMESİ: YÜKSEK ÇÖZÜNÜRLÜKLÜ DENİZ YÜZEY SICAKLIĞI (DYS) VERİLERİNİN KULLANIMI

Kılıçarslan, Berina Mina
Yüksek Lisans, İnşaat Mühendisliği
Tez Yöneticisi: Prof. Dr. İsmail Yücel

Şubat 2022, 146 sayfa

Bu çalışma, dört farklı deniz yüzeyi sıcaklığı (DYS) veri setinin uzaysal-zamansal doğruluğunun, Türkiye'nin Doğu Karadeniz (EBS) ve Akdeniz (MED) bölgelerinde gerçekleşen sel olayları sırasında Weather Research and Forecast (WRF)-Hydro sisteminin ürettiği hidrolojik tepki üzerindeki etkisi araştırılmaktadır. Günlük zamansal çözünürlüğe sahip ve aynı zamanda yüksek mekansal çözünürlüklü üç farklı DYS ürünü (GHRSST, Medspiration ve NCEP-DYS) ve bir düşük çözünürlüklü ve zamansal olarak sabit DYS ürünü (sırasıyla EBS ve MED bölgeleri için ERA5- ve GFS- DYS) test edilmiştir. WRF modelinin başlangıç ve sınır koşulu girdileri, WRF aracılığıyla yüzeye yakın atmosferik değişkenlerin türetilmesinde kullanılır. Her iki bölgede de saatlik veya günlük akış verileri kullanılarak WRF-Hydro sistemi için olay bazlı kalibrasyon gerçekleştirilmiştir. Aşamalı kalibrasyon yaklaşımına ek olarak, otomatik kalibrasyon algoritması olarak değerlendirilen Dinamik Olarak Boyutlandırılmış Arama (DDS) Algoritması, hidrograf simülasyonlarındaki olabilecek iyileştirmeyi araştırmak amacıyla kullanılmıştır. Takiben farklı DYS verileri ile simüle edilmiş yağış verilerinin akım üzerindeki etkisini değerlendirilmiştir. DYS temsiline bağlı olarak gözlemlere göre taşkın olaylarının meydana gelmesinde bazı mekansal ve zamansal farklılıklar göze çarpmaktadır. Daha yüksek çapraz korelasyona sahip (GHRSST ve Medspiration) DYS ürünleri, her iki bölge için taşkın hidrograflarında önemli bir gelişme sağlamıştır. GHRSST veri seti EBS bölgesi üzerinde simüle edilmiş akımlarda düşük çözünürlüklü DYS 'ye (ERA5) göre, NSE'de (~%70) ve KGE'de (0.06'dan 0.3'a) önemli gelişmeler gözlemlenmiştir. RMSE'de %20'ye varan azalma ve korelasyonda 0,3'ten 0,8'e bir artış sağlamıştır. Yüksek mekansal-zamansal korelasyona sahip GHRSST ve Medspiration DYS verilerinin kullanımı, MED bölgesinde düşük çözünürlüklü DYS simülasyonunda (GFS) görülen aşırı tahmini azaltarak, 300 m^3/s olarak gözlemlenen tepe akım noktasıyla tam olarak eşleşen akım değeri üretmiştir. WRF modeli içerisinde tanımlanmış yüksek çözünürlüklü DYS temsili yağış simülasyonlarını geliştirmiş GHRSST ve Medspiration simülasyonlarının hidrograflarının, gözlem yağışı ile simüle edilmiş hidrograflara kıyasla daha iyi performans gösterdiğini kanıtlamıştır. Ek olarak yüksek çözünürlüklü DYS simülasyonları sonucunda elde edilen hidrograflar için KGE değerlerinde 0.1'den 0.3'a bir artış gözlemlenmiştir.Manuel kalibrasyonun daha az yineleme sayısı ile daha iyi performans göstermesine rağmen, parametrelerin hidrolojik davranış üzerindeki etkisini dikkate alan süreç bazlı otomatik kalibrasyon yaklaşımı ile kalibre edilmiş model çıktıları özellikle EBS bölgesinde umut verici bir performans sergilediği gözlemlenmiştir.

Anahtar Kelimeler: Kalibrasyon, Deniz Yüzey Sıcaklığı, WRF, WRF-Hydro

To my family

and

to the memory of

my time traveler father

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to Prof. İsmail Yücel for letting me start this academic journey with him. For me, he is a rare example of an academician who loves his job/research and cares about his team; like a farmer, loves and cares about his garden and trees. He has supported and enlightened us about a new world unknown to us and waited with endless patience, watching us grow and flourish. When I was going through the hardest times of my life, he just stayed with all his respect and encouraged me to back to life. I will never forget his guidance throughout my life.

I would like to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for providing funding for this study under the project entitled 'Evaluation of Integrated Atmospheric-Hydrologic Modeling System for Flood Events at Two Different Basins' with project number 116Y193.

Thank you to Middle East Technical University for being my *home*. I will miss all your trees, your sunset, your cobblestone pavements, and your bare concrete buildings telling me the stories of resistance and endurance of many generations for years.

My sincere thanks to Assoc. Prof. M. Tuğrul Yılmaz for his invaluable contributions to my paper and thesis. He always did his best to provide helpful feedback.

I sincerely appreciated the valuable suggestions and comments of Assoc. Prof. Koray K. Yılmaz.

I would like to thank Prof. Elçin Kentel, who is always an inspiring person for me with her sincere attitude and stance as a professor.

I would also like to thank Assoc. Prof. Marouane Temimi for being part of my jury members and allowing me to continue this journey as a Ph.D. student in his lab.

I am deeply grateful to my teammates Eren Düzenli and Heves Pilatin for their endless friendship and help. They are the most talented people with whom I`ve ever had the pleasure of working.

I would like to express my deepest gratitude to my father, Muharrem Kılıçarslan. His love and furious advocacy of nature, his METU-ian mind, and his care for every single living creature on earth created the person I am today. Thank you for giving me the courage to pursue my dreams. It was with a heavy heart that I saw him leave, but I know he will be with me forever.

I would like to thank my dearest mother, Meryem Kılıçarslan, the strongest woman I have ever known. Her endless love is the greatest support that a daughter could ever ask for. She tried to find an exit path all the time and took me out of the worst situations. Her invaluable guidance and motivation led incredible things to happen in my life.

I would like to thank my dearest brother-my sidekick, Akanda Sina Kılıçarslan. We have been through many waves together in this ocean. We will continue to sail, and I know whenever I am happy, exclaim or murmur, he will continue to urge me to notice and remember Mr. Vonnegut`s words: "If this isn't nice, I don't know what is?"

Special thanks to my grandparents, Sabriye and Şevket Büyükyıldız, for all their prayers. They always be my *shelter from the storm*. I am so thankful for their presence.

Thank you to my dearest friend, my Buddha, Okan Toygun Öztürkler, for his patience, advice, and faith. Because he always understood.

My special thanks to Beyza Özel Tindoğan for her lifelong friendship and endless support. I always feel stronger with her presence in my life.

I am thankful to my dear friends Nur Çelik, Merve Yücel, and my neighbor Totoro-Elif Urhan Gül for being there any time.

I am grateful to my cousins Koray, Kutay, and Nartan Büyükyıldız for their love, friendship, and support.

I would like to thank Barış Mert Pehlivan for his support and care. He believed in me before I believed in myself and before he believed in himself too.

Last but not least, I would like to thank my modern world Rumi - Dr. Ali Bayramoğlu. His wisdom saved my life like many others. He taught me how to live a fulfilled and meaningful life with all my wounds and losses. I will always be grateful for this knowledge of life he gave to me.

TABLE OF CONTENTS

CHAPTERS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CERSAT : The Centre European Remote Sensing d'Archivage et de Traitement
DDS : Dynamically Dimensioned Search Algorithm
EBS : Eastern Black Sea
ECMWF : The European Centre for Medium-Range Weather Forecasts
ERA5 : ECMWF Reanalysis 5th Generation
GFS : The Global Forecast System
GHRSST : The Group for High Resolution Sea Surface Temperature Level 4 Ultra High Resolution
KGE : Kling-Gupta Efficiency
LKSATFAC : Saturated Hydraulic Conductivity Factor
LSM : Land Surface Model
MAE : Mean Absolute Error
MANN : Manning Roughness Scaling Factor
MED : Mediterranean
MODIS : Moderate Resolution Imaging Spectroradiometer
NCAR : National Center for Atmospheric Research
NCEP : National Centers for Environmental Prediction
NOAA : National Oceanic and Atmospheric Administration
NSE : Nash-Sutcliffe Efficiency
NWP : Numerical Weather Prediction
OVROUGHRTFAC : Surface Roughness Coefficient Factor
REFKDT : Infiltration Factor
RETDEPRTFAC : Surface Retention Depth Factor
RMSE : Root Mean Square Error
RR : Correlation Coefficient
RTG_SST_HR : The real-time, global, sea surface temperature
SLOPE : Deep Drainage Coefficient

SST : Sea Surface Temperature

WRF : Weather Research Forecasting

WRF ARW : Advanced Research WRF

WRF-Hydro : WRF Hydrological Modeling System

CHAPTER 1

INTRODUCTION

1.1 Introduction

Extreme weather events result in devastating flood catastrophes with casualties in today's world. As the climate warms, more water vapor is pumped into the atmosphere, increasing the severity of rainfall events. With the changing climate, the impact of flood disasters may be worsened over time (Allen & Ingram, 2002; Hirabayashi et al., 2013; Trenberth, 1999).

As a result, reliable flood forecasting is critical for various operational purposes. The most crucial components of an accurate flood forecasting system are considered as an accurate estimate of the spatial distribution and the intensity of heavy rainfall events and its associated hydrologic response (Ryu et al., 2017; Yucel & Onen, 2014). In this regard, Flood forecasting applications based on a hydro-meteorological modeling framework that integrates the atmosphere and the hydrology are becoming more prevalent (Kunstmann & Stadler, 2005). Selection of the Numerical Weather Prediction (NWP) model and the input datasets driving the boundary and the initial conditions have a profound effect on the accuracy of the short-term predictions; hence, better operational flood forecasts clearly require improved NWP-based forecasts (Done et al., 2004).

The ocean/sea-atmosphere interaction is the starting point that determines the climate distribution of the entire Earth. With this interaction, change in the atmosphere determines the regional climate types and the distribution of these climates on the planet with the effect of the Earth's daily and annual movement, ocean currents, winds, and topography (Bigg et al., 2003). SST is one of the most influential variables in this mentioned interaction. The rays of the sun, which passes the atmosphere and reach a few meters deep in the seas, are stored in the water as heat energy. However, this stored energy is transferred back to the atmosphere by wind, waves, and vertical atmospheric changes (temperature, humidity, momentum) on the sea surface. Especially when sudden temperature increases are seen in the air, the temperature, humidity, momentum flows at the sea-atmosphere interface may become so fast and intense and result in heavy precipitation events around the existing area.

Studies conducted particularly over the coastline with steeply varying topography address the fact that an accurate parameterization of the linkage between the land and atmosphere systems is required to demonstrate changes in water and energy fluxes and states for a more realistic flood prediction. To improve the reliability of such modeling systems, it is necessary to consider factors affecting this linkage. Therefore, the SST representation in the modeling system is regarded as one of the factors having a pronounced effect on water and energy fluxes in the lower-level atmosphere. Therefore, the condition of the SST has a substantial impact on the state of meteorological forcing variables in such NWP models, as the sea acts as a water and energy source for both the atmosphere and the land surface (Lebeaupin et al., 2006; Vecchi & Harrison, 2002). Accordingly, providing high accuracy SST input to the lower boundary is crucial for accurate precipitation modeling, hence for better flood forecasts through hydrological models.

The WRF-Hydro is a distributed, multi-physics hydrometeorological model system created by the United States National Center for Atmospheric Research (NCAR) to address major water challenges, including operational flash flood monitoring. The WRF-Hydro modeling system is capable of running in both offline (one-way between atmosphere-land) and online (two-way) modes, which connects the atmospheric and hydrological processes. According to its overall coupling goal, WRF/WRF-Hydro modeling system is designed to simulate overland flow, subsurface flow, channel routing, and vertical energy fluxes between the land-atmosphere through conceptual and physically-based concepts. Many studies have utilized this modeling system to examine the model's performance and applicability, including flood predictions, water balance, and water management studies across the globe (Kerandi et al., 2018; Li et al., 2017; Naabil et al., 2017; Wehbe et al., 2019). Despite the studies compare the impact of different sources of precipitation input (i.e., comparing observed and simulated) on runoff simulation and agree that further improvement in the precipitation simulation skills is still needed (Givati et al., 2016; Senatoreet al., 2015; Sun et al., 2020), not many studies have investigated the impact of the spatio-temporal resolution of various SST sources over the runoff predictions of WRF-Hydro modeling system via the improvements in the simulated precipitation.

Another factor having an impact on the accuracy of the flood forecast is the proper calibration of the hydrological modeling system. The calibration process is defined as tuning the parameters defined in the model system to capture the real system behavior. Unlike the parameters that can be measured with the in-situ techniques, most of the parameters within the hydrological models conceptually represent the real physical processes that are susceptible to the basin characteristics (Gupta et al., 1998). The complexity of the hydrological models increases with the factors including parameterizations, fundamental calculation unit definition, domain size, resolution, etc. Therefore, the calibration of the model parameters for an accurate representation of the hydrological processes is a salient challenge, especially for physical-based and distributed hydrological models, including WRF-Hydro. In addition, due to the complexity increase, the longer computational time demand of the model becomes another issue.

The organization of this study is as follows: In Chapter 1, the introduction part, the literature review as a background for the study, and the objectives of the study are provided. In Chapter 2, the methodology followed throughout the study, including the study area, dataset, and model descriptions, are, explained. In Chapter 3, precipitation results simulated with the WRF model are given in the first section. Then the WRF-Hydro calibration and the SST analysis are included in the following sections. In Chapter-4, the discussion based on the findings of the study is provided. Lastly, In Chapter-5, the key outcomes of the study are highlighted, and the recommendation for the future path is indicated.

1.2 Literature Review

SST primarily affects the heat and the water fluxes at the lower boundary of the atmosphere; hence there is a significant relationship between the variabilities of SST and convective extremes. A higher SST value generally indicates more moisture in the air and that the lower layers of the atmosphere are warmer. This frequently causes increased precipitation and convection totals (Ferrari et al., 2020). Even the small fluctuations in SST (±1 K) might significantly and nonlinearly alter the intensity of supercell growth throughout the seas (Miglietta et al., 2017). It is also stated that 1 ℃ of SST increase resulted in a 2 mm/d precipitation increase over the monsoon region (Roxy, 2014). Accordingly, enhanced modeling of SST fields in the Weather Research and Forecasting (WRF) system significantly improves the simulation of atmospheric flow dynamics and boundary layer processes (Senatore, Furnari, et al., 2020) most notably precipitation forecasts (Zhang et al., 2015).

Since the precipitation falling on the ground surface is generally driven by lower atmospheric boundary conditions, fluctuations in SST are crucial in determining the heavy precipitation intensity (Baltaci, 2017; Turuncoglu, 2015). Even the steady rise in SST may result in amplification of the precipitation extremes with the convective characteristics over the coastline (Meredith, Maraun, et al., 2015), whereas the rapid and anomalous oscillations in SSTs, may result in mesoscale convective system thunderstorms (Senatore et al., 2014). As a result, reliable precipitation predictions from NWP models may be improved by giving high-accuracy SST input to the lower boundary condition.

Many places in Turkey face severe flood risks because of the country's unique geography, which is bordered by water on three sides and has one of the most complicated terrains anywhere in the world. The Eastern Black Sea (EBS) and the

Mediterranean (MED) regions have been among the most susceptible places in terms of flood risk in Anatolia (Gurer & Ucar, 2009). Using high-resolution NWP models in the EBS and MED regions is important for flood prediction (Camera et al., 2020), where a steady increase in SST may cause abrupt augmentation of the extreme convective sourced events over coastlines (Meredith et al., 2015). Studies performed in the Anatolian Peninsula agreed that the SST fluctuations have a significant effect on observed heavy rainfalls (Turuncoglu, 2015). Bozkurt and Sen (2011) state that increased SST in winter and spring has no impact on precipitation formation in the EBS, while the MED shows a strong response to SST increase during the same seasons. Conversely, the EBS shows strong sensitivity to increased SST compared to the MED in the autumn season. Additionally, the study conducted for flood events that occurred in the EBS region, Arhavi, on 24 August 2015 elucidates that as a consequence of the increased SST, strong moisture convergence resulted in heavy precipitation activity generating the flood event (Baltaci, 2017). SST forcing is often seen as a crucial element in the development of heavy rainfall events, notably over the MED area (Cisneros et al., 2016; Furnari et al., 2018; Rebora et al., 2013; Robinson et al., 2012). Mediterranean SST also has an impact on severe precipitation and flooding in Central Europe, according to Volosciuk et al. (2016).

Despite its importance and possible influence on the accuracy of runoff predictions, there has been a scarcity of studies evaluating alternative SST input datasets and their effect on anticipated runoff accuracy (Chen et al., 2009; McCabe & Wolock, 2008). Studies that utilized high-resolution SST inputs and implemented parameter calibration in the WRF-Hydro modeling system runoff predictions have particularly remained limited with the study of Senatore, Furnari, et al. (2020) and Senatore, Davolio, et al. (2020).

WRF-Hydro model is mainly calibrated using streamflow observations (Liu et al., 2020; Varlas et al., 2019). There are different approaches followed in the literature for the calibration of the WRF-Hydro modeling system. The step-wise calibration approach that minimizes the number of runs to address the computational time issue was suggested by Yucel et al. (2015). This approach was widely used for the WRF-

Hydro calibration in the literature (Duan & Kumar, 2020; Galanaki et al., 2021; Givati et al., 2016; Kerandi et al., 2018; Kilicarslan et al., 2021; Li et al., 2017; Liu et al., 2020; Naabil et al., 2017; Sun et al., 2020; Wang et al., 2020). Additionally, various automated calibration techniques were implemented to find the optimal parameter set such as the PEST algorithm (Fersch et al., 2020; Silver et al., 2017; J. Wang et al., 2019), Stepwise Line Search Algorithm (Kim et al., 2021), and the Dynamically Dimensioned Search (DDS) Algorithm (Abbaszadeh et al., 2020; Cerbelaud et al., 2022; Lahmers et al., 2019; Yin et al., 2020; J. Zhang et al., 2020). In addition, the combination of step-wise and automated calibration approaches is utilized by Senatore et al. (2015). In this study, first, the WRF-Hydro model is calibrated with the step-wise approach to refine the parameters that the model is most sensitive to, then the PEST algorithm is used to calibrate this set of parameters by automated approach.

DDS algorithm is a broadly used optimization algorithm for distributed hydrological models similar to WRF-Hydro with high computational time and resource demand. The major advantage of the DDS algorithm is that the only constraint defined in the algorithm is the iteration number specified by the user (Tolson & Shoemaker, 2007). Due to its motivation to converge to the near-optimal set of the parameter within the user-defined computational time and with fewer iterations compared to the broadly used Shuffled Complex Evolution algorithm (Duan et al., 1992), it is preferred for the physical and distributed models with a high degree of complexity and extended run time (Abbaszadeh et al., 2020; Arsenault et al., 2014; Lahmers et al., 2019; Lespinas et al., 2018; J. Zhang et al., 2020).

1.3 Objectives of the Study

In this study, to expand on the argument, different SST products are utilized in the WRF/WRF-Hydro modeling frameworks to see the impact of the temporal and spatial resolution-wise improved representation of the SST on the simulations of extreme precipitation causing significant floods over the catchments of coastal regions in Turkey. In this context, basins and events from different geographic regions, as EBS and MED represented with humid and semi-arid climates, respectively, are chosen to see the SST impact on the hydrological response.

Another objective of this study is to compare two different calibration methodologies from the literature for the calibration process of the WRF-Hydro in terms of the resultant performance of the model. Accordingly, first, the calibration of the model is manually employed with a step-wise approach suggested by Yucel et al. (2015). Then, the DDS algorithm is selected as an automated optimization technique for the calibration application with the same set of parameters and ranges used in manual step-wise calibration. Depending on the recorded data availability, hourly or daily streamflow observations are used for the calibration of the selected past flood events in the Eastern Black Sea and Mediterranean regions.

CHAPTER 2

DATA & METHOD

2.1 Study Area & Events

In this study, flood events caused by heavy rainfall events that took place in the EBS and MED regions, which have different climatic characteristics, are simulated. Figure 1 depicts 3-km domains (d02) of the WRF model encompassing both regions, as well as chosen basins, associated channel networks and the locations of rain and stream gauge stations.

The mountains that run parallel to the sea in the north-eastern section of Turkey block humid air currents, where the EBS area is located. They reach elevations above 3000 m, creating a complex landscape with steep-sloped features. River systems respond swiftly to moderate rainfall events due to their small basin characteristics and steepy features, resulting in flash floods. The climate of the area has humid characteristics receiving precipitation in all seasons (Turkes, 1996). According to Baltaci (2017), it has the highest annual mean precipitation recorded in Turkey, surpassing 2200 mm. The typical Mediterranean climate is seen over the MED region, with predominant semi-humid features, with a wet winter and spring and a dry summer (Turkes, 1996). Average annual precipitation is recorded as 800 mm at the MED coastlines; however, it may reach up to 1500 mm in the Taurus Mountains (Turkes, 1999). In both regions, the effect of the sea and the complex topographic characteristics indicate that both locations have a significant orography effect reliance and effective energy supplies for convective precipitation. Due to the fact that higher SST affects the water content in the air, it is critical in the development of flash floods in locations with steep terrains, such as coastal areas.

In this study, two significant past flood events are selected for the following modeling practices. On August 24th, 2015, it was reported 32.4 millimeters of hourly precipitation, a total of 135 mm of rain fell during 24 hours over Artvin-Arhavi province in the EBS region. As an alternative event, on December 16th, 2018, the Antalya-Ovacik station reported hourly precipitation of 53.1 mm, while receiving a total of 651.7 millimeters on the same day. This was the heaviest rainstorm ever recorded over entire Turkey (Kaya et al., 2019). This recorded value was around three times greater than Antalya city's monthly average rainfall in December (265.3 mm) (Pilatin et al., 2021). A typical mesoscale convective signal was observed for the event that occurred over the EBS area in the summer, whereas a frontal system was dominating for the event that occurred over the MED region in the winter.

For the model performance evaluation, streamflow data from the EBS stations of D22A049, D22A079, D22A089 (Arhavi province), and D22A147 (Hopa province), as well as the MED stations of D08A071, D09A095, and E08A008 are selected (Figure 1 and Table 1). Each calibration event in Table 1 is run for 10 days to allow the model to warm up. Average streamflow observations (in m^3/s) collected by the State Hydraulic Works (SHW) of Turkey are provided at daily timesteps for five of the stream gauge stations (D22A079, D22A089, and D22A147 over EBS region; D09A095 and E08A008 over MED region) and at hourly timesteps for two of the stations (D22A049 over EBS and D08A071 over MED). However, the hourly streamflow data at these two stations are limited and only available after 2016.

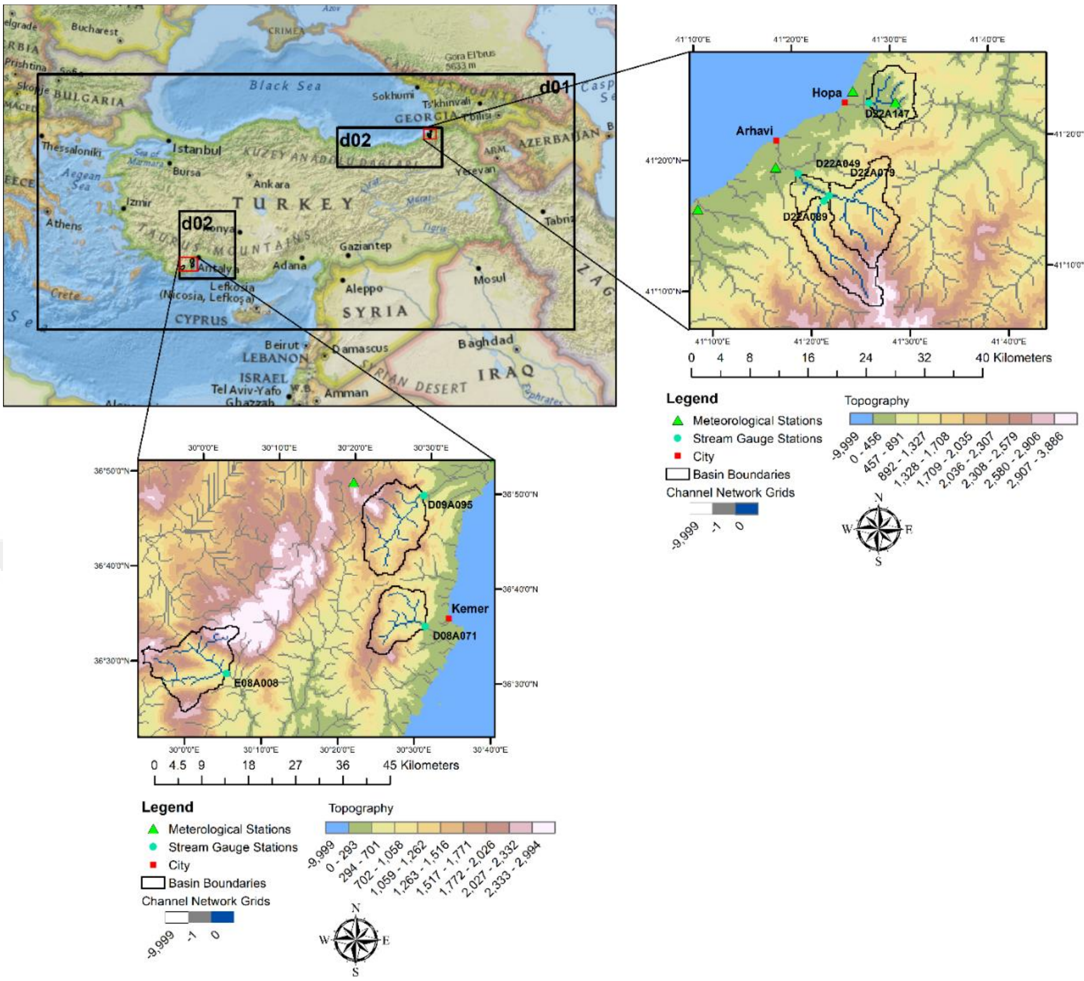


Figure 1 The outer and nested domains (d01 and d02) of the WRF model for the EBS and MED regions are displayed in the top-left. The boundaries of the selected basin, their outlet points (stream gauge stations denoted as blue dots), channel network grids in the WRF-Hydro model, and the meteorological station (denoted as a green triangle) are shown in the zoomed maps with the high-resolution topography layer in the background for the EBS region (top-right) and the MED region (bottom).

Table 1 Drainage areas and calibrated event periods of each selected basin over EBS and MED regions.

Region	Station	Drainage Area (km^2)	Calibration Event Period	
			Start	End
EBS	D22A049	175.8	08/27/2016	09/06/2016
			09/20/2017	09/30/2017
			10/19/2016	10/29/2016
	D22A079	85.8	10/19/2016	10/29/2016
			10/01/2018	01/11/2018
			06/24/2019	07/04/2019
	D22A089	71.5	08/27/2016	09/06/2016
			09/20/2017	09/30/2017
			10/19/2016	10/29/2016
	D22A147	41.9	08/27/2016	09/06/2016
			09/20/2017	09/30/2017
			10/19/2016	10/29/2016
MED	D08A071	98.3	01/09/2015	01/19/2015
			03/07/2017	03/17/2017
			03/23/2015	04/02/2015
	E08A008	164.5	01/09/2015	01/19/2015
			03/07/2017	03/17/2017
			03/23/2015	04/02/2015
	D09A095	164.6	01/21/2014	01/31/2014
			01/09/2015	01/19/2015
			03/23/2015	04/02/2015

2.2 Weather Research and Forecast (WRF) Model

The Advanced Research WRF model version 4.0 (Skamarock et al., 2019), created by National Center for Atmospheric Research (NCAR), is utilized to simulate the forcing data of the WRF-Hydro model for the chosen rainfall events. This forcing dataset includes hourly precipitation, pressure, temperature, wind speed, incoming long- and short-wave radiation, and humidity outputs.

WRF model is a non-hydrostatic medium-scale weather forecasting model that has been recognized as suitable for extreme weather applications with advanced numerical techniques, multi-clustering capability, and multiple physics options. It encompasses a number of interconnected physical approaches that calculate meteorological processes, such as the model in which convective clouds and precipitation are calculated, long- and short-wave radiation calculations to determine atmospheric and surface warming, the boundary layer level where the land-atmosphere interaction is modeled, and the land surface hydrometeorological processes. The microphysical model used for large-scale precipitation calculations has equations that predict air temperature and different hydrometeor classes (snow, rain, ice, and hail). The corresponding bands of the electromagnetic spectrum are used in long and short wavelength radiation calculations. At the lowest level and limit of the WRF model, the Noah–Multi Parameterization (Noah-MP) land surface model (LSM) calculates the soil-plant-atmosphere relationships between the land and the atmosphere.

A two-way nesting model setup with spatial resolutions of 9-km for the outer domain (d01) and 3-km for the inner domain (d02) is implemented in this study, as illustrated in Figure 1. Kilicarslan et al., (2021) provides comprehensive information on the domain extents for both regions.

The sensitivity study performed by Duzenli et al. (2020) shows that the GFS initial and boundary dataset (NOAA, 2015) was the most representative for the MED region, while the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 (ECMWF, 2020) provided the most representative dataset for the EBS region for the WRF model. Therefore, these two datasets are also utilized in the current study in both regions.

2.2.1 Sea Surface Temperature Field Update in WRF Model

Along with the time-invariant SSTs (ERA5 and GFS), three additional daily-updated SST datasets are employed in this sensitivity analysis: 1) Medspiration Ultra-High-Resolution Foundation Sea Surface Temperature (CERSAT, 2012); 2) The Group for High-Resolution Sea Surface Temperature Level 4 Ultra-High Resolution (GHRSST) (Team GHRSST, 2010a, 2010b); 3) Real-Time, Global, Sea Surface Temperature (RTG_SST_HR) represented by the National Centers for Environmental Prediction (NCEP), National Oceanic and Atmospheric Administration (NOAA) (NCEP & NOAA, 2014). Medspiration, GHRSST, and NCEP SST datasets are available in daily time step and have high spatial resolutions of 0.022°, 0.01°, and 0.083°, respectively. Medspiration, GHRSST, NCEP, ERA5, and GFS will be used to refer to the simulations performed with different SST products utilized in this work henceforth. Table 2 depicts the simulation dates for the WRF model utilizing these SST products for each area of study.

Table 2 Forcings used in the SST simulations and their run periods.

Region	Forcings		Run Periods	
	SST Products	Initial and Boundary Conditions	Start	End
EBS	ERA5 GHR Medspiration NCEP	ERA5 Reanalysis	08/17/2015	08/27/2015
MED	GFS GHR Medspiration NCEP	GFS Forecast	12/10/2018	12/20/2018

Given the absence of buoy measurements across the study regions to verify the accuracy of SST datasets, it is proposed to utilize the average cross-correlations as an indication of true signal (Yilmaz et al., 2014), presuming no other shared spatial and temporal signal exists between the SST datasets (i.e., higher average cross-correlations means a better product).

Additionally, to examine the precipitation simulated by various SST products, the root mean squared error (RMSE) and the mean absolute error (MAE) of the precipitation for each SST simulation are determined compared to the point-based precipitation records.

2.3 WRF-Hydro Model

The integrated hydrological model system named WRF-Hydro (version 5.1.1) is operated in this study (Gochis et al., 2020). The WRF-Hydro hydrological modeling system is an enhanced version of the above-mentioned WRF model with the traditional 1-dimensional Noah-MP LSM that incorporates overland, saturated subsurface, channel, and groundwater flow into a modeling structure. The LSM in the WRF-Hydro model is the same as the one-dimensional LSM (Noah-MP) operated in the WRF model. Accordingly, the hydrological modeling system is operated over the nested domain of WRF (d02) with a 3 km resolution. There are two options available in the modeling system configuration for the feedback mechanisms between WRF and WRF-Hydro model as uncoupled and coupled modes (Figure 2). There is only one-way feedback from the atmosphere to land in uncoupled mode. The meteorological inputs for this mode can be created by the WRF model or other sources such as radar and satellite precipitation can be used. A two-feedback mechanism is created between the atmospheric and the hydrological model for the coupled mode. Energy and moisture fluxes are generated from the WRF-Hydro model to the WRF model. In this case, surface hydrological processes contribute to the formation of the precipitation mechanisms. This linkage between WRF and WRF-Hydro models is established through the LSM.

WRF-Hydro model disaggregates the LSM grids into high-resolution routing grids after the moisture states are computed for the land surface column. In this study, LSM grid resolution is defined as 3 km (same as the WRF) and disaggregated into routing grids with a 250 m grid size in both regions.

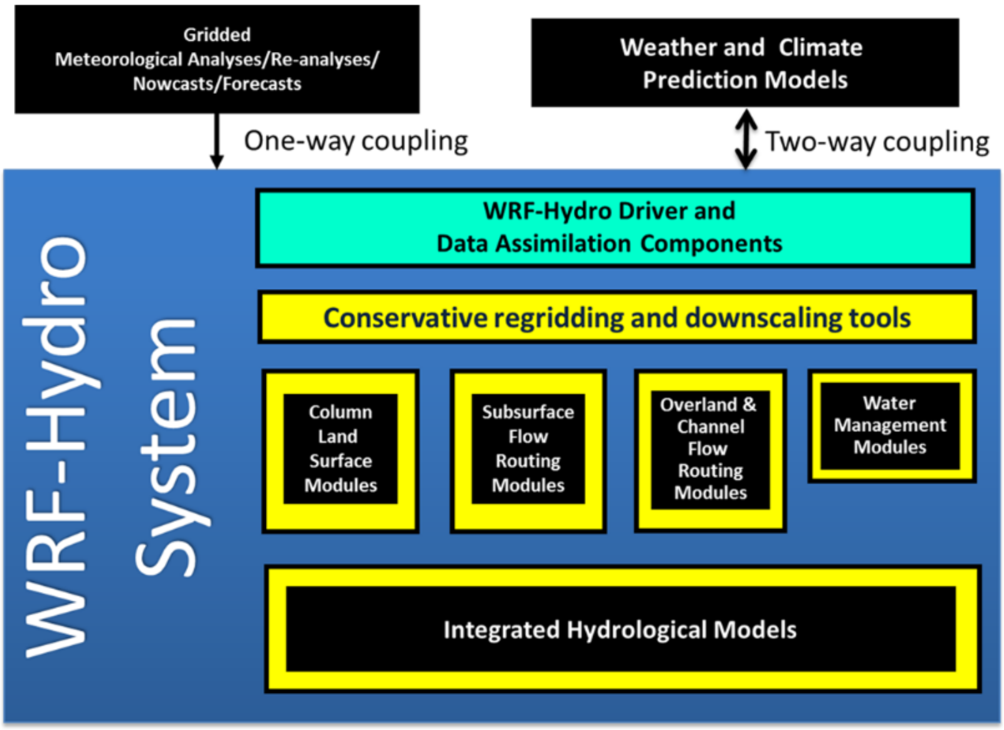


Figure 2 Integrated hydrological modeling system structure of WRF-Hydro model Gochis et al., 2020).

The system also has a simple reservoir or lake offset algorithm. Soil moisture, soil temperature, surface temperature, snow water equivalent, intercepted water amount, energy and moisture fluxes on the surface are calculated with LSM (Figure 3). Surface runoff is described in the LSM as excess filtration (net precipitation), whereas subsurface runoff is defined as flow within the soil columns. These variables are calculated by instantaneously solving the energy and water balance equations for the 1-dimensional soil/plant column. In the model, the surface soil structure consists of four different soil layers of 10, 30, 60, and 100 cm, respectively. It represents the soil characteristics in 16 different categories, obtained from the 1-km STATSGO database. Land cover classes are generated from the Moderate Resolution Imaging Spectroradiometer (MODIS) Modified IGBP 20-category land cover dataset. Overland and subsurface routing schemes for the entire domains are activated, while the channel routing scheme is only turned on over the defined basins. For the conceptual groundwater module, the pass-through option for the bucket model is activated.

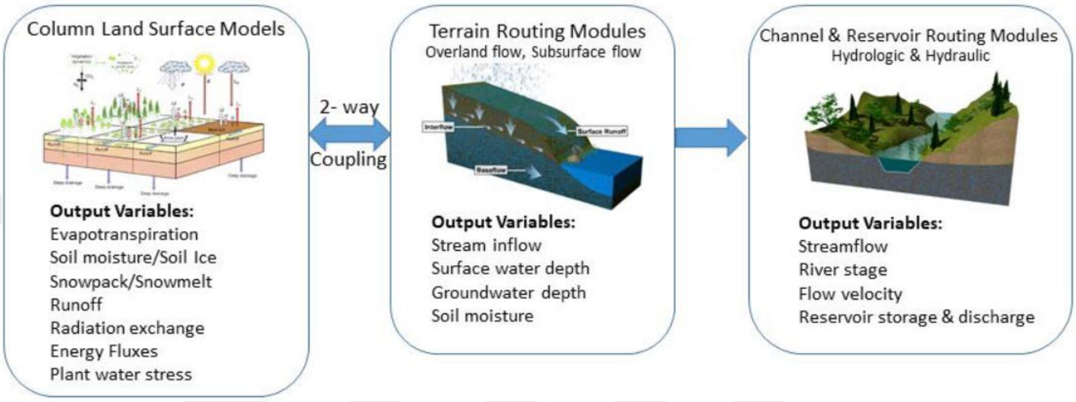


Figure 3 WRF-Hydro Physics components and output variables list for each module Gochis et al., 2020).

By using the WRF-Hydro Preprocessing tools of (Sampson & Gochis, 2015) in an ArcGIS environment, a high-resolution routing stack including flow accumulation, flow direction, channel definition, etc. is produced with the hydrologically conditioned digital elevation model (DEM) called HydroSHED (Lehner, Verdin, & Jarvis, 2008).

2.3.1 Calibration of the WRF-Hydro Model

WRF-Hydro calibration runs are performed with the WRF meteorological forcings updated with the observed precipitation. Using the hourly point observation precipitation data from the Turkish State Meteorological Service (MGM), the spatial distribution of the observed precipitation is produced with the Inverse Distance Weighting method, and map layers are created in netCDF format. The precipitation field in the WRF model outputs is replaced with the hourly precipitation layers. NetCDF commands (NCO) are utilized during the course of this process. The leave-one-out method is used to evaluate the interpolated precipitation. In this method, one rain gauge is removed from the interpolation at each step, and the interpolated value is calculated over the position of the removed rain gauge. For each rain gauge, the same procedure is performed. Interpolated precipitation data are compared to rain gauge point measurements by calculating the RMSE and MAE statistical metrics.

Calibration of the WRF-Hydro is carried out for three events in each basin (seven basins, Table 1) at hourly or daily time steps, depending on the availability of streamflow data. The validation of the calibrated parameters sets is performed with the SST simulations. In the D22A049 basin of the EBS region and D08A071 basin of the MED region, hourly streamflow data for two heavy precipitation events in the 2016 and 2017 hydrologic years are used to calibrate the model while other events are calibrated on a daily basis (Table 1). For the validation process (with SST events), hourly streamflow data is only available for the MED region, whereas the daily streamflow is used for the EBS region.

For the accuracy evaluation of the model, statistical metrics such as bias, root mean square error (RMSE), correlation coefficient (RR), Nash-Sutcliffe Efficiency (NSE), and Kling-Gupta Efficiency (KGE) are calculated between observed and simulated runoff to find the optimal model parameter sets. Bias quantifies the degree to which hydrograph volume is underestimated or overestimated compared to the observation hydrograph. RR represents the linearity between observed and simulated runoff and determines the hydrograph performance in terms of timing and shape. Likewise, KGE, NSE, and RMSE are the metrics that are sensitive to the volume and the distribution of the hydrograph, and they indicate the hydrograph's overall fit (Gupta, Kling, Yilmaz, & Martinez, 2009). This statistical analysis is conducted using hourly or daily time steps, depending on the temporal resolution of the observed streamflow data.

The calibrated parameter set of the WRF-Hydro model is then validated for SST simulations over both regions. WRF-Hydro model is forced by meteorological forcings constructed using the ERA5 and GFS SSTs, as well as three additional daily-updated SST datasets, namely Medspiration, GHRSST, and NCEP. Additionally, similar SST events are modeled by substituting observed precipitation in WRF precipitation.

2.3.1.1 Manual Step-wise Calibration

First, the model is calibrated manually using a step-wise approach allowing the model to first simulate the water balance in the basin and then distribute the amount of water accurately over time (Yucel et al., 2015). At each step, a common parameter value is determined for all three events, and then it is passed to the calibration of the following parameter. This was done based on the visual scanning of hydrographs and the calculated statistics.

Depending on the hydrological function of the parameters in the model, they are divided into two groups as the ones responsible from the hydrograph volume and the shape. At the initial stage, the parameters that determine the hydrograph volume, including the infiltration factor (REFKDT), the surface retention depth factor (RETDEPRTFAC), and the deep drainage coefficient (SLOPE), are calibrated. Then, at the second stage, the surface roughness coefficient factor (OVROUGHRTFAC), the channel Manning roughness scaling factor (MANN), and the saturated hydraulic conductivity factor (LKSATFAC) controlling the time of the peak and the temporal distribution (shape of the hydrograph) are calibrated.

Certain parameters (REFKDT, SLOPE, and MANN) are classified as the global parameters, defined as a fixed value for the whole WRF-Hydro domain and specified in tabular value format. Other parameters are classified as pixel-based parameters (RETDEPRTFAC, OVROUGHRT, LKSATFAC), allowing parameter values to be changed for each pixel throughout the model domain. In this study, those pixel-based parameters are tunned over the defined basin boundaries.

First, the calibration of the infiltration factor (REFKDT) is performed. The REFKDT significantly affects the infiltration of the precipitation through the soil column;

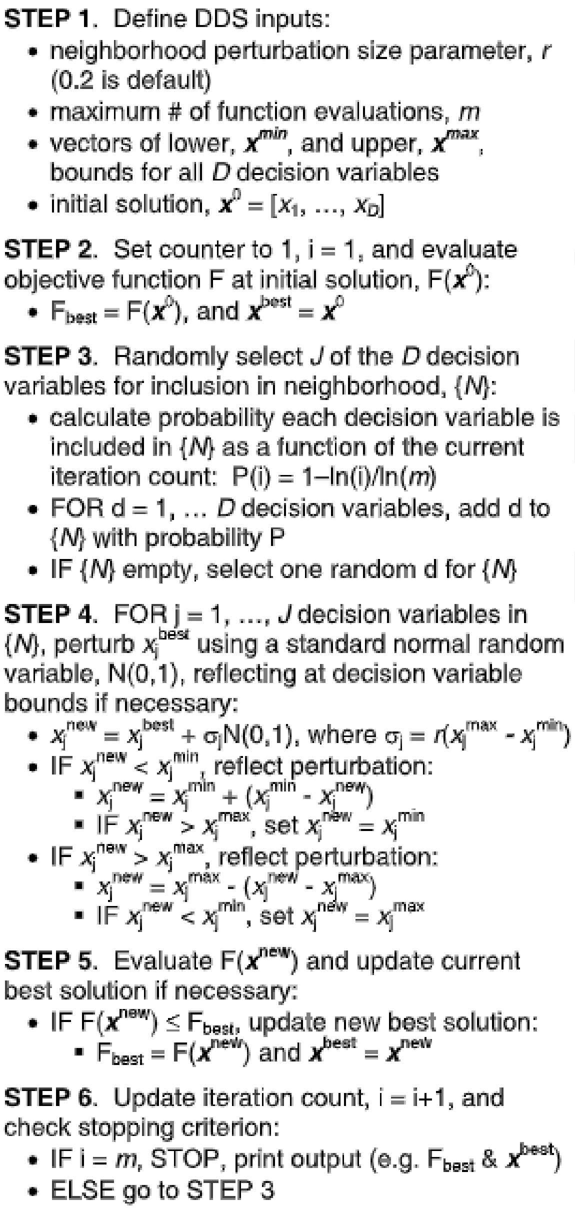
therefore, it plays a decisive role in distributing the total surface runoff into the surface and subsurface runoff (Schaake et al., 1996). The soil absorbs the precipitation until it reaches to saturation level. The parameter value is defined in a table and assigned as constant for the whole model domain. The default REFKDT value is defined as 3.0 in the model, while the value range is between 0.5 and 5.0. Secondly, another parameter that controls the total runoff volume named as the surface retention depth parameter (REDEPRTFAC) is calibrated. After the saturation level is exceeded, rest of the precipitation falling on the ground starts to build up on the surface. When the ponded water (extra-filtration) exceeds the predetermined water retention depth, it is diverted as the surface runoff. The REDEPRTFAC, on the other hand, is defined in the model as the multiplier of this water retention depth value, which is assumed to be 1 mm over the whole domain. While the default REDEPRTFAC value in the model is 1.0, the value range determined for calibration is 0.0–10 (in increments of 0.1). The last parameter controlling the runoff volume included in the calibration is the deep drainage coefficient (SLOPE). This parameter determines the rate of the water passage between the soil column and the conceptual bucket model developed to represent groundwater within the model. Its default value is defined as 0.1, and the calibration process is performed in increments of 0.3 in the range of 0.1-1.0.

At the last step of the manual step-wise calibration, the parameters controlling temporal distribution of the runoff volume are calibrated. The surface roughness parameter (OVROUGHRT) impacts the speed of the infiltration excess water transmitted over the surface until it reaches the channel network grids. The surface roughness values are defined in the model according to the MODIS-20 land use categories. Calibration of this parameter was performed in increments of 0.3 in the range of 0.1-1.0. Second, calibration of the tabular Manning`s roughness coefficients for all stream orders is performed with a scaling factor (MANN) with 0.5 increments in the 0.5-2.0 value range. MANN parameter is responsible from the time of flow conveyance throughout the channel grids. The higher the MANN value, the longer it takes for the water to reach the basin outlet, resulting in a delay of the peak flow value at the basin outlet. Finally, the calibration of the saturated hydraulic conductivity parameter (LKSATFAC), which controls the redistribution of the infiltrated water in the lateral plane, is performed. The default value of this parameter is defined as 1000 in the model. Calibrations is performed for the values of 10, 100, 1000, and 10000.

2.3.1.2 Automated Calibration

Besides stepwise calibration, the automated calibration method is also implemented in the scope of this study to take into account many distinct combinations of values between the parameters. Since the WRF-Hydro model is a complex and grid-based model, the simulation time is longer compared to other hydrological models in the literature. Therefore, one of the concerns in this study is capturing the optimal model parameter set in the possible shortest time. In this regard, the Dynamically Dimensioned Search (DDS) optimization algorithm, commonly preferred for complex hydrological models and preferred by NCAR for WRF-Hydro model calibration, is selected for the model calibration (Tolson & Shoemaker, 2007). The main factor in deciding on the DDS algorithm is that it can give better results with a limited number of iterations compared to other algorithms (Zhang et al., 2020). The pseudo-code of the DDS algorithm is given in Figure 4. As an objective function, KGE is calculated and optimized between the simulated and observed streamflow in each iteration.

As a summary of the algorithm given above: The only stopping criterion in the DDS algorithm is the number of iterations specified by the user. The algorithm starts with a global search, then adjusts this search size dynamically, reducing it to a more local size as the iteration sequence gets closer to the end. This transition is achieved by reducing the number of elements in the parameter set both dynamically and probabilistically. In this way, it aims to find the best possible global result according to the maximum number of iterations determined by the user depending on the available time and computational resources. It is worth noting that the purpose of this algorithm is not to achieve the global optimum result (Tolson & Shoemaker, 2007).



```
STEP 1. Define DDS inputs:
  • neighborhood perturbation size parameter, $r$ (0.2 is default)
  • maximum # of function evaluations, $m$
  • vectors of lower, $\boldsymbol{x}^{min}$, and upper, $\boldsymbol{x}^{max}$, bounds for all $D$ decision variables
  • initial solution, $\boldsymbol{x}^0 = [x_1, \ldots, x_D]$
STEP 2. Set counter to 1, i = 1, and evaluate objective function F at initial solution, $F(\boldsymbol{x}^0)$:
  • $F_{best} = F(\boldsymbol{x}^0)$, and $\boldsymbol{x}^{best} = \boldsymbol{x}^0$
STEP 3. Randomly select $J$ of the $D$ decision variables for inclusion in neighborhood, {$N$}:
  • calculate probability each decision variable is included in {$N$} as a function of the current iteration count:  $P(i) = 1 - \ln(i)/\ln(m)$
  • FOR d = 1, ... $D$ decision variables, add d to {$N$} with probability P
  • IF {$N$} empty, select one random d for {$N$}
STEP 4. FOR j = 1, ..., $J$ decision variables in {$N$}, perturb $x_j^{best}$ using a standard normal random variable, N(0,1), reflecting at decision variable bounds if necessary:
  • $x_j^{new} = x_j^{best} + \sigma_j N(0,1)$, where $\sigma_j = r(x_j^{max} - x_j^{min})$
  • IF $x_j^{new} < x_j^{min}$, reflect perturbation:
      ▪ $x_j^{new} = x_j^{min} + (x_j^{min} - x_j^{new})$
      ▪ IF $x_j^{new} > x_j^{max}$, set $x_j^{new} = x_j^{min}$
  • IF $x_j^{new} > x_j^{max}$, reflect perturbation:
      ▪ $x_j^{new} = x_j^{max} - (x_j^{new} - x_j^{max})$
      ▪ IF $x_j^{new} < x_j^{min}$, set $x_j^{new} = x_j^{max}$
STEP 5. Evaluate $F(\boldsymbol{x}^{new})$ and update current best solution if necessary:
  • IF $F(\boldsymbol{x}^{new}) \leq F_{best}$, update new best solution:
      ▪ $F_{best} = F(\boldsymbol{x}^{new})$ and $\boldsymbol{x}^{best} = \boldsymbol{x}^{new}$
STEP 6. Update iteration count, i = i+1, and check stopping criterion:
  • IF i = $m$, STOP, print output (e.g. $F_{best}$ & $\boldsymbol{x}^{best}$)
  • ELSE go to STEP 3
```

Figure 4 Pseudocode for DDS algorithm (Tolson & Shoemaker, 2007).

The same parameter ranges determined for the incremental calibration are also defined in the DDS algorithm. For cases with a similar number of calibrated parameters set (with six-parameters), the recommended iteration number for the algorithm to find the optimum set is noted as 200 (Tolson & Shoemaker, 2007). The number of model iterations performed in previously performed stepwise calibration method is approximately 60. In order to compare the two different calibration methods in terms of the number of iterations, the DDS algorithm was started with 60 iterations, and this number is increased to 100, 150, and 200. The automated calibration is performed for one event for the D22A049 basin from the EBS region and the D08A071 basin from the MED region.

CHAPTER 3

RESULTS

3.1 WRF Model Results

3.1.1 Evaluation of SST Datasets

In this study, two types of SST products are used: first, SST products (ERA5 and GFS) with coarse spatial (0.25°) resolution and they remained constant throughout the run time; second, products with higher spatial (0.022° for Medspiration, 0.01° for GHRSST, and 0.083° for NCEP) and temporal resolution (daily updated). The spatial distribution of each SST product during the course of the 10-day run is represented in Figure 5, while the temporal change of the spatially averaged SST values for each product is shown in Figure 6. All products represent higher temperatures (~300 K) over the western part of the EBS compared to the eastern part of the sea (~297 K), except the NCEP SST, which shows an even distribution of temperatures (~305 K) over the sea surface. Temperature variability between the products is greater over MED compared to EBS region (Figure 5). However, the temporal distribution of the SST products is consistent with each other, except for ERA5 and GFS, which have constant temperature values throughout the run time (Figure 6).

The cross-correlation between each product and others is calculated and averaged. Results show that average temporal cross-correlations over the MED (0.79-0.84) are greater than EBS (0.24-0.60), whereas the spatial average cross-correlations over the EBS (0.48-0.83) are greater than the MED (0.11-0.39). Among the high-resolution products, GHRSST (0.61) has the highest spatiotemporal cross-correlation among other products. Medspiration (0.54) and NCEP (0.36) follow the GHRSST product.

This order is the same for both regions. Overall, the results show that the GHRSST is the best while the NCEP is the least performing SST products for the events and regions examined in this study.

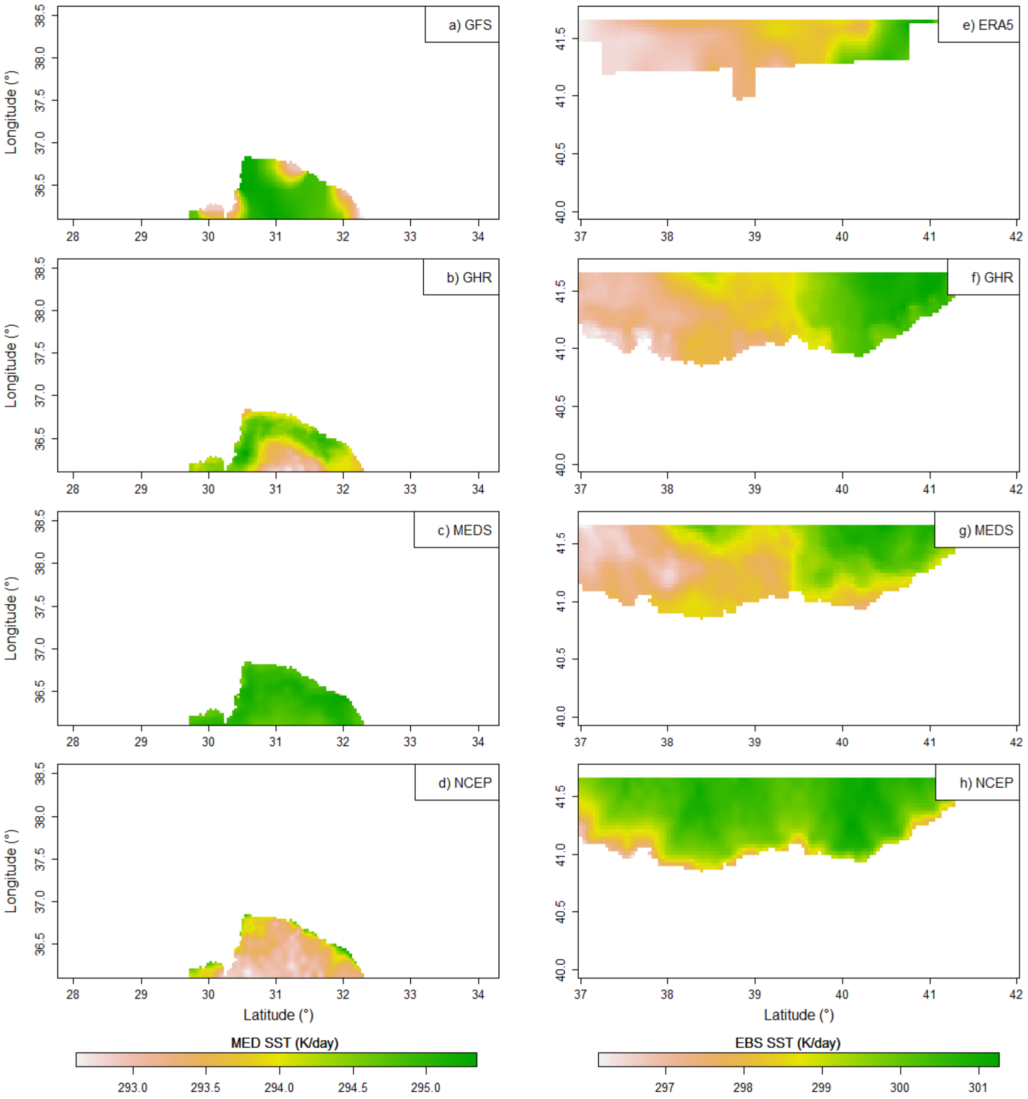


Figure 5 Temporally averaged spatial distribution of SST products of ERA5/GFS, GHRSST, Medspiration, and NCEP: (a)–(d) across MED and (e)–(h) over EBS regions.

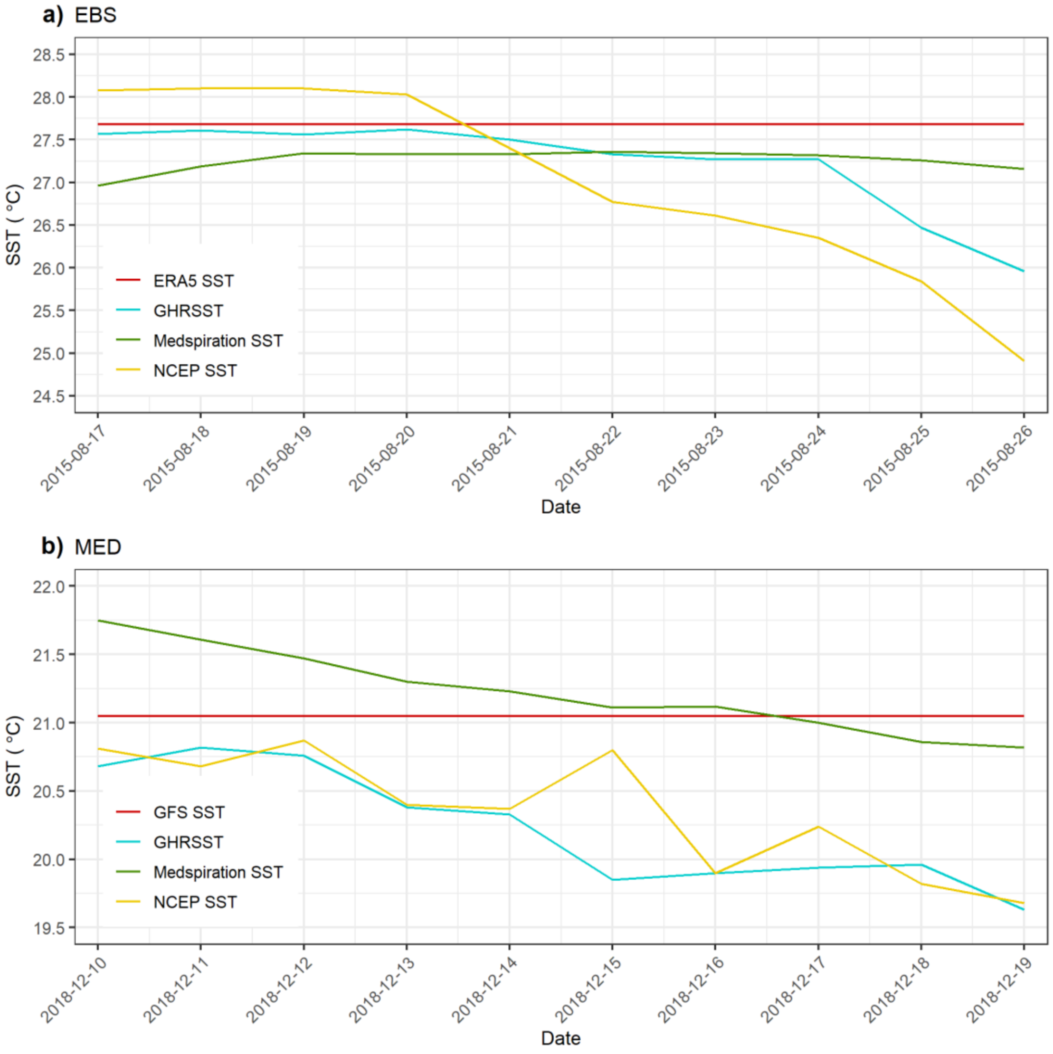


Figure 6 Spatially-averaged temporal distribution of SST products of ERA5/GFS, GHRSST, Medspiration, and NCEP: (a) across EBS and (b) MED regions.

3.1.2 Evaluation of Precipitation Simulations for SST Events

The mean of RMSE and MAE statistics are calculated between the interpolated (produced by IDW method), simulated precipitations with different SST products, and the available point observations over the EBS and MED regions (Table 3). Both statistics of the interpolated precipitation for EBS and MED are better compared to the simulated precipitations. Among the high-resolution SST simulations, GHRSST precipitation has the lowest MAE (~0.81) and RMSE (~1.89) values in the MED region. The RMSE values of the simulated precipitations are greater than the interpolated precipitation over the MED region by 41%, whereas it is %29 greater than that of the EBS region. Overall, both the RMSE and MAE values indicate that the precipitation produced by the interpolation is more accurate compared to the simulated ones with the different SST products. Therefore the uncertainty added to the precipitation input of WRF-Hydro is higher for the WRF model compared to the IDW interpolation technique.

Table 3 Comparison of mean absolute error (MAE) and root mean square error (RMSE) values calculated for precipitation observations from rain gauges, interpolated and modeled (with ERA5/GFS, GHR, MED, and NCEP SSTs) precipitation data extracted from the same rain gauge locations.

Region	Precipitation source	MAE	RMSE
EBS	Interpolated Observation	0.240	0.954
	ERA5-SST	0.334	1.297
	GHR-SST	0.347	1.369
	MED-SST	0.341	1.370
	NCEP-SST	0.327	1.302
MED	Interpolated Observation	0.466	1.165
	GFS-SST	0.868	2.129
	GHR-SST	0.808	1.885
	MED-SST	0.825	1.943
	NCEP-SST	0.812	1.901

Time series of the basin averaged simulated and observed precipitation for the SST events over D22A147 and D08A071 basins are given in Figure 7 (a) and (b), respectively, while the statistical measures obtained for each SST event in both basins are shown in Table 4. Figure 7 (a) shows the precipitation time series from 08/17/2015 00:00:00 UTC to 08/27/2015 00:00:00 UTC, for 241-hours. The highest precipitation at D22A147 is recorded at the 178th hour, or 08/24/2015 09:00:00 UTC as 26.3 mm of rain. However, the highest precipitation quantity over the whole EBS region is reported as 32.4 mm at 08/24/2015 00:00:00 UTC within the boundary of D22A049. In spite of this, as shown in Figure 7 (a), the D22A147 basin-average precipitation is calculated as 16.1 mm during the 169th hour, which coincides with the event peak timing over the EBS region. Also, it can be interpreted that the simulations with alternative SST datasets are able to approximate the broad pattern, except the primary peaks are generated a few hours earlier than observed peak However, notwithstanding the poor statistical measures (correlation of 0.01-0.03 and large RMSE of 3.19-5.30 in Table 4), it is clear that incorporating a high-resolution SST product increases the simulation accuracy corresponding to the actual precipitation. This result is valid, particularly for the Medspiration simulation. In addition, it is seen that the GHRSST simulation underestimates the recorded peak precipitation. Besides, GHRSST simulation overestimates the observed peak precipitation. Other simulated peaks are lower than the GHRSST simulation, but they are closer to the observed peak. Medspiration surpasses the response in all calculated statistics compared to the observed precipitation (Table 4).

The time series of the basin-averaged precipitation for the event between 12/10/2018 00:00:00 UTC-12/20/2018 00:00:00 UTC (241-hours) are shown in Figure 7 (b). Peak precipitation is recorded as 53.1 mm during the 162nd hour, corresponding to 12/10/2018 at 17:00:00 UTC. At the same time step, the peak basin-average precipitation for D08A071 is recorded as 15.7 mm. On the whole, all simulated precipitations reveal an almost exact match to the observed patterns, with modest overestimations. On the other hand, high-resolution SST simulations seem to be able to enhance the volume with decreased bias in line with the observation. Simulations performed with GFS SST, GHRSST, and NCEP SST all have slight delays in their peak timing (about 1-2 hours). However, Medspiration captures the precise peak moment. GFS SST has the biggest overestimation, ~ 17 mm compared to the peak observation, and a positive bias (0.56 m3/s) considering the whole model run period (Table 4).

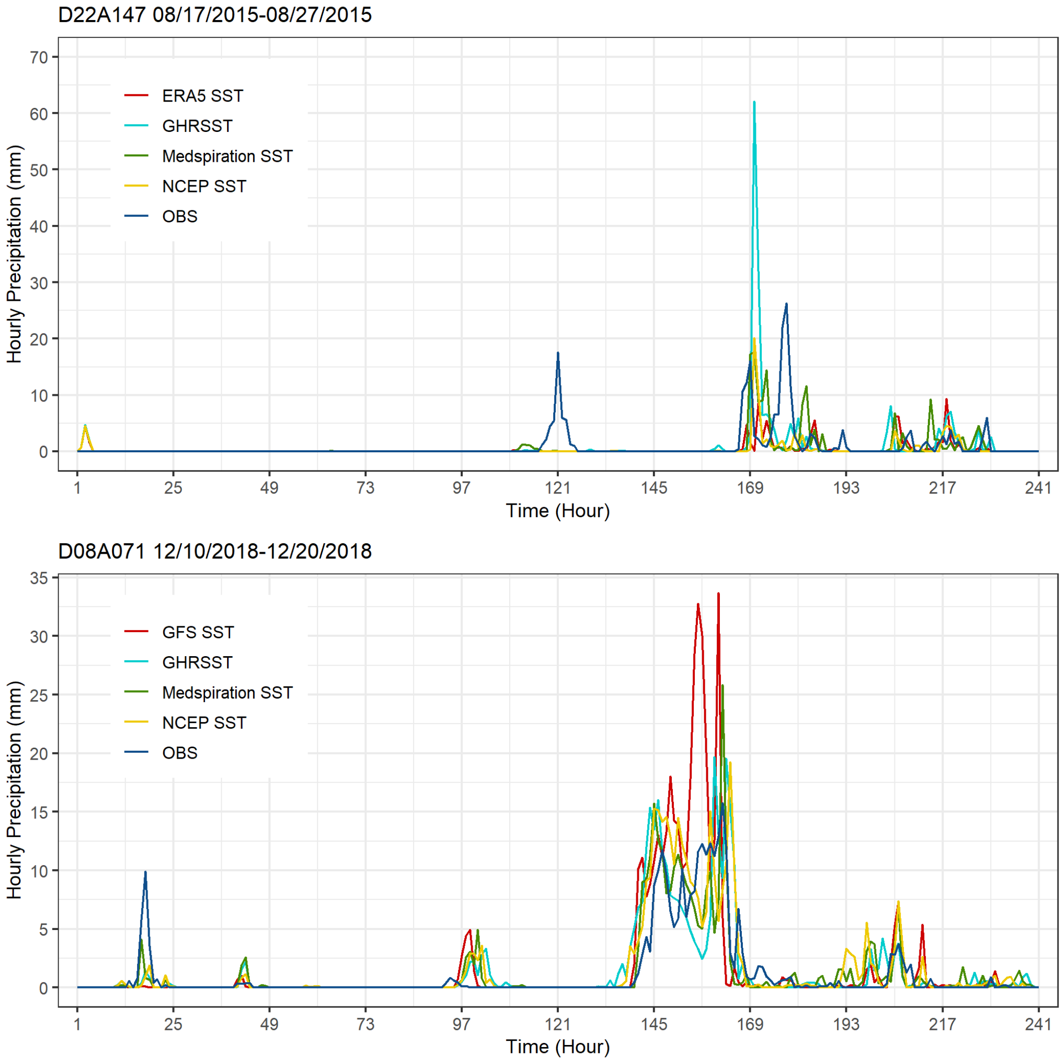


Figure 7 Time series of hourly precipitation that a) D22A147 basin over EBS region receives during the event occurred in 08/17/2015-08/27/2015 and b) D08A071 basin over MED region receives during the event occurred in 12/10/2018-12/20/2018 for ten days. Outputs are generated from the WRF model with the native SST field from ERA5 Reanalysis data (ERA5-SST) for the EBS region and GFS Forecast data (GFS-SST) for the MED region different SST products: GHRSST, Medspiration, and NCEP.

Table 4 Statistics of Bias, RMSE, and RR between observed and modeled precipitations with different SST datasets of ERA5, GHRSST, Medspiration, and NCEP for D22A147 over EBS and GFS, GHRSST, Medspiration, and NCEP for D08A071 over MED regions. Statistics calculated between the WRF-Hydro model calibration and validations outputs using the corresponding precipitation dataset as input and the observation hydrograph (Bias, RMSE, RR, NSE, and KGE).

Station			WRF Precipitation			WRF-Hydro Streamflow				
			Bias	RMSE	RR	Bias	RMSE	RR	NSE	KGE
D22A147 (Hopa)	Calibration	Observed Precip.	-	-	-	0.47	2.22	0.75	0.72	0.61
	Validation	Observed Precip.	-	-	-	-7.11	14.25	0.83	0.37	0.063
		ERA5-SST	-0.54	3.19	0.03	-9.92	20.13	0.42	-0.26	-0.39
		GHR-SST	-0.06	5.30	0.01	-6.16	10.83	0.83	0.63	0.30
		MED-SST	-0.24	3.55	0.03	-8.32	15.86	0.86	0.22	-0.08
		NCEP-SST	-0.54	3.38	0.01	-9.60	18.98	0.82	-0.12	-0.29
D08A071 (Kemer)	Calibration	Observed Precip.	-	-	-	-5.57	26.30	0.46	-1.26	-0.75
	Validation	Observed Precip.	-	-	-	-57.48	84.53	0.40	-0.13	0.11
		GFS-SST	0.56	3.45	0.60	-24.98	128.81	0.18	-1.62	0.02
		GHR-SST	0.18	2.35	0.52	-42.73	83.25	0.30	-0.09	0.31
		MED-SST	0.13	1.86	0.67	-43.63	83.57	0.31	-0.10	0.30
		NCEP-SST	0.33	2.23	0.60	-40.76	81.50	0.32	-0.05	0.33

Figure 8 depicts the peak day (08/24/2015) point observations of precipitation spatially distributed with the IDW method and the precipitation estimates from different SST simulations (ERA5, GHRSST, Medspiration, and NCEP) with WRF model for the SST event in the EBS region. It is worthy of note that the GHRSST simulation (Figure 8 (c)) overestimates the observed precipitation (Figure 8 (a)) over the D22A147 basin. In terms of the spatial distribution of the precipitation. Medspiration produces the closest pattern compared to the observation (Figure 8 (d)). This result is consistent with the temporal precipitation distribution results that the Medspiration products enhance the accuracy of the precipitation more compared to the coarse resolution ERA5 SST product. Furthermore, Medspiration and GHRSST simulations overestimate the precipitation towards the coastline, where they produce more than 140 mm of daily precipitation (Figure 8 (c and d)). On the other hand,

Figure 8 (e) shows that the NCEP SST simulation causes a significant underestimation of the precipitation over the selected basins. Figure 8 (c) demonstrates that the GHRSST is more accurate than other simulations in terms of reproducing the actual location of the SST event.

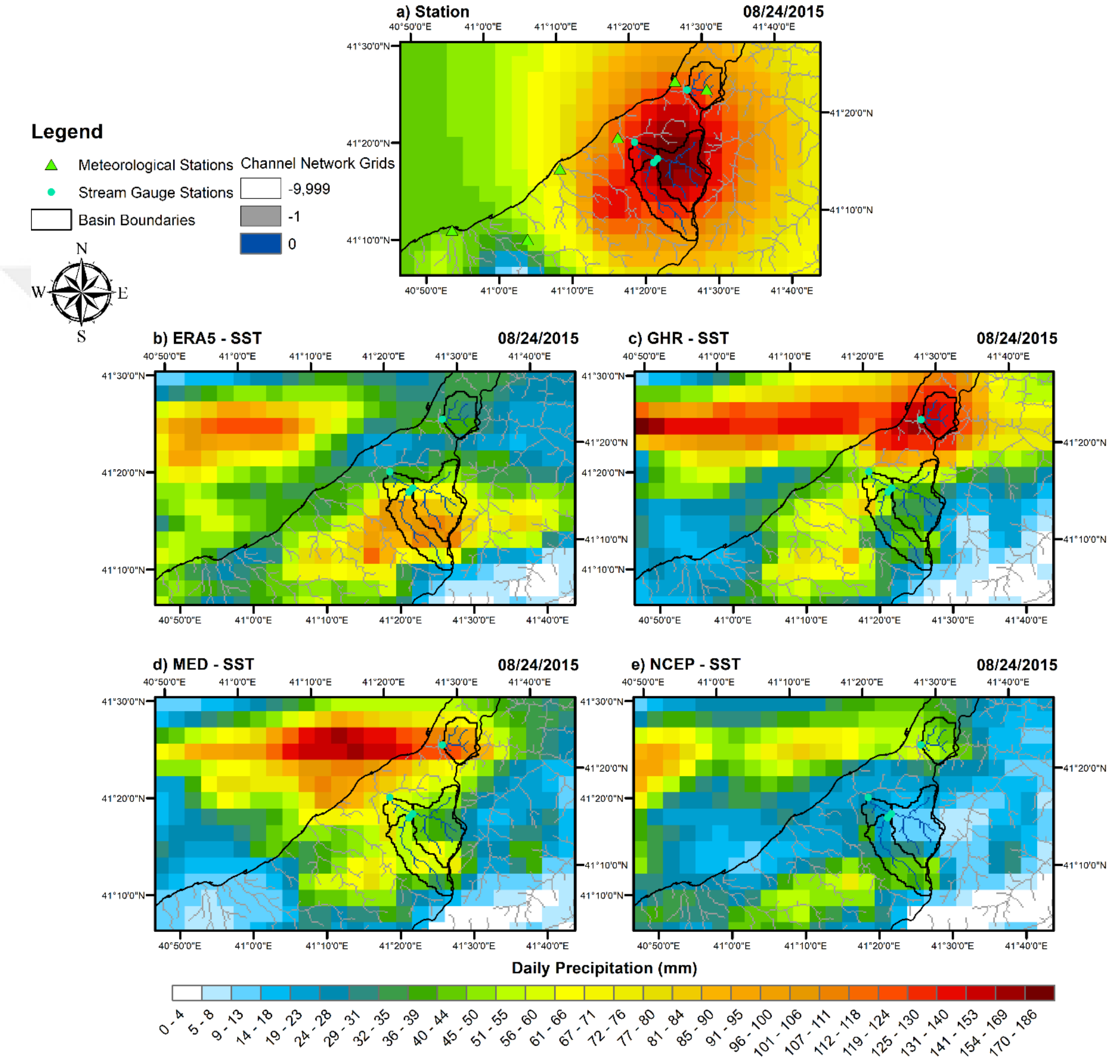


Figure 8 Spatial distribution of daily precipitation on the peak day (08/24/2015) for the run period of 08/17/2017 – 08/27/2017 over the EBS region. (a) The map at the top shows the interpolated observed precipitation map obtained from meteorological station data (green triangles). The black line indicates the boundaries of the selected basins for this study, while the blue dots show the corresponding stream gauge

stations. The four maps in the sub-panels refer to the simulated precipitations by the WRF model derived from different SST data sources for the peak hour: (b) ERA5, (c) GHRSST, (d) Medspiration, and (e) NCEP, respectively.

In Figure 9, the peak hour (12/16/2018 17:00:00) spatial distribution of the observed and simulated precipitations with different SST products (GFS, GHRSST, Medspiration, and NCEP) are illustrated for the MED region with the selected basins. Figure 9 (b) shows that the GFS simulation overestimates the actual event peak value (53.1 mm) and its location. Instead of creating the peak precipitation over the coastline, it shifts the events center through the sea. Compared to the GFS simulation, high-resolution SST simulations generate better results regarding the overall distribution patterns of the precipitation and capture observed depth. GHRSST simulation is successful in capturing the precipitation extent. However, it misses the observed precipitation depth with the amount of 16-18 mm. This remark may be owed to the previously mentioned delay of the peak time in the GHRSST simulation. NCEP and Medspiration simulations represent much closer precipitation depths during the peak time. In particular, Medspiration catches the event center and the observed precipitation depth (Figure 7 (b)) but creates an overestimation of ~8 mm depth corresponding to the peak precipitation (25.8 mm) denoted as a dark-orange pattern over the D22A071 basin (Figure 9 (d)).

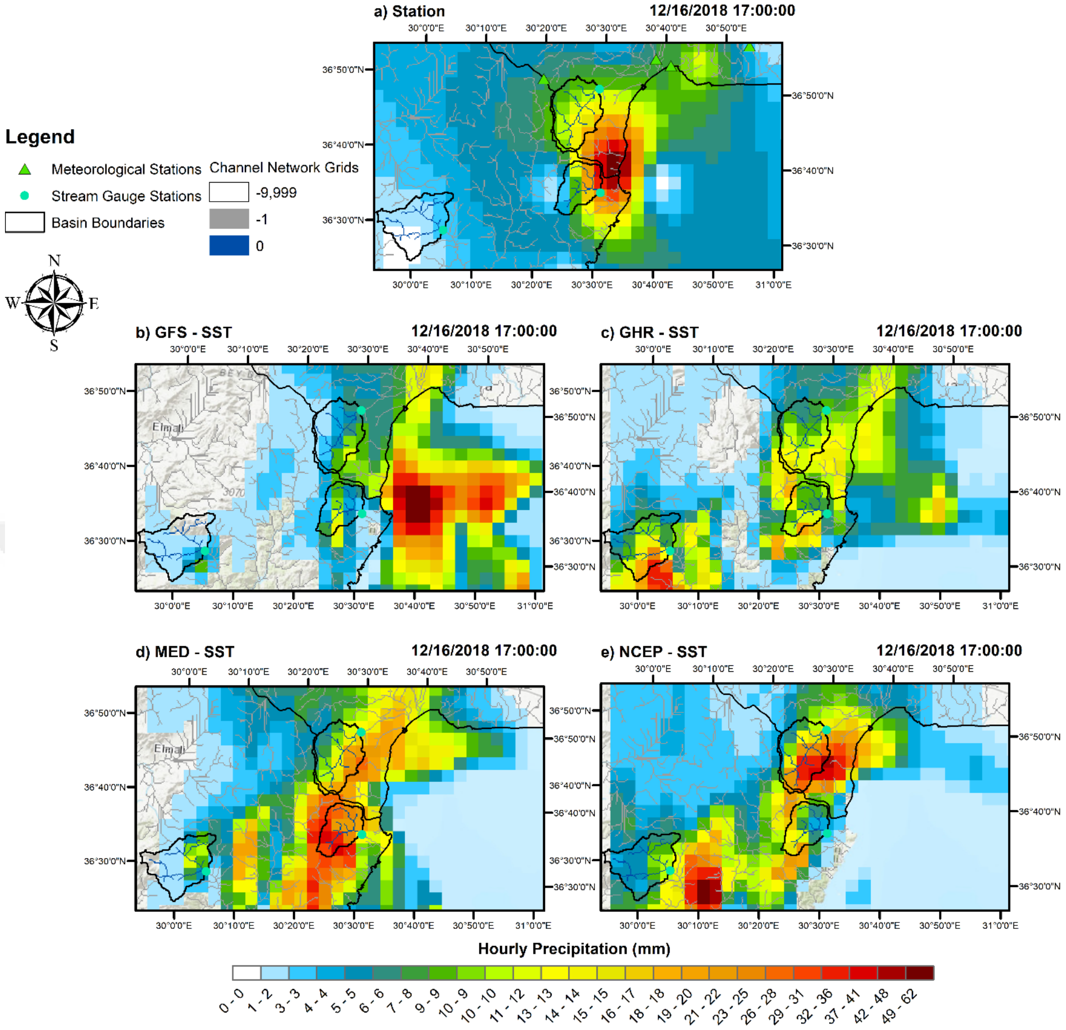


Figure 9 Spatial distribution of hourly precipitation at the peak hour (12/16/2018, 17:00:00 UTC) for the run period of 12/10/2018–12/20/2018 over the MED region. (a) The map at the top shows the interpolated observed precipitation map obtained from meteorological station data (green triangles). The black line indicates the boundaries of the selected basins for this study, while the blue dots show the corresponding stream gauge stations. The four maps in the sub-panels refer to the simulated precipitations by the WRF model derived from different SST data sources for the peak hour: b) GFS, c) GHRSST, d) Medspiration, and e) NCEP, respectively.

3.2 WRF-Hydro Model Results

3.2.1 Model Calibration Results

3.2.1.1 Manual Stepwise Calibration

This section evaluates the stepwise calibration outputs of three different events determined for selected sub-basins in the EBS and MED regions. The average values for statistics of the three events calculated for the parameter value determined at the end of the calibration of each parameter of all sub-basins are given in Table 5. Only the hydrograph results of two events and sub-basins of Arhavi and Kemer (D22A049 and D08A071, respectively) representing each study region of EBS and MED are given as example.

First, the calibration of the REFKDT parameter, which is in the group of parameters affecting the hydrograph volume, is calibrated for all the sub-basins and related events mentioned before. For the event that took place between 10/19/2016-10/29/2016, the hydrograph of the Arhavi basin is shown in Figure 10, while the hydrograph of the Kemer basin for the event that took place between 03/07/2017-03/17/2017 is shown in Figure 11. Since the default value of the REFKDT parameter is 3.0, the hydrograph with the REFKDT value of 3.0 in Figure 10 corresponds to that of the WRF-Hydro model run with the default parameter set.

The REFKDT value affects the infiltration capacity of the soil column. As it is seen from the graphs, as the REFKDT value increases, the hydrograph volume decreases because the penetration capacity of water into the soil increases. Considering the average statistics and hydrographs, the REFKDT value of 0.5, the lower limit for both basins, is selected as the optimum value. However, it has been observed that the hydrograph volume is still underestimated compared to the observation values in the Arhavi basin (Figure 10). As the calibration of this parameter progressed, the delay in the peak value in the model hydrograph is reduced, and the hydrograph volume is fed towards the timing of the observed peak value. Contrary to the hydrograph in Figure 10, a positive bias (3.72 m3/s) is observed in the average bias values of the three events in Table 5. A similar situation is also observed for the Kemer basin. Even though there is an overestimation in the representative hydrograph compared to the observation flow (Figure 11), a negative bias is calculated for the event averages (Table 5). In addition, a significant improvement (from -0.16 to 0.47) was noticed in the KGE value calculated for the Hopa basin (D22A147) compared to the default model outputs (Table 5).

Figure 12 and Figure 13 represent the calibration runs for the REDEPRTFAC parameter for the Arhavi and Kemer basins, respectively. It is observed that the model is not showing a response to RETDEPRTFAC parameter in all basins. This is because the EBS and MED regions have complex topographic characteristics, so the water is immediately directed as the overland flow without accumulating on the surface. For this reason, it was decided that the value of 0.0, which gave the best result, albeit with a small margin, was representative for both basins.

The SLOPE parameter is calibrated as the last parameter that determines the hydrograph volume in the model. As mentioned above, the SLOPE parameter determines the rate of transition of the water in the soil column to the bucket model. Therefore, as the parameter value increases, the hydrograph volume is expected to decrease. It is seen that although the SLOPE parameter has an improvement effect on the model performance, it does not have a significant effect on the event averages; therefore, it is decided that the default parameter value (0.1) is the optimal value (Figure 14 & Figure 15). Nevertheless, an improvement in bias, correlation, and RMSE is observed at the D22A147 and D09A095 basins (Table 5).

After the calibration of the parameters affecting the runoff volume is completed, the parameters affecting the temporal distribution and shape of the hydrograph are assessed. First, the calibration of OVROUGHRT is performed. For the represented basins in Figure 16 and Figure 17, there is no significant improvement is observed corresponding to the tuning of OVROUGHRTFAC. The average statistic results of the calibration of other basins also show that hydrographs are not significantly sensitive to the OVROUGHRTFAC parameter (Table 5). Therefore, the default parameter value (1.0) is selected for all basins, except the value of 0.1 is selected for the D09A095 basin.

Figure 18 and Figure 19 represent the calibration hydrographs of the MANN parameter for both basins. It is determined that the NSE value increased up to ~ 0.62 between the range of 1.5 and 2.0 in the Arhavi basin (Figure 18). A significant improvement is not observed in the model performance corresponding to the MANN in the Kemer basin (Figure 19). Based on the event averages, the value of 2.0 shows the most significant improvement at the D22A147 basin (~ 15% increase in KGE). In general, a decrease in the RMSE values is observed for all basins for the value of 2.0. A similar trend is observed when the MANN value is determined as 0.5 for the E08A008 basin in contrast to other basins (Table 5).

Finally, the calibration of the LKSATFAC parameter, which controls the saturated hydraulic conductivity in the lateral flow direction, is performed. Considering all the calibrated parameters, LKSATFAC stands out as the parameter that the model shows the most sensitivity in both regions. It significantly adjusts the peak value and improves the temporal distribution of the hydrograph, especially in the Kemer basin (Figure 21). A similar improvement is seen in the Arhavi basin (Figure 20). It is observed that the correlation value increases up to ~ 0.85, while the NSE value reaches ~ 0.73.

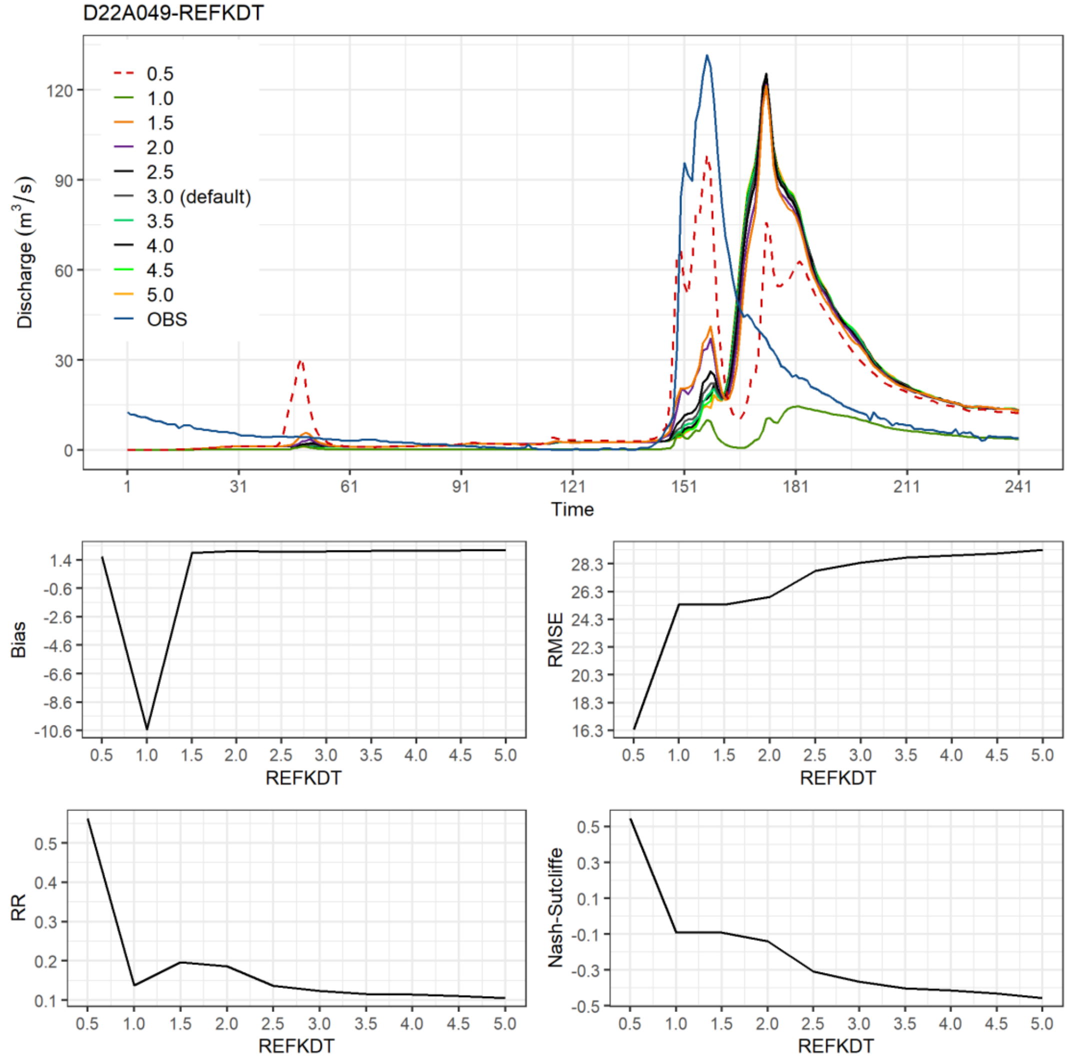


Figure 10 Arhavi (D22A049) basin REFKDT parameter calibration runs for the event that took place between 10/19/2016-10/29/2016.

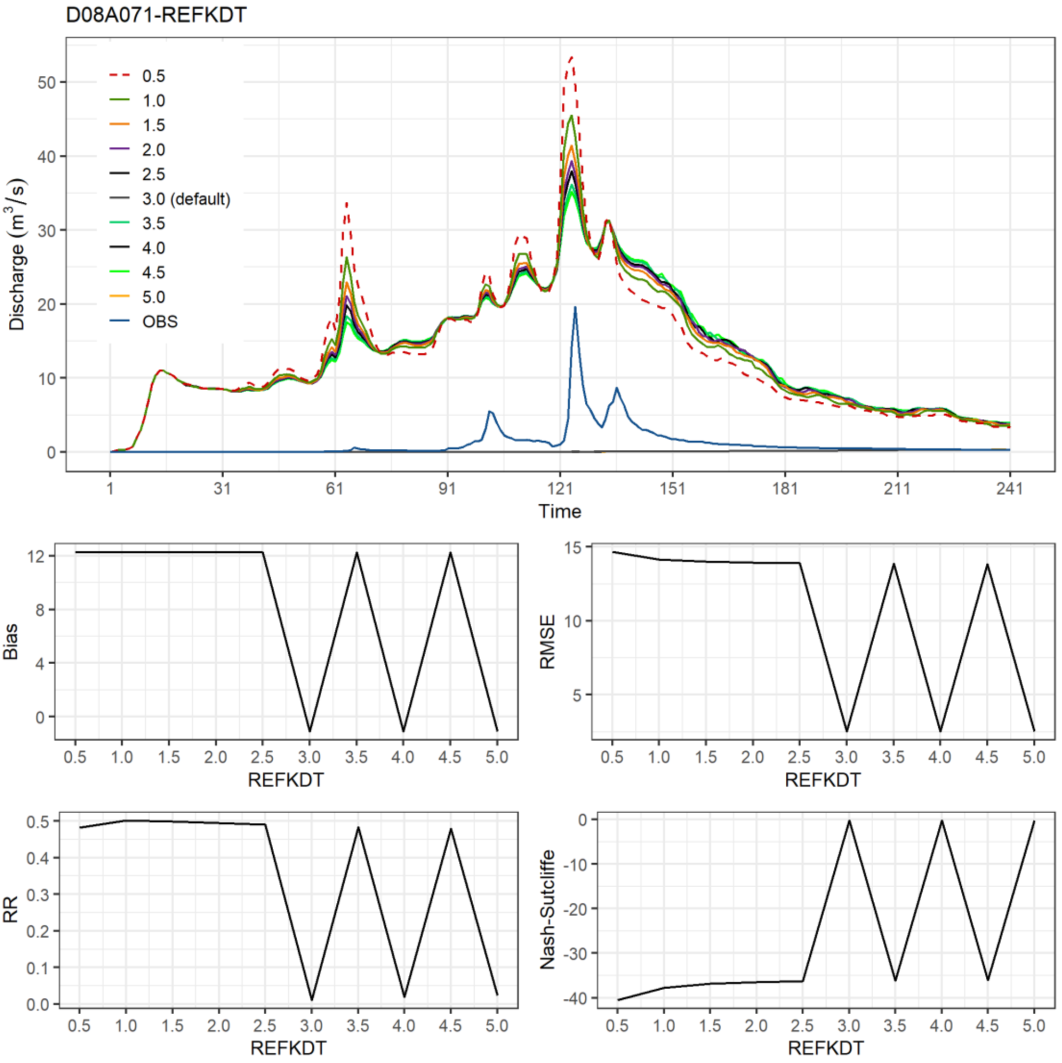


Figure 11 Kemer (D08A071) basin REFKDT parameter calibration runs for the event that took place between 03/07/2017-03/17/2017.

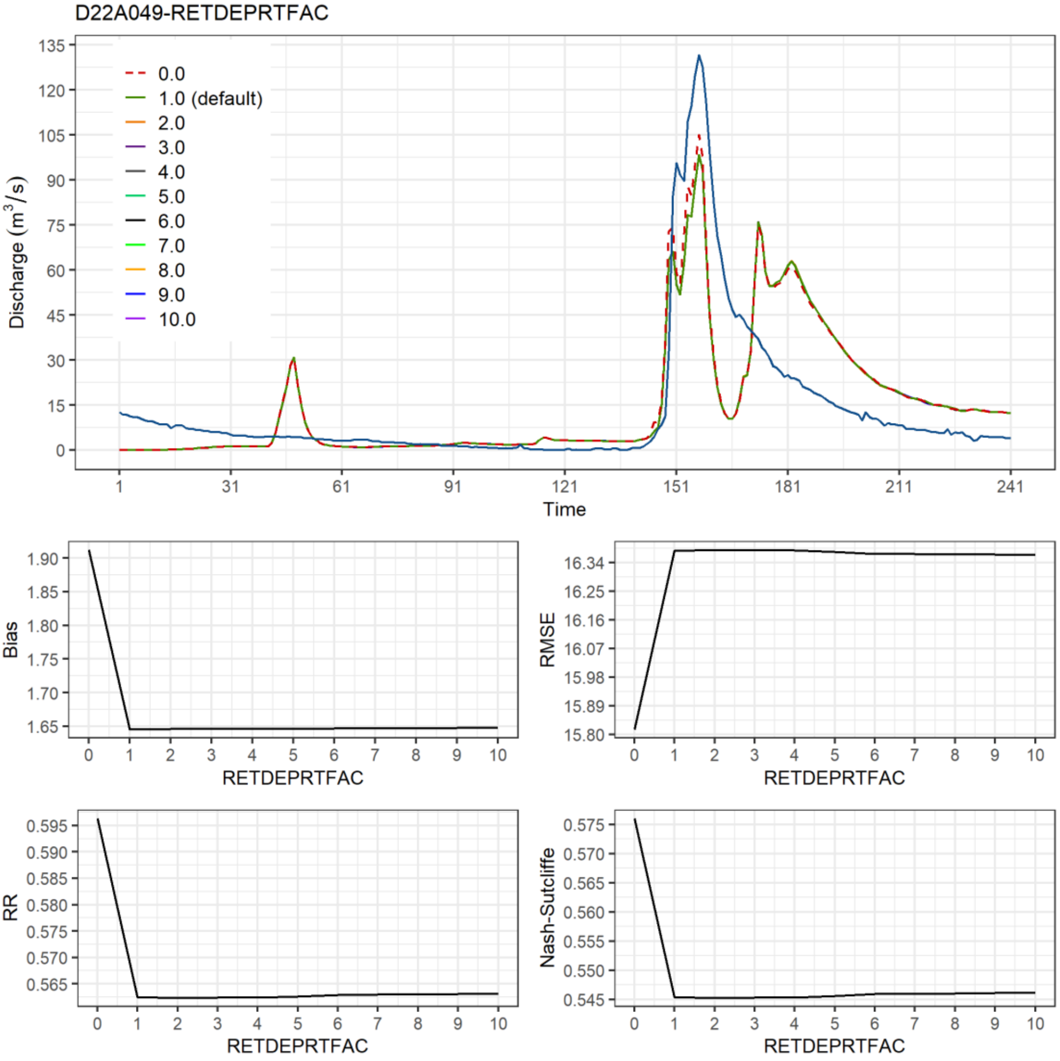


Figure 12 Arhavi (D22A049) basin RETDEPRTFAC parameter calibration runs for the event that took place between 10/19/2016-10/29/2016.

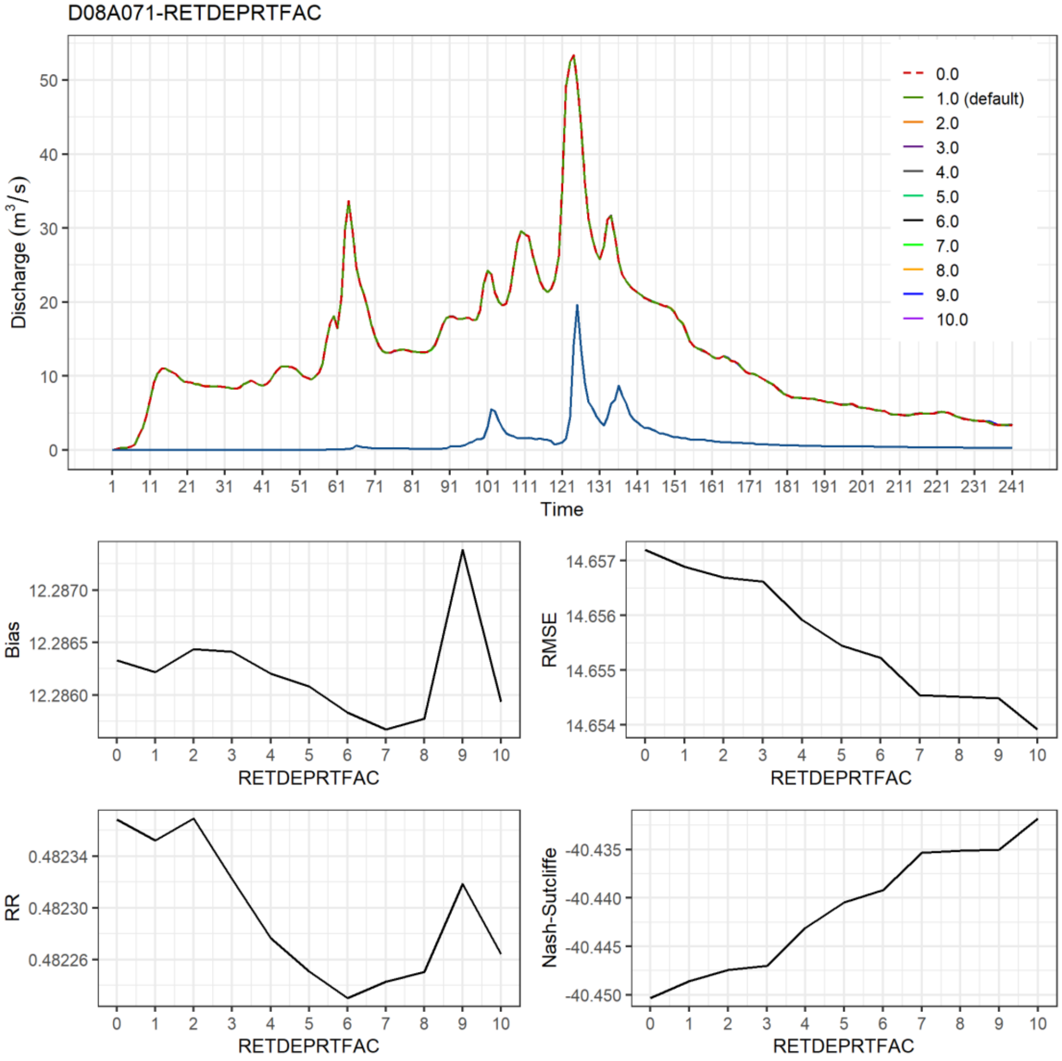


Figure 13 Kemer (D08A071) basin RETDEPRTFAC parameter calibration runs for the event that took place between 03/07/2017-03/17/2017.

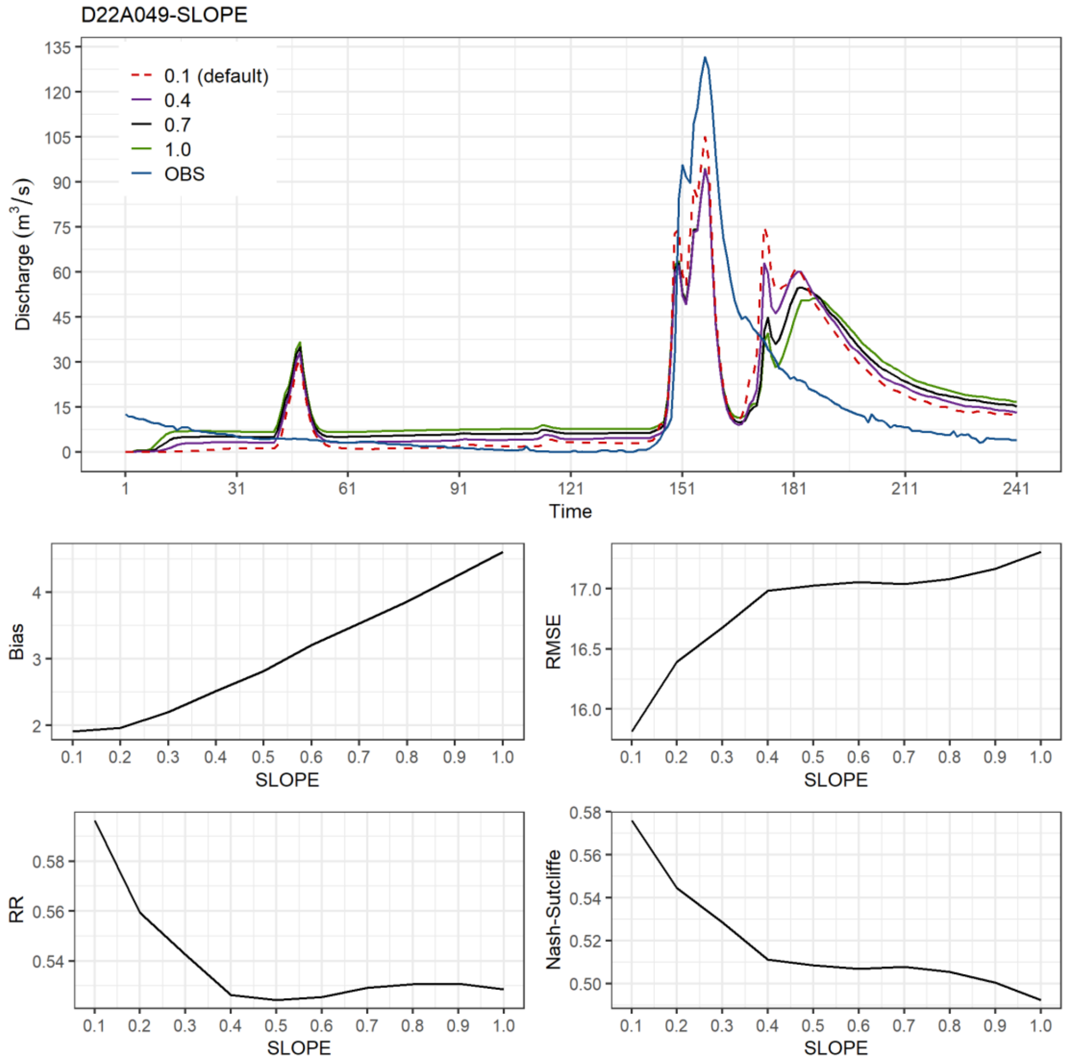


Figure 14 Arhavi (D22A049) basin SLOPE parameter calibration runs for the event that took place between 10/19/2016-10/29/2016.

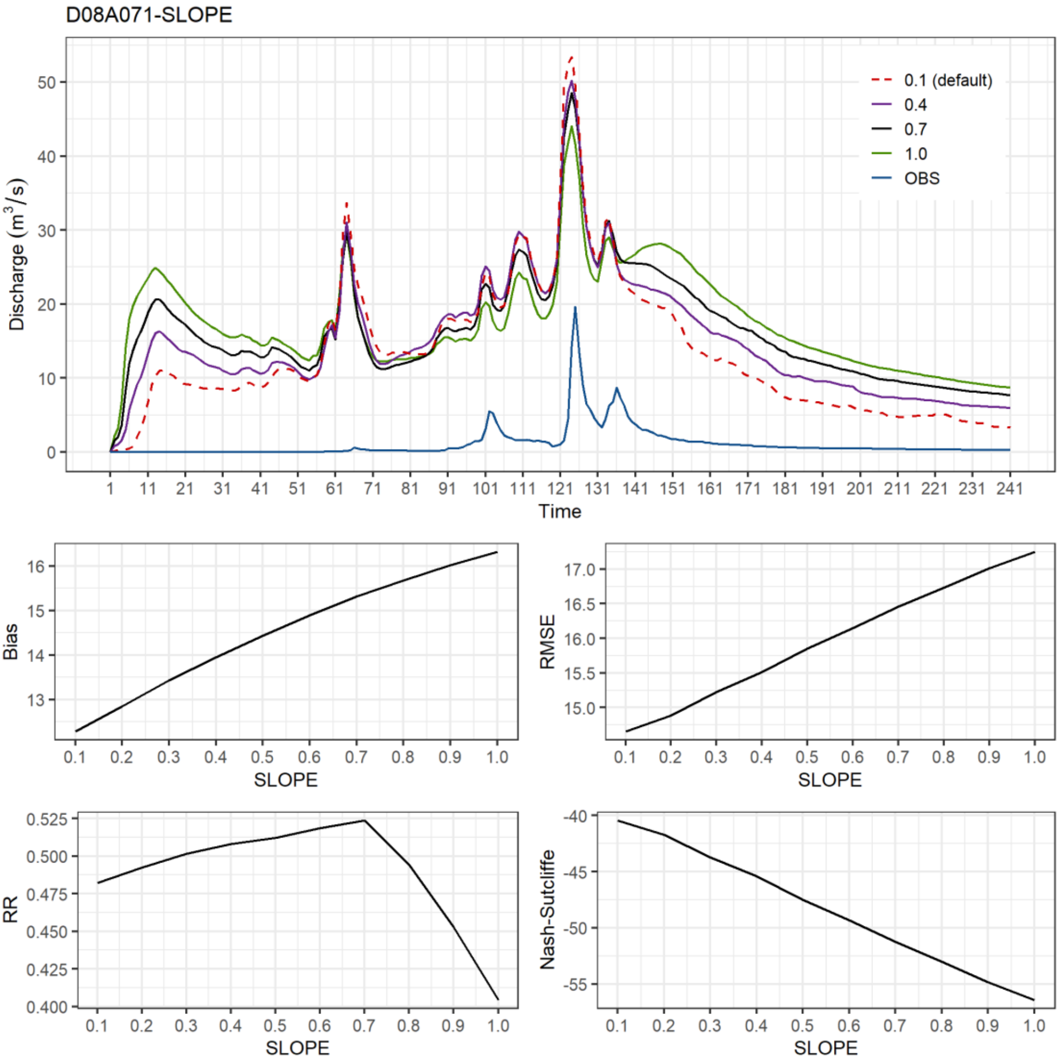


Figure 15 Kemer (D08A071) basin SLOPE parameter calibration runs for the event that took place between 03/07/2017-03/17/2017.

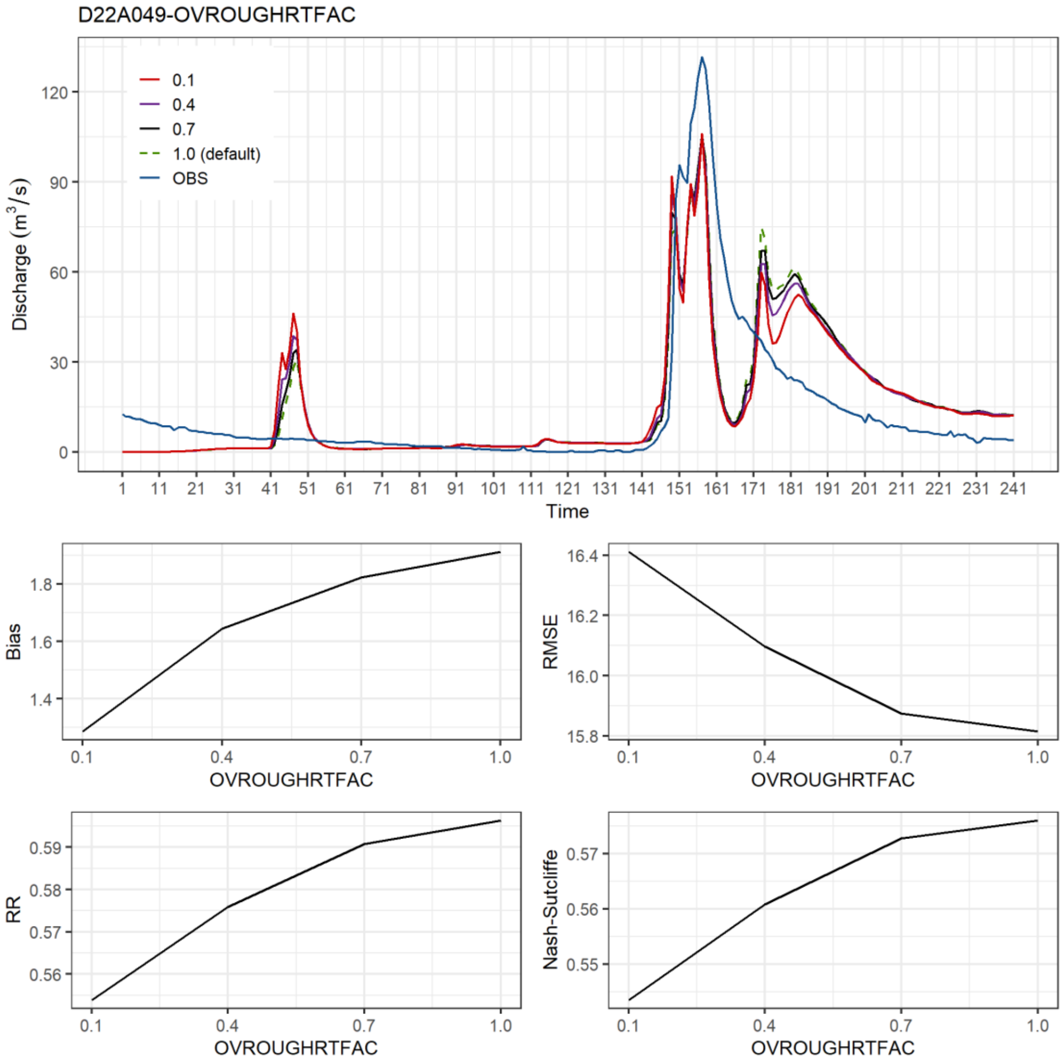


Figure 16 Arhavi (D22A049) basin OVROUGHRTFAC parameter calibration runs for the event that took place between 10/19/2016-10/29/2016.

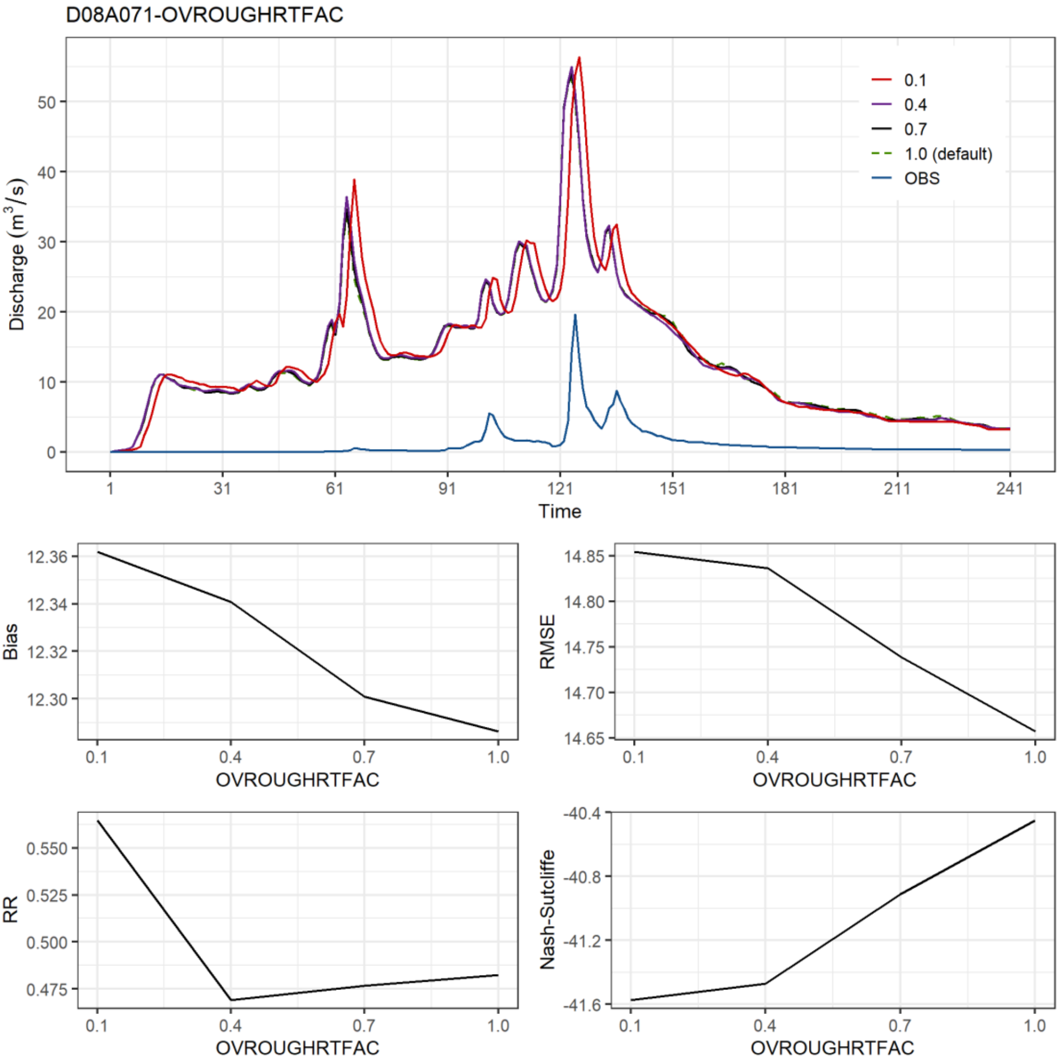


Figure 17 Kemer (D08A071) basin OVROUGHTRFAC parameter calibration runs for the event that took place between 03/07/2017-03/17/2017.

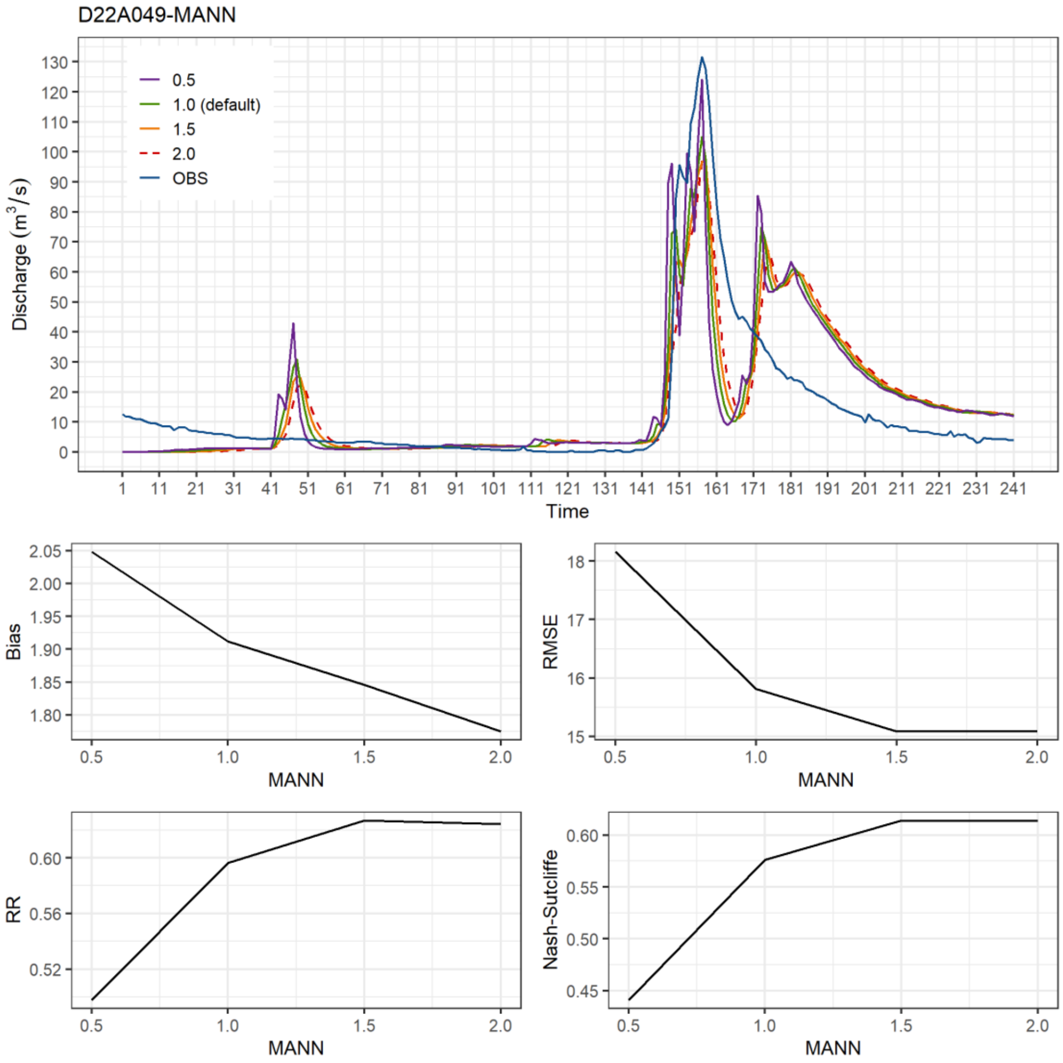


Figure 18 Arhavi (D22A049) basin MANN parameter calibration runs for the event that took place between 10/19/2016-10/29/2016.

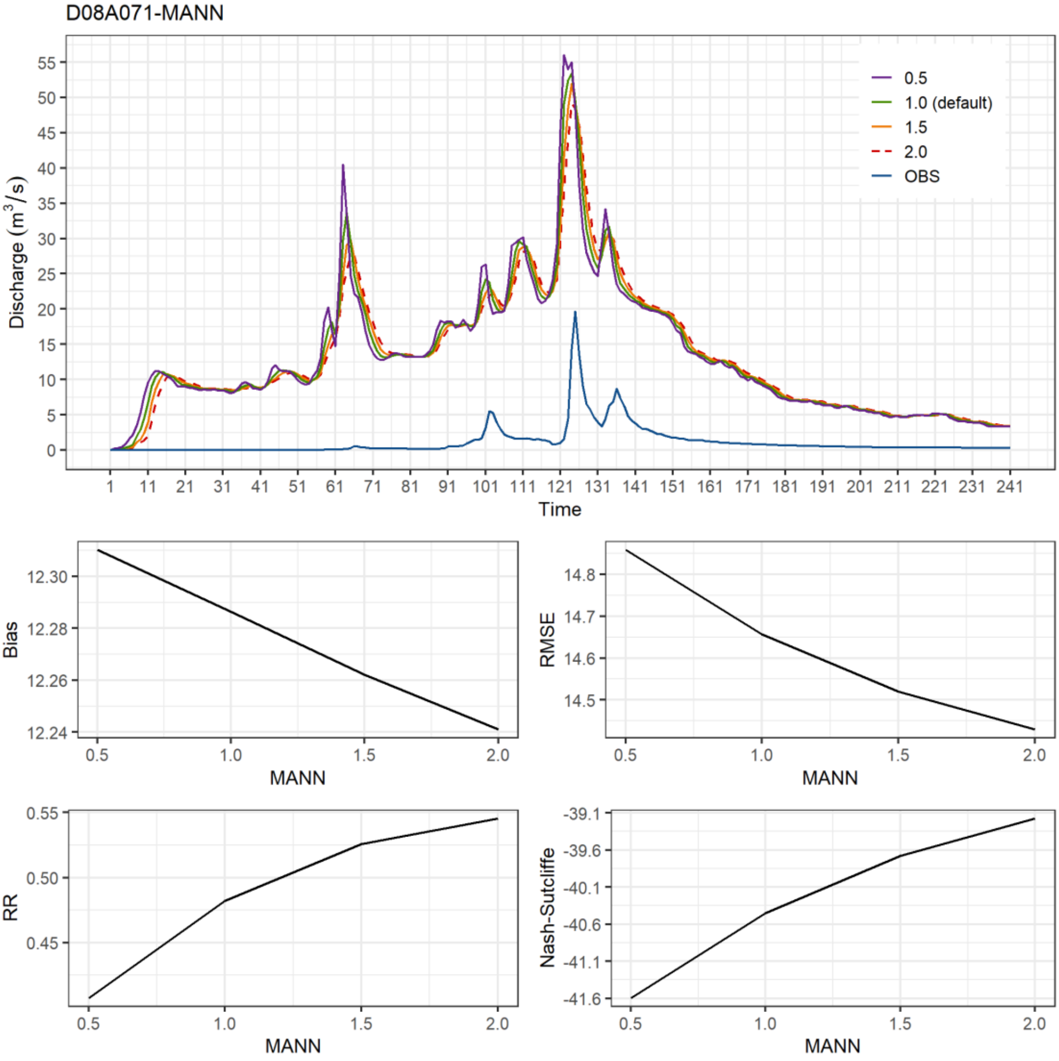


Figure 19 Kemer (D08A071) basin MANN parameter calibration runs for the event that took place between 03/07/2017-03/17/2017.

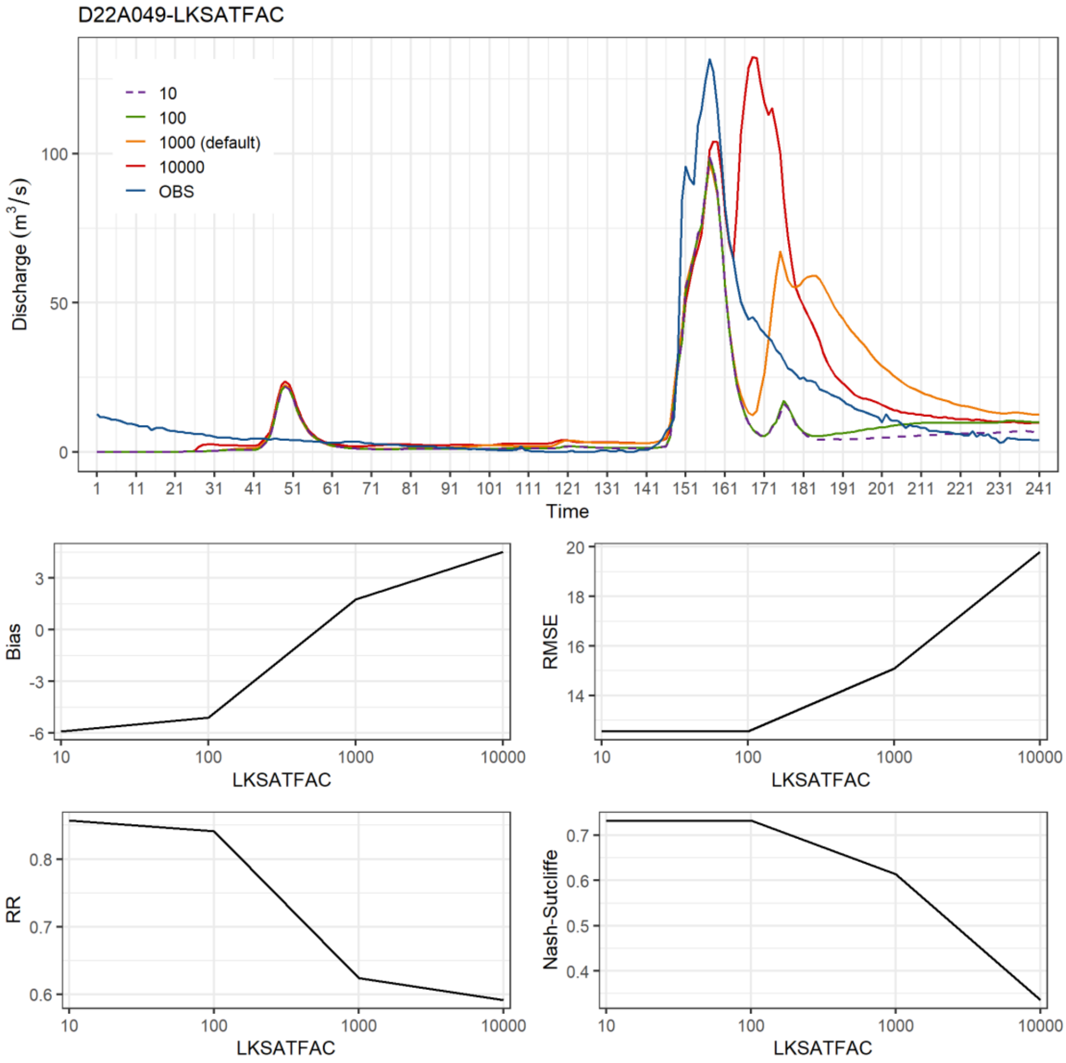


Figure 20 Arhavi (D22A049) basin LKSATFAC parameter calibration runs for the event that took place between 10/19/2016-10/29/2016.

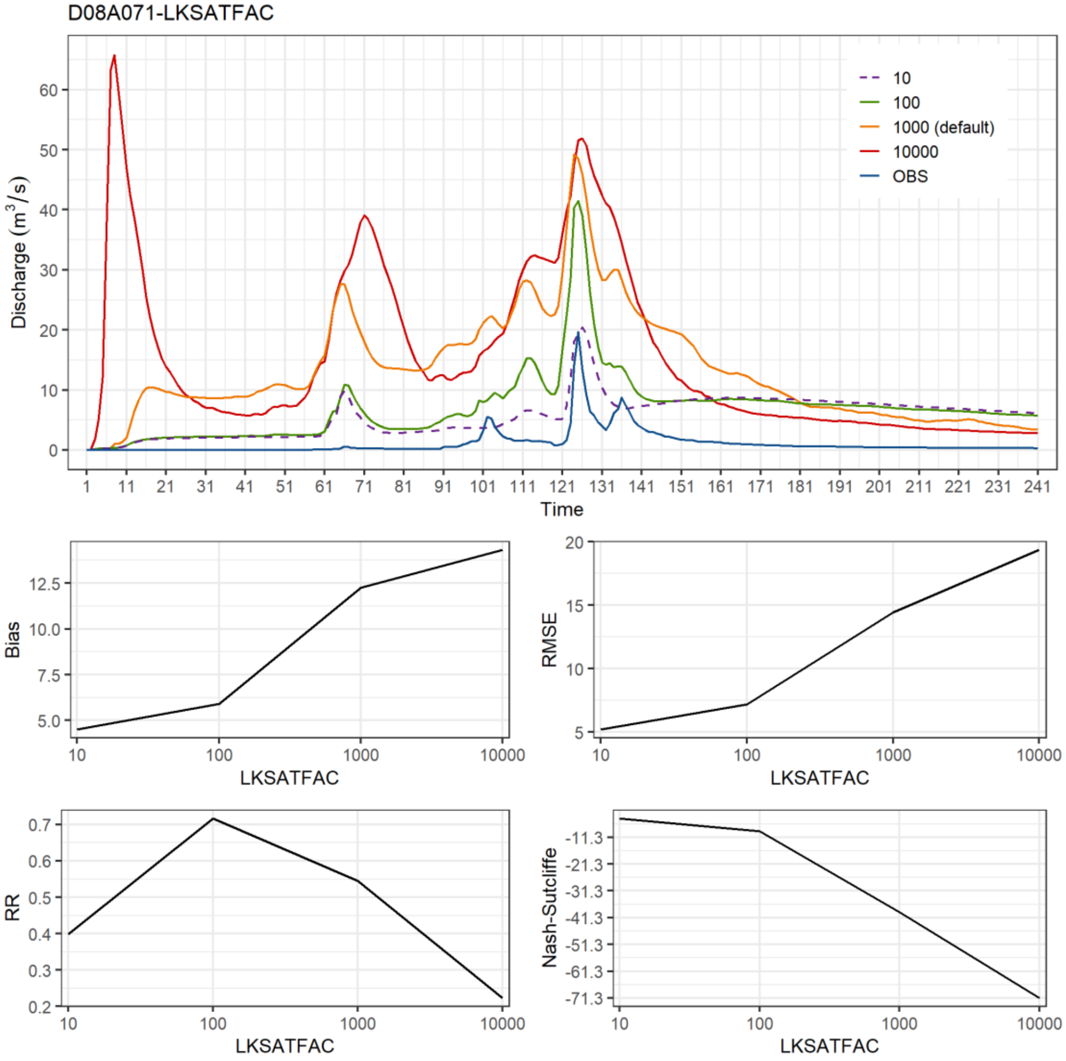


Figure 21 Kemer (D08A071) basin LKSATFAC parameter calibration runs for the event that took place between 03/07/2017-03/17/2017.

Table 5 Average statistics (Bias, correlation coefficient (RR), root mean square error (RMSE) and Nash-Sutcliffe Adequacy (NSE) and Kling-Gupta Efficiency (KGE)) of three events based on the calibrated parameter sets for each selected basin in MED and EBS regions.

Region			D22A049				
		Parm. Values	Bias	RMSE	RR	NSE	KGE
EBS	Default Parameter Set		4.24	40.55	0.13	-0.39	0.13
	REFKDT	0.5	3.72	40.48	0.38	-0.33	0.30
	RETDEPRTFAC	0.0	4.00	40.45	0.39	-0.35	0.31
	SLOPE	0.1	4.00	40.45	0.39	-0.35	0.31
	OVROUGHRTFAC	1.0	4.00	40.45	0.39	-0.35	0.31
	MANN	2.0	3.69	37.54	0.39	-0.14	0.40
	LKSATFAC	10	-2.34	32.16	0.56	0.16	0.40
			D22A147				
		Parm. Values	Bias	RMSE	RR	NSE	KGE
	Default Parameter Set		0.48	5.75	0.38	-0.58	-0.16
	REFKDT	0.5	0.58	3.20	0.63	0.51	0.47
	RETDEPRTFAC	0.0	0.97	3.16	0.63	0.52	0.47
	SLOPE	1.0	1.01	2.88	0.67	0.58	0.47
	OVROUGHRTFAC	1.0	1.01	2.88	0.67	0.58	0.47
	MANN	2.0	0.78	2.65	0.68	0.63	0.54
	LKSATFAC	10	0.47	2.22	0.75	0.72	0.61
MED			D08A071				
		Parm. Values	Bias	RMSE	RR	NSE	KGE
	Default Parameter Set		-5.28	16.68	0.43	-0.03	-0.12
	REFKDT	0.5	-1.02	30.12	0.44	-13.53	-3.18
	RETDEPRTFAC	0.0	-0.47	30.53	0.44	-13.54	-3.17
	SLOPE	0.1	-0.47	30.53	0.44	-13.54	-3.17
	OVROUGHRTFAC	1.0	-0.48	29.84	0.46	-13.52	-3.16
	MANN	2.0	-0.50	29.85	0.49	-13.09	-3.14
	LKSATFAC	10	-5.57	26.30	0.46	-1.26	-0.75
			D09A095				
		Parm. Values	Bias	RMSE	RR	NSE	KGE
	Default Parameter Set		-0.10	12.40	0.59	-1.69	-0.55
	REFKDT	0.5	1.31	9.67	0.73	-1.21	-0.29
	RETDEPRTFAC	0.0	1.59	9.51	0.65	-1.16	-0.25
	SLOPE	1.0	1.66	8.77	0.70	-0.78	-0.14
	OVROUGHRTFAC	0.1	1.69	8.55	0.70	-0.59	-0.06
	MANN	2.0	1.70	8.35	0.81	-0.54	-0.04
	LKSATFAC	10	2.29	9.02	0.77	-0.93	-0.14

Table 5 (Cont’d)

		E08A008					
		Parm. Values	Bias	RMSE	RR	NSE	KGE
MED	Default Parameter Set		19	23.45	0.35	-283.11	-13.75
	REFKDT	0.5	18.93	23.17	0.51	-289.21	-13.89
	RETDEPRTFAC	0.0	18.95	23.14	0.51	-288.88	-13.87
	SLOPE	0.1	18.95	23.14	0.51	-288.88	-13.87
	OVROUGHRTFAC	1.0	18.92	23.14	0.51	-288.89	-13.88
	MANN	0.5	18.97	22.99	0.50	-289.34	-13.86
	LKSATFAC	10	6.22	8.07	0.40	-54.98	-5.14

3.2.1.2 Automated Calibration

This section includes the results of the WRF-Hydro model calibration part utilizing an automated calibration algorithm, called DDS, and the algorithm test runs.

3.2.1.2.1 DDS Algorithm Test Results

Before the DDS algorithm is utilized for the calibration process of the WRF-Hydro model, the algorithm is tested with a hypothetical hydrograph with the known set of the same parameters, which are calibrated in the previous section. Instead of the observed streamflow data, the hydrograph obtained from this set is given as the input to the algorithm. The optimization of the model is performed based on this given hydrograph. To see the capability of the algorithm to solve the problem in terms of the number of decision variables which is defined as the number of parameters, this test is performed with the different parameters set sizes. First, one parameter (REFKDT) is included in the algorithm, then the size of the parameter set is systematically increased by adding one parameter at each step. At the last stage, all parameters are included in the algorithm, and the dimension of the problem is increased to six. The iteration number is defined as 60 throughout the test.

The results of the DDS algorithm test are given in Figure 22 - Figure 27. In Figure 22, the DDS algorithm is run with one parameter, which is REFKDT, and the other parameters are fixed to the parameter values defined in the hypothetical hydrograph. Results indicate that the model can find exactly the same hydrograph around the 40th iteration (Figure 22). Similar performance is seen for the calibration with the parameter size of two (Figure 23). After the set size is increased to three, the deflection from the hypothetical graph starts. Statistics show that the algorithm still can find the set of parameters within the acceptable statistics range as the KGE value between 0.6-0.9 until the addition of the last parameter (Figure 24-Figure 26). At the last step, the KGE value drastically decreases to -0.47. It is seen that the algorithm is still successful in terms of finding the hydrograph timing (Figure 27). Therefore, the correlation values are still in the acceptable range with a value of 0.8. However, the overestimation seen in the hydrograph volume propagates to the model error in the statistics.

DDS algorithm test shows that the algorithm is working effectively for optimizing the simulated hydrograph by WRF-Hydro model with respect to the objective function of KGE. Although a dramatic increase is seen when the size of the parameter set is increased to six, in order to perform an even comparison between the stepwise and the automated calibration methods, the same set of parameters (with the size of six) are used for the calibration process with the DDS algorithm. Since the number of iterations defined in the algorithm is positively correlated with the dimension of the problem definition, which is the size of the parameter set, it is assumed that the calibration started with the higher number of iterations may solve this issue.

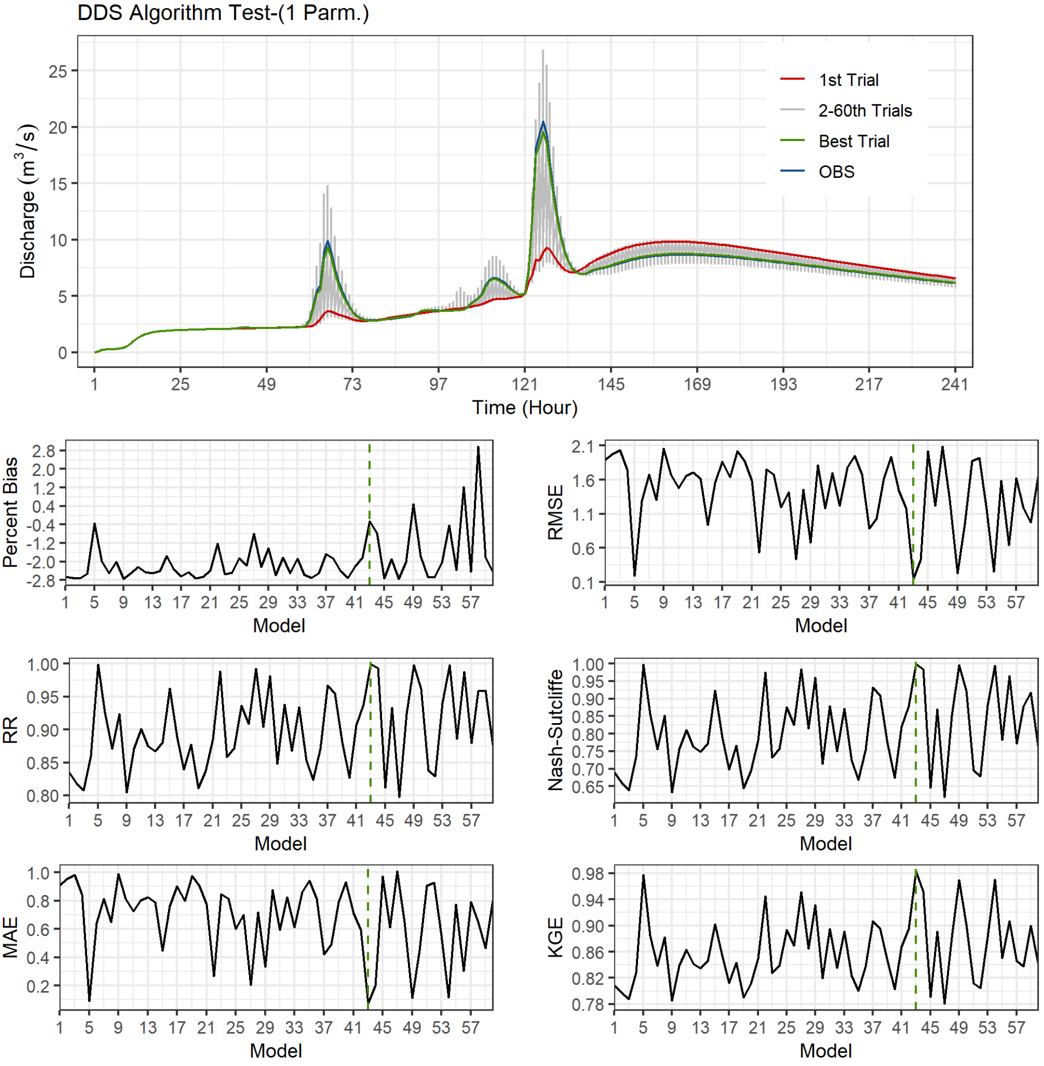


Figure 22 The calibration hydrographs performed using the DDS algorithm, including one decision variable (i.e., REFKDT parameter) for a hypothetical event in the Kemer basin and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

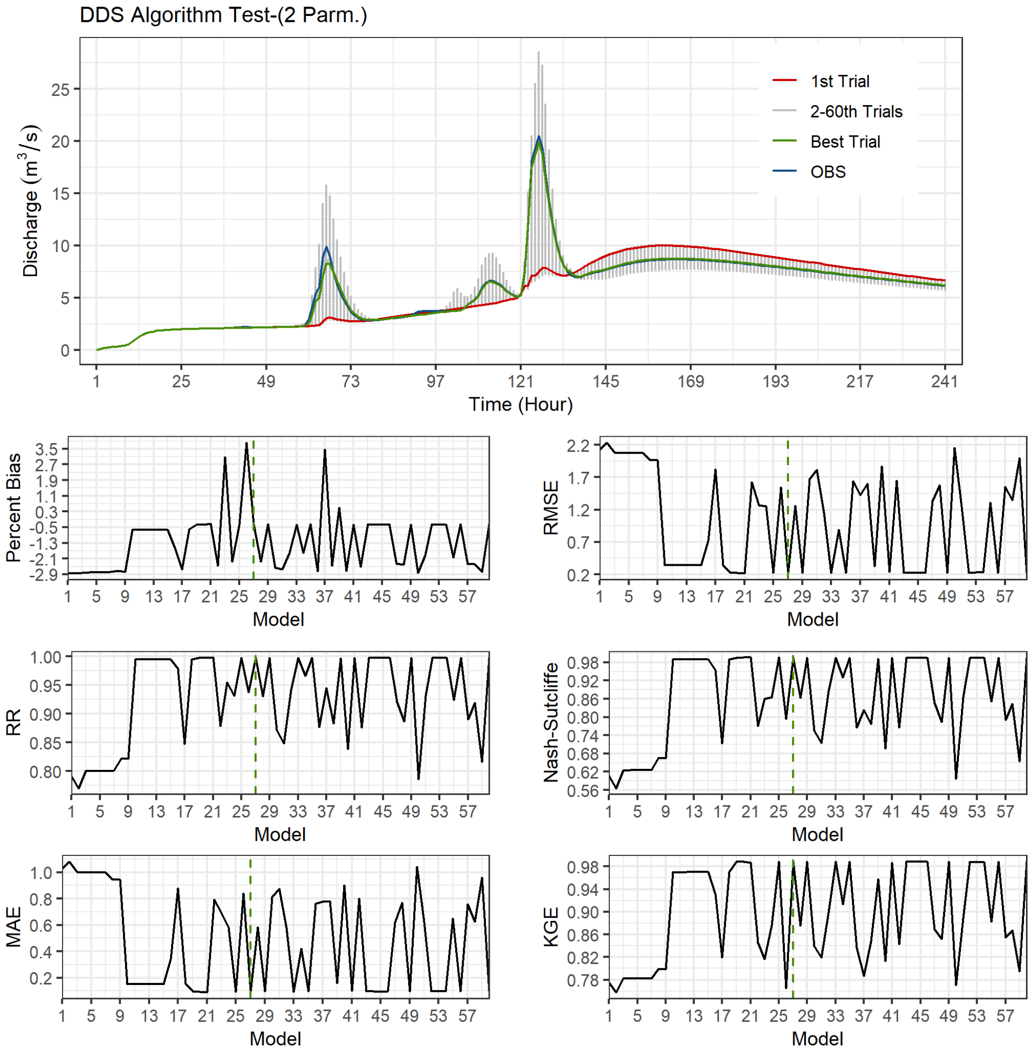


Figure 23 The calibration hydrographs performed using the DDS algorithm including two decision variables (i.e., REFKDT and RETDEPRTFAC parameters) for a hypothetical event in the Kemer basin and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

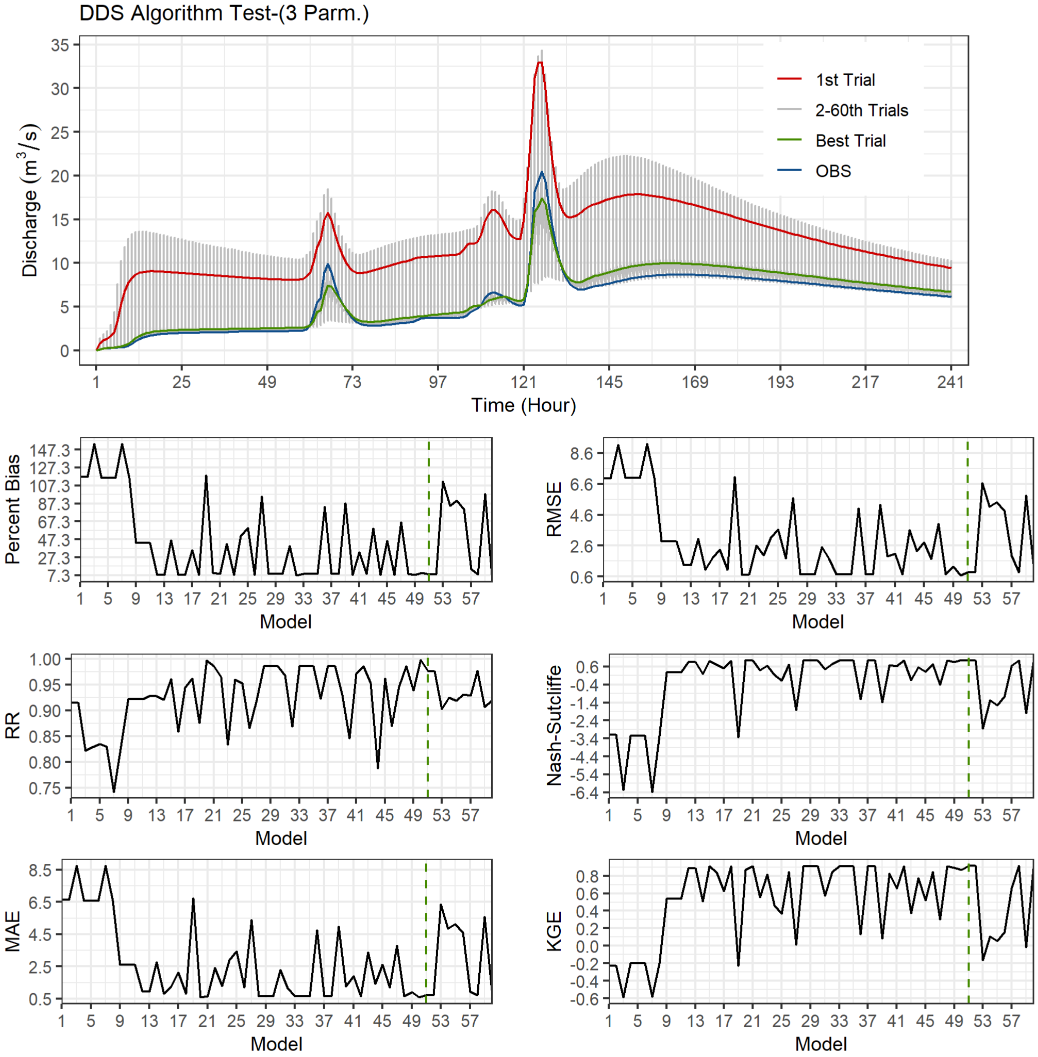


Figure 24 The calibration hydrographs performed using the DDS algorithm including three decision variables (i.e., REFKDT, RETDEPRTFAC, and SLOPE parameters) for a hypothetical event in the Kemer basin and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

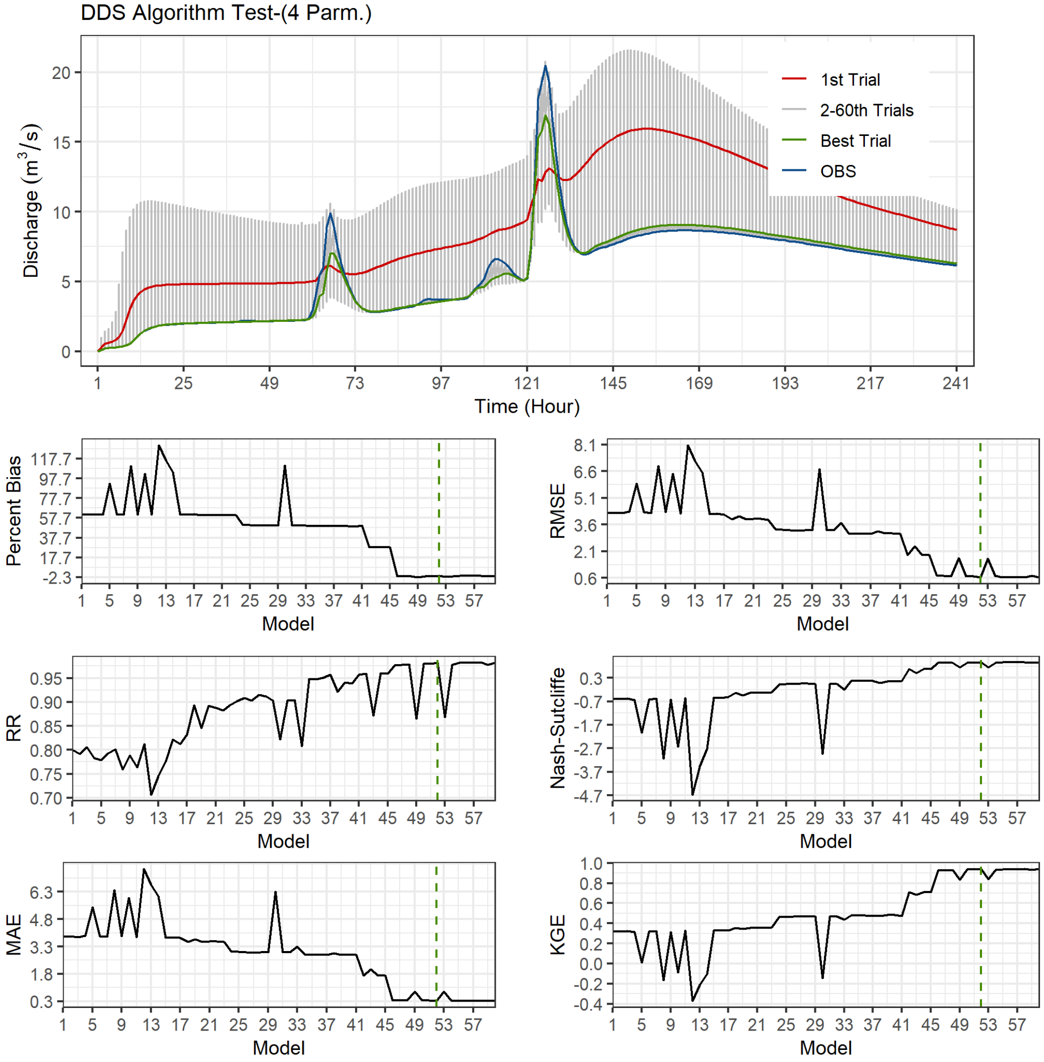


Figure 25 The calibration hydrographs performed using the DDS algorithm including four decision variables (i.e., REFKDT, RETDEPRTFAC, SLOPE, and OVROUGHRTFAC parameters) for a hypothetical event in the Kemer basin and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

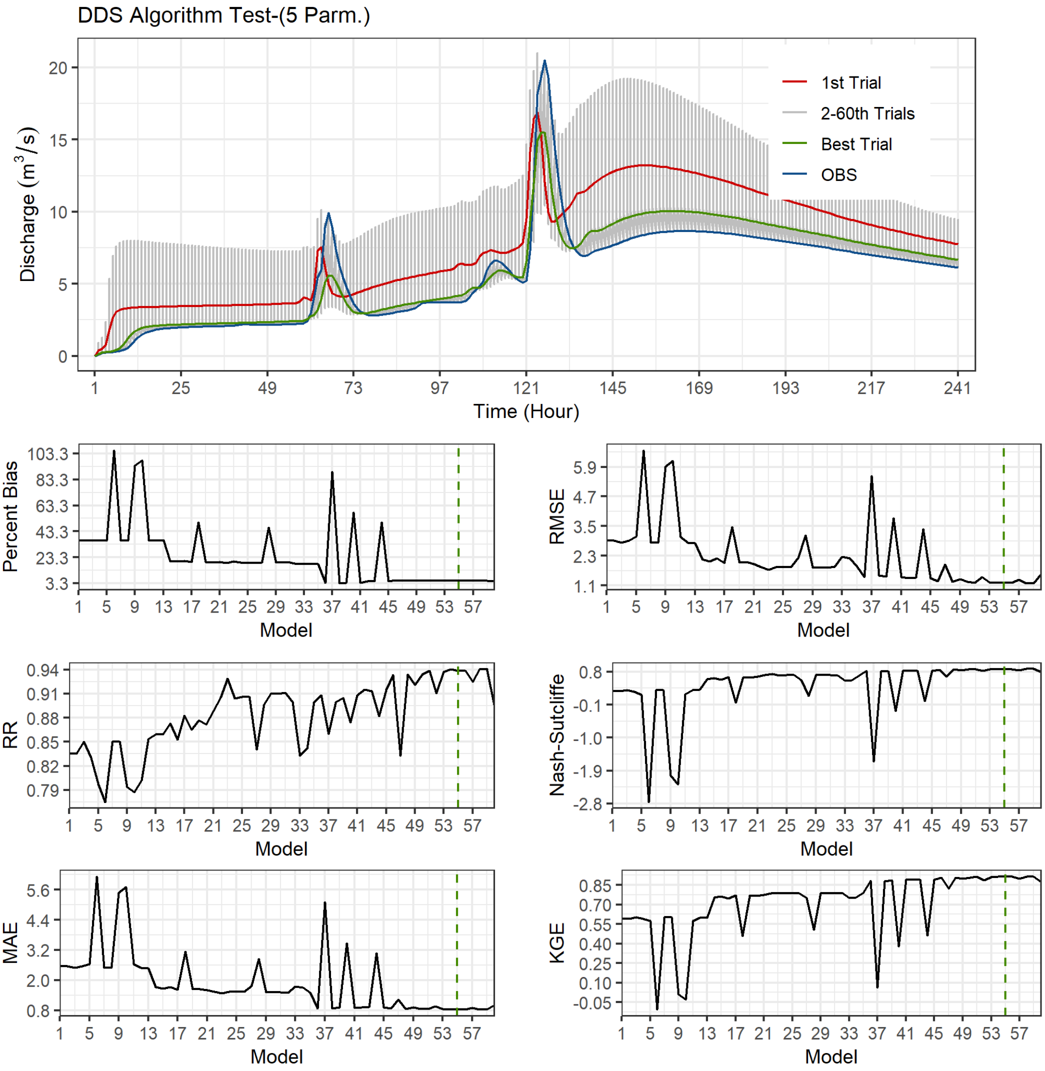


Figure 26 The calibration hydrographs performed using the DDS algorithm including five decision variables (i.e., REFKDT, RETDEPRTFAC, SLOPE, OVROUGHRTFAC, and MANN parameters) for a hypothetical event in the Kemer basin and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

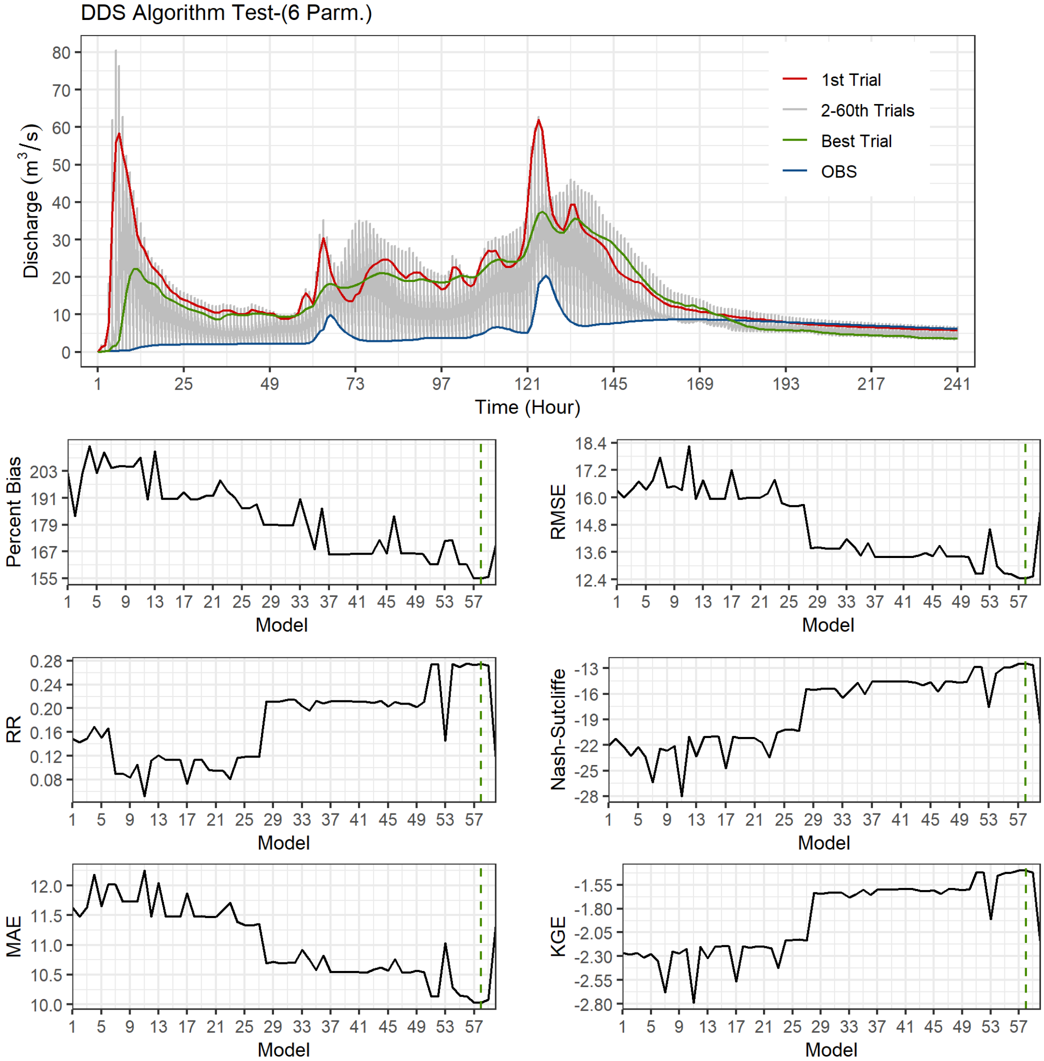


Figure 27 The calibration hydrographs performed using the DDS algorithm including six decision variables (i.e., REFKDT, RETDEPRTFAC, SLOPE, OVROUGHRTFAC, MANN, and LKSATFAC parameters) for a hypothetical event in the Kemer basin and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

3.2.1.2.2 DDS Algorithm Calibration Results

The calibration outputs generated with the DDS algorithm chosen as the optimization algorithm for automatic calibration are represented in this section. At this stage, two sub-basin and events from both regions are selected to be studied for the automated calibration process. Arhavi basin and the event that occurred between 10/19/2016-10/29/2016 are selected for the EBS region while the calibration was carried out with the Kemer basin and the event that occurred between 03/07/2017-03/17/2017 in the MED region. These events belong to the same basins that are also used in manual calibration in the previous section. In order to see the effect of the number of iterations on the performance of the model, four different iteration numbers (60, 100, 150, and 200) are determined for the automated calibration process.

In this context, the DDS calibration trials carried out for the Arhavi basin and 10/19/2016-10/29/2016 event are represented from Figure 28 to Figure 31. The calibration attempts with 60- and 100- iterations reveal that the model can gradually capture the peak timing compared to the first trial (Figure 28 & Figure 29). KGE value is improved around 21% when the iteration number is increased from 60 to 100. In general, as the number of iterations increases, it is seen that the success of the model in solving the event progressively increases. However, in calibration trials with 150-iterations, it is observed that the performance decreased compared to the calibration with 100-iterations, and the algorithm continues to give a more successful result in 200-iterations compared to the calibration with 100 iterations (Table 6). Although the highest KGE value (0.74) is achieved in the 200-iterations, no improvement is seen in the RMSE, RR, and NSE values compared to the 100-iterations. However, the hydrograph of the best trial for 200-iteration shows that the model can capture the peak time, as well as a gradual decrease trend (Figure 31), instead of a sudden decrease observed in the falling limb of the hydrographs of the best results with other iteration numbers (Figure 28 & Figure 29).

When Kemer basin simulations are examined, it is observed that the model is overestimating systematically, and the calibration is making progress in reducing this difference between the simulated and the observed hydrographs (Figure 32-Figure 33). Similar to the case in the Arhavi basin, it is determined that there is a decrease in the statistical metrics for the calibration with 150 iterations for the Kemer basin (Table 6).

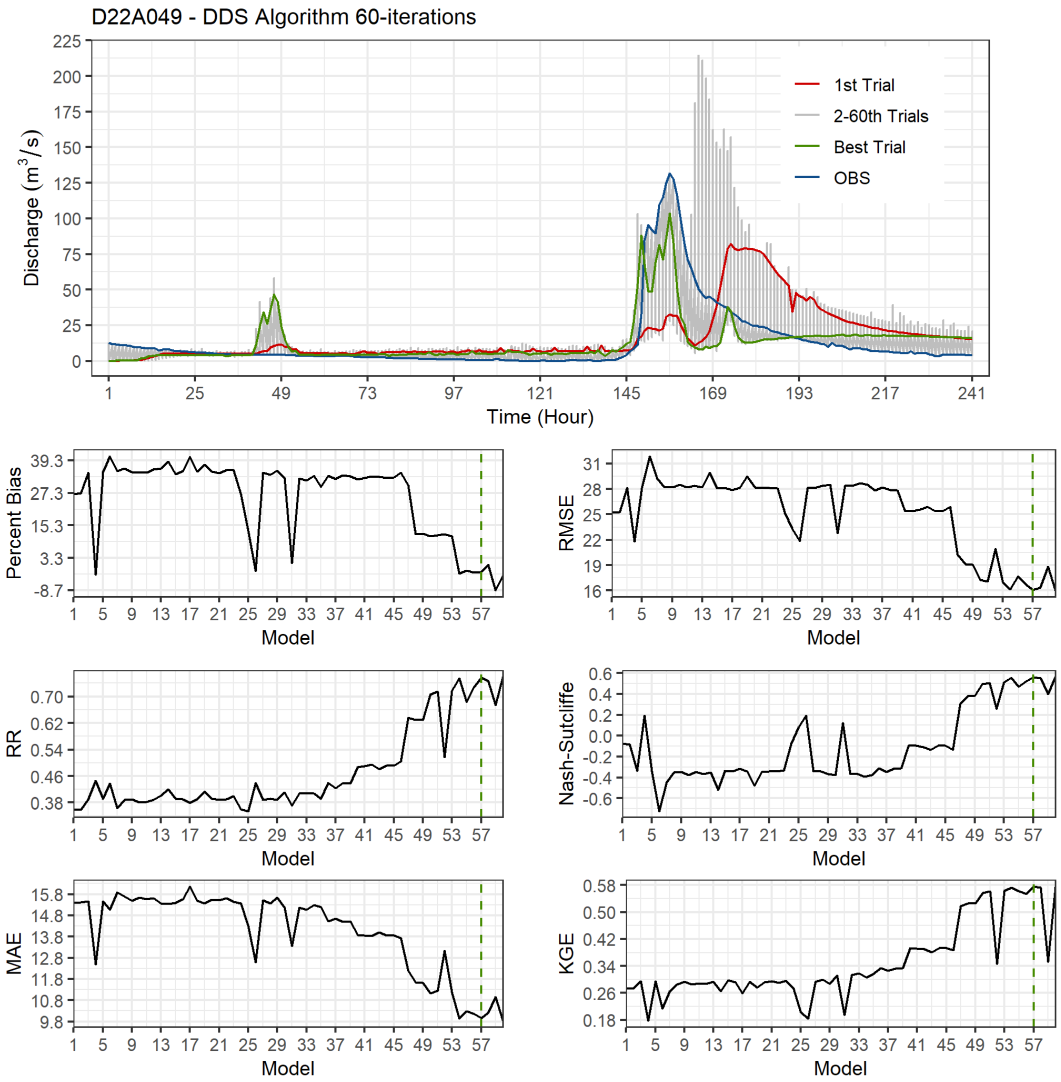


Figure 28 The calibration hydrographs performed with a total of 60 iterations using the DDS algorithm for the Arhavi basin and the event that took place between

10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE)).

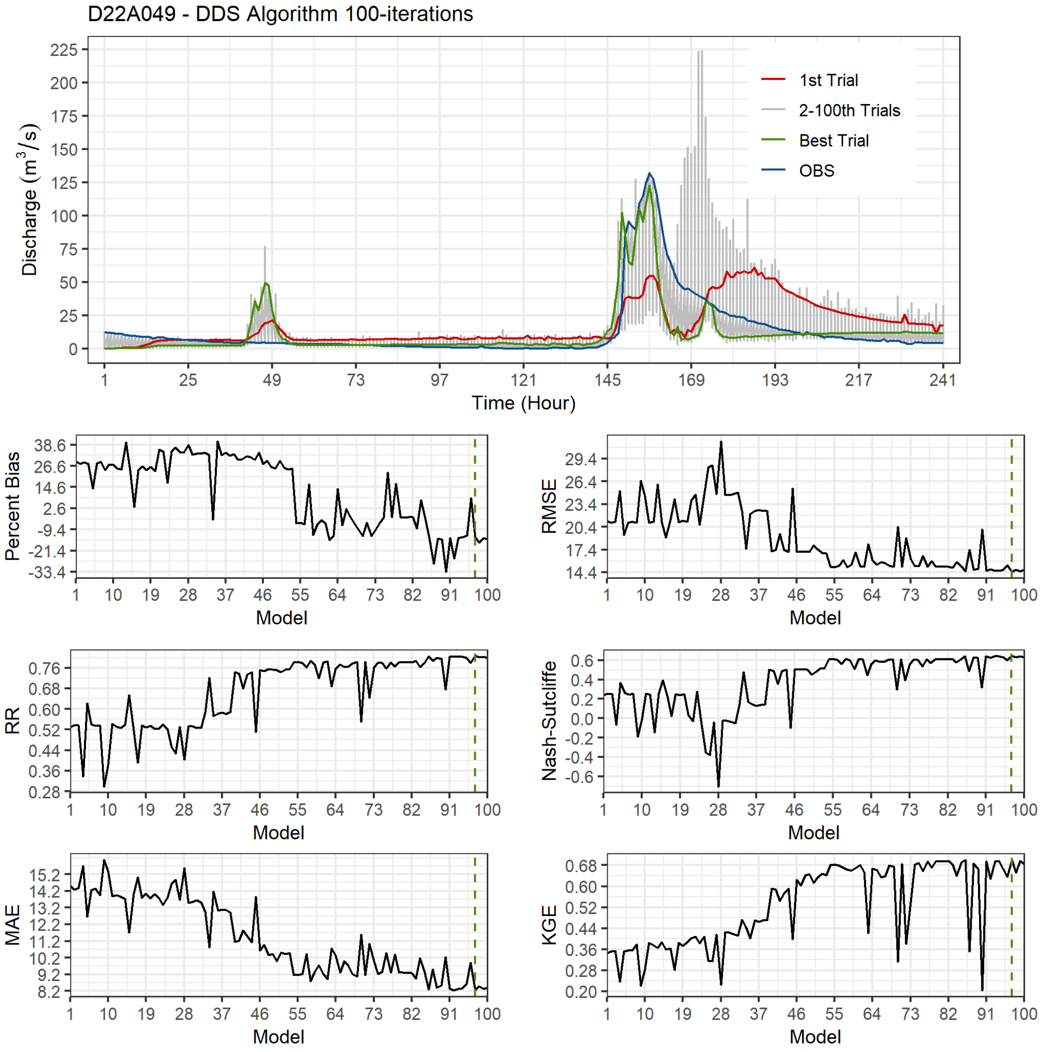


Figure 29 The calibration hydrographs performed with a total of 100 iterations using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR),

Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

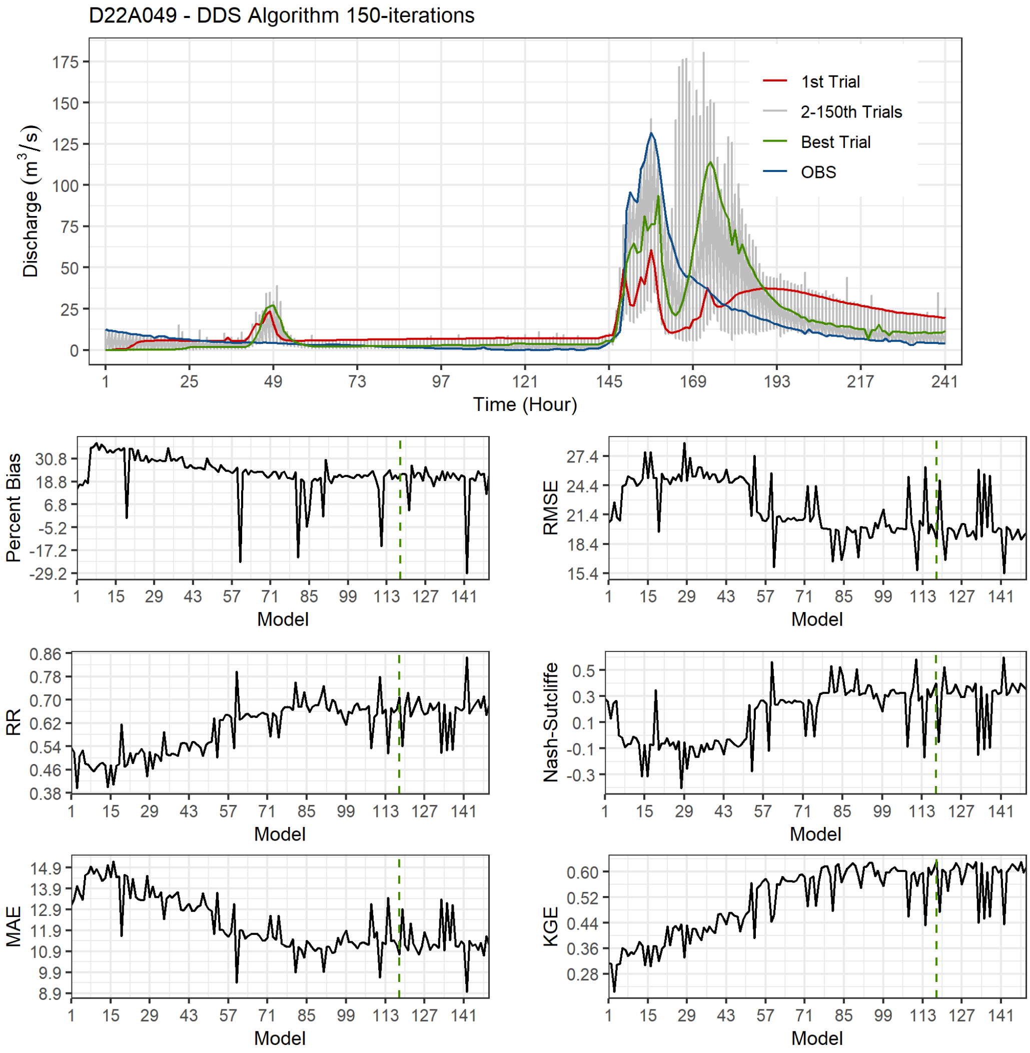


Figure 30 The calibration hydrographs performed with a total of 150 iterations using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE)).

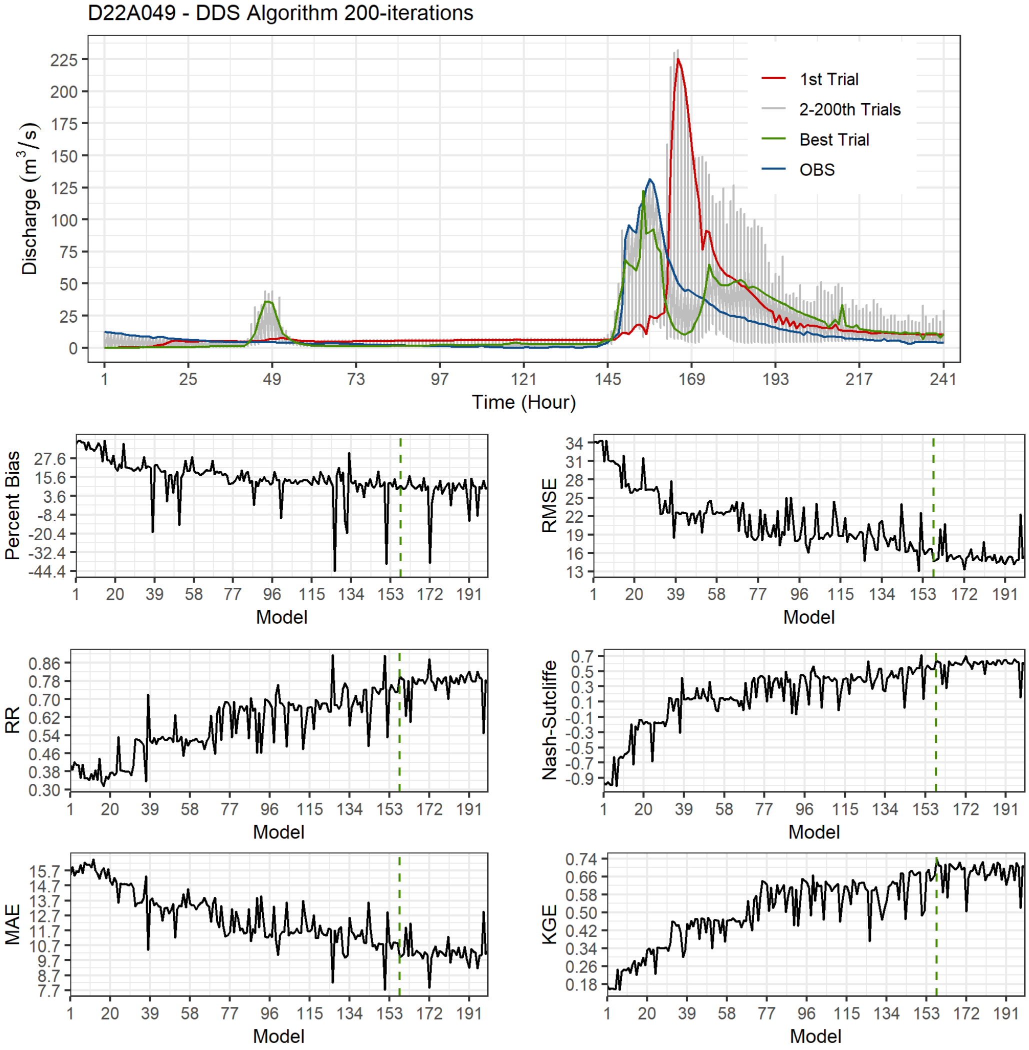


Figure 31 The calibration hydrographs performed with a total of 200 iterations using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

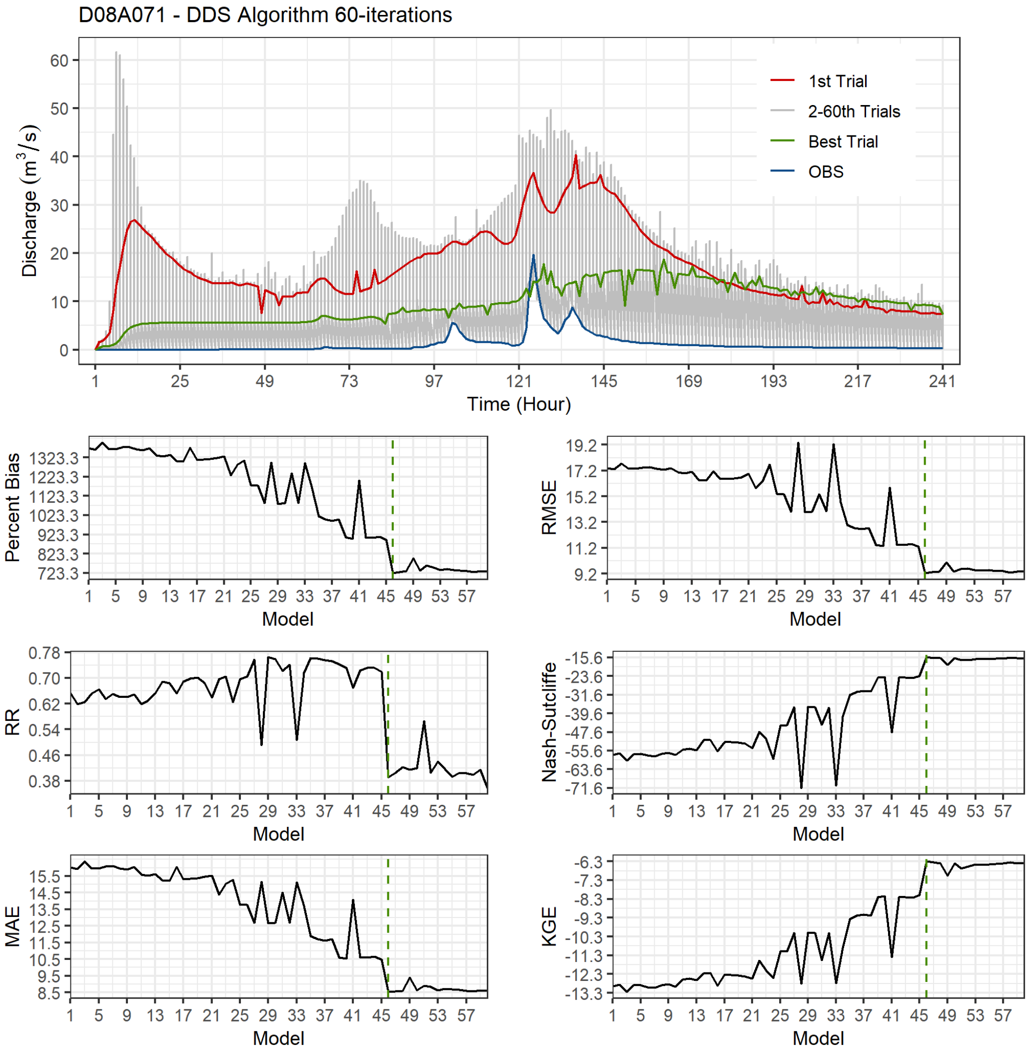


Figure 32 The calibration hydrographs performed with a total of 60 iterations using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

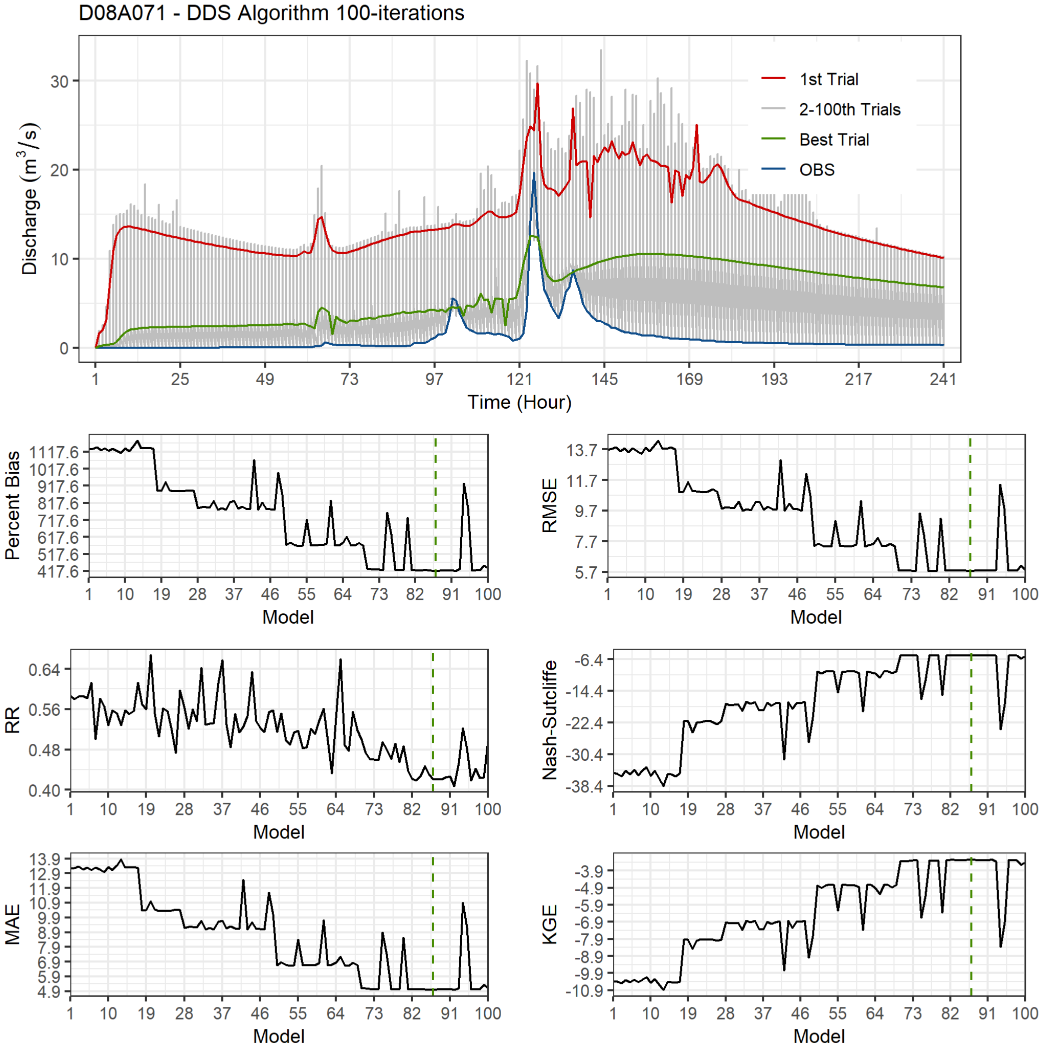


Figure 33 The calibration hydrographs performed with a total of 100 iterations using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

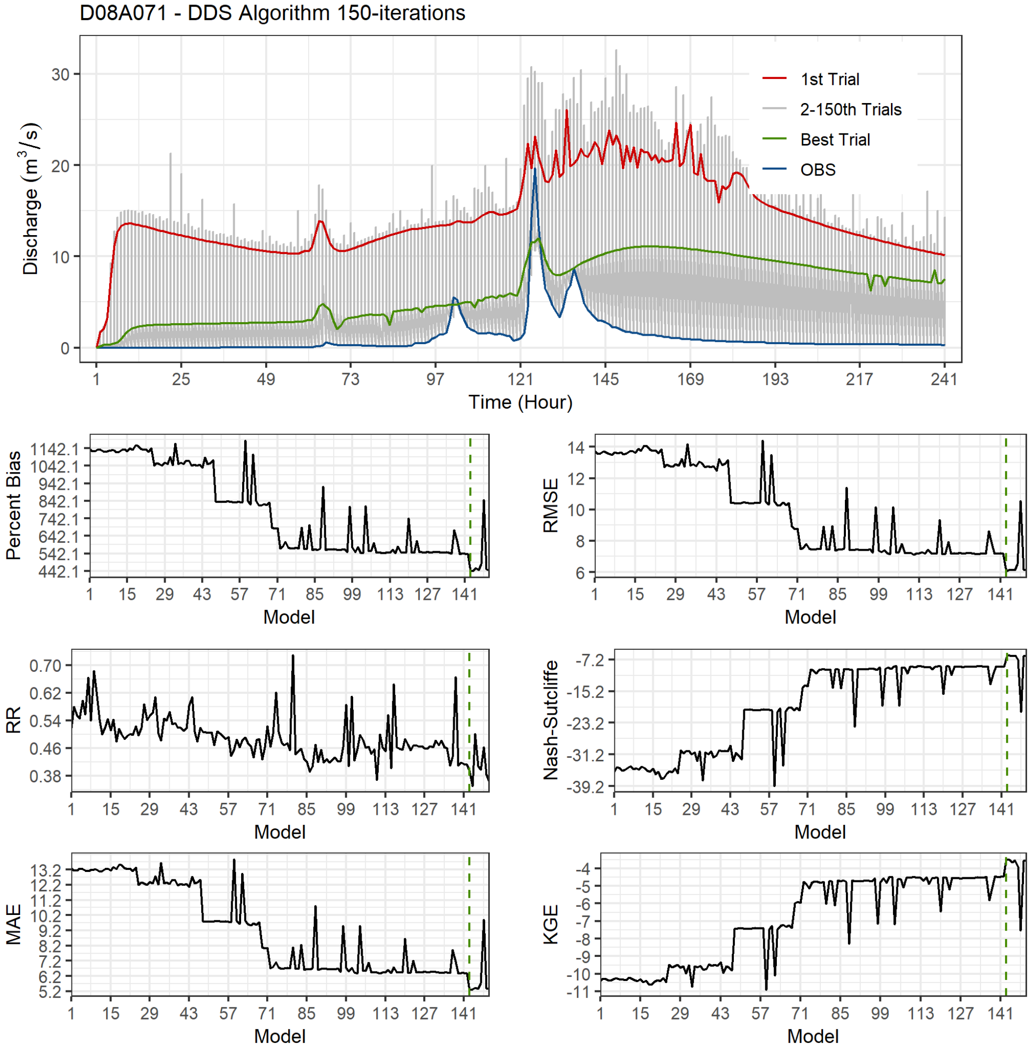


Figure 34 The calibration hydrographs performed with a total of 150 iterations using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

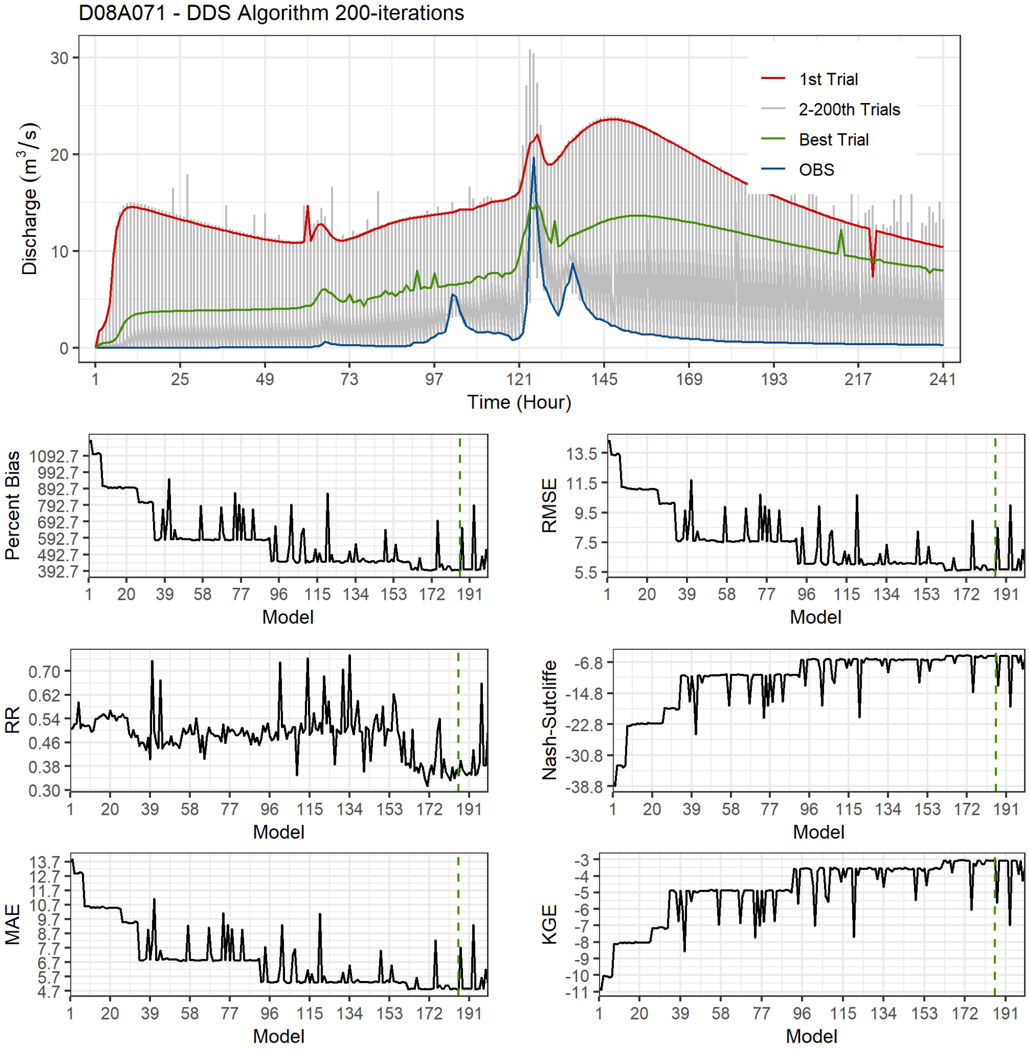


Figure 35 The calibration hydrographs performed with a total of 200 iterations using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

Table 6 Statistical metrics (PBias, RMSE, RR, NSE, KGE) of the best results achieved in automated calibration runs using different iteration numbers (60, 100, 150, 200)

Region	Station	Iteration Number	PBias	RMSE	RR	NSE	KGE
EBS	D22A049 (Arhavi)	60	-2.00	16.08	0.76	0.56	0.58
		100	-14.0	14.48	0.81	0.64	0.70
		150	22.0	18.84	0.71	0.40	0.63
		200	10.0	14.62	0.80	0.64	0.74
MED	D08A071 (Kemer)	60	72.3	9.26	0.39	-15.54	-6.30
		100	41.8	5.80	0.42	-5.42	-3.27
		150	44.2	6.07	0.39	-6.12	-3.49
		200	39.3	5.57	0.35	-4.99	-2.99

3.2.1.2.3 Processes-based Automated Calibration

A similar calibration approach suggested by Yucel et al. (2015) was utilized in this section while considering the impacts of parameters on hydrological behavior or in other words the water balance and the temporal distribution of the water amount throughout the time series (Yilmaz et al., 2008). The calibration parameter set is divided into two groups based on how their impact on the resultant hydrographs. The first group consists of the ones affecting the hydrograph volume (REFKDT, RETDEPRTFAC, and SLOPE), and the second group consists of those affecting the timing of the hydrograph (OVROUGHRTFAC, MANN, and LKSATFAC). First, the group of volume parameters is calibrated with the different iteration numbers (60, 100, 150, and 200 iterations) selected in the previous section. After this volume set is calibrated, the calibrated values for the first group of parameters are fixed. At this stage, only the second group of parameters are calibrated with the same iteration numbers, while the values of the parameters in the volume group are fixed within the model. The same objective function (i.e., KGE) is selected for the processed-based automated calibration. With this approach, it is proposed that automated calibration can be performed while taking into account the major impact of the parameters on the hydrological processes instead of following the method of calibrating the whole set of parameters at once. The same sub-basins and the events from the previous section are selected for the process-based calibration as Arhavi basin and the event that occurred between 10/19/2016-10/29/2016 are selected for the EBS region, whereas the Kemer basin and the event that occurred between 03/07/2017-03/17/2017 for the MED region.

Results of the automated calibration in the Arhavi basin for the parameter set of volume are shown between Figure 36 and Figure 39. A gradual improvement is observed in most of the statistical metrics while the iteration number increases (Table 7). The KGE value changes between 0.68 and 0.75. Especially for the 150- and 200-iterations, a substantial improvement is seen in the event peak in terms of the volume. However, the model is creating a second peak at the recession part instead of reaching the baseflow level again. From this test, the parameter set of the best trial of 150-iteration runs is selected, and the calibration process is continued with the set of timing parameters (Figure 40-Figure 43). A 12% improvement is observed in the KGE value for the 60- and 150-iterations. Besides, there is still a recession issue in the falling limb part of the hydrographs. On the other hand, for the 100- and 200-iterations, there is a drastic decrease in the performance statistics (Table 7).

Figure 44 to Figure 47 show the process based calibration results of over the Kemer basin located in MED region. In general, for this event there is an overestimation problem throughout the simulated hydrographs, as mentioned in the previous section. After the volume parameter set is calibrated, the set of 60-iteration runs is selected for the calibration of the timing set. All calculated NSEs and KGEs show negative values being close to each other. 150-iterations shows the best results in terms of KGE and Pbias values, while for the other statistical metrics 200-iterations cretaes the best results.

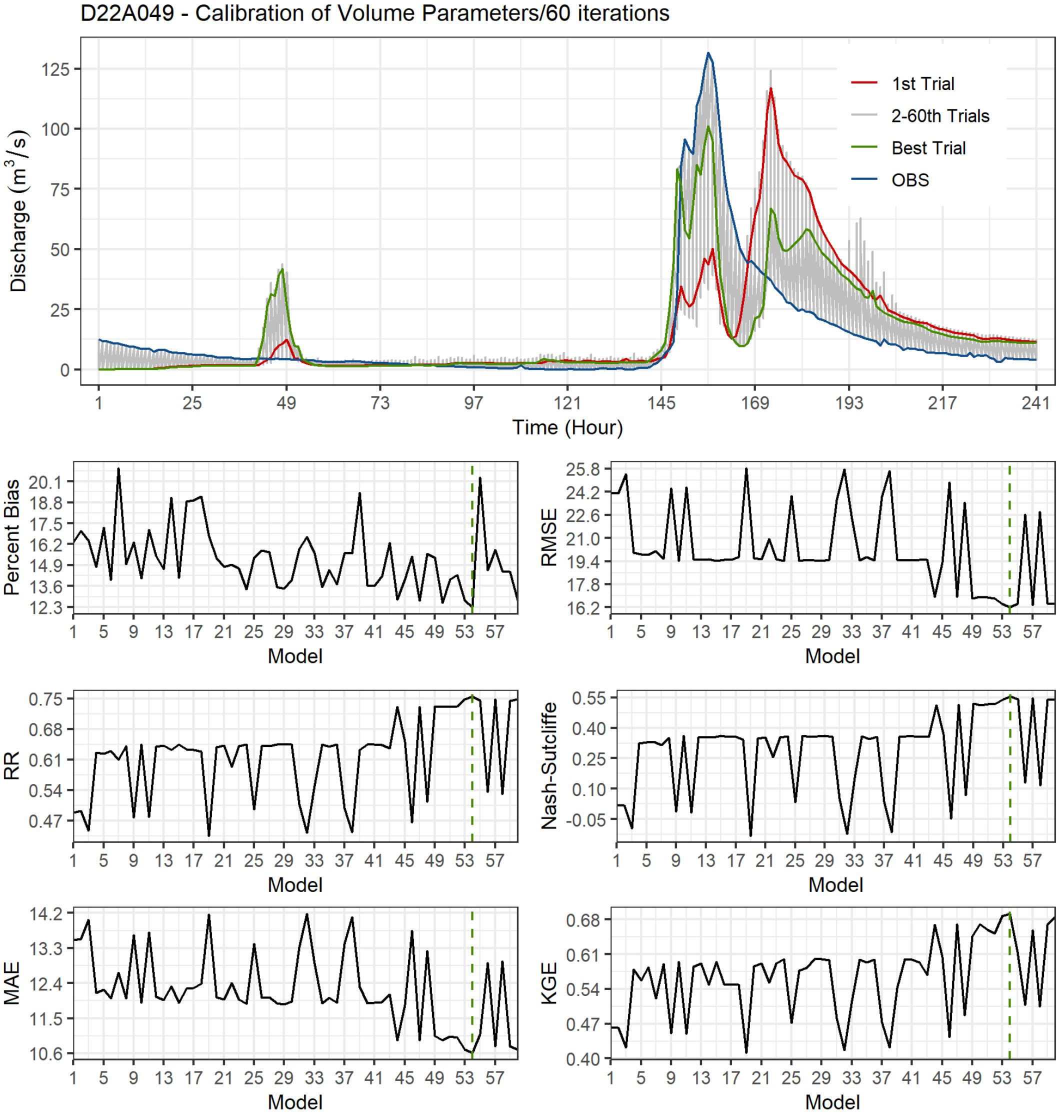


Figure 36 The calibration hydrographs performed with a total of 60 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

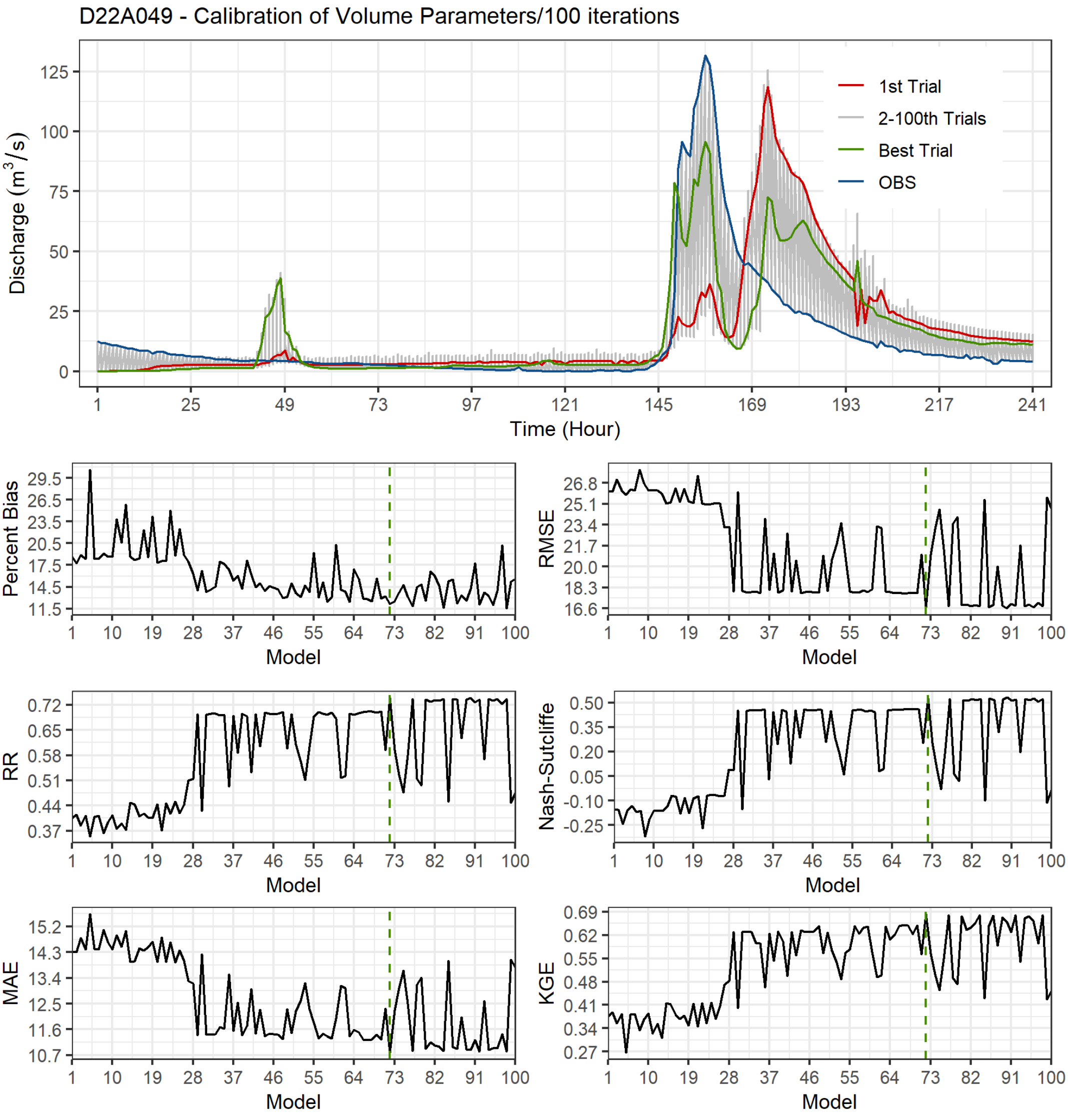


Figure 37 The calibration hydrographs performed with a total of 100 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

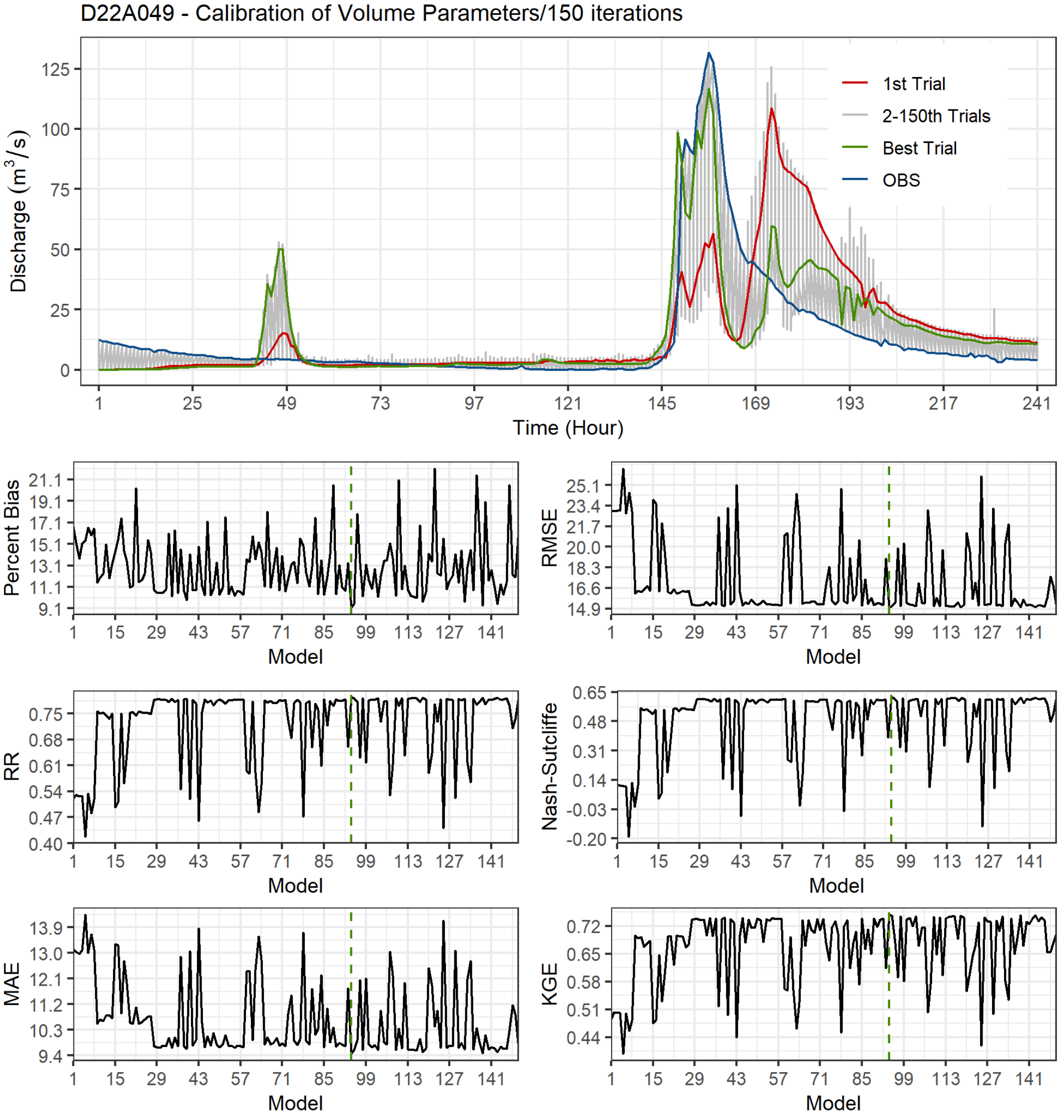


Figure 38 The calibration hydrographs performed with a total of 150 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

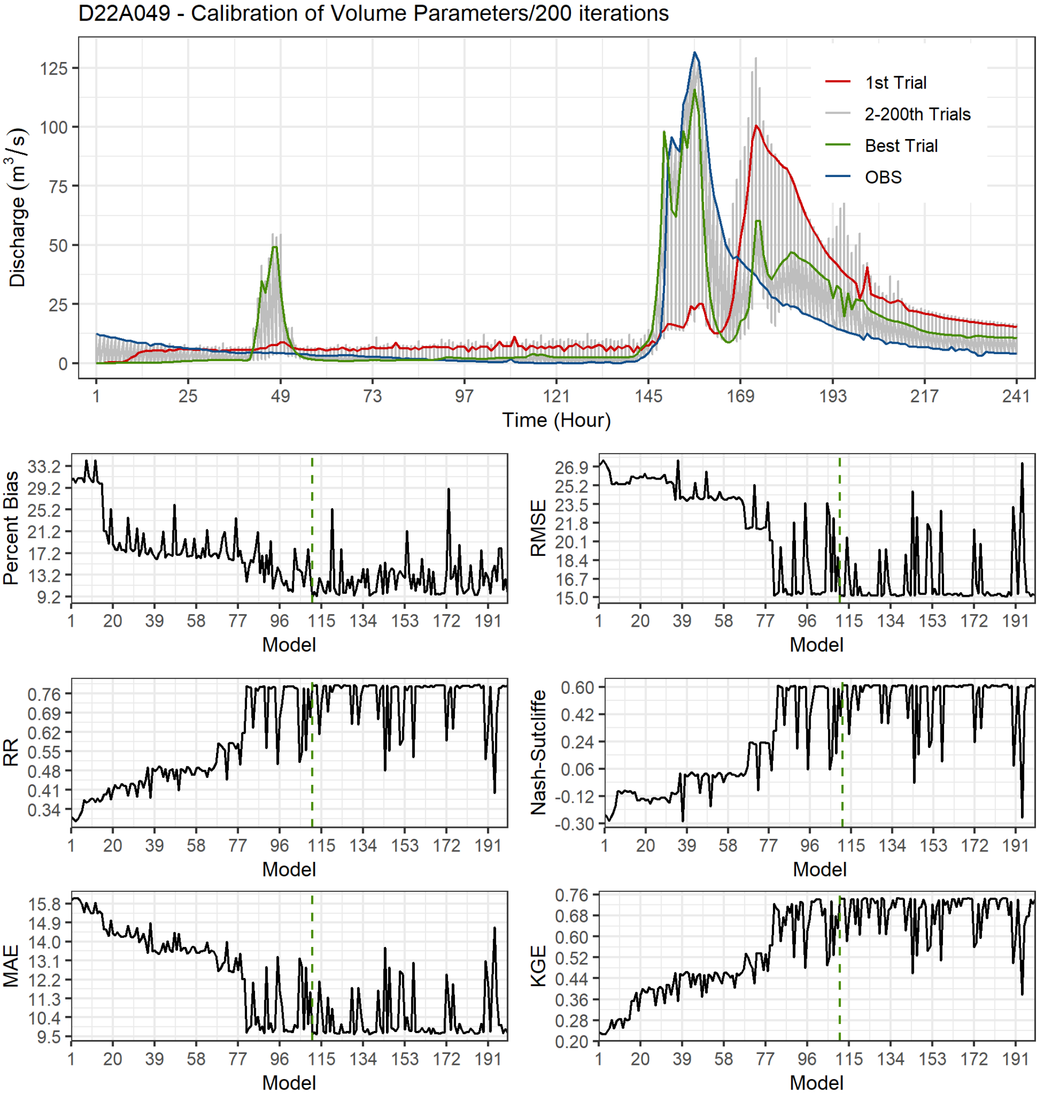


Figure 39 The calibration hydrographs performed with a total of 200 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

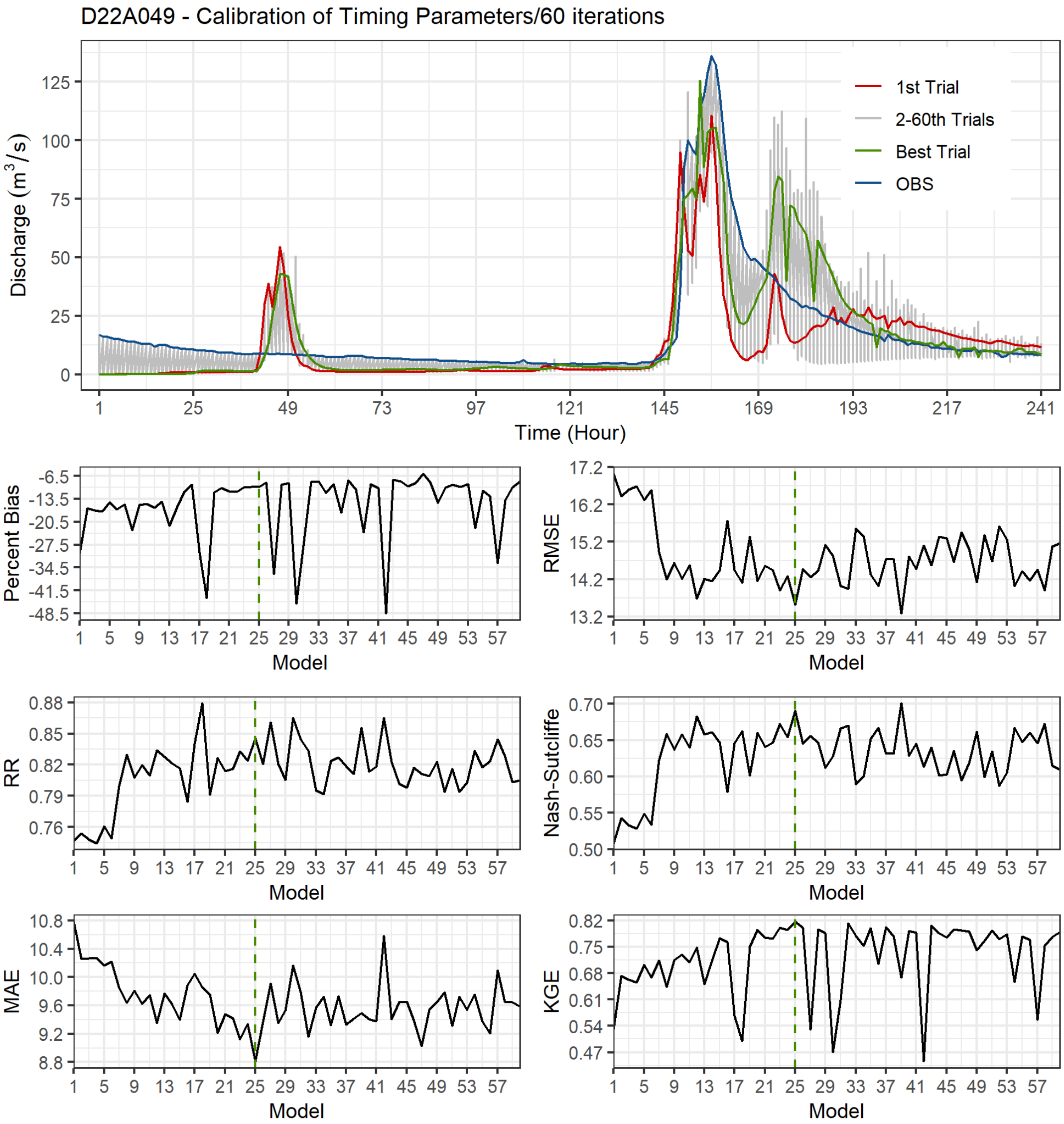


Figure 40 The calibration hydrographs performed with a total of 60 iterations and volume parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

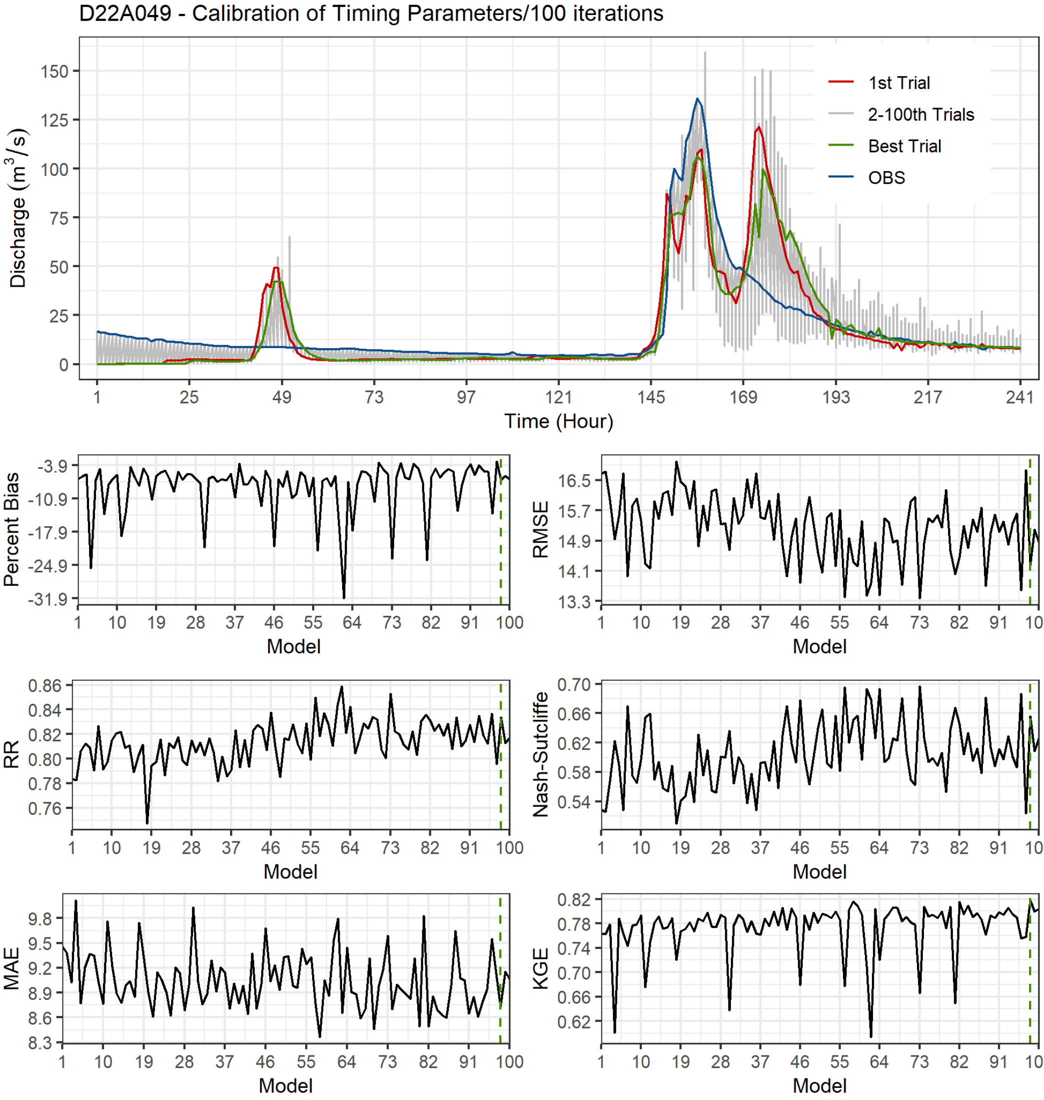


Figure 41 The calibration hydrographs performed with a total of 100 iterations and volume parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

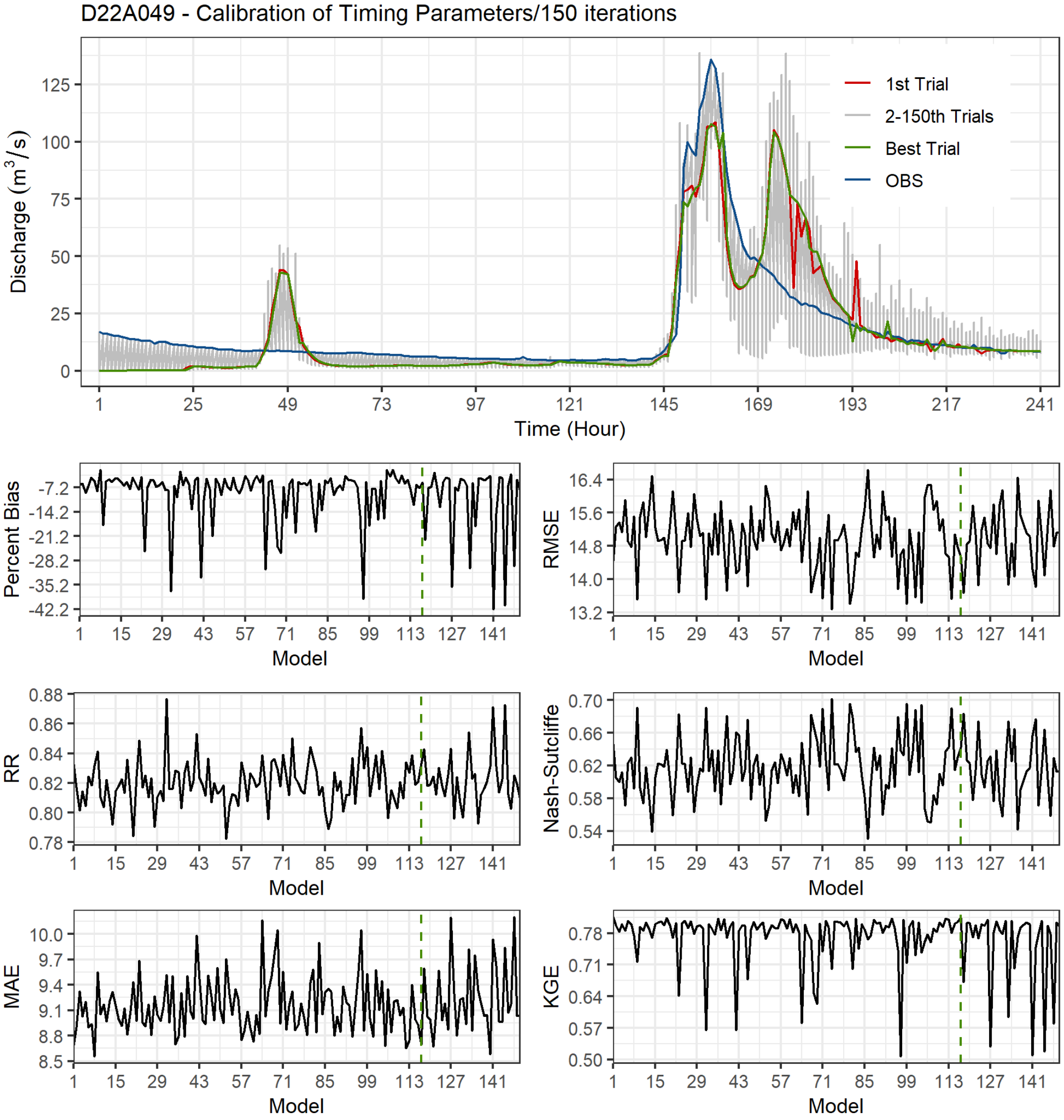


Figure 42 The calibration hydrographs performed with a total of 150 iterations and volume parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

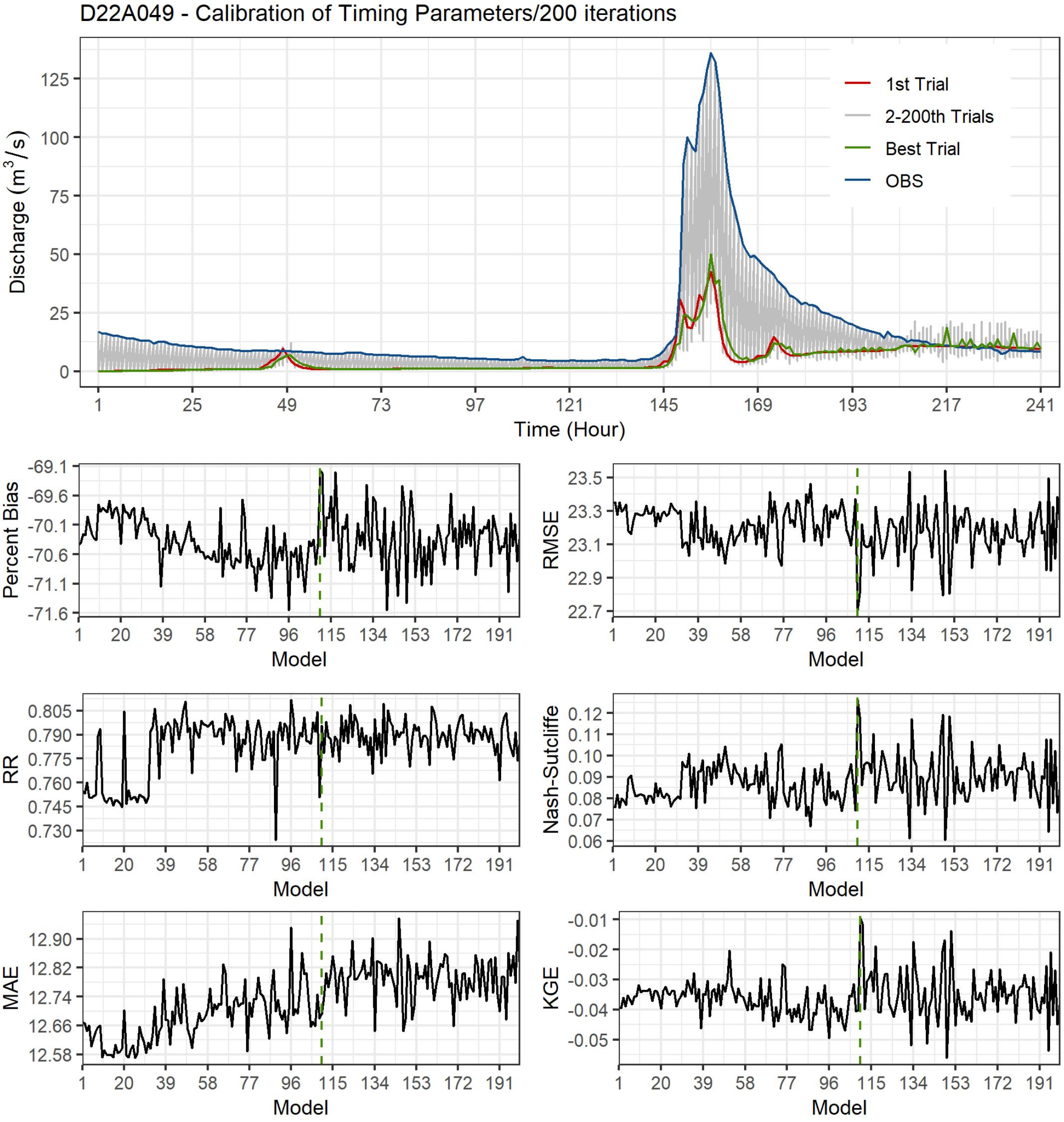


Figure 43 The calibration hydrographs performed with a total of 200 iterations and volume parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Arhavi basin and the event that took place between 10/19/2016-10/29/2016, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

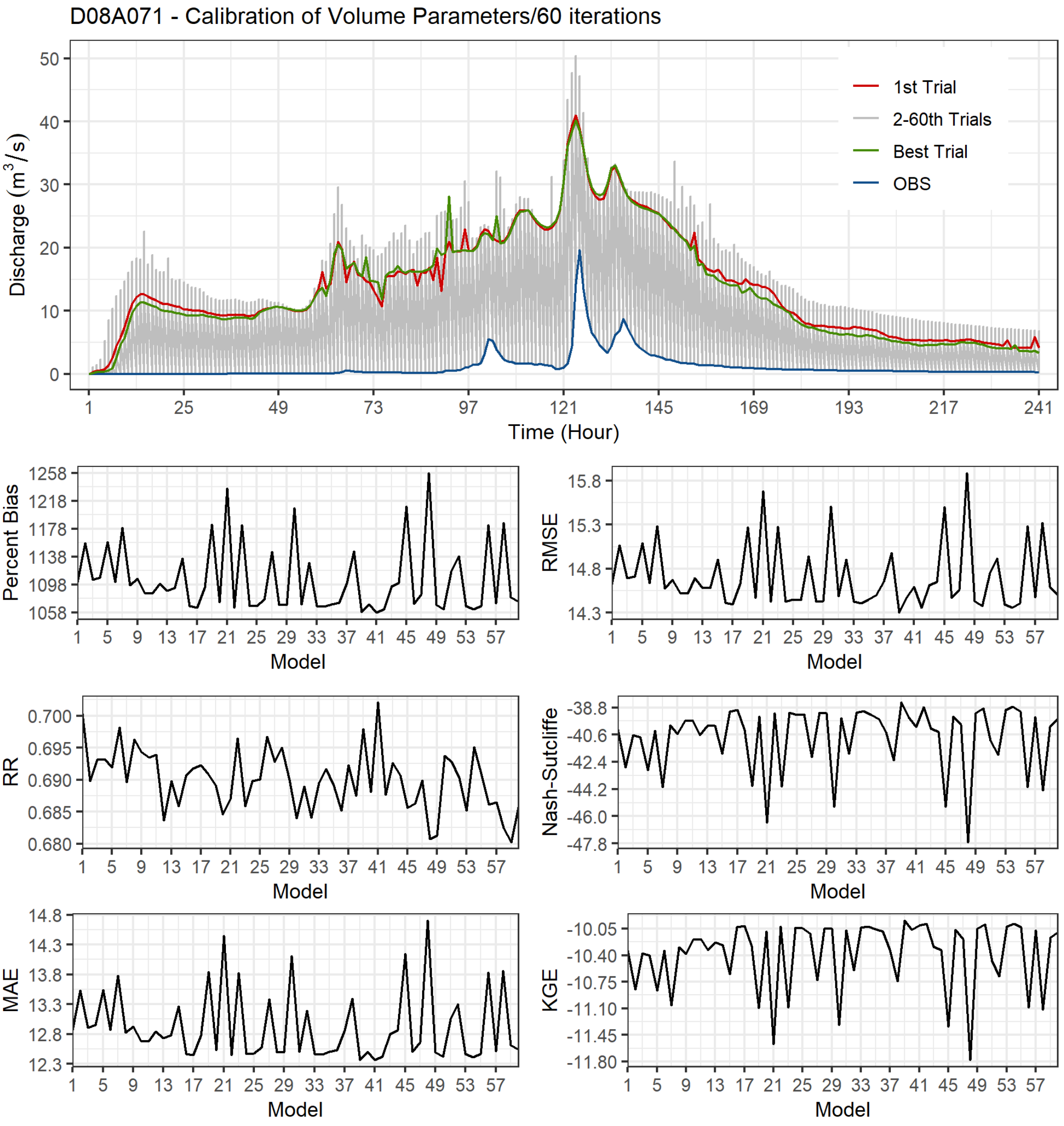


Figure 44 The calibration hydrographs performed with a total of 60 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

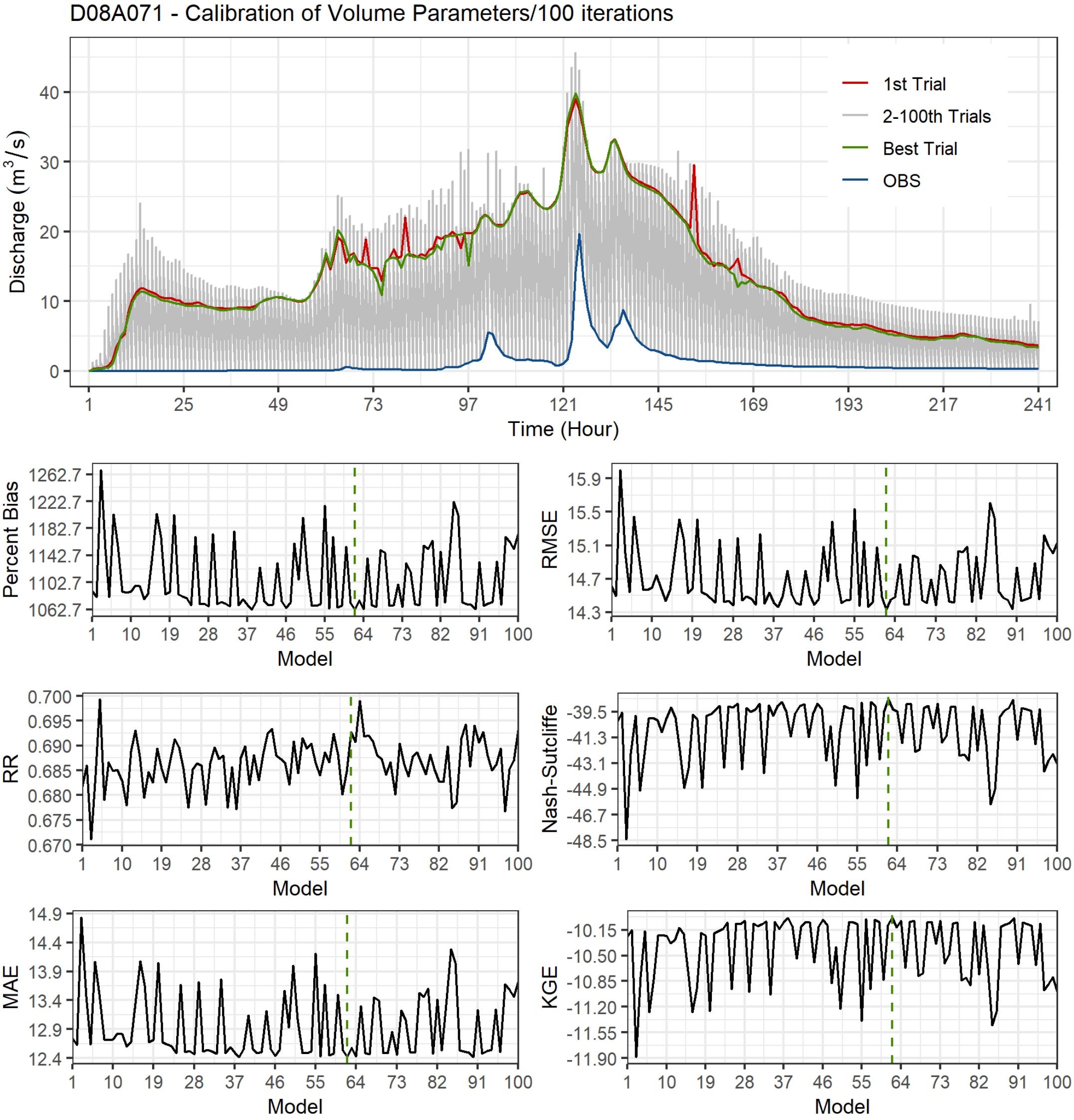


Figure 45 The calibration hydrographs performed with a total of 100 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

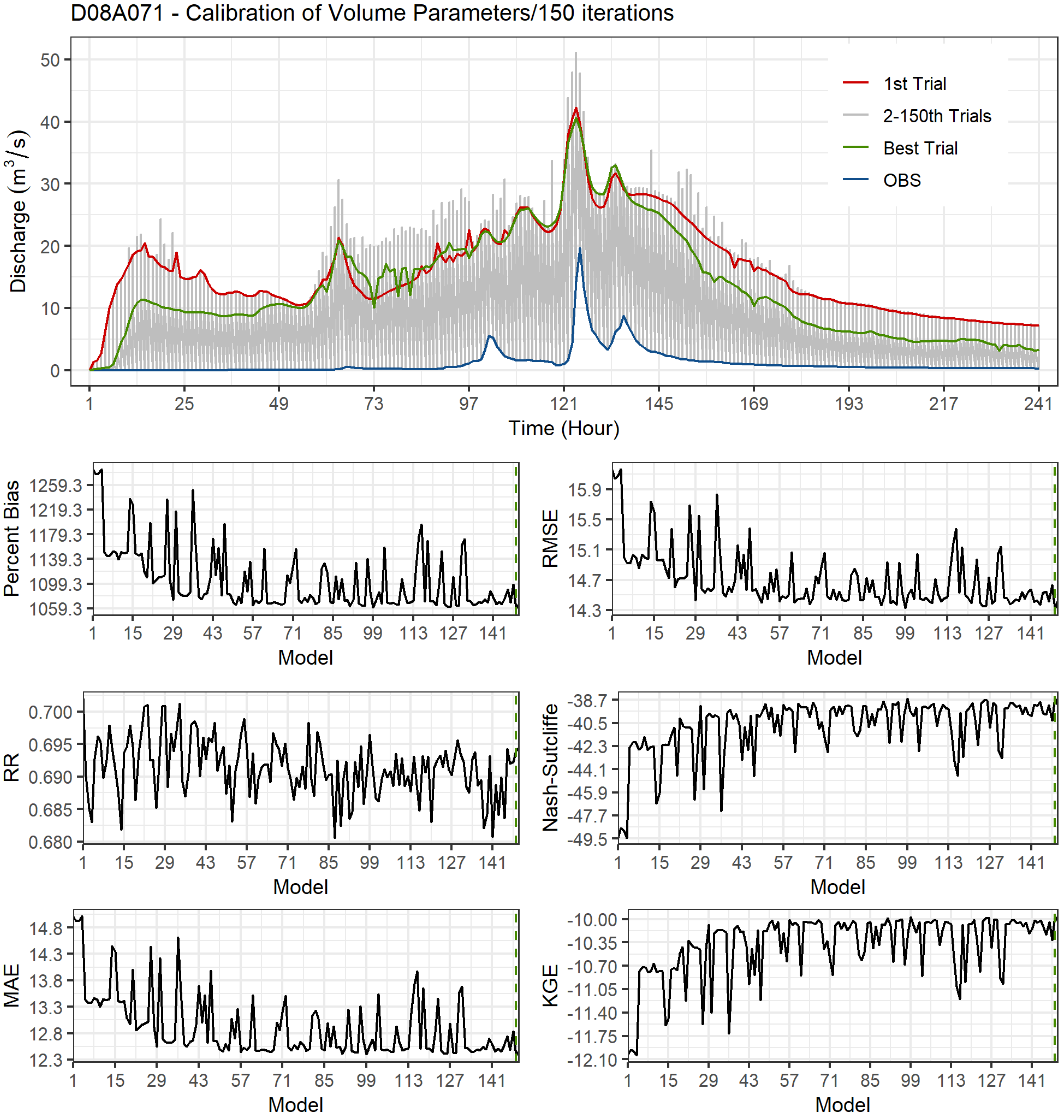


Figure 46 The calibration hydrographs performed with a total of 150 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

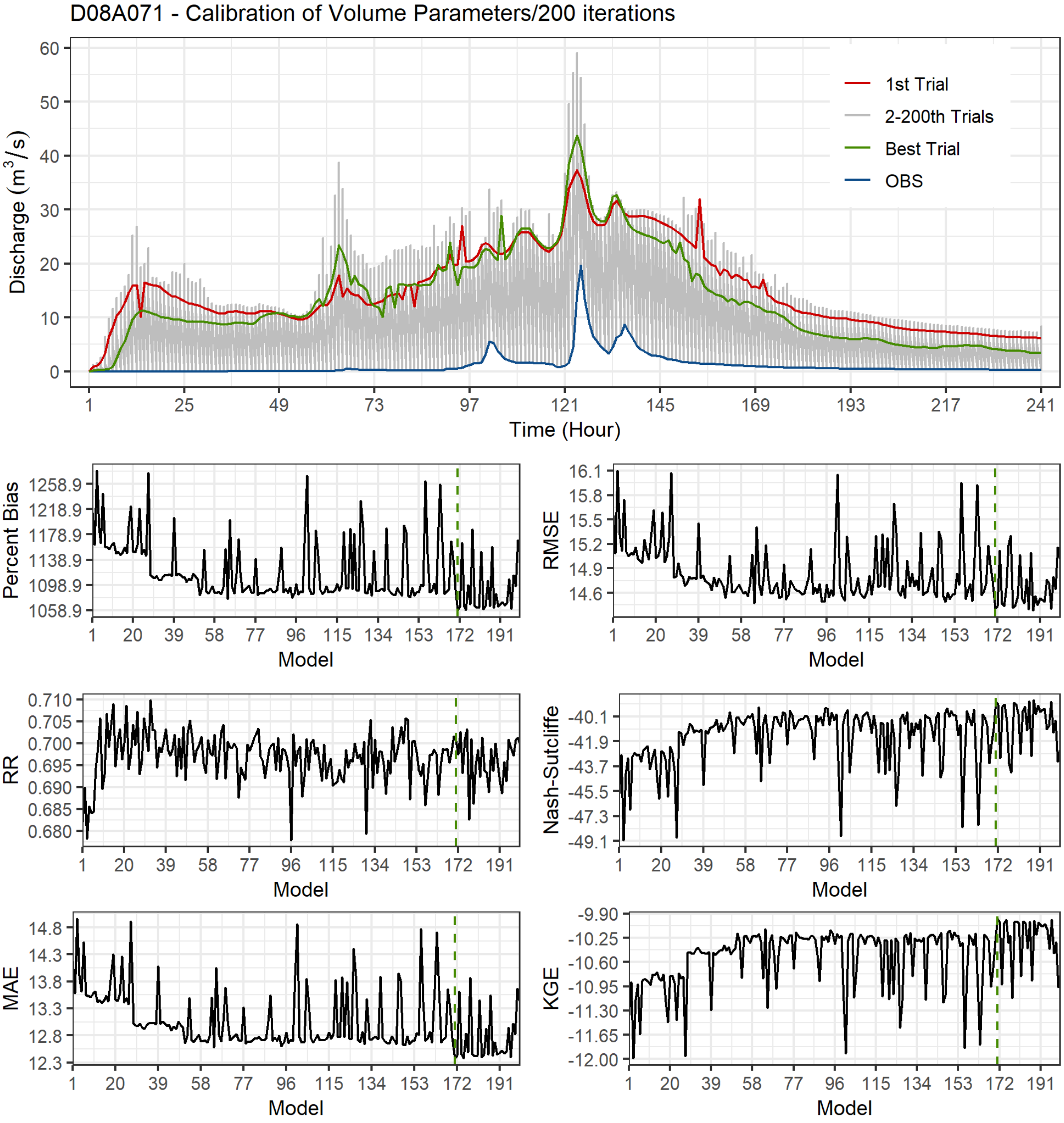


Figure 47 The calibration hydrographs performed with a total of 200 iterations and volume parameters (REFKDT, RETDEPRTFAC, and SLOPE) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

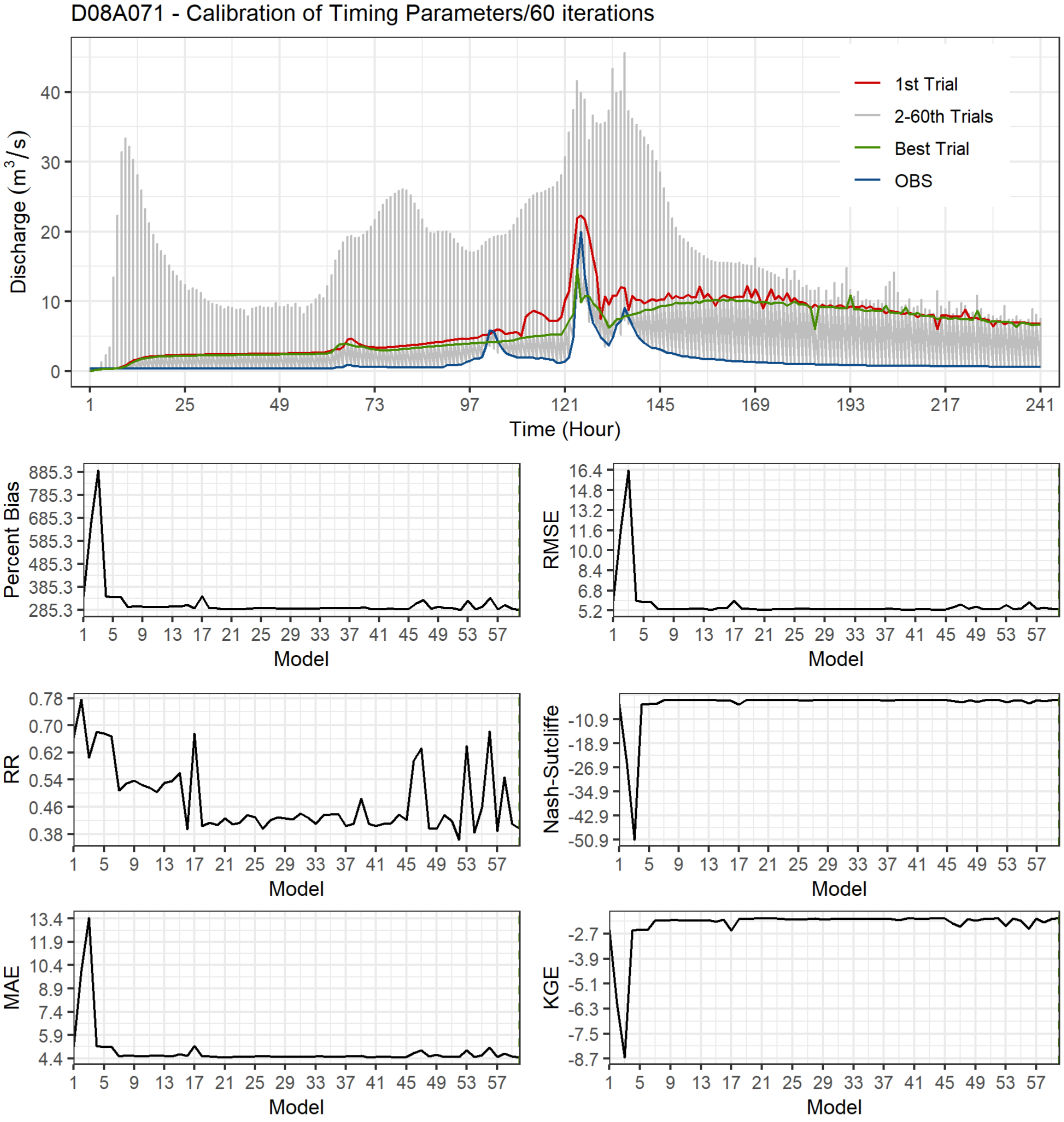


Figure 48 The calibration hydrographs performed with a total of 60 iterations and timing parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

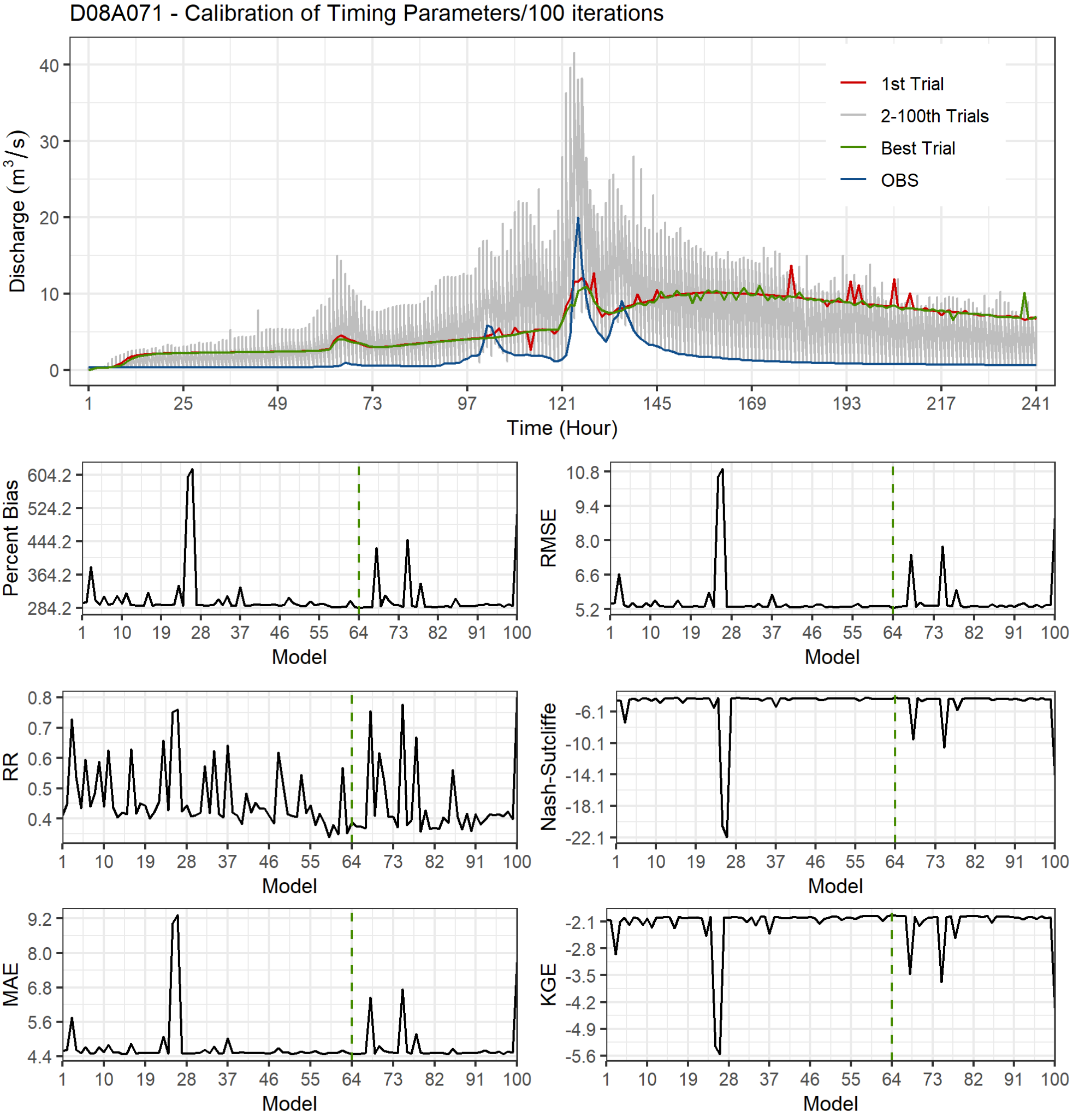


Figure 49 The calibration hydrographs performed with a total of 100 iterations and timing parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

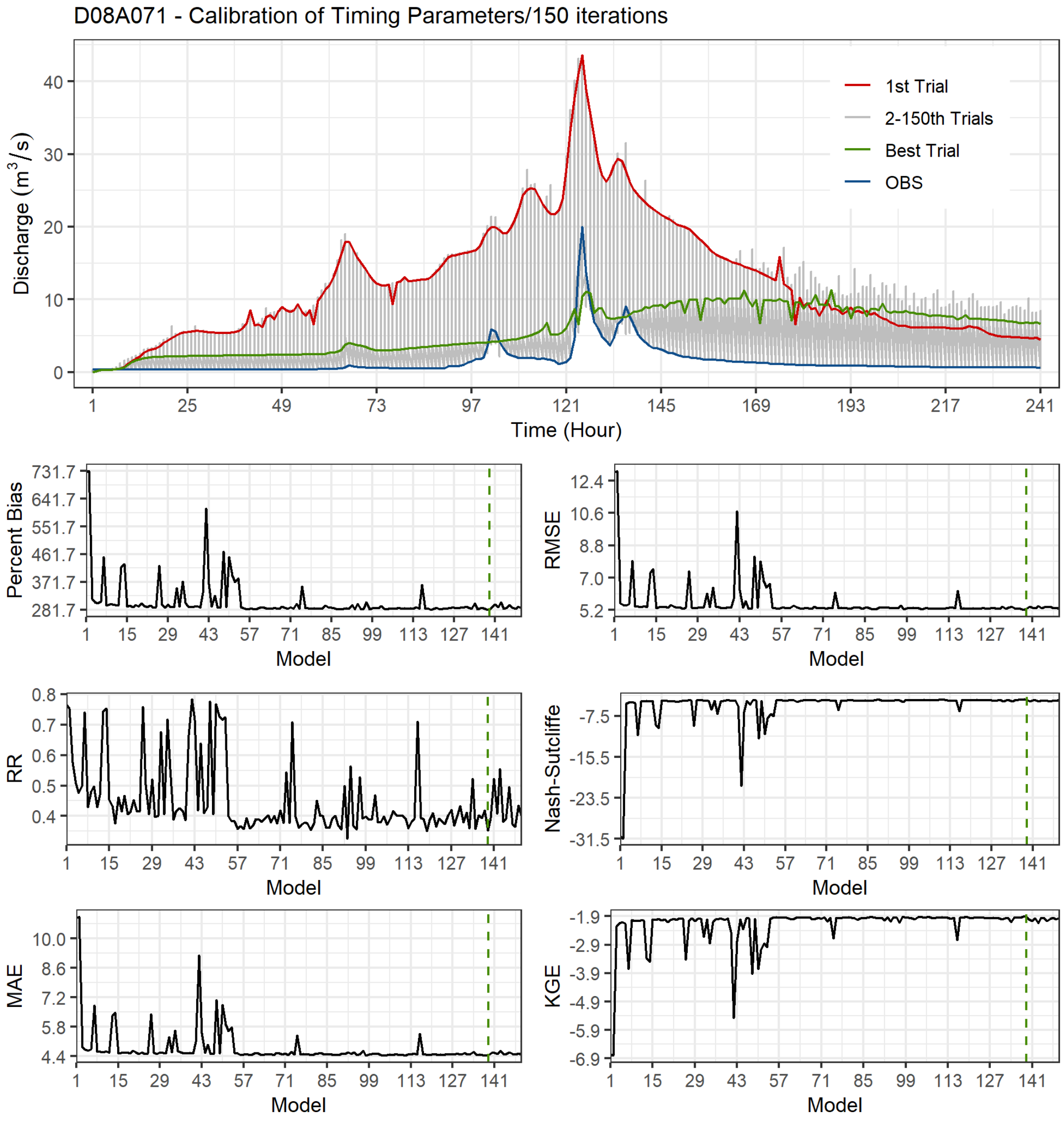


Figure 50 The calibration hydrographs performed with a total of 150 iterations and timing parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

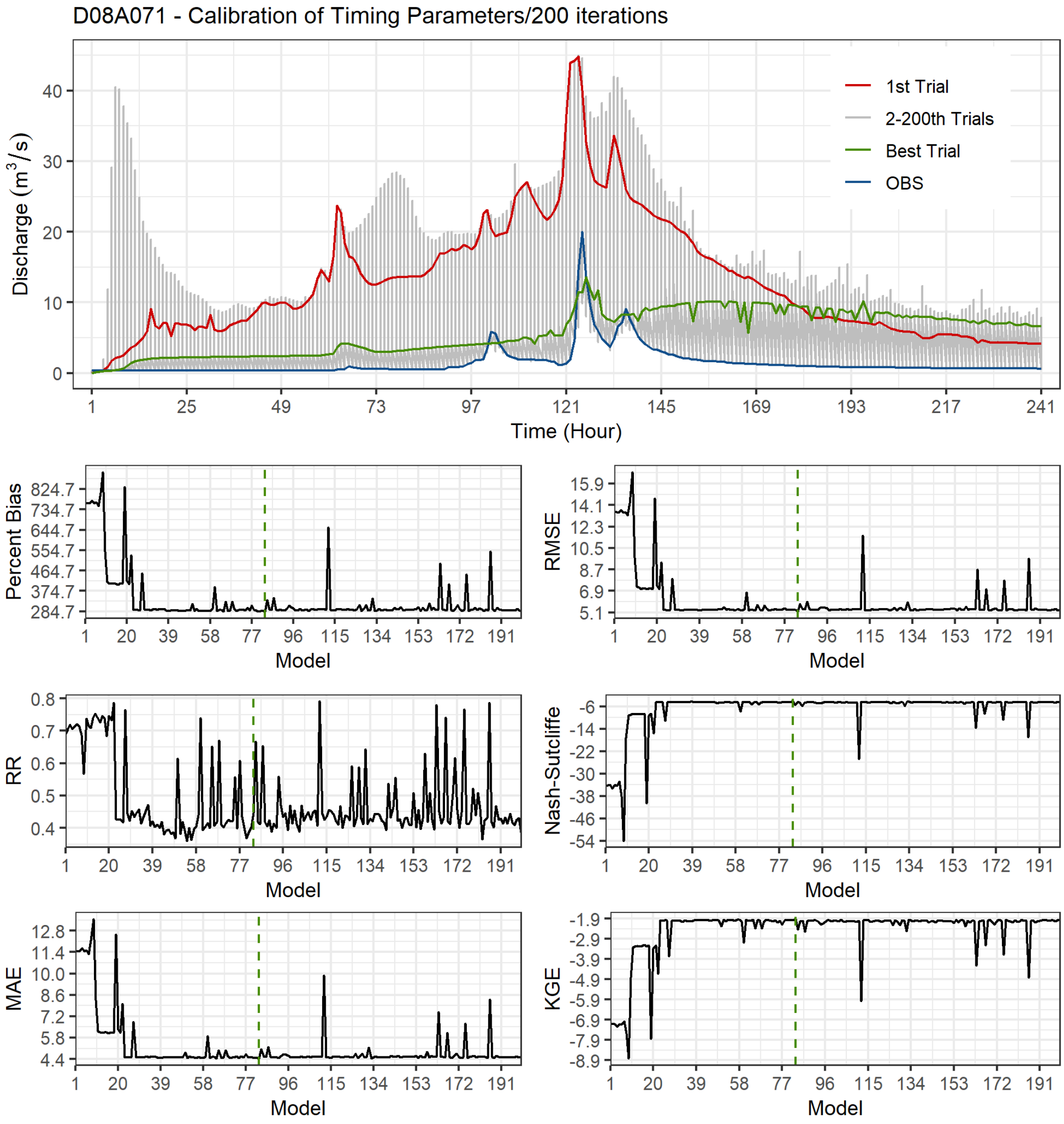


Figure 51 The calibration hydrographs performed with a total of 200 iterations and timing parameters (OVROUGHRTFAC, MANN, and LKSATFAC) using the DDS algorithm for the Kemer basin and the event that took place between 03/07/2017-03/17/2017, and the statistical indices calculated for each iteration (Percent Bias, Root Mean Square Error (RMSE), Correlation Coefficient (RR), Nash-Sutcliffe Efficiency, Mean Absolute Error (MAE), Kling-Gupta Efficiency (KGE))

Table 7 Statistical metrics (PBias, RMSE, RR, NSE, KGE) of the best results achieved in automated calibration runs using different iteration numbers (60, 100, 150, 200) by means of the process-based approach.

Region	Station	Parameter Set	Iteration Number	PBias	RMSE	RR	NSE	KGE
EBS	D22A049 (Arhavi)	Volume	60	12.3	16.23	0.75	0.55	0.69
			100	12.2	16.67	0.74	0.53	0.68
			150	9.19	14.99	0.79	0.62	0.75
			200	9.20	15.09	0.79	0.61	0.75
		Timing	60	-10.0	13.52	0.85	0.69	0.82
			100	-52.9	15.47	0.88	0.59	0.39
			150	-6.45	14.44	0.83	0.65	0.82
			200	-6.92	22.71	0.80	0.13	-0.01
MED	D08A071 (Kemer)	Volume	60	10.58	14.30	0.70	-38.47	-9.94
			100	10.63	14.33	0.69	-38.64	-9.98
			150	10.59	14.32	0.69	-38.58	-9.96
			200	10.59	14.41	0.70	-39.04	-9.97
		Timing	60	28.6	5.29	0.40	-4.41	-1.95
			100	28.4	5.25	0.39	-4.32	-1.93
			150	28.2	5.26	0.35	-4.34	-1.91
			200	28.5	5.19	0.45	-4.20	-1.93

3.2.1.3 Comparison of the Calibration Approaches

DDS algorithm with calibration of the whole parameter set gives better performance in the EBS region compared to manual calibration results (Stepwise KGE=0.40 and NSE=0.16), while the manual calibration in the MED region gives slightly better results (Stepwise KGE=-0.75, NSE=-1.26) than the DDS algorithm (Table 8 & Table 9). While both calibration approaches give negative results for KGEs and NSEs in the MED region, manual calibration decreases these values slightly. It is demonstrated that the number of iterations used to reach the optimum values in DDS is an effective factor in the performance of the algorithm.

Compared to the DDS algorithm results, including the whole parameter set, the model results are enhanced by ~%11 in KGE and ~%8 in NSE in processed-based calibration for the Arhavi basin.

In Kemer basin, the process-based calibration is improves the KGE results by ~30 % obtained by the calibration of the whole parameter set (Table 8). However, the stepwise calibration improvement steps forward compared to the automated approaches in terms of the statistics (Stepwise KGE=-0.75, NSE=-1.26) in the Kemer basin (Table 8 & Table 9). For both basins, it can be inferred that the process-based automated calibration approach shows better performance compared to the automated calibration of the whole parameter set. Arhavi basin shows better performance in terms of automated calibration, while the stepwise calibration approach improves the model results more compared to the automated calibration techniques in Kemer basin.

Table 8 Comparison of the KGE performance values for calibration with stepwise, DDS-whole set and DDS- processed-based approaches.

Region	Station	Calibration Method	Number of Iterations			
			60	100	150	200
EBS	D22A049 (Arhavi)	Stepwise	0.40	-	-	-
		DDS-whole set	0.58	0.70	0.63	0.74
		DDS-processed based	0.82	0.39	0.82	-0.01
MED	D08A071 (Kemer)	Stepwise	-0.75	-	-	-
		DDS-whole set	-6.30	-3.27	-3.49	-2.99
		DDS-processed based	-1.95	-1.93	-1.91	-1.93

Table 9 Comparison of the NSE performance values for calibration with stepwise, DDS-whole set and DDS- processed-based approaches.

Region	Station	Calibration Method	Number of Iterations			
			60	100	150	200
EBS	D22A049	Stepwise	0.16	-	-	-
		DDS-whole set	0.56	0.64	0.40	0.64
		DDS-processed based	0.69	0.59	0.65	0.13
MED	D08A071	Stepwise	-1.26	-	-	-
		DDS-whole set	-15.54	-3.27	-3.49	-2.99
		DDS-processed based	-4.41	-4.32	-4.34	-4.20

3.2.2 Evaluation of SST Events for WRF-Hydro Model

WRF-Hydro model performances of the simulations forced by the WRF meteorological data generated with different SST datasets are examined in this section. The calibrated parameter set with the manual stepwise calibration method is utilized for Hopa (D22A147) and Kemer (D08A071) basins. The simulated hydrograph based on the observation precipitation forcing is also compared with the hydrographs forced by the WRF model to be able to see the errors caused by the interpolation process and the model itself.

The calibrated parameter set which is obtained by the manual stepwise approach, is validated for SST events in Hopa and Kemer basins. Table 4 shows the statistical measures calculated for calibration and validation events in both basins. Hydrological response to selected SST events over Hopa and Kemer basins are simulated, and the hydrographs derived by various SST simulations and observed precipitation are demonstrated in Figure 52. Based on the simulated hydrograph of the Hopa basin, it is observed that the ERA5 and NCEP simulations are underestimated compared to the observed hydrograph (Figure 52 (a)). The negative bias found for the WRF precipitation simulations in Figure 7 (a) may explain this understimation issue for ERA5 and NCEP (datasets showing the highest bias error). The hydrograph volumes produced by the Medspiration simulation and the observed precipitation simulation are marginally better than those produced by the ERA5 and NCEP simulations. Despite the fact that the GHRSST simulation overestimates precipitation and skips the peak time for Hopa, the daily mean discharge of the GHRSST is the most accurate simulation in predicting discharges. Because the WRF simulation of the GHRSST provided the most accurate volume of water that the Hopa received on peak day, as seen in Figure 8 (c). Accordingly, the daily average of water volume conveyed to the channel creates the closest volume to the observed one with the lowest negative bias and RMSE values. The NSE value of this simulation is calculated as 0.63, and the KGE value is calculated as 0.03. It is observed that the correlation coefficient value is increased from 0.42 to 0.80 on average for the simulations performed with high-resolution SST datasets compared to the the ERA5 simulation. Likewise, the GHRSST simulation shows an improvement of approximately 70 % in NSE compared to ERA5 simulation (Table 4). In line with the results of cross-correlation calculations, it is determined that GHRSST is the best SST product that represents the Arhavi basin and the corresponding event.

In the Kemer basin, it is seen that the GFS simulation overestimated the peak value and missed the timing trend of the observation hydrograph (812.9 m^3/s) (Figure 52 (b)). The simulations derived using high-resolution SST and observation precipitation shows that although the hydrographs follow a similar trend with the observation in the falling limb part, they decrease abruptly in the rising limb part. Minor delays are observed in the primary peak time for the high-resolution SST simulations. GFS SST hydrograph has the lowest bias (-24.98 m^3/s) but the largest RMSE (129.81) and the lowest correlation coefficient (0.18) (Table 4). A better representation of hydrograph peak timing and volume is seen for the meteorological forcings updated by high-resolution SST datasets, as parallel to the closer spatial distribution of precipitation to the observation at peak time over the Kemer basin (Figure 9). In particular, the simulation forced with the observed precipitation correctly identifies the time and magnitude of the first peak but produces a negative bias in the volume of the second peak. It creates a hydrograph volume bias of -53.48 m3/s compared to the high-resolution SST simulations, however, it achieves a better correlation (0.40). High-resolution SST simulations improve the ERA5 results by ~550 m^3/s volume reduction and successfully capture the peak timing.

In both basins, the findings are consistent with the accuracy analysis based on the mean cross-correlation of each SST product. When the influence of the GHRSST and Medspiration products on the WRF-Hydro model performance is examined, the highly cross-correlated GHRSST and Medspiration products also provide the best results in the model among other simulations.

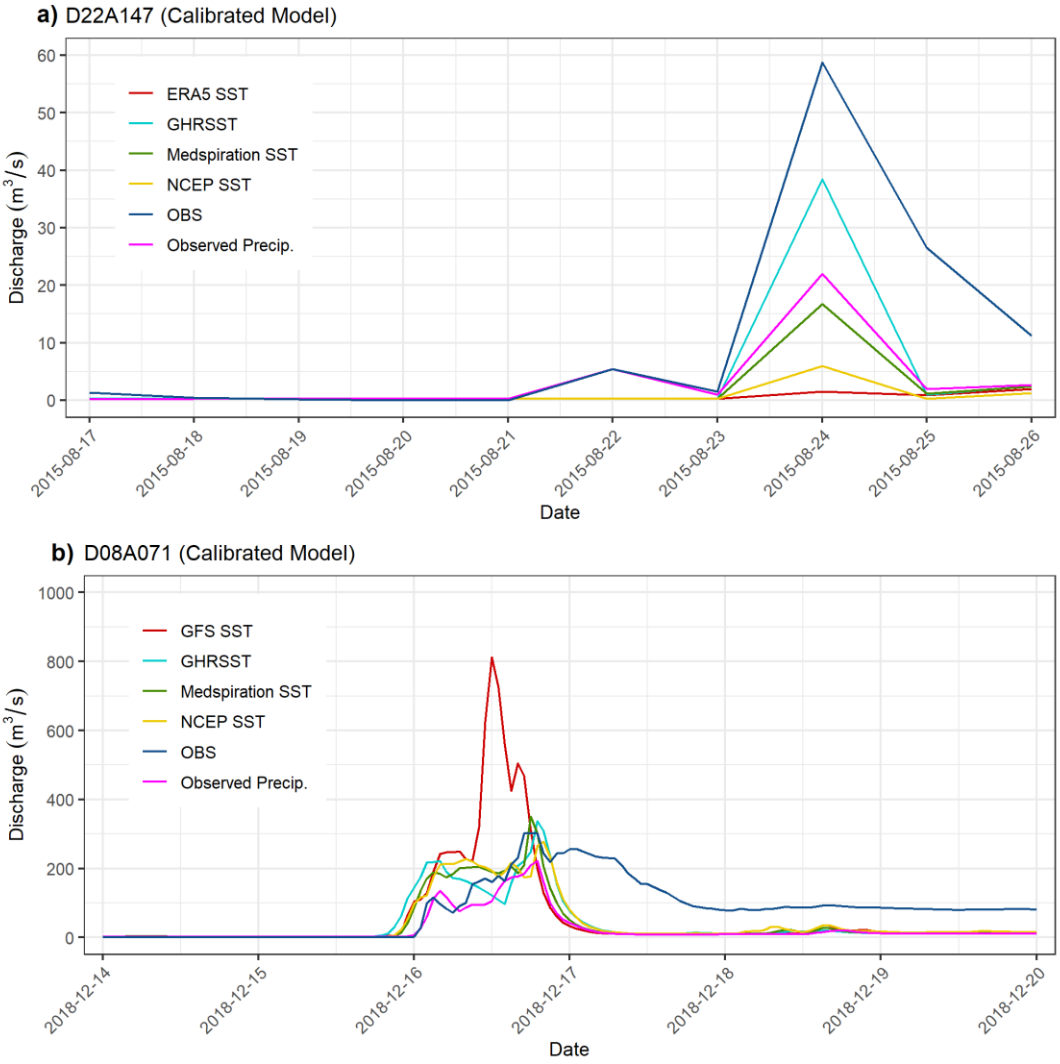


Figure 52 a) 17/08/2015-27/08/2015 event and D22A049 basin b) 03/07/2017-03/17/2017 (last six days) event and D08A071 basin observation hydrograph, modeled hydrographs with the precipitation data generated using ERA5/GFS GHRSST, Medspiration, NCEP SST products, and the modeled hydrograph with the precipitation data based on the interpolation of the observation points.

3.2.3 Evaluation of Rainfall-Runoff Spatial Distributions

Accumulated precipitation estimates from four different SST simulations and the runoff estimates on channel grid cells calculated by the WRF-Hydro model forced with the forcing of these SST simulations derived by the WRF model are matched and displayed gridded wise for each region in Figure 53 and Figure 54. WRF model precipitation output at 1 km resolution and the streamflow output of the WRF-Hydro model over the defined channel grids at 250 m resolution are overlapped in these figures.

Figure 53 represents the grid-based dynamic representation of ERA5, GHRSST, Medspiration, and NCEP simulations over the EBS region. The streamflow results (m^3/s) over the channel grids of the two basins (D22A049 and D22A147 stations from Figure 1) are plotted over the accumulated WRF precipitation (mm). Green dot denote basin outlets. The precipitation over the basins at the time step right before the precipitation event started is given in the first time step, 08/23/2015 23:00:00 (Figure 53 (a-d)). Streamflow distribution over the river networks shows the baseflow levels before the beginning of the event. In the second time step, it is seen that the highest precipitation depth is generated in the GHRSST simulation (Figure 53 (f)). This result is consistent with the remark from the previous section that the GHRSST is overestimating the SST event in the EBS region (Figure 7 (a)). Both basins react to heavy precipitation very quickly due to the hydrological characteristics of the basins related to the region`s topography (steepy-sloped). This response is explicitly seen in Figure 53 (e and g) for D22A049 in Figure 53 (f and h) for D22A147. For the last time step, streamflow values decrease due to less precipitation depth received by the basins compared to the previous time step. Baseflow conditions prevail again in some tributaries through the upstream parts of both river networks in one hour (Figure 53 (i and j)).

A grid-based dynamic representation of GFS, GHRSST, Medspiration, and NCEP simulations over the MED region is given in Figure 54. The streamflow results (m^3/s)

over the channel grids of the three basins (D08A071, D09A095, and E08A008 stations from Figure 1) are displayed over the accumulated WRF precipitation (mm). The first time step shows the baseflow conditions in streamflow and the accumulated precipitation distribution over the area at 02:00:00 on 12/16/2018, shortly before the SST event starts. In the second time step (16:00:00), there is a substantial increase in accumulated precipitation depth towards the eastern part of the area. This precipitation amount builds up and conveys through the channel grids of the D09A095 and D08A071 basins. For the D08A071 basin, an overestimated volume of 370 m^3/s discharge is generated and conveyed to the outlet in 14 hours as 516 m^3/s (Figure 54 (a and e)). In Figure 54 (k), a discharge value of 698 m^3/s is observed for the D09A095 basin due to the higher precipitation depth produced over the upper basin by the Medspiration simulation compared to other simulations. Figure 54 (j and l) demonstrates that the streamflow network is still on the rising limb part of the hydrograph owing to the slight delay in the primary peak timing of the hydrographs of GHRSST and NCEP simulations given in Figure 7 (b).

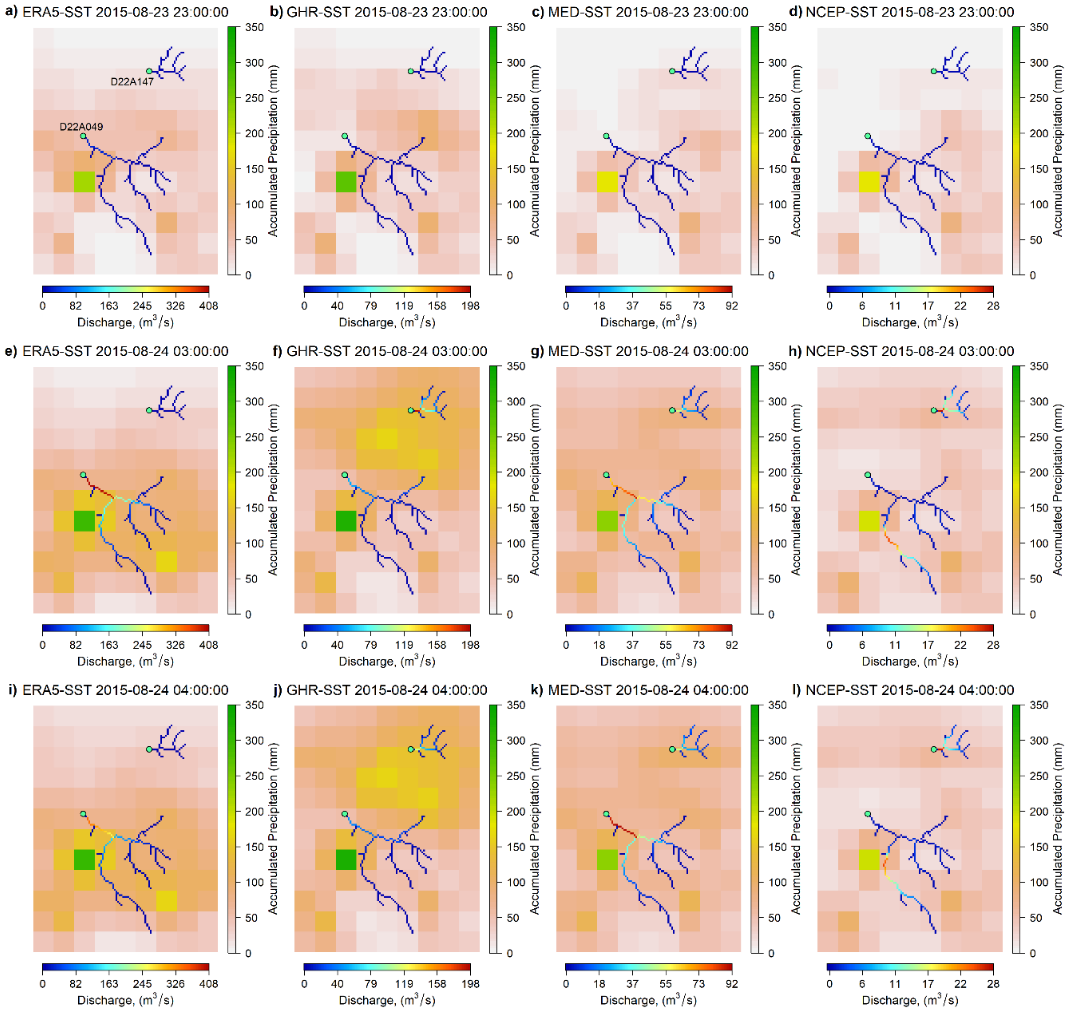


Figure 53 Overlapped dynamic maps of accumulated precipitation simulated by the WRF model (3-km) operated with four different SST datasets (ERA5, GHRSST, Medspiration, and NCEP) and discharge simulated by the WRF-Hydro model (250-m) over the EBS region at 08/23/2015 23:00:00, 08/24/2015 03:00:00, and 08/24/2015 04:00:00. Stream gauges are denoted by green dots.

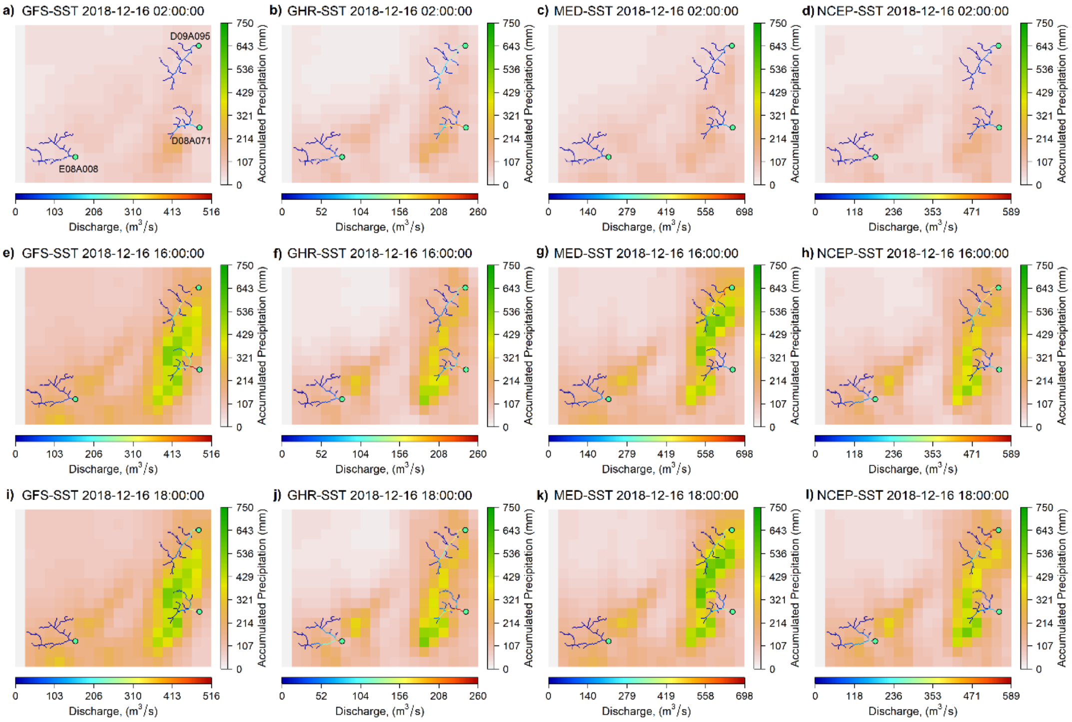


Figure 54 Overlapped dynamic maps of accumulated precipitation simulated by WRF model (3-km) operated with four different SST datasets (GFS, GHRSST, Medspiration, and NCEP) and discharge simulated by WRF-Hydro model (250-m) over the MED region at 12/16/2018 02:00:00, 12/16/2018 16:00:00, and 12/16/2018 19:00:00. Stream gauges are denoted by green dots.

CHAPTER 4

DISCUSSION

Different SST products resulted in a diverse range of spatial and temporal precipitation variability, which substantially impacts the accuracy of runoff simulations concerning timing and volume. Duzenli et al. (2020) showed that physical parameterization, rather than the initial and boundary conditions, had a more significant impact on precipitation variability in extreme precipitation events that are occurred in the MED and EBS. However, in WRF model simulations, the initial state changes and time-varying boundary conditions imposed by GFS and ERA5 might have distinct effects. These diverse datasets further influence the outcomes of SST impact on the findings. Therefore, it is important to note this remark is related to the inherent uncertainty of the findings before drawing any conclusions. As it is expected, the uncertainty raised by the error sources of the lower boundary representation through different SST datasets is of primary importance in weather prediction efforts. In this study, the results show that examining the time-variant and high-resolution SST on the complex terrains over the coastline like the Mediterranean and Eastern Black Sea regions (Ferrari et al., 2020; Papanastasiou et al., 2010; Senatore et al., 2014) may enhance the precipitation estimation accuracy and decrease the discrepancies of the modeling in mesoscale. The exchange process of water and energy between the land/sea surface and the atmosphere's boundary layer involves complex linkages. A better depiction of the lower boundary condition in terms of SST in the atmospheric-hydrological processes is essential in formulating hydrometeorological predictions for flooding and flash floods at various spatial and temporal resolutions. Similarly, updating the lower boundary with more appropriate SST helps to reduce the overall error in storm hydrographs derived using the one-way coupled system. Overall, the GHRSST product yields the highest spatio-temporal correlation, while NCEP yields the least among the temporally variable SST products. Consistent with this result, GHRSST-based runoff simulations yield the highest accuracy, while NCEP shows the lowest among the temporally variable SST products. These results clearly show the significance of using higher spatio-temporal resolution SST products in the simulation of heavy rainfall and extreme runoff.

Cross-correlation is employed as a validation technique since there are no dense buoy observations over the research areas. Hence, the agreement between the accuracy assessments with the cross-correlation and the streamflow simulations implies that this approach may be used to evaluate SST datasets in remote sites that do not have any in-situ measurements.

Hydrographs generated using different SST datasets reveal significant differences between individual events, but they are susceptible to the precipitation inputs modeled with different SST datasets. According to Jee and Kim (2017), detailed SST representation in the WRF model shows a negligible influence on the simulated precipitation for short-term events. Nevertheless, this result is consistent with the findings of previous studies agreeing that the high-resolution SST representation improves the performance of short-term simulations (Cassola et al., 2016; Ivatek-Šahdan et al., 2018; Lebeaupin et al., 2006). In general, the detailed SST representation in the WRF model is not integrated for the short-term events due to the fact that the fluctuations of the SST are low over a short time period. Nonetheless, as a result of climate change, unexpected SST fluctuations are likely to lead to the development of severe storms (Pilatin, 2020). It was found that estimated precipitation depths vary substantially for the two short-term events over EBS and MED regions corresponding to the integration of time-variant (GHRSST, Medspiration, and NCEP) or -invariant (GFS and ERA5) SST products. The time-variant SST products enhanced the accuracy of the predictions in terms of the center locations, the peak timing, and the extent of the SST events, notably in the MED region. Conversely, GFS SST creates the highest overestimation in basin-average precipitation, ~17 mm for the D08A071 basin, and produces precipitation over the sea (agreeing with the study of Cassola et al. (2016) conducted in MED region). The simulations with the high-resolution SST products (Medspiration and GHRSST) were able to address the problem of the large variability in precipitation and hydrological response for the SST event having a mesoscale and convective characteristic occurred in the EBS region. Thus, in the MED and EBS study regions, the Medspiration and GHRSST produced the best basin-averaged precipitation representation, which consequently translates into the improved surface runoff estimations in small coastal catchments.

Additionally, the results corroborate the findings of Senatore et al. (2015), Givati et al. (2016), and Sun et al. (2020) that the WRF model ran with observed precipitation performed better in simulated hydrographs than the simulations performed with the coarse SSTs (ERA5 and GFS) in both locations. On the other hand, models which were driven with the high-resolution SST products may produce better hydrological state representation than those performed with the observed precipitation. According to the results, the interpolation error is 30 - 40 %, which is less than the simulated precipitation for both regions. To further reduce this error, it is essential to have a higher density of meteorological observation stations networks, especially for the regions with complex terrain characteristics. A possible cause of this error resulting in observed precipitation simulation may be due to the fact that almost all of the rain gauges are situated on valley floors, which do not accurately reflect the orographic precipitation features of the region (Eris & Agiralioglu, 2018). Consequently, this may be the reason for the underestimation observed in the hydrographs simulated with the observed precipitation.

It is indicated that WRF-Hydro and its calibration process function reasonably well in that calibration tends to improve model simulations when appropriate precipitation inputs are used. The infiltration-excess, interflow, and channel routing processes influence the discharge estimations in the model. As a result of the infiltration excess process prevailing in the semi-arid areas like the basins in the MED region, plus the non-homogenized precipitation rate distribution is causing the infiltration excess to be underestimated or missed. The overestimation of the discharge during the falling limb of the hydrographs may be attributed to the inadequacies in the interflow mechanism of the model, which has been reported in this study and the earlier studies involving humid (Ryu et al., 2017; Senatore et al., 2015), semi-arid (Lahmers et al., 2019; Silver et al., 2017), and arid regions (Silver et al., 2017). Nevertheless, it is worth mentioning that the calibration method showed better performance in the humid EBS region than that in the semi-arid MED region. The statistics particularly for NSE and KGE, are represented by high positive values (0.2–0.7) in the EBS region. Moreover, with sharp and steep small catchments over the EBS, the hydrologic response is very fast, and overland flow is quickly joined to the river networks and pours to the outlets within a 1-h period. The high-resolution gridded rainfall-runoff coupling greatly benefits from monitoring the water excess condition for a given storm over topographically complex and steeply small watersheds.

Ultimately, identifying the appropriate parameter sets prior to the operational model run for the discharge forecast purposes is critical for more accurate forecasts (Senatore et al., 2015; Silver et al., 2017; Yucel et al., 2015). WRF-Hydro model was calibrated with two different methods, automated and stepwise, and the performance of the related methods were compared. For the stepwise calibraton, among the parameters, REFKDT, SLOPE, MANN and LKSATFAC revealed an important impact on making reliable runoff predictions in both regions, but especially the saturated hydraulic conductivity parameter factor (LKSATFAC) became substantially critical over the MED region.

The model reaches similar optimum parameter values with a lower iteration number for manual calibration compared to the automated calibration (DDS algorithm).

However, the manual calibration requires expert knowledge and manual screening for the decision process of the parameter sets. In addition, the parameters should be sampled from the set of discrete values within the parameter range. In particular, the DDS method, which reaches the optimum value by considering many combinations of all parameters together by optimizing the defined objective function (i.e., KGE), can be used successfully in event-based studies as conducted in the EBS region.

A widely used automated calibration algorithm called DDS is utilized as an alternative for the stepwise calibration procedure. Since the DDS algorithm is started with a randomly selected initial set and continues to select the parameters set in the same way, sudden drops observed in the performances regardless of the increasing iteration number may be attributed to the algorithm structure itself. This depicts that the higher number of iterations defined in the algorithm may not necessarily improve the calibration results. However, as an initial condition, starting the model with a better-guessed parameter set (e.g., default set) may enhance the result systematically with the increasing iteration number. Besides, the process-based automated calibration performed by dividing the parameter set into two groups regarding their impact on hydrological behavior improves the traditional automated calibration approach by up to 30 %. Instead of utilizing the automated calibration for the whole decision variables (i.e., model parameters) at once, considering the funtions of the parameters inside the model and accordingly grouping them for the calibration procedure may bring a new perspective to the automated calibration techniques.

4.2 Added Utility Compared Against Earlier SST Studies

Simulation of the hydrological response after the extreme precipitation events is guided by hydrometeorological modeling systems and local terrain characteristics. Very few papers (Senatore, Davolio, et al., 2020; Senatore, Furnari, et al., 2020) have examined the improved representation of SST representation inside the hydrometeorological modeling system, driving the model predictions linked with a hydrological model. While investigating the influence of SST products on the precipitation estimates, these two studies concurrently examined the impact of SST products on hydrological response simulations during various intense storms. For the first time, SST sensitivity analysis of the WRF and WRF-Hydro modeling system, either with or without data assimilation, was explored by means of mesoscale modeling uncertainty associated with initial/boundary conditions (Senatore, Furnari, et al., 2020). Parallel to mentioned studies, the influence of the SST field on hydrological responses by using the same modeling structure. In relation to Senatore, Davolio, et al. (2020) and Senatore, Furnari, et al. (2020), this study provides the following advancements:

- Due to other relevant sources of uncertainty like initial/boundary conditions, the influence of high-resolution SST datasets is not seen to be substantial in the investigations of Senatore, Davolio, et al. (2020) and Senatore, Furnari, et al. (2020). Conversely, using high-resolution SST datasets in the WRF model showed considerable improvements in runoff simulations in this study compared to previous studies.
- Senatore, Davolio, et al. (2020) and Senatore, Furnari, et al. (2020) studies have been performed over the basins along the Mediterranean Sea, but this present study focuses on basins located in two different regions located over the coastline of the Mediterranean and the Black Sea regions. Accordingly, the sensitivity of the hydrological responses to high-resolution SST products is investigated over a different region with significantly different climatic properties and has not been explored before (i.e., EBS). The results revealed that the high-resolution SST datasets improved hydrological response over the basins in the EBS region more than the MED basins.
- Three different SST products, namely as Medspiration, GHR, and NCEP SST products with 0.022°, 0.01°, and 0.083° resolution, respectively, were used as compared to previous studies which solely utilized Medspiration SST product (0.022°) (Senatore, Davolio, et al., 2020; Senatore, Furnari, et al., 2020). This current study, it was aimed to expand the choices of the different

high-resolution datasets in order to arise a more comprehensive conclusion in terms of the SST product comparison.

- Senatore, Furnari, et al. (2020) used gathered buoy records to conduct temporal accuracy evaluations of the high-resolution SST product. In this study, the SST cross-correlation values of four different SST products were calculated by considering both the temporal and the spatial distribution of the datasets in the absence of buoy observations over the study area. By using additional evaluation of the spatial accuracy of the SST products, this present study expands on the accuracy analysis performed by Senatore, Furnari, et al. (2020).

CHAPTER 5

CONCLUSION

Investigation of hydrologic response for extreme precipitation events simulated by different SST datasets integrated into the WRF model with coarse, high, and time-variant and -invariant resolutions in small catchments with a complex topography and experiencing coastal orography was carried out in this study. The WRF-Hydro model setup is established with one-way coupling from the WRF 3-km domain to the distributed and physical WRF-Hydro model for the simulations of flood hydrographs during significant rainfall events. Although both the GFS and ERA5 (i.e., initial and boundary conditions) for MED and EBS regions contain SST fields that are considered coarse resolution and time-invariant, the GHRSST, Medspiration, and NCEP SST datasets utilized in both regions as external SST datasets with high temporal and spatial resolution. To further enhance the WRF-Hydro model's performance, stepwise and automated calibration methods are used for two separate sets of parameters affecting hydrograph volume and shape. Using higher spatio-temporal resolution SST products (Medspiration and GHRSST), prevailing higher average cross-correlation is also highly influential in capturing the temporal and spatial variability of precipitation in small catchments. This effect is variable from region to region. The key findings of this study are listed as the following:

- Both manual stepwise and automated calibration methods showed better performance in the humid EBS region than that in the semi-arid MED region.
- Process-based automated calibration showed that although the automated calibration techniques are utilized to obtain the values for all decision variable at once for the hydrological model, considering the hydrological behavior of the

parameters during the calibration procedure is still an important factor for further improvement of the results.

- Improved peak discharge estimates in simulated hydrographs are one of the main advantages of updating the SST field with the high spatiotemporal resolution SST datasets in both regions. GHRSST and Medspiration have improved in capturing peak discharge timing and volume than other SST datasets for flood hydrographs simulated over both regions.
- Using time-invariant SSTs for runoff simulations resulted in low performance depending on the spatial and temporal distribution of the precipitation. The flood peak magnitude was simulated to be 2.5 times greater than the observed peak magnitude in the MED region, while the EBS region does not exhibit any hydrological response.
- SST-driven simulations show better performance than those simulations with the observed precipitations, particularly for higher resolution SSTs.

According to the results of precipitation and hydrograph simulations, high-resolution SST products may be used as initial and lower boundary conditions for operational forecast purposes for heavy precipitation events. Under the consideration of abnormal SST changes exacerbated by changing climate, time-variant SST features characterized with high spatio-temporal resolution should be accounted for extreme weather event evaluations in complex coastal topographical regions. Alternative to the commonly preferred stepwise calibration approach, utilizing an automated calibration technique taking into account the hydrological behaviors of the model parameters may be helpful in the calibration of such complex and computationally expensive models.

For the future path, the coupled configuration of the WRF/WRF-Hydro modeling system may be helpful to see the impact of the improved hydrological response on the two-way feedback mechanism between the atmosphere and the land surface model in order to build a more reliable modeling system.

REFERENCES

Abbaszadeh, P., Gavahi, K., & Moradkhani, H. (2020). Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. *Advances in Water Resources*, *145*. https://doi.org/10.1016/j.advwatres.2020.103721

Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature.Com*, *419*(September), 225–232. https://www.nature.com/articles/nature01092

Arsenault, R., Poulin, A., Pascal Côté, ;, & Brissette, F. (2014). *Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration*. https://doi.org/10.1061/(ASCE)HE.1943-5584

Baltaci, H. (2017). Meteorological analysis of flash floods in Artvin (NE Turkey) on 24 August 2015. *Natural Hazards and Earth System Sciences*, *17*(7), 1221–1230. https://doi.org/10.5194/nhess-17-1221-2017

Bigg, G. R., Jickells, T. D., Liss, P. S., & Osborn, T. J. (2003). The role of the oceans in climate. *International Journal of Climatology*, *23*(10), 1127–1159. https://doi.org/10.1002/joc.926

Bozkurt, D., & Sen, O. L. (2011). Precipitation in the Anatolian Peninsula: Sensitivity to increased SSTs in the surrounding seas. *Climate Dynamics*, *36*(3), 711–726. https://doi.org/10.1007/s00382-009-0651-3

Camera, C., Bruggeman, A., Zittis, G., Sofokleous, I., & Arnault, J. (2020). Simulation of extreme rainfall and streamflow events in small Mediterranean watersheds with a one-way-coupled atmospheric-hydrologic modelling system. *Natural Hazards and Earth System Sciences*, *20*(10), 2791–2810. https://doi.org/10.5194/nhess-20-2791-2020

Cassola, F., Ferrari, F., Mazzino, A., & Miglietta, M. M. (2016). The role of the sea on the flash floods events over Liguria (northwestern Italy). *Geophysical Research Letters*, *43*(7), 3534–3542. https://doi.org/10.1002/2016GL068265

Cerbelaud, A., Lefèvre, J., Genthon, P., & Menkes, C. (2022). Assessment of the WRF-Hydro uncoupled hydro-meteorological model on flashy watersheds of the Grande Terre tropical island of New Caledonia (South-West Pacific). *Journal of Hydrology: Regional Studies*, *40*, 101003. https://doi.org/10.1016/j.ejrh.2022.101003

CERSAT. (2012). *Medspiration*. http://cersat.ifremer.fr/thematic-portals/projects/medspiration

Chen, W., Wang, L., Xue, Y., & Sun, S. (2009). Variabilities of the spring river runoff system in East China and their relations to precipitation and sea surface temperature. *International Journal of Climatology*, *29*(10), 1381–1394. https://doi.org/10.1002/joc.1785

Cisneros, M., Cacho, I., Frigola, J., Canals, M., Masqué, P., Martrat, B., Casado, M., Grimalt, J. O., Pena, L. D., Margaritelli, G., & Lirer, F. (2016). Sea surface temperature variability in the central-western Mediterranean Sea during the last 2700 years: A multi-proxy and multi-record approach. *Climate of the Past*, *12*(4), 849–869. https://doi.org/10.5194/cp-12-849-2016

Done, J., Davis, C. A., & Weisman, M. (2004). The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmospheric Science Letters*, *5*(6), 110–117. https://doi.org/10.1002/asl.72

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. In *WATER RESOURCES RESEARCH* (Vol. 28, Issue 4).

Duan, Y., & Kumar, S. (2020). Predictability of seasonal streamflow and soil moisture in national water model and a humid alabama–coosa–tallapoosa river basin. *Journal of Hydrometeorology*, *21*(7), 1447–1467. https://doi.org/10.1175/JHM-D-19-0206.1

Duzenli, E., Yucel, I., Pilatin, H., & Yilmaz, M. T. (2020). Evaluating the performance of a WRF initial and physics ensemble over Eastern Black Sea and Mediterranean regions in Turkey. *Atmospheric Research*, *248*(August 2020), 105184. https://doi.org/10.1016/j.atmosres.2020.105184

ECMWF. (2020). *ERA5*. https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

Eris, E., & Agiralioglu, N. (2018). Streamflow map of the Eastern Black Sea Region, Turkey. *Scientia Iranica*, *25*(3A), 1048–1056. https://doi.org/10.24200/sci.2017.4261

Ferrari, F., Cassola, F., Tuju, P. E., Stocchino, A., Brotto, P., & Mazzino, A. (2020). Impact of model resolution and initial/boundary conditions in forecasting flood-causing precipitations. *Atmosphere*, *11*(6), 1–20. https://doi.org/10.3390/atmos11060592

Fersch, B., Senatore, A., Adler, B., Arnault, J., Mauder, M., Schneider, K., Völksch, I., & Kunstmann, H. (2020). High-resolution fully coupled atmospheric-hydrological modeling: A cross-compartment regional water and energy cycle evaluation. *Hydrology and Earth System Sciences*, *24*(5), 2457–2481. https://doi.org/10.5194/hess-24-2457-2020

Furnari, L., Senatore, A., & Mendicino, G. (2018). *Impact of sea surface temperature representation uncertainty on operational hydrometeorological forecasts over coastal Mediterranean catchments Impact of sea surface temperature representation uncertainty on operational hydrometeorological forecasts over*. *August 2019*, 19–20.

Galanaki, E., Lagouvardos, K., Kotroni, V., Giannaros, T., & Giannaros, C. (2021). Implementation of WRF-Hydro at two drainage basins in the region of Attica, Greece, for operational flood forecasting. *Natural Hazards and Earth System Sciences*, *21*(7), 1983–2000. https://doi.org/10.5194/nhess-21-1983-2021

Givati, A., Gochis, D., Rummler, T., & Kunstmann, H. (2016). Comparing One-Way and Two-Way Coupled Hydrometeorological Forecasting Systems for Flood Forecasting in the Mediterranean Region. *Hydrology*, *3*(2), 19. https://doi.org/10.3390/hydrology3020019

Gochis, D. J., Barlage, M., Cabell, R., Casali, M., Dugger, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., & Zhang, Y. (2020). The WRF-Hydro modeling system technical description, (Version 5.1.1). *NCAR Technical Note*, 107.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97WR03495

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gurer, I., & Ucar, I. (2009). Flood Disasters’ Inventory in Turkey in 2009. *Eleventh International Symposium on Water Management and Hydraulic Engineering*, *September*, 371–380.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

Ivatek-Šahdan, S., Stanešić, A., Tudor, M., Odak Plenković, I., & Janeković, I. (2018). Impact of SST on heavy rainfall events on eastern Adriatic during

SOP1 of HyMeX. *Atmospheric Research*, *200*(October 2017), 36–59. https://doi.org/10.1016/j.atmosres.2017.09.019

Jee, J. B., & Kim, S. (2017). Sensitivity study on high-resolution WRF precipitation forecast for a heavy rainfall event. *Atmosphere*, *8*(6). https://doi.org/10.3390/atmos8060096

Kaya, O. F., Guler, M., Altan, A., & Yorganci, I. (2019). 15-16-17 Aralık 2018 Tarihlerinde Antalya İli Kemer/Ovacık Mevkiinde Görülen Aşırı Yağış Hadisesinin Aktüel Hava Haritaları, Uzaktan Algılama ve Sayısal Hava Tahmin Ürünleri ile Sinoptik ve Hidrolojik Analizi. *10. Ulusal Hidroloji Kongresi 9-12, November*.

Kerandi, N., Arnault, J., Laux, P., Wagner, S., Kitheka, J., & Kunstmann, H. (2018). Joint atmospheric-terrestrial water balances for East Africa: a WRF-Hydro case study for the upper Tana River basin. *Theoretical and Applied Climatology*, *131*(3–4), 1337–1355. https://doi.org/10.1007/s00704-017-2050-8

Kilicarslan, B. M., Yucel, I., Pilatin, H., Duzenli, E., & Yilmaz, M. T. (2021). Improving WRF-Hydro runoff simulations of heavy floods through the sea surface temperature fields with higher spatio-temporal resolution. *Hydrological Processes*, *35*(9), 1–23. https://doi.org/10.1002/hyp.14338

Kim, S., Shen, H., Noh, S., Seo, D. J., Welles, E., Pelgrim, E., Weerts, A., Lyons, E., & Philips, B. (2021). High-resolution modeling and prediction of urban floods using WRF-Hydro and data assimilation. *Journal of Hydrology*, *598*(March), 126236. https://doi.org/10.1016/j.jhydrol.2021.126236

Kunstmann, H., & Stadler, C. (2005). High resolution distributed atmospheric-hydrological modelling for Alpine catchments. *Journal of Hydrology*, *314*(1–4), 105–124. https://doi.org/10.1016/j.jhydrol.2005.03.033

LaCasse, K. M., Splitt, M. E., Lazarus, S. M., & Lapenta, W. M. (2008). The impact of high-resolution sea surface temperatures on the simulated nocturnal

Florida marine boundary layer. *Monthly Weather Review*, *136*(4), 1349–1372. https://doi.org/10.1175/2007MWR2167.1

Lahmers, T. M., Gupta, H., Castro, C. L., Gochis, D. J., Yates, D., Dugger, A., Goodrich, D., & Hazenberg, P. (2019). Enhancing the structure of the WRF-hydro hydrologic model for semiarid environments. *Journal of Hydrometeorology*, *20*(4), 691–714. https://doi.org/10.1175/JHM-D-18-0064.1

Lebeaupin, C., Ducrocq, V., & Giordani, H. (2006). Sensitivity of torrential rain events to the sea surface temperature based on high-resolution numerical forecasts. *Journal of Geophysical Research Atmospheres*, *111*(12), 1–19. https://doi.org/10.1029/2005JD006541

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos*, *89*(10), 93–94. https://doi.org/10.1029/2008EO100001

Lespinas, F., Dastoor, A., & Fortin, V. (2018). Performance of the dynamically dimensioned search algorithm: Influence of parameter initialization strategy when calibrating a physically based hydrological model. *Hydrology Research*, *49*(4), 971–988. https://doi.org/10.2166/nh.2017.139

Li, L., Gochis, D. J., Sobolowski, S., & Mesquita, M. D. S. (2017). Evaluating the present annual water budget of a Himalayan headwater river basin using a high-resolution atmosphere-hydrology model. *Journal of Geophysical Research*, *122*(9), 4786–4807. https://doi.org/10.1002/2016JD026279

Liu, Y., Liu, J., Li, C., Yu, F., Wang, W., & Qiu, Q. (2020). Parameter Sensitivity Analysis of the WRF-Hydro Modeling System for Streamflow Simulation: a Case Study in Semi-Humid and Semi-Arid Catchments of Northern China. *Asia-Pacific Journal of Atmospheric Sciences*. https://doi.org/10.1007/s13143-020-00205-2

McCabe, G. J., & Wolock, D. M. (2008). Joint variability of global runoff and global sea surface temperatures. *Journal of Hydrometeorology*, *9*(4), 816–824. https://doi.org/10.1175/2008JHM943.1

Meredith, E. P., Semenov, V. A., Maraun, D., Park, W., & Chernokulsky, A. v. (2015). Crucial role of Black Sea warming in amplifying the 2012 Krymsk precipitation extreme. *Nature Geoscience*, *8*(8), 615–619. https://doi.org/10.1038/ngeo2483

Miglietta, M. M., Mazon, J., Motola, V., & Pasini, A. (2017). Effect of a positive Sea Surface Temperature anomaly on a Mediterranean tornadic supercell. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-13170-0

Naabil, E., Lamptey, B. L., Arnault, J., Kunstmann, H., & Olufayo, A. (2017). Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa. *Journal of Hydrology: Regional Studies*, *12*(June), 196–209. https://doi.org/10.1016/j.ejrh.2017.05.010

NCEP, & NOAA. (2014). *NCEP SST Analysis*. https://polar.ncep.noaa.gov/sst/ophi/verification.shtml

NOAA. (2015). *NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D65D8PWK

Papanastasiou, D. K., Melas, D., & Lissaridis, I. (2010). Study of wind field under sea breeze conditions; an application of WRF model. *Atmospheric Research*, *98*(1), 102–117. https://doi.org/10.1016/j.atmosres.2010.06.005

Pastor, F., Valiente, J. A., & Palau, J. L. (2018). Sea Surface Temperature in the Mediterranean: Trends and Spatial Patterns (1982–2016). *Pure and Applied Geophysics*, *175*(11), 4017–4029. https://doi.org/10.1007/s00024-017-1739-z

Pilatin, H. (2020). *Sensitivity and Sea Surface Temperature Analyses of WRF Model for Predicting Heavy Rainfall Events Observed in Eastern Black Sea and Mediterranean Regions of Turkey*.

Rebora, N., Molini, L., Casella, E., Comellas, A., Fiori, E., Pignone, F., Siccardi, F., Silvestro, F., Tanelli, S., & Parodi, A. (2013). Extreme rainfall in the mediterranean: What can we learn from observations. *Journal of Hydrometeorology*, *14*(3), 906–922. https://doi.org/10.1175/JHM-D-12-083.1

Robinson, I., Piollé, J. F., LeBorgne, P., Poulter, D., Donlon, C., & Arino, O. (2012). Widening the application of AATSR SST data to operational tasks through the Medspiration Service. *Remote Sensing of Environment*, *116*, 126–139. https://doi.org/10.1016/j.rse.2010.12.019

Roxy, M. (2014). Sensitivity of precipitation to sea surface temperature over the tropical summer monsoon region-and its quantification. *Climate Dynamics*, *43*(5–6), 1159–1169. https://doi.org/10.1007/s00382-013-1881-y

Ryu, Y., Lim, Y. J., Ji, H. S., Park, H. H., Chang, E. C., & Kim, B. J. (2017). Applying a coupled hydrometeorological simulation system to flash flood forecasting over the Korean Peninsula. *Asia-Pacific Journal of Atmospheric Sciences*, *53*(4), 421–430. https://doi.org/10.1007/s13143-017-0045-0

Sampson, K., & Gochis, D. (2015). *WRF Hydro GIS Pre-Processing Tools , Version 2 . 2 Documentation*. 1–39.

Senatore, A., Davolio, S., Furnari, L., & Mendicino, G. (2020). Reconstructing flood events in mediterranean coastal areas using different reanalyses and high-resolution meteorological models. *Journal of Hydrometeorology*, *21*(8), 1865–1887. https://doi.org/10.1175/JHM-D-19-0270.1

Senatore, A., Furnari, L., & Mendicino, G. (2020). Impact of high-resolution sea surface temperature representation on the forecast of small Mediterranean catchments' hydrological responses to heavy precipitation. *Hydrology and*

Earth System Sciences, *24*(1), 269–291. https://doi.org/10.5194/hess-24-269-2020

Senatore, A., Mendicino, G., Gochis, D. J., Yu, W., Yates, D. N., & Kunstmann, H. (2015). Fully coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *Journal of Advances in Modeling Earth Systems*, *7*(4), 1693–1715. https://doi.org/10.1002/2015MS000510

Senatore, A., Mendicino, G., Knoche, H. R., & Kunstmann, H. (2014). Sensitivity of modeled precipitation to sea surface temperature in regions with complex topography and coastlines: A case study for the Mediterranean. *Journal of Hydrometeorology*, *15*(6), 2370–2396. https://doi.org/10.1175/JHM-D-13-089.1

Silver, M., Karnieli, A., Ginat, H., Meiri, E., & Fredj, E. (2017). An innovative method for determining hydrological calibration parameters for the WRF-Hydro model in arid regions. *Environmental Modelling and Software*, *91*, 47–69. https://doi.org/10.1016/j.envsoft.2017.01.010

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Zhiquan, L., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., & Huang, X.-Y. (2019). A Description of the Advanced Research WRF Model Version 4. *NCAR Technical Note NCAR/TN-475+STR*. http://library.ucar.edu/research/publish-technote

Sun, M., Li, Z., Yao, C., Liu, Z., Wang, J., Hou, A., Zhang, K., Huo, W., & Liu, M. (2020). Evaluation of flood prediction capability of the WRF-hydro model based on multiple forcing scenarios. *Water (Switzerland)*, *12*(3). https://doi.org/10.3390/w12030874

Team GHRSST. (2010a). *Black Sea Ultra High Resolution SST L4 Analysis 0.01 deg Resolution | PO.DAAC / JPL / NASA*.

https://podaac.jpl.nasa.gov/dataset/OISST_UHR_NRT-GOS-L4-BLK-v2.0?ids=TemporalResolution&values=Daily&search=l4

Team GHRSST. (2010b). *Mediterranean Sea Ultra High Resolution SST L4 Analysis 0.01 deg Resolution | PO.DAAC / JPL / NASA*. https://podaac.jpl.nasa.gov/dataset/OISST_UHR_NRT-GOS-L4-MED-v2.0?ids=TemporalResolution&values=Daily&search=l4

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1), 1–16. https://doi.org/10.1029/2005WR004723

Trenberth, K. E. (1999). Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climatic Change*, *42*, 327–339.

Trenberth, K. E., & Shea, D. J. (2005). Relationships between precipitation and surface temperature. *Geophysical Research Letters*, *32*(14), 1–4. https://doi.org/10.1029/2005GL022760

Turkes, M. (1996). Spatial and Temporal Analysis of Annual Rainfall Variations in Turkey. *International Journal of Climatology*, *16*(9), 1057–1076. https://doi.org/10.1002/(SICI)1097-0088(199609)16:9<1057::AID-JOC75>3.0.CO;2-D

Turkes, M. (1999). Vulnerability of Turkey to desertification with respect to precipitation and aridity conditions. *Turkish Journal of Engineering and Environmental Sciences*, *23*(5), 363–380.

Turuncoglu, U. U. (2015). Identifying the sensitivity of precipitation of Anatolian peninsula to Mediterranean and Black Sea surface temperature. *Climate Dynamics*, *44*(7–8), 1993–2015. https://doi.org/10.1007/s00382-014-2346-7

Varlas, G., Anagnostou, M. N., Spyrou, C., Papadopoulos, A., Kalogiros, J., Mentzafou, A., Michaelides, S., Baltas, E., Karymbalis, E., & Katsafados, P. (2019). A multi-platform hydrometeorological analysis of the flash flood event of 15 November 2017 in Attica, Greece. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010045

Vecchi, G. A., & Harrison, D. E. (2002). Monsoon breaks and subseasonal sea surface temperature variability in the Bay of Bengal. *Journal of Climate*, *15*(12), 1485–1493. https://doi.org/10.1175/1520-0442(2002)015<1485:MBASSS>2.0.CO;2

Volosciuk, C., Maraun, D., Semenov, V. A., Tilinina, N., Gulev, S. K., & Latif, M. (2016). Rising mediterranean sea surface temperatures amplify extreme summer precipitation in central europe. *Scientific Reports*, *6*(February), 1–7. https://doi.org/10.1038/srep32450

Wang, J., Wang, C., Rao, V., Orr, A., Yan, E., & Kotamarthi, R. (2019). A parallel workflow implementation for PEST version 13.6 in high-performance computing for WRF-Hydro version 5.0: a case study over the midwestern United States. *Geoscientific Model Development*, *12*(8), 3523–3539. https://doi.org/10.5194/gmd-12-3523-2019

Wang, W., Liu, J., Li, C., Liu, Y., Yu, F., & Yu, E. (2020). An evaluation study of the fully coupled WRF/WRF-Hydro modeling system for simulation of storm events with different rainfall evenness in space and time. *Water (Switzerland), 12*(4). https://doi.org/10.3390/W12041209

Wehbe, Y., Temimi, M., Weston, M., Chaouch, N., Branch, O., Schwitalla, T., Wulfmeyer, V., Zhan, X., Liu, J., & al Mandous, A. (2019). Analysis of an extreme weather event in a hyper-arid region using WRF-Hydro coupling, station, and satellite data. *Natural Hazards and Earth System Sciences*, *19*(6), 1129–1149. https://doi.org/10.5194/nhess-19-1129-2019

Yilmaz, K. K., Gupta, H. v., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9), 1–18. https://doi.org/10.1029/2007WR006716

Yilmaz, M. T., Anderson, M. C., Zaitchik, B., Hain, C. R., Crow, W. T., Ozdogan, M., Chun, J. A., & Evans, J. (2014). Comparison of prognostic and diagnostic surface flux modeling approaches over the Nile River basin. *Water Resources Research*, *50*(1), 386–408. https://doi.org/10.1002/2013WR014194

Yin, D., Xue, Z. G., Gochis, D. J., Yu, W., Morales, M., & Rafieeinasab, A. (2020). A process-based, fully distributed soil erosion and sediment transport model for WRF-hydro. *Water (Switzerland), 12*(6). https://doi.org/10.3390/w12061840

Yucel, I., & Onen, A. (2014). Evaluating a mesoscale atmosphere model and a satellite-based algorithm in estimating extreme rainfall events in northwestern Turkey. *Natural Hazards and Earth System Sciences*, *14*(3), 611–624. https://doi.org/10.5194/nhess-14-611-2014

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*, 49–66. https://doi.org/10.1016/j.jhydrol.2015.01.042

Zhang, J., Lin, P., Gao, S., & Fang, Z. (2020). Understanding the re-infiltration process to simulating streamflow in North Central Texas using the WRF-hydro modeling system. *Journal of Hydrology, 587*(December 2019), 124902. https://doi.org/10.1016/j.jhydrol.2020.124902

Zhang, X., Cornuelle, B., Martin, A., Ralph, M., & Weihs, R. (2015). West-WRF Sensitivity to Sea Surface Temperature Boundary Condition in California Precipitation Forecasts of AR Related Events A33H-2496 Model forecast RMS errors Conclusions : *AGU E-Poster*, 10.