Multisource Remote Sensing and Long Short-Term Memory for River Discharge Prediction in the Lancang Mekong River

by

xxxxxxxxx

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Water Engineering and Management

Examination Committee: Prof. xxxxxxxx (Chairperson)
Dr. xxxxxxx
Dr. xxxxxx
Dr. xxxxxx

Nationality: xxxxxx
Previous Degree: Bachelor of Engineering in Agricultural Engineering
Institute of Engineering
xxxxxxx

Scholarship Donor: ADB-JSP Scholarship

Asian Institute of Technology
School of Engineering and Technology
Thailand
April 2023

AUTHOR'S DECLARATION

I, xxxxxxx, declare that the research work carried out for this thesis was in accordance with the regulations of the Asian Institute of Technology. The work presented in it are my own and has been generated by me as the result of my own original research, and if external sources were used, such sources have been cited. It is original and has not been submitted to any other institution to obtain another degree or qualification. This is a true copy of the thesis, including final revisions.

Date: 21 April 2023

Name: xxxxxxx

Signature:

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisor Prof. Sangam Shrestha, whose persistent support, guidance, and expert knowledge has been crucial to the successful completion of my research. His valuable insights and feedback have played a pivotal role in steering the course of my work and have provided me with the motivation to achieve my goals.

I would also like to extend my sincere appreciation to my examination committee Dr. Mohana Sundaram Shanmugam, Dr. Ho Huu Loc and Dr. Kavinda Gunasekara, whose thorough evaluation of my work has been instrumental in ensuring the quality and rigor of my research. Their constructive feedback and insightful suggestions were instrumental in reaching new levels of academic achievement and thoughtful critiques and probing questions challenged me to think more deeply about my subject matter and approach.

I am particularly grateful to the ADB-JSP scholarship programs that have provided me with the financial support necessary to pursue my academic and research goals. Without their support, I would not have been able to undertake this important work.

I would also like to acknowledge the assistance and support of all the Water Engineering Management faculty, staff, seniors and my colleagues, whose inspiration and feedback helped me to stay focused and motivated and have enabled me to approach my work with confidence and enthusiasm.

Finally, I am grateful for the unwavering love, care, and support that my husband Mr. Salik Bhusal has provided me with during my academic journey. His belief in me has been a constant source of motivation, and I am fortunate to have him by my side. Additionally, the support and encouragement from my family and friends have been invaluable. Their love and support have been the foundation upon which my success has been built, and I am deeply grateful for their ongoing support.

ABSTRACT

The spatiotemporal variation of river discharge plays a crucial role in effective water resources management, as it provides crucial information for decision making in various domains such as irrigation, hydropower, flood forecasting, and environmental conservation. Traditional ground-based gauging stations, although providing accurate and reliable measurements, are typically limited in their spatial coverage and prone to data gaps. Satellite based remote sensing can provide a feasible alternative method for river monitoring, as it offers spatially distributed and temporally continuous data that can be utilized to infer river water level and discharge. In this context, this study aimed to develop a framework using a Long Short-Term Memory (LSTM) model that incorporates multiple satellite missions to predict daily river discharge and evaluate its effectiveness for ungauged locations.

First, water level at eight reach of Lancang Mekong River was obtained utilizing altimetric water level measurements from Jason-1/2, ENVISAT, and Saral altimetry missions. Optical images from MODIS Aqua and Terra were used to obtain the time series surface reflectance ratio of dry to wet pixels as a proxy for water discharge for each location. However, careful selection of the location of the dry and wet pixels near the gauging station is essential for accurate prediction, and the study determined the best location by computing the correlation between surface reflectance ratio and discharge for each pixel.

LSTM models were developed for each station and trained using single and combined input features; surface reflectance ratio from MODIS Aqua and Terra and altimetry derived water level to assess their influence on the accuracy of daily discharge prediction. The inclusion of features from multi mission satellite showed improved and better performance in discharge prediction, with predictive accuracy ranging from 0.54 to 0.92 in terms of NSE. The spatial transferability of the LSTM model to predict daily river discharge at ungauged locations was further assessed. The result showed model performs better for certain reaches of the river, but performance decreases while moving farther from the trained reach indicating the model can be developed for certain reaches of the river that can predict discharge at ungauged locations within that reach with good and reliable accuracy.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	= Artificial Intelligence
AMR	= Advanced Microwave Radiometer
AMSR-E	= Advanced Microwave Scanning Radiometer for the Earth Observing System
ANN	= Artificial Neural Network
CNES	= Centre National d'Etudes Spatiales
CTOH	= Center for Topographic Studies of the Ocean and Hydrosphere
DEM	= Digital Elevation Model
DORIS	= Doppler Orbitography and Radio-positioning Integrated by Satellite
ENVISAT	= Environmental Satellite
EOS	= Earth Observing System
ERS	= European Remote Sensing satellite
ESA	= European Space Agency
EUMETSAT	= Exploitation of Meteorological Satellites
GDR	= Geophysical Data Record
GEE	= Google Earth Engine
GEOS-3	= Geodynamics Experimental Ocean Satellite-3
GFO-RA	= GEOSat Follow-On Radar Altimeter
GHz	= Gigahertz
GNSS	= Global Navigation Satellite System
GPS	= Global Positioning System
GRDC	= Global Runoff Data Centre
HY-2A	= Hai Yang 2A
JPL	= Jet Propulsion Laboratory
km	= Kilometer
Ku band	= Kurtz-under band
LRA	= Laser Retroflector Array
LRM	= Low Resolution Mode
LSTM	= Long Short-Term Memory
MLP-ANN	= Multilayer perceptron Artificial Neural Network

mm	= millimeter
MODIS	= Moderate Resolution Imaging Spectroradiometer
MRC	= Mekong River Commission
MWR	= Microwave Radiometer
NASA	= National Aeronautics and Space Administration
NIR	= Near Infrared
NOAA	= National Oceanic and Atmospheric Administration
NSE	= Nash Sutcliffe efficiency
OSTM	= Ocean Surface Topography Mission
PBIAS	= percent bias
PLRM	= Pseudo Low Resolution Mode
POD	= Probability of Detection
R^2	= Coefficient of Determination
RMSE	= Root Mean Square Error
RNN	= Recurrent Neural Network
RS	= Remote Sensing
SAR	= Synthetic Aperture Radar
SARAL	= Satellite with ARgos and ALtiKa
SDR	= Stage Discharge Rating
SRAL	= SAR Radar Altimeter
SWOT	= Surface Water and Ocean Topography
TOPEX	= The Ocean Topography Experiment
US	= United States
VS	= Virtual Station

CHAPTER 1
INTRODUCTION

1.1 Background of the Study

Water is nature's most valuable resource, and due to rising demand, fresh water supply is in risk of becoming non-renewable (Kim et al., 2008). The water on the Earth spans over 71 percent of the surface area. However, only 3% of this is fresh water, with the remaining 2.5% trapped in ice and glaciers. As a result, people are dependent on only 0.5% of the available freshwater for their various requirements (Kashid & Pardeshi, 2014). These needs are fulfilled by sources of inland surface water like lakes, reservoirs, creeks, streams, and rivers, which are crucial for our daily lives. These sources of water are responsible for supplying most of the water needed for thermoelectric generation, public use, agriculture, mining, and industrial applications (Kenny et al., 2009).

Understanding the distribution patterns of water resources spatial- temporally is crucial for effective water resource management and evaluating water vulnerability. River discharge is an important variable in study of hydrological process and applications such as global change monitoring, flood risk assessments, water supply management, irrigation projects and many more (Van Dijk et al., 2016). Measurement of the discharge of rivers is vital to gain a comprehensive understanding of how hydrological processes and changes are occurring at both basin and global scale. This is particularly important due to the impact of natural disasters and anthropogenic activities on the hydrological cycle (Robert Brakenridge et al., 2012). The level of water bodies is a major concern when it comes to managing and monitoring inland water resources.

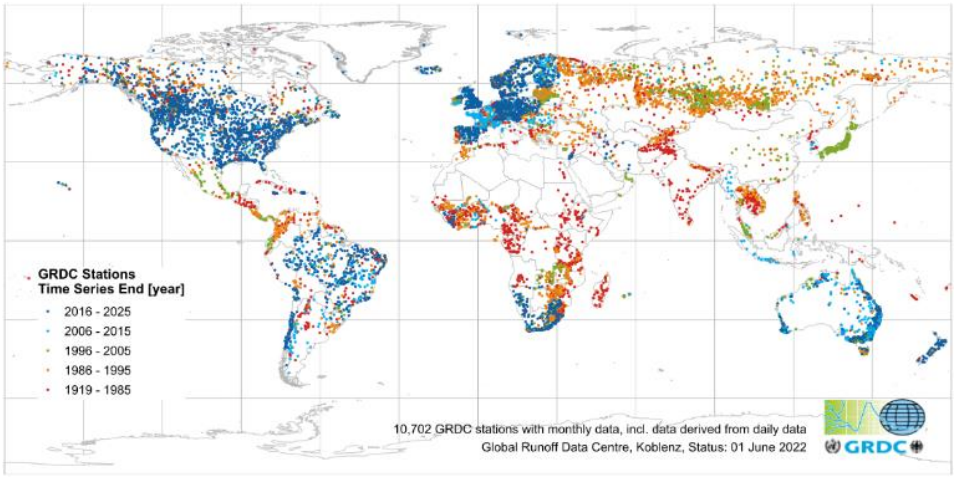
Traditionally, the measurement of river discharge has been based on a system of gauging stations that record water levels at specific time intervals (Shi et al., 2020). These gauges are typically located at fixed points in rivers and lakes, which limits the number of measurements that can be taken. Additionally, the distribution of these gauges is not uniform and can vary depending on location accessibility and national policies. Setting up and maintaining these gauges in remote areas can also be both difficult and expensive. The precision of measurements obtained from gauge stations is heavily dependent on the processing techniques employed and the condition of the

water bodies being monitored. Hence, during extreme events such as floods, gauge stations are particularly susceptible to inaccuracies (Biancamaria et al., 2010).

Figure 1-1 depicts the global distribution of hydrologic gauges and their insufficient monitoring capability.

Figure 1-1

Global Distribution of Hydrologic Gauges (*GRDC Data Portal*)



To monitor changes in river discharge accurately, it is crucial to have a reliable and repeatable approach, especially when in-situ gauge stations are insufficient or unavailable. Utilizing satellite-measured water level and inundation area to estimate river stage and discharge is a feasible alternative method that is both accessible and straightforward (Huang et al., 2018). Broadly, use of remote sensing in discharge estimation are grouped into three categories: first category also called rating curve method uses stage-discharge relationship developed from remotely sensed water level or inundated areas with simultaneously collected in-situ data for discharge estimation, second category uses river water level or width from remote sensing data as input parameter or reference for calibration in hydrological or hydraulic models and third category also called remote sensing driven method uses remote sensing data as major input for estimation of river discharge (Shi et al., 2020).

Various studies have been carried out demonstrating applicability of satellite altimetry water level for river discharge prediction (Birkinshaw et al., 2010; Bogning et al., 2018; Calmant & Seyler, 2006; Jiang et al., 2017). However, the temporal resolution of water level and discharge is limited to ten days (Topex/Poseidon, Jason 1/2/3) or thirty five

days (EnviSat, SARAL/ALTIKA) due to repeat period of satellite orbit (Tarpanelli, Barbetta, et al., 2013). Moreover, use of optical sensors and passive microwave sensors in discharge estimation cannot be discarded since they have less revisit time almost daily and large coverage area. The reflectance ratio of land and water body (C/M) from Near Infrared (NIR) band of Moderate Resolution Imaging Spectroradiometer (MODIS), Landsat, Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and Global Flood Detection System (GFDS) has been effectively utilized as proxy for river discharge estimation (Brakenridge et al., 2005; Hou et al., 2018; Tarpanelli et al., 2019). Cloudy sky hinders the usability of optical sensors data which may result in data outages in measurements and the possible way to overcome is to use multi-mission satellite products to increase sample size (Tarpanelli et al., 2019).

Thus, combining optical sensor data with altimetry data can help overcome the issue of cloudy skies that may affect optical sensors and improve temporal sampling. Artificial Neuron Networks (ANNs) are widely used in the area of remote sensing since they provide a simple and effective way to integrate features from different data sources into the common retrieval algorithm.

1.2 Statement of the Problem

The precise determination of river discharge is crucial for a numerous hydrological purpose, including water supply management and planning, hydropower generation, reservoir control and management, flood control and prediction, as well as comprehending the global water cycle. To get an accurate measurement of discharge, most rivers require a series of flow velocity measurements in each sector of the cross section. Also, river discharge is frequently obtained from river stage measurements using the stage-discharge rating (SDR) curve to save expense, effort, and time. At several points along the river, in-situ gauge station would offer continuous and dependable river stage measurements (Liu et al., 2016).

However, the ongoing erosion and sediment deposition activities within river channels and banks result in a continual modification of channel morphology, which causes alterations to the stage-discharge relationship over time. As a result, a regular comparison of the stage discharge rating curves to direct observations is required and hence river gauging stations are expensive, laborious, and time-consuming to maintain

(Pan, 2013). On other hand, the number of gauging stations has dropped on a global basis, which is especially problematic in the face of hydrological regimes that have altered dramatically in recent years or will do so in the future due to climate change (Arnell & Gosling, 2013; Fekete & Vörösmarty, 2007.). Mekong is one of the such basin whose hydrology is altered due to impact of climate change and hydropower developments (Hoang et al., 2016). The Lancang Mekong River is a lengthy and remote waterway, which would make it challenging to establish and sustain sufficient gauge stations to monitor the entire river.

In transboundary river basins, issues like political or economic constraints and data delays may lead to the inaccessibility of data in specific regions. Consequently, the difficulty of obtaining regular and reliable observations at numerous sites along the river persists as an ongoing challenge (Liu et al., 2016). Conventional methods of estimating river discharge have limited spatial and temporal coverage, resulting in several ungauged river reaches. To address this, spaceborne sensors can be utilized due to the rise in the number of earth monitoring satellites and their enhanced temporal and spatial resolution. Hence, this research aims to develop a reliable methodology for estimating river discharge using multiple satellite missions, including optical sensors and altimeter products, and extend this methodology to ungauged river reaches. Overall, this research aims to contribute to the development of a practical and robust framework for estimating river discharge using remote sensing technologies, which can support water resource management activities in both gauged and ungauged river reaches.

1.3 Research Questions

1. Is it feasible to estimate river discharge accurately and reliably by integrating data from multiple satellite missions, such as optical sensors and altimeter data?
2. What is the accuracy of river discharge estimates obtained from multiple satellite missions compared to those obtained through traditional methods?
3. What are the unique characteristics and constraints of different satellite missions and how they can impact on performance of river discharge prediction?
4. How the methodology for river discharge estimation using remote sensing can be extended to ungauged river reaches, where there are no direct measurements of discharge available for calibration and validation?

1.4 Objectives of the Study

The main objective of this study is to develop and evaluate a framework using multisource remote sensing datasets and Long Short-Term Memory (LSTM) for river discharge prediction in the Lancang Mekong River Basin and assess its applicability in ungauged reach.

To accomplish this main objective, following specific objectives are set:

i. To obtain water level time series coping with the limitations of temporal and spatial resolution of satellite altimetry.
ii. To develop LSTM model for predicting daily river discharge using multisource remote sensing data and evaluate its performance.
iii. To analyze spatial transferability of LSTM model for discharge prediction at ungauged locations in multiple river reach.

1.5 Scope of the Study

i. This study focuses on using remote sensing data from multiple sources like satellite altimetry data and satellite optical sensor data to predict river discharge.
ii. The study aims to develop a Long Short-Term Memory (LSTM) for discharge prediction at different reaches of Lancang Mekong River.
iii. The study intends to evaluate the spatial transferability of the developed model to predict discharge at ungauged reaches.
iv. The study contributes to improving water management practices in the Lancang Mekong river basin by providing a reliable method for discharge estimation.

1.6 Limitations of the Study

i. The accuracy of remote sensing data, particularly satellite optical sensor data, is affected by cloud covering, which limits the obtainability and quality of data for the study.
ii. The quality of satellite altimetry data is affected by topography, particularly in areas with steep slopes, which restricts the quantity and quality of data available for discharge estimation.

CHAPTER 2
LITERATURE REVIEW

2.1 Importance of Monitoring River Flow

Monitoring and quantifying river flow is crucial for future forecasting as well as for the sustainable management of this valuable resource (E. Zakharova et al., 2020). Importance of more accurate, high frequency, and accessible water data is increasing because of growing populations and rival water priorities, including the preservation and restoration of aquatic ecosystems.(Hirsch & Costa, 2004; Arnell & Gosling, 2013). River discharge is a crucial parameter for various applications at different levels, which includes global water balance, design of engineering structures, flood forecasting, managing reservoirs, navigation, water supply planning, restoration, and environmental management (Gravelle, 2015). River flow regimes, which encompass factors like low flows, long-term average, seasonality of flow, high flows, including other forms of flow variability, have a vital role in ecosystem of freshwater (Arnell & Gosling, 2013).

In the late 1800s, advancements in technology led to the development of hydroelectricity, and subsequently, the widespread establishment of gauging stations at strategic locations for the purpose of gathering data to aid in river discharge forecasting. This was important for ensuring a consistent supply of water volumes in accordance with demand, as well as maintaining uninterrupted power production (Depetris, 2021). However, the current environmental and climate crisis has presented new challenges to the original straightforward objectives. Apart from changes in precipitation patterns and an increase in extreme weather events, climate change also brings non-stationary characteristics to river flow series (Arnell & Gosling, 2013). Therefore, in the face of climate change and its impact on river flows, it is even more essential to establish reliable hydrological databases. Such databases can serve as a foundation for river management that focuses on flood control, water scarcity, and water quality maintenance. By carefully monitoring the frequency and intensity of river flows, informed decisions can be made to manage water resources effectively (Derecki & Quinn, 1987).

It is crucial to highlight that Milliman & Farnsworth's (2011) synthesis indicates that over half of the rivers with no hydrological data are situated in Oceania and Africa.

Milliman & Farnsworth, (2011). Maintaining global river discharge databases is a challenging task and despite major improvements since river flow measurement began, the extensive gauging network is still far from perfect (Milliman & Farnsworth, 2011). For instance, the Global Runoff Data Centre (GRDC) only includes a limited number of stations that are mostly determined by the participating nations. Additionally, hydrological data is often insufficient or partial, and changes that occur over time in gauging networks could be undetected (Depetris, 2021).

2.2 Measurement Methods of Stream Discharge

Discharge measurements are conducted in natural watercourses to quantify the surface outflow of a basin, assess its periodic variability, and characterize its outflow dynamics. Data on stream flow must be collected in a cohesive way, for a consistent period, with an assessment of the accuracy and uncertainty involved (Tazioli, 2011). There are several techniques available for estimating river discharge, including stream gauging through the use of stage and rating curves, dilution gauging, velocity point measurements, and the use of current meters (Gravelle, 2015). Most hydrometric stations have a staff gauge, which continually records data during predetermined time intervals. The water level data may be transformed into data on river discharge using a stage-discharge relationship. While analytical formulae may be used to explain these relationships, it is preferable to collect experimental data and calibrate it using actual river discharge measurements (Perumal et al., 2007).

To determine the mean flow velocity of a river cross-section, the conventional approach is to submerge a current meter in multiple locations (Tazioli, 2011). However, in situations where water depth is too shallow for the equipment to be submerged, or flow velocity is too low due to turbulence, traditional methods of obtaining mean flow velocity using a current meter may not be effective. To address these issues, an alternative method known as the artificial tracing method can be used, which involves the use of radioactive substances or chemicals like sodium chloride (Florkowski et al., 1969).

To address the gap of on-site measurements, alternative methods of streamflow prediction and monitoring; modeling and satellite observations, have been rapidly developed. While satellite observation may give global views, hydrological modeling is a powerful tool for providing insights at regional or basin scales (E. Zakharova et al.,

2020). Since river discharge cannot be monitored directly due to its nature, observations of other hydraulic variables, such as water elevation, river flow velocity, water surface extent, and slope, both satellite and conventional monitoring methods is used to calculate river discharge (Bjerklie et al., 2005).

2.3 Applicability of Remote Sensing in Stream Discharge Estimation

Remote sensing has developed as a reliable source of observations during the recent years, especially in regions of the world with scarce in situ networks (Bjerklie et al., 2018). Many researchers have investigated the use of remotely sensed hydraulic characteristics, such as width of water bodies, gradient, and elevation, to estimate stream discharge since there is currently no direct satellite-based technique for measuring it (Bjerklie et al., 2005; Dingman & Bjerklie, 2005; Leon et al., 2006; Zakharova et al., 2020). Thus, satellite remote sensing is used to estimate a variety of hydrological status variables and fluxes. Though satellite remote sensing systems can provide extensive coverage with continuous and updated measurements, but their accuracy mainly depends on the available in-situ observations for algorithm development and validation (Tang et al., 2009).

Contrary to popular belief, remote sensing does not aim to completely replace gauges in discharge estimation (Gleason & Durand, 2020). Remote sensing for discharge estimation is therefore used for substitute of different purpose:

- At the mercy of politics and economics, gauges that exist now might not exist tomorrow.
- Remote Sensing signals are able to expand point gauge measurements in space and time due to high gauge calibration data.
- Water resources may be documented using Remote Sensing for Discharge in a variety of channel configurations and during times of flooding, which are challenges for gauges.
- Another area where Remote Sensing for Discharge may excel is with discontinued gauges, as these gauges can be calibrated and then utilized to parameterize models as they advance into the future.
- When creating a primary data source for usage in locations that are not well monitored, remote sensing might be utilized to prepare inaccurate data.

Bjerklie et al. (2003) assessed the possibilities of the various satellite data sources, described a variety of methods for using remote sensing to calculate river flow, and thought about the possibility of computing river flow rate exclusively from remote sensing data sources. In comparison to other models that merely contain width and slope, slope and velocity, the authors contend that models which relies on these parameters are often more accurate, especially for big rivers.

Lin et al. (2019) created a cogent worldwide reanalysis of daily river discharge at approximately three million river reaches using a considerable amount of ground and remote sensing data (mostly for precipitation and evapotranspiration) along with the most recent advances in remote sensing hydrography. The power of remote sensing for global hydrologic modeling is illustrated by the level of temporal and spatial precision that has never before been attained. Lin et al. (2019) employed a calibration strategy that takes uncertainty into account, calibrating with gauges when available, remote sensing products when gauges are not, and reanalysis data when remote sensing and gauges are both accessible. This makes it feasible to produce discharge that wouldn't be possible without the use of remote sensing by making the best use possible of both in situ and remote sensing data.

Remote sensing has become an increasingly popular tool for predicting discharge due to its extensive spatial coverage and long-term monitoring capabilities, as demonstrated by recent studies (Garkoti & Kundapura, 2021; Huang et al., 2018; Mengen et al., 2020; Sichangi et al., 2016).

2.4 Estimation of Water Level and Discharge Using Satellite Altimetry

Water levels in big rivers, lakes, and floodplains have been continuously monitored by satellite altimetry, and a time series spanning more than 25 years is now available (Papa et al., 2010). Multiple research works have shown the potential of satellite altimetry to estimate the stream flow in medium to large rivers having a width of a few kilometers, as demonstrated in studies carried out on the Po River in Italy (Tarpanelli, Barbetta, et al., 2013), Niger River (Tourian et al., 2017), Ob River (Kouraev et al., 2004) and several location is Amazon river (E. A. Zakharova et al., 2006) with the use of rating curve developed by correlating altimetry river water level with observed measurements taken for river discharge or development of models.

Thus, recent improvements in radar altimetry technology have led to better monitoring of river water levels, which makes it possible to use these sensors to validate various applications, ranging from simple routing strategies to complex hydraulic models. Though the altimetry mission's spatial-temporal sampling is a limitation (Tarpanelli et al., 2019). Satellite missions with ground tracks that repeat, such as TOPEX/Poseidon and Jason (10-day) or ERS-2 and ENVISAT (35-day), can provide a virtual gauge of river channel stages at different times. This approach increases the temporal and spatial coverage of measurements. The upcoming SWOT satellite mission is expected to further improve this capability in the future (Birkinshaw et al., 2014).

Chen et al. (1998) demonstrated how models may be integrated with remote sensing to produce discharge using data from the Topex/Poseidon satellite to observe sea level changes and comprehend anomalies in sea surface heights. According to their argument, monitoring ocean anomalies can help us comprehend the necessary adjustments to the global hydrologic cycle that led to such anomalies if the ocean is the ultimate repository for all terrestrial water.

Birkinshaw et al. (2010) proposed remote sensing techniques to be used such as SAR to predict discharge at ungauged locations by combining altimetry data and observed measurements of channel cross sections. The method relies on information about the upstream and downstream cross-sectional areas, which could be determined using altimetry data and river geomorphology data from either in situ or remote sensing sources. River width data from remote sensing can be used to track changes in cross-sectional area over time, which when combined with altimetry data, allows for the estimation of discharge at sub-satellite sites.

In the research by Bogning et al., (2018), the evaluation of five altimetry missions was assessed by comparing their results to gauge station measurements. By combining data from all radar altimetry missions, a long term and improved water level time series was obtained. This led to better characterization of water level peaks and more accurate estimation of annual discharge throughout the shared observation period, as the increased data sampling in the river basin contributed to improved accuracy in comparison to in situ mean annual discharge.

This method possesses several limitations which includes dependence on altimetry data quality in the continental water bodies, unavailability of in-situ observations to develop rating curves, assumption of static rating curve while extending time series of discharge and temporal sampling rate (Papa et al., 2010).

2.4.1 Satellite Altimetry: Introduction and Working Principle

During the late 1960s, a new technique called satellite radar altimetry was introduced. This method involves using active microwave observations to determine the distance between the Earth's surface and a satellite. By transmitting an electromagnetic pulse in the nadir direction, the radar sensor can accurately calculate the signal's two-way travel time (Δt). To do this, the radar echo or waveform is recorded using pulse compression and de-ramping methods. Information about the surface's composition can be obtained from the waveform's amplitude and shape, which are linked to the backscattering coefficient. The altimeter range (R), or the distance between the satellite and the surface, is determined when the received power reaches the midpoint of the leading edge at mid-height. It is estimated as

$$R = \frac{c\Delta t}{2} \qquad \text{Equation 2-1}$$

The value of 'c' represents the velocity of light in vacuum.

To ensure precise calculation of surface topography using radar altimetry, it is necessary to adjust the range measurement for several factors. These include atmospheric propagation delays, instrument corrections, and surface geophysical adjustments, as noted by Frappart et al. (2017). The sea-surface height (SSH) over the ocean can be represented using an ellipsoid, as illustrated in Figure 2-1. The height of the reflecting surface (h) is then determined based on this representation:

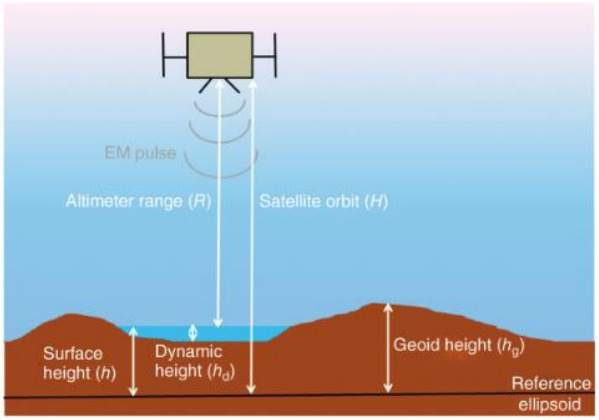
$$h = H - R - \Delta R_{ion} - \Delta R_{dry} - \Delta R_{wet}(-\Delta R_{ssb}) \qquad \text{Equation 2-2}$$

The height of the satellite's center of mass above the ellipsoid, H, is estimated using precise orbit determination techniques. The nadir altimeter range, R, is the distance from the satellite's center of mass to the surface, adjusted for instrumental corrections. The corrections applied to the range, ΔR_i are applied over all types of surfaces. The atmospheric correction, ΔR_{ion}, accounts for the atmospheric refraction range delay caused by the ionosphere's dielectric characteristics and free electron content. The

atmospheric refraction range delay caused by the troposphere's dry gas component is accounted for by the atmospheric correction ΔR_{dry}. The atmospheric refraction range delay caused by the tropospheric water content in clouds and water vapor is accounted for by the atmospheric correction ΔR_{wet}. The range correction over oceans and great lakes, ΔR_{ssb} is the result of the interaction of the altimeter's electromagnetic pulse with the scatterers within the footprint, including wave and surface roughness effects.

Figure 2-1

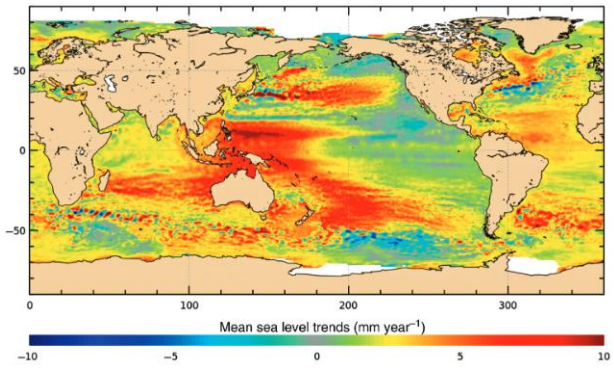
The Principle of Radar Altimetry Measurement (Frappart et al., 2017)



Initially, altimeter products were utilized to observe the surface elevation of oceans and analyze the non-uniform rise in sea level. It has been found that the sea level increase rate is not consistent worldwide, and in specific areas such as the western Pacific, it can be up to three times higher than the global average rate which is shown in Figure 2-2. Later these altimetry satellite missions are being used for observation of large rivers and lakes for long term monitoring of fluctuation in water level (Frappart et al., 2012).

Figure 2-2

Spatial Trend Patterns in Sea Level over January 1993-2014 based on Multi-mission Satellite Altimetry (https://climate.esa.int/en/projects/sea-level/)



Satellite altimetry can help with hydrological research since it keeps track of surface water heights (Calmant & Seyler, 2006). Using specialized software like the Multi-mission Altimetry Processing Software, more accurate altimetry data could be obtained in order to enhance the accuracy of the water stages generated from altimetry. The accuracy and bias of water stages estimated from radar altimetry were evaluated in several studies and found to have good accuracy compared with in-situ observations (Biancamaria et al., 2017).

2.4.2 Satellite Altimetry Products

Satellite altimetry measurements have been continuously extending the time series of inland water level observations since the launch of ERS-1 in 1991 and TOPEX/Poseidon in 1992. With the availability of altimetry data from ERS-2, ENVISAT, Jason-1, and Jason-2, the time series has been further extended (Birkinshaw et al., 2010).

Previous and existing satellite altimetry missions, such as GEOS-3, SeaSat, Geosat, TOPEX/Poseidon, Geosat Follow-on, and Jason-1/2/3 from NASA, as well as ERS-1/2, ENVISAT, CryoSat-2, and Sentinel-3 from ESA, have contributed to the field of satellite altimetry. Other missions, like the Chinese-planned HY-2A or the combined French Indian SARAL/AltiKa project, are also in operation. The summary of satellite altimetry missions is presented in Table 2-1.

Table 2-1

Summary of Satellite Altimetry Missions (Jiang et al., 2017)

Satellite	**Agency**	**Period**	**Altitude (km)**	**Altimeter**	**Frequency Used**	**Repetitivity (Day)**	**Equatorial Inter-Track Distance (km)**
Skylab	NASA	May1973–February1974	435	S193	Ku-band		
GEOS3	NASA	April1975–July1979	845	ALT	Ku and C-band		
SeaSat	NASA	July–October1978	800	ALT	Ku-band	17	
Geosat	US Navy	October1985–January1990	800		Ku-band	17	

Satellite	Agency	Period	Altitude (km)	Altimeter	Frequency Used	Repetitivity (Day)	Equatorial Inter-Track Distance (km)
ERS-1	ESA	July1991–March2000	785	RA	Ku-band	35	80
Topex/ Poseidon	NASA/ CNES	September1992 – October2005	1336	Poseidon	Ku and C-band	10	315
ERS-2	ESA	April1995-July2011	785	RA	Ku-band	35	80
GFO	US Navy/ NOAA	February1998–October2008	800	GFO-RA	Ku-band	17	165
Jason-1	CNES/ NASA	December2001–June2013	1336	Poseidon-2	Ku and C-band	10	315
Envisat	ESA	March2002–April2012	800	RA-2	Ku and S-band	35	80
OSTM/ Jason-2	CNES/ NASA/ Eumetsat/ NOAA	Jun2008–present	1336	Poseidon-3	Ku and C-band	10	315
CryoSat-2	ESA	April2010–present	720	SIRAL	Ku-band	369	7.5
HY-2	China	August2011–present	971		Ku and C-band	14168	
Saral	ISRO/ CNES	February2013–present	800	AltiKa	Ka-band	35	80
Jason-3	CNES/ NASA/ Eumetsat/ NOAA	January2016–present	1336	Poseidon-3B	Ku and C-band	10	315
Sentinel-3A	ESA	February2016–present	814	SRAL	Ku and C-band	27	104

2.5 Multispectral Satellite Data for Monitoring River Discharge

The frequent revisit time and large spatial coverage of satellite sensors have led to their increased use in estimating river flow, particularly given the need for daily measurements of river discharge (Filippucci et al., 2022). The feasibility of monitoring and estimating discharge from space has been demonstrated in numerous studies in difficult river sites ranging medium-sized to large catchments from various sensors, including MODIS, MERIS, Landsat and OLCI datasets (Brakenridge & Anderson,

2006; Shi et al., 2020; Tarpanelli, Brocca, et al., 2013; Tarpanelli et al., 2020). Remote sensing products can be used to estimate hydraulic variables, such as river channel widths and surface water extent, which are essential for river flow estimation. Synthetic aperture radar (SAR) and visible spectrum digital imagery are among the remote sensing techniques used to collect this information and they have been used to monitor the extent of a floodplain's inundation, cross section geometry and variation of water level with river extent (L. Smith & Pavelsky, 2008; Sun et al., 2010).

To estimate river discharge using multispectral images, there are two main approaches available. One approach for estimating discharge from space involves creating a regression model between the reflectivity of the inundated channel, which includes both the inundated and potentially inundated areas, and in-situ measured discharge. A second approach is to create a regression model that relates the width of the river to in-situ discharge measurements. These methods have proven to be successful in estimating river discharge in various rivers (Bjerklie et al., 2003; Li et al., 2019; L. C. Smith & Pavelsky, 2008; Tarpanelli, Brocca, et al., 2013).

Brakenridge & Anderson (2006) demonstrated the usefulness of the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor in hydrology. They showed that the sensor can be used for river discharge estimation along with characterization and detecting the floods. However, they only presented one early case study where MODIS was used to estimate discharge over a short period of time (approximately 30 data points).

Sichangi et al. (2018) computed temporal river width data from the MODIS images to derived river discharge. Only remote sensing data with NSE values larger than 0.5 at the analyzed sites were used to estimate discharge. The calculated velocity's values, which range from 0.63 to 1.2 m/s and 0.81 to 1.35 m/s, were found to be within the bounds of the expected velocity 0.96 m/s. The methodology for calculating discharge is a viable strategy for usage on rivers all across the world because it exclusively relies on global satellite databases.

Sahoo et al. (2020) presented an integrating MODIS and Landsat methodology and tested the method in eastern India's Brahmani River Basin to estimate high-frequency discharge. To increase the temporal frequency of the single datasets, a copula-based

model was suggested, and improved results were then shown. This study on the use of many satellites for different missions demonstrated their value in improving flow monitoring, particularly in highly forested areas.

2.5.1 NIR Band Reflectance as Proxy of Discharge

Optical remote sensing sensors are crucial instruments which detect different spectral signatures based on the wavelengths they observe, allowing them to measure reflected or emitted energy. However, optical remote sensing images are impacted by clouds, haze, and cloud shadows, making it difficult to distinguish between them and other dark objects like water and shadows (Zhu & Woodcock, 2012). Water has low reflectance in near infrared (NIR) wavelengths and absorbs more energy, while non-water has high reflectance due to the strong water absorption in NIR (Ahn & Park, 2020). Any location that experiences flooding will have an increase in water surface, which will lower the area's NIR reflectance value as seen by the satellite. Detecting variations in river discharge can be achieved by analyzing the spectral characteristics of water bodies and other objects in near-infrared (NIR) images (Shanlong et al., 2010).

The Calibration/Measurement (C/M) technique involves using the changes in surface water extent to estimate river flow. This method relies on the idea that the ratio of reflectivity from non-inundated areas and the inundated area is linked to river discharge. First, Brakenridge et al. (2007) used data from the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) at 37 GHz to estimate river flows by comparing the brightness temperature of a pixel over the river with that of an unaffected pixel. Subsequently, MODIS sensor had been used successfully exploiting different behavior of land and water in NIR band for discharge estimation (Sahoo et al., 2020a; Van Dijk et al., 2016). In contrast, if light, atmospheric scattering, or vegetation characteristics change and recorded radiances change simultaneously, the ratio stays constant. Utilizing MODIS optical data to detect surface water changes economically and effectively is the paired-measurement method (Brakenridge et al., 2005).

Conventional sensors can effectively capture surface water dynamics, but the use of surface reflectance ratio (C/M) technique requires less data and offers distinct advantages. There are two primary approaches for implementing the C/M technique. The first approach employs optical satellite data, particularly the near-infrared (NIR) band of the Moderate Resolution Imaging Spectroradiometer (MODIS), to estimate

discharge and detect floods in rivers that are not gauged or are inaccessible. The second approach involves the use of microwave satellite data, such as the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), which is a more complex process requiring additional datasets, such as precipitation data (Temini et al., 2011).

Tarpanelli, Brocca, et al. (2013) utilized this method to estimate river discharge for different sections of the Po River situated in the northern region of Italy. The researchers successfully applied this approach to ungauged river sites and medium-sized basins, exhibiting significant temporal variability. The results of the study revealed the accuracy of the MODIS data in determining flow estimates and highlighted the reliability of this method in providing flow estimates for rivers.

In a study conducted by Shi et al. (2020), the Harmonized Landsat and Sentinel-2 (HLS) surface reflectance product was utilized to estimate the river discharge in small rivers with widths ranging from 30 to 100 meters in the Murray Darling Basin, Australia. The approach employed linear regression between C/M and observed discharge data, and the findings indicated a high level of consistency with the observed discharge.

2.5.2 Satellite Optical Data Sets

Multispectral image data can be obtained for different satellite missions. This data availability depends upon the orbital period of satellite and their spatial and temporal resolution. Terra and Aqua satellites both have the MODIS instrument in operation. Every one to two days, it observes the whole surface of the Earth within a 2330-kilometer viewing swath. The sensors of the instrument are designed to measure 36 spectral bands covering a range of wavelengths from 0.405 to 14.385 m, and provide data for three different spatial resolutions: 250 m, 500 m, and 1000 m (MODIS Web). Likewise, the details of other satellite missions are provided in Table 2-2.

Table 2-2

Details of Multi-Spectral Satellite Missions

S.N.	Data	Agency	Equatorial cross-track separation (km)	Spatial Resolution	Temporal Resolution	Period	Source
1	MODIS AQUA	NASA	2330	250 m, 500 m, 1000 m	1 – 2 days	May 2002 - Present	https://lpdaac.usgs.gov/
2	MODIS TERRA	NASA	2330			Dec 1999 - Present	
3	Landsat ETM +	NASA and USGS	185	30m	16 days	Apr 1999 – Sep 2021	
4	Sentinel 2	ESA	290	10 - 60 m	5 days	June 2015 - Present	https://sentinels.copernicus.eu/
5	Meris	ESA	1150	300 m	3 days	2002 - 2012	https://www.esa.int/

2.6 Multi-Mission Satellite Approach for River Discharge Estimation

By integrating data from various space-based parameters, including river width and river bed velocity from optical sensors and Synthetic Aperture Radar (SAR), along with river stage level data from satellite altimetry, it is anticipated that more precise estimations of river discharge can be obtained compared to relying solely on a single parameter (Sichangi et al., 2016). The spatial sampling of satellite altimetry is limited due to its ground track pattern, which varies from 10 days for the Jason series to 369 days for the CryoSat-2 missions, with the ERS, Envisat, and Saral/AltiKa missions falling in between with a 35-day pattern. Thus, the multi-mission approach has been widely accepted and used in many studies to cope with the limitation of spatiotemporal resolution (Table 2-3) of satellite altimetry and optical images.

The approach involving use of altimetry and optical sensors for river discharge estimation has shown better performance in Niger and Po rivers with NSE of 0.98 and 0.83 (Tarpanelli et al., 2019).

Table 2-3

Previous Studies Using Multi-Mission Approach for Discharge Estimation

Author	Study Area	Data Used	Method	Key Results
Sichangi et al. (2016)	Eight of the largest rivers in the world	Envisat, Jason 2, MODIS and DEM	The empirical equation proposed by (Bjerklie et al., 2003) is utilized to compare the effective river width combined with the river stage to estimate river discharge.	Improved prediction accuracy in most of the river. NSE varied between 0.6 and 0.97 with $R^2 > 0.90$.
Huang et al. (2018)	Upper Brahmaputra River, China	Jason 2/3, Saral AltiKa Landsat 5/7/8 and Sentinel 2 images	Modified Manning's equation and Rating Curve	The NSE vary between 0.65 to 0.97 and RMSE vary between 32.97 to 695 m^3/s.
Bjerklie et al. (2018)	Yukon River, North America	Jason 2 Landsat	Manning's equation and Prandtl-Von Karman equation	The calibrated discharge estimate showed the accuracy of +/- 2%
Garkoti & Kundapura, (2021)	Krishna River, India	Jason 3, Sentinel 3A, Sentinel 3B, Sentinel 1 and Sentinel 2 images	Modified Manning's equation	The NSE vary between 0.53 to 0.62 and R^2 vary between 0.83 to 0.97.
Scherer et al. (2020)	Lower Mississippi River	Envisat, Jason-2/3, Sentinel-3A/3B, Landsat-4/5/7/8, Sentinel 2A/2B	Manning's equation	Time series of river discharge by utilizing satellite altimetry and remote sensing data with an NRMSE ranging from 7 to 35%.

2.7 Application of Artificial Intelligence in Hydrology

John McCarthy is credited with coining the term "artificial intelligence" in 1956, which is now a field of computer science that aims to develop intelligent agents capable of comprehending their surroundings and making decisions that increase their likelihood of success. The term "artificial intelligence" encompasses the ability to hold two different ideas at once while maintaining functionality. However, for an AI system to be effective, it must also possess the ability to reason, respond quickly, and learn from previous experiences. (Singh et al., 2013).

Machine learning is an application to AI. The concept of teaching a computer to learn and improve on its own without direct instruction is known as machine learning, which involves the use of mathematical models to process data. To train a computer to emulate human reasoning, one technique involves using neural networks. These are a collection of algorithms that are modeled after the structure and behavior of the human brain. This allows the computer system to continuously acquire new skills and enhance its performance. Machine learning has been effectively used in remote sensing for source separation, classification, regression, clustering, and coding (Camps-Valls, 2009).

To predict different hydrological parameters, artificial neural networks (ANNs) are often and accurately employed; yet they are typically only created in one or two hidden layers. Deep learning networks have been developed to handle complex problems, and their architecture has been expanded with multi-layered designs. Deep learning is a subfield of machine learning that utilizes artificial neural networks (ANNs) to represent complex concepts by learning at various depths and levels. The unpredictable, complex, and nonlinear character of streamflow makes it viable to utilize AI-based modeling techniques. ANNs can mimic streamflow in particular using additional hydrological factors (Bengio, 2009; Ghumman et al., 2011).

Identifying an appropriate deep-learning network can be a difficult task, as there are many options available with varying degrees of adaptability. It is crucial to choose the correct type of network, as selecting the wrong one could lead to suboptimal results. One type of neural network that has shown promising results for processing continuous (time-series) data such as streamflow is the Recurrent Neural Network (RNN), as per recent studies (Elumalai et al., 2017). Long-term dependencies between the network's inputs and outputs could be discovered through Long Short-Term Memory (LSTM).

Better model performance has been achieved using this method, highlighting the LSTM's potential for application in hydrological modeling applications (Kratzert et al., 2018).

Several research have been done to address the suitability of machine learning in the field of hydrology. Some of the research are discussed below:

Esmaeilzadeh et al. (2017) employed various data-mining methods including ANNs, support vector regression, the M5 tree method, and a hybrid Wavelet-ANN approach to predict daily flow to the Sattarkhan Dam in Iran. Based on the study's findings, the Wavelet artificial neural network (WANN) method proved to be the most effective approach for flow estimation.

Kratzert et al. (2018) discovered that the LSTMs, when applied to 291 catchments to model daily streamflow using the publicly available CAMELS dataset, were more successful as individual catchment models in snow-driven catchments and less effective in arid basins. This is due to the greater importance of long-term dependencies in snow-driven processes, such as snow accumulation and snowmelt. Additionally, the LSTM model outperformed the benchmark conceptual model by a small margin.

Zhang et al. (2018) calculated water tables in agricultural areas using LSTM-RNN. They contrasted the simulation produced by LSTM-RNN with that produced by Multilayer Perceptron (MLP) and discovered that the former performed better than the latter. So, they suggested model can be used as a substitute for existing methods of estimating water table depth, particularly in locations where hydrogeological data are challenging to come by.

To reduce model uncertainty in hourly streamflow projections in the Chinese watersheds of Longquan Creek and Jinhua River, Chiang et al. (2018) utilized ensemble approaches in artificial neural networks. The researchers combined multiple neural networks, which led to a 19-37% improvement in the accuracy of streamflow predictions compared to a single neural network approach.

According to Kao et al. (2020) the Shihmen Reservoir watershed in Taiwan experienced multi-step-ahead flood forecasting using the LSTM-based Encoder-Decoder (LSTM-ED) model,. Results indicated that the suggested model may boost the interpretability

of the model internals and raise the reliability of flood forecasting by converting and connecting the rainfall sequence with the runoff sequence.

Cheng et al. (2021) developed three machine learning models using ANN, Support Vector Regression (SVR), and LSTM to anticipate discharge variation in North China and adequate accuracy was attained. The results showed that MLP performed marginally better than LSTM-RNN and much better than SVR. Additionally, it was shown that ANNs were more effective than other machine learning techniques for modeling and forecasting karst spring outflow.

2.7.1 Long Short-Term Memory (LSTM)

A unique variety of RNN, known as the LSTM architecture, a type of model or framework for sequential data, was created to address the typical RNN's inability to learn long-term dependencies. It implements gates that control "memory cells" using a unique combination of hidden units, elementwise products, and sums between units. These cells are made to store data unaltered over extended periods of time (Hochreiter & Schmidhuber, 1997). The key advantage of LSTM is that it can learn long-term dependence, which is not achievable with straightforward RNNs.

Runoff was modelled for Fen River basin, China by Hu et al. (2018) using traditional ANN and LSTM models by and results showed LSTM performed better in terms of time series discharge simulation. The LSTM model is more cognitive than the ANN model, and the ANN model is more sensitive when compared to simulations of flooding events.

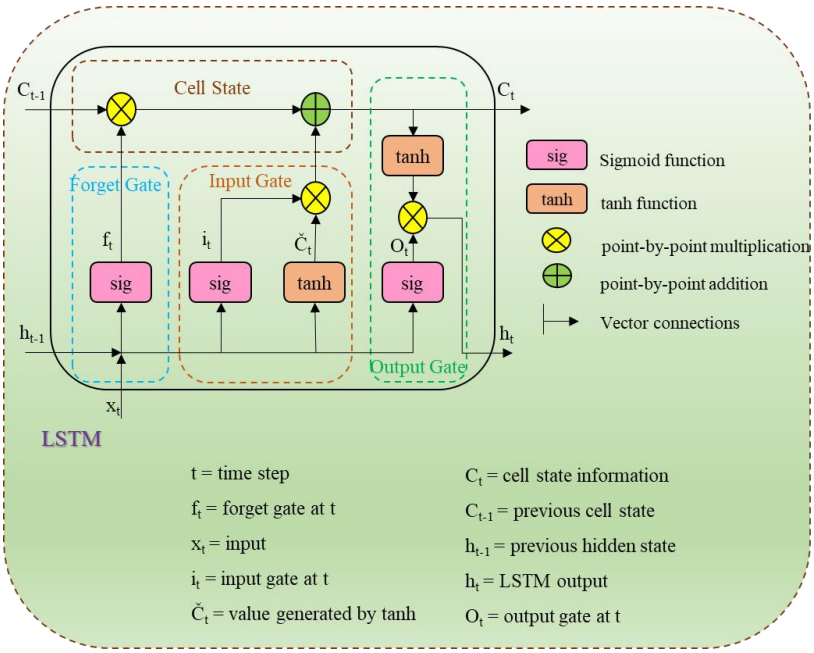
Recently, Nogueira Filho et al. (2022) conducted a study to investigate the performance of LSTM in comparison to FFNN and conceptual models for a regional method in a catchment with limited data. The study found that the LSTM model outperformed the Feedforward model in the specific catchment, showcasing its ability to model the hydrological dynamics of streamflow in semiarid areas. Additionally, the neural networks trained using regionalization data performed better than those trained for individual catchments, demonstrating the capacity to incorporate process understanding from various watersheds.

LSTM Architecture

Figure 2-3 depicts the basic LSTM unit, which consists of a cell with an input gate, output gate, and forget gate. To handle the disappearing or ballooning gradient problem, LSTMs employ the idea of gating (Greff et al., 2017). The three gates present in a recurrent neural network can be considered as artificial neurons that compute an activation, using an activation function, based on a weighted sum of the current data x_t, the previous hidden state h_{t-1}, and a bias term b. The main function of the cell is to retain information for a flexible amount of time (Ordóñez & Roggen, 2016).

Figure 2-3

Illustration of a Long-Short Term Memory Unit



Forget gate functions as a filter that decides what information is significant and what can be disregarded. It utilizes a sigmoid function to compute the importance of the old output, which is a weighted sum of the current input X_t, the previous hidden state h_{t-1}, and any bias b. The resulting values, ranging from 0 to 1, are used for point-by-point multiplication.

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

Input gate performs the following operations to update the cell status.

The second activation function in the LSTM architecture is responsible for determining the importance of the current input and the previous hidden state. Its output is also within the range of 0 (not important) to 1 (important). Next, a hyperbolic tangent (tanh) function is applied to the current input and the previous hidden state, producing a vector (Ct) that regulates the network by holding all potential values between -1 and 1. The point-wise multiplication of the outputs from the activation functions follows this step.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)$$

$$C_t = tanh(W_c.[h_{t-1}, x_t] + b_c)$$

Output gate computes the value of the next hidden state, which holds information about prior inputs. It employs a sigmoid function to evaluate the significance of the current and prior hidden states. The tanh function is then applied to the new cell state, generated from the original cell state, and multiplied point-by-point with the results from the sigmoid function. The resulting value from the activation of the gates determines which information should be retained in the hidden state, which is then used for prediction.

In RNNs, the updated hidden and cell states are used for prediction in the next time step.

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

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

At conclusion, the forget gate selects whatever pertinent information from the earlier processes is required to conclude. The output gates contribute to the next hidden state, while the input gate determines which relevant information should be incorporated from the current stage.

2.8 Relevant Studies in Mekong River Basin

Several studies have been conducted in Lancang Mekong River Basin for the estimation of river discharge using remote sensing. Some of them are discussed.

(Kim et al., 2019) used multi-mission radar altimetry data to estimate discharge at three different locations in the Lower Mekong River by employing Ensemble Learning Regression (ELQ) technique, which generated multiple rating curves. The study found

that ELQ improved Q estimation by using various H values derived from Jason-2 altimetry, regardless of the distance from in-situ Q stations and provided more accurate results than using a single rating curve.

Birkinshaw et al. (2014) utilized Landsat satellite imagery along with ERS-2, ENVISAT, and satellite altimetry data to estimate the discharge at Nakhon Phanom and Vientiane on the Mekong River for ten years. The combination of these data sources allowed for the creation of a time series of river channel water levels and the channel slope, resulting in a Nash-Sutcliffe efficiency value of 0.90 for Nakhon Phanom and 0.86 for Vientiane to obtain river discharge.

A novel decile thresholding method was developed by Mengen et al. (2020) using the concept of at-many-stations hydraulic geometry (AMHG) in Mekong River. Although the decile thresholding method performs well for low-flow rates, it often leads to underestimation of peak flows. The decile thresholding discharge estimation outperformed the Gleason and Wang's optimized AMHG technique, with an RRMSE of 19.5 percent for the overall examined period and 16 percent for just the dry seasons.

However, the discharge estimated by above and other studies are site specific only and no studies have been carried out to develop regional model for whole river. This study is intended to develop a single regional model which can be used to predict discharge at ungauged sites of the river.

CHAPTER 3
STUDY AREA AND DATA

3.1 Study Area

The Tibetan Plateau serves as the starting point for several major rivers, including the Yangtze, Salween, Irrawaddy, Red River, and Mekong. The Mekong River basin spans approximately 795,000 km^2, originating from the Tibetan Plateau and flowing for 5000 km southwards before reaching the sea. The river passes through China (Yunnan Province), Myanmar, Laos, Thailand, Cambodia, and Vietnam. Table 3-1shows the respective areas in the Mekong River Basin for each country. The Greater Mekong River can be divided into two basins: the Upper Basin, where the river is called Lancang and is located in Tibet and China; and the Lower Mekong Basin, which runs from Yunnan to the sea (as shown in Figure 3-1).

Figure 3-1

Map of Study Area

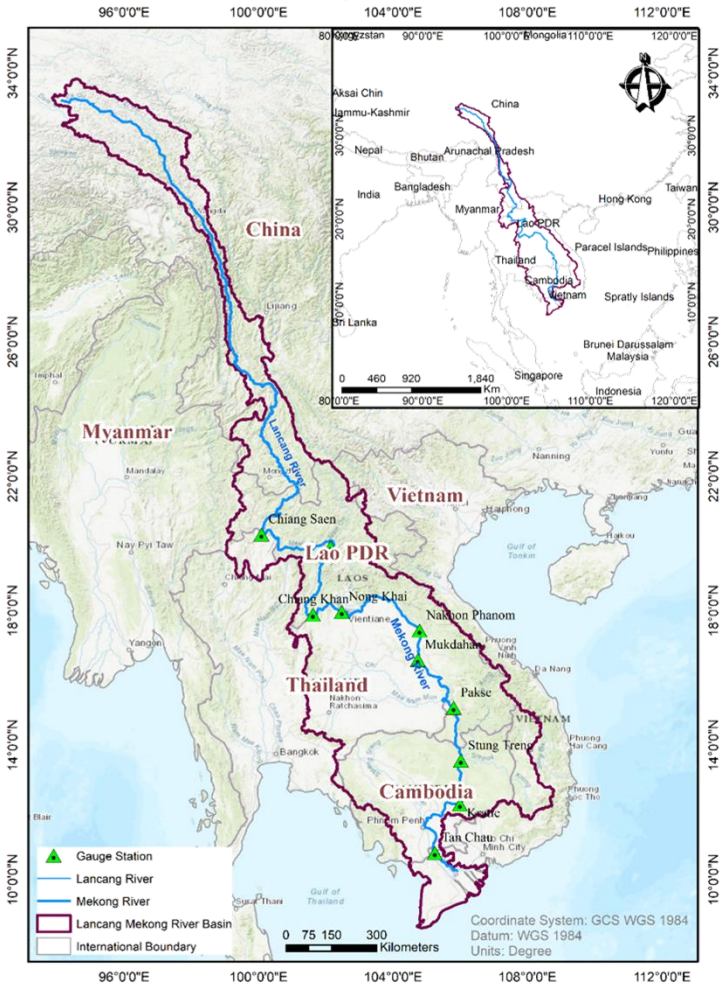


Table 3-1

Country Wise Area in the Lancang Mekong River Basin

Basin	Area (km^2)	Countries	Area of country in basin (km^2)	% of the total area of the basin	% of the total area of the country
Mekong	795,000	China	165,000	21	2
		Myanmar	24,000	3	4
		Lao PDR	202,000	25	85
		Thailand	184,000	23	36
		Cambodia	155,000	20	86
		Vietnam	65,000	8	20

The Upper Basin (China and Myanmar), which contributes around 15% of the flow, accounts for 24% of the entire catchment. There are no large tributaries, therefore all future water resource development will follow the main Mekong. On the other hand, the Lower Mekong Basin is also nourished by sizable tributaries, which account for around 85% of the annual flow.

3.1.1 Climate

In the region upstream, the climate is cold, whereas in the region downstream, the climate is tropical. Yunnan province in southern China's Upper Basin also experiences monsoons, though there are significant topographic variations. Yunnan province experiences different climate zones ranging from subtropical and tropical monsoons in the south to temperate monsoons in the north. However, there is a significant variation in the timing of the start of the southwest monsoon in Yunnan province from year to year, which affects the rainfall pattern in the Upper Basin of China. While the yearly precipitation level drops to as low as 600 mm in the north, the seasonal rainfall pattern is similar to that in the Lower Basin. The variations in climatic zones of the Lancang Mekong Basin are shown in Figure 3-2.

The southwest monsoon, which typically lasts from May through late September or early October and coincides with the Lower Basin's flood season, dominates the region's climate. In most areas of the basin, there are frequent periods of intense rainfall lasting one or two days. The wettest months of the year are August, September, and in the delta even October due to tropical cyclones that affect most of the region later in the season. Less than 1,500 mm of rain fall each year averages throughout the floodplain of Cambodia and the Mekong Delta, while more than twice that amount falls in the

Central Highlands of the Lao People's Democratic Republic and inside the major valley at Pakse (MRC 2005). The is variation of precipitation between Upper Mekong and Lower Mekong is presented in Table 3-2.

Figure 3-2

The Climatic Zones of the Lancang Mekong River Basin (He et al., 2009)

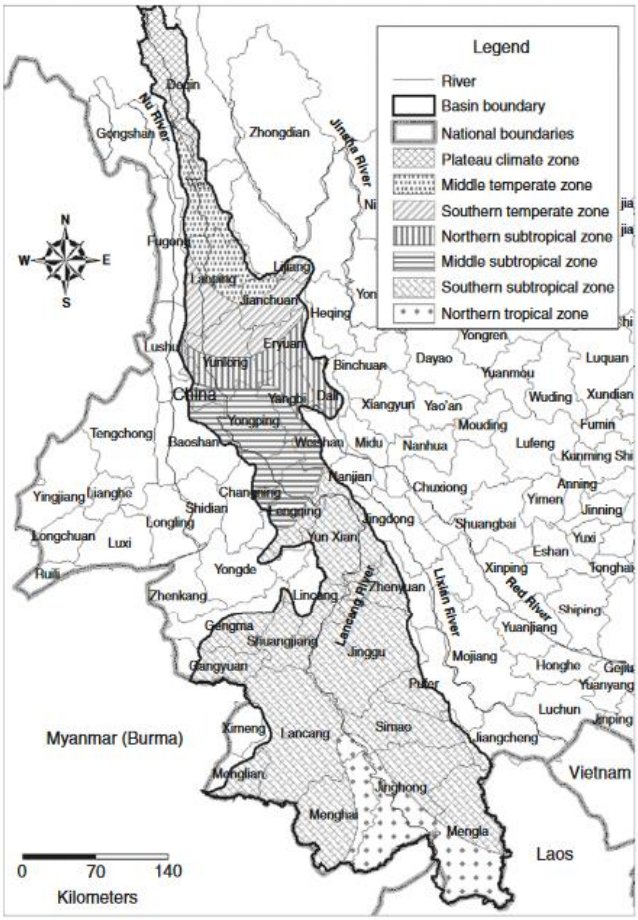


Table 3-2

Monthly Average Precipitations in the Mekong River Basin (in mm) (He et al., 2009)

Basin	Month	January	February	March	April	May	June	July	August	September	October	November	December	Annual
Upper Mekong	Rainfall	17	19	27	41	98	174	224	228	132	102	43	19	1125
	%	1.52	1.65	2.43	3.67	8.70	15.5	19.9	20.3	11.7	9.10	3.81	1.72	100
Lower Mekong	Rainfall	8	15	40	77	198	241	269	292	299	165	54	14	1672
	%	0.5	0.9	2.4	4.6	11.8	14.4	16.1	17.5	17.9	9.9	3.2	0.8	100

3.1.2 Hydrology

The Mekong River ranks eighth in the global basins with a mean annual flow into the South China Sea of around 14,500 m^3/s. Although the flow from the Upper Basin makes up just 13% of the normal annual flow, it may make up to 30% of the flow during the

dry season (MRC 2010). A large network of tributaries makes up the Mekong River basin, creating several sub-basins. In the Lower Basin, significant tributary networks emerge. These systems may be split into two categories: those that contribute to the main rainy season flow and those that drain low relief areas with lesser rainfall. Tropical monsoonal regions are typically characterized by an abundance of water and a relatively consistent flow pattern. However, the flow continues to vary dramatically from year to year (MRC, 2018). The estimated mean annual flow of the basin is almost 460 km^3 and of the total annual flow, in an average year about 75 per cent occurs within just four months between July and October.

The mean annual runoff in the basin is shown in Figure 3-3. In the Lancang Mekong watershed, river discharge rises sharply from north to south, with significant regional variations which is shown in Figure 3-4.

Figure 3-3

Mean Annual Runoff in the Mekong River Basin (MRC, 2010)

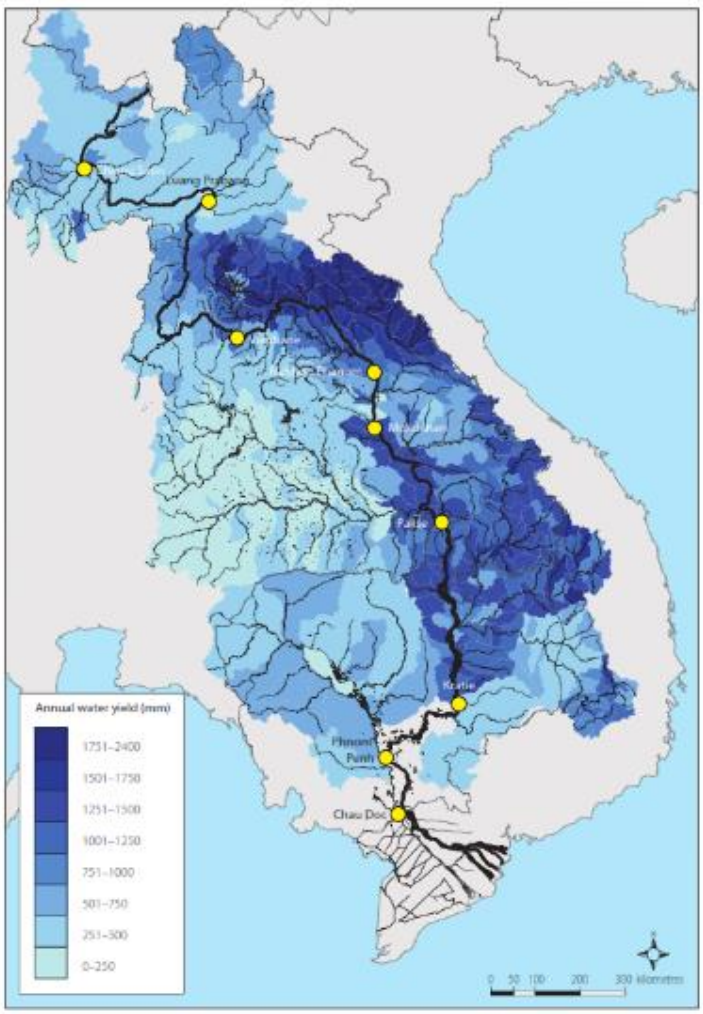
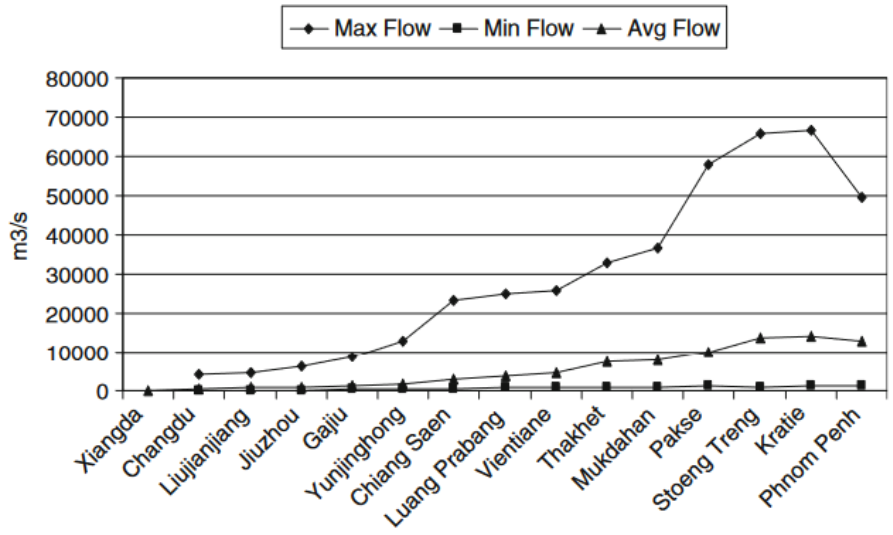


Figure 3-4

Maximum, Minimum and Mean Annual Flows at Sites Along the Lancang Mekong River (He et al., 2009)



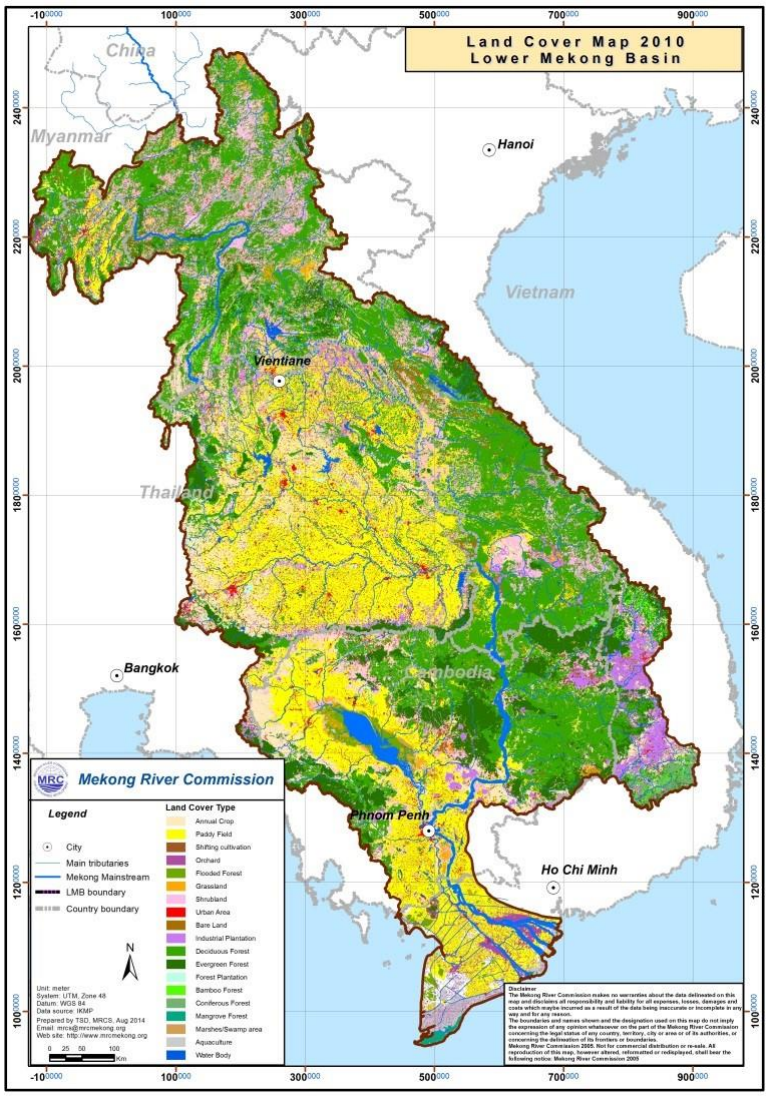
3.1.3 Land Use and Land Cover

The river and its tributaries are constrained by narrow, steep gorges in the upper Mekong River Basin. In this region of the basin, there are few tributary river systems. The river alters as the floodplain widens, the valley expands, and the river widens and slows. Soil erosion is the main issue here. The primary determinants of the river's hydrology are the climate, geography, and land use of the lower Mekong Basin. The soil recovers after upland shifting agriculture (slash and burn), but the vegetation takes much longer. Similar to other areas of the basin, shifting and permanent cultivation have gradually decreased the amount of forest cover over the past three decades. Over the past 50 years, Thailand's Lower Basin regions have had the greatest rate of forest cover loss among all the countries along the Lower Mekong. The land use and land cover map of the Mekong Basin is presented in Figure 3-5.

The most prevalent soil types are salty and sandy, making a large portion of the area unsuitable for growing wet rice. However, agriculture is intense despite the low fertility. The main crops are cassava, maize, and glutinous rice. The main hydrological risk in this area is drought.

Figure 3-5

Land Use and Land Cover Map of Mekong River Basin (MRC, 2010)



3.2 Data

3.2.1 In-situ River Discharge

In this study, daily in-situ discharge data was obtained from Mekong River Commission (MRC) data portal (https://mrcmekong.org/) both for training and validation purposes. Table 3-3 presents the details of gauged stations and their locations on Lancang Mekong River which were used in the study. These gauged stations represent the different hydrological characteristics in the basin. This data was used to develop the LSTM model for discharge prediction and evaluate the performance on gauged and ungauged reaches. The location of these stations is indicated in Figure 3-1.

Table 3-3

List of In-Situ Gauge Stations in the Mekong Mainstream used in this Study

S.N.	Name of Station	Station Code	Basin Coverage	Location		Streamflow Record
			km^2	Latitude (°N)	Longitude (°E)	
1	Chiang Saen	TH_010501	213027	20.273	100.083	Jan 1, 2003 to Dec 31, 2018*
2	Luang Prabang	LA_011201	237938	19.892	102.137	
3	Nong Khai	TH_012001	304990	17.887	102.739	
4	Nakhnon Phanom	TH_013101	353753	17.40	104.800	
5	Mukdahan	TH_013402	399746	16.540	104.737	
6	Pakse	LA_013901	510889	15.117	105.800	
7	Stung Treng	KH_014501	624016	13.545	106.017	
8	Kratie	KH_014901	726763	12.240	105.987	
9	Tan Chau	VN_019803	768492	10.803	105.243	
10	Jinghong	CN_092600	153386	22.00	100.770	
11	Chiang Khan	TH_011903	295193	17.884	101.612	

**The period varies based on data availability*

3.2.2 Satellite Altimetry Data for Water Level

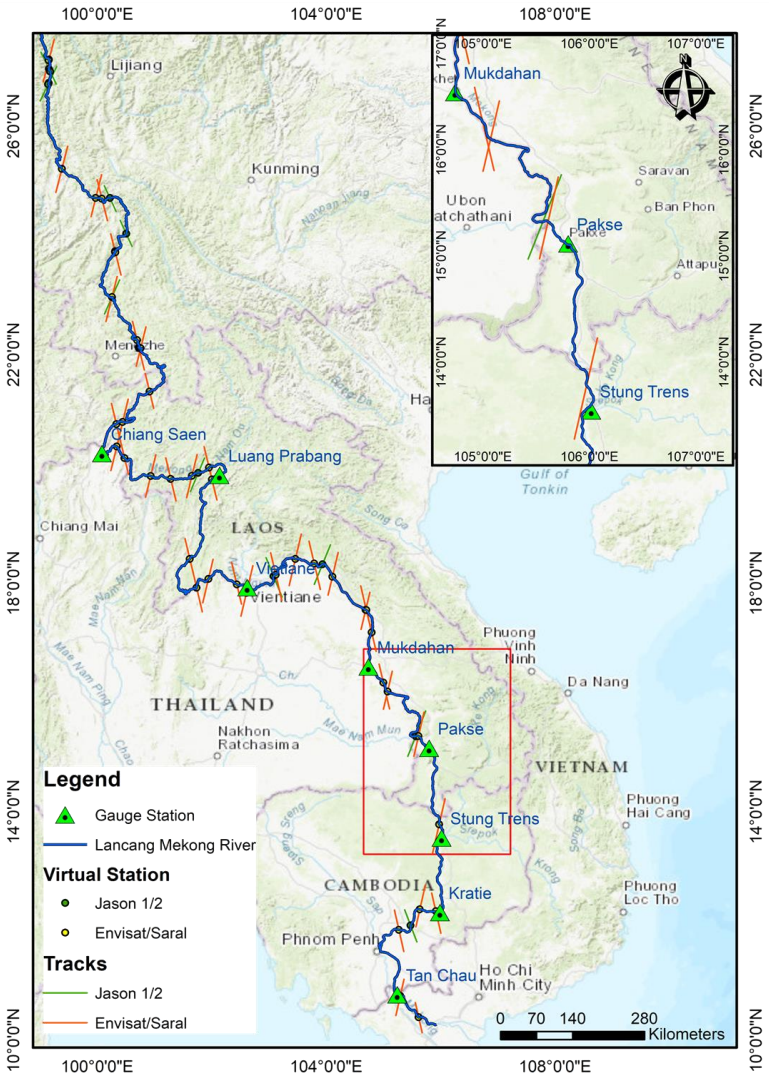
This study used Jason-1/2, Envisat and Saral altimetry products to derive time series water level from 2003 to 2018 since these datasets are freely available by the service provider. Also, from these altimetry products, sufficient virtual stations can be located along the Lancang Mekong River. Figure 3-6 shows the altimeter tracks crossing Lancang Mekong River and Virtual Stations. Envisat provides data from 2002 to 2010 and has an orbital period of 35 days. For Envisat, the publicly available Geophysical Data Record (GDR) from the Center for Topographic Studies of the Ocean and Hydrosphere (CTOH; http://ctoh.legos.obs-mip.fr/data/alongtrack-data/datarequest) includes 18-Hz along-track range data.

Meanwhile, Jason-1 and Jason-2 satellites offer surface observations of the Earth between 66 degrees North and 66 degrees South. These satellites have a 1336 km altitude and a 66-degree inclination. They follow a 10-day repeat cycle and have a ground-track spacing of around 315 km at the equator.

In December 2001, CNES and NASA launched the Jason-1 mission, which employs the Poseidon-2 altimeter, a two-frequency altimeter with C (5.3 GHz) and Ku (13.575 GHz) bands used in previous missions like Topex/Poseidon. Jason-1 also features a Microwave Radiometer from NASA, a DORIS instrument from CNES, a Black Jack Global Positioning System receiver from NASA, and a Laser Retroflector Array (LRA) from NASA/Jet Propulsion Laboratory (JPL) to enable precise orbit determination. It provides Earth surface observations between 66° latitude North and 66° latitude South at an altitude of 1336 km and an inclination of 66° on a 10-day repeat cycle with an equatorial ground-track spacing of approximately 315 km. Jason-1 operated until June 21, 2013, and remained in its intended orbit until January 26, 2009.

Figure 3-6

Virtual Stations Track Along the Lancang Mekong River



In June 2008, CNES, NASA, EUMETSAT, and NOAA collaborated to launch the Jason-2 mission. Its payload includes the Advanced Microwave Radiometer (AMR) from JPL/NASA, the Global Navigation Satellite System (GNSS), and a Laser Retroflector Array (LRA) from JPL/NASA. It also features the DIODE real-time monitoring system of DORIS device from CNES, a GNSS transceiver from NASA/JPL, and the Poseidon-3 radar altimeter from CNES. Jason-2 operated in its nominal orbit until July 3, 2016.

The SARAL (Satellite with ARgos and ALtiKa) mission is a collaboration between the Indian Space Research Organisation (ISRO) and the Centre National d'Études Spatiales (CNES) that aims to examine ocean circulation and sea surface elevation using altimetric measurements. The mission was launched on 25 February 2013, and its main instrument, the ALtiKa altimeter, is the first of its kind to operate in the Ka-band in space. The ALtiKa altimeter was designed and built by CNES and is a key component of the SARAL mission. Table 3-4 provides the major characteristics of altimetry missions used in this study.

Table 3-4

Major Features of the Radar Altimetry Missions used in this Study

Mission	**Jason-1/2/3**	**ENVISAT**	**Saral**
Instrument	Poseidon-2 / Poseidon-3 / Poseidon-3B	Radar Altimeter (RA-2)	Altika
Space agency	Centre National d'Etudes Spatiales (CNES) / National Aeronautics and Space Administration (NASA)	European Space Agency (ESA)	Centre National d'Etudes Spatiales (CNES) / Indian Space Research Organization (ISRO)
Operation	2001–2013 / Since 2008 / Since 2016	2002–2012	Since 2013
Acquisition mode	Low Resolution Mode (LRM)	LRM	Pulse Limited Mode
Acquisition	Along-track	Along-track	Along-track
Frequency (GHz)	13.575 (Ku) / 5.3 (C)	13.8 (Ku) / 13.575 (Ku) / 3.2 (S)	35.75 (Ka)
Altitude (km)	1315	800	800
Orbit inclination (°)	66	98.55	98.55
Repetitively (days)	9.9156	35	35
Equatorial cross-track separation (km)	315	80	80

3.2.3 Satellite Optical Data Sets

The different satellite missions used in this study for the optical images were moderate-resolution imaging spectroradiometer (MODIS); MODIS AQUA and TERRA. Due to the coarse spatial resolution, MODIS has provided better results in large rivers with a width of more than 100 m for surface reflectance ratio estimation and thus selected for Mekong River. The optimal location for selecting wet pixel is the river area and inundation sensitive area during flood events ensuring minimum vegetation coverage while for dry pixel, area should be located far from river area preferably urban areas (Shi et al., 2020; Tarpanelli, Brocca, et al., 2013).

The sensors on the Earth Observing System (EOS) Terra and Aqua satellites, including the widely used MODIS sensor, are used for monitoring various terrestrial, atmospheric, and oceanic phenomena. MODIS is known for its moderate spatial resolution, high spectral resolution, and high temporal resolution, providing data every 1-2 days with two passes a day at mid-latitudes. MODIS has a total of 36 bands with wavelengths ranging from 0.4 m to 14.4 m. In this study, the daily level-2 products MOD09GQ and MYD09GQ from Terra and Aqua, respectively, are utilized.

Table 3-5 provides the major characteristics of optical satellite sensors used in this study.

Table 3-5

Major Features of the Optical Satellite Sensors Used in this Study

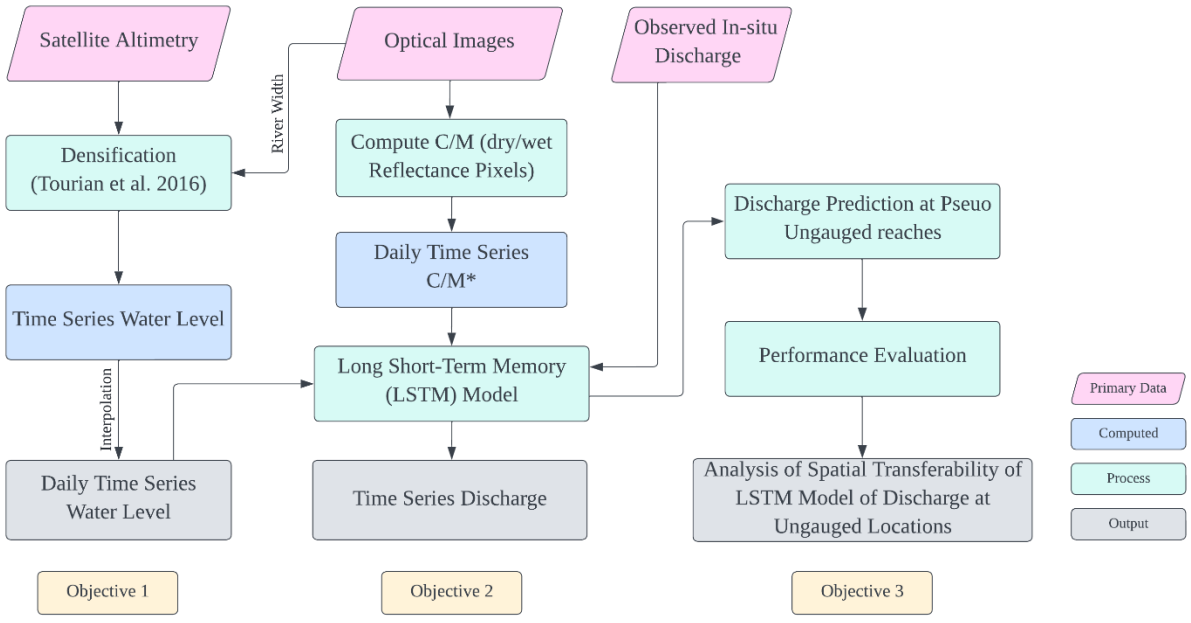
Optical Sensor	Product	Band (Spectral range in nm)	Temporal Resolution (days)	Spatial Resolution (m)
MODIS AQUA	MYD09GQ	2 (841-876)	1-2	250
MODIS TERRA	MOD09GQ	2 (841-876)	1-2	250

CHAPTER 4
METHODOLOGY

The overall methodology adopted for this research is presented in Figure 4-1. The details of the methodology for each objective are explained further in this chapter. Long Short-Term Memory machine learning algorithm was employed for predicting river discharge and evaluated the performance of the developed model. The input to the model are water level data obtained from satellite altimetry and reflectance ratio of dry and wet pixels from optical images. The performance of the model to simulate river discharge is evaluated using different statistical indicators. Further, the applicability of developed framework for discharge prediction at ungauged reaches of the Mekong River was evaluated by considering some pseudo ungauged reaches to represent the ungauged reach. Also, spatial-temporal variation of discharge was analyzed within the Mekong River.

Figure 4-1

Overall Methodological Framework for the Study

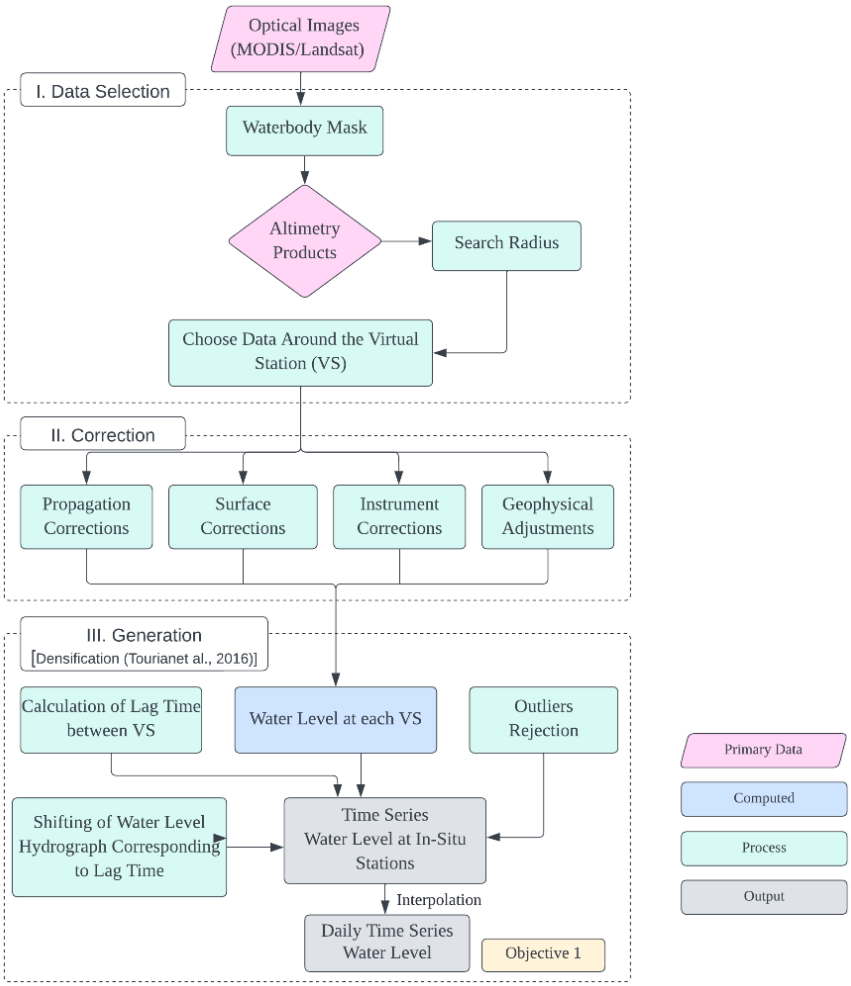


4.1 Obtaining Water Level at Different Stations in Mekong River

The methodological approach to obtain the water level time series at different reach of river using multi-mission satellite altimetry is shown in Figure 4-2.

Figure 4-2

Methodological Approach to Obtain Water Level Using Satellite Altimetry



4.1.1 Water Level Time Series from Satellite Altimetry at Virtual Stations

Satellite remote sensing provides an excellent opportunity to obtain freely accessible water level heights, especially in remote and complex areas. Several altimetric missions, such as TOPEX/Poseidon, Jason-1/2/3, ERS-1/2, Envisat, CryoSat-2, SARAL/AltiKa, and Sentinel 3A, 3B, have collected data over the past twenty years that can be used to monitor the behavior of water bodies on Earth (Garkoti & Kundapura, 2021). The improved precision of altimetry sensors has led to their widespread use as a means of verifying a variety of applications, ranging from simple routing techniques to intricate hydraulic models. In addition, recent technological developments in radar altimetry have become invaluable in monitoring river water levels (Domeneghetti et al., 2015; Schneider et al., 2017).

Following the ERS-2 mission, ENVISAT and SARAL missions were launched with the same orbital configuration featuring a repeat cycle of 35 days and a cross-track

resolution of less than 80 km. ENVISAT carried a dual frequency radar altimeter operating in both Ku and S bands. However, its orbital height was lowered in 2010, which changed its repeat cycle and ground tracks. The study focused on the high frequency (18 Hz) Ku-band data collected during the regular operation mode from 2002 to 2010, in the Sensor Geophysical Data Records (SGDRs) format. The SGDRs contained altimetry data, including satellite position and timing, distance between the satellite and surface, and corrections. ENVISAT had four ranges, Ocean, ICE-1, ICE-2, and SEAICE, calculated by standard waveform retracking algorithms. Although the ICE-1 retracker was designed to observe ice sheets, it was also used in the study to determine range measurements for inland waterbodies due to their similar waveform.

Jason-2, which was launched in June 2008, followed the orbital configuration of its predecessors Topex/Poseidon and Jason-1, with a cross-track resolution below 315 km and a repeat cycle of 10 days. High-frequency ranges from the Sensor Geophysical Data Records (SGDR) altimeter products were used to estimate water levels. However, changing retracking algorithms along a single satellite track can lead to height offsets. To avoid these errors, the ICE1 retracker was used to obtain all altimeter range measurements of a specific inland water body. The height of the reflecting water surface relative to an ellipsoid or a geoid is estimated using a formula that considers the propagation delays of electromagnetic waves due to their interaction with the atmosphere and geophysical entities.

$$h = H - R - C_{ionosphere} - C_{drytroposphere} - C_{wettroposphere} - Cs_{olidEarthtide} - C_{poletide}$$

where 'h' represents the height of the reflecting water surface, which can be a water body or river section. 'H' refers to the altitude of the satellite relative to an ellipsoid, while 'R' represents the altimeter range. '$C_{ionosphere}$' is a correction factor that accounts for delayed propagation through the ionosphere. '$C_{drytroposphere}$' and '$C_{wettroposphere}$' are corrections for delayed propagation in the troposphere, taking into consideration pressure and humidity variations. Additionally, '$C_{solidEarthtide}$' and '$C_{poletide}$' are corrections that consider crustal vertical motions from solid and polar tides, respectively

One of the primary obstacles encountered was the challenge of comparing water level measurements obtained from different satellite missions, taken at various locations and

with distinct characteristics. The issue of intersatellite biases made it challenging to combine water level measurements from different altimetry missions. Furthermore, errors in atmospheric corrections and data interpolation along the track can lead to significant inaccuracies, especially in inland water bodies and rivers where the topography is non-uniform (Tourian et al., 2016). Radial errors from a multi-mission analysis are used to correct range biases for altimeter measurements over land. The errors are interpolated over land and the same retracker is used for all measurements to minimize differences between altimeter tracks and enable the use of multiple missions as a virtual system. To address the issue of intersatellite bias, relevant studies were consulted to obtain the necessary information, which was then applied accordingly. The mean range bias with respect to Topex for ENVISAT, Jason-1, Jason-2 and SARAL are taken as 450.8 ± 7.9 mm, 97.3 ± 1.3 mm, -4.7 ± 1.0 mm and -67.5 ± 1.7 mm respectively from Schwatke et al., 2015.

Using this procedure, the water surface level is calculated from multiple satellite altimetry data for each virtual station at the corresponding time. The accuracy of obtained water level at virtual stations was evaluated comparing with the water level obtained from DAHITI water level database (https://dahiti.dgfi.tum.de/en/) for selected virtual stations.

4.1.2 Densification of Altimetry Derived Time Series Water Level for In-Situ Stations

Multi-mission altimetry products were used to increase the temporal resolution of water level in this study. The time series water level was derived at various Virtual Stations (VS) from different satellite missions. The water level obtained from different satellite missions had bias within due to mission's geophysical data and these intersatellite bias was estimated and removed and other errors of the altimetry data was corrected as mentioned. Then, the virtual stations were connected hydraulically and statistically using an algorithm developed by (Tourian et al., 2016, 2017) using the time series water level data, river width and slope. For this, time lag between virtual stations was also estimated.

The time lag between VSs was determined using the anticipated river width and the slope obtained from satellite altimetry using the Equation 4-1.

$$T_L = \frac{L}{C} = \frac{L}{bV} \qquad \text{Equation 4-1}$$

Where, T_L is travel time, L is the distance between the virtual stations, c is celerity, b = constant=5/3 and V is the velocity. The velocity is computed by the Equation 4-2 developed by (Tourian et al., 2016).

$$V = 1.48W^{-0.8}S^{0.6} \qquad \text{Equation 4-2}$$

where, W refers to river width and S is the slope of the river.

Then, at any location along the river, the combination of all the altimetry data involved aligning the water level data from various virtual stations to a reference location by adjusting the time lag. The data was combined by normalizing the time series and the readings were then rescaled down to their actual water level values after an outlier identification procedure.

River Width Estimation:

The study utilized the RivWidthCloud algorithm developed by Yang et al. (2020) in the Google Earth Engine (GEE) to extract the width and centerline of rivers from remotely sensed images. This algorithm is advantageous due to its ease of use within the GEE cloud computing environment and its ability to automatically minimize the negative effects of clouds and shadows. These benefits eliminate the need for users to download and locally process remote sensing data, making it easier for them to extract common width statistics and time series.

Initially, the river mask was extracted from the satellite image. From the river mask, the centerline of the river was derived, and the river width was obtained at each section along the centerline. The width was calculated by determining the direction orthogonal to the local centerline for each pixel, and then computing the river width along these orthogonal directions. The extracted centerline was also utilized to define the along-river coordinate distance between virtual stations and to compute the slope of the river.

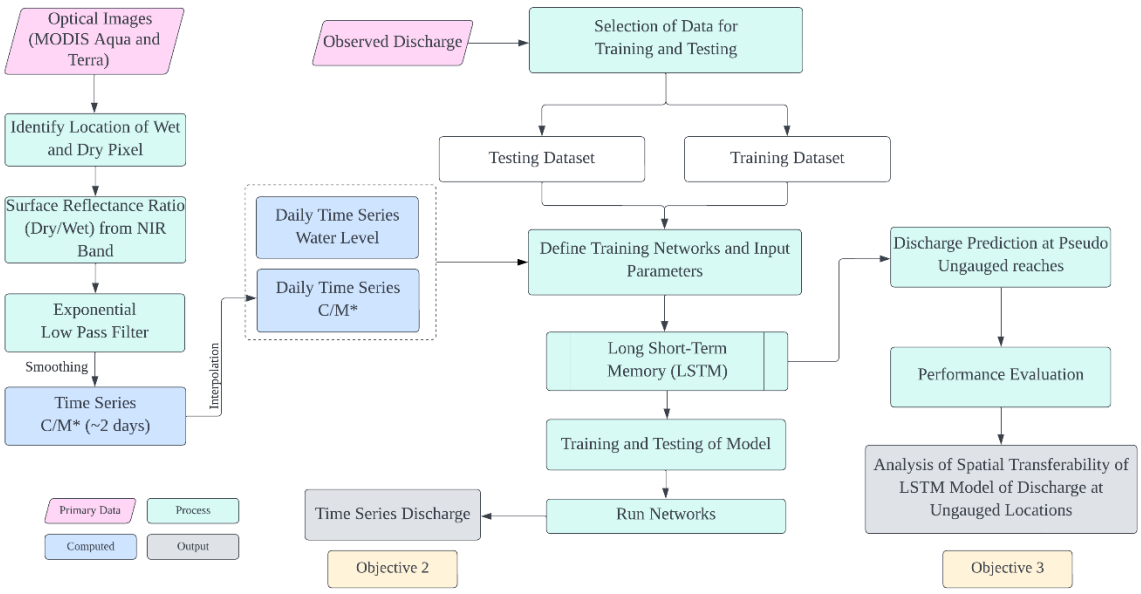
4.2 Developing LSTM Model for Predicting Daily River Discharge Using Surface Reflectance Ratio and Altimetry Derived Water Level

The methodological framework to accomplish this objective is shown in Figure 4-3.

The Near Infra-Red (NIR) band of optical sensors can distinguish between water and land surfaces based on their surface reflectance, with water pixels (M) having lower reflectance values than land pixels (C). The reflectance values of water pixels decrease with flooding events, while those of land pixels remain relatively constant. Therefore, measuring the variation in the time series of water pixel reflectance with respect to land pixel reflectance indirectly provides information on river discharge (Tarpanelli et al., 2017). The ratio of the spectral reflectance values of C and M pixels theoretically minimizes atmospheric effects, as atmospheric disturbances affect both C and M in the same way (Tarpanelli, Brocca, et al., 2013). The C/M ratio increases with water presence and discharge, making it proxy for measurement of river discharge.

Figure 4-3

Methodological Framework for Discharge Prediction using Optical Sensor Data and Altimetry Data



A cloudy sky, on the other hand, renders the optical pictures useless and causes data outages in the measurements. Also, the selected location of C and M pixels may also affect the accuracy of discharge estimation. The data from more satellite missions of the same kind might be gathered and combined to provide the necessary information as a feasible solution in these circumstances. Thus, this study used Long Short-Term

Memory (LSTM) approach to combine observations from optical sensors with altimetry data for estimation of river discharge to overcome these issues.

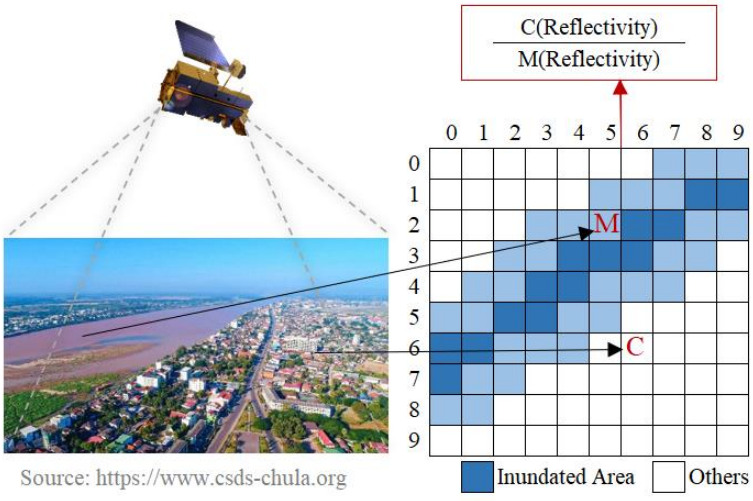
4.2.1 Surface Reflectance Ratio from Satellite Optical Data Sets

For each river reach, the surface reflectance ratio of a land pixel known as C (calibration), which was selected from an area near the river that remains free of surface water even during high flooding events, and a water pixel known as M (measurement), was chosen from within the river where water is permanently present, was calculated. Figure 4-4 shows the typical representation of wet and dry pixel reflectivity. First, the cloud-affected pixels need to be located and eliminated. Then, the location for selecting M and C pixels is chosen. The optimal site for wet pixel M is next to a river in a region that is both entirely submerged in water and sensitive to changes in the flooded area during flood occurrences. The locations of C and M pixels were determined adopting the method described by Tarpanelli, Brocca, et al., 2013.

An area with 38 x 38 pixels surrounding the gauging station was selected for each investigating station. At first, cloudy images were filtered; the pixel with a value of band 1 higher than 0.2 were excluded since optical images are contaminated by clouds and their shadows which impacts surface reflectance values whereas band 2 provides the river discharge signal. Also, visual inspection was done to analyze cloud free images.

Figure 4-4

A Representation of Wet and Dry Pixel Reflectivity



To mitigate the impact of atmospheric noise on the obtained time series of C/M, a low pass filter (averaging moving window) was applied to smooth the data. The resulting smoothed C/M values were referred to as C/M*.

4.2.2 LSTM Model Development

The Long Short-Term Memory (LSTM) networks are designed to remember and learn long-term patterns, overcoming problems related to the vanishing and exploding gradients. The LSTM architecture consists of an input layer having neurons equal to the number of explanatory variables, one or more memory cells, and an output layer (Zebin et al., 2018). The memory cells are the core feature of LSTM networks, and they contain three gates that control and adjust the cell state(s_t), namely the forget gate (f_t), input gate (i_t), and output gate (o_t).

During each time-step t, the three gates receive input x_t (one element of the input sequence) and the previous output h_{t-1} of the memory cells at the previous time-step t-1. The gates act as filters and serve different functions:

- The forget gate determines which information to discard from the cell state.
- The input gate decides which information to add to the cell state.
- The output gate determines which information from the cell state should be used.

The gates have numerous parameters, including weights and biases, which are learned or updated during the network's training process.

The LSTM model was developed for each selected station to predict discharge which uses surface reflectance ratio of dry and wet pixel and altimetry derived water level as input features. The utilization of a machine learning approach has proven beneficial as it allows for the integration of data from various sources into a single retrieval algorithm. This can be achieved by adding or removing input features from the LSTM configuration and updating the training process accordingly. The determination of the optimal number of hidden layers and neurons in these LSTM networks is typically done through a trial-and-error procedure. The training set and validation set were sampled randomly.

All the datasets should have a common time series for successful merging of different data from various sources. The time series of water level altimetry is limited to the satellite's flyover of the river, but the C/M* ratio obtained from optical satellite sensors can be affected by cloudy images. These circumstances result in missing data, which breaks up the time series. In addition, the accuracy of the altimetry multi-mission satellites might vary based on the altimeter utilized and the number of satellites available at the same time. The temporal resolutions of the time series produced by optical sensors vary due to the repetition cycle of the satellites. When compared to altimetric time series, MODIS-derived time series have a better temporal resolution. The expected temporal resolution of MODIS-derived data is approximately 2 to 3 days.

Therefore, to overcome this issue, all satellite data sets were interpolated at a daily scale to provide time series with the same frequency that are consistent with the temporal resolution of ground observations. Following pre-processing (a-c) of data was carried out before feeding to the LSTM model.

a) Data Scaling:

Scaling helps to standardize the range of independent variables and improves the performance and interpretability of ML algorithms. Additionally, it can speed up the training process and improve the numerical stability of optimization algorithms. The commonly used techniques in data scaling are Standardization and Normalization or Min-Max scaling. Standardization assumes that the data will have a bell-shaped distribution and changes the values to have a mean of zero and standard deviation of one. Min-Max scaling, on the other hand, transforms the values ranging between 0 and 1. In this research, input features were Min-Max scaled.

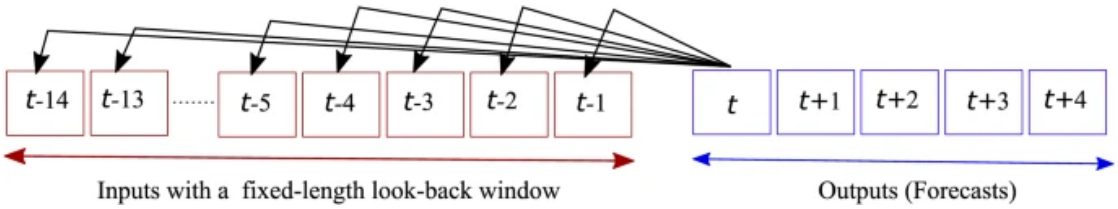
b) Windowing of Input Data:

The next step included preparing input data for predictions/forecasts using a sliding window technique where past observations were grouped to predict/forecast future discharge. Since the current state of a watershed is dependent on the past behavior of the system, the input features are reorganized using window size (Figure 4-5). This creates fixed-length vectors that machine learning models can use to find patterns and relationships in past observations to make future discharge predictions. Various configurations of window size were checked for discharge prediction during

hyperparameter tuning and look back window size of 5 days was adopted in this study for discharge prediction.

Figure 4-5

Schematic Representation of the Sliding Window Approach used in this Study



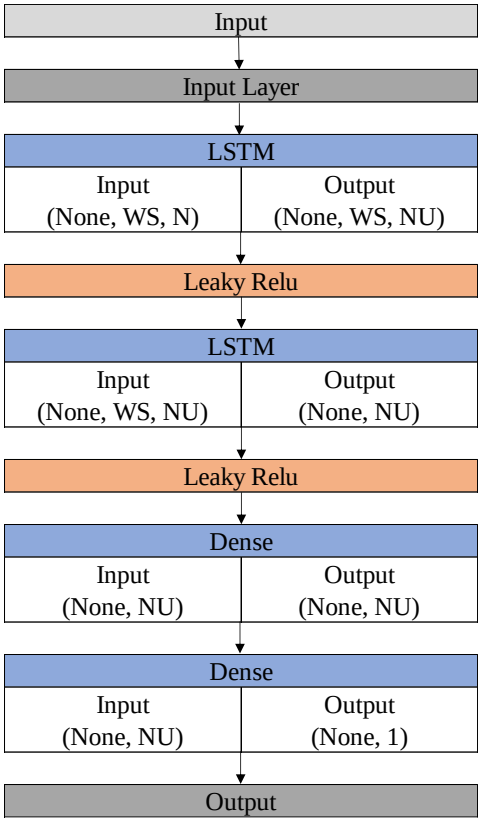
c) Data Splitting

Data splitting in an LSTM model is the process of dividing a dataset into multiple subsets that are used for training, validation, and testing. The training dataset represents the portion of the data that is utilized to train the machine learning algorithm by learning from historical data to estimate the parameters of the model, thus allowing for effective predictions when the machine learning model is subsequently exposed to previously unseen data. The validation dataset, on the other hand, is employed to optimize the model's hyperparameters. The test dataset, in contrast, is utilized solely for the evaluation of the algorithm after the model has been fully trained and optimized.

When splitting data for an LSTM model, it is important to ensure that the split preserves the temporal order of the data if the data is a time series. Thus, in this research training set and test set were prepared from each alternate days which accounted 50% of total data as training set and remaining as test set. Alternate days were selected to preserve all characteristics of input feature during the training process. A separate validation set was used taking 10% of training data to evaluate the model's performance during the training process before testing. The LSTM model was implemented in Keras with TensorFlow backend and computational code was implemented in Python 3.9 and the summary of LSTM model is shown in Figure 4-6.

Figure 4-6

Summary of LSTM Model Architecture used in this Study



(N: Number of Feature, NU: No. of Neurons)

4.2.3 Evaluation of the Performance of LSTM Model

The performances of the LSTM model to simulate discharge at various locations of Lancang Mekong River was assessed through various metrics such as correlation coefficient (R), coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE), and percent bias (PBIAS). Additionally, visual representations such as line diagrams and scatter diagrams were also used for interpretation. This was done by comparing the observed discharge at in-situ stations with simulated discharge.

$$R^2 = \frac{n\sum Q_i^{obs} Q_i^{pre} - \sum Q_i^{obs} \sum Q_i^{pre}}{\left(\sqrt{n(\sum Q_i^{obs})^2 - (\sum Q_i^{obs})^2}\right) * \left(\sqrt{n(\sum Q_i^{pre})^2 - (\sum Q_i^{pre})^2}\right)} \quad \text{Equation 4-3}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Q_i^{obs} - Q_i^{sim})^2}{\sum_{i=1}^{n} (Q_i^{obs} - \overline{Q_i}^{obs})^2} \quad \text{Equation 4-4}$$

$$PBIAS = \frac{\sum_{i=1}^{n} (Q_i^{obs} - Q_i^{pre}) * 100}{\sum_{i=1}^{n} (Q_i^{obs})} \quad \text{Equation 4-5}$$

where, Q_i^{obs} is observed discharge, Q_i^{pre} is predicted discharge and $\overline{Q}_i^{obs}$ is the average observed discharge.

Coefficient of Determination (R^2):

The coefficient of determination is a statistical measure that represents the proportion of variance in the dependent variable that is explained by the independent variables in a regression model. It ranges from 0 to 1, where 0 indicates that the model does not explain any of the variability of the response data and 1 indicates that the model perfectly explains all the variability. A high R^2 value suggests that the model is a good fit for the data, while a low R^2 value suggests that the model may not adequately capture the relationship between the variables.

Nash-Sutcliffe Efficiency (NSE):

Nash-Sutcliffe Efficiency (NSE) is a statistical metric used to evaluate the accuracy of hydrological models by comparing the simulated and observed values of a variable of interest. The NSE ranges from -infinity to 1, where values closer to 1 indicate better model performance. A value of 0 or less indicates that the model performs no better than a naïve mean model, while negative values indicate the model is performing worse than the mean model (Nash & Sutcliffe, 1970).

Percent Bias (PBIAS):

Percent bias (PBIAS) is a statistical measure used to evaluate the accuracy of a model by measuring the average tendency of the predicted values to be larger or smaller than the observed values. It is expressed as a percentage, with a perfect model having a PBIAS value of zero. Positive PBIAS values indicate that the model overestimates the observed values, while negative PBIAS values indicate an underestimation of the observed values.

A model with higher R^2 and NSE values and lower PBIAS values is indicative of a relatively better model for the simulation of discharge.

4.3 Assessment on Spatial Transferability of LSTM Model for Discharge Prediction at Ungauged Locations

There are limited number of hydrological gauging stations in Lancang Mekong River, and it is not always feasible to measure discharge at every location in the river reach. On other hand, it is not always easy to obtain gauge data due to geographical

accessibility, data sharing policy and many more. Spatial transfer of machine learning models is useful for transferring a trained model from one location to another. In the case of river discharge prediction at ungauged reaches, spatial transfer is required because it is not feasible to measure river discharge at every point along a river. Therefore, a model trained in one geographical location can be transferred to another to predict river discharge. The methodological framework is presented above in Section 4.2.

There are different methods for spatial transfer of machine learning models, including domain adaptation, transfer learning, and meta-learning. In this research, a commonly used method, transfer learning was used. Transfer learning involves reusing a pre-trained model in one location for a new location. This approach has shown effectiveness because pre-trained models have already learned useful features that can be applied to the new domain. However, in a large watersheds like Lancang Mekong River basin where different hydrological behavior prevails along different reaches, one model may not sufficiently represent whole basin characteristics further requiring dividing the whole basin into number of reaches and developing model for those reaches showing similar hydrological behavior.

Four models developed for Nong Khai, Nakhon Phanom, Pakse and Kratie were selected for spatial transfer assessment which represents different reaches of the river from upstream to downstream. To test the effectiveness of spatial transfer, each LSTM model was transferred to the other remaining stations, which were considered as pseudo-ungauged reaches. The performance evaluation was carried out to assess the accuracy of the river discharge predictions. Several performance metrics were used, including correlation coefficient (R), and coefficient of determination (R-squared) and Nash-Sutcliffe Efficiency (NSE).

CHAPTER 5
RESULTS AND DISCUSSIONS

5.1 Water Level using Satellite Altimetry

5.1.1 Virtual Stations over Lancang Mekong River

Satellite altimetry offers near-global profile data along tracks, confined solely by the satellite's orbital parameters, yet restricted to its ground track by the nadir-only measurement capability of the sensors. Thus, initially the location where the path of satellites intersected with Lancang Mekong river were focused. These points were designated as Virtual Stations (VS) and water level data was collected for each of these virtual stations using satellite altimeter missions from Jason-1, Jason-2, ENVISAT and SARAL. This data was then processed and analyzed to gain insight into the water levels at 54 different Virtual Stations. The location of virtual stations for different altimetry are presented in Figure 5-1 and details are presented in Table 5-1. Radar altimeters do not determine the range value at each point by direct measurement. Instead, they rely on the echo of a radar signal to extract the range value. The data rate for each of these missions differs, with ENVISAT having 18 Hz, Jason-2 having 20 Hz, and SARAL/AltiKa having 40 Hz, resulting in along-track distances of 380 m, 294 m, and 173 m respectively between altimeter measurements.

Figure 5-1

The Ground Track and the Selected Virtual Stations from Jason-1/2 (left) and ENVISAT, SARAL (right)

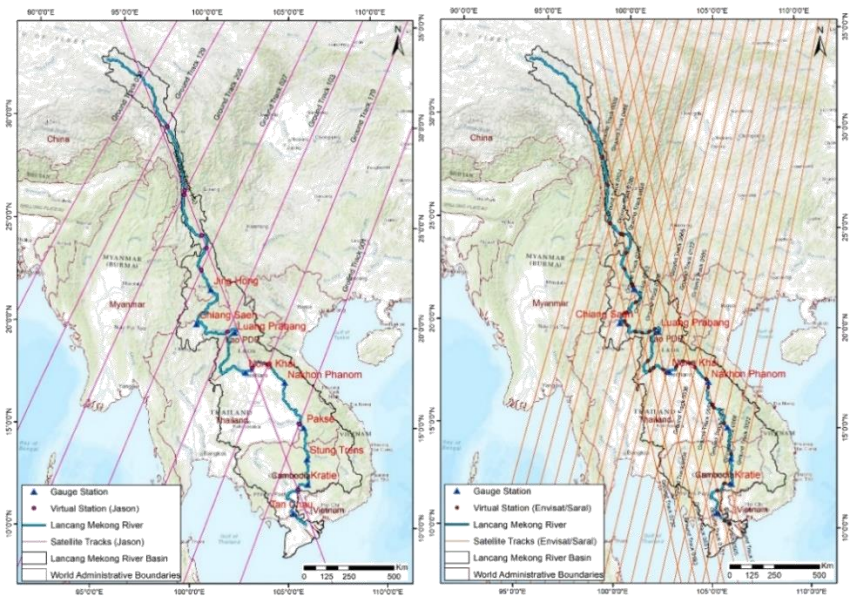


Table 5-1

List of Virtual Station with Chainage and Location Along Lancang Mekong River

SN	Chainage (km)	Satellite Mission	Track Number	Latitude	Longitude
1	0	Envisat/Saral	0737	30.231	97.922
2	18	Jason	129	30.105	97.992
3	255	Envisat/Saral	0008	28.544	98.797
4	464	Envisat/Saral	0466	27.161	99.152
5	495	Jason	140	26.955	99.176
6	523	Jason	205	26.735	99.140
7	742	Envisat/Saral	0924	25.254	99.379
8	836	Envisat/Saral	0380	24.744	99.970
9	847	Envisat/Saral	0193	24.730	100.078
10	873	Jason	140	24.744	100.218
11	980	Jason	140	24.119	100.504
12	1037	Envisat/Saral	0193	23.812	100.308
13	1159	Jason	027	23.009	100.254
14	1159	Envisat/Saral	0838	23.009	100.255
15	1282	Envisat/Saral	0193	22.256	100.693
16	1300	Envisat/Saral	0193	22.116	100.727
17	1303	Envisat/Saral	0294	22.116	100.755
18	1460	Envisat/Saral	0193	21.356	100.912
19	1586	Envisat/Saral	0294	20.825	100.441
20	1601	Envisat/Saral	0737	20.786	100.331
21	1712	Envisat/Saral	0294	20.393	100.337
22	1742	Envisat/Saral	0737	20.192	100.474
23	1840	Envisat/Saral	0752	19.879	100.931
24	1885	Envisat/Saral	0193	19.826	101.281
25	1933	Envisat/Saral	0208	19.883	101.652
26	1947	Jason	103	19.934	101.762
27	1970	Envisat/Saral	0651	20.027	101.951
28	2045	Envisat/Saral	0651	19.819	102.001
29	2233	Envisat/Saral	0193	18.427	101.614
30	2363	Envisat/Saral	0193	17.923	101.733
31	2403	Envisat/Saral	0666	18.078	101.940
32	2476	Envisat/Saral	0651	17.982	102.437
33	2509	Envisat/Saral	0122	17.833	102.603
34	2590	Jason	140	18.126	103.077

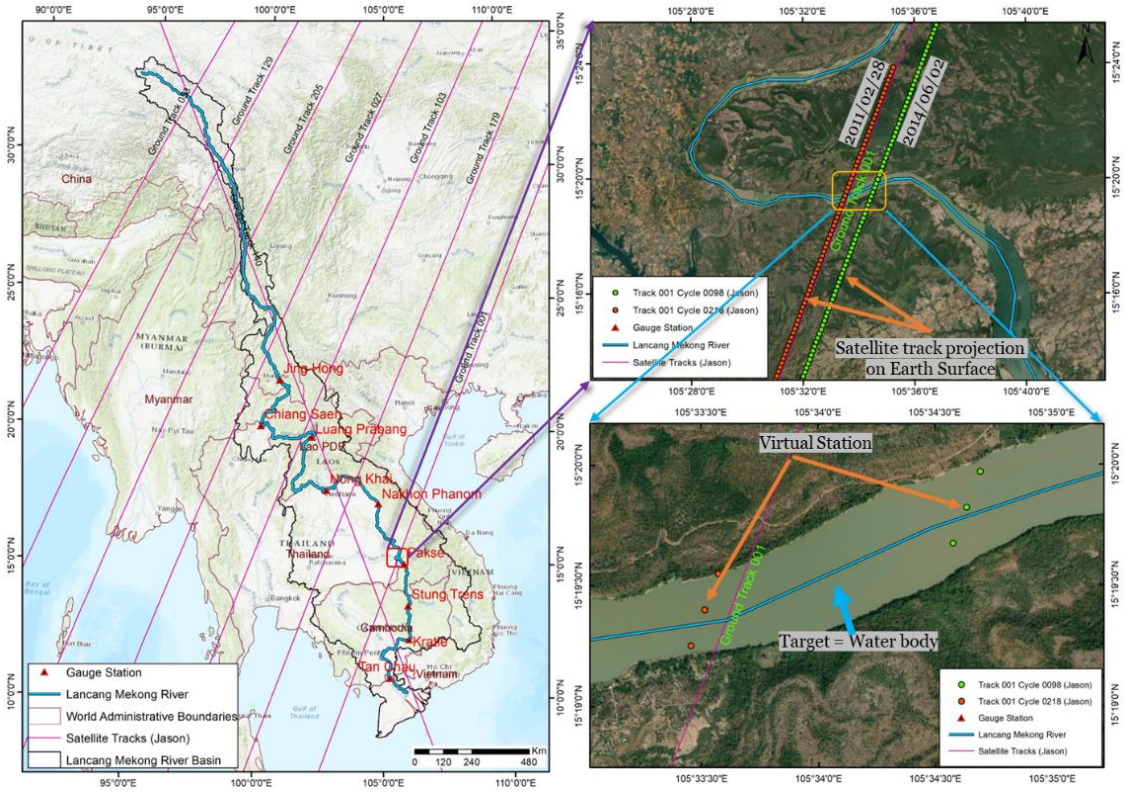
SN	Chainage (km)	Satellite Mission	Track Number	Latitude	Longitude
35	2595	Envisat/Saral	0107	18.151	103.117
36	2662	Envisat/Saral	0580	18.426	103.460
37	2699	Envisat/Saral	0565	18.346	103.788
38	2720	Jason	179	18.334	103.934
39	2753	Envisat/Saral	0036	18.117	104.107
40	2845	Envisat/Saral	0494	17.537	104.688
41	2847	Envisat/Saral	0021	17.531	104.700
42	2894	Envisat/Saral	0021	17.140	104.792
43	3006	Envisat/Saral	0021	16.262	104.996
44	3027	Envisat/Saral	0952	16.105	105.070
45	3190	Jason	001	15.323	105.561
46	3196	Envisat/Saral	0408	15.326	105.610
47	3394	Envisat/Saral	0866	13.782	105.972
48	3592	Envisat/Saral	0021	12.262	105.915
49	3624	Envisat/Saral	0866	12.296	105.634
50	3664	Jason	140	12.013	105.473
51	3698	Envisat/Saral	0565	11.938	105.268
52	3886	Envisat/Saral	0866	10.783	105.292
53	3947	Envisat/Saral	0565	10.409	105.613
54	3986	Envisat/Saral	0322	10.274	105.898

5.1.2 Water Level at Virtual Stations

Data of each cycle of satellite altimetry mission were downloaded for each virtual station and processed to obtain actual water level. In order to obtain water level time series at virtual station locations, each virtual station is created as a box with a size corresponding to the river width (Figure 5-2). However, due to non-perpendicular ground tracks crossing the river at most virtual stations, the size of virtual stations may differ from the river width. Therefore, only water-related measurements within the virtual station boxes are averaged to obtain water level time series. By utilizing the biases mentioned earlier on various missions, we obtain time series water level of different virtual stations along the Lancang Mekong River. Obtained water level at virtual station seems to have outliers which can be attributed to data outage due to topography or errors during data capture and pre-processing. Highly deviated water levels were filtered before merging to obtained water level time series at gauge stations.

Figure 5-2

The Ground Track of Two Cycles of Jason-2 Representing Water Related Measurement Around Virtual Station



The arrangement of all sequences of water level, from top to bottom, along the river is illustrated in Figure 5-3. The temporal resolution of obtained water level is higher for Jason-2 and lower for ENVISAT and SARAL. With this, water level is obtained for 54 virtual stations along the Lancang Mekong mainstream. The water levels obtained at upper reaches seems to be erratic and more bias can be observed compared to the downstream reaches. This presence of more bias in upper reaches may effect on the model accuracy for daily discharge prediction. In

Figure **5-4**, the average water level obtained from the defined virtual stations is presented. Additionally, the circle on the plot represents the average water level along the Lancang Mekong River. This plot provides an overall view of the water level variations in the virtual stations and guage stations in the river, allowing us to visualize the changes in water level over time at different points along the river. It can be seen that water level varies from 350m above mean sea level at Chiang Saen to 10m above mean sea level at Kratie within the reach length of 2000 km.

Figure 5-3

Altimetry Derived Water Level Time Series at Virtual Stations

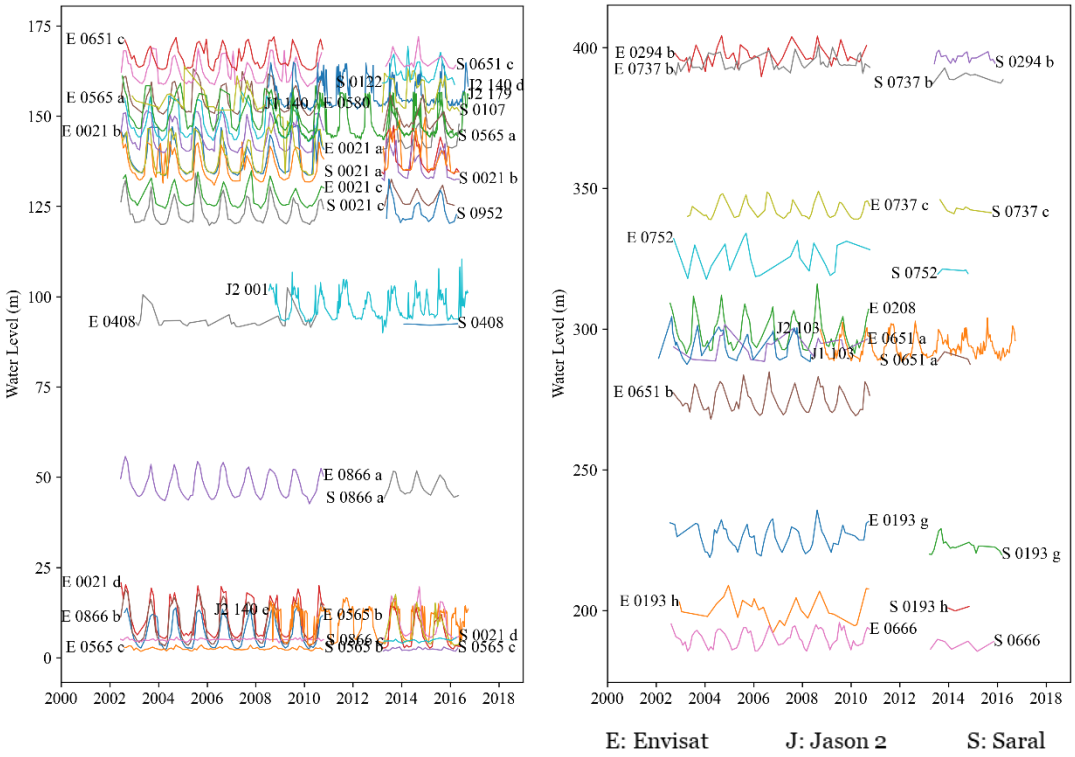
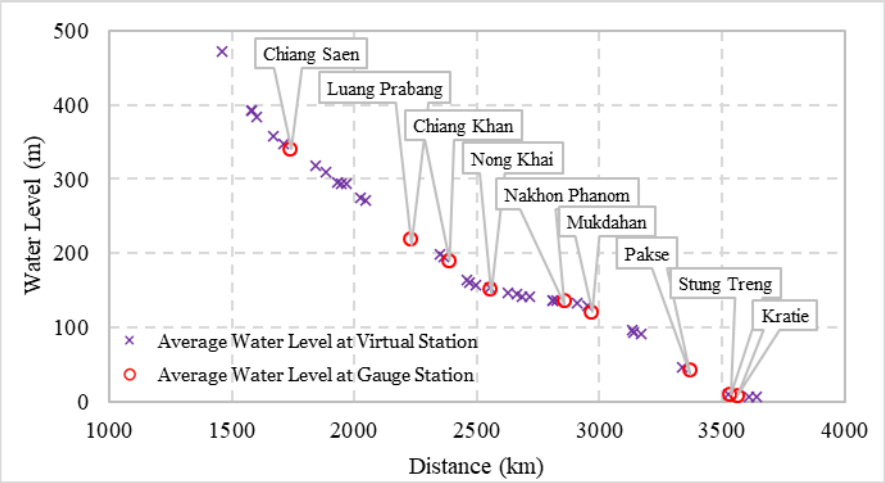


Figure 5-4

Average Altimetric Water Level at the Selected Virtual and Gauge Stations



To ensure the accuracy of the water level data obtained from the altimetric missions, a validation process was conducted using water level data obtained from DAHITI database at virtual stations along the river. DAHITI is a reliable database that provides global water level time series data derived from satellite radar altimetry and hydrological models.

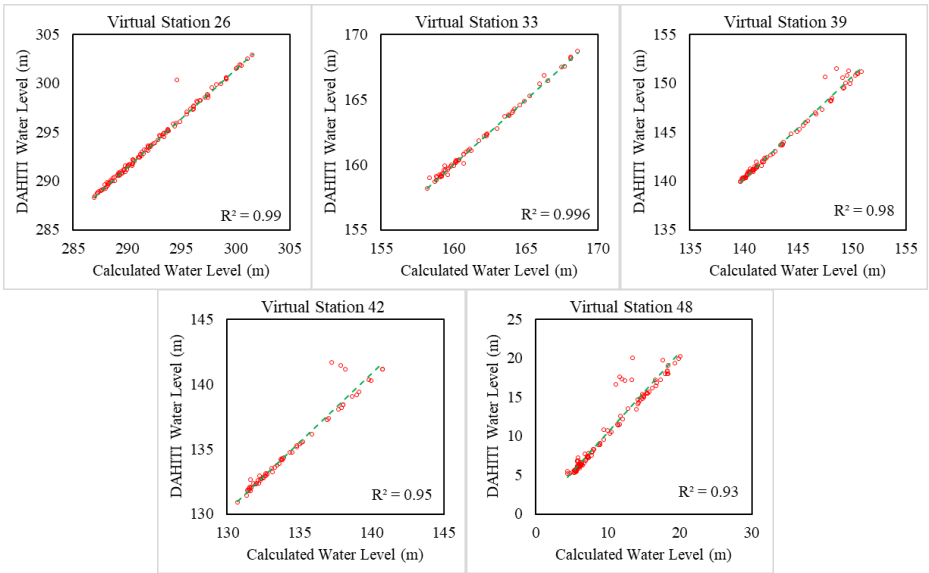
By comparing the water level data obtained from the altimetric missions with the water level data from DAHITI at the same virtual stations, we were able to assess the consistency and accuracy of the altimetric data. The comparison allowed us to identify any systematic errors or biases in the measurements and adjust to improve the accuracy of the data.

The scatterplot in Figure 5-5 displays the water level values of five virtual stations plotted against the corresponding DAHITI data. This type of graphical representation allows for a visual assessment of the goodness of fit between the two datasets and helps to identify any potential outliers or inconsistencies in the data. By analyzing the scatterplot, we can gain a better understanding of the relationship between the water level data obtained from the virtual stations and the DAHITI data.

The validation result shows that coefficient of determination ranges from 0.93 to 0.99 at five different virtual stations between altimetry derived water level and DAHITI water level. This high coefficient of determination indicates a strong relationship between the derived altimetric data and the DAHITI data, providing further evidence of the accuracy and reliability of the measurements obtained from the altimetric missions. The validated water level data can now be used with confidence for further calculations and analysis, including estimating the discharge of the river.

Figure 5-5

Comparison of Altimetry Derived Water Level with DAHITI Database for Selected Virtual Stations



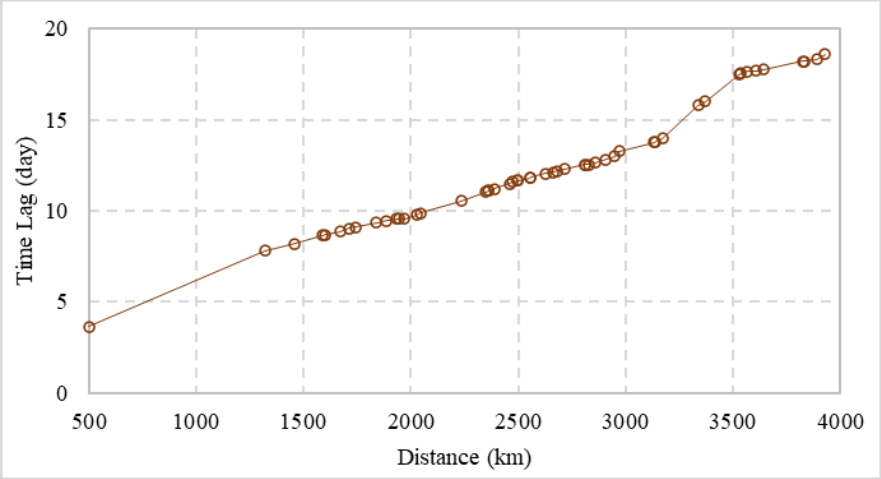
5.1.3 Water Level at Gauging Station Integrated from Virtual Stations

Since the objective of this study is to develop an LSTM model for discharge prediction at gauge stations utilizing the available observed river discharge, altimetric water level at those gauge stations needs to be obtained. Thus, the water levels obtained at virtual stations need to be transferred to the location of in-situ gauging stations. This was achieved following the densification procedure described by Tourian et al., 2016 where data from multiple satellite altimeters at virtual stations are connected and more detailed water level time series are obtained. This method helps to increase temporal resolution of altimetry derived water level by connecting each virtual station statistically and hydraulically.

A simple empirical hydraulic equation is employed in this approach, incorporating additional measurements of average river width and slope (based on Bjerklie et al., 2005). The provided equation computes the time lag and the mean flow velocity induced by the streamflow between the selected spot on the river and the altimetric virtual stations. Figure 5-6 shows the estimated time lag between each virtual station located along the Lancang Mekong River with respect to the initial virtual station.

Figure 5-6

The Estimated Time Lag Between Each Virtual Station Located along the Lancang Mekong River



After estimating the time lag, each water level measurement at the virtual stations is adjusted and shifted to the designated location. The resulting time series is then standardized, and outliers are removed. Finally, the time series is scaled to match the statistical distribution of water levels at the selected location.

Figure 5-7 shows the computed water levels at nine gauging stations namely Chiang Saen, Luang Prabang, Chiang Khan, Nong Khai, Nakhon Phanom, Mukdahan, Pakse, Stung Treng and Kratie presented in a single plot for easy comparison. Each gauging station is represented by a different line color. The y-axis represents water level values, while the x-axis denotes the time frame of the data. The graph highlights the variability in water levels across the different stations and provides insights into hydrological trends and patterns. The variation in water level is high on the downstream station with relatively river slope compared to upstream reaches.

Figure 5-7

Altimetry Derived Water Levels Across Gauging Stations

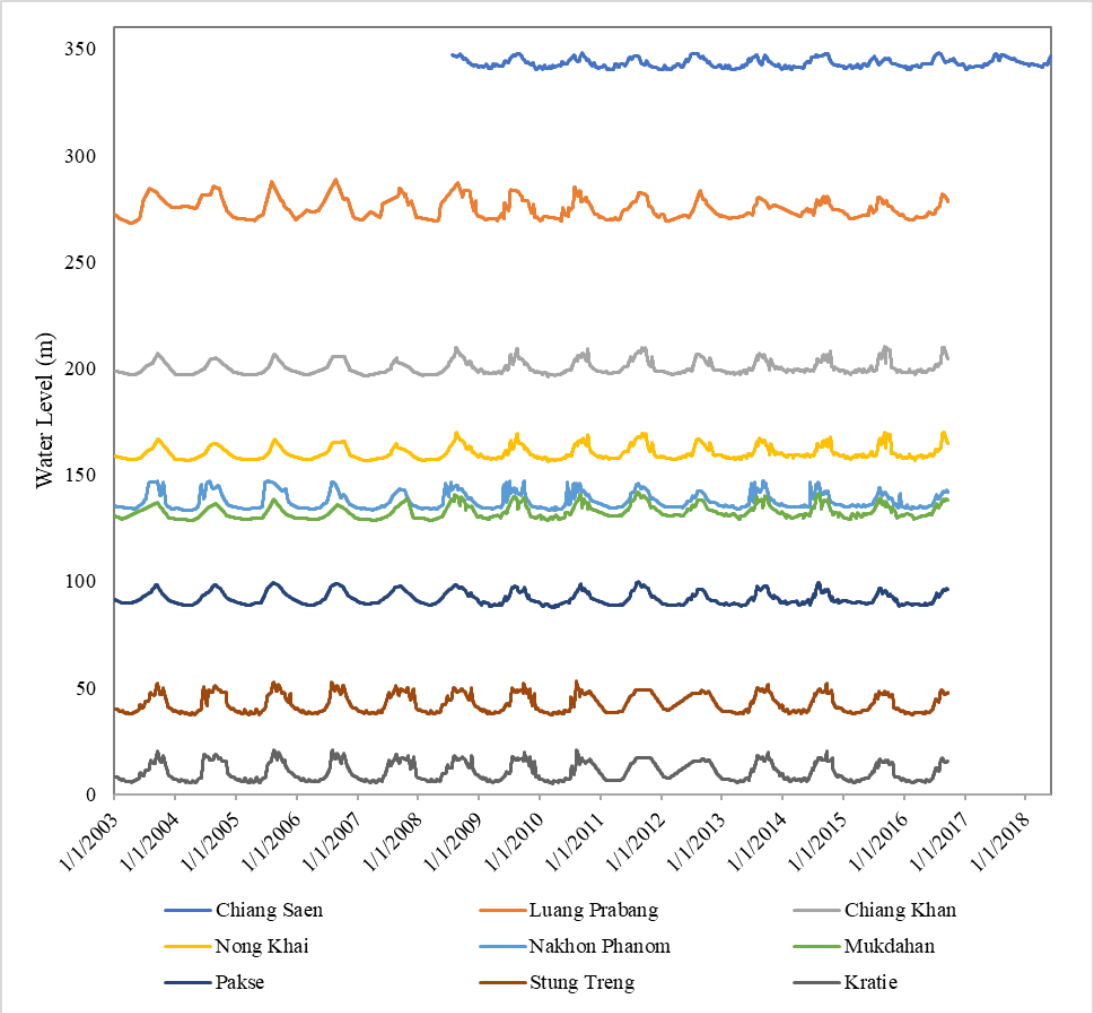
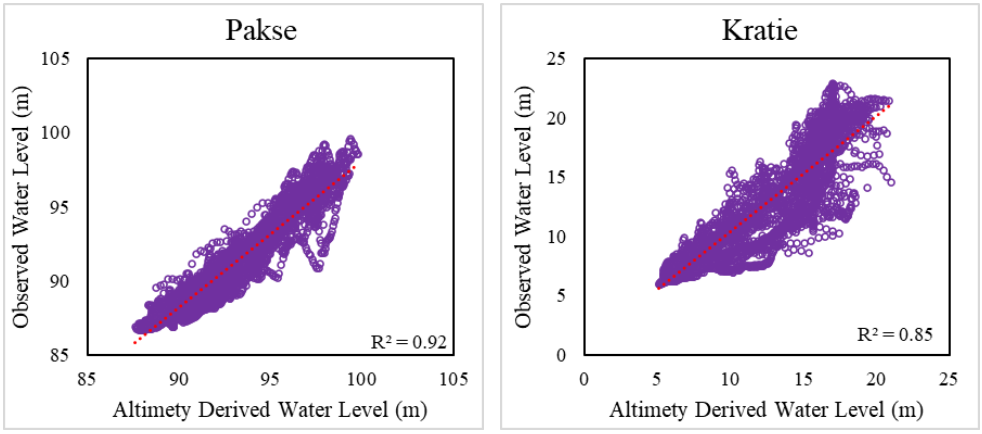


Figure 5-8 shows the comparison of the altimetry-derived water level with the observed water level at two different gauging stations, Pakse and Kratie. The comparison enables the assessment of the accuracy of the altimetry-derived water level measurements by contrasting them with the observed water level data. The coefficient of determination

is found to be 0.92 for Pakse and 0.85 for Kratie. This indicates a strong positive relationship between the two variables and supports the accuracy of the altimetry-derived water level measurements. Further, it justifies the reliability of use of densification procedure developed on another geographical region applied on Lancang Mekong River basin.

Figure 5-8

The Altimetry Derived Water Level Compared with Observed Water Level at Pakse and Kratie



5.2 Discharge Modelling Using Water Level and Optical Datasets

This study uses ratio of surface reflectance from dry and wet pixel (C/M) obtained from MODIS AQUA and TERRA satellite products and altimetry derived water level as inputs feature for river discharge prediction. Discharge prediction was carried out using each of these input features alone, as well as in combination with each other. The performance evaluation of discharge prediction with C/M ratio from AQUA and TERRA satellite and combined with altimetry water level is carried out by comparing with observed discharge as well.

5.2.1 Surface Reflectance Ratio Time Series from Optical Datasets

The ratio of reflectance from more stable land pixel (C) and variable water pixel (M) in Near Infrared (NIR) band of optical images serves as an indicator of discharge variation. This principle is adapted to train LSTM model and predict discharge based on remote sensing data. For each optical dataset AQUA and TERRA, all images were processed and analyzed using Google Earth Engine (GEE) platform and Python 3.9 for the period

2003-2021. The location of C and M pixel plays a pivotal role in discharge estimation, and this needs to be carefully determined.

Table 5-2 shows the number of images selected for analysis for each station. The total number of images after cloud filtering is reduced to almost 50%. The mean temporal resolution was decreased to ~2 days from ~1 day. So, linear interpolation was applied to obtain daily time series data.

Table 5-2

The Number of MODIS Images Selected for Analysis for Each Station.

Station	Period	Total Number of Images	Selected Images			
			Aqua		Terra	
			M Pixel	C Pixel	M Pixel	C Pixel
Chiang Saen	2003 - 2021	6939	4272	4318	3875	3804
Luang Prabang	2003 - 2018	5844	3366	3482	2929	2927
Nong Khai	2003 - 2021	6939	3952	3698	3717	3550
Nakhon Phanom	2003 - 2021	6939	3304	3526	3650	3706
Mukdahan	2003 - 2021	6939	3092	3141	3595	3396
Pakse	2003 - 2018	5844	2950	3158	3071	3261
Stung Treng	2003 - 2021	6939	3594	3743	3536	3683
Kratie	2003 - 2021	6939	3630	3878	3713	3998
Tan Chau	2003 - 2021	6939	3424	3387	3607	3498

To best locate C and M pixel for each gauging station, C/M ratio was computed for each pixel within the box obtaining [38 x 38] x [(38 x 38) - 1] time series C/M ratios. All the pixels in the box were used to calculate the time series for C/M by assuming that each pixel except one was C, and the remaining one was M. Similarly, the ratio was calculated for all pixels of the box i.e., 1443, assuming M and varying the C location, following the procedure as above.

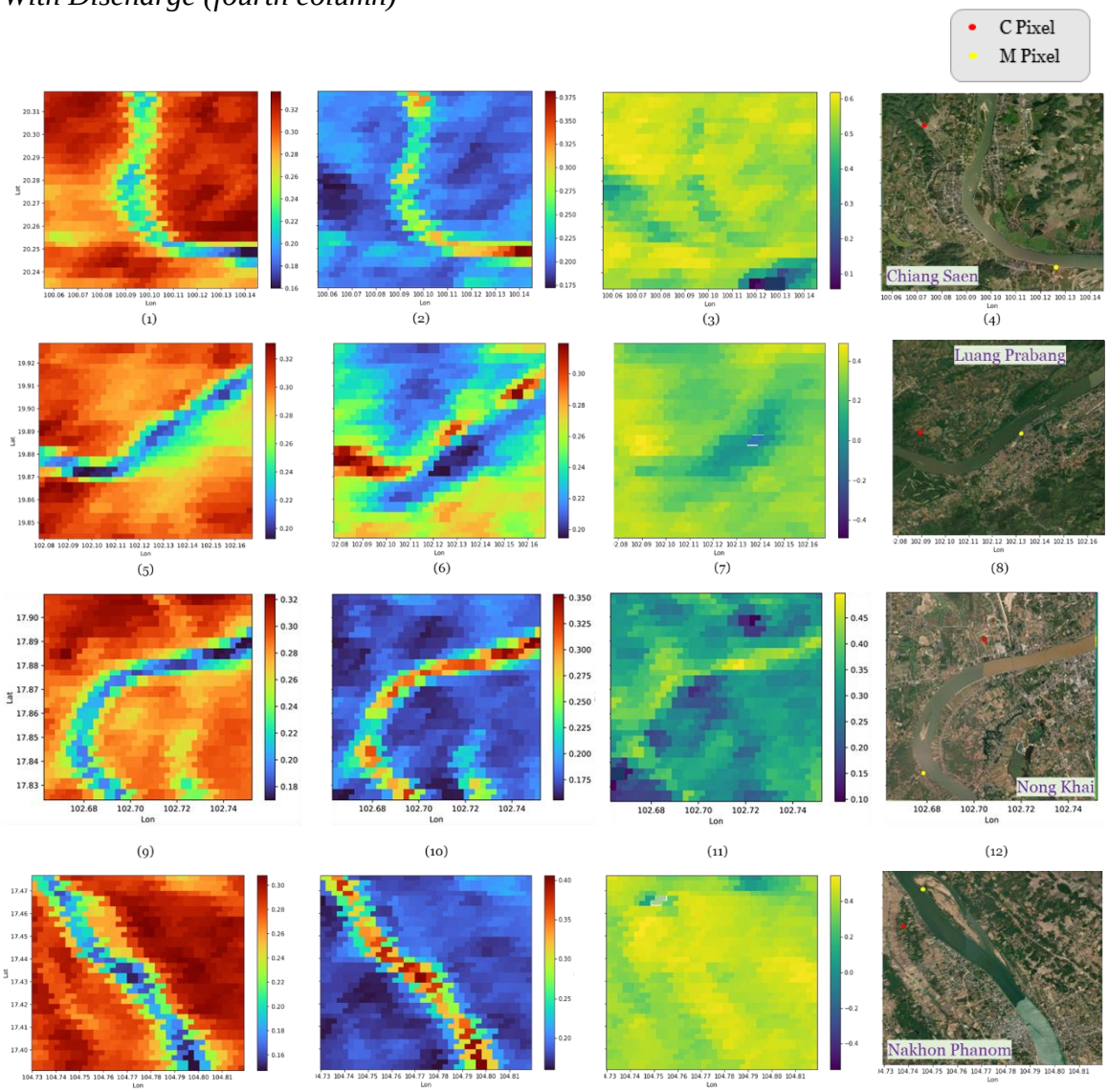
Obtained C/M ratios were correlated with in-situ discharge and maximum coefficient of correlation for each pixel was obtained. Finally, the best correlated combination of C and M pixel were selected for discharge estimation. Figure 5-9 shows the best location of C and M pixel for each gauging stations considered for this study. The

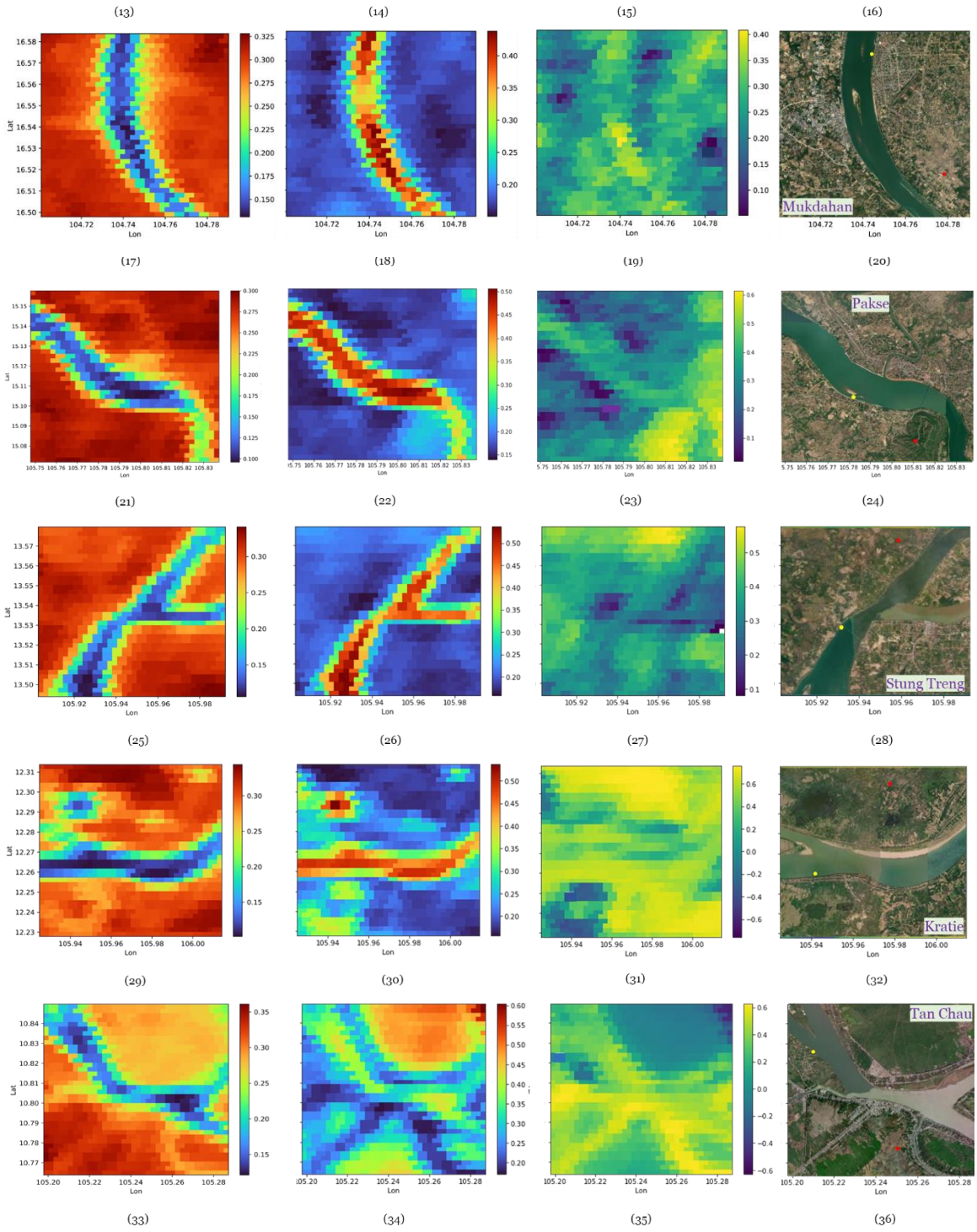
coefficient of variation for C pixel is below 0.3 and relatively stable while that for M pixel shows higher variability.

Best location for C Pixels were obtained near to the urban areas or densely vegetated areas with lower coefficient of variations of surface reflectance whereas best location for M Pixel were obtained near the river flood plain portions where higher sensitivity to water variability is expected.

Figure 5-9

Median Value of Reflectance of Band 2 (first column), Coefficient of Variation of Reflectance (second column), Coefficient of Correlation Between C/M and Discharge (third column) and the Location of the C and M Pixel Having Maximum Correlation With Discharge (fourth column)



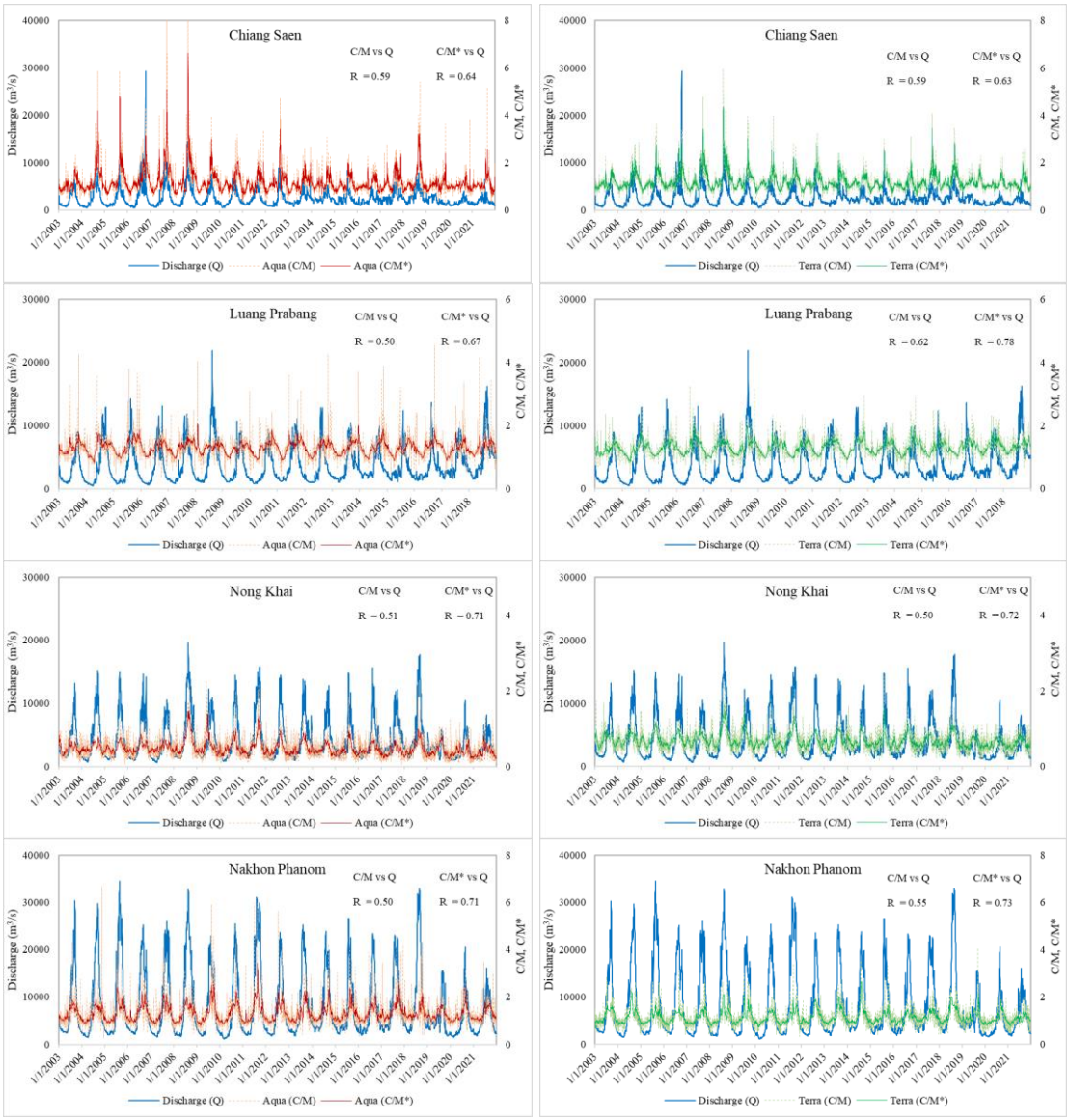


Although the C/M ratio helps to reduce the noise in the surface reflectance measurements, C/M fluctuates quickly over time and seems to be unstable. An exponentially smoothing filter was added to the C/M ratio to reduce the effects of short-term and observation noises and C/M* was computed for each gauging station. The correlation between satellite data and observed discharge for each gauging station is improved after filtering of C/M ratio. Figure 5-10 shows the comparison of correlation between observed data and C/M* with C/M at nine reaches of the Lancang Mekong

River from various geographical location and hydrological condition. It can be observed that the correlation coefficient is improved from the range 0.44 – 0.63 to 0.64 – 0.76 for AQUA products. However, for TERRA products, the coefficient of correlation improved from 0.55 – 0.76 to 0.63 – 0.83. It is notable to observe that TERRA products are highly correlated with observed discharge in the Lancang Mekong Basin.

Figure 5-10

Coefficient of Correlation Between Surface Reflectance Ratio and Observed Discharge at Gauging Stations Within the Lancang Mekong River for AQUA and TERRA Satellite Data



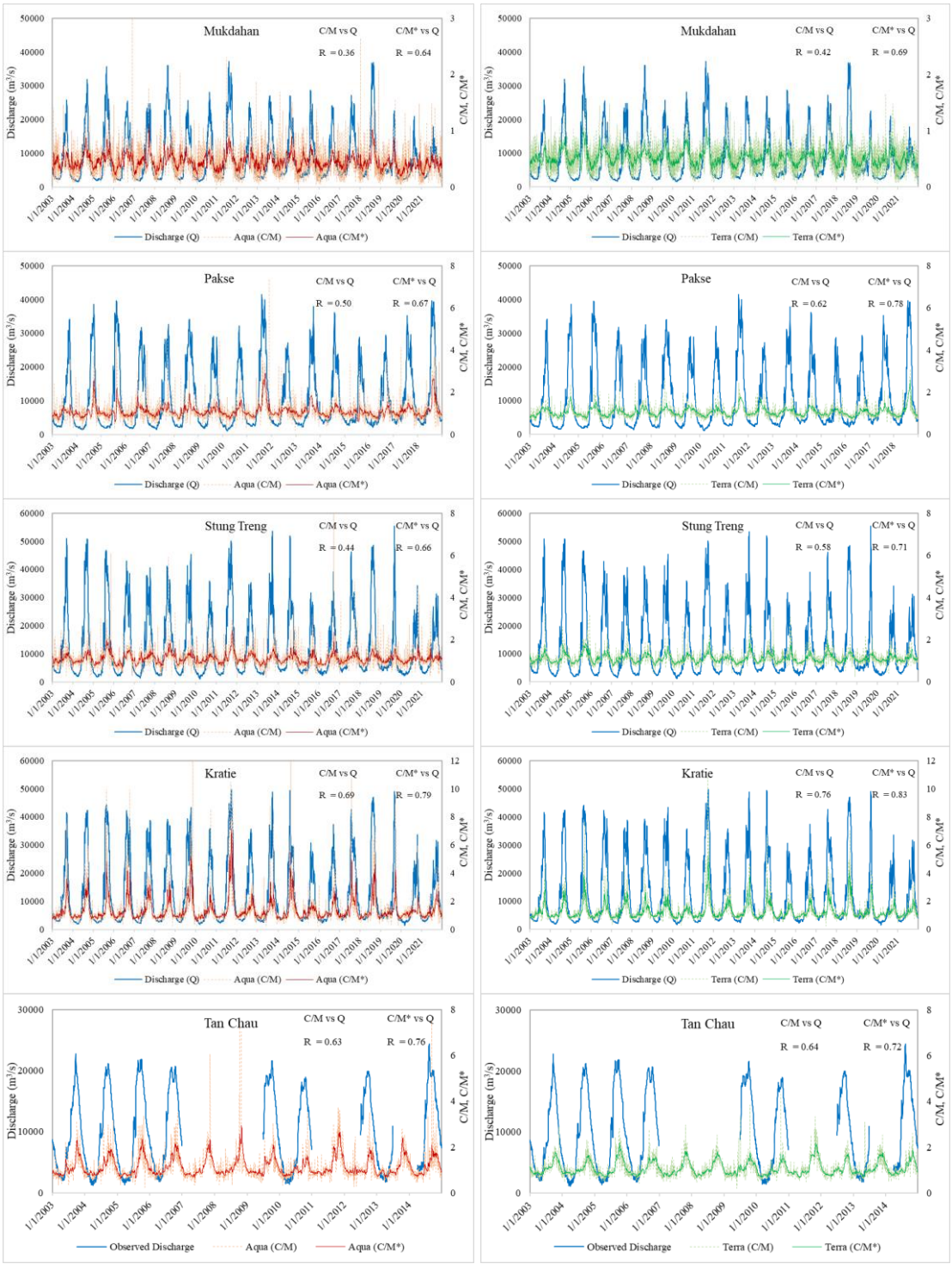


Table 5-3 summarizes the coefficient of correlation between observed discharge and surface reflectance ratio for each gauging stations. The highest correlation between discharge and C/M* is obtained for Kratie station for AQUA and TERRA product and lowest is obtained for Luang Prabang for AQUA product which is 0.43. Also, significant improvements are obtained for AQUA products after applying smoothing filter in most of the stations.

Table 5-3

Coefficient of Correlation Between C/M and C/M with Observed Discharge.*

Station	Aqua		Terra	
	C/M	C/M*	C/M	C/M*
Chiang Saen	0.59	0.64	0.59	0.63
Luang Prabang	0.29	0.43	0.49	0.66
Nong Khai	0.51	0.71	0.50	0.72
Nakhon Phanom	0.50	0.71	0.55	0.73
Mukdahan	0.36	0.64	0.42	0.69
Pakse	0.50	0.67	0.62	0.78
Stung Treng	0.44	0.66	0.58	0.71
Kratie	0.69	0.79	0.76	0.83
Tan Chau	0.63	0.76	0.64	0.72

Thus, C/M* time series is obtained for each gauging station on daily resolution for each AQUA and TERRA satellite products. This C/M* time series is used as input features from remote sensing for developing LSTM model for each of the gauging stations for discharge prediction.

5.2.2 Discharge Prediction with Surface Reflectance Ratio and LSTM Model

5.2.2.1 Data Preprocessing

In this study, three LSTM models are developed for gauging stations using AQUA, TERRA and combined AQUA-TERRA products as input features for LSTM model. Before using data with an LSTM model, it is important to preprocess it to make it suitable for the model to work with. This included cleaning and normalizing the data, breaking it down into smaller units, making sure all data has the same length, separating it into training and test sets, and creating vector representation of the input features. The techniques of data preprocessing are discussed in detail in the methodology chapter of this report. The data are split into training and testing dataset taking alternate days after min-max normalization.

5.2.2.2 LSTM Model Development and Hyperparameter Optimization

The LSTM model was implemented in Keras with TensorFlow backend and computational code was implemented in Python 3.9. The LSTM architecture consists

of three main components: the input gate, the forget gate, and the output gate. Each component is made up of a sigmoid layer and a dot product operation. The input gate manages the volume of information that can enter the cell state, the forget gate determines what information should be retained in the cell state, and the output gate decides what information is to be extracted from the cell state. The cell state is a memory unit that can retain information over a longer period. The output of the LSTM cell is passed through a fully connected layer with a linear activation function to produce the final output.

After preprocessing the input feature, LSTM network architecture was defined. Defining the architecture of an LSTM model involves several important decisions. The first included determining the number of layers to use in the model. LSTM models can have one or more layers, with each layer containing multiple LSTM cells. Each layer can extract different levels of abstraction from the input data. More layers can capture more complex patterns, but it also increases the risk of overfitting. In this study, two layers were implemented.

The next step was determining the number of neurons in each layer. The number of neurons in each layer will affect the model's capacity to learn and represent the underlying patterns in the data. Determining the number of neurons in each layer of an LSTM model is crucial for optimal performance. The complexity of the problem and the dimensionality of the input are key factors that should be considered when deciding on the number of neurons. Additionally, a larger number of neurons in a layer can increase the capacity of the model to learn more complex representations, however, it can also lead to increased computation and the risk of overfitting. Therefore, the number of neurons was carefully chosen to strike a balance between model performance and computational efficiency. The LSTM layer was provided with 32 neurons and hidden layer was provided with 16 neuron units after repetitive trail.

Another important decision is choosing the type of activation function to use in the LSTM cells. The activation function controls the output of each neuron, and it is used to introduce non-linearity in the model. The activation functions used in this study is Leaky ReLu. This activation function is selected because Leaky ReLU mitigates the "dying ReLU" problem by allowing a small negative slope for negative input values, thus preventing the gradients from becoming negligible. This facilitates the training

process by enabling the network to learn more complex representations, whilst remaining computationally efficient. Additionally, it is a simple and straightforward modification to implement.

Dropout regularization can be added to the LSTM model to prevent overfitting by randomly dropping out some neurons during the training process. Overfitting happens when a model becomes excessively complex and begins to memorize the training data instead of being able to generalize to new data. To address this issue, dropout is employed by randomly setting a certain percentage of the neurons to zero during training. This compels the remaining neurons to adapt and learn more resilient features, making the model less reliant on the particular training data. Additionally, dropout also increases the diversity of the models, which can lead to better generalization performance. By adding dropouts to LSTM models, it helps to prevent overfitting and improve the overall performance on unseen data.

In LSTM, hyperparameter tuning is the process of finding the optimal combination of parameters that results in the best performance on the validation set. These parameters are not learned from the data during training and can significantly impact the performance of the model. Hyperparameter tuning involves training the model multiple times with different combinations of parameters and evaluating the model's performance on a validation set to find the best combination. Grid search, random search, and Bayesian optimization are some methods for hyperparameter tuning in LSTM.

Hyperparameter tuning was performed in this research to optimize LSTM model performance. The hyperparameters used for tuning include the number of neurons per layer, the learning rate, the batch size, the dropout rate, number of epochs, window size and the number of layers. To perform the tuning, grid search method and manual trial and error method were explored for the different combinations of the hyperparameters. Grid search is a simple and exhaustive method of searching through a predefined range of hyperparameters by specifying a set of values for each hyperparameter and training the model for each combination of those values. This process allowed us to test a wide range of combinations and determine the optimal set of hyperparameters that resulted in the best performance on the validation set.

After the optimal combination of hyperparameters was determined, the final LSTM model was trained using these hyperparameters on the entire training set and evaluated on the test set to report the final performance of the model. It's worth noting that the process of hyperparameter tuning can be computationally time-consuming, but it is crucial for LSTM model to achieve the best performance. Table 5-4 summarizes the optimal parameters of LSTM model used in this research.

Figure 5-11 shows the performance of LSTM model in terms of loss and mean squared error with increasing number of epochs. It is observed the model performance has improved with increasing number of epochs and stabilizes with epochs about 800. Thus, the number of epochs considered for training of LSTM model in this study is taken as 800.

Figure 5-11

Model Loss and Mean Squared Error During Training and Validation

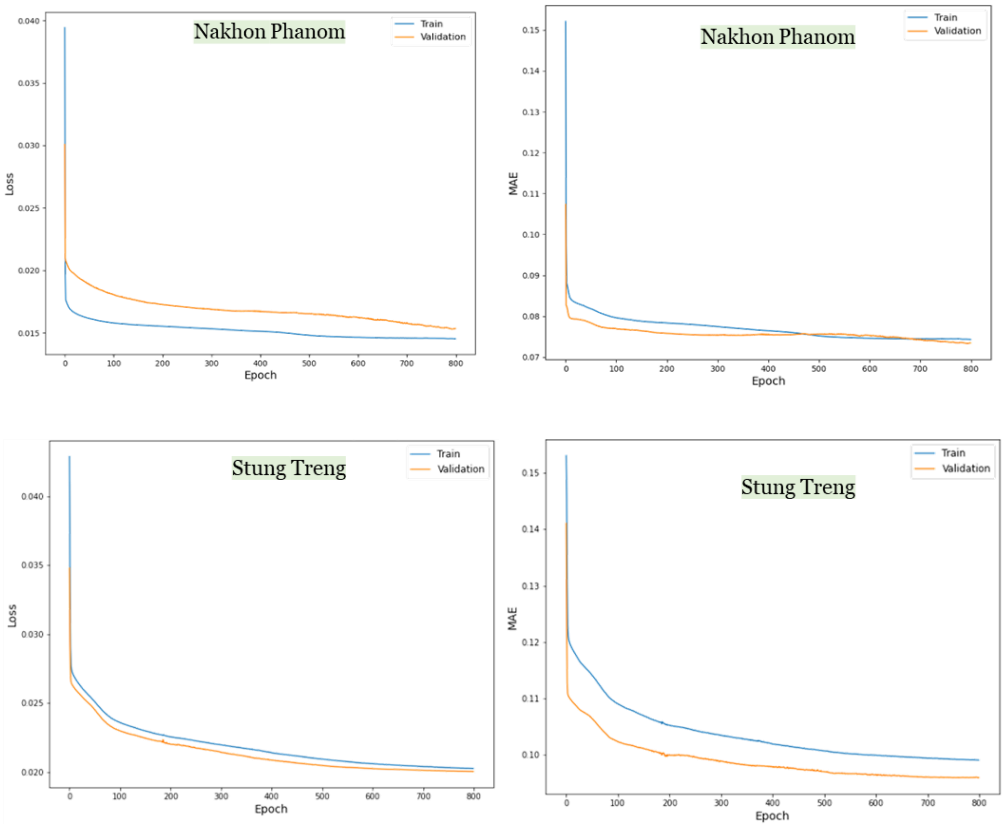


Table 5-4

Optimized Hyperparameters for LSTM Model Development.

Station	No. of Hidden Layer	No. of Units	Window Size	Batch Size	Optimizer / Activation Function
Chiang Saen	1	16	5	5	Adam / Leaky ReLu
Luang Prabang	1	16	4	5	
Nong Khai	1	16	5	5	
Nakhon Phanom	1	16	1	1	
Mukdahan	1	16	5	5	
Pakse	1	16	1	1	
Stung Treng	1	16	5	5	
Kratie	1	16	1	1	
Tan Chau	1	16	1	1	

5.2.2.3 Discharge Prediction using Surface Reflectance Ratio

After optimization of hyperparameters, the LSTM model was able to identify patterns and relationships in the data that were not immediately apparent, which has the potential to lead to new insights. For analyzing the performance of discharge simulated by single product and combined product, model was trained using Aqua and Terra products separately at first and then was trained using combined input from Aqua and Terra products. Then, the model was tested for the next set of data which was not used during the training period. To include all pattern and seasonality of the data, model was trained using data for alternate days and tested for next alternate days. Overall, the results from the model are promising and indicate robustness of using optical data for discharge estimation. Figure 5-12 shows the comparison between the observed and simulated discharge using LSTM for input features from MODIS AQUA, TERRA and combination of both.

The Table 5-5 provides the performance of different products (Aqua, Terra and combination of Aqua and Terra) in different stations (Chiang Saen, Luang Prabang, Nong Khai, Nakhon Phanom, Mukdahan, Pakse, Stung Treng, Kratie and Tan Chau) for predicting discharge compared with the observed discharge. The three input features used are Aqua, Terra, and combined Aqua and Terra products. The results for each station have been tabulated separately for both the training and testing periods. The evaluation has been done in terms of Coefficient of Correlation (R), Coefficient of Determination (R^2), Nash–Sutcliffe Efficiency (NSE), and Percent Bias.

In terms of the correlation coefficient, the results show moderate to strong positive correlations between simulated and observed discharge for most of the stations. The results indicate that combined input feature of Aqua and Terra product showed the highest correlation coefficient values between simulated and observed discharge, ranging from 0.70 to 0.88, for all the nine stations. For the Aqua product, the correlation coefficients range from 0.49 to 0.78, and for the Terra product, the correlation coefficients range from 0.63 to 0.86. The highest correlation coefficient of 0.88 was observed for the Kratie station with the Aqua+Terra product, while the lowest correlation coefficient of 0.49 was observed for the Luang Prabang station with the Aqua product.

The coefficient of determination values is positive for all stations and products, indicating that the simulated discharge data explains some portion of the variance in the observed discharge data. The coefficient of determination values ranges from 0.24 to 0.77 for Aqua, 0.39 to 0.73 for Terra, and 0.46 to 0.77 for combined Aqua and Terra. The combined Aqua and Terra product generally showed higher R-squared values than the individual Aqua and Terra products, indicating that the combination of both products leads to a better fit between the simulated and observed discharge data. However, the R-squared values are generally lower than the corresponding correlation coefficients, indicating that the simulated discharge data does not explain all the variance in the observed discharge data.

The Nash–Sutcliffe Efficiency (NSE) values for combined Aqua and Terra are also the highest among the three input features for most of the stations, indicating better performance in simulating the observed discharge. The value of NSE across all stations and products ranges from 0.23 to 0.76. The highest NSE value of 0.76 is achieved by combined Aqua and Terra products at Kratie station, while the lowest NSE value of 0.23 is achieved by Aqua at Luang Prabang station during both calibration and validation period. In general, the combined Aqua and Terra product outperforms both Aqua and Terra products separately in terms of NSE values.

The Percent Bias (PBIAS) values for all the three products are negative, indicating underestimation of the simulated discharge in Chiang Saen, Nakhon Phanom, Stung Treng and Kratie stations. However, for remaining stations, PBIAS values are positive indicating overestimation of simulated discharge compared to observed discharge. The

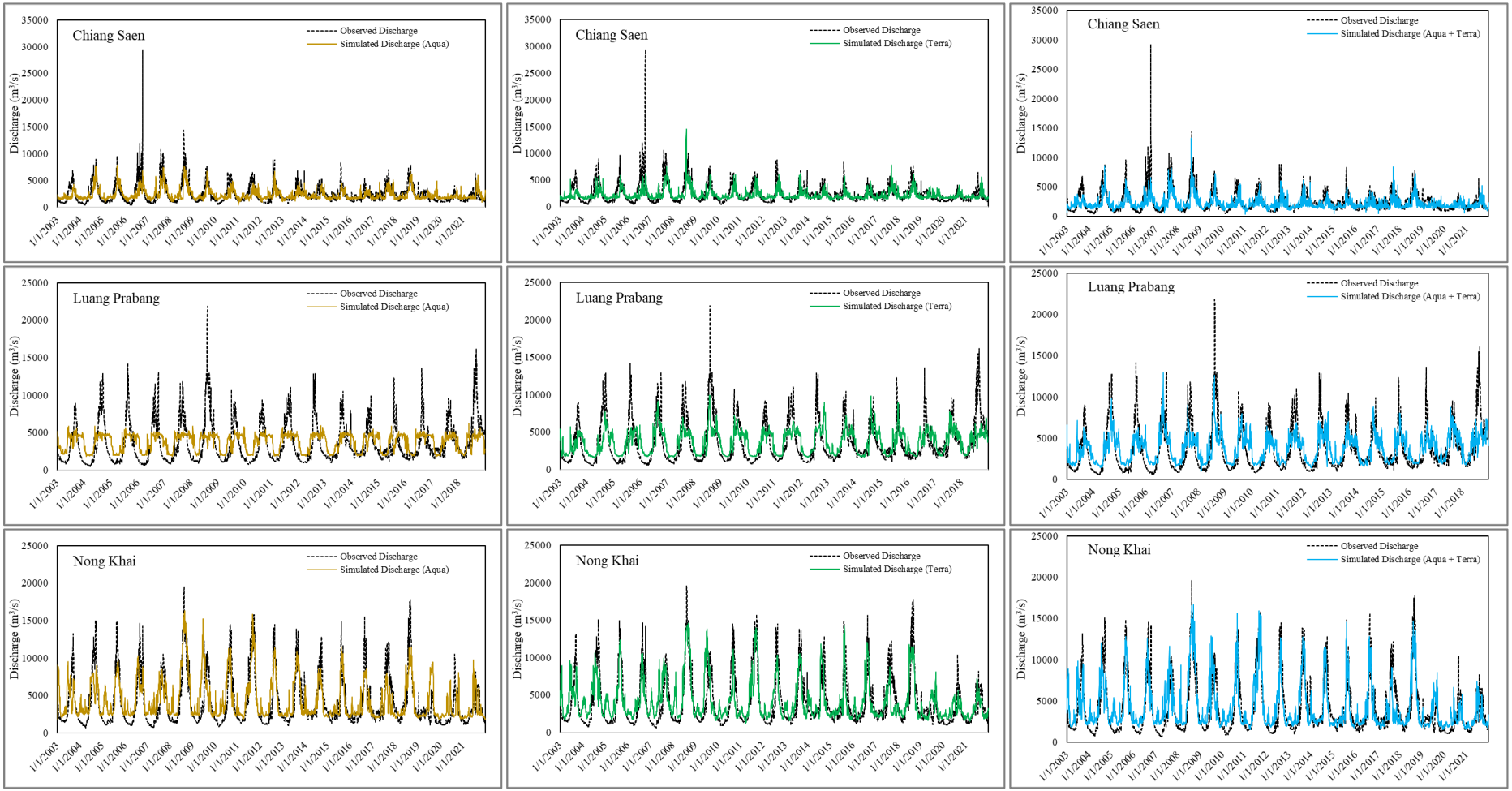
magnitude of the PBIAS values varies greatly across stations and products. The PBIAS varies across all stations and products range from -1.21 % to 25.12%. The highest PBIAS value of 25.12% is found for Aqua product at Mukdahan station, while the lowest PBIAS value of -1.21 % is found for Aqua product at Stung Treng station.

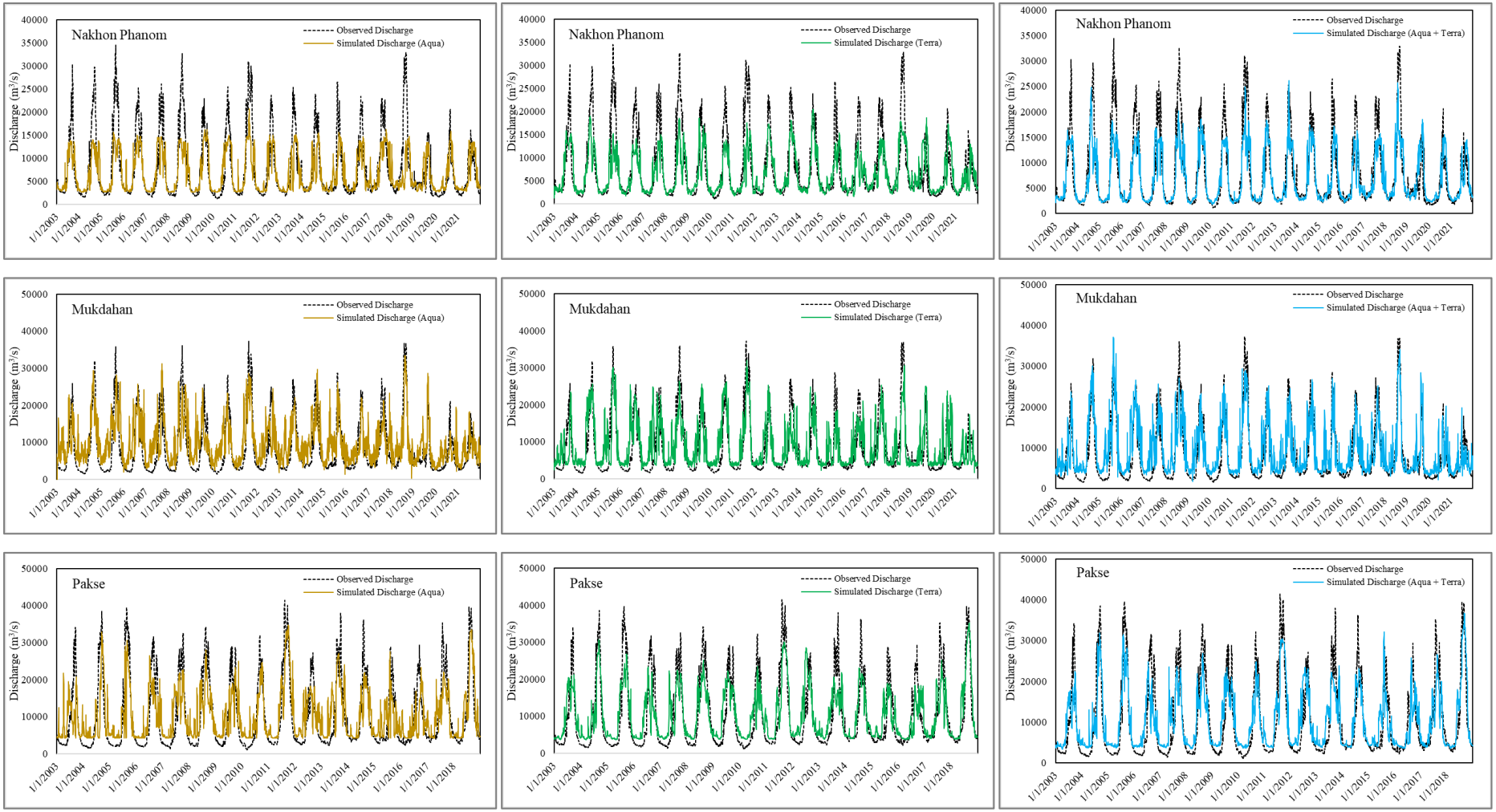
The model is able to predict the low flow closely matching with observation in Chiang Saen, Nakhon Phanom, Stung Treng and Kratie stations and model have overestimated low flow in remaining stations. On other hand, models have underestimated high flows in almost all the stations but with better prediction on Nong Khai, Mukdahan, Stung Treng, Kratie and Tan Chau compared to others. The model could not reproduce the seasonal pattern of discharge in Luang Prabang while considering Aqua product only. The combined use of Aqua and Terra input features has shown better performance on discharge prediction in the same station. The NSE varies from 0.42 to 0.76 for combined input feature indicating geographical location and different hydrological characteristics have dominant impact upon the performance of discharge prediction.

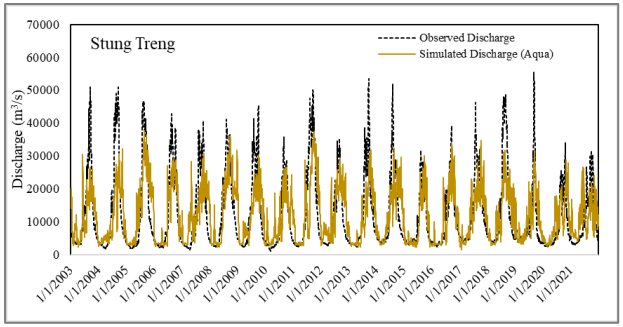
Hence, the results highlight the importance of using multiple remote sensing data as input features for simulating discharge and the benefits of combining them. The combined Aqua and Terra product showed the highest correlation coefficient, coefficient of determination, and Nash–Sutcliffe Efficiency values for all the nine stations. It is worth noting that the performance of discharge prediction is found to be better in the downstream reaches compared to upstream reaches Chiang Saen and Luang Prabang. Further this provides insight on exploring the potential of using other combinations of satellite products or incorporating other data sources to improve the accuracy of discharge simulation or improve the predictive capacity of the model.

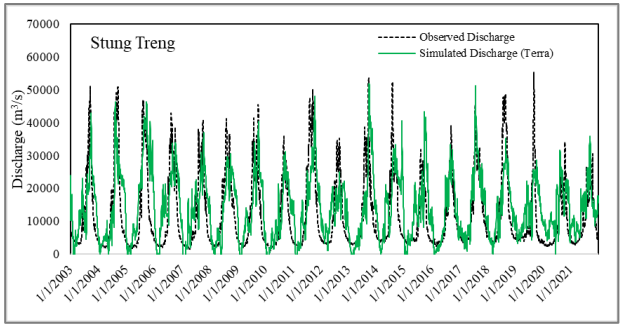
Figure 5-12

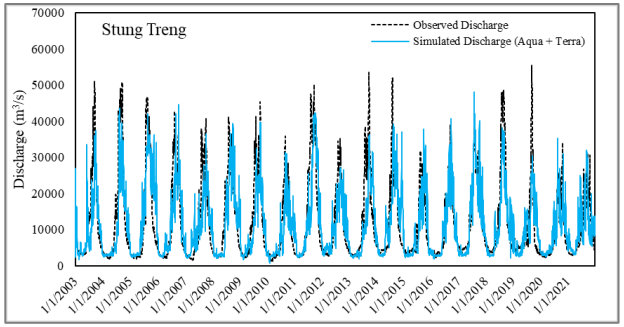
Observed and Predicted Discharge Using LSTM for MODIS-AQUA and MODIS-TERRA and Combination of Both.

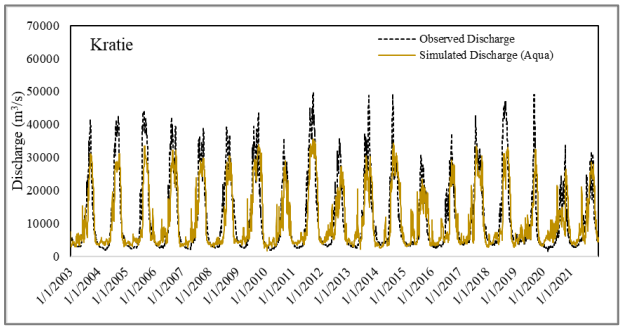


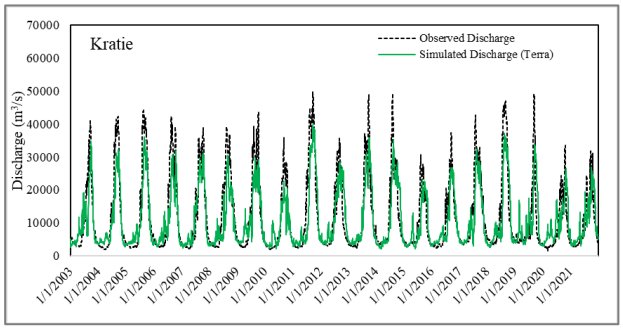

Nakhon Phanom
Observed Discharge
Simulated Discharge (Aqua)
Nakhon Phanom
Observed Discharge
Simulated Discharge (Terra)
Nakhon Phanom
Observed Discharge
Simulated Discharge (Aqua + Terra)
Mukdahan
Observed Discharge
Simulated Discharge (Aqua)
Mukdahan
Observed Discharge
Simulated Discharge (Terra)
Mukdahan
Observed Discharge
Simulated Discharge (Aqua + Terra)
Pakse
Observed Discharge
Simulated Discharge (Aqua)
Pakse
Observed Discharge
Simulated Discharge (Terra)
Pakse
Observed Discharge
Simulated Discharge (Aqua + Terra)
Discharge (m3/s)

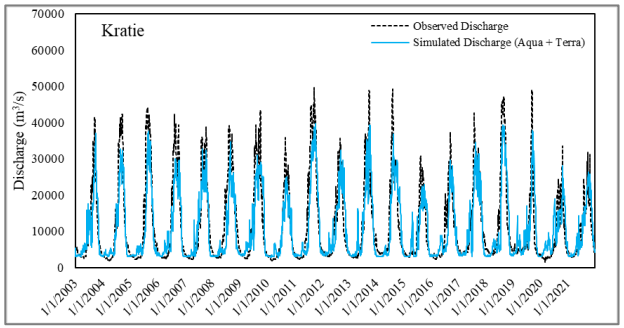

Stung Treng
Observed Discharge
Simulated Discharge (Aqua)
Discharge (m³/s)

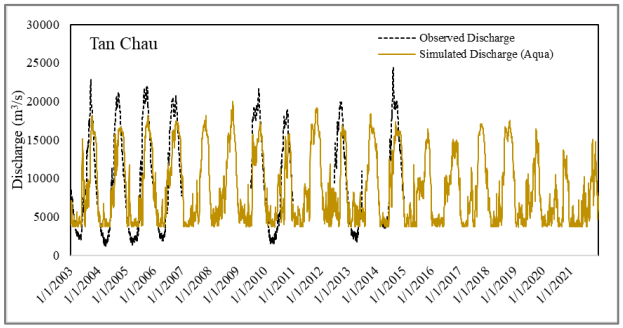

Stung Treng
Observed Discharge
Simulated Discharge (Terra)
Discharge (m³/s)

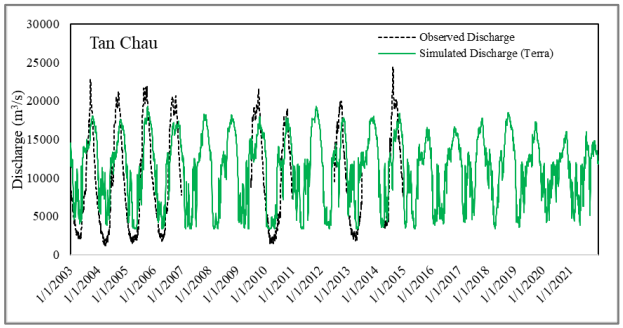

Stung Treng
Observed Discharge
Simulated Discharge (Aqua + Terra)
Discharge (m³/s)


Kratie
Observed Discharge
Simulated Discharge (Aqua)
Discharge (m³/s)


Kratie
Observed Discharge
Simulated Discharge (Terra)
Discharge (m³/s)


Kratie
Observed Discharge
Simulated Discharge (Aqua + Terra)
Discharge (m³/s)


Tan Chau
Observed Discharge
Simulated Discharge (Aqua)
Discharge (m³/s)


Tan Chau
Observed Discharge
Simulated Discharge (Terra)
Discharge (m³/s)

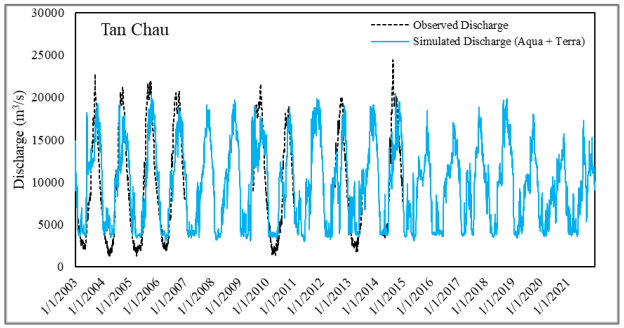

Tan Chau
Observed Discharge
Simulated Discharge (Aqua + Terra)
Discharge (m³/s)

Table 5-5

Performances of Discharge Prediction Using MODIS Aqua, Terra and Combination of Both

Station	Product	Training				Testing			
		R	R^2	NSE	PBIAS	R	R^2	NSE	PBIAS
Chiang Saen	Aqua	0.65	0.42	0.41	-5.97	0.64	0.41	0.40	-5.76
	Terra	0.63	0.40	0.38	-9.11	0.63	0.39	0.37	-9.33
	Aqua + Terra	0.70	0.49	0.47	-9.40	0.70	0.48	0.46	-9.53
Luang Prabang	Aqua	0.49	0.24	0.23	4.98	0.49	0.24	0.23	4.96
	Terra	0.67	0.45	0.44	4.48	0.67	0.45	0.44	4.22
	Aqua + Terra	0.71	0.50	0.48	6.37	0.71	0.51	0.49	5.96
Nong Khai	Aqua	0.75	0.56	0.55	6.95	0.75	0.56	0.55	7.03
	Terra	0.75	0.57	0.56	4.23	0.75	0.57	0.56	4.43
	Aqua + Terra	0.81	0.66	0.65	4.89	0.81	0.65	0.65	5.04
Nakhon Phanom	Aqua	0.73	0.54	0.50	-12.56	0.73	0.54	0.50	-12.64
	Terra	0.75	0.56	0.52	-13.74	0.75	0.56	0.52	-13.75
	Aqua + Terra	0.83	0.69	0.67	-9.96	0.83	0.69	0.67	-10.03
Mukdahan	Aqua	0.71	0.50	0.43	24.40	0.71	0.50	0.42	25.12
	Terra	0.75	0.57	0.55	10.03	0.76	0.57	0.55	9.95
	Aqua + Terra	0.82	0.68	0.66	13.29	0.82	0.68	0.65	13.61
Pakse	Aqua	0.74	0.55	0.55	2.56	0.74	0.55	0.55	2.63
	Terra	0.81	0.66	0.65	3.56	0.81	0.66	0.65	3.57
	Aqua + Terra	0.85	0.72	0.71	2.45	0.85	0.72	0.71	2.45
Stung Treng	Aqua	0.70	0.48	0.48	-1.21	0.70	0.49	0.49	-1.26
	Terra	0.70	0.49	0.43	17.78	0.70	0.49	0.43	17.75
	Aqua + Terra	0.78	0.61	0.61	-3.90	0.78	0.61	0.61	-3.84
Kratie	Aqua	0.84	0.71	0.70	-7.18	0.84	0.71	0.70	-7.25
	Terra	0.86	0.73	0.71	-9.18	0.86	0.73	0.71	-9.19
	Aqua + Terra	0.88	0.77	0.76	-9.41	0.88	0.77	0.76	-9.42
Tan Chau	Aqua	0.78	0.61	0.59	-8.30	0.78	0.61	0.59	-8.30
	Terra	0.77	0.60	0.54	15.26	0.77	0.60	0.54	15.22
	Aqua + Terra	0.80	0.64	0.63	1.92	0.80	0.64	0.63	2.04

Aqua: C/M from Aqua Product; Terra: C/M from Terra Product

5.2.3 Discharge Prediction Using Altimetry Derived Water Level and Surface Reflectance Ratio

Apart from using C/M* for discharge prediction, this study incorporated altimetry derived water level as input feature to assess the use of multi-source remote sensing products for reliable and accurate estimation of river discharge. Since the correlation between C/M* with discharge is already explored in previous topics, correlation between water level and observed discharge in different in-situ stations along the river reach is explored. Figure 5-13 shows the correlation between altimetry derived water level and in-situ discharge that suggests the changes in water level are associated with changes in discharge. The results indicate that there is a positive correlation between water level and in-situ discharge at all the stations.

Correlation coefficients range from 0.62 to 0.94, indicating a strong relationship between water level and discharge at each station. The station with the high correlation coefficient demonstrated a strong positive correlation between water level and in-situ discharge, indicating that as water level increases at this station, in-situ discharge also increases proportionally.

These findings have important implications for predicting discharge in rivers. Hence, it suggests that water level measurements can be used to predict discharge in a river system. The use of LSTM models to predict discharge based on water level measurements may have practical applications in water management and environmental monitoring. The results of the LSTM model simulations provide further evidence of the relationship between water level and discharge and demonstrate the potential of LSTM models for predicting discharge based on altimetric water level measurements.

Figure 5-13

Correlation Between Altimetry Derived Water Level and In-Situ Discharge at Gauge Stations

Now, LSTM models were developed for each station considered integrating input features from optical sensor remote sensing and altimetry satellites. Two scenarios were evaluated: first using altimetric water level only and second using combination of surface reflectance ratio for Aqua, Terra, and water level as input features. Figure 5-14

presents the graphical comparison between predicted discharge and observed discharge at Chiang Saen, Luang Prabang, Nong Khai, Nakhon Phanom, Mukdahan, Pakse, Stung Treng and Kratie for both cases. The results showed that the LSTM models were able to accurately predict discharge in both cases, with high R-squared values indicating a good fit between observed and simulated discharge. It can be noted, water level only has underestimated high flows but with combined input of multiple products, these peak flows have been better predicted. This can be due to missing water level attributed by temporal resolution of satellite altimetry and fulfilling this gap by use of surface reflectance ratio which have better temporal resolution. This also highlights the importance of using multisource remote sensing data for discharge prediction.

The Table 5-6 shows the results of predicted discharge compared with observed discharge for different stations. The comparison was done for two different cases: using water level only as input feature and using a combination of surface reflectance ratio for Aqua, Terra, and water level as input features. Four metrics were used to evaluate the accuracy of the simulated discharge: Coefficient of Correlation (R), Coefficient of Determination (R^2), Nash–Sutcliffe Efficiency (NSE), and Percent Bias.

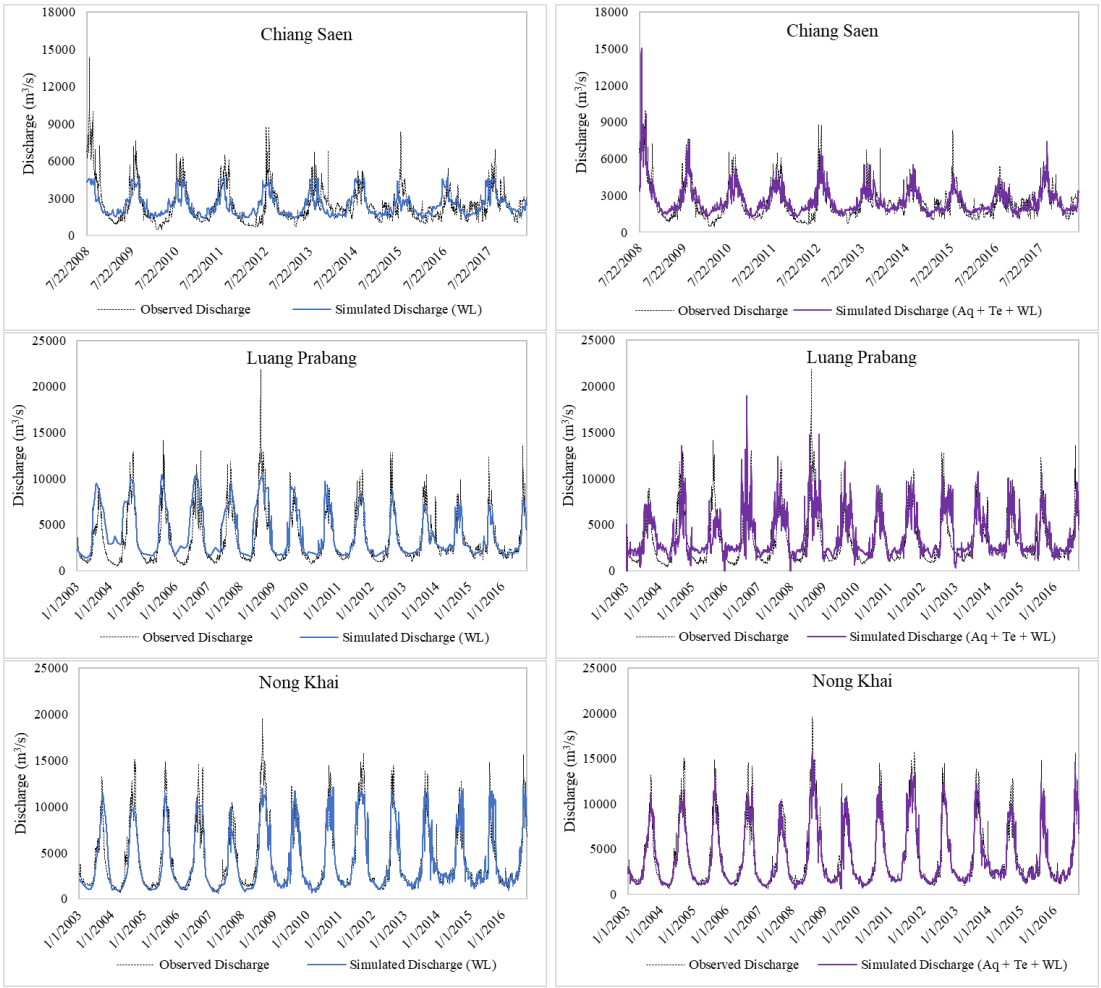
For the Pakse station, both cases produced similar results for training and testing periods, with R values of 0.96 and R^2 values of 0.92 for first case and R values of 0.97 and R^2 values of 0.94 for second case. The NSE values are above 0.90 for both cases, indicating good agreement between simulated and observed discharge. However, there was a difference in the PBIAS values, with values of 9.26 and 12.64 during the testing period and 9.37 and 12.81 for testing period for the water level only and the combined case, respectively. This suggests that the combined case may slightly overestimate the discharge. However, in terms of other performance indicators, combination of Aqua, Terra and water level outperformed use of single water level only for discharge estimation.

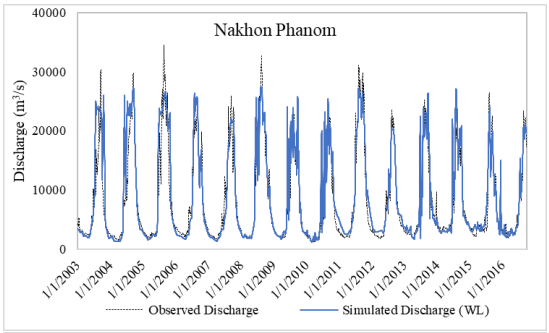
For the Kratie station, the trend of results is similar to Pakse in terms of correlation coefficient, coefficient of determination and Nash Sutcliff Efficiency but the magnitude of performance indicators are lower than Pakse. However, in terms of percent bias, the simulated discharge is slightly underestimated by about 0.25 % which is far better than Pakse for combined input features during both training and testing period. This suggests that the combined case may slightly underestimate the discharge in Kratie station.

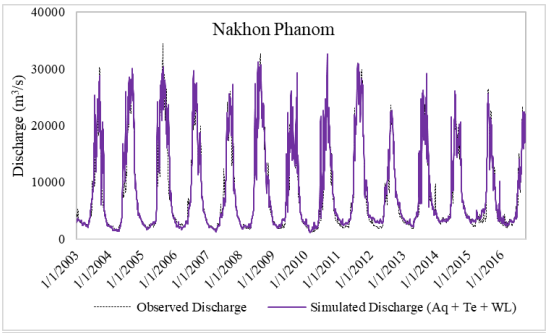
This suggests that altimetric water level measurements can be used as a reliable predictor of discharge, and that the addition of optical remote sensing data can improve the accuracy of discharge predictions. Overall, these findings have important implications for water management and environmental monitoring, as they demonstrate the potential of integrating multiple satellite missions into machine learning models for predicting water flow in rivers and streams. Overall, the results indicate that the combination of Aqua, Terra, and water level data produced better results than the single water level, as evidenced by higher R, R^2, and NSE values for both stations.

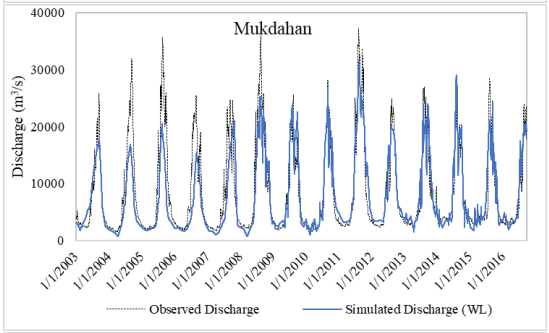
Figure 5-14

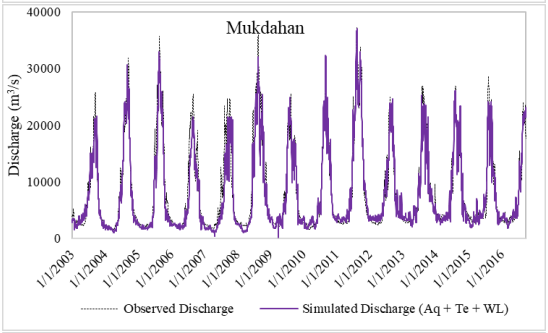
Observed and Simulated Discharge using LSTM for Water Level and Combination of Water Level, MODIS Aqua and Terra.

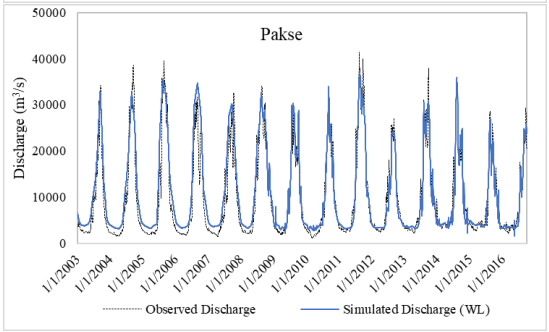


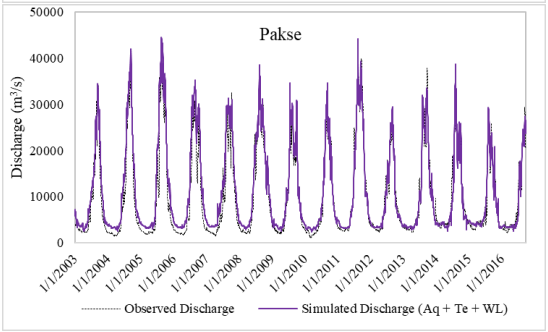

Nakhon Phanom
Discharge (m3/s)
Observed Discharge
Simulated Discharge (WL)

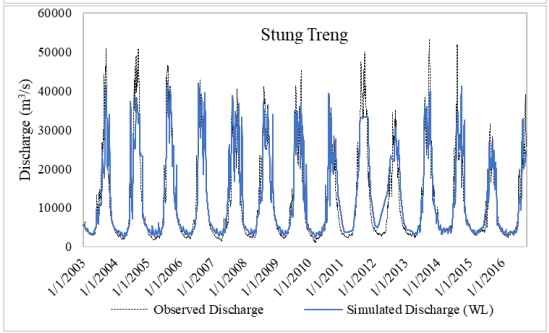

Nakhon Phanom
Discharge (m3/s)
Observed Discharge
Simulated Discharge (Aq + Te + WL)

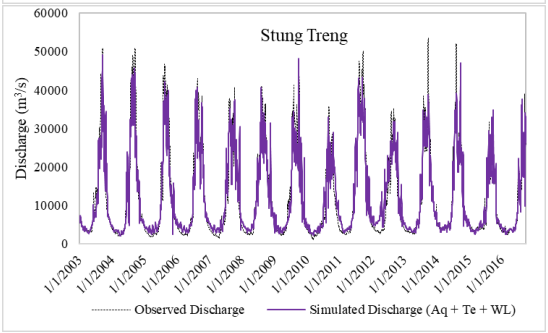

Mukdahan
Discharge (m3/s)
Observed Discharge
Simulated Discharge (WL)

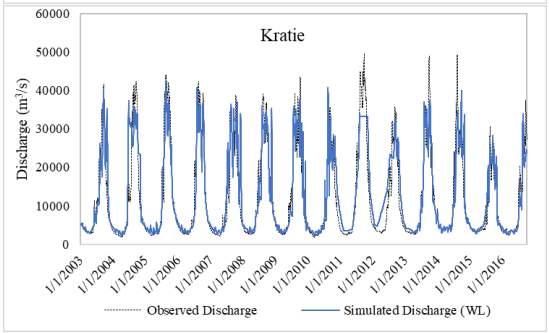

Mukdahan
Discharge (m3/s)
Observed Discharge
Simulated Discharge (Aq + Te + WL)


Pakse
Discharge (m3/s)
Observed Discharge
Simulated Discharge (WL)


Pakse
Discharge (m3/s)
Observed Discharge
Simulated Discharge (Aq + Te + WL)


Stung Treng
Discharge (m3/s)
Observed Discharge
Simulated Discharge (WL)


Stung Treng
Discharge (m3/s)
Observed Discharge
Simulated Discharge (Aq + Te + WL)


Kratie
Discharge (m3/s)
Observed Discharge
Simulated Discharge (WL)

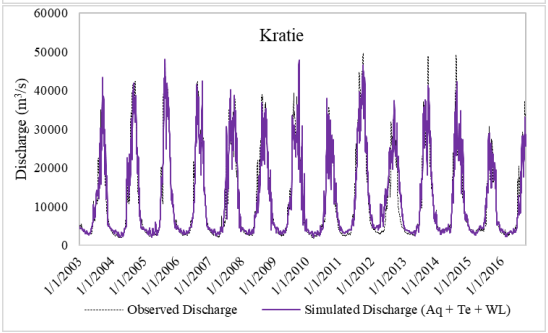

Kratie
Discharge (m3/s)
Observed Discharge
Simulated Discharge (Aq + Te + WL)

Table 5-6

Performances of Simulated Discharge with Observed Discharge Using Water level and Combination of Water Level, MODIS Aqua and Terra.

Station	Product	Training				Testing			
		R	R^2	NSE	PBIAS	R	R^2	NSE	PBIAS
Chiang Saen	WL	0.67	0.45	0.44	-2.22	0.66	0.44	0.44	-2.27
	Aq + Te + WL	0.81	0.66	0.65	4.26	0.81	0.66	0.66	4.56
Luang Prabang	WL	0.85	0.73	0.68	16.21	0.85	0.73	0.68	16.30
	Aq + Te + WL	0.75	0.57	0.54	11.41	0.76	0.57	0.54	12.13
Nong Khai	WL	0.92	0.85	0.85	-2.33	0.93	0.86	0.85	-2.16
	Aq + Te + WL	0.95	0.91	0.91	-3.76	0.95	0.91	0.91	-3.50
Nakhon Phanom	WL	0.93	0.86	0.84	6.76	0.93	0.86	0.84	6.78
	Aq + Te + WL	0.95	0.91	0.89	6.94	0.95	0.91	0.90	6.91
Mukdahan	WL	0.90	0.81	0.80	-9.61	0.90	0.81	0.80	-9.59
	Aq + Te + WL	0.95	0.91	0.90	-5.89	0.95	0.91	0.90	-5.78
Pakse	WL	0.96	0.92	0.91	9.26	0.96	0.92	0.91	9.37
	Aq + Te + WL	0.97	0.94	0.92	12.64	0.97	0.94	0.92	12.81
Stung Treng	WL	0.92	0.84	0.84	3.01	0.92	0.84	0.84	3.15
	Aq + Te + WL	0.93	0.87	0.86	-1.20	0.93	0.87	0.86	-0.96
Kratie	WL	0.92	0.84	0.84	2.78	0.92	0.84	0.84	2.89
	Aq + Te + WL	0.95	0.90	0.90	-0.26	0.95	0.90	0.90	-0.25

Aq: C/M from Aqua Product; Te: C/M from Terra Product; WL: Water Level from Altimeter

5.2.4 Influence of Input Features on the Accuracy of Discharge Prediction

Figure 5-15 illustrates the performance of LSTM model to predict daily river discharge at various locations of Lancang Mekong River using input features from MODIS Aqua, MODIS Terra and Altimeter individually and combined. Results indicate the combined use of Aqua, Terra, and altimetry derived water level data as input features leads to more accurate predictions of daily river discharge compared to using single products alone. This suggests that integrating data from multiple sources can enhance the accuracy of discharge predictions. Furthermore, it can be noted that the accuracy of the predictions tends to improve as the measurement location moves downstream from Chiang Saen to Kratie along the river reach.

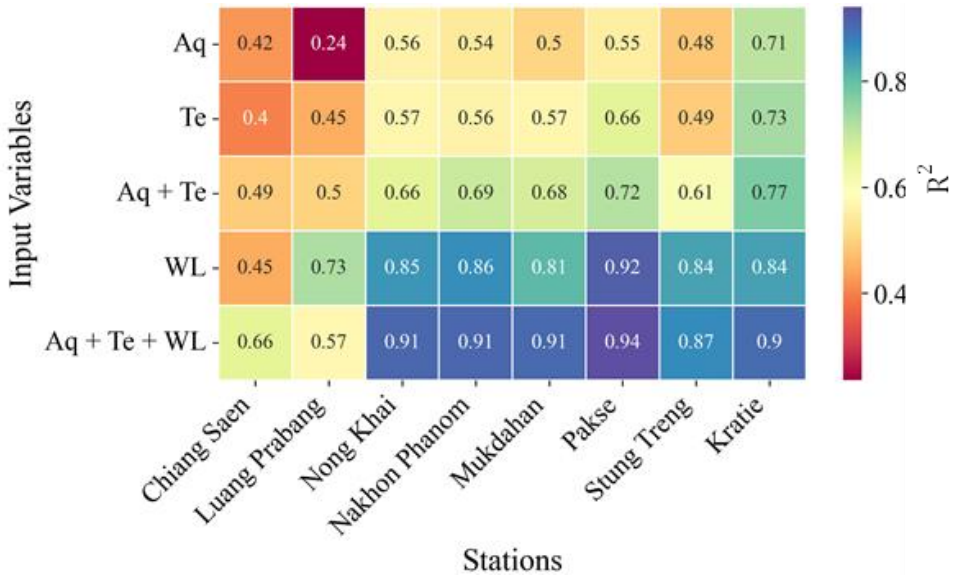
The use of Aqua and Terra satellite data, along with water level data, likely provides a more comprehensive understanding of the complex dynamics of river systems. Aqua and Terra satellites provides valuable information about various environmental parameters, such as temperature, precipitation, and vegetation, which can affect river discharge. Water level data, on the other hand, directly measures the height of the water surface, which is a key parameter for estimating discharge.

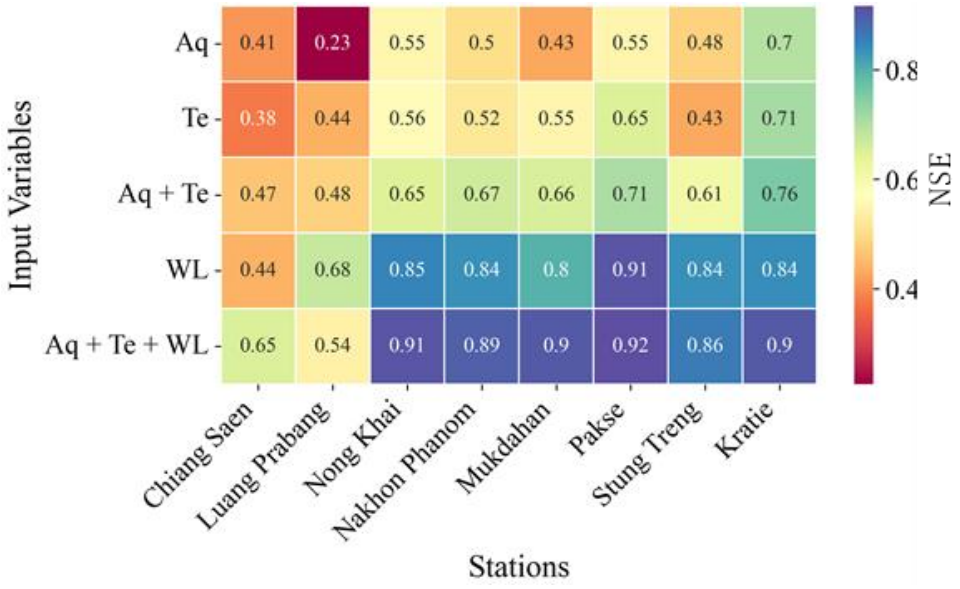
The finding that accuracy increases downstream along the river reach could be attributed to the availability of more reliable and accurate data from downstream locations due to flat topography and wider river, as upstream measurements can be affected by various local factors such as river morphology, land use, and measurement errors.

This demonstrates that using a combination of surface reflectance ratio from Aqua, Terra, and altimetry derived water level data resulted in improved accuracy in predicting river discharge compared to using single products alone. The accuracy of predictions tends to increase as the measurement location moves downstream along the river reach. This highlights the importance of considering multiple sources of data and their spatial context when predicting river discharge, and the potential for using remote sensing data to enhance such predictions.

Figure 5-15

Accuracy of Discharge Prediction with Different Input Features





5.3 Spatial Transferability of LSTM Models for Discharge Prediction at Ungauged Locations in Multiple River Reaches

The accurate prediction of discharge at ungauged locations is a crucial problem in hydrology, with important implications for water resource management and flood forecasting. Traditionally, discharge measurements have been obtained at a limited number of gauged locations along a river, and these measurements have been used to develop regression models to estimate discharge at ungauged locations based on physical and statistical methods. However, such models often suffer from poor performance due to the complexity and nonlinearity of the hydrological processes involved, as well as the spatial variability of the watershed (Samaniego et al., 2010).

In recent years, machine learning techniques such as artificial neural networks (ANNs) and long short-term memory (LSTM) models have shown promise for improving the accuracy of discharge predictions at ungauged locations. These models are capable of capturing the nonlinear and dynamic relationships between hydrological variables, and can learn from large datasets of historical observations to make predictions for new locations (Duan et al., 1992).

In this regard, the development and validation of an LSTM model for predicting discharge at ungauged reaches of a river is focused. Previous assessment showed improved performance in discharge estimation with use of altimetric water level and C/M* from combined Aqua and Terra products. Thus, these combined features are used

as input features for this application and validation. The model was trained using three input features and discharge measurements at four different locations within the river, which are representative of upstream, middle, and downstream reaches in Mekong River. We tested the model's ability to predict discharge at ungauged locations within these reaches' conclusions are made based on the predictive accuracy of model.

5.3.1 Evaluation of LSTM Model to Predict Discharge at Ungauged Reach of River

Four models developed for Nong Khai, Nakhon Phanom, Pakse and Kratie were considered for evaluating the spatial transferability of LSTM models to predict river discharge at ungauged reach. The model identification is given based on the location where it is trained. Thus, the four models are Model Nong Khai, Model Nakhon Phanom, Model Pakse and Model Kratie. Since ungauged reaches are the reach with no measurement of discharge and hence lack observation, these four models are used to predict discharge at other gauging stations considering them as pseudo-ungauged reach. This helps to validate and analyze the model's ability to predict discharge for ungauged reach.

Figure 5-16 presents the discharge predicted by Model Nong Khai during spatial transfer at pseudo-gauged reaches. The model is able to predict annual and seasonal patterns at all predicted stations. The high and low flows are well captured in all stations except Chiang Saen and Luang Prabang with over estimation. Table 5-7 illustrates the performance of model prediction during spatial transfer and predictive accuracy at those stations. The R values range from 0.72 to 0.90, indicating that there is a positive correlation between the predicted and observed river discharge at all nine locations and R^2 ranges from 0.51 to 0.80, showing better fit with observed discharge.

NSE values range from -0.16 to 0.79, with the highest value obtained Pakse and lowest obtained at Luang Prabang and Chaing Saen respectively. Negative values indicate that the model performs worse than a simple mean in Luang Prabang station. Likewise, PBIAS values range from -16.73 to 41.43. These results suggest that the performance of Model Nong Khai varies across the reaches during spatial transfer. The model performed well in terms of four performance indicators at Nakhon Phanom, Mukdahan, Pakse, Stung Treng and Kratie. However, the model could not perform well at Luang Prabang with negative NSE value and high PBIAS value.

Table 5-7

Performance of Model Nong Khai During Spatial Transfer at Pseudo Ungauged Locations

Stations	R	R^2	NSE	PBIAS
Chiang Saen	0.72	0.51	0.32	-14.09
Luang Prabang	0.78	0.60	-0.16	41.43
Chaing Khan	0.85	0.73	0.34	10.37
Nong Khai	0.95	0.91	0.91	-3.58
Nakhon Phanom	0.88	0.77	0.75	-11.41
Mukdahan	0.86	0.75	0.74	3.69
Pakse	0.90	0.80	0.79	3.05
Stung Treng	0.86	0.73	0.73	5.29
Kratie	0.88	0.78	0.70	-16.73

Figure 5-16

Predicted Discharge at Pseudo Ungauged Reaches using LSTM Model Nong Khai

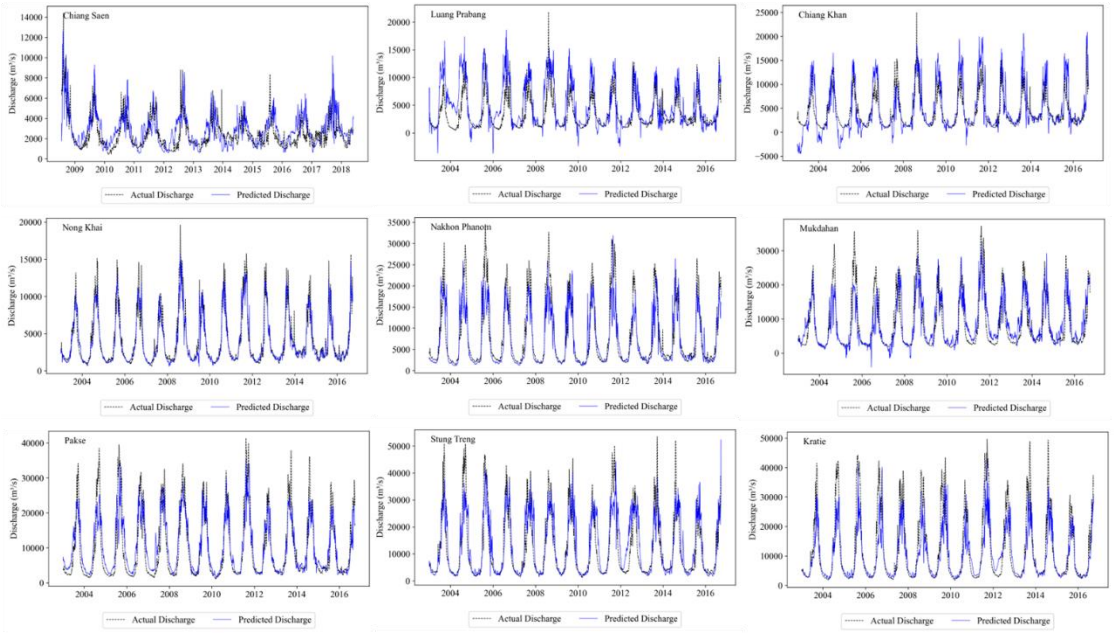


Figure 5-17 shows the discharge predicted by Model Nakhon Phanom in other pseudo gauge station considered. The model is able to predict annual and seasonal fluctuations in river discharge with very good performance except Chiang Saen, performing lower. The high and low flows are also well captured in Nong Khai, Mukdahan, Stung Treng and Kratie. The performance of model prediction during spatial transfer and predictive accuracy at those stations are presented in Table 5-8. The R values range from 0.65 to 0.93, which indicates a positive correlation between the predicted and observed river discharge at all nine locations. On the other hand, R^2 ranges from 0.42 to 0.86.

NSE values range from -1.76 to 0.78, with the highest value obtained Pakse and lower predictive accuracy obtained at Chiang Saen, Luang Prabang and Chaing Khan. Likewise, PBIAS values range from -2.83 % to 83 %. The model performance varied in discharge prediction along different reaches considered. The model performed good to very good performance in terms of all performance indicators at Nong Khai, Mukdahan, Pakse, Stung Treng and Kratie but percent bias are about 25 % in most of these stations. On the other hand, the model could not make reliable predictions at Chiang Saen, Luang Prabang and Chiang Khan supported by negative NSE and high PBIAS value but is able to capture the seasonality.

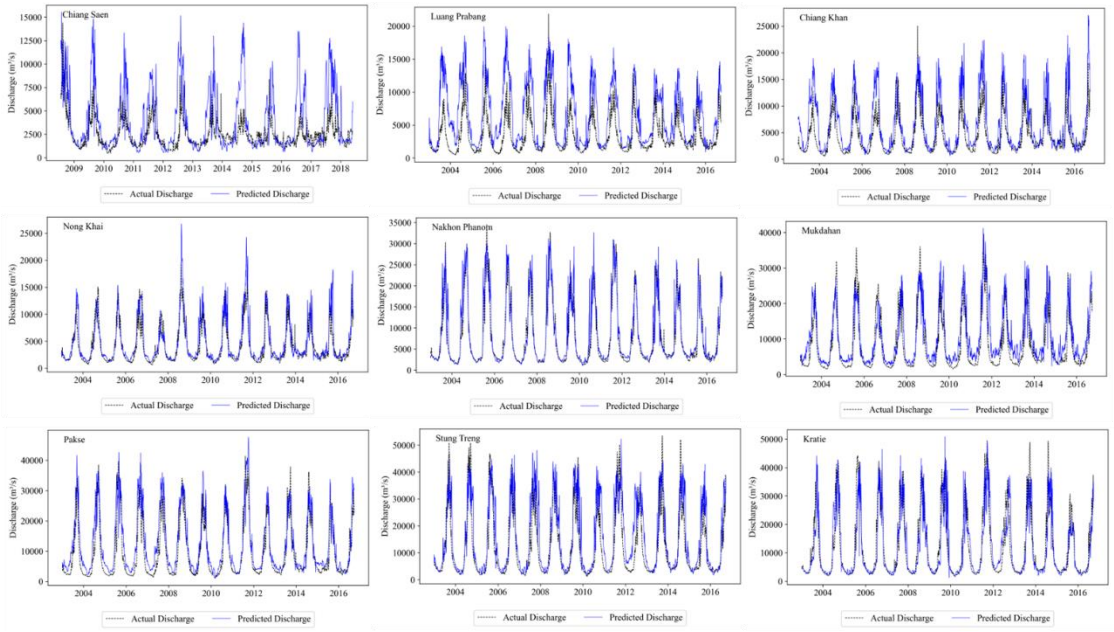
Table 5-8

Performance of Model Nakhon Phanom during Spatial Transfer at Pseudo Ungauged Locations

Stations	R	R^2	NSE	PBIAS
Chiang Saen	0.65	0.42	-1.76	-54.24
Luang Prabang	0.83	0.69	-1.38	93.00
Chaing Khan	0.88	0.77	-0.15	54.72
Nong Khai	0.90	0.81	0.69	15.44
Nakhon Phanom	0.95	0.91	0.90	6.92
Mukdahan	0.89	0.79	0.71	25.12
Pakse	0.93	0.86	0.78	24.50
Stung Treng	0.87	0.76	0.64	26.80
Kratie	0.88	0.78	0.77	-2.83

Figure 5-17

Predicted Discharge at Pseudo Ungauged Reaches using LSTM Model Nakhon Phanom



The discharge predicted by Model Pakse while transferring to other location are provided in Figure 5-18. Except Chiang Saen, the model simulated yearly and seasonal variations in river discharge with very good accuracy. The high and low flows are also well captured in Nong Khai, Mukdahan, Stung Treng and Kratie. The model's performance varied as it was transferred to other locations. In terms of all four performance measures, the model performed well to well in Nong Khai, Mukdahan, Nakhon Phanom, and Stung Treng. Despite negative NSE and high PBIAS values, the model was unable to accurately predict at Chiang Saen, Luang Prabang, Chiang Khan, and Kratie. However, it was able to detect seasonality in these locations.

Table 5-9 depicts the model's performance during spatial transfer and the stations' predictive accuracy. The R values at all nine locations range from 0.65 to 0.93, demonstrating a positive connection between the expected and actual river discharge. However, R2 ranges from 0.52 to 0.88, with Mukhdahan and Nong Khai showing the highest correlation.

NSE values vary between -1.41 to 0.76, with the highest predictive accuracy obtained at Mukdahan and the model was not able to predict reliable discharge at Chiang Saen and Luang Prabang. Similarly, PBIAS values range from -3.56% to 91%, with Nakhon Phanom recording the highest number.

The model's performance varied as it was transferred to other locations. In terms of all four performance measures, the model performed well to well in Nong Khai,

Mukdahan, Nakhon Phanom, and Stung Treng. Despite negative NSE and high PBIAS values, the model was unable to accurately predict at Chiang Saen, Luang Prabang, Chiang Khan, and Kratie. However, it was able to detect seasonality in these locations.

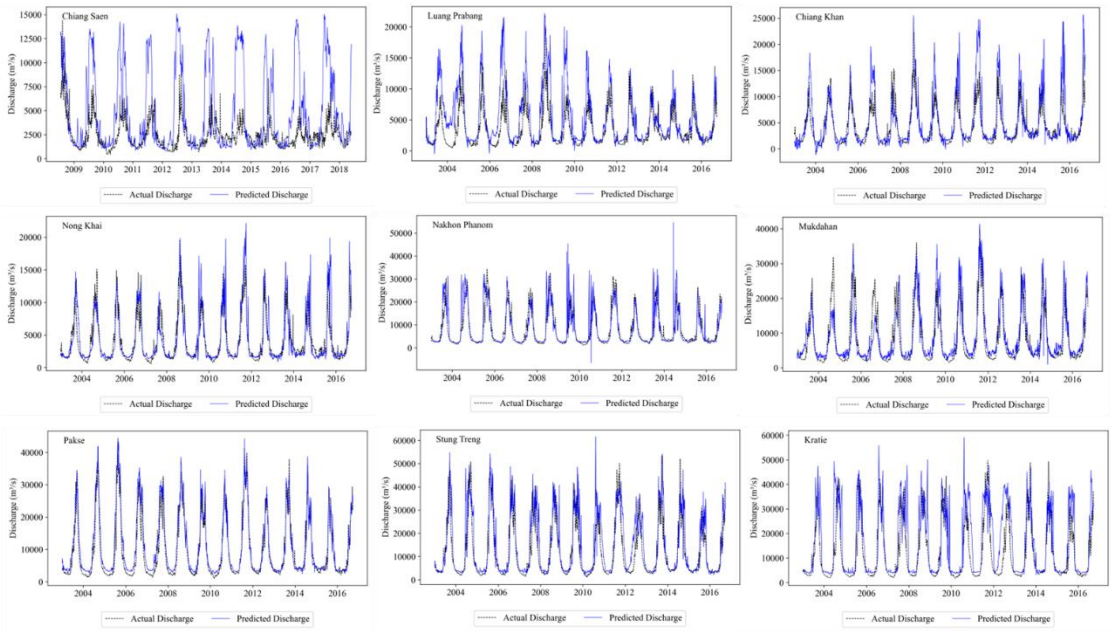
Table 5-9

Performance of Model Pakse During Spatial Transfer at Pseudo Ungauged Locations

Stations	R	R^2	NSE	PBIAS
Chiang Saen	0.52	0.27	-1.41	-91.43
Luang Prabang	0.84	0.71	-0.44	55.66
Chaing Khan	0.86	0.75	0.26	20.78
Nong Khai	0.88	0.77	0.69	4.18
Nakhon Phanom	0.84	0.71	0.65	3.56
Mukdahan	0.88	0.78	0.76	10.19
Pakse	0.97	0.94	0.92	12.72
Stung Treng	0.89	0.79	0.70	19.82
Kratie	0.82	0.68	0.44	32.71

Figure 5-18

Predicted Discharge at Pseudo Ungauged Reaches using LSTM Model Pakse



The discharge predicted by Model Kratie in other spatial locations and compared with observed discharge are provided in Figure 5-19. The model captured annual variation in discharge in all the stations, but the accuracy decreased while moving upstream reaches. The high and low flows are also well captured in Nong Khai, Mukdahan, Pakse and Stung Treng. Table 5-10 depicts the model's performance during spatial transfer and the stations' predictive accuracy. The R values at all nine locations range from 0.66

to 0.93, demonstrating a positive connection between the expected and actual river discharge. Also, R^2 ranges from 0.44 to 0.87, with Pakse and Stung Treng showing the better fit with observed discharge.

NSE varied between -1.46 to 0.82, with the highest predictive accuracy obtained at Pakse and the model was not able to predict reliable discharge at Chiang Saen, Luang Prabang and Chiang Khan. Similar to other models, the model's performance varied as it was transferred to other locations. In terms of all four performance measures, the model performed well in Nong Khai, Mukdahan, Nakhon Phanom, Stung Treng and Pakse. Despite negative NSE and high PBIAS values, the model was unable to accurately predict discharge at Chiang Saen, Luang Prabang, Chiang Khan but was able to predict annual and seasonal variability.

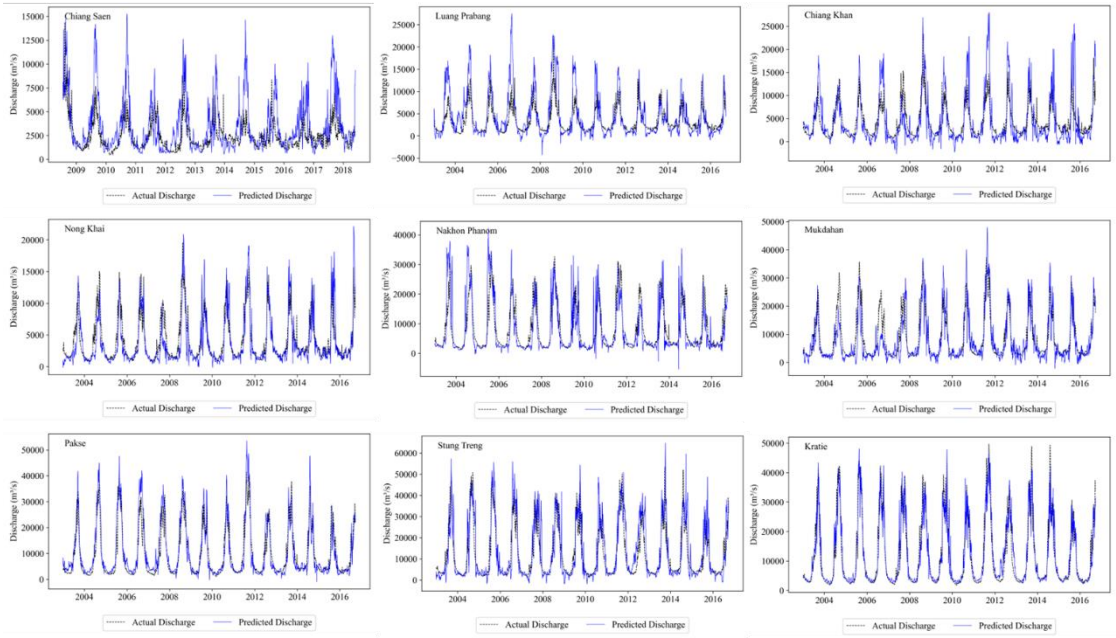
Table 5-10

Performance of Model Kratie During Spatial Transfer at Pseudo-Ungauged Locations

Stations	R	R^2	NSE	PBIAS
Chiang Saen	0.66	0.44	-1.46	-38.30
Luang Prabang	0.83	0.68	-0.49	28.54
Chaing Khan	0.81	0.66	0.005	3.09
Nong Khai	0.87	0.75	0.67	-8.38
Nakhon Phanom	0.84	0.70	0.61	-5.09
Mukdahan	0.85	0.73	0.68	-6.77
Pakse	0.93	0.87	0.82	7.09
Stung Treng	0.89	0.79	0.69	10.15
Kratie	0.95	0.90	0.90	-0.24

Figure 5-19

Predicted Discharge at Pseudo Ungauged Reaches using LSTM Model Kratie



The above results illustrate the spatial transferability of LSTM model using remote sensing data for discharge prediction at ungauged reaches of the river. However, the performance of discharge prediction was found reliable and accurate for certain reach length from the model location where it was trained. The performance of the LSTM model for each station is presented, showing the positive correlation coefficient between observed and predicted discharge.

Figure 5-20 provides the overview summary of four models used to analyze the spatial transferability in terms of performance indicators R^2 and NSE. The results along the Mekong River indicate that the model performs well for certain reaches, but not for others. The model has captured general pattern and trends in the data but struggling to accurately predict actual discharge value for Chiang Saen, Luang Prabang and Chiang Khan. This is due to the different hydrological condition between trained and predicted locations. Specifically, this suggests that the model can be developed for certain reaches of the river that can predict discharge at ungauged locations within that reach with good accuracy. However, this accuracy decreases as the distance from the model station increases.

The model developed for Kratie station can predict discharge accurately for nearby stations such as Stung Treng and Pakse, but not as accurately for stations further away, such as Chiang Khan, Luang Prabang. Similarly, the model developed for Pakse station can predict discharge accurately for nearby stations such as Mukdahan and Stung

Treng, but not as accurately for stations further away, such as Kratie and Chiang Khan. The performance of the LSTM model in predicting discharge is found to decrease as the distance between the trained location and test location increased. Therefore, it can be concluded that the LSTM model developed for a certain reach of the river works best for locations within that reach that have similar catchment properties, and that further improvements are necessary to enhance its accuracy for all stations.

Thus, this study recommends the use of the above four LSTM models for discharge prediction using remote sensing data at ungauged reaches of the Mekong River from Chiang Saen to Kratie depending upon the geographical locations. Figure 5-21 provides the model and its predictive range for discharge prediction at ungauged reach.

Figure 5-20

Model Performance of Spatial Transferability of LSTM Model

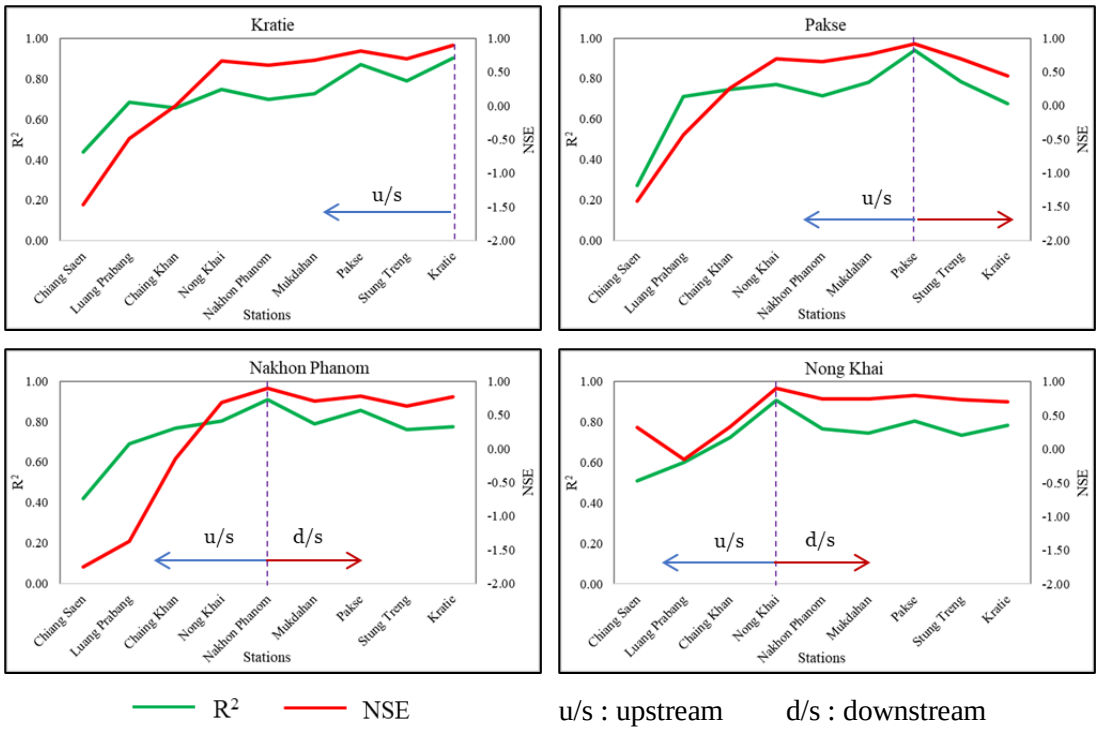
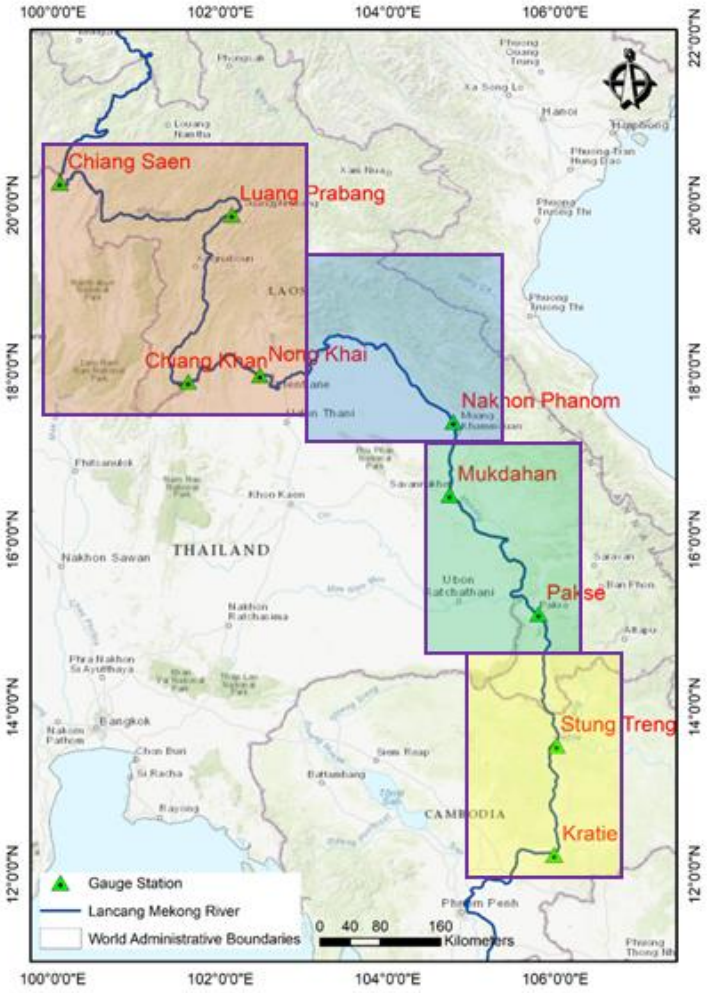


Figure 5-21

The Predictive Range for LSTM Model for Discharge Prediction at Ungauged Reach



5.3.2 Application of Developed Model for Discharge Prediction in Transboundary Locations

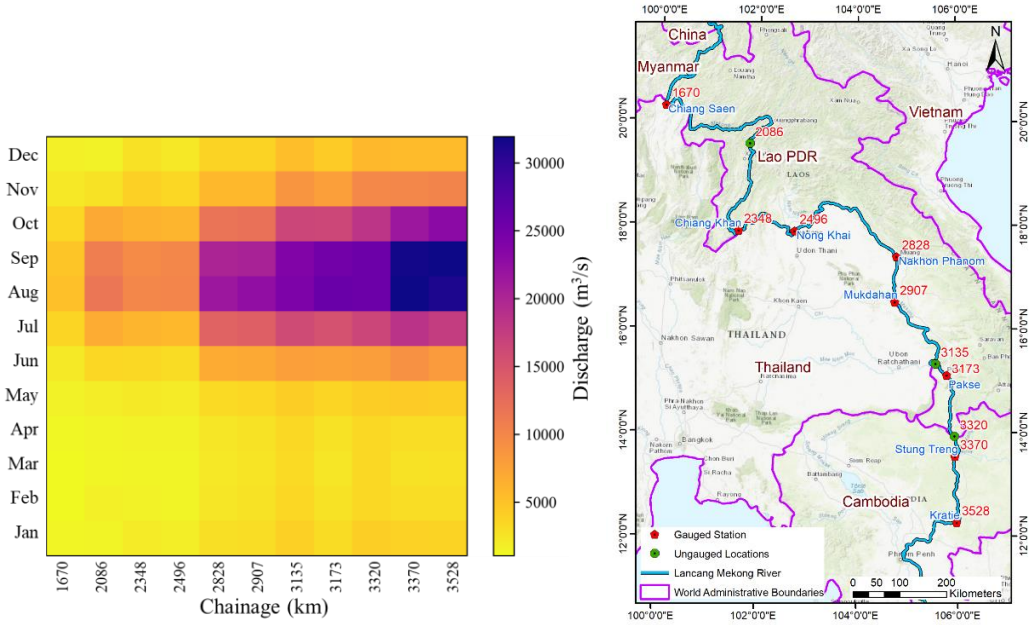
The developed model for predicting river discharge using a combination of surface reflectance ratio of Aqua and Terra products and water level data has shown promising results in predicting discharge at various gauged locations along the river. However, the model's true potential lies in its ability to make predictions in ungauged reaches, especially in transboundary river basins where data availability can be limited. The effective use of the developed model in predicting discharge in transboundary locations can have significant benefits, such as improving water resource management and planning and promoting cooperation between countries sharing the same river basin. The model's ability to make predictions in ungauged reaches using remote sensing data can help overcome data scarcity issues and provide valuable information for decision-makers.

To explore spatial-temporal trends in discharge in the Lancang Mekong River, data from various gauged stations along the river were analyzed. Further, discharge prediction have been made using the developed model at ungauged reaches to enrich the spatial variability along the river. Figure 5-22 shows average monthly spatial-temporal variation for these locations over the past 12 years. The river discharge tends to increase as it moves towards the downstream locations and during the wet season (July to October). Same spatial variability can be observed on the predicted discharge at ungauged locations proving effectiveness of the developed model. These findings have important implications for water resources management in the region, as they suggest that water availability may be higher during the wet season and that downstream areas may be more susceptible to flooding.

In conclusion, the application of the developed model for discharge prediction in transboundary locations can provide valuable insights into the water dynamics of shared river basins. The spatial and temporal prediction capabilities of the model, using machine learning algorithms, can help identify areas at risk of flooding or water scarcity, and facilitate improved water resource management and planning. The model's effective use can promote cooperation between countries and help ensure the sustainable use of shared water resources.

Figure 5-22

Spatio-Temporal Variation of Discharge Along Mekong River



CHAPTER 6
SUMMARY, CONCLUSION AND RECOMMENDATION

6.1 Summary

Estimating river discharge is a critical aspect of water resources management, as it provides valuable information for decision making related to irrigation, hydropower, flood forecasting, and environmental conservation. Traditionally, water discharge has been estimated using ground based gauging stations, which provide accurate and reliable measurements, but are often limited in their spatial coverage and data gaps. Satellite based remote sensing offers an alternative approach for water discharge prediction, as it provides spatially distributed and temporally continuous data that can be used to infer river water level and discharge. However, the development of accurate and reliable methods for water discharge estimation from satellite data remains a challenging task, particularly for ungauged locations where ground-based data is not available.

In this context, this research aims to develop a Long Short-Term Memory model to predict river discharge incorporating input features from multiple satellite missions and assess its applicability for ungauged locations.

First, the estimation of water level at gauging stations was carried out following the densification of water level procedure developed by Tourian et al. (2016). This procedure involves integrating altimetric water level measurements at virtual stations located between the actual gauging stations by transferring those water levels spatially. Since the temporal resolution of Jason-1/2 is 10 days and ENVISAT and Saral is 35 days, the densification procedure combined altimetric water level from various virtual stations to gauge station obtaining high temporal resolution that was used for estimating water discharge.

Further, reflectance ration of dry to wet pixel (C/M) from Near Infrared band of optical images was used as a proxy for water discharge, which is obtained from MODIS Aqua and Terra. Results demonstrate the C/M ratio is correlated with the discharge and further smoothing obtained C/M time series with exponential filter improved the degree of correlation with river discharge. However, the location of the C and M pixels near to the gauging station needs to be carefully selected, which is important for accurate

prediction. In this study, the best location for the C and M pixels was defined as the location where the C/M ratio was best correlated with observed discharge and identified by computing the correlation for each pixel.

Separate LSTM models were developed for each station considered which takes input features of C/M* from MODIS Aqua, MODIS Terra and altimetry derived water level and also utilizing combined input features to assess the influence of different input feature on accuracy of discharge prediction. The inclusion combined input feature i.e., C/M* from Aqua and Terra and altimetric water level showed improved and better performance in discharge prediction with predictive accuracy ranging from 0.54 to 0.92 in terms of NSE. The model was able to predict discharge with high accuracy in the downstream reaches compared to upstream reaches.

Further, assessment on spatial transferability of LSTM model to predict river discharge at ungauged location was carried out. The result showed model performs better for certain reaches of the river, but performance decreases while moving farther from the trained reach. This indicates the model can be developed for certain reaches of the river that can predict discharge at ungauged locations within that reach with good accuracy.

The results of the study demonstrate that the altimetry derived water level and the ratio of reflectance from dry to wet pixel obtained from optical sensor remote sensing data integrated into LSTM model performs well in predicting daily river discharge from multiple satellite missions and is able to provide reliable estimates for ungauged locations. The study also highlights the importance of using multiple satellite missions to increase the temporal resolution of altimetry derived water level as well as improving model performance, as each mission provides unique information that can be used to improve the accuracy and reliability of the predicted discharge.

The findings of the study have important implications for water resources management, particularly in regions where ground based gauging stations are scarce or unreliable. The use of satellite data for river discharge estimation can provide spatially distributed and temporally continuous data that can be used to support decision making related to water resources management. The LSTM model developed in this study provides a powerful tool for river discharge estimation from multiple satellite missions and can be used to improve our understanding of water flow dynamics in a range of settings.

6.2 Conclusion

The following conclusion are made based on this research:

1. Time series water level can be estimated from the satellite observations with good accuracy and use of multiple altimetry missions helps to increase the temporal resolution of water level in Lancang Mekong River.
2. The combination of multiple optical sensors provided better performance in simulating discharge compared to use of single sensor. This is demonstrated by combined use of MODIS Aqua and Terra products which have predicted discharge with higher performance indices at all reaches of Lancang Mekong River compared to individual Aqua and Terra products.
3. The multi mission approach which uses optical and radar altimetry data within LSTM model demonstrated its effectiveness in predicting river discharge along the Lancang Mekong River.
4. The LSTM model predictive accuracy can vary between different locations with some stations exhibiting lower accuracy despite model's ability to capture underline patterns.
5. LSTM model trained on discharge data with remote sensing inputs can be transferred to ungauged reaches for discharge prediction with good accuracy. However, performance decreases as the distance between trained and predicted location increases.
6. Overall, LSTM with remote sensing inputs has potential for discharge prediction in ungauged reaches, but limitations exist.

6.3 Recommendation and Further Research

1. Use of freely available Remote sensing data and publicly available hydrological data can be a cost effective and accessible alternative to traditional gauging method for river discharge monitoring.
2. Policy makers and water resources managers could benefit from considering the use of remote sensing and machine learning for predicting river discharge, particularly in regions where gauging infrastructures are limited or unavailable.
3. Exploring the potential benefit of incorporating additional remote sensing data such as LANDSAT which could help to fill data gaps and improve model

performance particularly with areas with frequent cloud cover or image distortion.

4. Investigation on the applicability of the LSTM model with remote sensing inputs to other river systems and regions can be carried out.
5. The potential of other machine learning models for discharge prediction with remote sensing inputs can be explored.
6. The potential of using remote sensing data to estimate additional hydrological variables, such as water quality and sediment transport can be carried out.

REFERENCES

Ahn, J.-H., & Park, Y.-J. (2020). Estimating Water Reflectance at Near-Infrared Wavelengths for Turbid Water Atmospheric Correction: A Preliminary Study for GOCI-II. *Remote Sensing*, *12*(22), Article 22. https://doi.org/10.3390/rs12223791

Arnell, N. W., & Gosling, S. N. (2013). The impacts of climate change on river flow regimes at the global scale. *Journal of Hydrology*, *486*, 351–364. https://doi.org/10.1016/j.jhydrol.2013.02.010

Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

Biancamaria, S., Andreadis, K. M., Durand, M., Clark, E. A., Rodriguez, E., Mognard, N. M., Alsdorf, D. E., Lettenmaier, D. P., & Oudin, Y. (2010). Preliminary Characterization of SWOT Hydrology Error Budget and Global Capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1), 6–19. https://doi.org/10.1109/JSTARS.2009.2034614

Biancamaria, S., Frappart, F., Leleu, A.-S., Marieu, V., Blumstein, D., Desjonquères, J.-D., Boy, F., Sottolichio, A., & Valle-Levinson, A. (2017). Satellite radar altimetry water elevations performance over a 200m wide river: Evaluation over the Garonne River. *Advances in Space Research*, *59*(1), 128–146. https://doi.org/10.1016/j.asr.2016.10.008

Birkinshaw, S. J., Moore, P., Kilsby, C. G., O'Donnell, G. M., Hardy, A. J., & Berry, P. A. M. (2014a). Daily discharge estimation at ungauged river sites using remote sensing: DAILY DISCHARGE ESTIMATION USING REMOTE SENSING. *Hydrological Processes*, *28*(3), 1043–1054. https://doi.org/10.1002/hyp.9647

Birkinshaw, S. J., Moore, P., Kilsby, C. G., O'Donnell, G. M., Hardy, A. J., & Berry, P. A. M. (2014b). Daily discharge estimation at ungauged river sites using remote sensing: DAILY DISCHARGE ESTIMATION USING REMOTE SENSING. *Hydrological Processes*, *28*(3), Article 3. https://doi.org/10.1002/hyp.9647

Birkinshaw, S. J., O'Donnell, G. M., Moore, P., Kilsby, C. G., Fowler, H. J., & Berry, P. A. M. (2010). Using satellite altimetry data to augment flow estimation techniques on the Mekong River. *Hydrological Processes*, *24*(26), 3811–3825. https://doi.org/10.1002/hyp.7811

Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W., & Garambois, P.-A. (2018). Satellite remote sensing estimation of river

discharge: Application to the Yukon River Alaska. *Journal of Hydrology*, *561*, 1000–1018. https://doi.org/10.1016/j.jhydrol.2018.04.005

Bjerklie, D. M., Lawrence Dingman, S., Vorosmarty, C. J., Bolster, C. H., & Congalton, R. G. (2003). Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, *278*(1–4), 17–38. https://doi.org/10.1016/S0022-1694(03)00129-X

Bjerklie, D. M., Moller, D., Smith, L. C., & Dingman, S. L. (2005a). Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology*, *309*(1–4), 191–209. https://doi.org/10.1016/j.jhydrol.2004.11.022

Bjerklie, D. M., Moller, D., Smith, L. C., & Dingman, S. L. (2005b). Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology*, *309*(1–4), Article 1–4. https://doi.org/10.1016/j.jhydrol.2004.11.022

Bogning, S., Frappart, F., Blarel, F., Niño, F., Mahé, G., Bricquet, J.-P., Seyler, F., Onguéné, R., Etamé, J., Paiz, M.-C., & Braun, J.-J. (2018). Monitoring Water Levels and Discharges Using Radar Altimetry in an Ungauged River Basin: The Case of the Ogooué. *Remote Sensing*, *10*(2), Article 2. https://doi.org/10.3390/rs10020350

Brakenridge, G. R., Nghiem, S. V., Anderson, E., & Chien, S. (2005). Space-based measurement of river runoff. *Eos, Transactions American Geophysical Union*, *86*(19), 185–188. https://doi.org/10.1029/2005EO190001

Brakenridge, G. R., Nghiem, S. V., Anderson, E., & Mic, R. (2007). Orbital microwave measurement of river discharge and ice status. *Water Resources Research*, *43*(4). https://doi.org/10.1029/2006WR005238

BRAKENRIDGE, R., & ANDERSON, E. (2006). MODIS-BASED FLOOD DETECTION, MAPPING AND MEASUREMENT: THE POTENTIAL FOR OPERATIONAL HYDROLOGICAL APPLICATIONS. In J. Marsalek, G. Stancalie, & G. Balint (Eds.), *Transboundary Floods: Reducing Risks Through Flood Management* (pp. 1–12). Springer Netherlands. https://doi.org/10.1007/1-4020-4902-1_1

Calmant, S., & Seyler, F. (2006). Continental surface waters from satellite altimetry. *Comptes Rendus Geoscience*, *338*(14–15), 1113–1122. https://doi.org/10.1016/j.crte.2006.05.012

Camps-Valls, G. (2009). Machine learning in remote sensing data processing. *2009 IEEE International Workshop on Machine Learning for Signal Processing*, 1–6. https://doi.org/10.1109/MLSP.2009.5306233

Cheng, S., Qiao, X., Shi, Y., & Wang, D. (2021). Machine learning for predicting discharge fluctuation of a karst spring in North China. *Acta Geophysica*, *69*(1), 257–270. https://doi.org/10.1007/s11600-020-00522-0

Chiang, Y.-M., Hao, R.-N., Zhang, J.-Q., Lin, Y.-T., & Tsai, W.-P. (2018). Identifying the Sensitivity of Ensemble Streamflow Prediction by Artificial Intelligence. *Water*, *10*(10), 1341. https://doi.org/10.3390/w10101341

Depetris, P. J. (2021). The Importance of Monitoring River Water Discharge. *Frontiers in Water*, *3*. https://www.frontiersin.org/articles/10.3389/frwa.2021.745912

Derecki, J. A., & Quinn, F. H. (1987). Use of current meters for continuous measurement of flows in large rivers. *Water Resources Research*, *23*(9), 1751–1756. https://doi.org/10.1029/WR023i009p01751

Dingman, S. L., & Bjerklie, D. M. (2005). Estimation of River Discharge. In M. G. Anderson & J. J. McDonnell (Eds.), *Encyclopedia of Hydrological Sciences* (p. hsa069). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470848944.hsa069

Domeneghetti, A., Castellarin, A., Tarpanelli, A., & Moramarco, T. (2015). Investigating the uncertainty of satellite altimetry products for hydrodynamic modelling. *Hydrological Processes*, *29*(23), 4908–4918. https://doi.org/10.1002/hyp.10507

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/10.1029/91WR02985

Elumalai, V., Brindha, K., Sithole, B., & Lakshmanan, E. (2017). Spatial interpolation methods and geostatistics for mapping groundwater contamination in a coastal area. *Environmental Science and Pollution Research*, *24*(12), 11601–11617. https://doi.org/10.1007/s11356-017-8681-6

Esmaeilzadeh, B., Sattari, M. T., & Samadianfard, S. (2017). Performance evaluation of ANNs and an M5 model tree in Sattarkhan Reservoir inflow prediction. *ISH Journal of Hydraulic Engineering*, *23*(3), 283–292. https://doi.org/10.1080/09715010.2017.1308277

Fekete, B. M., & Vörösmarty, C. J. (n.d.). *The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements*. 8.

Filippucci, P., Brocca, L., Bonafoni, S., & Tarpanelli, A. (2022). *River Discharge estimation from optical satellite data: Latest advances using NIR sensors* [Other]. display. https://doi.org/10.5194/egusphere-egu22-3513

Florkowski, T., Davis, T. G., Wallander, B., & Prabhakar, D. R. L. (1969). The measurement of high discharges in turbulent rivers using tritium tracer. *Journal of Hydrology*, *8*(3), 249–264. https://doi.org/10.1016/0022-1694(69)90001-8

Frappart, F., Blumstein, D., Cazenave, A., Ramillien, G., Birol, F., Morrow, R., & Rémy, F. (2017). Satellite Altimetry: Principles and Applications in Earth Sciences. In J. G. Webster, *Wiley Encyclopedia of Electrical and Electronics Engineering* (pp. 1–25). John Wiley & Sons, Inc. https://doi.org/10.1002/047134608X.W1125.pub2

Frappart, F., Papa, F., Silva, J. S. da, Ramillien, G., Prigent, C., Seyler, F., & Calmant, S. (2012). Surface freshwater storage and dynamics in the Amazon basin during the 2005 exceptional drought. *Environmental Research Letters*, *7*(4), 044010. https://doi.org/10.1088/1748-9326/7/4/044010

Garkoti, A., & Kundapura, S. (2021). Deriving water level and discharge estimation using satellite altimetry for Krishna River, Karnataka. *Remote Sensing Applications: Society and Environment*, *22*, 100487. https://doi.org/10.1016/j.rsase.2021.100487

Ghumman, A. R., Ghazaw, Y. M., Sohail, A. R., & Watanabe, K. (2011). Runoff forecasting by artificial neural network and conventional model. *Alexandria Engineering Journal*, *50*(4), 345–350. https://doi.org/10.1016/j.aej.2012.01.005

Gleason, C., & Durand, M. (2020). Remote Sensing of River Discharge: A Review and a Framing for the Discipline. *Remote Sensing*, *12*(7), 1107. https://doi.org/10.3390/rs12071107

Gravelle, R. (2015). *Discharge Estimation: Techniques and Equipment*.

GRDC Data Portal. (n.d.). Retrieved June 16, 2022, from https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser#dataDownload/Home

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

He, D., Lu, Y., & Li, Z. (n.d.). *Chapter 14—Watercourse Environmental Change in Upper Mekong*. 28.

Hirsch, R. M., & Costa, J. E. (2004). U.S. stream flow measurement and data dissemination improve. *Eos, Transactions American Geophysical Union*, *85*(20), 197–203. https://doi.org/10.1029/2004EO200002

Hoang, L. P., Lauri, H., Kummu, M., Koponen, J., van Vliet, M. T. H., Supit, I., Leemans, R., Kabat, P., & Ludwig, F. (2016). Mekong River flow and hydrological extremes under climate change. *Hydrology and Earth System Sciences*, *20*(7), 3027–3041. https://doi.org/10.5194/hess-20-3027-2016

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hou, J., van Dijk, A. I. J. M., Renzullo, L. J., & Vertessy, R. A. (2018). Using modelled discharge to develop satellite-based river gauging: A case study for the Amazon Basin. *Hydrology and Earth System Sciences*, *22*(12), 6435–6448. https://doi.org/10.5194/hess-22-6435-2018

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water*, *10*(11), 1543. https://doi.org/10.3390/w10111543

Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., Hou, A., & Hong, Y. (2018a). Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River. *Remote Sensing of Environment*, *219*, 115–134. https://doi.org/10.1016/j.rse.2018.10.008

Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., Hou, A., & Hong, Y. (2018b). Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River. *Remote Sensing of Environment*, *219*, 115–134. https://doi.org/10.1016/j.rse.2018.10.008

Jiang, L., Schneider, R., Andersen, O., & Bauer-Gottwein, P. (2017). CryoSat-2 Altimetry Applications over Rivers and Lakes. *Water*, *9*, 211. https://doi.org/10.3390/w9030211

Kao, I.-F., Zhou, Y., Chang, L.-C., & Chang, F.-J. (2020). Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. *Journal of Hydrology*, *583*, 124631. https://doi.org/10.1016/j.jhydrol.2020.124631

Kashid, S. G., & Pardeshi, S. A. (2014). A survey of water distribution system and new approach to intelligent water distribution system. *2014 First International Conference on Networks & Soft Computing (ICNSC2014)*, 339–344. https://doi.org/10.1109/CNSC.2014.6906645

Kenny, J. F., Barber, N. L., Hutson, S. S., Linsey, K. S., Lovelace, J. K., & Maupin, M. A. (2009). *Estimated use of water in the United States in 2005* (Report No. 1344;

Circular, p. 60). USGS Publications Warehouse. https://doi.org/10.3133/cir1344

Kim, Lee, Chang, Bui, Jayasinghe, Basnayake, Chishtie, & Hwang. (2019). Daily River Discharge Estimation Using Multi-Mission Radar Altimetry Data and Ensemble Learning Regression in the Lower Mekong River Basin. *Remote Sensing*, *11*(22), Article 22. https://doi.org/10.3390/rs11222684

Kim, Y., Schmid, T., Charbiwala, Z. M., Friedman, J., & Srivastava, M. B. (2008). NAWMS: Nonintrusive autonomous water monitoring system. *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems - SenSys '08*, 309. https://doi.org/10.1145/1460412.1460443

Kouraev, A. V., Zakharova, E. A., Samain, O., Mognard, N. M., & Cazenave, A. (2004). Ob' river discharge from TOPEX/Poseidon satellite altimetry (1992–2002). *Remote Sensing of Environment*, *93*(1), 238–245. https://doi.org/10.1016/j.rse.2004.07.007

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). *Rainfall-Runoff modelling using Long-Short-Term-Memory (LSTM) networks* [Preprint]. Catchment hydrology/Modelling approaches. https://doi.org/10.5194/hess-2018-247

Leon, J. G., Calmant, S., Seyler, F., Bonnet, M.-P., Cauhopé, M., Frappart, F., Filizola, N., & Fraizy, P. (2006). Rating curves and estimation of average water depth at the upper Negro River based on satellite altimeter data and modeled discharges. *Journal of Hydrology*, *328*(3), 481–496. https://doi.org/10.1016/j.jhydrol.2005.12.006

Li, H., Li, H., Wang, J., & Hao, X. (2019). Extending the Ability of Near-Infrared Images to Monitor Small River Discharge on the Northeastern Tibetan Plateau. *Water Resources Research*, *55*(11), 8404–8421. https://doi.org/10.1029/2018WR023808

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Liu, K.-T., Tseng, K.-H., Shum, C., Liu, C.-Y., Kuo, C.-Y., Liu, G., Jia, Y., & Shang, K. (2016). Assessment of the Impact of Reservoirs in the Upper Mekong River Using Satellite Radar Altimetry and Remote Sensing Imageries. *Remote Sensing*, *8*(5), 367. https://doi.org/10.3390/rs8050367

Mengen, D., Ottinger, M., Leinenkugel, P., & Ribbe, L. (2020). Modeling River Discharge Using Automated River Width Measurements Derived from Sentinel-1 Time Series. *Remote Sensing*, *12*(19), Article 19. https://doi.org/10.3390/rs12193236

Milliman, J. D., & Farnsworth, K. L. (2011). *River Discharge to the Coastal Ocean: A Global Synthesis* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511781247

MODIS Web. (n.d.). Retrieved July 30, 2022, from https://modis.gsfc.nasa.gov/data/

Nogueira Filho, F. J. M., Souza Filho, F. de A., Porto, V. C., Vieira Rocha, R., Sousa Estácio, Á. B., & Martins, E. S. P. R. (2022). Deep Learning for Streamflow Regionalization for Ungauged Basins: Application of Long-Short-Term-Memory Cells in Semiarid Regions. *Water*, *14*(9), Article 9. https://doi.org/10.3390/w14091318

Ordóñez, F. J., & Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, *16*(1), Article 1. https://doi.org/10.3390/s16010115

Pan, F. (2013). Remote sensing of river stage and discharge. *SPIE Newsroom*. https://doi.org/10.1117/2.1201212.004611

Papa, F., Durand, F., Rossow, W. B., Rahman, A., & Bala, S. K. (2010). Satellite altimeter-derived monthly discharge of the Ganga-Brahmaputra River and its seasonal to interannual variations from 1993 to 2008. *Journal of Geophysical Research: Oceans*, *115*(C12). https://doi.org/10.1029/2009JC006075

Perumal, M., Moramarco, T., Sahoo, B., & Barbetta, S. (2007). A methodology for discharge estimation and rating curve development at ungauged river sites: RATING CURVE ESTIMATION AT UNGAUGED RIVER SITES. *Water Resources Research*, *43*(2). https://doi.org/10.1029/2005WR004609

Robert Brakenridge, G., Cohen, S., Kettner, A. J., De Groeve, T., Nghiem, S. V., Syvitski, J. P. M., & Fekete, B. M. (2012). Calibration of satellite measurements of river discharge using a global hydrology model. *Journal of Hydrology*, *475*, 123–136. https://doi.org/10.1016/j.jhydrol.2012.09.035

Sahoo, D. P., Sahoo, B., & Tiwari, M. K. (2020a). *Performance Evaluation of Remote Sensing-based High Frequent Streamflow Estimation Models at the Bramhani River Basin Outlet*. *2020*, H011-0003.

Sahoo, D. P., Sahoo, B., & Tiwari, M. K. (2020b). Copula-based probabilistic spectral algorithms for high-frequent streamflow estimation. *Remote Sensing of Environment*, *251*, 112092. https://doi.org/10.1016/j.rse.2020.112092

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale: MULTISCALE PARAMETER REGIONALIZATION. *Water Resources Research*, *46*(5). https://doi.org/10.1029/2008WR007327

Scherer, D., Schwatke, C., & Dettmering, D. (n.d.). *Estimation of River Discharge using Multi-Mission Satellite Altimetry and Optical Remote Sensing Imagery*. 15.

Schneider, R., Godiksen, P., Ranndal, H., Madsen, H., & Bauer-Gottwein, P. (2017). Application of CryoSat-2 altimetry data for river analysis and modelling. *Hydrology and Earth System Sciences*, *21*, 751–764. https://doi.org/10.5194/hess-21-751-2017

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, *19*(10), 4345–4364. https://doi.org/10.5194/hess-19-4345-2015

Sea Level. (n.d.). ESA Climate Office. Retrieved August 10, 2022, from https://climate.esa.int/en/projects/sea-level/

Shanlong, L., Bingfang, W., Nana, Y., Fapeng, L., Meiping, W., & Jing, W. (2010). Progress in River Runoff Monitoring by Remote Sensing□□. *Advances in Earth Science*, *25*(8), 820. https://doi.org/10.11867/j.issn.1001-8166.2010.08.0820

Shi, Z., Chen, Y., Liu, Q., & Huang, C. (2020). Discharge Estimation Using Harmonized Landsat and Sentinel-2 Product: Case Studies in the Murray Darling Basin. *Remote Sensing*, *12*(17), Article 17. https://doi.org/10.3390/rs12172810

Sichangi, A. W., Wang, L., & Hu, Z. (2018). Estimation of River Discharge Solely from Remote-Sensing Derived Data: An Initial Study Over the Yangtze River. *Remote Sensing*, *10*(9), Article 9. https://doi.org/10.3390/rs10091385

Sichangi, A. W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., Zhou, J., Liu, W., & Kuria, D. (2016a). Estimating continental river basin discharges using multiple remote sensing data sets. *Remote Sensing of Environment*, *179*, 36–53. https://doi.org/10.1016/j.rse.2016.03.019

Sichangi, A. W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., Zhou, J., Liu, W., & Kuria, D. (2016b). Estimating continental river basin discharges using multiple remote sensing data sets. *Remote Sensing of Environment*, *179*, 36–53. https://doi.org/10.1016/j.rse.2016.03.019

Singh, G., Mishra, A., & Sagar, D. (n.d.). *SBIT JOURNAL OF SCIENCES AND TECHNOLOGY ISSN 2277-8764 VOL-2, ISSUE 1, 2013*. 4.

Smith, L. C., & Pavelsky, T. M. (2008). Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia: RIVER DISCHARGE AND HYDRAULIC GEOMETRY. *Water Resources Research*, *44*(3), Article 3. https://doi.org/10.1029/2007WR006133

Smith, L., & Pavelsky, T. (2008). Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia. *Water Resour. Res*, *44*. https://doi.org/10.1029/2007WR006133

Sun, W. C., Ishidaira, H., & Bastola, S. (2010). Towards improving river discharge estimation in ungauged basins: Calibration of rainfall-runoff models based on satellite observations of river flow width at basin outlet. *Hydrology and Earth System Sciences*, *14*(10), 2011–2022. https://doi.org/10.5194/hess-14-2011-2010

Tang, Q., Gao, H., Lu, H., & Lettenmaier, D. P. (2009). Remote sensing: Hydrology. *Progress in Physical Geography: Earth and Environment*, *33*(4), 490–509. https://doi.org/10.1177/0309133309346650

Tarpanelli, A., Amarnath, G., Brocca, L., Massari, C., & Moramarco, T. (2017). Discharge estimation and forecasting by MODIS and altimetry data in Niger-Benue River. *Remote Sensing of Environment*, *195*, 96–106. https://doi.org/10.1016/j.rse.2017.04.015

Tarpanelli, A., Barbetta, S., Brocca, L., & Moramarco, T. (2013). River Discharge Estimation by Using Altimetry Data and Simplified Flood Routing Modeling. *Remote Sensing*, *5*(9), 4145–4162. https://doi.org/10.3390/rs5094145

Tarpanelli, A., Brocca, L., Lacava, T., Melone, F., Moramarco, T., Faruolo, M., Pergola, N., & Tramutoli, V. (2013a). Toward the estimation of river discharge variations using MODIS data in ungauged basins. *Remote Sensing of Environment*, *136*, 47–55. https://doi.org/10.1016/j.rse.2013.04.010

Tarpanelli, A., Brocca, L., Lacava, T., Melone, F., Moramarco, T., Faruolo, M., Pergola, N., & Tramutoli, V. (2013b). Toward the estimation of river discharge variations using MODIS data in ungauged basins. *Remote Sensing of Environment*, *136*, 47–55. https://doi.org/10.1016/j.rse.2013.04.010

Tarpanelli, A., Iodice, F., Brocca, L., Restano, M., & Benveniste, J. (2020). River Flow Monitoring by Sentinel-3 OLCI and MODIS: Comparison and Combination. *Remote Sensing*, *12*(23), Article 23. https://doi.org/10.3390/rs12233867

Tarpanelli, A., Santi, E., Tourian, M. J., Filippucci, P., Amarnath, G., & Brocca, L. (2019). Daily River Discharge Estimates by Merging Satellite Optical Sensors and Radar Altimetry Through Artificial Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 329–341. https://doi.org/10.1109/TGRS.2018.2854625

Tazioli, A. (2011). Experimental methods for river discharge measurements: Comparison among tracers and current meter. *Hydrological Sciences Journal*, *56*(7), Article 7. https://doi.org/10.1080/02626667.2011.607822

Temini, M., Lacava, T., Tarendra, L., Tramutoli, V., Ghedira, H., & Riadh, A. (2011). A multi-temporal analysis of AMSR-E data for flood and discharge monitoring during the 2008 flood in Iowa. *Hydrological Processes*, *25*(16), 2623. https://doi.org/10.1002/hyp.8020

Tourian, M. J., Schwatke, C., & Sneeuw, N. (2017). River discharge estimation at daily resolution from satellite altimetry over an entire river basin. *Journal of Hydrology*, *546*, 230–247. https://doi.org/10.1016/j.jhydrol.2017.01.009

Tourian, M. J., Tarpanelli, A., Elmi, O., Qin, T., Brocca, L., Moramarco, T., & Sneeuw, N. (2016). Spatiotemporal densification of river water level time series by multimission satellite altimetry: SPATIOTEMPORAL DENSIFICATION OF ALTIMETRY OVER RIVERS. *Water Resources Research*, *52*(2), 1140–1159. https://doi.org/10.1002/2015WR017654

Van Dijk, A. I. J. M., Brakenridge, G. R., Kettner, A. J., Beck, H. E., De Groeve, T., & Schellekens, J. (2016a). River gauging at global scale using optical and passive microwave remote sensing. *Water Resources Research*, *52*(8), 6404–6418. https://doi.org/10.1002/2015WR018545

Van Dijk, A. I. J. M., Brakenridge, G. R., Kettner, A. J., Beck, H. E., De Groeve, T., & Schellekens, J. (2016b). River gauging at global scale using optical and passive microwave remote sensing. *Water Resources Research*, *52*(8), Article 8. https://doi.org/10.1002/2015WR018545

Yang, X., Pavelsky, T. M., Allen, G. H., & Donchyts, G. (2020). RivWidthCloud: An Automated Google Earth Engine Algorithm for River Width Extraction From Remotely Sensed Imagery. *IEEE Geoscience and Remote Sensing Letters*, *17*(2), 217–221. https://doi.org/10.1109/LGRS.2019.2920225

Zakharova, E. A., Kouraev, A. V., Cazenave, A., & Seyler, F. (2006). Amazon River discharge estimated from TOPEX/Poseidon altimetry. *Comptes Rendus Geoscience*, *338*(3), 188–196. https://doi.org/10.1016/j.crte.2005.10.003

Zakharova, E., Nielsen, K., Kamenev, G., & Kouraev, A. (2020). River discharge estimation from radar altimetry: Assessment of satellite performance, river

scales and methods. *Journal of Hydrology*, *583*, 124561. https://doi.org/10.1016/j.jhydrol.2020.124561

Zebin, T., Sperrin, M., Peek, N., & Casson, A. J. (2018). Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1–4. https://doi.org/10.1109/EMBC.2018.8513115

Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, *561*, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*, 83–94. https://doi.org/10.1016/j.rse.2011.10.028