Master thesis

Hydrological response Karnali Basin under different wet scenarios

Bart Bravenboer

Delft University of Technology

TU Delft

Master thesis

Hydrological response Karnali Basin under different wet scenarios

by

Bart Bravenboer

Bravenboer	Student Number
Bart	4531728

Supervisor TU Delft:	Prof Dr. T.A. Bogaard
Supervisor University Utrecht:	Prof Dr. W. Immerzeel
Supervisor TU Delft:	Dr. Ir. Gerrit Schoups
Supervisor University Utrecht:	P. Pokhrel

TU Delft

Preface

Nepal with all its pristine nature, but so vulnerable for the wreaths of climate change. To see the endurance in the eyes from the Nepali people, and the way they are adapting to the changing nature, gives you the idea that we can overcome climate change as humanity. Even the slightest contribution to understanding the phenomenon that are happening in these mountain ranges, gave me a deep sense of fulfilment. It is there that I figured out, that in the upcoming years I want to contribute to the academic field.

The work in front of you is the final product of 8 years of studying. I had the incredible opportunity to commit myself for 8 months on this inspiring topic. This thesis contributes in better understanding the drivers behind extreme discharge events in the Karnali River Basin. It emphasizes that only a slight change in temperature, will alter the timing and magnitude of extreme events. Created by the melting processes of snow accumulation on high mountain ranges.

One of professors ones told me, to keep being inspired you have to surround yourself with people that know more then you. That is when you truly grow. That is precisely how I felt the last months.

I want to thank the 'Save the Tiger, Save the Grasslands! Save the Water' project for making this thesis possible. Keep continuing all the incredible work you are doing for Nepal. My gratitude to the extremely helpful supervision I got from my committee. It was truly inspiring to work closely together with people that have so much passion for what they are doing. One of the reasons I would like to continue working in this field.

I want to thank my fieldwork buddies for our beautiful campaign in Nepal and all the hilarious moments we had. Where you really showed me the love you have for your country.

To all the friends I made in the last 8 years, I feel like a rich man and will cherish my student time for the rest of my life. I want to give special words of gratitude to all my friends living at t' Duintje. Where home really feels like home. You were there for me, every step of the way.

I want to thank my beautiful family, who taught me among many thinks, to separate essentials from shadows and dust. But most of all, for there undying support in whatever I choose to do in life.

This thesis is in memorial from my dearest grandmother, who was and will always be, a source of inspiration for me.

Bart Jan Sebastiaan Bravenboer
Delft, August 2024

Nomenclature

List of acronyms

List of acronyms in alphabetical order, including meaning in English. The page number is the in-text location where the acronym is first mentioned.

KRB Karnali River Basin 5

DEM Digital Elevation Model 4

SPHY Spatial Processes in Hydrology 5

RCP Representative Concentration pathway 7

SSP Shared Socioeconomic pathway 7

FDC Flow Duration Curves 6

ERA5 ECMWF Reanalysis 5th generation 5

PDF Probabilty density function 6

ECMWF European Centre for Medium-Range Weather Forecasts 14

MSL Mean Sea level 14

SPYDER Scientific PYthon Development Environment 15

NetCDF Network Common Data Form 21

ldd local drain direction 21

ArcSWAT2012 ArcGIS Soil and Water Assessment 2012 26

Symbols

Symbol	Definition	Unit
V	Velocity	[m/s]
Q	River discharge	[m^3/s]
T	Temperature	[$^{\circ}$C]
P	Precipitation	[mm/day]
ρ	Density	[kg/m^3]
S	Snow accumulation	km^3)

Symbol	Definition	Unit
E	Elevation	(m)
ΔT	Temperature change	[°C]
A	Area	km^2
ρ	Density	[kg/m^3]
λ	Latent heat of fusion	[J/kg]

Abstract

The Karnali River Basin, one of Nepal's major watersheds, faces significant challenges due to climate change, which is expected to intensify temperature fluctuations and alter precipitation patterns. Research has shown that snow mass in this region is highly sensitive to even minor temperature fluctuations, making it a critical factor in hydrological dynamics. This study aims to investigate how snow accumulation across different elevation ranges, combined with temperature fluctuations, influences river discharge in the Karnali Basin, with a particular focus on identifying the drivers of extreme flooding events.

Using the SPHY hydrological model, forced with high-resolution ERA5 climate data, this study modeled river discharge and, with the aid of a custom Python code based on SPHY's snow module logic, analyzed how snow is accumulating across various elevation bands within 111 delineated sub-basins in the Karnali Basin. The analysis was further refined using a RandomForestClassifier to identify key relationships between snow accumulation, temperature fluctuations, and extreme discharge events.

The study revealed that snow accumulation at elevations between 4000 and 5000 meters, combined with temperature increases 5 to 15 days prior to an event, are critical predictors of extreme discharge in the Humla Karnali and Mugu Karnali rivers. The Flow Duration Curves (FDCs) indicated a 41% increase in high-flow segments under snow-dominated scenarios, underscoring the significant role of snowmelt in driving extreme flooding events. Additionally, the machine learning model predicted with up to 96% accuracy when the discharge would exceed critical thresholds, particularly in key indicator regions within the Upper Karnali.

The findings highlight the critical importance of monitoring snow accumulation and temperature trends in the Karnali Basin, as their interplay is a significant driver of extreme hydrological events. This study contributes to a better understanding of the factors driving extreme discharge events in snow-fed river systems, providing valuable insights for improving flood prediction and water resource management in the region. Future research should focus on refining predictive models and incorporating additional climatic variables to enhance accuracy under changing climate conditions.

Contents

List of Figures

1

Introduction

1.1. Research area

The KRB, with a surface of more then 46,100 km², is located in western Nepal between 28°20' and 30°41' north latitude and 80°33' and 83°40' east longitude (Khatiwada & Pandey, 2019).

The KRB, one of Nepal's three major basins alongside the Gandaki and Koshi basins to the central and east, features the Karnali River entering from the Humla-Karnali near Khojarnath, with its primarily snow-fed tributaries (Khatiwada et al., 2016). The basin further divides into five significant watersheds: West Seti, Humla Karnali, Mugu Karnali, Tila, and Bheri, with all except the Humla Karnali, which originates in China, being sourced within Nepal (Khatiwada et al., 2016). Unlike most Nepali rivers that flow north to south, the Mugu Karnali flows east to west, and the Humla Karnali flows west to east. The Karnali River traverses western Nepal before joining the Mahakali River in India, where it continues as the Ghaghara River in India's lower reaches. (Khatiwada et al., 2016).

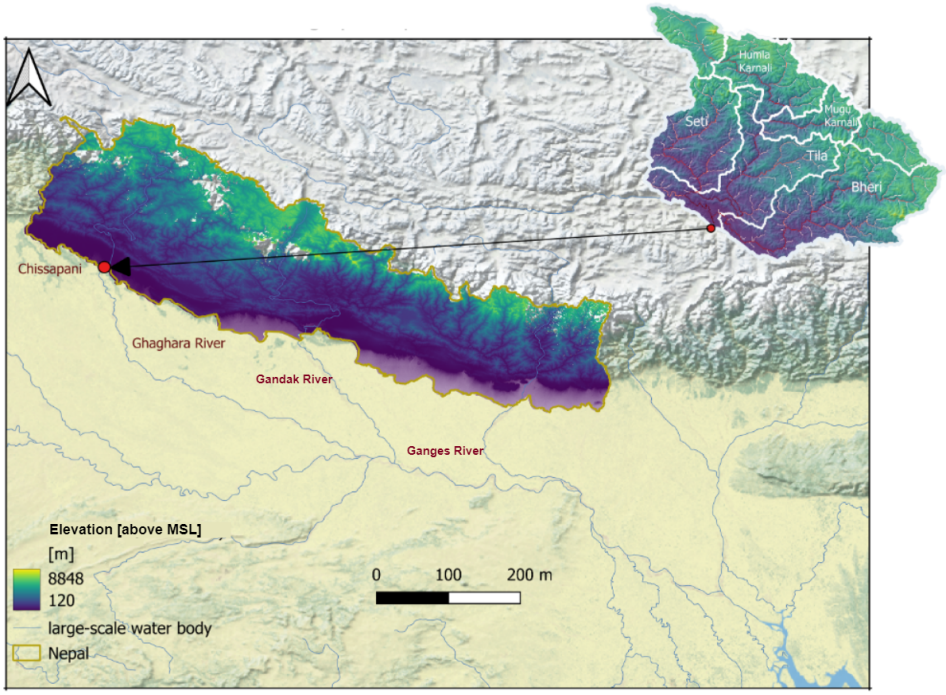


Figure 1.1: Geographical position KRB in Nepal. Illustrating the heterogeneous elevation distribution. The Karnali River Basin configures at Chisapani, crossing the border into India. Where is flows into the Ghaghara River eventually emerging into the Ganges River.

1.2. Climate

The climate of the KRB is shaped by its complex topography and the interplay between monsoonal and westerly wind systems. The towering Mahabharata range, with elevations between 1,500 and 2,700 meters, plays a critical role in disrupting the flow of monsoon winds within the basin. As these moist winds from the Indian Ocean ascend the southern slopes of the Himalayas, they cool and condense, releasing most of their moisture as precipitation on the southern flanks (Bookhagen & Burbank, 2006). This process, known as orographic precipitation, is responsible for the significant rainfall observed in the southern regions of the basin during the summer monsoon season, which lasts from June to September. During this period, Nepal receives about 80% of its annual rainfall, with an average of 1,530 mm (Shrestha, 2000).

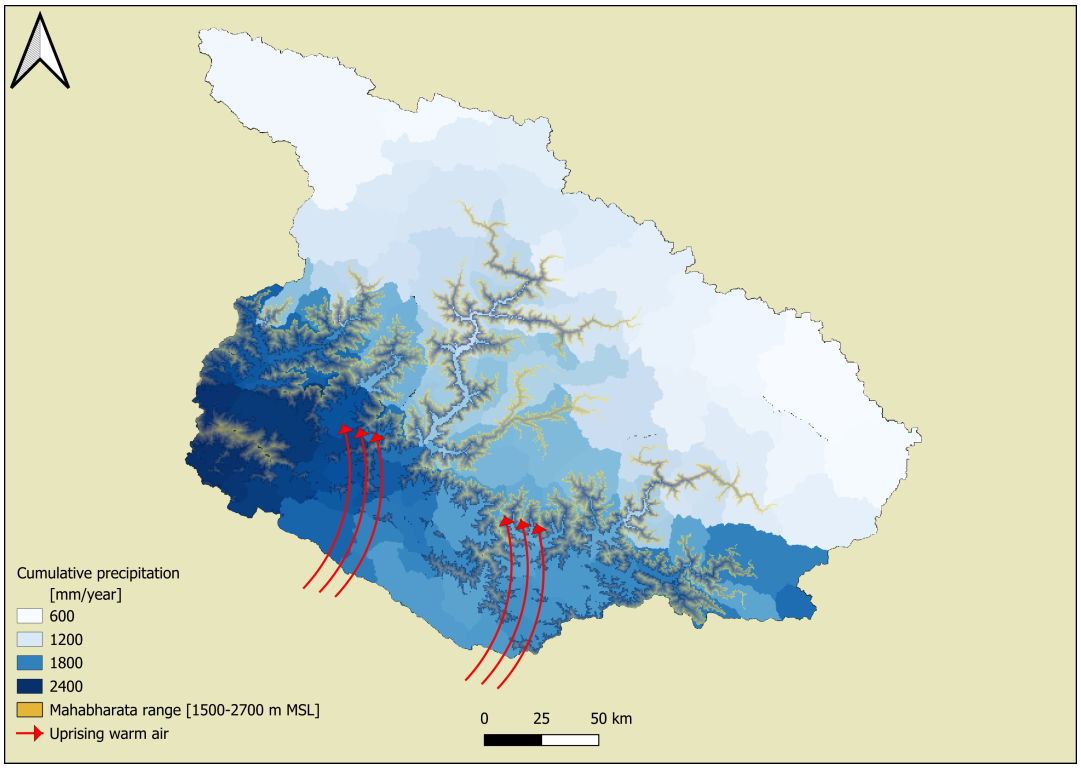


Figure 1.2: Illustrates rain shadow effect and the heterogeneous distribution of cumulative precipitation due to orthographic lifting by the Mahabharata range.

However, the impact of the orographic barrier results in a stark contrast in precipitation between the southern and northern regions of the KRB. The northern areas, situated in the rain shadow of the Himalayas, receive significantly less precipitation. As the monsoon winds lose much of their moisture before crossing the high peaks, the northern regions experience a marked reduction in rainfall, with annual totals often falling below 600 mm (Palazzi et al., 2013). This rain shadow effect, combined with the high elevation and the complex interaction between monsoonal and westerly wind systems, contributes to the arid conditions observed in these northern parts (Anders et al., 2006) (Treichler et al., 2019).

In contrast, localized pockets within the mountainous areas can receive over 2,400 mm of rainfall annually, due to the orographic relief that enhances precipitation in certain areas (Palazzi et al., 2013). Overall, river discharge within the KRB is primarily driven by summer monsoon precipitation, but is also supplemented by baseflow and the melt of snow and glaciers during the winter, when westerly winds play a more significant role in the region's climate (Dahal et al., 2020). This complex climatic pattern underscores the diverse hydrological responses observed across the basin.

1.3. Basin characteristics

Snow-fed sub-basins

Humla Karnali and Mugu Karnali

These sub-basins are primarily fed by snowmelt from the high-altitude regions of the Himalayas. The discharge in these rivers typically peaks in late spring to early summer (April to June) as temperatures rise and snowmelt accelerates. Historically, these sub-basins have exhibited relatively stable flow patterns with moderate seasonal fluctuations. However, there has been a slight increase in variability due to changes in snowmelt timing, influenced by rising temperatures over the past few decades (Budhathoki et al., 2023).

Rainfall-fed sub-basins

Bheri and Seti

These sub-basins are more directly influenced by the monsoonal rainfall, with discharge patterns characterized by pronounced seasonal variability. Peak flows occur during the monsoon season, particularly in August, when intense rainfall causes significant increases in river discharge. For instance, the Bheri River contributes about 439 m³/s and the Seti around 302 m³/s to the overall flow at Chisapani during this period . These high-flow events are often followed by much lower discharges during the dry season (winter months), with flows dropping as low as 335 m³/s in February (Lamichhane et al., 2024).

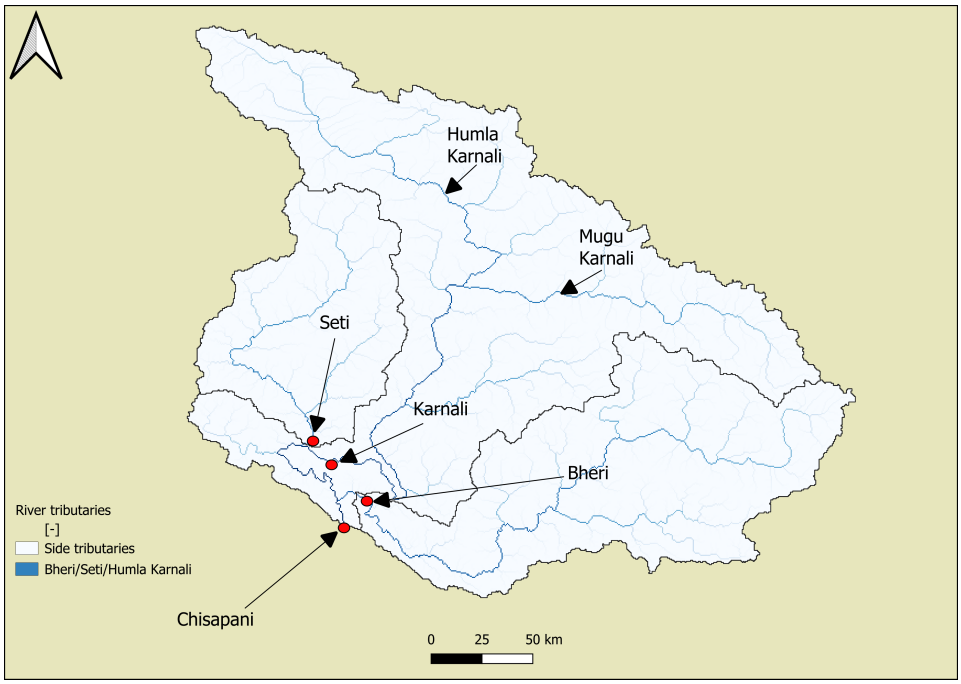


Figure 1.3: Different sub-basins find their outlet points downstream and converge together at Chisapani.

High flow and low flow patterns

High Flow (Monsoon Season)

The monsoon season (June to September) dominates the hydrology of the KRB, particularly in the rainfall-fed sub-basins. The intensity of the monsoonal rainfall can lead to rapid increases in river discharge, often resulting in peak flows that are several times higher than the baseflow observed during the dry season. The highest recorded discharge occurs in August, a direct response to the concentrated rainfall during the monsoon. This pattern is critical for water resources, agriculture, and flood management in the region.

Low Flow (Dry Season)

During the dry season (November to March), the flow in the rivers across the KRB significantly diminishes. The snow-fed rivers, such as the Humla Karnali and Mugu Karnali, maintain a relatively steady

baseflow during this period due to the gradual melting of glacial ice and snow. However, the rainfall-fed rivers like the Bheri and Seti exhibit much lower discharges during these months, reflecting the lack of significant precipitation. The low-flow periods are critical for understanding water availability for drinking, irrigation, and ecological sustainability in the region.

Observed changes over the past decades

Over the past three decades, the KRB has experienced a noticeable increase in the extremity of its flow regimes. The frequency and magnitude of high-flow events during the monsoon have risen, partly due to more intense and erratic rainfall patterns. Meanwhile, the low-flow periods have also shown increased variability, with some years experiencing significantly lower discharges during the dry season. These changes have been attributed to broader climatic shifts, including rising temperatures and changes in precipitation patterns, which have altered both the timing and volume of snowmelt and rainfall (Lamichhane et al., 2024). In summary, the KRB hydrology is characterized by marked seasonal variability, with distinct high-flow and low-flow periods influenced by the interaction of snowmelt and monsoonal rainfall.

1.4. Introduction to hydrological modeling in the KRB

Modeling the hydrological processes in the KRB is crucial for understanding and managing the water resources in this complex region. The basin's diverse topography, climate, and hydrological dynamics necessitate the use of sophisticated models that can capture the variability in water flow and storage across different sub-basins. Hydrological models used in this context can be broadly categorized into three types: lumped models, distributed models, and flood models. Each of these models is founded on different principles and is suited to specific aspects of hydrological simulation.

1. Lumped Models

Lumped models simplify the representation of a watershed by treating it as a single or aggregated set of units, without spatial differentiation within the watershed. The model inputs, such as precipitation and temperature, are averaged across the entire basin or sub-basin, and the outputs, such as runoff, are also aggregated.

- **Fundamental Principles**

  Lumped models are based on empirical relationships between inputs and outputs. They often rely on calibrated parameters to simulate hydrological processes, without detailed spatial representation.

- **Application in Tamakoshi basin, eastern Nepal**

  An example of a lumped model used in the KRB is the Snowmelt Runoff Model (SRM), which has been applied to predict snowmelt-driven runoff in the Himalayan regions, including the KRB. SRM simplifies the complex snowmelt processes into basin-wide averages, making it practical for large-scale, data-scarce environments (Budhathoki et al., 2023).

2. Flood Models

Flood models are specialized tools designed to simulate the dynamics of flood events, including the movement, extent, and timing of floodwaters. These models can be either lumped or distributed but are specifically tailored to predict flooding risks.

- **Fundamental Principles**

  Flood models incorporate detailed hydraulic simulations, often using high-resolution Digital Elevation Model (DEM) and hydrodynamic equations to predict how floodwaters will spread across the landscape. They also integrate real-time meteorological data to simulate the progression of flood events.

- **Application in KRB**

  In the KRB, the Hydrologic Engineering Center's River Analysis System (HEC-RAS) model has been widely applied for flood hazard mapping and risk assessment. The model was used to simulate flood events of varying return periods, ranging from 2-year to 1000-year floods. For

instance, a study conducted on a 38 km segment of the Karnali River downstream from Chisapani used HEC-RAS to model flood scenarios based on historical discharge data. The study found that flood depths could reach up to 23 meters at Chisapani bridge during extreme events like the 1000-year return period flood, with significant impacts on both infrastructure and agricultural lands. The simulation of the 2014 flood, which was one of the most severe in the basin's history, showed that the floodwaters inundated critical infrastructure, including schools, health facilities, roads, and an airport, highlighting the model's importance in flood risk management and emergency planning. (Aryal et al., 2020).

3. Distributed Models

Distributed models provide a detailed spatial representation of the watershed by dividing it into smaller grid cells or units, each with its own set of hydrological parameters. These models account for variations in topography, land use, soil properties, and climatic conditions across the basin.

- **Fundamental Principles**

  Distributed models are based on physically-based equations that describe the movement of water through the landscape. They integrate spatial data, such as DEM, to simulate hydrological processes at a fine scale.

- **Application in KRB**

  An example of a distributed model used in the KRB is SPHY (Spatial Processes in Hydrology) model, specifically the version calibrated by P. Pokhrel. This research is part of a larger project "Save the Tiger, Save the Grasslands, Save the Water". Recognizing the intricate link between climate change, land use, river dynamics, and tiger populations, the Save the Tiger project seeks to gain a comprehensive understanding of the complex factors affecting Himalayan grasslands. Their focus includes examining the impact of changing weather patterns, analyzing water flow within the Himalayas, investigating potential changes in river behavior due to climate change, and studying factors like groundwater availability and floodplain dynamics in the plains. By studying these combined influences on vegetation and deer populations, the project aims to identify the conditions most conducive to thriving tiger populations. This knowledge will be instrumental in developing effective conservation strategies for both tigers and their critical grassland habitat in the face of a changing climate. Additionally, this knowledge will aim to improve effective and sustainable water management in the KRB

  SPHY is particularly well-suited for the KRB due to its ability to capture the spatial variability in hydrological processes such as snowmelt, infiltration, and runoff. By dividing the basin into smaller grid cells, SPHY accounts for the differences in topography, land cover, and climatic conditions across the basin. This detailed spatial representation allows for a more nuanced and accurate simulation of the basin's hydrological dynamics, making it an ideal tool for studying complex regions like the KRB.

Assumptions and Challenges

The SPHY model is based on the principle of mass conservation, ensuring that the total amount of water entering, leaving, and being stored within a system is accurately accounted for—a concept often referred to as the water balance. The model defines dominant hydrological processes through specific equations and parameters, and it employs a subgrid variability approach to represent these processes with precision at finer spatial scales. Each grid cell can vary in glacier cover and land use, affecting key processes like interception, effective precipitation, and potential evaporation. This detailed approach allows SPHY to effectively model the complex interactions and variations in the hydrological cycle across diverse landscapes.

However, SPHY also operates under key assumptions that can impact its accuracy. One such assumption is the lumped parameter approach, which implies that parameters are uniformly distributed within each model element. While this simplifies the modeling process, it can introduce inaccuracies due to the natural heterogeneity of catchments. Additionally, the model assumes a deterministic nature of processes, presuming that physical processes can be represented in

a predictable manner given specific initial conditions. The inherent complexity and variability of real-world catchments can challenge these assumptions, potentially affecting the model's overall accuracy. (Budhathoki et al., 2023; Lamichhane et al., 2024).

Calibrated SPHY model by P.Pokhrel

The Pearson correlation coefficients across different exceedance intervals in the KRB demonstrate the strong performance of the SPHY model in simulating discharge. With correlations consistently high, particularly around a mean of 0.98, the model effectively captures the overall hydrological behavior of the basin. Even at the extreme ends, where the correlation slightly decreases, the SPHY model still shows considerable robustness. This slight dip at the highest exceedance intervals is a common challenge for distributed models and does not detract from the model's overall accuracy and reliability in predicting discharge. The results affirm that SPHY is well-suited for hydrological simulations in the KRB, with only minor adjustments needed to enhance its performance in extreme high-flow events.

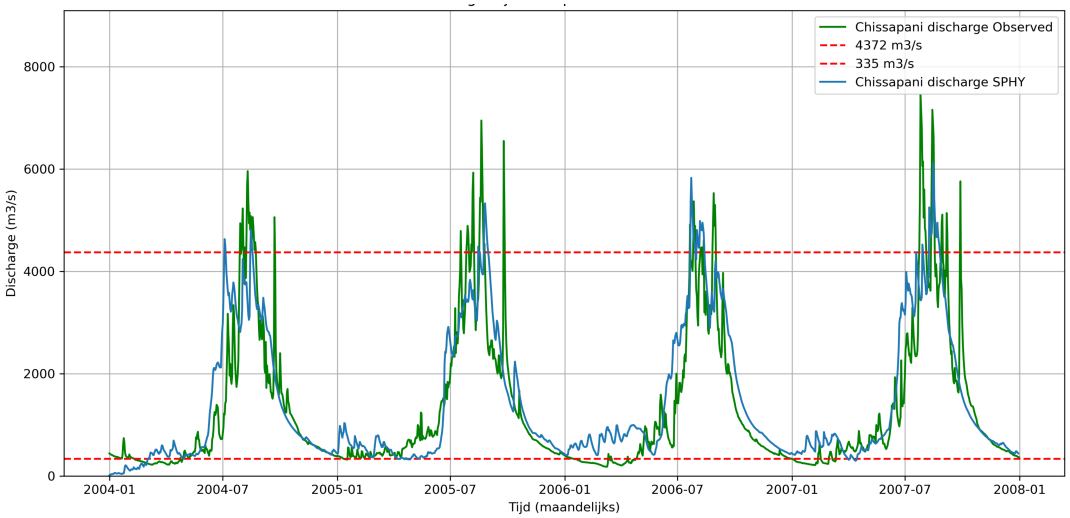


Figure 1.4: The observed discharge at Chisapani is shown in green, overlaid with the SPHY-simulated discharge for the same location.

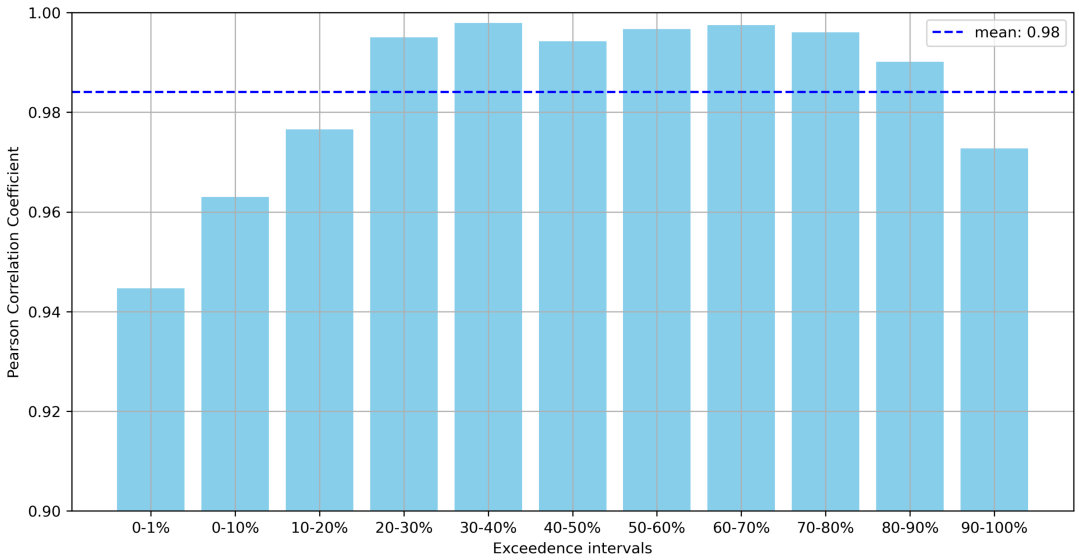


Figure 1.5: Pearson correlation coefficient comparing the SPHY-simulated discharge with the observed discharge at Chisapani for the different exceedence intervals.

2

Problem statement

2.1. Climate change and jeopardizing water resources

The Himalayan region, widely recognized as one of the most vulnerable areas on Earth to climate change, is experiencing significant warming that surpasses the global average (Bhutiyani et al., 2007). This warming is accompanied by pronounced fluctuations in precipitation (Palazzi et al., 2013), which include an increase in extreme weather events (Goswami et al., 2006) and accelerated glacial retreat (Bolch et al., 2011). These ongoing trends, observed over recent decades, have had substantial consequences for Himalayan glaciers and water resources (Immerzeel, 2010). Studies suggest that climate change will significantly impact annual water supplies, with even more pronounced effects on seasonal availability (Singh & Bengtsson, 2004). The resulting alterations in river discharge from Himalayan rivers have profound and long-term implications for water resources, which directly affect the livelihoods of both resident and downstream populations (Akhtar et al., 2008). This potential threat extends beyond water scarcity, posing significant risks to food and energy security, increasing the likelihood of natural hazards, and degrading environmental quality, ultimately impacting overall livelihoods and quality of life (Rasul, 2014).

Future climate change scenarios for the Himalayas present a complex picture of precipitation patterns, posing significant challenges for managing water resources in the region. Studies by (Eklabya P, 2009) and others highlight a concerning trend of decreasing minimum river flows, which are crucial for maintaining water availability during dry periods. This decline in base flow is further compounded by projected disruptions to established precipitation patterns. While overall summer precipitation in the Himalayas is expected to increase, this trend comes with two key features:

1. **Increased Intensity**
   The frequency of extreme precipitation events and the overall intensity of daily precipitation are projected to rise. This implies shorter bursts of heavier rainfall, which can lead to flash floods and increased erosion, while doing little to replenish water reserves effectively.
2. **Fewer rainy days**
   Paradoxically, even with the projected increase in total summer precipitation (an average increase of 0.8 to 1.2 mm per day between 2006 and 2100 under Representative Concentration pathway (RCP) 8.5), the number of rainy days is expected to decrease (approximately 8 days over the same period). This translates to longer intervals of dry weather interspersed with intense downpours, creating a more unpredictable and potentially less reliable water supply.

Recent research (Lamichhane et al., 2024) introduces critical new insights into the growing threat of increased flooding and more pronounced temperature fluctuations in the region. These findings build upon the existing understanding of climate change effects on the Himalayan region, particularly within the context of water resource management. This study also highlights a significant rise in both precipitation and temperatures under future climate scenarios. Specifically, (Lamichhane et al., 2024) projects a substantial increase in river discharge under the Shared Socioeconomic pathway (SSP)—which are scenarios used to model future climate impacts. Under SSP245, which represents a moderate climate

change scenario, extreme discharge events are expected to increase by 27.12% to 54.88%. Even more alarming is the projection under SSP585, a scenario representing more extreme climate change, where extreme discharge events could rise by 45.4% to 93.3%.

This increase in discharge directly correlates with heightened flood risks, as more intense and frequent precipitation events are likely to overwhelm river systems, leading to more frequent and severe flooding. Moreover, the study emphasizes the rapid rise in temperature, particularly minimum temperatures, which are expected to increase by 0.049°C to 0.97°C per year under SSP245, and by 0.057°C to 0.187°C per year under SSP585 across different future periods. These elevated temperature fluctuations contribute to accelerated snowmelt and glacial retreat, further exacerbating the risk of flooding, especially during the monsoon season when precipitation is already at its peak.

Given the projected increase in temperature fluctuations due to climate change, as highlighted in recent studies, it becomes crucial to understand the relationship between snow accumulation in the KRB, temperature variations, and river discharge. Recents studies (Pandey et al., 2024) underscore the significance of snowmelt as a critical component of river discharge, particularly in regions like the Himalayas where snowmelt can contribute significantly to annual water flow. In the context of the KRB, where snowmelt already plays a vital role in the hydrological cycle, the increased temperature fluctuations anticipated under future climate scenarios pose a significant risk. Large snow accumulations, combined with sudden temperature spikes, could lead to rapid snowmelt, thereby overwhelming river systems and causing extreme flooding events. This is particularly concerning during the pre-monsoon and early monsoon seasons, when snowmelt is a major contributor to river discharge. Studies by (Pandey et al., 2024) reveals that even slight changes in temperature can significantly alter the timing and volume of snowmelt, potentially leading to a mismatch between peak snowmelt and rainfall, which could exacerbate flood risks. The KRB, with its heavy reliance on snowmelt for river discharge, is especially vulnerable to such dynamics. Therefore, understanding the interplay between snow accumulation, temperature fluctuations, and river discharge is critical for predicting and mitigating flood risks in the region.

In conclusion, as climate change intensifies temperature fluctuations, the potential for spontaneous and extreme flooding due to rapid snowmelt in the KRB increases. This necessitates a deeper exploration of snow accumulation patterns, temperature variations, and their combined impact on river discharge, to better inform flood management and water resource strategies in the region.

3

Objectives and Research questions

Overarching goal

The primary objective of this study is to gain a better understanding of the interplay between snow accumulation, temperature fluctuations, and river discharge in the KRB. This research aims to investigate both the spatial and temporal dynamics of snow accumulation across various elevation ranges. Specifically, the study seeks to predict how discharge patterns will change in response to different levels of snow accumulation in specific areas of the KRB under varying temperature scenarios. By achieving this, the study will provide insights into the conditions under which snow accumulation and temperature changes lead to significant alterations in river discharge, thereby enhancing our understanding into the drivers of extreme flooding events.

Research questions

Main Research Question:
How do snow accumulation, temperature fluctuations, and their interplay influence river discharge in the KRB, particularly in understanding the drivers of extreme flooding events?

Sub-Questions:

- How can snow accumulation across different elevation ranges in the pre-monsoon period be effectively modeled and visually represented to identify snow-dominated scenarios?
- How do the FDC) of the KRB vary under different combinations of snow-dominated and rain-dominated scenarios compared to the climatological baseline?
- What is the relationship between snow accumulation at various elevation ranges, lagged temperature, and river discharge in the KRB, as identified by the RandomForestClassifier model?
- Can specific areas within the KRB be identified as indicator regions for extreme discharge events, and can the trained machine learning model generate probabilities of exceedance based on snow accumulation at certain elevations, combined with lagged temperature conditions prior to these events?

4

Methodology

This study uses a structured methodological approach to investigate the hydrological responses of the KRB under extreme weather conditions. The primary aim is to analyze how the basin reacts to these extremes, thereby gaining a deeper understanding of the key climatological parameters influencing these responses.

The study focuses on understanding the interplay between snow accumulation, lagged temperature, and extreme discharge events. After provoking these reactions in the basin, the study further explores whether predictions can be made about when critical thresholds of accumulated snow, combined with elevated temperatures, might lead to extreme discharge, thereby enhancing our understanding of these complex dynamics. Each research question is addressed through specific analytical approaches.

Research Question 1: How can snow accumulation across different elevation ranges in the pre-monsoon period be effectively modeled and visually represented to identify snow-dominated scenarios?
To answer this question, the study utilizes the ERA5 dataset, which provides high-resolution climate data across different elevation ranges in the KRB. During the pre-monsoon period (February to May), the dataset is reshuffled to create a proxy for years with high snow accumulation. This data is then processed using a custom interactive Python code, developed specifically for this study, to model and visually represent snow accumulation across various elevation ranges. The calibrated SPHY model is subsequently employed to simulate these dynamics, producing detailed visual and quantitative outputs that identify snow-dominated scenarios within the basin.

Research Question 2: How do the FDC of the KRB vary under different combinations of snow-dominated and rain-dominated scenarios compared to the climatological baseline?
This question is addressed by generating and comparing FDC on a yearly temporal scale for the KRB under different extreme weather conditions. Using the ERA5 dataset and the SPHY model, the study simulates both snow-dominated and rain-dominated scenarios. These scenarios are then used to create FDC, which graphically depict the frequency and magnitude of river discharge under various flow conditions. By comparing these curves to a climatological baseline, the study identifies how different combinations of snowmelt and rainfall influence discharge patterns, offering insights into the basin's hydrological behavior under extreme conditions.

Research Question 3: What is the relationship between snow accumulation at various elevation ranges, lagged temperature, and river discharge in the KRB, as identified by the RandomForest-Classifier model?
In this phase, the study expands its scope by dividing the KRB into 111 delineated areas, allowing for a more granular analysis of snow accumulation across all elevation ranges. The interactive Python code developed for Research Question 1 is again utilized to model snow accumulation in these newly delineated areas. This snow accumulation data, along with lagged temperature data, is analyzed using the RandomForestClassifier, a machine learning algorithm. This approach helps to understand

the relationship between snow accumulation at different elevations, preceding temperature variations, and subsequent river discharge, thereby enhancing our understanding of how these factors interact to influence hydrological outcomes.

Research Question 4: Can specific areas within the KRB be identified as indicator regions for extreme discharge events, and can the trained machine learning model generate probabilities of exceedance based on snow accumulation at certain elevations, combined with lagged temperature conditions prior to these events?
Building on the previous analysis, the study identifies key sub-basins within the 111 delineated areas that significantly contribute to extreme discharge events. The RandomForestClassifier model is applied to snow accumulation and temperature data across these areas, with a focus on identifying those regions most strongly associated with high discharge events. The model is then used to generate probabilities of exceedance, based on snow accumulation and lagged temperature conditions, thereby pinpointing specific indicator regions that can serve as early warnings for potential flooding.

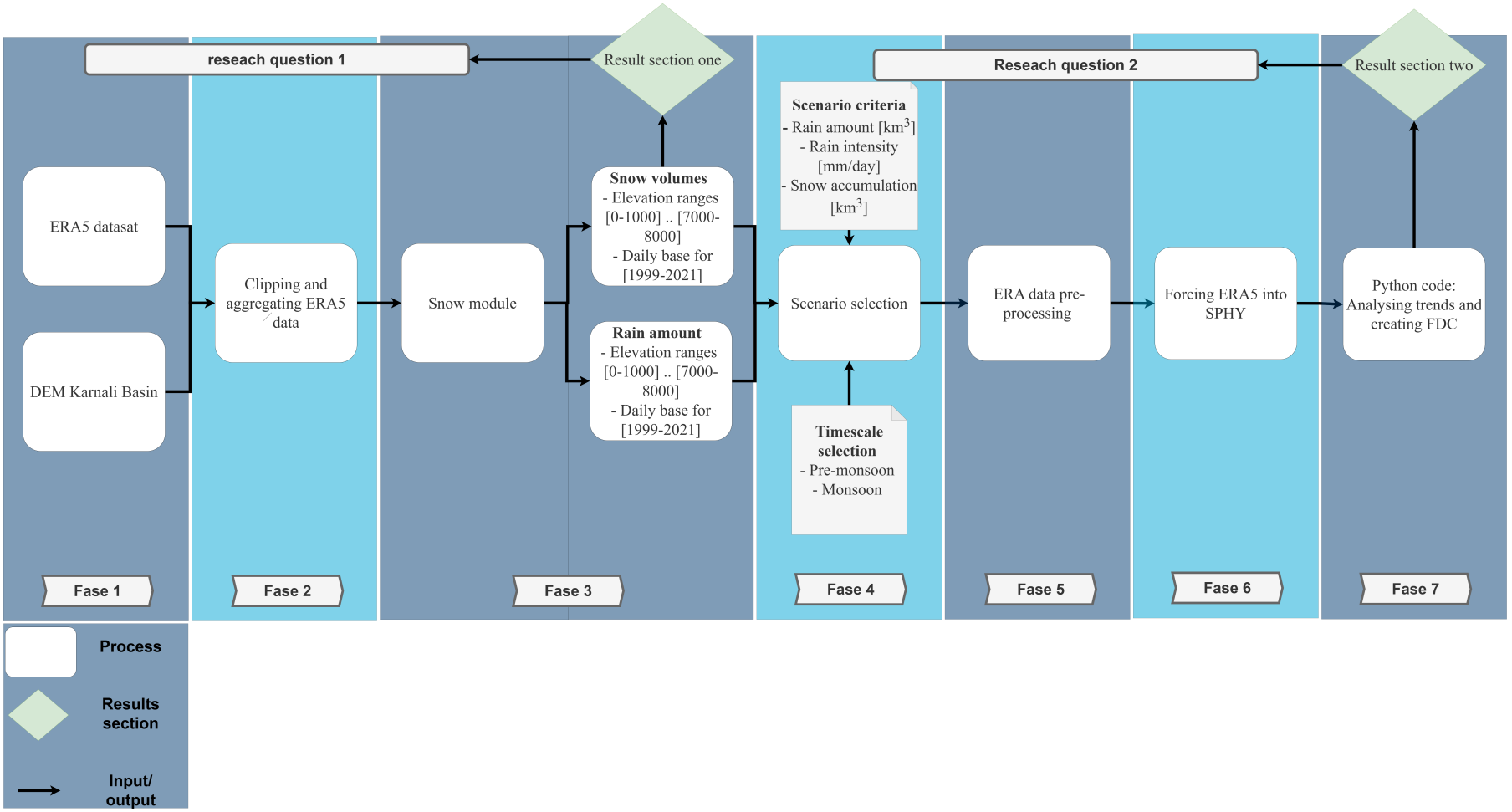


Figure 4.1: Flowchart answering research question one and two

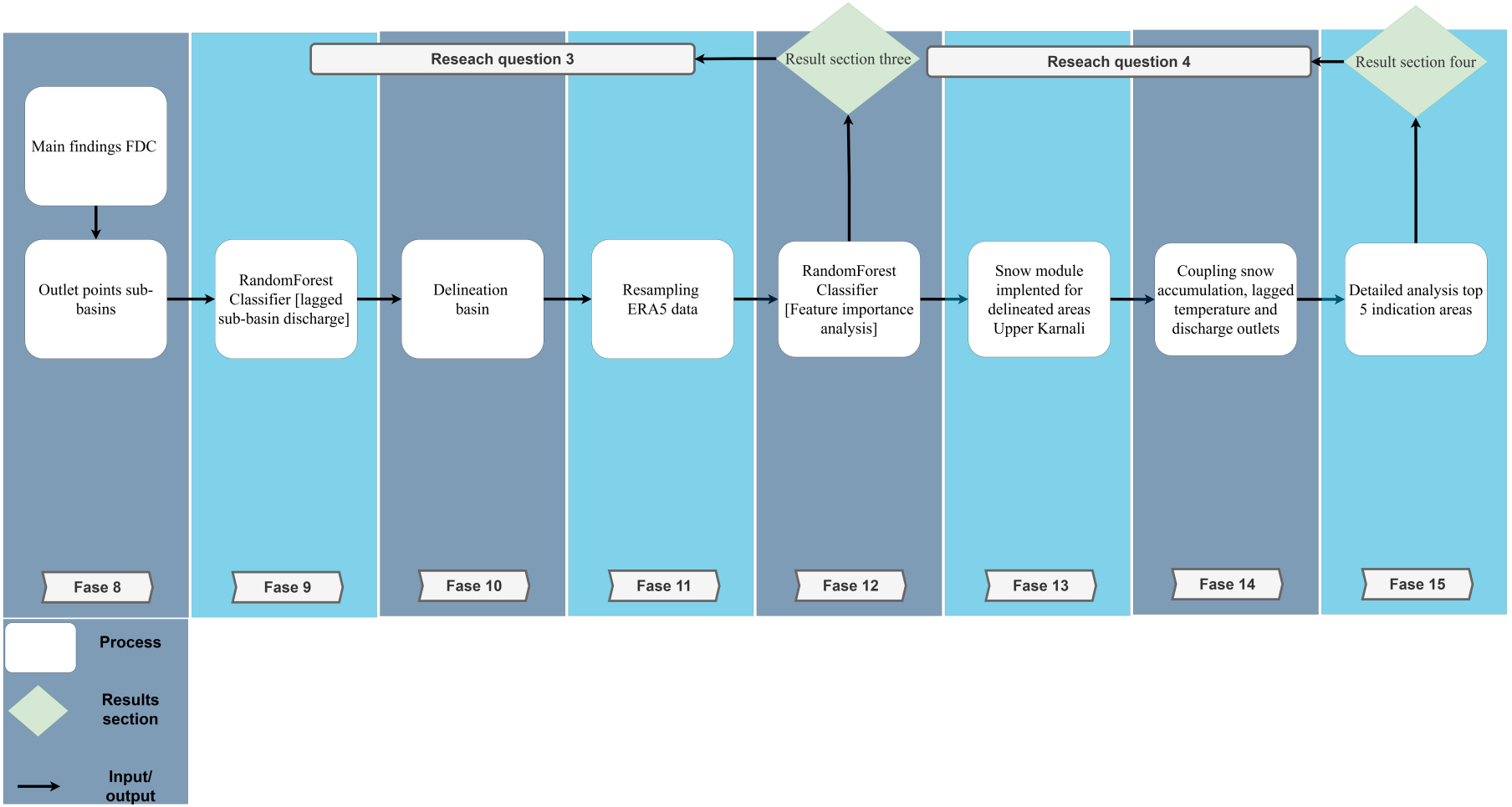


Figure 4.2: Flowchart answering research question three and four

4.1. Fase 1: ERA5 Dataset

Process

The ERA5 dataset, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides global climate and weather data at a high temporal resolution. This dataset integrates observations from various sources using advanced data assimilation systems and offers data at an hourly time step with a spatial resolution of approximately 31 km (0.25 degrees), making it highly valuable for regional and local-scale climate studies. ERA5 includes parameters such as precipitation, minimum temperature, and maximum temperature. For this study, the ERA5 data will be utilized on an averaged daily scale, with daily average temperature measured in degrees Celsius and daily average precipitation measured in mm/day. It is important to note that these are not basin-wide averages but are specific to each 31 km² cell within the basin, with the ECMWF assigning average values to each cell.

The data is available as NetCDF (Network Common Data Form) files, which are commonly used for array-oriented scientific data. To illustrate the type of data we are dealing with, figure 4.3 demonstrates the spatial distribution of precipitation and average daily temperature across the KRB for a specific date, July 6, 1991. This figure serves as an example of how the data is structured and input into the model, showing the spatial resolution and daily values assigned to each cell.

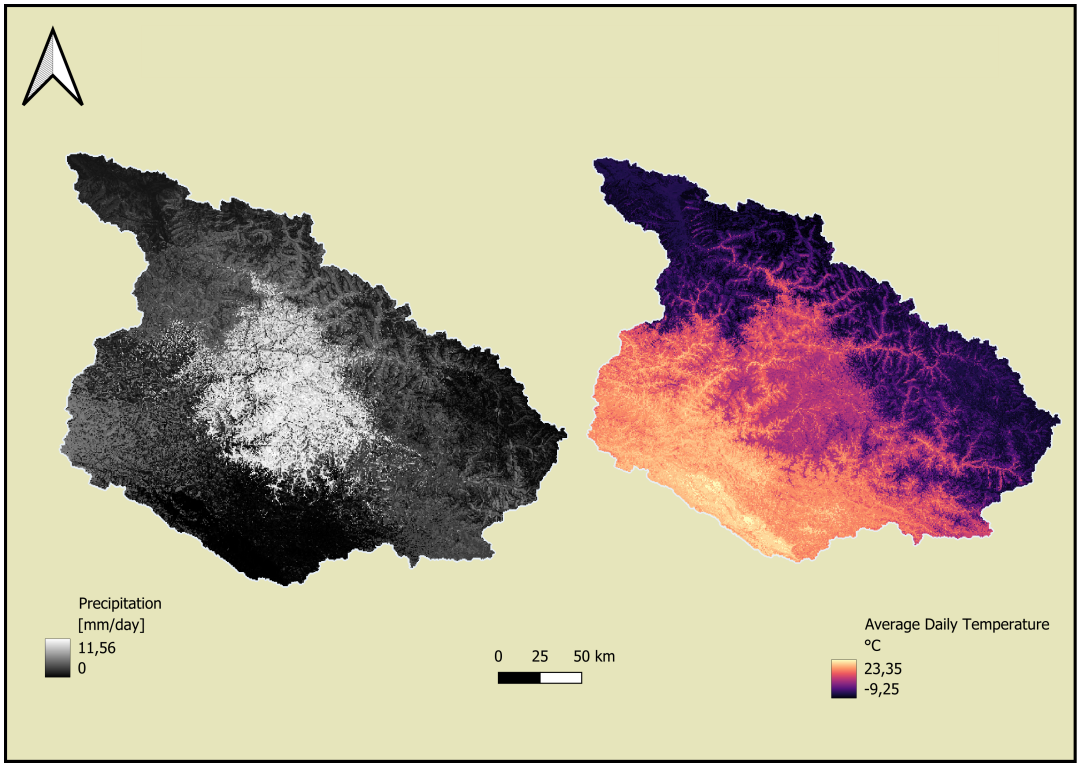


Figure 4.3: left figure: showing the spatial distribution from the ERA5 precipitation data on the 6th of July, 1991. Rigth figure: showing the spatial distribution from the ERA5 average temperature data on the 6th of July, 1991

Additionally, we use a DEM to account for the topographical variation within the basin. The DEM has an extent ranging from 454,534.611 to 767,034.611 in the x-direction and from 993,300 to 993,962 in the y-direction. The map has a width of 625 pixels and a height of 555 pixels. The data type is Float32, a thirty-two bit floating point. The map uses the PCRaster Raster File format. Band 1 of the map shows various statistics: the maximum elevation is 8,165.9419 meters, the mean elevation is 3,451.2985 meters above Mean Sea level (MSL), the minimum elevation is 132.0146 meters above MSL, and the standard deviation is 1,760.0297 meters. The valid data percentage for Band 1 is 99.66%. The pixel size is 500 by 500 meters.

This DEM is divided into elevation ranges to facilitate the analysis of climatic variables across dif-

ferent topographical features. By segmenting the basin according to these elevation ranges using the DEM Raster Calculator in QGIS, we can better understand how different elevations impact climatic and hydrological processes within the basin.

Importance of Elevation Ranges

In this study, the diverse topography of the KRB, which ranges from lowland areas to high mountainous regions, requires careful consideration of different elevation ranges. This heterogeneity significantly affects climatic conditions within the basin. Higher elevations typically experience more snowfall, impacting snow accumulation and melt processes, while lower elevations are more prone to rainfall. The temperature gradient across these elevations also influences runoff patterns and the overall hydrological response. Thus, accurately capturing these variations is essential for understanding the basin's response to different climatic scenarios. This approach is supported by studies such as those by (Viviroli et al., 2007) (Garbrecht & Schneider, 2008) The elevation ranges used in this study (0-1000, 1000-2000, ..., 7000-8000 meters) are consistent with classifications used in similar research, confirming their appropriateness for evaluating climatic and hydrological responses (Painter et al., 2010).

Output

1. A NetCDF file with a spatial resolution of 31 km², which assigns average daily values for precipitation (P) and temperature to each cell. This dataset spans 30 years (1991-2022), with P given in mm/day and temperature (T_{avg}, T_{min}, T_{max}) in degrees Celsius (°C).
2. The basin is divided into different elevation ranges based on the input DEM map, facilitating further analysis.

4.2. Fase 2: Clipping and Aggregating ERA5 Data

Input

ERA5 dataset and DEM segmented by elevation ranges

Process

The ERA5 dataset with a spatial resolution of 31 km² is processed to assign each cell to its corresponding elevation range based on the DEM segmentation. Once the cells are categorized by elevation range, the data within each range is averaged. This results in a single average value of temperature and a single average value of precipitation for each elevation range and for each day. Specifically, the process involves:

1. **Assigning Cells to Elevation Ranges**

   Each 31 km² cell from the ERA5 dataset is assigned to an elevation range: 0-1000m, 1000-2000m, 2000-3000m, 3000-4000m, 4000-5000m, 5000-6000m, 6000-7000m, 7000-8000m above MSL based on the DEM. This categorization is done by comparing the cell's location with the segmented DEM elevation ranges.

2. **Averaging Data Within Each Elevation Range**

   The process is facilitated by a custom code executed in Scientific PYthon Development Environment (SPYDER), which reads the ERA5 data and performs zonal statistics to compute mean values for each elevation range. The temperature data (T_{avg}, T_{min}, T_{max}) is similarly averaged for all cells within each elevation range. The specific code used for this process can be found in the Appendix.

3. **Producing Daily Average Values**

   The daily average values for temperature and precipitation for each elevation range are computed for the period from January 1, 1991, to December 31, 2022.

Output

The output is a set of processed datasets that include daily averaged temperature (T_{min}, T_{avg}, T_{max}) and precipitation for each elevation range. These datasets provide one average value for temperature and one for precipitation for each elevation range and each day, prepared for further analysis in subsequent steps. Additionally, the basin is divided into different elevation ranges based on the input DEM. This detailed segmentation allows for a precise analysis of the climatic variables across different topographical features.

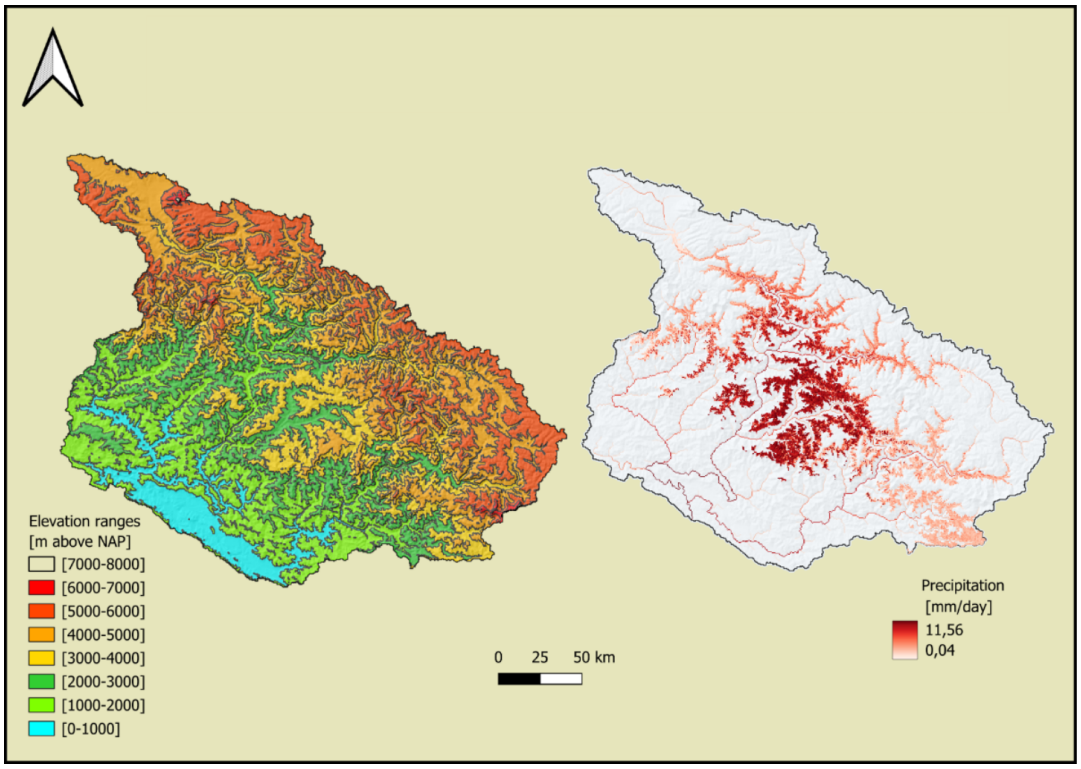


Figure 4.4: Left figure: clipped elevation ranges KRB on basis of DEM. Right figure: clipped precipitation by elevation range [4000-5000 m MSL] on 6th of July, 1991

4.3. Fase 3: Snow module

Input

The input for the snow module is derived from the previously processed datasets, which provide daily average temperature and precipitation data for each elevation range. These datasets form the basis for further analysis and modeling of snow accumulation and melt processes within the study area.

Process

The Snow Module consists of eight components: the determination of the snow/rain threshold value, precipitation solid/liquid, potential snow melt, actual snow melt, snow store update, refrozen water, total snow storage and elevation range scaling.

1. **Snow/rain threshold value**
   The snow/rain threshold value is a critical parameter used to distinguish precipitation falling as either snow or rain. This threshold is determined based on the average daily temperature. In the calibrated SPHY model for the KRB, the threshold is set at 2°C, aligning with findings in the literature (Rajagopal & Harpold, 2016). The threshold value directly influences the timing and amount of snowmelt. If the threshold is set too high, precipitation that should be classified as snow may be incorrectly classified as rain, leading to underestimation of the snowpack and pre-

mature snowmelt predictions. Conversely, a threshold set too low may result in overestimation of the snowpack and delayed snowmelt predictions (Wen et al., 2013).

Maps of precipitation and snow borders for March 18, 1993, and May 29, 1993, illustrate significant inter-seasonal temperature differences within the basin, highlighting its heterogeneity. By the monsoon season, precipitation levels increase across the basin, particularly in the southern and central regions. The changes in the snow border reflect seasonal warming and the rise in the snow/rain threshold, showing substantial differences in snow coverage between the beginning and end of the monsoon period. While these maps are snapshots in time, similar patterns are observed when comparing other years. Literature suggests this pattern is likely due to topography, as moist air masses are uplifted by the Mahabharata range (1500–2700 m.a.s.l), causing high precipitation in localized areas due to orthographic effects

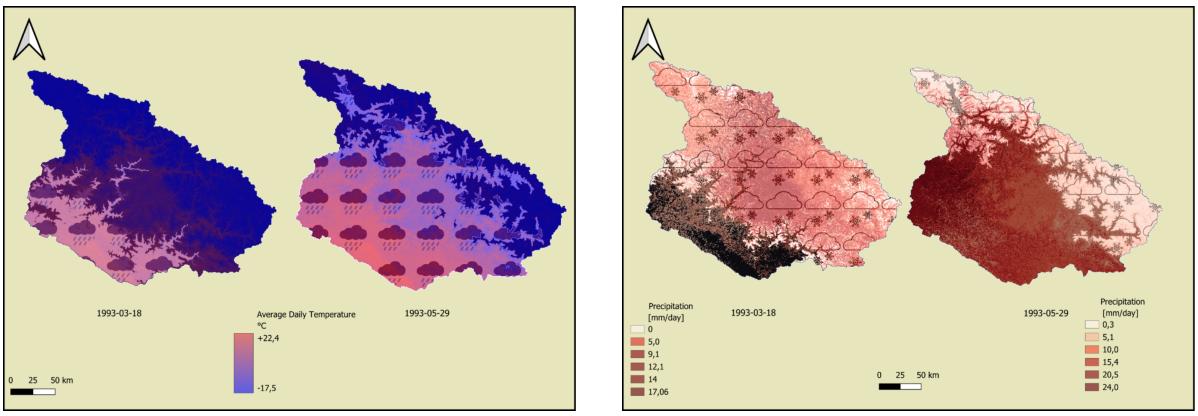


Figure 4.5: Left figure: Rain border comparison 18 th of February vs. 29th of May changing due to the average daily temperature. Right figure: spatial distribute precipitation on the 18th of February vs. 29th of May.

The module classifies precipitation as snow or rain based on the average temperature (T_{avg}). If T_{avg} is below a certain threshold (e.g., 2°C), the precipitation is classified as snow; otherwise, it is classified as rain.

2. **Precipitation solid and liquid**
   Columns for solid precipitation ($P_{s,t}$) and liquid precipitation ($P_{l,t}$) are created based on the temperature threshold.

$$P_{s,t} = \begin{cases} P_{et}, & \text{if } T_{avg,t} \leq T_{crit} \\ 0, & \text{if } T_{avg,t} > T_{crit} \end{cases}$$

$$P_{l,t} = \begin{cases} P_{et}, & \text{if } T_{avg,t} > T_{crit} \\ 0, & \text{if } T_{avg,t} \leq T_{crit} \end{cases}$$

3. **Potential snow melt (A_{pot})**
   The potential snow melt is calculated for days when the average temperature is above 0°C, using the following formula:

$$M(t) = M(t-1) \cdot e^{k \cdot (T - T_{threshold})}$$

   where:

   - $M(t)$ is the snow mass at time t (km^3),
   - $M(t-1)$ is the snow mass at the previous time step (km^3),
   - k is the degree-day factor (°C^{-1}),
   - T is the daily average temperature (°C),
   - $T_{threshold}$ is the threshold temperature (°C) above which snowmelt starts.

This formula incorporates several critical factors that influence the snowmelt process. The degree-day factor (k) correlates air temperature with the rate of snowmelt, typically expressed in mm/day per degree Celsius. Adjusting the k value allows for the calibration of the model to accurately simulate snowmelt under varying conditions. For instance, during a warm period followed by snowfall, the ground releases heat, accelerating the melting of newly fallen snow, thus requiring an increased k value. Conversely, during a cold period followed by snowfall, minimal ground heat flux results in more stable snowpack, maintaining a standard k value.

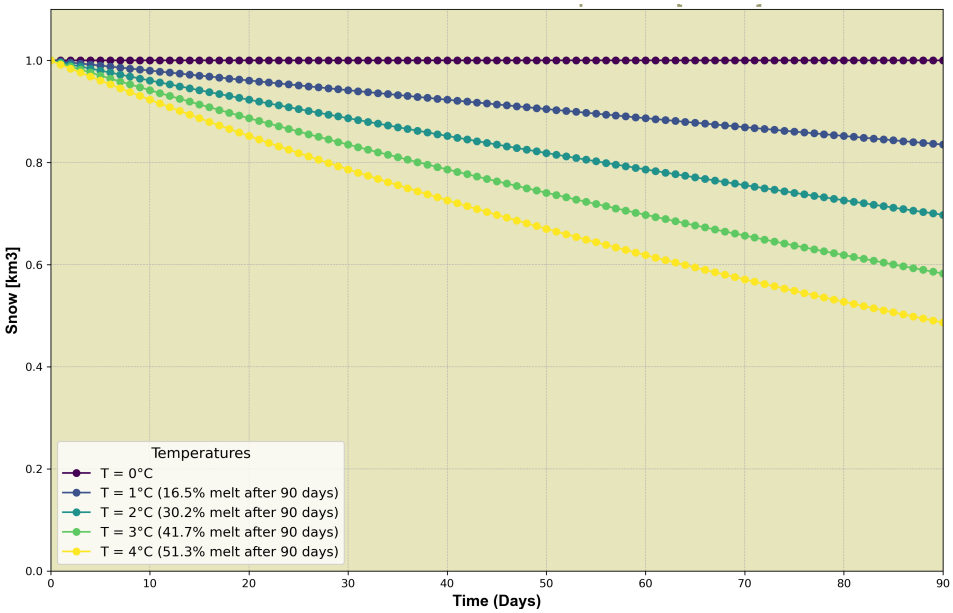


Figure 4.6: Snow melt over time at different temperatures with a sensitivity parameter [k] = 0.002

Temperature (T) is directly input into the snow melt formula, typically using a rolling mean of daily average temperatures from the moment snow falls to the end date of the calculation, which can be variably set. The threshold temperature ($T_{threshold}$) is set at 2°C, a scientifically grounded choice that effectively represents the transition between snow and rain in various climatic conditions (Braun, 1993)(Hock, 2003)

4. **Actual snow melt (A_{act})**
   The actual snow melt is computed by taking the minimum of the potential snow melt and the previous day's snow store.

$$A_{act,t} = \min(A_{pot,t}, SS_{t-1})$$

5. **Snow store (SS) update**
   The snow store is updated daily based on the solid precipitation and actual snow melt.

$$SS_t = \begin{cases} SS_{t-1} + P_{s,t} + SSW_{t-1}, & \text{if } T_{avg,t} < 0 \\ SS_{t-1} + P_{s,t} - A_{act,t}, & \text{if } T_{avg,t} \geq 0 \end{cases}$$

6. **Refrozen water (SSW) calculation**
   The maximum refrozen water is determined using a snow storage capacity coefficient (SSC), and the actual refrozen water is calculated accordingly.

$$SSW_{max,t} = SSC \times SS_t$$

$$SSW_t = \begin{cases} 0, & \text{if } T_{avg,t} < 0 \\ \min(SSW_{max,t}, SSW_{t-1} + P_{l,t} + A_{act,t}), & \text{if } T_{avg,t} \geq 0 \end{cases}$$

7. **Total snow storage (SST)**
   This is the sum of the snow store and the melt water that has refrozen within it.

$$SST_t = (SS_t + SSW_t)$$

8. **Snow accumulation (km³)**
   This scaling process, conducted in QGIS, uses vectorized elevation ranges as an overlay for the DEM. The results provide area-scaled elevation ranges, allowing for the determination of snow or rain volumes accumulating at different elevations. The majority of the surface lies within the 0-6000 meters above MSL range, with higher elevations having progressively narrower ranges due to the tapering of mountain slopes.

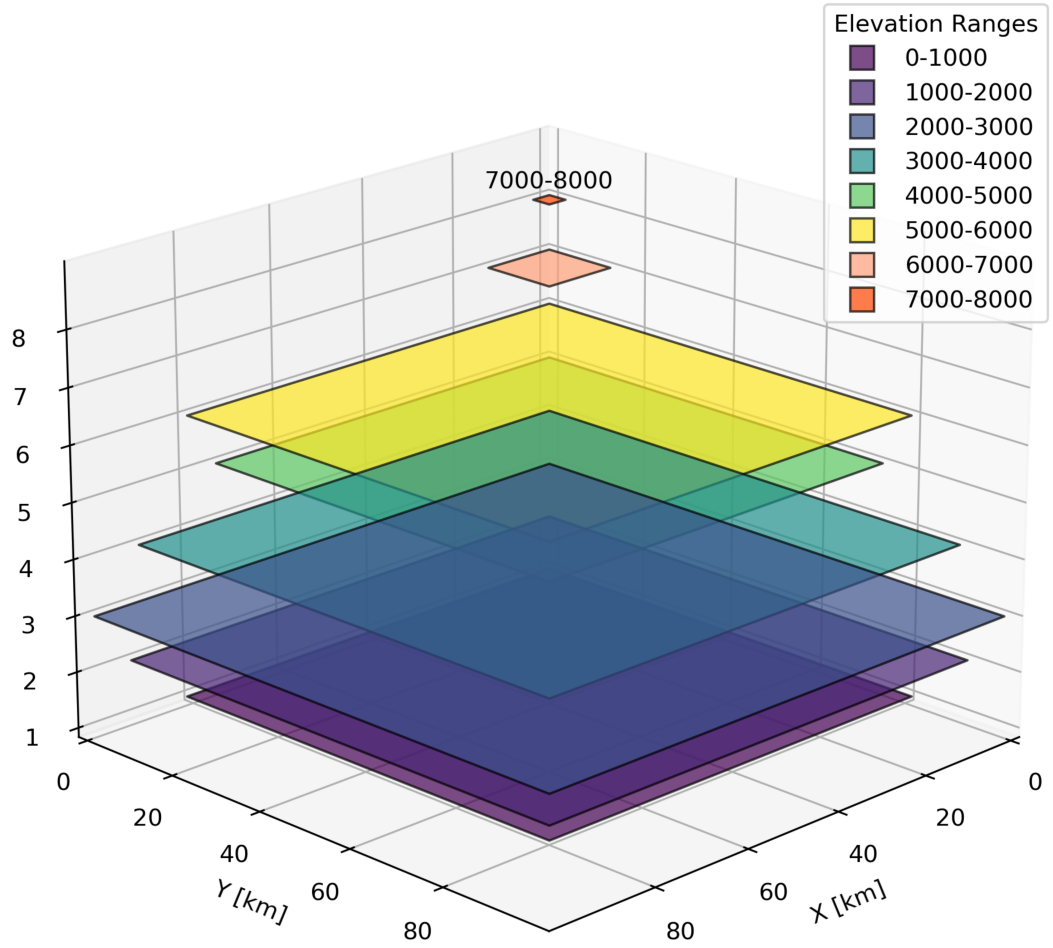


Figure 4.7: Surface areas for different elevation ranges

Finally, the snow accumulation is calculated by converting the total snow storage to cubic kilometers, adjusted for the area of the region.

$$\text{Snow\_accumulation (km}^3) = \left(\frac{SST}{1,000,000}\right) \times \text{Area}$$

Output

For each elevation range, an average temperature value for a specific day is scaled by area and input into the melt module, where the output is the initial snow mass minus the melted quantity. This process

can be applied to any start and end dates, providing flexibility in the analysis period. The output of the snow module is the snow accumulation per elevation range on a given date.

4.4. Fase 4: Scenario selection

Input

The input for this stage is a function designed to calculate snow accumulation in the KRB between two specified dates. This function leverages the ERA5 dataset and the outputs from previous steps to compute the necessary metrics, taking both dates and elevation ranges as inputs.

Process

The process involves establishing criteria for ranking the years based on three metrics: rain amount, snow accumulation, and rain intensity. A custom Python script, developed for this thesis, processes these metrics as follows:

1. **Input handling**
   The script takes two dates and elevation ranges as inputs to define the analysis period. The chosen periods are February to May for the pre-monsoon season and June to September for the monsoon season. While these periods are not rigid and can vary yearly, they are fixed in this study for comparative purposes.The chosen periods align with standard climatological seasons recognized in the region. According to research, the pre-monsoon season in India spans from March to May, and the monsoon season extends from June to September (Roxy et al., 2017). This segmentation allows for consistent comparison across years, despite natural variability in season onset and duration.

2. **Data processing**
   The script reads the ERA5 dataset, applies classification logic to distinguish between rain and snow, implements the snow module for every day in each elevation range. It visualizes these ranges, providing insights into the periods of the year when most precipitation occurs at different elevations.

3. **Calculation of metrics**

   - **Rain amount**
     Aggregates rainfall over the specified period and scales it by surface area to compute the volume in km³.
   - **Snow accumulation**
     Calculates the total snow accumulated over the period, including initial and end-date accumulations.
   - **Rain intensity**
     Computes the average daily rainfall intensity in mm/day.

Additionally, the script identifies the top 20 precipitation events to provide insights into extreme weather patterns. This information will be valuable in interpreting the FDC after integrating the data into the SPHY model.

Output

The output consists of three scenarios designed for further analysis:

- **High snow accumulation scenario**
  Identifies the top four years with the highest snow accumulation in the pre-monsoon period.
- **High rain amount scenario**
  Identifies the top three years with the highest total rain amount during the monsoon period.

- **Rain-on-snow scenario**
  Identifies two years with significant snow accumulation followed by two years with intense monsoon rainfall.

These scenarios will be used in subsequent steps to analyze their impact on runoff and water discharge in the basin, providing a comprehensive understanding of extreme events and their implications.

4.5. Fase: 5 Reprocessing ERA5 data for scenario analysis

Input

The input for this stage consists of four-year sequences derived from different scenarios. Before these sequences can be used to force the SPHY model, they need to be preprocessed to be ready for the three scenario runs.

Process

The preprocessing involves consolidating ERA5 data, which serves as the forcing data for SPHY. This includes precipitation (P) , minimum temperature (T_{min}), average temperature (T_{avg}), and maximum temperature (T_{max}). The entire dataset needs to be organized into 1460 individual Network Common Data Form (NetCDF) files, corresponding to 365 days for each of the four years, sequentially numbered from 1 to 1460. The Python code developed for this thesis outlines the following steps in the process:

1. **File Extraction**
   - The script begins by extracting the relevant files for each of the specified years from the ERA5 dataset. This is done by iterating over each row in a DataFrame that lists filenames and constructing a filename pattern based on the specified parameter (e.g. P, for precipitation).
   - The script checks if each file exists in the source directory and, if found, copies it to the destination directory. This ensures that all necessary data files are available for further processing.
2. **File renaming**
   - Once the files are extracted, they need to be renamed according to a specific pattern. This renaming is crucial for ensuring that the files are correctly recognized and processed during the SPHY model runs.
   - The script uses regular expressions to match the old filename pattern and then constructs new filenames. This involves incrementing a counter and adjusting the filename components to adhere to the required format. The new filenames are sequentially numbered from 1 to 1460.
3. **Consolidation into individual NetCDF files**
   - Each renamed file represents a single day of data. These files are stored as individual NetCDF files, ready to be used as input for the SPHY model runs.The Python code for these steps can be found in the Appendix.

Output

The output of this stage is the restructured ERA5 data, ready for use in the three different scenarios in the SPHY model.

4.6. Fase 6: Forcing ERA5 Data in the calibrated SPHY Model

Input

The input for this stage consists of the preprocessed NetCDF files from the previous stage, specifically 1460 files each for P, t_{min}, t_{avg}, and t_{max}. These files serve as the forcing data for the SPHY model. Additionally, various area characteristics need to be loaded to get the model running, including DEM, slope maps, local drain direction (ldd) map, sub-basin outlines, and stations maps.

Process

The SPHY model, developed by FutureWater and calibrated for the KRB by P. Pokhrel, is a spatially distributed, leaky bucket type model applied on a cell-by-cell basis. It does not include energy balance calculations, focusing instead on water balance to minimize complexity and runtime. This lumped conceptual hydrological model treats the entire catchment as a single unit, assuming homogeneity in the hydrological response. This simplification allows for ease of use and lower computational demand, making lumped models ideal for regions with limited data availability.

A key feature of SPHY is its modular structure, allowing users to toggle modules on or off depending on the specific requirements of the study. This adaptability makes SPHY a versatile tool for various hydrological applications. The following modules are being implemented in the calibrated model SPHY model for the KRB.

Key Modules in SPHY

1. **Snow Module**
   The snow module simulates snow accumulation and melt processes on a daily basis. It uses a degree-day approach where potential snowmelt is proportional to the temperature above a specific threshold. The model also accounts for refreezing of meltwater within the snowpack, influencing overall snow storage and runoff.

2. **Routing Module**
   The routing module calculates water flow from each cell to its downstream neighbor using a flow accumulation scheme. For cells without lakes, the accumulated flow includes contributions from the cell itself and upstream cells. If lakes are present, a fractional flow accumulation scheme adjusts the lake storage available for routing based on its actual storage.

3. **Groundwater Module**
   This module simulates water movement through different soil layers and the groundwater reservoir. It includes three compartments: the root zone, sub-soil, and groundwater store. Water balance equations simulate interactions among these compartments, including surface runoff, lateral flow, percolation, and base flow.

Area characteristics Required

To run the SPHY model, several static input maps and parameters need to be specified:

- **DEM**
  Represents the elevation of each cell and is crucial for generating other maps like slope and ldd.
- **Slope map**
  Created from the DEM to determine the steepness of each cell.
- **ldd** map
  Indicates the flow direction from each cell to its steepest downslope neighbor, essential for routing.
- **Sub-basins map**
  Divides the basin into smaller areas for detailed hydrological analysis.
- **Stations map**
  Identifies locations for reporting time-series data of various model fluxes. For this fase the coordinates from Chisapani will be used.
- **Climate variable forcing data**
  Include $precipitation$, t_{avg}, t_{min}, and t_{max} NetCDF files.
- **Soil and land use maps**
  Define the soil hydraulic properties and land use characteristics affecting infiltration and evapotranspiration.

SPHY reports back on a wide range of parameters, including those related to the snow module. However, it does not provide snow accumulation data per elevation range, which is essential for our in-depth analysis of the correlation between snow accumulation at specific ranges and extreme discharge

events. Therefore, we developed our own code to calculate these metrics.

The logic of the snow module in SPHY is consistent with the logic implemented in our custom code to maintain coherence. This ensures that we can draw clear correlations between snow accumulation in our model and discharge data from the SPHY model.

Output

The output of this stage is the calculated discharge at the Chisapani outlet point, derived from the processed ERA5 data using the SPHY model. This discharge data will be crucial for analyzing the hydrological response of the basin under different scenarios.

4.7. Fase 7: Analyzing Trends and Creating FDC

Input

The input for this stage is the discharge data output from the previous chapter, specifically at the Chisapani outlet point.

Process

A FDC is a graphical representation of stream flow variability within a specific period at a given location. It plots the percentage of time that discharge values are equaled or exceeded, offering insights into the flow regime's variability and the frequency of different flow magnitudes. FDC are widely used in hydrology for water resource management, environmental flow assessments, hydro power potential estimation, and flood and drought analysis. They provide a comprehensive overview of the river flow characteristics, allowing for effective planning and management of water resources (Ridolfi et al., 2020) (Smakhtin, 2001).

The custom Python code, developed specifically for this purpose, performs several steps to generate an FDC and analyze its characteristics for a specified year.

1. **Validation and data loading**
   The function `FDC_plot(year)` begins by verifying that the input year is an integer. It then opens an Excel file containing discharge data and checks if the specified year is present as a sheet. The discharge data for the specified year is read from the Excel sheet.
2. **Data sorting and preparation**
   The discharge data is sorted in descending order, and the index is reset to ensure a sequential index. The total number of data points is calculated, and a new column for percentage exceedence is created, representing the percentage of time a particular discharge is exceeded.
3. **Discharge binning and frequency analysis**
   The discharge data is divided into bins, and the frequency of discharge values within each bin is calculated. The bin with the highest frequency is identified, and the minimum, maximum, and average discharge values for this bin are computed. The cumulative sum of counts for each bin is calculated to determine the exceedence probability for the minimum and maximum discharge values.
4. **Statistical measures**
   The mean, median, and mode of the discharge data are calculated. Indices corresponding to these values are identified to extract the percentage exceedence values for each.
5. **Plotting the FDC**
   The FDC is plotted, showing the relationship between discharge and percentage exceedence. Statistical measures such as mean, median, and mode are represented as lines on the plot. Shaded areas highlight specific exceedence ranges (0-1% and 1-10%), providing a visual representation of the variability in discharge.

   On the x-axis, the FDCshows the percentage of time that a specific discharge value is exceeded. This gives an idea of the flow duration or frequency of different flow magnitudes. On the y-axis, it displays the discharge values (in cubic meters per second), indicating the flow rate at different exceedance probabilities.

Output
The output is a flexible Python function capable of generating FDC for any given year within the 30-year period, along with key statistical characteristics such as mean, median, and mode discharges. This allows for a comprehensive analysis of discharge trends and their variability over time.

4.8. Fase 8: Outlet Points

Input
FDC

Process
Building on the insights gained from the analysis of the FDC, the next step involves identifying extreme discharge events at key outlet points within the basin. Outlet points are positioned at the end of sub-basins, including Upper Karnali, Seti, and Bheri. Additionally, we maintain the outlet point at Chisapani to understand the timing and magnitude of discharges from various sub-basins in correlation with extreme events at Chisapani. We are particularly interested in understanding which tributaries contribute significantly to these high discharge rates, the timing of these discharges, and whether the various sub-basins exhibit synchronous behavior with extreme events observed at Chisapani.

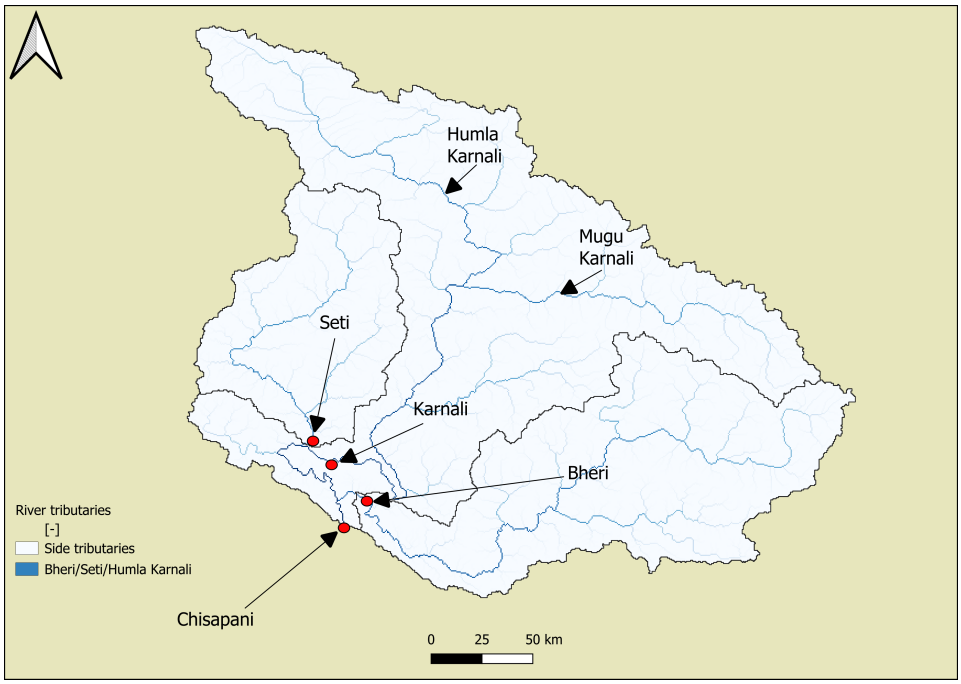


Figure 4.8: Outlet points Seti, Bheri, Upper Karnali and Chisapani

By rerunning the scenarios with this integrated outlet map, we can focus on the behavior of the different outlet points at the end of the sub-basins under various conditions. This analysis will help determine the specific contributions of each sub-basin to the overall discharge during extreme events.

Output
Discharge data for sub-basins (Upper Karnali, Bheri, Seti, Tila)

4.9. Fase 9: RandomForestClassifier

Input

Discharge data from outlet points

Process

The next step in the methodology involves implementing a RandomForestClassifier to analyze the lagged discharge data from various sub-basins. RandomForestClassifier is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks (Cutler et al., 2012). This approach begins by generating a multitude of decision trees, each trained on a different subset of the dataset. These subsets are created through a process called bootstrapping, where random samples are taken from the original data, allowing some data points to be repeated while others are omitted. This variation ensures that each decision tree sees a slightly different view of the data, promoting diversity among the trees.

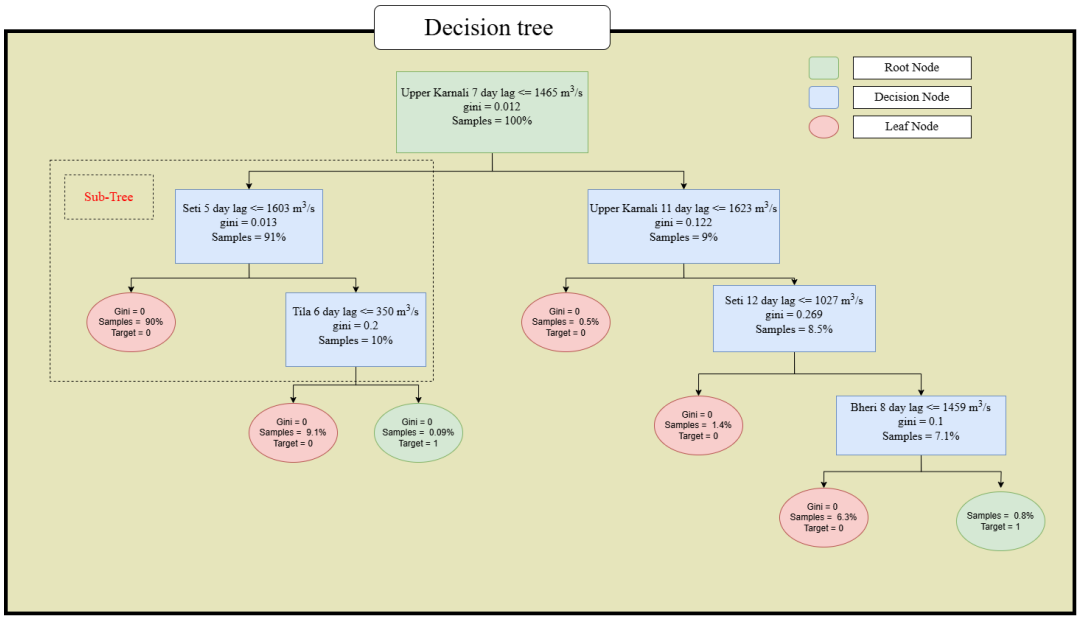


Figure 4.9: Example from a decision tree generated by the RandomForestClassifier machine learning model. Building a path from the root to the leaves that represents the decision-making process of the model.

In a decision tree, there are three main components: Root Node, Decision Node, and Leaf Node.

1. **Root Node**
   The root node is the topmost node in a decision tree. It represents the first feature that is used to split the data. In this tree, the root node is ”Upper Karnali 7 day lag <= 1465 m³/s.” This node splits the entire dataset based on whether the Upper Karnali 7-day lagged discharge is less than or equal to 1465 m³/s.

2. **Decision Node**
   Decision nodes are points where the data is further split based on different features. These nodes represent decisions or tests on the features that divide the data into smaller subsets. In the diagram, decision nodes are shown in blue. For example, ”Seti 5 day lag <= 1603 m³/s” is a decision node that splits the subset of data where the Upper Karnali 7-day lag is less than or equal to 1465 m³/s.

3. **Leaf Node**
   Leaf nodes, shown in red, are the endpoints of the tree where no further splitting occurs. These

nodes represent the final decision or outcome based on the features. Each leaf node provides a classification or a target value. For example, the leaf node at the bottom right with "Gini = 0, Samples = 0.9%, Target = 1" indicates that for the given path through the tree, the outcome is predicted with a certain level of purity (Gini impurity of 0).

These nodes work together to classify or predict outcomes based on the input features, building a path from the root to the leaves that represents the decision-making process of the model.

Each decision tree in the forest independently makes a prediction based on the input data. In a classification context, these trees will classify the input into one of the possible categories. The RandomForestClassifier aggregates these individual predictions to make a final decision. It does this by taking a majority vote among all the trees: the class that receives the most votes from the individual trees becomes the final output of the forest. This voting mechanism enhances the robustness and accuracy of the model by mitigating the risk of overfitting, which can occur when relying on a single decision tree that might capture noise or anomalies in the data.

Moreover, RandomForestClassifier also provides a measure of feature importance. During the training process, it evaluates the contribution of each feature to the model predictive power by analyzing how the inclusion or exclusion of specific features affects the accuracy of the trees. Features that consistently improve the decision trees' accuracy are deemed more important. This ability to rank features by importance is valuable for understanding the underlying data patterns and for making informed decisions about which features are most relevant for the classification task.

Random forests are also effective in handling large datasets with higher dimensionality. According to (Biau & Scornet, 2016), the method ensemble nature allows it to efficiently manage and process large volumes of data while maintaining high predictive accuracy. The algorithm's robustness and flexibility make it suitable for various applications, from image classification to medical diagnosis. Furthermore, studies (Strobl et al., 2007) highlight that random forests can handle missing values and maintain performance with imbalanced datasets, which are common challenges in real-world data scenarios. These attributes underscore the versatility and reliability of RandomForestClassifier in diverse fields of study.

In this context, we will use the RandomForestClassifier to determine how the discharge data from different sub-basins 1, 5, and 10 days prior to an extreme event correlate with discharge events exceeding 8000 m^3/s at Chisapani. This analysis will help us understand the contribution of each sub-basin to extreme discharge events and whether there are synchronous patterns in their behavior leading up to significant events.

Output

A detailed understanding of the lagged discharge from the various sub-basins in relation to extreme events at Chisapani. This will provide insights into how the sub-basins behave in the days leading up to significant discharge events, helping to clarify the dynamics of high discharge occurrences in the KRB. Additionally, we will attempt to identify threshold data for 1, 5, and 10 days prior to the event and their standard deviations. This will provide insight into the accuracy of the prediction and how well these thresholds can function as indicators. A lower standard deviation indicates a better threshold value because it reflects less variability and higher reliability in predicting extreme discharge events, making it a more consistent and dependable predictor.

4.10. Fase 10: Delineation Basin

Process

Based on the work of Pandey (2020), the KRB was delineated into 111 sub-basins using ArcGIS Soil and Water Assessment 2012 (ArcSWAT2012). A threshold area of 3000 hectares was defined to generate the river network. The variation in sub-basin sizes, ranging from 44 to 3182 square kilometers, arises due to several factors.

- **Topographical Features**
  The natural topography of the basin, including mountains, valleys, and plains, influences how the land is divided. Steeper areas may require smaller sub-basins to accurately capture the rapid changes in elevation and flow direction (Zomer et al., 2008).

- **Hydrological Connectivity**
  Sub-basins are delineated to ensure hydrological connectivity, meaning that they are defined by natural drainage patterns. This ensures that each sub-basin drains into a specific point in the river network, maintaining the integrity of the watershed. (Kennard et al., 2007).

- **Threshold Area for Stream Delineation**
  The threshold area, set at 3000 hectares, determines the minimum drainage area required to form a stream. This value is used to initiate the river network and influences the number and size of sub-basins. Areas with higher drainage density may have smaller sub-basins, while flatter regions with less drainage density might have larger sub-basins (Kennard et al., 2007).

- **Spatial Heterogeneity**
  To accurately model the diverse hydrological processes within the basin, the delineation process accounts for spatial heterogeneity. This means that regions with varying land use, soil types, and climatic conditions are divided into appropriately sized sub-basins to ensure that each area's unique characteristics are adequately represented in the model (Zomer et al., 2008).

After delineating the 111 sub-basins, the DEM Raster Calculator in QGIS, integrated with a Python code, was used to further subdivide these areas into different elevation ranges. The same tool was also employed to scale the areas according to their specific elevations, ensuring precise categorization and assignment of areas. This processed and stored data provides detailed information for subsequent analysis.

Output

111 delineated areas with distinct elevation bands. Each area is subdivided into elevation ranges and scaled according to the surface area present at each elevation. Given the substantial elevation heterogeneity within the basin, each delineated area has been further divided into eight elevation ranges: 0-1000 meters, 1000-2000 m, 2000-3000 m, 3000-4000 m, 4000-5000 m, 5000-6000 m, 6000-7000 m, and 7000-8000 meters above MSL. This ensures that the data is ready for the next step. Figure 1 shows one of these delineated areas in detail.

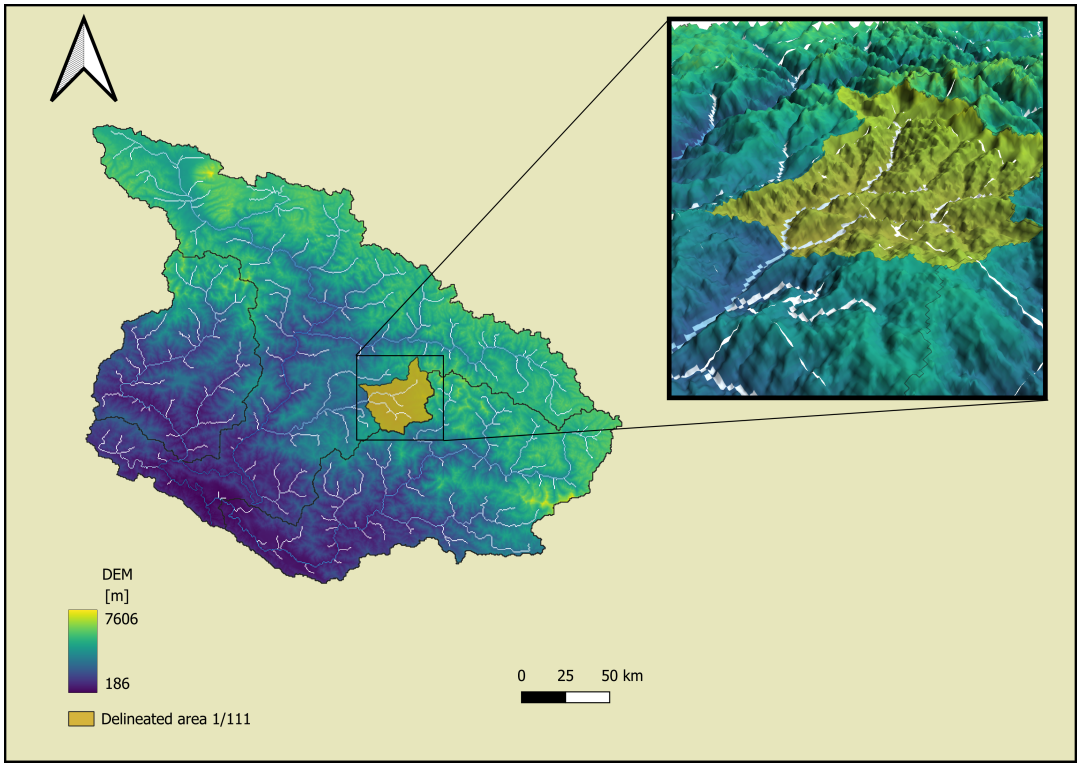


Figure 4.10: Showing one of the 111 delineated areas within the KRB

4.11. Fase 11: Resampling ERA5 Data

Input

111 delineated areas

Process

The next step in the methodology involves resampling ERA5 data to generate cumulative rainfall and lagged temperature data for different elevation bands. This process is crucial for identifying indicators of extreme discharge events within the KRB. The use of cumulative rainfall and lagged temperature as predictors is supported by extensive research in hydrology and climatology.

Justification for Cumulative Rainfall

- **Hydrological responses to rainfall**

  Research has shown that cumulative rainfall is a critical factor in triggering extreme discharge events. The total amount of rainfall over a period significantly influences the soil moisture and saturation levels, which in turn affect the runoff and river discharge. Discussed by literature (Smith et al., 2015) is the concept of contributing areas and how rainfall accumulation affects runoff and discharge. By incorporating cumulative rainfall data, we can better understand and predict peak discharge events.

- **Event-based rainfall-runoff modeling**

  Studies have utilized event-based models to predict peak discharge based on cumulative rainfall. These models take into account the temporal distribution of rainfall and its accumulation to forecast river discharge during extreme weather events. Literature (Smith et al., 2015) provides an overview of rainfall-runoff relationships and the impact of cumulative rainfall on extreme discharge.

Justification for lagged temperature

- **Temperature influence on snowmelt and runoff**

In regions where snowmelt significantly contributes to river discharge, lagged temperature (temperatures from previous days or weeks) plays an important role in determining the timing and magnitude of snowmelt. This, in turn, affects river discharge. Studies (Kang & Lee, 2014) highlight the importance of temperature patterns preceding sno wmelt events and their effect on runoff.

- **Thermal Dynamics and River Discharge**

  Temperature variations influence evapotranspiration rates, soil moisture dynamics, and subsequently, river discharge. Understanding the lagged effects of temperature helps in predicting the hydrological response of a basin to changing thermal conditions. Literature (Viviroli & Weingartner, 2004) discusses the impact of temperature on hydrological processes in alpine regions, providing insights into how lagged temperature data can be used as a predictor for extreme discharge events.

  By resampling ERA5 data, we can generate cumulative rainfall and lagged temperature data for the 111 delineated areas, categorized by different elevation bands. This detailed dataset enables us to analyze the spatial and temporal patterns in these indicators, enhancing our ability to predict extreme discharge events using the RandomForestClassifier.

Temperature variations influence evapotranspiration rates, soil moisture dynamics, and subsequently, river discharge. Understanding the lagged effects of temperature helps in predicting the hydrological response of a basin to changing thermal conditions. The impact of temperature on hydrological processes in alpine regions, particularly how lagged temperature data can be used as a predictor for extreme discharge events, has been discussed extensively in the literature (D. Viviroli Weingartner, 2004).

By resampling ERA5 data, we can generate cumulative rainfall and lagged temperature data for the 111 delineated areas, categorized by different elevation bands. This detailed dataset enables us to analyze the spatial and temporal patterns in these indicators, enhancing our ability to predict extreme discharge events using the RandomForestClassifier. Specifically, we examine lagged temperature and cumulative precipitation over varying time windows, ranging from 1 to 20 days prior to each extreme discharge event. The selection of this range is based on the understanding that hydrological responses to temperature and precipitation can vary significantly depending on the duration of these antecedent conditions. For instance, shorter lag times (1-5 days) may capture immediate responses such as snowmelt or rapid runoff, while longer lag times (15-20 days) are critical for identifying slower processes like soil saturation and gradual snowmelt contributions. By incorporating this range of temporal scales, we can extract and examine both spatial and temporal patterns, allowing us to identify critical thresholds and relationships that may act as indicators of impending extreme discharge events. These three elements—cumulative precipitation over various windows, lagged temperature over the same windows, and the extreme discharge events at the Chisapani bridge—are then integrated into the RandomForestClassifier model. This model will be used to uncover key predictors and their interactions, ultimately enhancing our ability to forecast extreme discharge events within the Karnali River Basin.

Output

Cumulative rainfall and lagged temperature data for 111 areas on a daily scale. This dataset will serve as the input for subsequent analyses, allowing us to identify correlations between these indicators and extreme discharge events.

4.12. Fase 12: RandomForestClassifier (Feature Importance Analysis)

Input

Cumulative rainfall and lagged temperature

Process
The next step involves using the RandomForestClassifier to analyze cumulative rainfall and lagged temperature data for the 111 delineated areas within the KRB. This analysis focuses on determining the significance of these two predictors in forecasting extreme discharge events in each region.

Output
In a hypothetical analysis, the RandomForestClassifier is expected to identify lagged temperature, particularly the average temperature (t_{avg}), as a potentially significant feature for predicting extreme discharge events in the Upper Karnali region. This is anticipated because the Upper Karnali is predominantly snow-fed. Conversely, for the Bheri and Seti regions, which are primarily rain-fed, cumulative precipitation is likely to emerge as a key feature.

By establishing this hypothetical distinction, the methodology can proceed with a focused approach:

- For the Upper Karnali, subsequent analysis will hypothetically concentrate on lagged temperature data in combination with snow accumulation.
- For the Bheri and Seti regions, the analysis will likely emphasize cumulative rainfall data.

This proposed bifurcation in the methodology aims to enable tailored predictive modeling, which should hypothetically enhance the accuracy and relevance of predictions for each specific region.

4.13. Fase 13: Snow Module SPHY Logic implemented for delineated areas

Input
Lagged temperature data for Upper Karnali

Process
The entire process described in fase 3 will be repeated for the 48 delineated areas within the Upper Karnali. The process is performed in a Python code developed for this master thesis.

Output
Snow accumulation data for 47 regions in Upper Karnali across different elevation ranges on a daily scale for the snow scenario making use from the logic described in the SPHY modules.

4.14. Fase 14: Upper Karnali: Coupling Snow Accumulation, lagged temperature and Discharge Outlets

Input
Snow accumulation data and discharge data from SPHY (at the different outlet points) and lagged temperature for the individual delineated areas

Process
The next step involves analyzing the relationship between snow accumulation, lagged temperature, and discharge across different elevation ranges in the Upper KRB. The snow accumulation data for 47 regions in Upper Karnali, categorized by different elevation ranges and scenarios (rain/snow/rain-on-snow), will be used. This data is coupled with discharge data from the 47 outlet points generated in the SPHY model outlet map.

Each of these outlet points is strategically placed at the end of each sub-basin, providing a comprehensive overview of the discharge patterns. The discharge data is generated by forcing the ERA5

snow-scenario into the calibrated SPHY model for the KRB. This process aims to find the best correlation between snow accumulation, lagged temperature, and discharge at the sub-basin level. The coupling process in the RandomForestClassifier involves correlating the snow accumulation and lagged temperature at each elevation range within each region with the discharge generated at the corresponding outlet point.

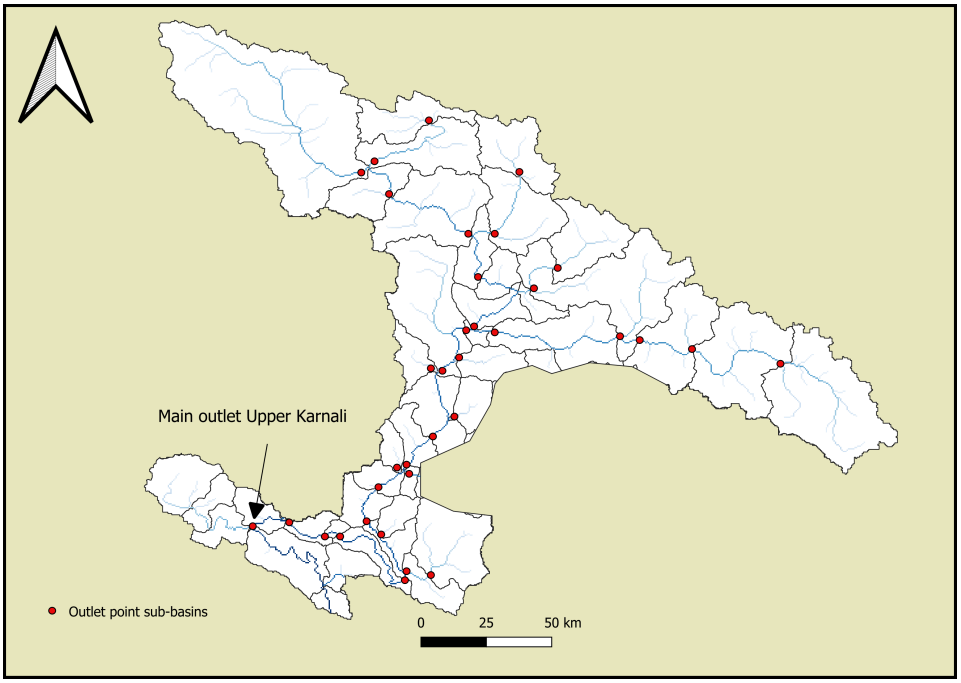


Figure 4.11: Different outlet points from the 47 delineated areas within the Upper Karnali. To calculate their individual contribution to the extreme event at the main outlet downstream.

The objective is to identify which combinations of snow accumulation, lagged temperature, and discharge patterns are most predictive of the overall discharge observed at the main outlet point downstream in the Upper Karnali. This analysis aims to find the best predictors for discharge events at the main outlet point by investigating the discharge contributions from various sub-basins.

Steps

1. **Data preprocessing**
   Combine snow accumulation data, discharge data, and lagged temperature data into a single dataset for analysis. Ensure that the data is aligned by date and elevation range.
2. **Feature selection**
   Use the RandomForestClassifier to determine which snow accumulation features (e.g., specific elevation ranges) and lagged temperature data are most strongly correlated with the discharge data.
3. **Correlation analysis**
   Analyze the relationships between snow accumulation, lagged temperature, and discharge for each elevation range and sub-basin.
4. **Model training**
   Train the RandomForestClassifier on the combined dataset to identify patterns and make predictions about discharge based on snow accumulation and lagged temperature.
5. **Identification of key areas**
   Determine which of the 47 regions show the strongest correlation between snow accumulation, lagged temperature, and discharge patterns, and identify five areas that closely align with the overall discharge trends at the main outlet point.

Output

Identification of five areas that closely correlate with discharge patterns. These areas will provide insights into the regions within the Upper KRB that are most influential in predicting discharge events, contributing to a better understanding of how extreme events are generated.

4.15. Fase 15: Upper Karnali: Detailed Analysis of the Five Areas

Input

Five identified areas, each with a time frame of average snow accumulation before an extreme event and a time frame of average lagged temperature, serving as the best indicators for predicting high discharge.

Process

Building on the results from the previous step, which identified the best indicators and key regions for predicting extreme discharge events, this step aims to further refine the analysis by focusing on the five most indicative areas. The previous analysis identified snow accumulation and lagged temperature as significant predictors, with specific time periods showing strong correlations. In this step, we will investigate the exact values of snow accumulation and lagged temperature required to predict the probability of exceeding discharge thresholds at the main outlet point. We will conduct an examination of snow accumulation across various elevation ranges and its correlation with extreme discharge events downstream in the Upper Karnali. Specifically, we will quantify the probability of exceeding specific discharge thresholds based on the constructed snow accumulation and lagged temperature data from these key areas.

Steps

1. **Data collection**
   Construct snow accumulation and lagged temperature data for the different elevation ranges from the five identified key areas. This is done by resampling the ERA5 data and implementing the snow module, as described in the previous steps. The data will include values over various specified periods, such as 10-15 days or 25-35 days before an event for snow accumulation, and 5-10 days or 10-20 days for lagged temperature, depending on the findings from the previous step.
2. **Statistical analysis**
   Conduct a statistical analysis to establish correlations between snow accumulation, lagged temperature, and discharge events exceeding 2000 m^3/s at the main outlet. This step involves using the RandomForestClassifier to quantify these relationships and determine the probability of exceeding the discharge threshold.
3. **Probability calculation**
   Calculate the probability of discharge exceeding 2000 m^3/s based on the snow accumulation and lagged temperature patterns observed in the various elevation ranges. This can be done by using the trained model, where we input hypothetical snow accumulation values for the different elevation ranges of the high indicator areas, along with their lagged temperature. The trained machine learning model (RandomForestClassifier) will provide the probability of exceedance based on historical correlations.

Output

The expected outcome is a trained model where hypothetical snow accumulation values for different elevation ranges in high indicator areas can be inputted, providing the probability of exceedance for a specified discharge downstream at the Main outlet of the Upper Karnali.

5 Results

5.1. General application code

The Python code developed for this thesis offers tools for analyzing snow mass and melt patterns. It allows for flexible comparisons between any two dates within the 1991-2022 period, enabling a detailed examination of temporal changes. Additionally, the code quantifies where the most significant melting has occurred and identifies the elevations with the most snow accumulation. It also determines the snow/rain boundary, providing important insights into where snowfall transitions to rainfall.

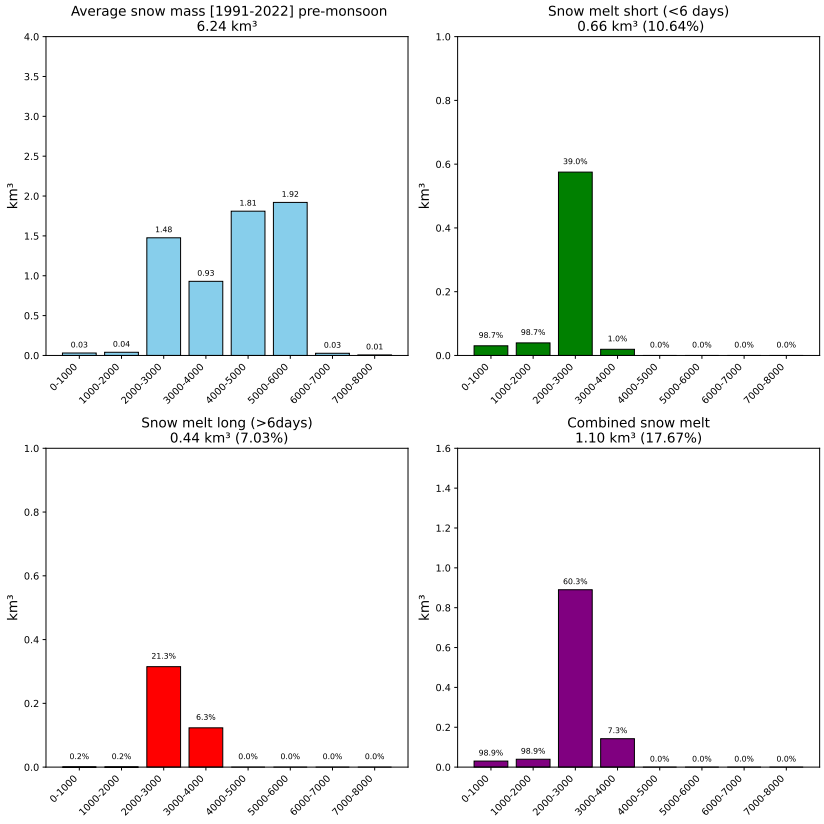


Figure 5.1: Average snow mass and melt distribution during the pre-monsoon period, illustrating the proportion of rapid versus gradual snowmelt.

In addition to temporal analysis, the code generates spatial distributions of snow accumulation, as shown in the provided figure. These spatial outputs can be easily integrated into GIS platforms like QGIS, facilitating detailed mapping and analysis of snow storage across different regions. This

combination of temporal and spatial capabilities makes the code a valuable tool for exploring the factors that influence snow mass and melt, supporting the broader research objectives of the thesis.

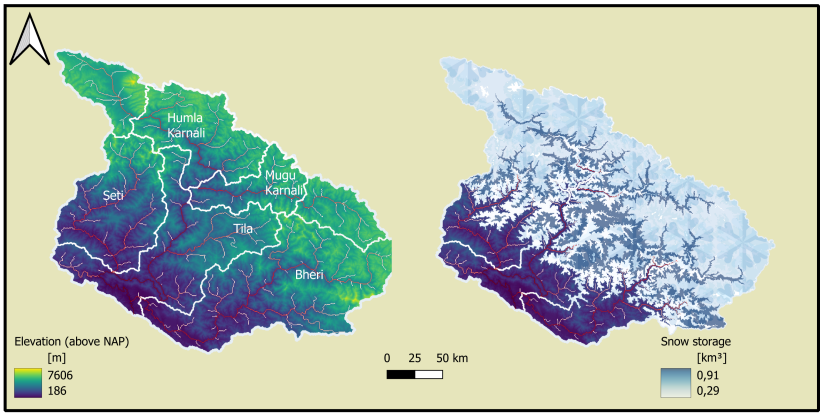


Figure 5.2: Automatically generate QGIS map after implementing interactive python code to visualize snow accumulation in pre-monsoon period

Comparison snow mass pre-monsoon versus monsoon

During the monsoon period, there is a notable shift in both average snow mass and melt patterns compared to the pre-monsoon season. Over the period from 1991 to 2022, the data consistently shows a significant decrease in average snow mass during the monsoon, indicating substantial melting driven by the season's elevated temperatures and rainfall. While the short-term melt remains similar to that observed in the pre-monsoon period, the long-term melt becomes more pronounced, accounting for the majority of the total melt. This results in a considerably higher overall melt volume during the monsoon.

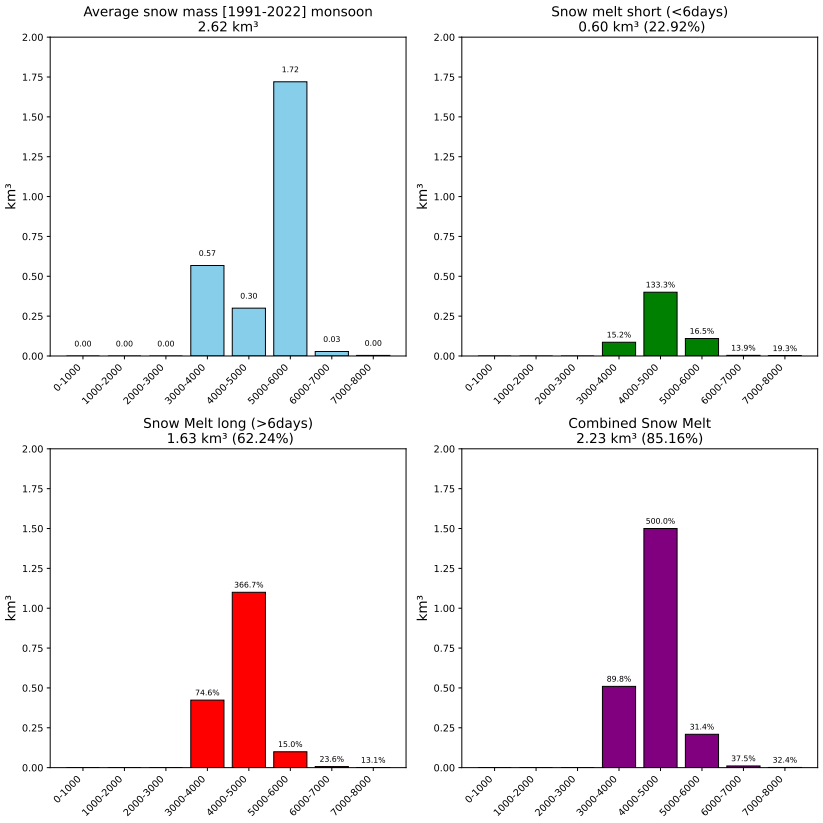


Figure 5.3: Average snow mass and melt distribution during the pre-monsoon period, illustrating the proportion of rapid versus gradual snowmelt.

In addition to the broader trends observed across different elevations, the data for the monsoon period indicates a significant increase in the melting process at altitudes between 4000 and 5000 meters above MSL. Here, the melt rate increases by 500%, meaning the snow mass at these elevations decreases to one-fifth of its initial volume during the monsoon. This reduction is reflected in the snow accumulation, which starts at approximately 1.8 km³ and decreases to 0.3 km³ by the end of the period.

This notable decrease highlights the pronounced ablation processes at these lower elevations during the monsoon, emphasizing the susceptibility of snow reserves in this altitude range to seasonal climatic conditions.

Comparison rain volume/intensity pre-monsoon versus monsoon

The analysis of the figures, spanning 1991 to 2022, reveals that during the pre-monsoon period, a larger proportion of total rainfall is concentrated in the top 20 events, highlighting the season's intense rainfall episodes. In contrast, the monsoon period shows a significant increase in overall rainfall, with a higher percentage of total precipitation occurring at higher elevations during these top events. This suggests that at greater altitudes, rainfall becomes increasingly dependent on a few heavy events, contributing significantly to the overall precipitation and runoff during the monsoon season. Additionally, the rain/snow boundary during the monsoon is identified around 3000 meters, where precipitation transitions from rainfall to snowfall.

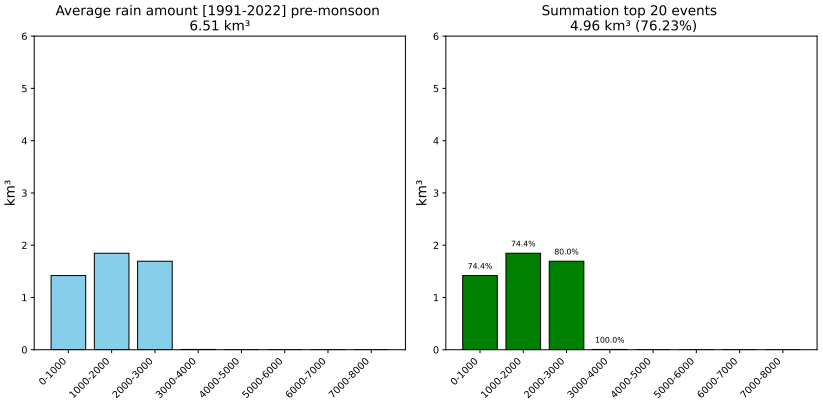


Figure 5.4: Average rainfall distribution during the pre-monsoon period (1991-2022), highlighting total rainfall and the contribution of the top 20 rainfall events to the overall volume.

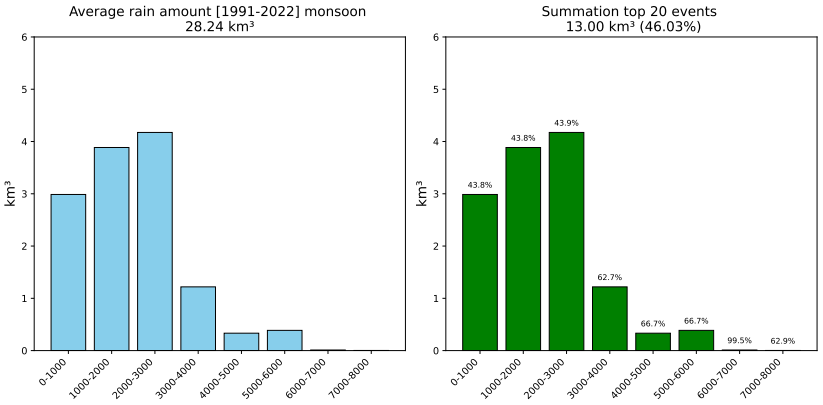


Figure 5.5: Average rainfall distribution during the monsoon period (1991-2022), illustrating the significantly higher total rainfall and the influence of the top 20 rainfall events on the total precipitation volume.

These patterns, verified by literature, indicate that the increased rainfall volume and intensity during the monsoon are primary drivers of accelerated runoff, leading to heightened discharge and potential flooding (Ougahi, 2024). The spatial distribution further underscores the role of elevation in influencing hydrological responses, where lower elevations receive direct rainfall, and higher elevations contribute through snowmelt (Clemenzi et al., 2023). The amplified runoff due to intense monsoon events highlights the need for hydrological models to account for both rainfall volume and intensity to predict extreme discharge events effectively (Ougahi, 2024).

5.2. Scenario selection

Based on the flowchart, scenario selection involves evaluating and categorizing climatic data to identify years with extreme snow and rain events. This systematic approach ensures that we capture the most significant variations in hydrological responses, providing a comprehensive understanding of how the system reacts to different stressors. To achieve this, we used the output from the interactive Python code, which analyzed the ERA5 dataset to rank years based on snow accumulation and rainfall volume. The highest-ranked years were selected to create distinct scenarios for further study.

Scenario 1: Snow-dominated scenario

In the first scenario, we focus on the years with the highest accumulated pre-monsoon snow accumulation. By selecting the four years with the highest snow volumes, we aim to observe the amplified effects of prolonged snow accumulation. Literature suggests that consecutive years with high snow accumulation can lead to a saturated hydrological system, making it more sensitive to temperature shifts. For instance, Viviroli et al. (2007) highlight that repeated high snow years increase the vulnerability of the hydrological system to temperature variations, potentially leading to more extreme runoff events. This scenario was designed to investigate the hydrological system's response to consecutive high-snow years and understand how it influences sensitivity to temperature changes. The selected years are:

Year	Snow Storage (km³)	Rank
2021	8.4 (pre-monsoon)	1
1991	8.1(pre-monsoon)	2
2007	7.9 (pre-monsoon)	3
2017	7.5 (pre-monsoon)	4

Table 5.1: Top snow storage values for pre-monsoon periods ranked by year.

In the next step this consecutive years will be used as forcing data in the calibrated SPHY model for the KRB to observe the effects on the discharge cycle.

Scenario 2: Rain-dominated scenario

The second scenario examines the hydrological response to years with the highest accumulated monsoon rainfall. We sequence the four years with the highest rainfall volumes to observe their amplified effects. Studies suggest that multiple years of high rainfall can saturate the hydrological system, increasing its sensitivity to additional rainfall. Consecutive high rainfall years can enhance the system's vulnerability to runoff events, potentially causing more extreme flooding. This scenario aims to investigate the hydrological system's response to consecutive high-rainfall years and understand how it influences sensitivity to extreme rainfall events (Fang, 2021). The selected years are:

Year	Rain Volume (km³)	Rank
2018	43.3 (monsoon)	1
2013	41.3 (monsoon)	2
2010	40.1 (monsoon)	3
2007	39.7 (monsoon)	4

Table 5.2: Top rain volumes for monsoon periods ranked by year.

Scenario 3: Rain-on-snow scenario

To broaden our understanding, we also create an rain-on-snow scenario that combines two years of the highest snow accumulation with two years of the highest monsoon rainfall. This scenario uses the rank 1 and 2 years for both snow and rain to provide insights into the combined effects of extreme snow and rainfall on the hydrological system. By evaluating these combined scenarios, we aim to understand how the hydrological system responds to simultaneous snow and rain extremes, offering a more holistic view of the potential impacts on the system. The selected years are:

Year	Snow Storage (km³)	Snow Rank	Rain Volume (km³)	Rain Rank
2021	8.4 (pre-monsoon)	1	35.8 (monsoon)	9
1991	8.1 (pre-monsoon)	2	27.7 (monsoon)	31
2018	2.1 (pre-monsoon)	30	43.3 (monsoon)	1
2013	1.9 (pre-monsoon)	32	41.3 (monsoon)	2

Table 5.3: Comparison of snow storage and rain volume for selected years, ranked by year.

FDC

The manipulated ERA5 dataset, now differentiated into high rain volumes and large snowfall scenarios, is used to analyze these changes. By artificially generating high discharge events through the strategic alignment of climatological phenomena, the model examines how different precipitation scenarios affect the FDC at Chisapani bridge. Chisapani bridge is selected as the outlet point since it is where the five sub-basins converge; this is illustrated in figure .

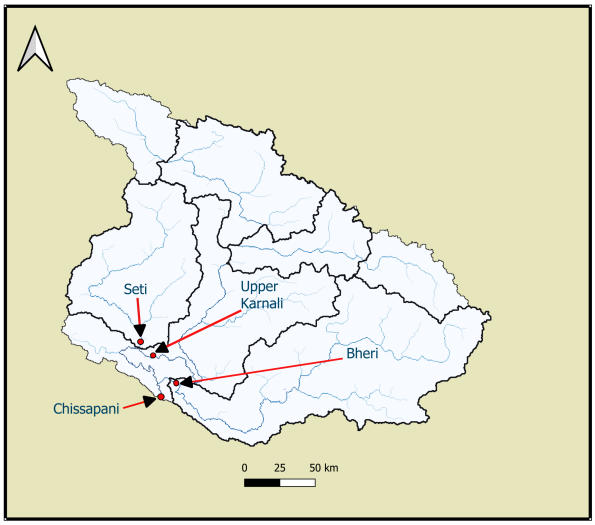


Figure 5.6: Average snow mass and melt distribution during the pre-monsoon period, illustrating the proportion of rapid versus gradual snowmelt.

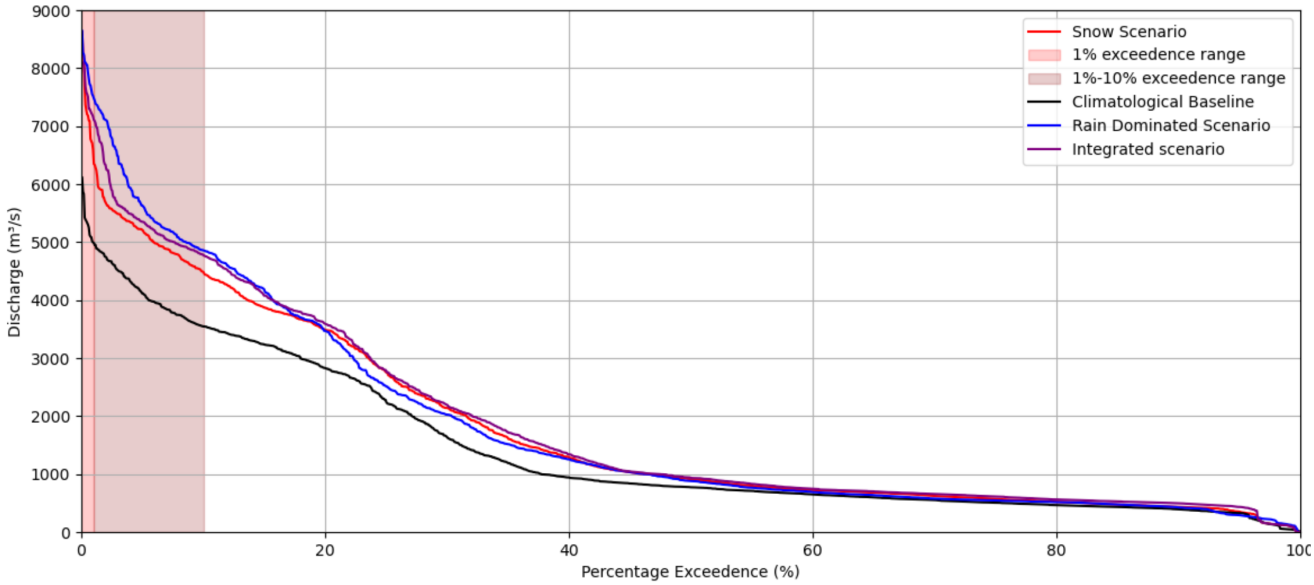


Figure 5.7: On the x-axis the percentage of exceedence versus discharge on the y-axis. The FDCis based on the outcome from the calibrated SPHY model while forces with ERA5 data resampled for the rain-dominated (red line), snow-dominated (blue line) and rain-on-snow scenario (purple line).

1. **High flow segment (1-10% Exceedence Range)**

    - The rain-dominated scenario shows the highest increase in high flows, with 41% for high flow mean and 41% for high flow top 1%. This is followed by the snow-dominated scenario with a 26% increase in high flow mean and a 41% increase in high flow top 1%. The rain-on-snow scenario, while also showing a significant increase, is slightly lower with 33% and

38%. The primary reason for the highest increase in the rain-dominated scenario is the direct impact of intense and frequent rainfall events, which cause rapid runoff and peak discharges. The snow-dominated scenario benefits from significant snowmelt, but this is more gradual compared to the immediate impact of rainfall.

2. **Mid range segment (10-90% Exceedence Range)**

   - The rain-on-snow scenario shows the highest increase in mid-range flows, with 27% for the mid-range mean and 34% for the mid-range top 1%. This is followed by the snow-dominated scenario with 22% and 25%, and then the rain-dominated scenario with 20% and 37%. The combined effects of snowmelt and rainfall in the rain-on-snow scenario provide a more stable and sustained flow, ensuring consistent discharge levels. This balance is less pronounced when only one factor (snow or rain) is dominant.

3. **Low flow segment (90-100% Exceedence Range)**

   - The rain-on-snow scenario exhibits the highest increase in low flow values, with 32% for the low flow mean and 27% for the low flow top 1%. This is more significant than the increases seen in either the snow-dominated or rain-dominated scenarios. The snow-dominated scenario shows a 15% increase in the low flow mean and an 11% increase in the low flow top 1%, while the rain-dominated scenario shows a 10% increase in the low flow mean and an 9% increase in the low flow top 1%. The rain-on-snow scenario benefits from continuous contributions from both snowmelt and rainfall, preventing the river from reaching very low discharge levels and ensuring higher baseflow during typically dry periods.

5.3. RandomForestClassifier

The RandomForestClassifier model identified key features contributing to extreme discharge events at Chisapani. By analyzing the mean discharge from different sub-basins, it became evident that the Upper Karnali sub-basin demonstrated a more stable mean discharge with a lower standard deviation compared to others. This stability, combined with lower variability, indicated that the Upper Karnali sub-basin is a better predictor of extreme events, making it a crucial area for further analysis.

Density Plot for Upper Karnali for the 10 day discharge lag in predicting an extreme discharge event at Chisapani [>8000 m³/s]. The blue area indicates 'Target 0', meaning that 10 days later at Chisapani there is no extreme discharge event above 8000 m³/s, and 'Target 1' meaning that 10 days later at Chisapani there will be an extreme discharge event above 8000 m³/s.

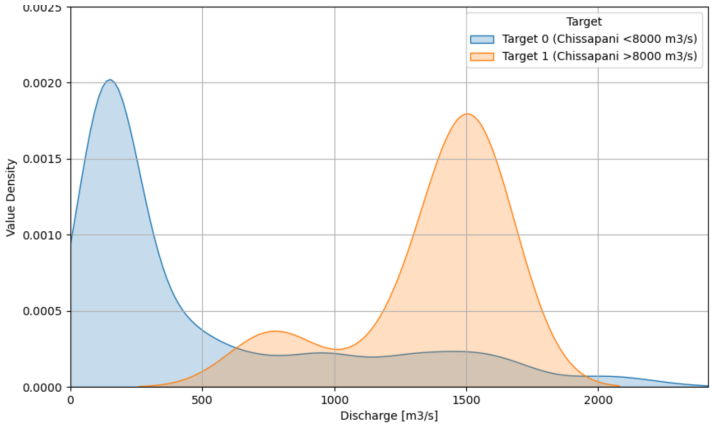
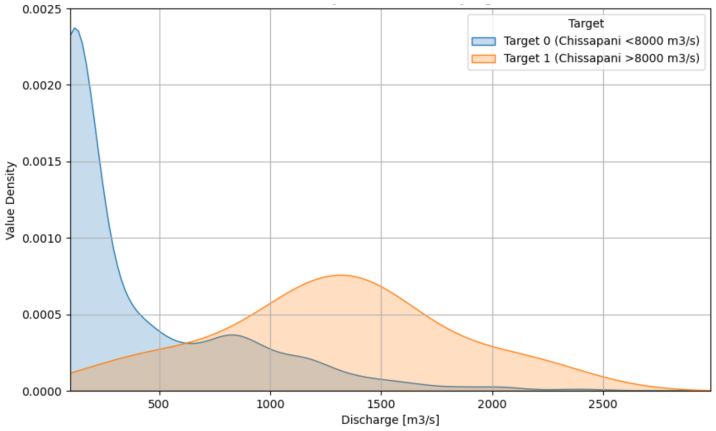


Table 5.4: Mean and Standard Deviations for the lagged discharge days [1, 5, and 10] at Upper Karnali and Seti. This table highlights the differences in variability between the snow-fed Upper Karnali and the rain-fed Seti.

River	Lag Day	Mean (m³/s)	Standard Deviation (m³/s)
Upper Karnali	10	1606	291
	5	1772	242
	1	1910	242
Seti	10	990	650
	5	1747	630
	1	2348	330

Density Plot for Seti with a discharge lag of 10 days before the extreme event at Chisapani (above 8000 m³/s). The blue area indicates 'Target 0', meaning that 10 days later at Chisapani there is no extreme discharge event above 8000 m³/s, and 'Target 1' meaning that 10 days later at Chisapani there will be an extreme discharge event above 8000 m³/s.



Focus on upper Karnali: lagged temperature as a predictor

Upon further analysis, we identified that the most influential climatological parameter for predicting extreme events in the Upper Karnali is lagged temperature. The model highlighted that average temperature, especially when lagged, serves as a critical indicator for forecasting high discharge events. This finding aligns with the hydrological characteristics of the Upper Karnali, where temperature variations significantly affect snowmelt timing and magnitude.

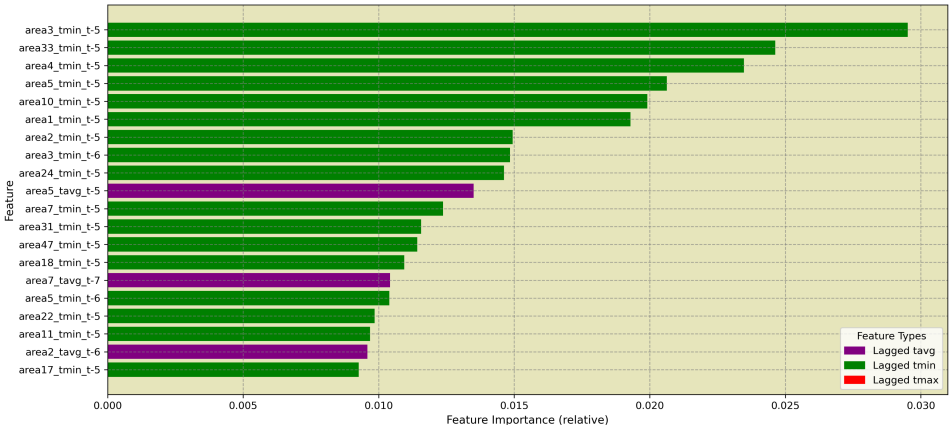


Figure 5.8: Top 20 feature importance generated by the machine learning model for the Upper Karnali. The model assigns high feature importance to lagged temperature variables, particularly lagged tmin, making it a strong predictor of discharge trends in this snow-fed river.

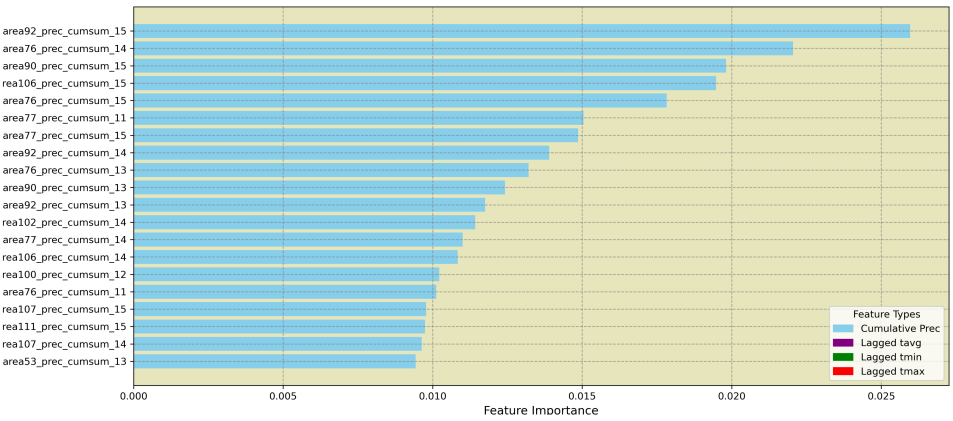


Figure 5.9: Top 20 feature importance generated by the machine learning model for the Bheri. The model assigns high feature importance to cumulative precipitation, which aligns with Bheri's characteristics as a rain-fed river, where rainfall directly influences discharge patterns.

Following the identification of lagged temperature as a key predictor in the Upper Karnali, our analysis then shifted focus to the Bheri River, a rain-fed system where different climatic factors are expected to influence discharge patterns. Given its reliance on direct rainfall, we anticipated that cumulative precipitation would play a more significant role in predicting extreme discharge events. To validate this hypothesis, we applied the same machine learning model to assess the relative importance of various climatological parameters in forecasting high discharge events in the Bheri River.

Having established that lagged temperature, particularly average temperature, is a critical predictor for extreme events in the Upper Karnali, we now turn our attention to identifying specific areas within the basin that are most indicative of these events. By analyzing the spatial distribution of snow accumulation and melt across the delineated sub-basins, we aim to determine whether certain regions exhibit stronger correlations with extreme discharge events. This detailed examination will help pinpoint the top five areas that not only align closely with the extreme conditions observed but also offer valuable insights into the localized hydrological responses within the Upper Karnali.

5.4. Detailed analysis top 5 indicating areas

After completing phases 12, 13, 14, and 15, we identified five key areas that closely correlate with extreme downstream discharge events at the Upper Karnali. These regions, shown in the accompanying maps, provide critical insights into the spatial distribution of snow accumulation and melt, which are significant predictors of high discharge events.

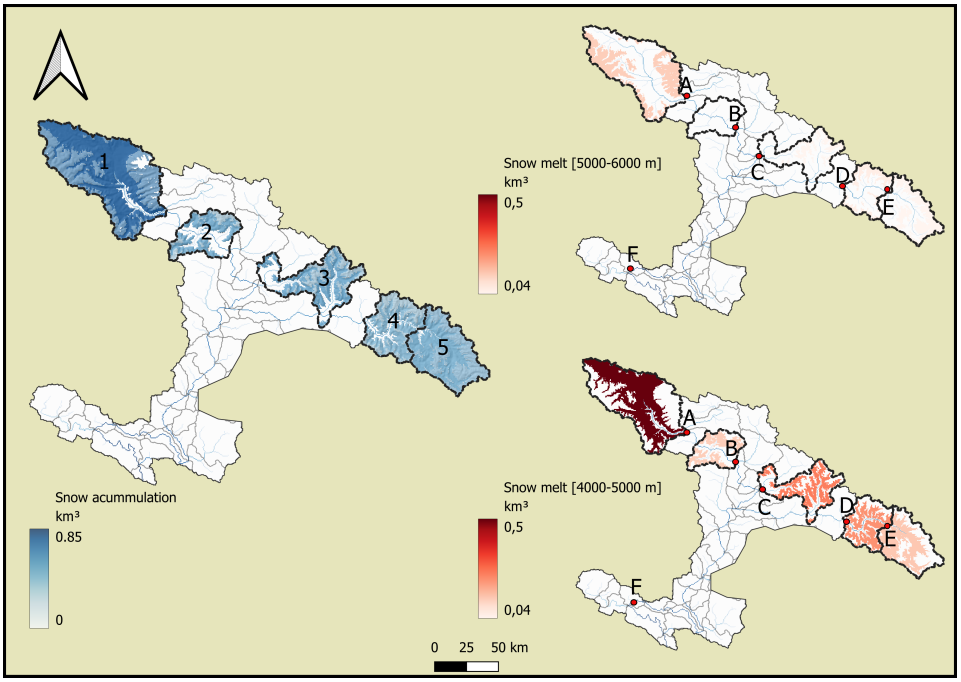


Figure 5.10: This figure illustrates the average snow accumulation during the period of 25-35 days before an event where the Upper Karnali at outlet F (downstream) exceeds a discharge value of 2000 m³/s. Additionally, the two figures on the right show the average snow melt occurring in the weeks preceding the high discharge event.

We observed five identical patterns across the key areas:

1. **Pattern 1**

   Consistent snow accumulation trends are seen across different elevation ranges, with significant snow accumulation observed 25-35 days before the high discharge event. The snow volumes steadily increase, reaching their peak just before the discharge event.
2. **Pattern 2**

   Lagged temperature data shows a rise in average temperatures 10-20 days before the event. This temperature increase coincides with the onset of snow melt, indicating a strong correlation between rising temperatures and subsequent discharge events.
3. **Pattern 3**

   The snow melt predominantly occurs in the elevation range of 4000-5000 meters. This elevation range consistently shows the highest volumes of snow melt, which directly contributes to the downstream discharge.
4. **Pattern 4**

   The relationship between snow melt and discharge is evident in all five plots. As the snow melts at higher elevations, the discharge at outlet F increases, highlighting the impact of snow melt on river discharge levels.
5. **Pattern 5**

   Each plot demonstrates a clear temporal sequence where snow accumulation peaks first, followed by rising temperatures and snow melt, ultimately leading to increased river discharge. This sequence is consistent across all five areas, underscoring the reliability of these indicators in predicting high discharge events.

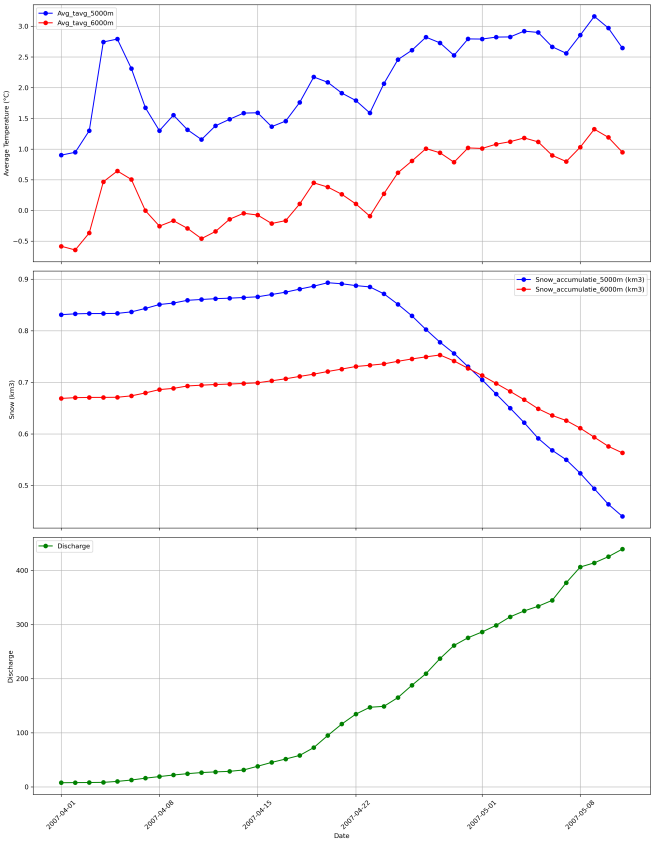


Figure 5.11: Figure top: x-axis average values from 04-01 till 05-08 versus temperature in °C for 4000-5000 meter MSL elevation range and 5000-6000 meter MSL elevation range. Middel figure: x-axis average values from 04-01 till 05-08 versus snow accumulation in km^3 for 4000-5000 meter MSL elevation range and 5000-6000 meter MSL elevation range. Bottom figure: x-axis average values from 04-01 till 05-08 versus discharge in m^3/s at the outlet points from the delineated areas.

Probability Density Functions

The combined analysis of snow accumulation and temperature reveals critical insights into the conditions that lead to high discharge events in the Upper Karnali basin. Specifically, within the 4000-5000 meter elevation range, the probability of high discharge (exceeding 2000 m^3/s) is strongly influenced by both factors, as illustrated by the following findings.The first set of figures demonstrates the predictive power of snow accumulation within this range. The probability density functions clearly show that when the average snow accumulation 35-25 days prior to the event exceeds certain thresholds, the likelihood of high discharge events increases significantly. These areas fall within the >90% prediction interval, emphasizing their importance as predictors.

However, snow accumulation alone is not sufficient to trigger these extreme events. The accompa-

nying temperature plot for the same elevation range shows that high discharge events are only probable when the average temperature 10-20 days before the event also falls within the critical >90% range. Specifically, temperatures above 2°C significantly increase the probability of high discharge, as shown by the density contours. These findings underscore the necessity of monitoring both snow accumulation and temperature data in the 4000-5000 meter range. Only when both conditions align within their respective >90% ranges does the probability of high discharge exceed 90%, making these areas particularly vulnerable to extreme events. This highlights the need for integrated monitoring and modeling efforts to effectively predict and manage flood risks in the Upper Karnali basin.

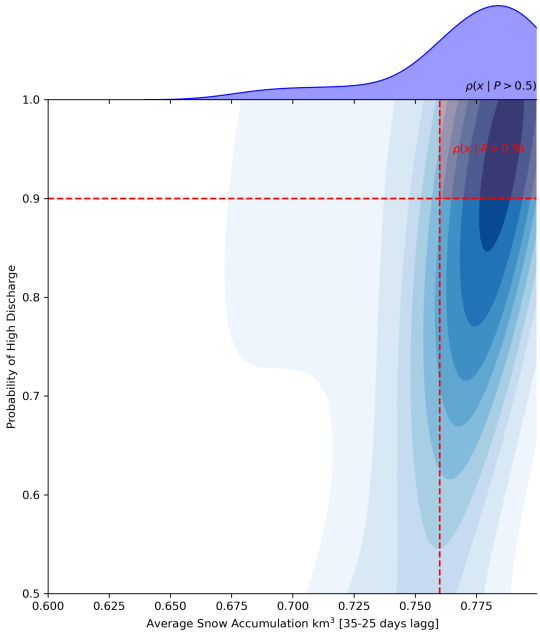


Figure 5.12: Area 1, 4000-5000 m: Snow accumulation in this range significantly increases high discharge probability, nearing 90% at 0.75 km³. The top curve shows most cases lead to over 50% discharge probability.

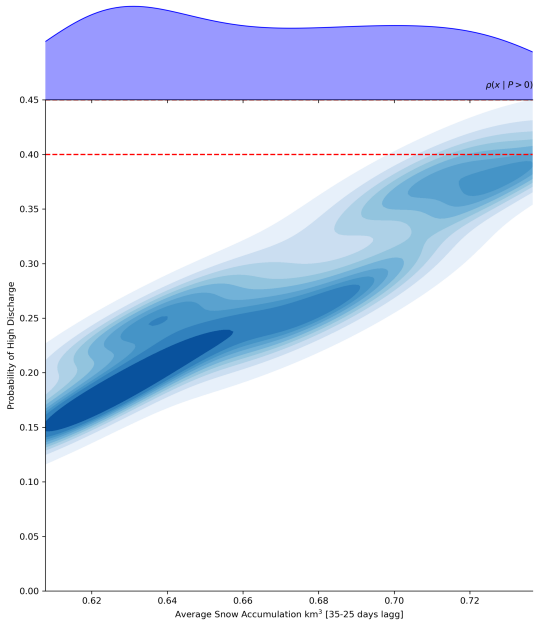


Figure 5.13: Area 1, 5000-6000 m: Here, snow accumulation is a weaker predictor, with discharge probability maxing out around 45%. The flatter top curve indicates fewer instances exceed 50% probability.

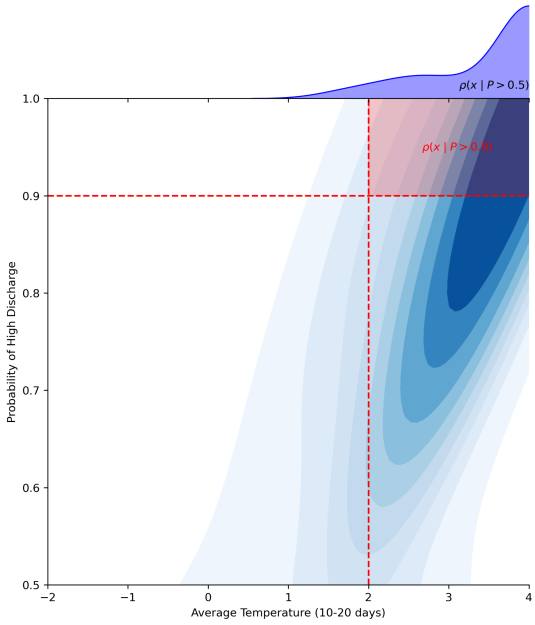


Figure 5.14: High discharge probability increases sharply when the average temperature 10-20 days prior exceeds 2°C, with a >90% probability of extreme discharge events.

Evaluation of machine learning model performance and predictive accuracy

The evaluation of the machine learning model's predictive performance when forecasting whether the river discharge will exceed 2000 m³/s downstream at the Upper Karnali, based on snow accumulation, demonstrates significant insights into both its strengths and limitations.

Precision, recall, and F1-score analysis

The performance metrics indicate a high overall accuracy of 97%, showcasing the model's strong ability to correctly predict the outcomes. However, when delving into the specific metrics for each class, the distinction becomes evident:

Table 5.5: Classification report for high discharge prediction

Metric	Class 0 (No High Discharge)	Class 1 (High Discharge)
Precision	0.99	0.81
Recall	0.98	0.91
F1-Score	0.98	0.86
Support	366 instances	43 instances

The model shows excellent precision and recall for Class 0, meaning it is highly effective at predicting when the discharge will not exceed 2000 m³/s. This indicates that the model is highly reliable in not over-predicting extreme events, thus minimizing false alarms, which is crucial in operational settings where unnecessary flood warnings can lead to public distrust and operational inefficiencies.

For Class 1, the model exhibits a slightly lower precision of 0.81, suggesting that when it predicts a high discharge event, it is correct 81% of the time. The recall of 0.91 is strong, indicating that the model successfully identifies 91% of actual high discharge events, though it does miss a small percentage of such events. The F1-score of 0.86 provides a balanced view of precision and recall, confirming the model's robustness but also highlighting areas for improvement, particularly in reducing false positives (where the model predicts a high discharge that does not occur).

6 Discussion

The study of the hydrological responses of the KRB under extreme weather conditions has relied on the use of the SPHY (Spatial Processes in Hydrology) model and the machine learing model RandomForestClassifier. These tools, supported by the ERA5 reanalysis dataset, have facilitated detailed simulations and predictions. However, several methodological and technical challenges must be acknowledged to fully understand the limitations and implications of the findings.

6.1. ERA5 Data

The ERA5 dataset, which underpins much of the modeling in this study, offers high temporal resolution and comprehensive coverage. However, its spatial resolution of 31 km poses significant challenges for accurately capturing localized climatic phenomena. This coarse resolution may obscure microclimatic variations that are crucial for detailed hydrological modeling, particularly in the heterogeneous terrain of the KRB. The generalization required at this resolution can lead to the averaging out of important climatic variables, thereby reducing the accuracy of predictions.

Additionally, ERA5 is a reanalysis product that integrates various observational datasets and model outputs. While this process improves data consistency, it also introduces uncertainties related to the quality and representativeness of the underlying observations. These uncertainties can propagate through the hydrological modeling process, affecting the reliability of the results. The generalization inherent in the ERA5 data may not fully capture the complexity of the climatic drivers in the region, potentially leading to inaccuracies in the SPHY model simulations and the Random Forest Classifier predictions.

6.2. SPHY

The SPHY model simulation of snow accumulation and melt dynamics is central to understanding the basin's hydrological behavior. The degree-day approach, while effective in modeling temperature-driven snowmelt, has notable limitations. The fixed temperature threshold used to distinguish between snow and rain may oversimplify the complex conditions that dictate precipitation type (Kienzle, 2008). Additionally, other critical factors like atmospheric pressure, wind speed, and solar radiation are not incorporated into the snowmelt calculations. This exclusion could lead to inaccuracies, particularly in regions where these factors have a significant impact on the rate and timing of snowmelt (Lee et al., 2023).

Moreover, the model approach to flow routing, which calculates water movement across the basin terrain, simplifies the hydrodynamic processes that occur in reality. While the model successfully simulates general flow patterns, it may fail to capture the nuances of water flow in areas with complex topography. For instance, the interaction between channel flow and floodplain dynamics is not fully represented, which could result in underestimations or overestimations of peak flows during extreme weather events.

The process of area scaling within the SPHY model is crucial for estimating snow and rain volumes across different elevation ranges. However, this method introduces several assumptions that can impact the accuracy of the model's predictions. The two-dimensional area scaling approach assumes that snow accumulation and precipitation are uniformly distributed across each elevation band, regardless of terrain heterogeneity. This assumption overlooks the significant influence of topographic features such as slope, aspect, and surface roughness on snow distribution and retention (Lee et al., 2023).

In mountainous regions like the KRB, steeper slopes often retain less snow compared to flatter areas, where snow can accumulate more readily. Additionally, the aspect of a slope—its orientation relative to the sun—can affect the amount of solar radiation it receives, further influencing snowmelt rates. The simplified area scaling method used in SPHY does not fully account for these variations, potentially leading to inaccuracies in the estimation of snow water equivalent (SWE) and, subsequently, river discharge.

Research indicates that more advanced scaling methods, which incorporate three-dimensional terrain characteristics, can provide a more accurate representation of snow distribution. However, implementing such methods requires more detailed topographic data and increases the complexity of the model. While the current approach in SPHY is practical for broad-scale analysis, it may not capture the detailed spatial variability needed for precise hydrological predictions, especially in regions with highly varied terrain.

Predicting the future based on the past

The methodology of predicting future hydrological responses based on historical data is inherently limited by the assumption that the factors influencing river discharge remain constant over time. This assumption is increasingly problematic in the context of climate change, where shifts in precipitation patterns, temperature trends, and land use practices can significantly alter the hydrological characteristics of a basin. The potential for such changes to invalidate historical data as a reliable predictor of future conditions is a critical concern that limits the applicability of the current models.

The use of a fixed temperature threshold for snowmelt, the assumption of uniform soil properties in the groundwater module, and the reliance on historical flow data for generating FDC all contribute to an oversimplified understanding of the basin's hydrological processes. While these approaches are necessary to operationalize the models, they introduce significant uncertainties that must be considered when interpreting the results.

Pre-monsoon and monsoon date selection

The pre-monsoon period in this study was defined from February to May (beginning of February to end of May), and the monsoon period from June to September (beginning of June to end of September). This definition is based on general climatic patterns observed in Nepal, where the monsoon season typically starts around June 13 and ends around September 23, bringing about 80% of the annual rainfall.

While the chosen periods are practical for delineation, alternative definitions could provide more precise insights into climatic impacts on snow accumulation and melt processes. For instance, the monsoon could be defined based on the percentage of total annual rainfall received. Another method could involve identifying the monsoon onset when rainfall exceeds a certain threshold over several consecutive days. These approaches might better capture the variability and onset of monsoon conditions, potentially leading to more accurate modeling of hydrological responses.

The current assumption that pre-monsoon snow accumulation is only considered up to the end of May may overlook significant events. It is possible that substantial snow accumulation could occur in early June, which can still heavily influence monsoon dynamics. This oversight could result in missing critical years where high snow accumulation in June could trigger intense monsoon reactions. Therefore, incorporating a more dynamic and responsive definition of the monsoon period might enhance the accuracy of the study's predictions. The climatic data indicates that monsoon variability and onset can significantly affect hydrological processes. Research shows that the onset and intensity of the

monsoon are critical factors in determining the water availability and snowmelt dynamics in Himalayan regions. According to literature (Nepal & Shrestha, 2015), monsoon onset variability can lead to significant differences in water resource management outcomes. Similarly, a delayed or early onset of the monsoon season can substantially impact snowmelt and river discharge patterns. This has been highlighted in previous studies. (Jain & Singh, 2020).

6.3. FDC

When classifying flow regimes using exceedance percentages such as 1% exceedance (high flows), mid-range, and low-range exceedance (low flows), several limitations and discussion points arise.

- **Representation of flow variability**
  FDC provide a simplified but strong representation of flow variability by plotting the percentage of time specific discharges are equaled or exceeded. However, they do not account for the timing and sequence of flows, which can be critical in understanding hydrological and ecological dynamics. For instance, high flows that occur consecutively versus sporadically can have different impacts on riverine ecosystems and water resource management.
- **Impact of data length and quality**
  The accuracy of an FDC is highly dependent on the length and quality of the historical flow data used. Shorter datasets or those with gaps can lead to less reliable FDC, potentially misrepresenting the true flow characteristics of the river. This can be particularly problematic in regions with limited hydrological monitoring data (Pumo et al., 2018) (Liucci et al., 2014).
- **Climate and land use changes**
  FDC often use historical data to predict future flow regimes. However, changes in climate and land use can significantly alter these regimes, making historical FDC less reliable for future predictions. As climate change progresses, alterations in precipitation patterns, snow melt timings, and evapotranspiration rates can all impact river flows in ways that historical data may not fully capture (Langat et al., 2019).
- **Climate and land use changes**
  Human activities such as dam construction, water withdrawals, and land use changes can significantly alter natural flow regimes. FDC constructed from historical data may not accurately reflect these anthropogenic impacts, leading to potential misinterpretations of flow conditions and their ecological consequences (Langat et al., 2019).

6.4. RandomForestClassifier

The RandomForestClassifier, employed to predict high discharge events, is a powerful tool for handling large datasets and capturing complex relationships between variables. However, its application in this study comes with challenges. The model's reliance on historical data assumes that future conditions will closely mirror past patterns—a premise that may not hold true given the dynamic nature of climate and land-use changes. The limited temporal span of the dataset (30 years) further constrains the model's ability to generalize to future scenarios, potentially limiting its effectiveness in predicting extreme events.

Another significant limitation of the RandomForestClassifier is its "black-box" nature, which makes it difficult to interpret the decision-making process behind its predictions. Although techniques such as decision tree visualization can provide some insight, the overall opacity of the model reduces the transparency of the results, which is a critical issue in scientific research where interpretability is often as important as predictive power. Moreover, the model purely focuses on identifying patterns in data without considering the underlying physical processes. For instance, when analyzing cumulative precipitation, the model might find a strong correlation with a 20-day lag, even if the maximum time it takes for rainfall to impact discharge at Chisapani is only 10 days. The model does not have the capability to recognize the physical impossibility of such correlations; it solely identifies statistical relationships. Therefore, it is crucial to complement the model's findings with a strong understanding of the physical processes involved to ensure that the predictions are not only statistically valid but also physically plausible.

Final remarks discussion

The combination of SPHY and the Random Forest Classifier offers valuable tools for modeling and predicting the hydrological responses of the KRB under extreme weather conditions. However, the limitations discussed highlight the need for cautious interpretation of the results. The assumptions embedded in the models, the coarse spatial resolution of the input data, and the reliance on historical scenarios all contribute to uncertainties that must be acknowledged. The area scaling approach, while practical, may oversimplify the complex interactions between topography and snow distribution, leading to potential inaccuracies. Future research should aim to address these limitations by incorporating higher-resolution data, extending the temporal span of the datasets, and exploring more complex interactions between climatic variables. Such efforts would enhance the reliability and applicability of hydrological predictions in this and similar regions.

7

Conclusion

Main research question: How do snow accumulation, temperature fluctuations, and their interplay influence river discharge in the KRB, particularly in understanding the drivers of extreme flooding events?

When the model was forced with snow-dominated scenarios, the FDC showed significantly more extreme behavior compared to the climatological baseline. The high-flow segments, particularly within the top 1% of discharge values, increased by 41%, demonstrating how rapid snow melt driven by rising temperatures can lead to extreme peaks in river discharge. This underscores the critical role of snow dynamics in shaping the basin's hydrological response during years of significant snow accumulation.

Further analysis revealed that snow accumulation and temperature fluctuations are pivotal in influencing river discharge, especially in the context of extreme flooding events. In snow-fed rivers like the Humla Karnali and Mugu Karnali, snow accumulation at elevations between 4000 and 5000 meters, followed by significant temperature increases, were identified as key predictors of extreme discharge events. This elevation range is crucial because it is where snow accumulates during the pre-monsoon period and melts during the monsoon, directly contributing to river discharge. The absence of the rain-shadow effect at this altitude allows for substantial snow accumulation, making it a reliable indicator for potential flooding.

The study also demonstrated that these factors interact dynamically, with snow accumulation providing a baseline for potential discharge and temperature fluctuations serving as the trigger for converting stored snow into runoff. This interplay was particularly evident in years dominated by snow accumulation, where lagged temperature increases drove rapid snowmelt, leading to extreme river discharges. The machine learning model, incorporating these elements, was able to predict with 90% accuracy whether the river discharge would exceed the critical threshold of 2000 m^3/s downstream at the Karnali river. When snow accumulation in identified predictor areas exceeded individual thresholds, coupled with prolonged temperature increases, the model's accuracy rose to 96%.

These findings underline the importance of closely monitoring snow accumulation and temperature trends in the KRB, as their interplay is a significant driver of extreme hydrological events. The main research question—how snow accumulation, temperature fluctuations, and their interplay influence river discharge in the KRB—has been effectively addressed through the following four sub-questions.

1. How can snow accumulation across different elevation ranges in the pre-monsoon period be effectively modeled and visually represented to identify snow-dominated scenarios?

   The visual representation of snow accumulation across different elevations provided valuable insights into the spatial distribution of snow and its potential impact on downstream hydrology. The modeling showed that snow accumulation was not uniform across the basin, with higher elevations experiencing more significant snowpack, which in turn influenced the timing and magnitude of snowmelt and subsequent river flow. An inverse relationship was observed between

snow accumulation and subsequent monsoon rainfall: in years with high pre-monsoon snow accumulation (2021 and 1991), monsoon rainfall was relatively low, whereas in years with extreme monsoon rainfall (2018 and 2013), pre-monsoon snow accumulation was notably lower. This inverse relationship likely arises because high snow accumulation is often associated with cooler and more stable atmospheric conditions, which can limit the convective activity needed for intense monsoon rains. Conversely, years with lower snow accumulation may experience warmer pre-monsoon conditions, which can enhance atmospheric instability and lead to stronger monsoon rainfall. The Python code developed for this analysis was subsequently used to address the other sub-questions in the study.

2. How do the FDC of the KRB vary under different combinations of snow-dominated and rain-dominated scenarios compared to the climatological baseline?

   The FDC of the KRB reveal distinct variations when comparing snow-dominated, rain-dominated, and rain-on-snow scenarios to the climatological baseline. These variations were derived from hydrological simulations conducted using the SPHY model, driven by ERA5 climate data, which included daily averages of temperature (t_{avg}, t_{min}, t_{max}) and precipitation ($prec$). The scenarios were constructed using selected years—2021, 1991, 2007, and 2017 for the snow-dominated scenario; 2018, 2013, 2010, and 2007 for the rain-dominated scenario; and 2021, 1991, 2018, and 2013 for the rain-on-snow scenario—based on earlier analyses that identified the years with the most extreme conditions of snow accumulation and precipitation patterns. These scenarios provided a clear basis for comparing the hydrological impacts of varying weather patterns on the KRB.

   The rain-dominated scenario demonstrated a 41% increase in high-flow values, indicative of the intense and often unpredictable nature of rainfall events in the region. These events resulted in high surface runoff, as precipitation exceeded the soil's infiltration capacity, leading to rapid and significant peaks in river discharge. The steep initial slope of the FDC reflects the basin's quick hydrological response to extreme rainfall events, underscoring the challenges in managing water resources and the heightened flood risks during such periods.

   Similarly, the snow-dominated scenario revealed a marked increase in river discharge, particularly in the higher flow segments. The mean discharge within the 1-10% exceedance range increased by 26%, with the top 1% of discharge values rising by 41% compared to the baseline. This sharp increase is attributed to rapid snowmelt during the early monsoon period, driven by rising temperatures. The steep slope of the FDC indicates a quick decline in discharge following the peak snowmelt period, emphasizing the critical role that snowmelt plays in shaping river discharge during years with significant snow accumulation. This scenario highlights the implications for flood risk, water resource management, and ecosystem health in snow-dominated years.

   The rain-on-snow scenario, which combines the effects of snowmelt and direct rainfall, resulted in FDC that produced significant peaks in discharge, especially in the high-flow segments. The heavy snow accumulation during 2021 and 1991, followed by the intense monsoon rains of 2018 and 2013, led to substantial river discharge. These years were selected based on significant rain-on-snow events, where heavy rainfall occurred on top of an existing snowpack, leading to rapid snowmelt and substantial river discharge. When rain falls on snow, it not only contributes directly to runoff but also accelerates snowmelt by increasing the energy available to melt snow. The ground, often saturated or frozen during these events, had reduced infiltration capacity, further amplifying surface runoff. The steep curve at the beginning of the FDC for this scenario indicates that a small percentage of time corresponds to very high discharge levels, typical of the rapid onset of high flows during rain-on-snow events. This scenario underscores the potential for extreme discharge events that are much higher than what would be expected from either snowmelt or rainfall alone. By maintaining the natural seasonal cycles, this analysis provides a crucial understanding of potential flood risks in regions where both significant snowfall and strong monsoon rainfall are common, even when these factors do not coincide within the same year.

3. What is the relationship between snow accumulation at various elevation ranges, lagged temperature, and river discharge in the KRB, as identified by the Random Forest Classifier model?

   The analysis revealed that, for the Humla Karnali and Mugu Karnali rivers, snow accumulation at elevations between 4000 and 5000 meters above a certain threshold, occurring 35 to 25 days before an event, coupled with a significant increase in temperature 5 to 15 days prior, proved to be key indicators of extreme discharge, defined as combined flows exceeding 2000 m³/s. These correlations are particularly relevant for these two rivers because they are primarily snow-fed, making snow dynamics and temperature variations critical in predicting extreme discharge events. This finding aligns with existing literature, which emphasizes the snow-fed nature of the Humla Karnali and Mugu Karnali rivers, confirming the importance of snow accumulation and temperature fluctuations in driving their hydrological responses.

   The elevation range of 4000 to 5000 meters is particularly significant because it is where snow typically accumulates during the pre-monsoon period and melts during the monsoon, directly contributing to river discharge. This altitude is still low enough to avoid the significant impact of the rain-shadow effect, which can limit precipitation at higher elevations. In contrast, at elevations between 5000 and 6000 meters, temperatures generally remain above the melting threshold, making these higher altitudes less relevant for predicting significant discharge events. Therefore, the 4000-5000 meter range is crucial for understanding and forecasting the hydrological behavior of these snow-fed rivers.

   Conversely, in areas where cumulative precipitation was the highest, the most reliable predictors of extreme events were found in the Bheri and Seti rivers. This observation is consistent with the hydrological nature of these rivers, which are primarily rain-fed, meaning that cumulative precipitation plays a more significant role in driving extreme discharge. These findings highlight the necessity of considering the distinct hydrological characteristics of each sub-basin when predicting extreme hydrological events in the KRB.

4. Can specific areas within the KRB be identified as indicator regions for extreme discharge events, and can the trained machine learning model generate probabilities of exceedance based on snow accumulation at certain elevations, combined with lagged temperature conditions prior to these events?

   To explore this, we calculated snow accumulation over time using the code developed in response to Research Question 1, focusing on the 48 delineated areas within the Upper Karnali for the snow-dominated scenario (years 2021, 1991, 2004, and 2007). We then examined snow accumulation in the 4000-5000 meter and 5000-6000 meter elevation ranges across these areas, analyzing how these accumulations correlated with downstream discharge from the Karnali River. This analysis was done in conjunction with ERA5 temperature data, specifically clipped for these individual areas, to assess their predictive power.

   The results revealed that five of these areas, which coincidentally had the highest snow accumulations, showed trends that closely matched the discharge patterns downstream. The machine learning model, trained to provide probabilities of exceedance, used average snow accumulation from 35 to 25 days prior to the event and average lagged temperature from 5 to 15 days before the event as inputs. With these factors, the model was able to predict with 90% certainty whether the 2000 m³/s discharge threshold would be exceeded downstream in the Karnali River. Furthermore, when these five predicting areas exceeded their individual snow accumulation thresholds, combined with a prolonged period of lagged temperature increase, the model's accuracy improved even further, correctly predicting in 96% of cases whether the 2000 m³/s threshold would be surpassed.

   These outcomes underscore the effectiveness of identifying specific sub-regions within the basin that can serve as strong predictors of extreme discharge events, offering valuable insights for flood forecasting and water resource management.

Final remarks conclusion

This study provides a nuanced understanding of the hydrological dynamics of the KRB under extreme weather conditions. By integrating the SPHY model with advanced machine learning techniques, the research detailed how snow accumulation and temperature trends influence river discharge, particularly in critical elevation ranges. The findings emphasize the importance of monitoring snow and temperature patterns, as these factors are significant drivers of hydrological extremes in the basin.

Future research should focus on refining these models, enhancing data resolution, and incorporating additional climatic variables to improve predictive reliability.

References

Akhtar, M., Ahmad, N., & Booij, M. (2008). The impact of climate change on the water resources of Hindukush–Karakorum–Himalaya region under different glacier coverage scenarios. *Journal of Hydrology*, *355*(1-4), 148–163. https://doi.org/10.1016/j.jhydrol.2008.03.015

Anders, A. M., Roe, G. H., Hallet, B., Montgomery, D. R., Finnegan, N. J., & Putkonen, J. (2006). Spatial patterns of precipitation and topography in the Himalaya. In *Tectonics, climate, and landscape evolution*. Geological Society of America. https://doi.org/10.1130/2006.2398(03)

Aryal, D., Wang, L., Adhikari, T. R., Zhou, J., Li, X., Shrestha, M., Wang, Y., & Chen, D. (2020). A Model-Based Flood Hazard Mapping on the Southern Slope of Himalaya. *Water*, *12*(2), 540. https://doi.org/10.3390/w12020540

Bhutiyani, M. R., Kale, V. S., & Pawar, N. J. (2007). Long-term trends in maximum, minimum and mean annual air temperatures across the Northwestern Himalaya during the twentieth century. *Climatic Change*, *85*(1-2), 159–177. https://doi.org/10.1007/s10584-006-9196-1

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bolch, T., Pieczonka, T., & Benn, D. I. (2011). Multi-decadal mass loss of glaciers in the Everest area (Nepal Himalaya) derived from stereo imagery. *The Cryosphere*, *5*(2), 349–358. https://doi.org/10.5194/tc-5-349-2011

Bookhagen, B., & Burbank, D. W. (2006). Topography, relief, and TRMM-derived rainfall variations along the Himalaya. *Geophysical Research Letters*, *33*(8). https://doi.org/10.1029/2006GL026037

Braun, L. (1993). *Application of a Conceptual Precipitation Runoff Model in the* (tech. rep.). https://www.researchgate.net/publication/242567643

Budhathoki, B. R., Adhikari, T. R., Shrestha, S., & Awasthi, R. P. (2023). Application of hydrological model to simulate streamflow contribution on water balance in Himalaya river basin, Nepal. *Frontiers in Earth Science*, *11*. https://doi.org/10.3389/feart.2023.1128959

Clemenzi, I., Gustafsson, D., Marchand, W.-D., Norell, B., Zhang, J., Pettersson, R., & Allan Pohjola, V. (2023). Impact of snow distribution modelling for runoff predictions. *Hydrology Research*, *54*(5), 633–647. https://doi.org/10.2166/nh.2023.043

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In *Ensemble machine learning* (pp. 157–175). Springer New York. https://doi.org/10.1007/978-1-4419-9326-7{_}5

Dahal, P., Shrestha, M. L., Panthi, J., & Pradhananga, D. (2020). Modeling the future impacts of climate change on water availability in the Karnali River Basin of Nepal Himalaya. *Environmental Research*, *185*. https://doi.org/10.1016/j.envres.2020.109430

Eklabya P. (2009). *Climate Change Impacts and Vulnerability in the Eastern Himalayas Contents* (tech. rep.).

Garbrecht, J. D., & Schneider, J. M. (2008). Case Study of Multiyear Precipitation Variations and the Hydrology of Fort Cobb Reservoir. *Journal of Hydrologic Engineering*, *13*(2), 64–70. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:2(64)

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., & Xavier, P. K. (2006). Increasing Trend of Extreme Rain Events Over India in a Warming Environment. *Science*, *314*(5804), 1442–1445. https://doi.org/10.1126/science.1132027

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*(1-4), 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9

Immerzeel, W. (2010). Climate Change Will Effect the Asian Water Towers. *Science*, *328*(5984), 1379–1382. https://doi.org/10.1126/science.1187443

Jain, C. K., & Singh, S. (2020). Impact of climate change on the hydrological dynamics of River Ganga, India. *Journal of Water and Climate Change*, *11*(1), 274–290. https://doi.org/10.2166/wcc.2018.029

Kang, K., & Lee, J. H. (2014). Hydrologic modelling of the effect of snowmelt and temperature on a mountainous watershed. *Journal of Earth System Science*, *123*(4), 705–713. https://doi.org/10.1007/s12040-014-0423-2

Kennard, M. J., Olden, J. D., Arthington, A. H., Pusey, B. J., & Poff, N. L. (2007). Multiscale effects of flow regime and habitat and their interaction on fish assemblage structure in eastern Australia. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(10), 1346–1359. https://doi.org/10.1139/f07-108

Khatiwada, K. R., & Pandey, V. P. (2019). Characterization of hydro-meteorological drought in Nepal Himalaya: A case of Karnali River Basin. *Weather and Climate Extremes*, *26*. https://doi.org/10.1016/j.wace.2019.100239

Khatiwada, K. R., Panthi, J., Shrestha, M. L., & Nepal, S. (2016). Hydro-climatic variability in the Karnali River Basin of Nepal Himalaya. *Climate*, *4*(2). https://doi.org/10.3390/cli4020017

Kienzle, S. W. (2008). A new temperature based method to separate rain and snow. *Hydrological Processes*, *22*(26), 5067–5085. https://doi.org/10.1002/hyp.7131

Lamichhane, M., Phuyal, S., Mahato, R., Shrestha, A., Pudasaini, U., Lama, S. D., Chapagain, A. R., Mehan, S., & Neupane, D. (2024). Assessing Climate Change Impacts on Streamflow and Baseflow in the Karnali River Basin, Nepal: A CMIP6 Multi-Model Ensemble Approach Using SWAT and Web-Based Hydrograph Analysis Tool. *Sustainability*, *16*(8), 3262. https://doi.org/10.3390/su16083262

Langat, P. K., Kumar, L., Koech, R., & Ghosh, M. K. (2019). Hydro-Morphological Characteristics Using Flow Duration Curve, Historical Data and Remote Sensing: Effects of Land Use and Climate. *Water*, *11*(2), 309. https://doi.org/10.3390/w11020309

Lee, J., Choi, J., Seo, J., Won, J., & Kim, S. (2023). Exploring Climate Sensitivity in Hydrological Model Calibration. *Water*, *15*(23), 4094. https://doi.org/10.3390/w15234094

Liucci, L., Valigi, D., & Casadei, S. (2014). A New Application of Flow Duration Curve (FDC) in Designing Run-of-River Power Plants. *Water Resources Management*, *28*(3), 881–895. https://doi.org/10.1007/s11269-014-0523-4

Martinet, J., Rango, A., & Major, E. (2008). *The Snowmelt-Runoff Model (SRM) User's Manual* (tech. rep.).

Nepal, S., & Shrestha, A. B. (2015). Impact of climate change on the hydrological regime of the Indus, Ganges and Brahmaputra river basins: a review of the literature. *International Journal of Water Resources Development*, *31*(2), 201–218. https://doi.org/10.1080/07900627.2015.1030494

Ougahi. (2024). Combining Hydrological Models and Remote Sensing to Characterize Snowpack Dynamics in High Mountains. *Remote Sensing*, *16*(2), 264. https://doi.org/10.3390/rs16020264

Painter, T. H., Deems, J. S., Belnap, J., Hamlet, A. F., Landry, C. C., & Udall, B. (2010). Response of Colorado River runoff to dust radiative forcing in snow. *Proceedings of the National Academy of Sciences*, *107*(40), 17125–17130. https://doi.org/10.1073/pnas.0913139107

Palazzi, E., Von Hardenberg, J., & Provenzale, A. (2013). Precipitation in the hindu-kush karakoram himalaya: Observations and future scenarios. *Journal of Geophysical Research Atmospheres*, *118*(1), 85–100. https://doi.org/10.1029/2012JD018697

Pandey, A., Parashar, D., Palni, S., Sarkar, M. S., Mishra, A. P., Singh, A. P., Costache, R., Abdulqadim, T. J., Pande, C. B., Tolche, A. D., & Khan, M. Y. A. (2024). Spatiotemporal snowline status and climate variability impact assessment: a case study of Pindari River Basin, Kumaun Himalaya, India. *Environmental Sciences Europe*, *36*(1), 104. https://doi.org/10.1186/s12302-024-00924-7

Pumo, D., Francipane, A., Cannarozzo, M., Antinoro, C., & Noto, L. V. (2018). Monthly Hydrological Indicators to Assess Possible Alterations on Rivers' Flow Regime. *Water Resources Management*, *32*(11), 3687–3706. https://doi.org/10.1007/s11269-018-2013-6

Rajagopal, S., & Harpold, A. A. (2016). Testing and Improving Temperature Thresholds for Snow and Rain Prediction in the Western United States. *JAWRA Journal of the American Water Resources Association*, *52*(5), 1142–1154. https://doi.org/10.1111/1752-1688.12443

Rasul, G. (2014). Food, water, and energy security in South Asia: A nexus perspective from the Hindu Kush Himalayan region□. *Environmental Science & Policy*, *39*, 35–48. https://doi.org/10.1016/j.envsci.2014.01.010

Ridolfi, E., Kumar, H., & Bárdossy, A. (2020). A methodology to estimate flow duration curves at partially ungauged basins. *Hydrology and Earth System Sciences*, *24*(4), 2043–2060. https://doi.org/10.5194/hess-24-2043-2020

Roxy, M. K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P., & Rajeevan, M. (2017). A threefold rise in widespread extreme rain events over central India. *Nature Communications*, *8*(1), 708. https://doi.org/10.1038/s41467-017-00744-9

Shrestha. (2000). *Interannual variation of summer monsoon rainfall over Nepal and its relation to Southern Oscillation Index* (tech. rep.).

Singh, P., & Bengtsson, L. (2004). Hydrological sensitivity of a large Himalayan basin to climate change. *Hydrological Processes*, *18*(13), 2363–2385. https://doi.org/10.1002/hyp.1468

Smakhtin, V. (2001). Low flow hydrology: a review. *Journal of Hydrology*, *240*(3-4), 147–186. https://doi.org/10.1016/S0022-1694(00)00340-1

Smith, Baeck, M. L., & Miller, A. J. (2015). Exploring storage and runoff generation processes for urban flooding through a physically based watershed model. *Water Resources Research*, *51*(3), 1552–1569. https://doi.org/10.1002/2014WR016085

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Treichler, D., Kääb, A., Salzmann, N., & Xu, C.-Y. (2019). Recent glacier and lake changes in High Mountain Asia and their relation to precipitation changes. *The Cryosphere*, *13*(11), 2977–3005. https://doi.org/10.5194/tc-13-2977-2019

Viviroli, Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*, *43*(7). https://doi.org/10.1029/2006WR005653

Viviroli & Weingartner. (2004). The hydrological significance of mountains: from regional to global scale. *Hydrology and Earth System Sciences*, *8*(6), 1017–1030. https://doi.org/10.5194/hess-8-1017-2004

Wen, L., Nagabhatla, N., Lü, S., & Wang, S.-Y. (2013). Impact of rain snow threshold temperature on snow depth simulation in land surface and regional atmospheric models. *Advances in Atmospheric Sciences*, *30*(5), 1449–1460. https://doi.org/10.1007/s00376-012-2192-7

Zomer, R. J., Trabucco, A., Bossio, D. A., & Verchot, L. V. (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment*, *126*(1-2), 67–80. https://doi.org/10.1016/j.agee.2008.01.014

8

Appendix

8.1. Appendix A: Supplement scenario analysis

Climatological baseline vs. Snow-dominated scenario

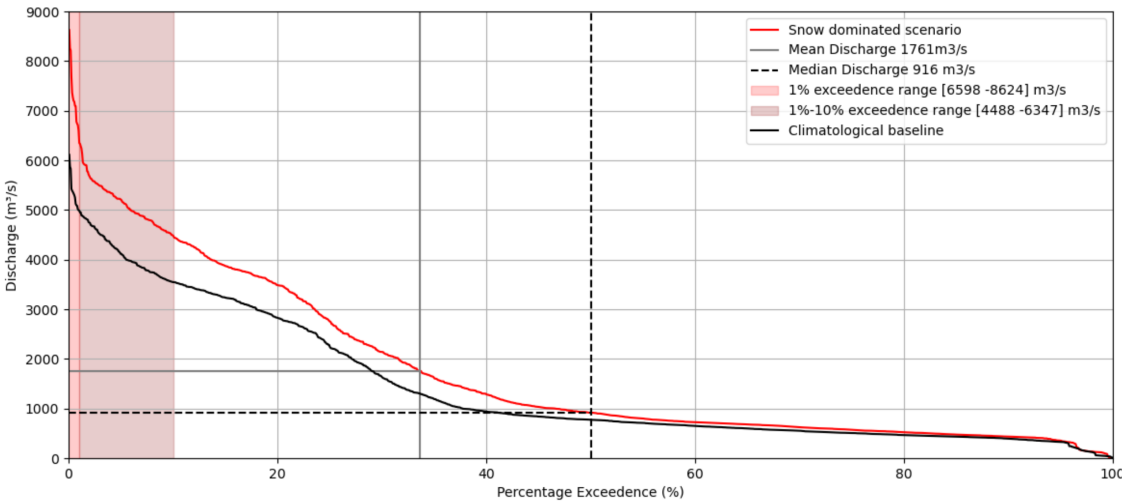


Figure 8.1: On the x-axis the percentage of exceedence versus discharge on the y-axis. The FDCis based on the outcome from the calibrated SPHY model while forces with ERA5 data resampled for the snow-dominated scenario.

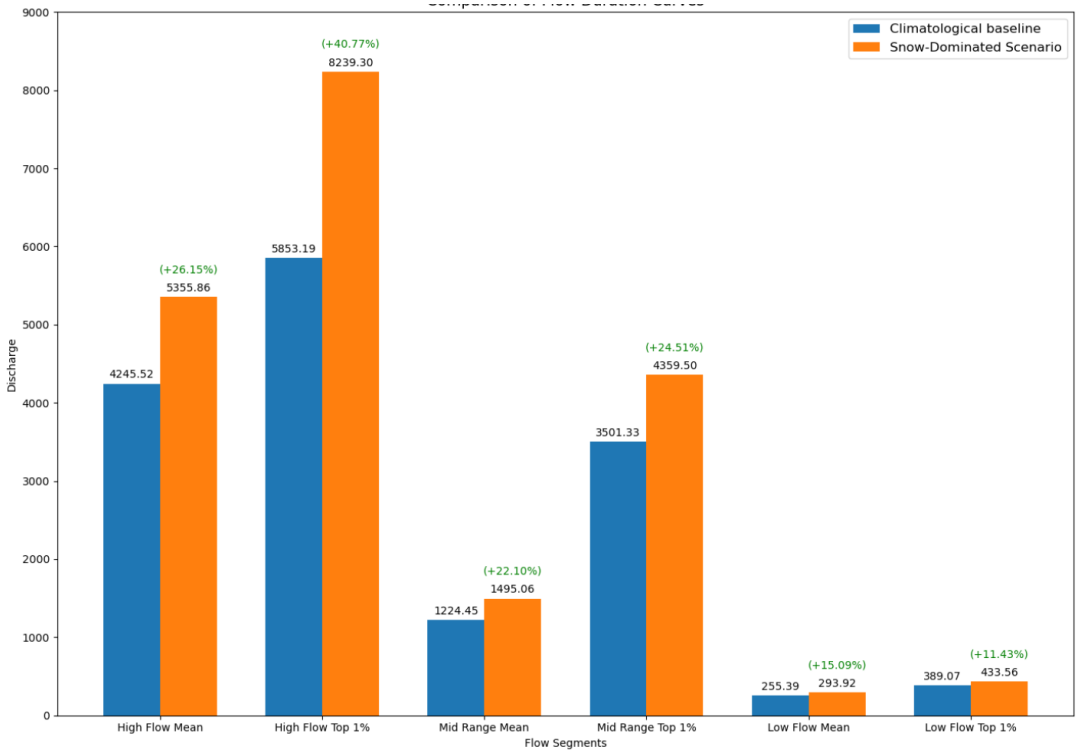


Figure 8.2: On the x-axis the different flow segments versus the discharge in m³/s in the y-axis. The comparison was made for the snow-dominated scenario versus the climatological baseline for the high/mid/low flows

1. **High Flow Segment (1-10% Exceedence Range)**
   - The high flow values in the snow-dominated scenario show a notable increase compared to the climatological baseline. There is a 26% increase in the high flow mean and a 41% increase in the high flow top 1%. This significant increase is attributed to the extensive snowmelt contributing to runoff, especially during warmer periods. High snow accumulation years lead to substantial runoff increases when temperatures rise, due to the large volumes of stored snow that melt and flow into rivers. This is supported by studies (Viviroli et al., 2007) and (Hock, 2003).
2. **Mid Range Segment (10-90% Exceedence Range)**
   - The mid-range flows also increase, with a 22% increase in the mid-range mean and a 25% increase in the mid-range top 1%. The snow-dominated scenario stabilizes the mid-range flow as the snow melt provides a consistent source of water. Snowmelt provides a reliable source of water that maintains streamflow during periods where precipitation alone would not be sufficient (Martinet et al., 2008).
3. **Low Flow Segment (90-100% Exceedence Range)**
   - There is an increase in the low flow values. The low flow mean increases by 15% and the low flow top 1% increases by 11%. This indicates that during snow-dominated years, the persistent melting of snow prevents very low discharge levels. The findings are supported by studies (Braun, 1993) that indicate snowmelt can maintain baseflow levels in rivers, reducing the frequency of low flow periods.

Climatological baseline vs. Rain dominated scenario

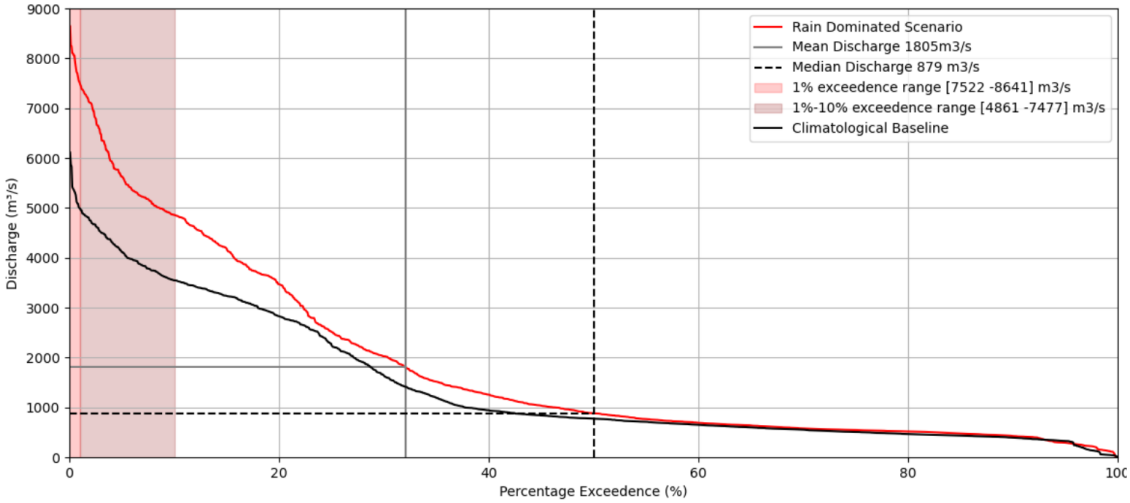


Figure 8.3: On the x-axis the percentage of exceedence versus discharge on the y-axis. The FDCis based on the outcome from the calibrated SPHY model while forces with ERA5 data resampled for the rain-dominated scenario.

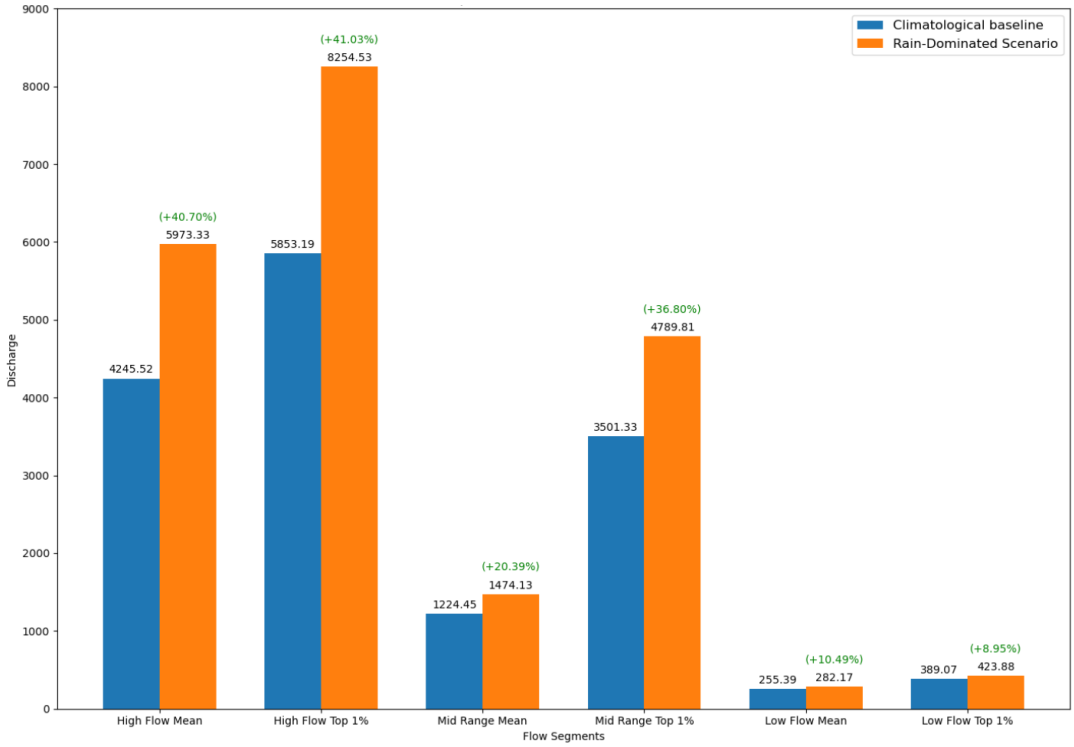


Figure 8.4: on the x-axis the different flow segments versus the discharge in m³/s in the y-axis. The comparison was made for the rain-dominated scenario versus the climatological baseline for the high/mid/low flows

1. **High Flow Segment (1-10% Exceedence Range)**

   - The high flow values in the rain-dominated scenario show a significant increase compared to the climatological baseline. There is a 41% increase in the high flow mean and a 41% increase in the high flow top 1%. This indicates that intense and frequent rainfall significantly boosts peak flow events, contributing to higher risks of flooding during these periods. Litera-

ture is (Ougahi, 2024) confirming that high rainfall volumes and intensities during monsoon seasons amplify runoff, leading to extreme discharge events.

2. **Mid Range Segment (10-90% Exceedence Range)**

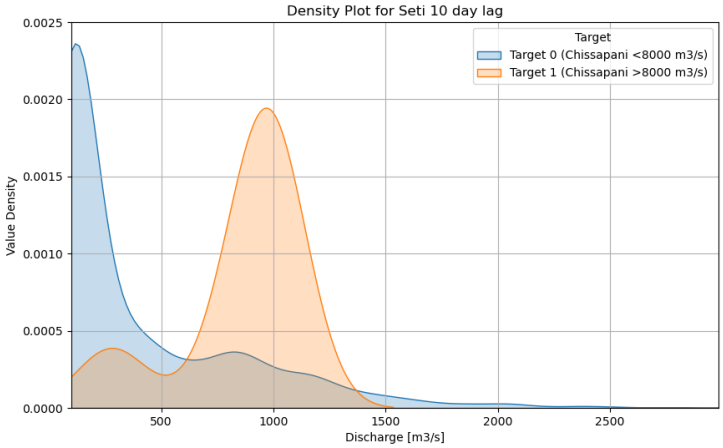
    - The mid-range flows exhibit a 20% increase in the mid-range mean and a 37% increase in the mid-range top 1%, reflecting the consistent impact of heavy rainfalls on river discharge. This suggests that during rain-dominated years, rivers maintain higher than usual flow rates, which can support water supply needs but also indicate a shift in river regime. Studies (Clemenzi et al., 2023) supports the finding that sustained heavy rainfall impacts typical flow conditions, often leading to higher baseline flows.

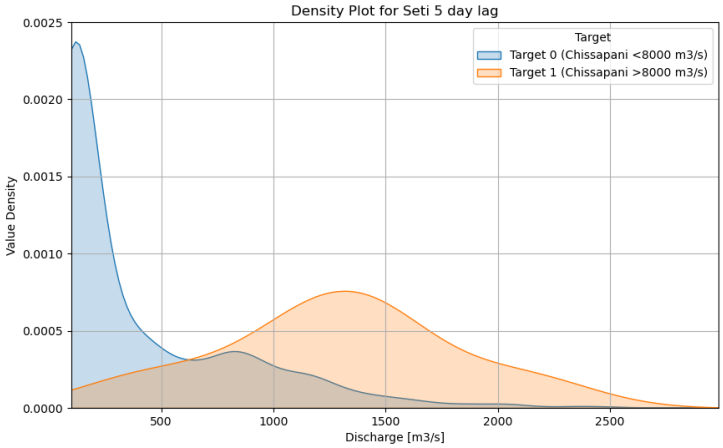
3. **Low Flow Segment (90-100% Exceedence Range)**

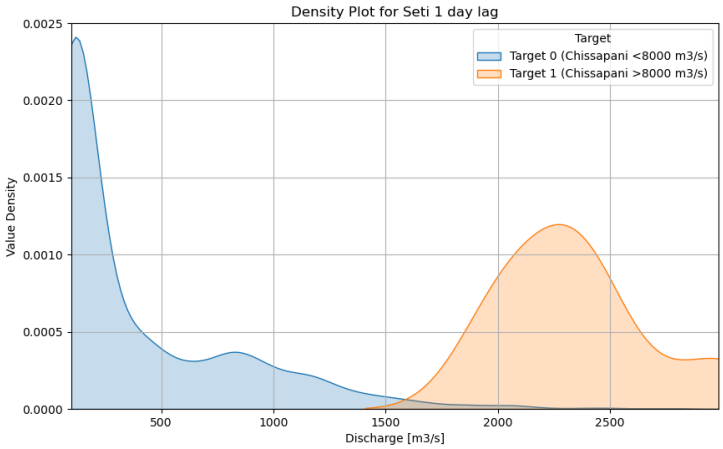
    - The low flow values in the rain-dominated scenario also increase, with a 10% increase in the low flow mean and an 9% increase in the low flow top 1%. This increase suggests that the constant input from rainfall minimizes periods of low discharge, enhancing water availability during typically dry periods. Studies indicate that prolonged and intense rainfall periods reduce the frequency and severity of low flows (Viviroli et al., 2007).

8.2. Appendix B: supplement RandomForestClassifier

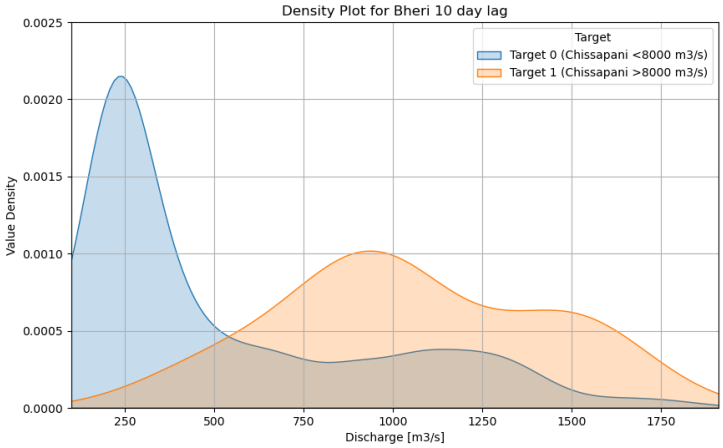
Density plot feature with highest deviation [Seti]

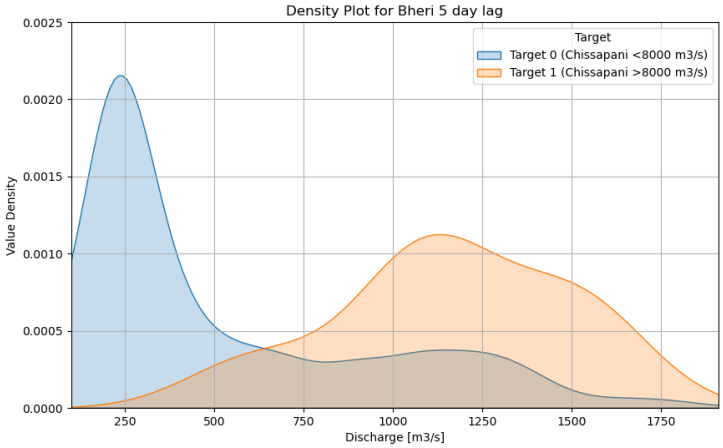


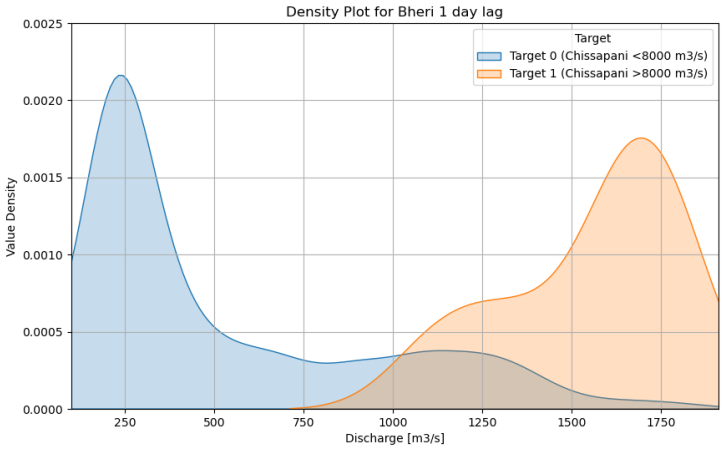




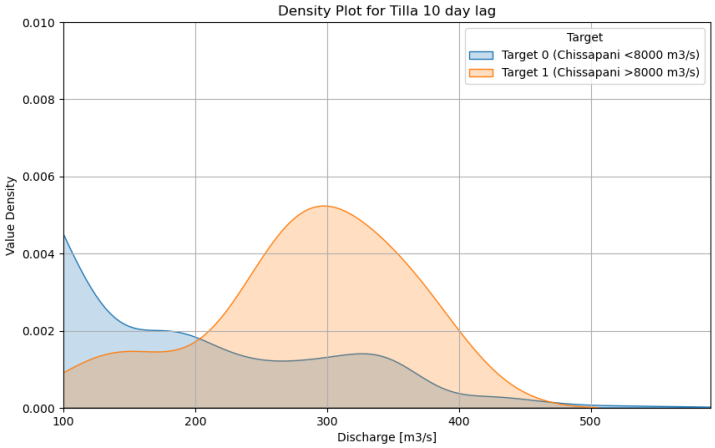
Density plot [Bheri]

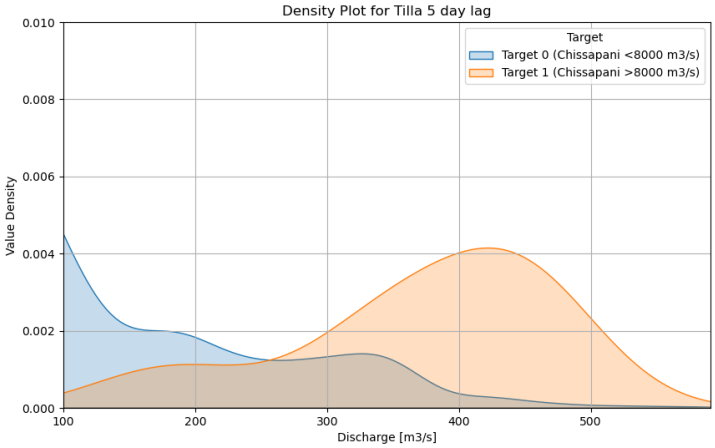


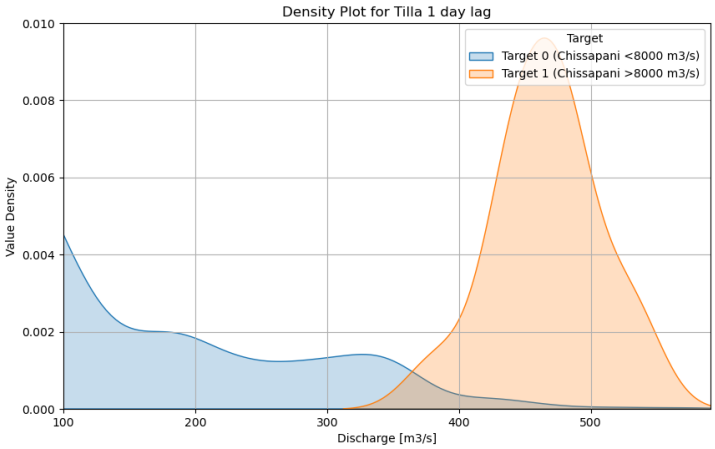




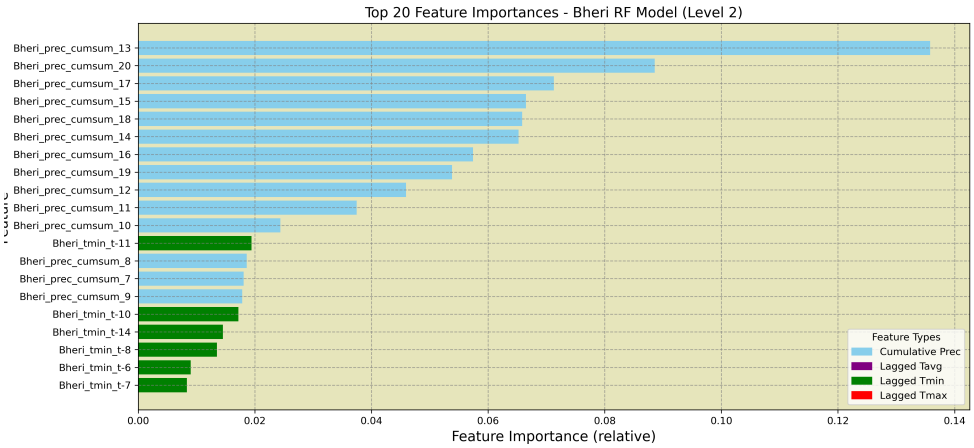
Density plot [Tilla]



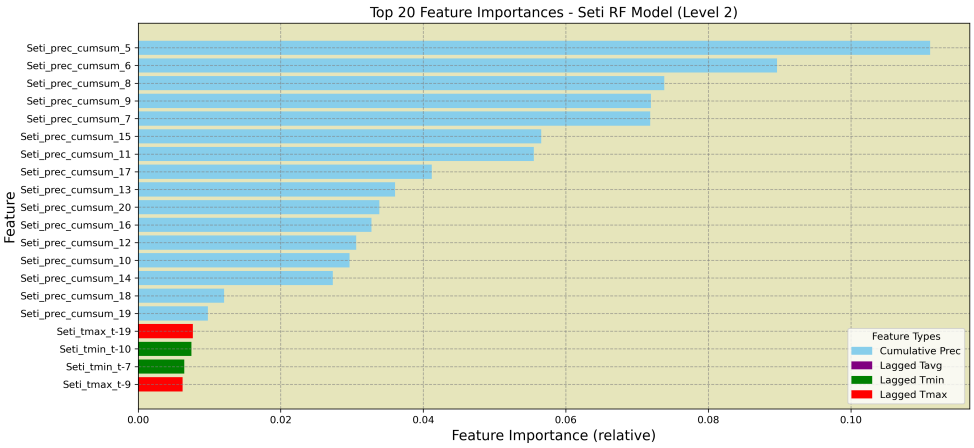




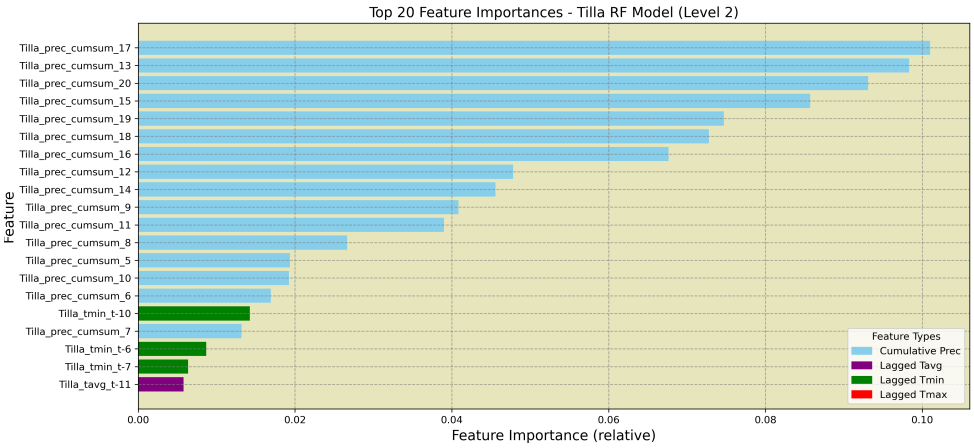
Sub-basin: Bheri



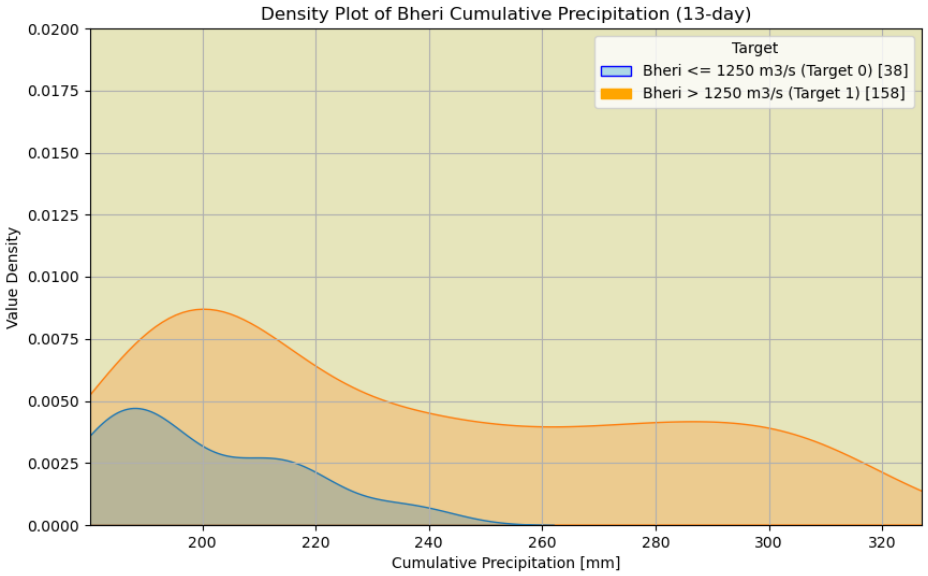
Sub-basin: Seti



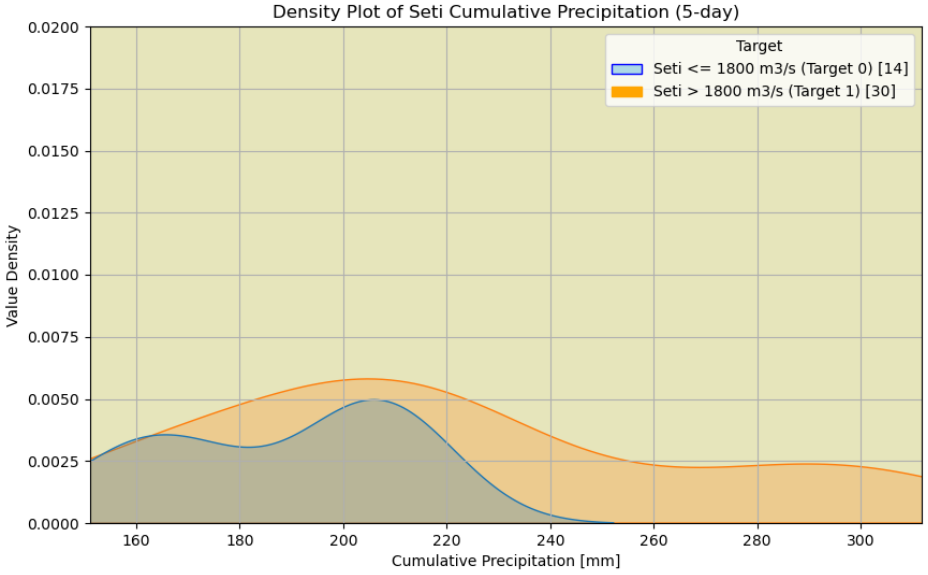
Sub-basin: Tilla



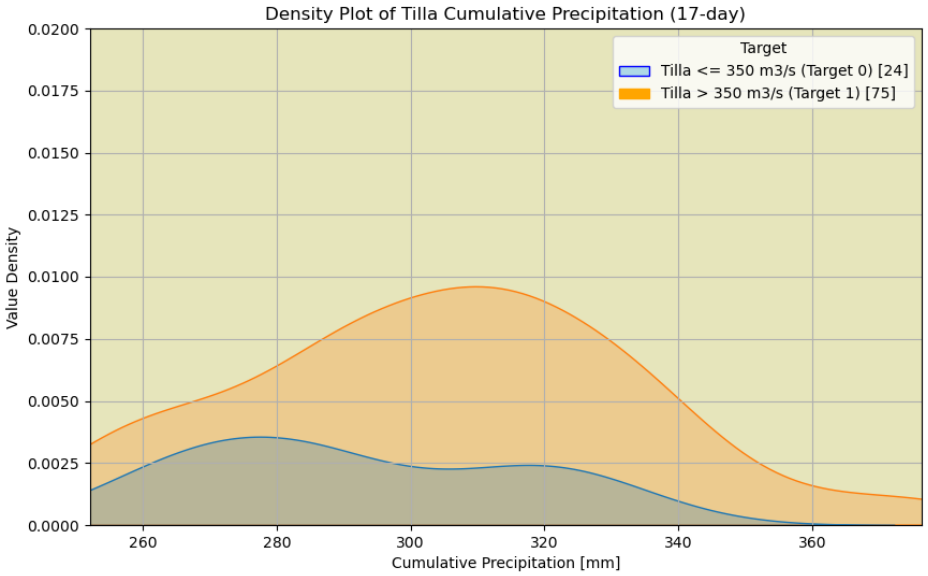
Density plot Bheri



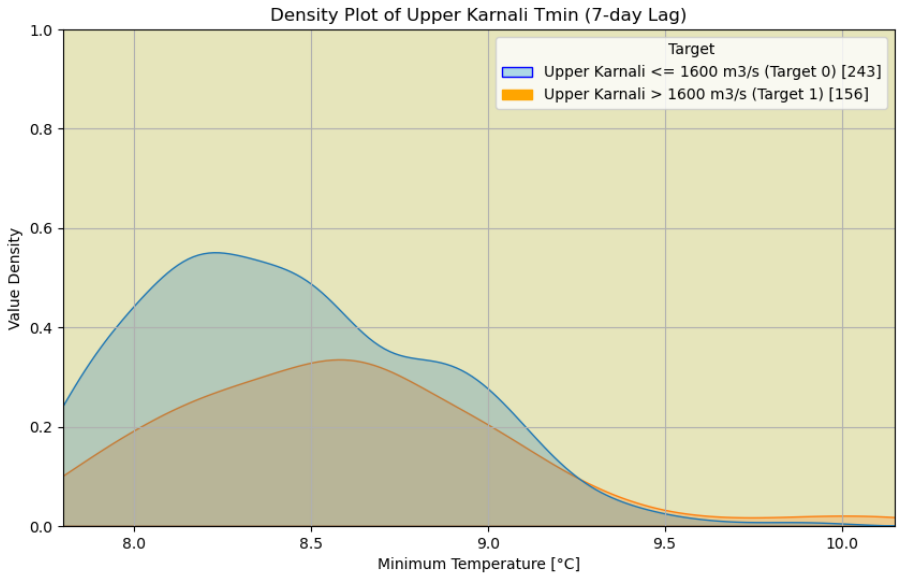
Density plot Seti

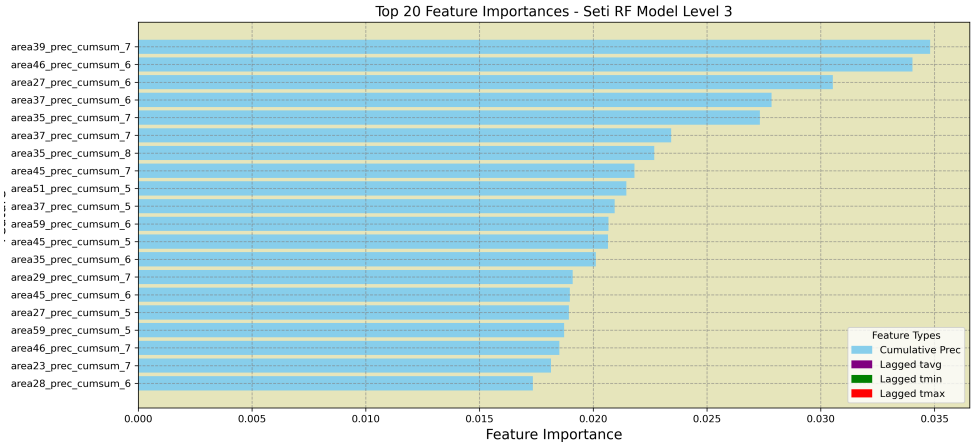


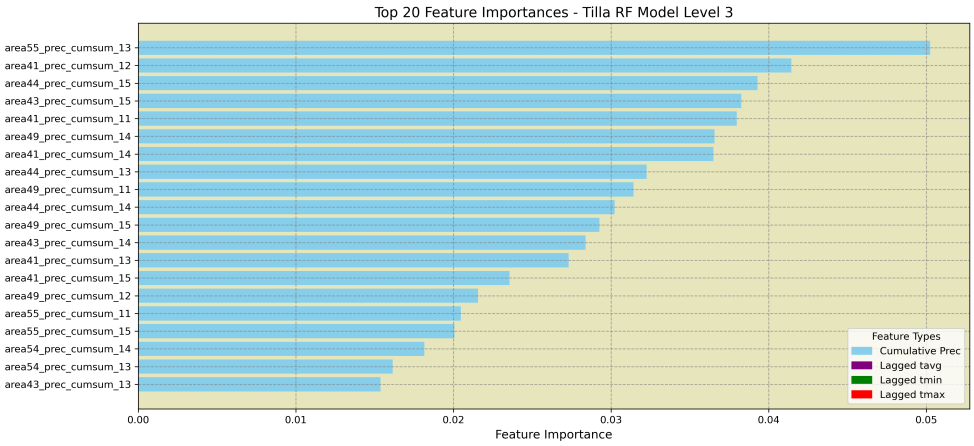
Density plot Tilla

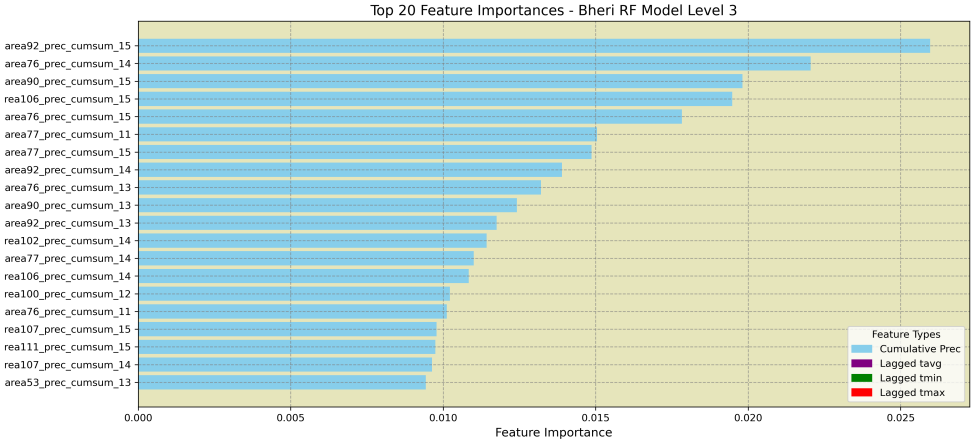


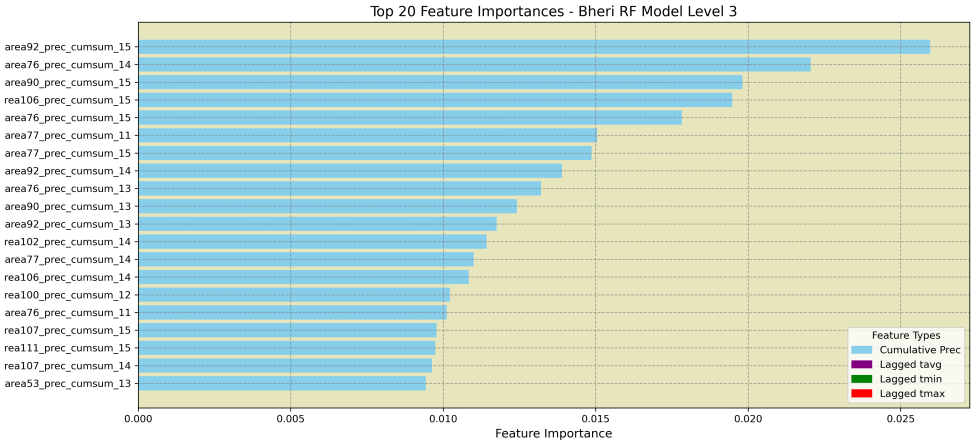
Density plot Upper Karnali




Top 20 Feature Importances - Seti RF Model Level 3
area39_prec_cumsum_7
area46_prec_cumsum_6
area27_prec_cumsum_6
area37_prec_cumsum_6
area35_prec_cumsum_7
area37_prec_cumsum_7
area35_prec_cumsum_8
area45_prec_cumsum_7
area51_prec_cumsum_5
area37_prec_cumsum_5
area59_prec_cumsum_6
area45_prec_cumsum_5
area35_prec_cumsum_6
area29_prec_cumsum_7
area45_prec_cumsum_6
area27_prec_cumsum_5
area59_prec_cumsum_5
area46_prec_cumsum_7
area23_prec_cumsum_7
area28_prec_cumsum_6
Feature Types
Cumulative Prec
Lagged tavg
Lagged tmin
Lagged tmax
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
Feature Importance


Top 20 Feature Importances - Tilla RF Model Level 3
area55_prec_cumsum_13
area41_prec_cumsum_12
area44_prec_cumsum_15
area43_prec_cumsum_15
area41_prec_cumsum_11
area49_prec_cumsum_14
area41_prec_cumsum_14
area44_prec_cumsum_13
area49_prec_cumsum_11
area44_prec_cumsum_14
area49_prec_cumsum_15
area43_prec_cumsum_14
area41_prec_cumsum_13
area41_prec_cumsum_15
area49_prec_cumsum_12
area55_prec_cumsum_11
area55_prec_cumsum_15
area54_prec_cumsum_14
area54_prec_cumsum_13
area43_prec_cumsum_13
Feature Types
Cumulative Prec
Lagged tavg
Lagged tmin
Lagged tmax
0.00
0.01
0.02
0.03
0.04
0.05
Feature Importance


Top 20 Feature Importances - Bheri RF Model Level 3
area92_prec_cumsum_15
area76_prec_cumsum_14
area90_prec_cumsum_15
rea106_prec_cumsum_15
area76_prec_cumsum_15
area77_prec_cumsum_11
area77_prec_cumsum_15
area92_prec_cumsum_14
area76_prec_cumsum_13
area90_prec_cumsum_13
area92_prec_cumsum_13
rea102_prec_cumsum_14
area77_prec_cumsum_14
rea106_prec_cumsum_14
rea100_prec_cumsum_12
area76_prec_cumsum_11
rea107_prec_cumsum_15
rea111_prec_cumsum_15
rea107_prec_cumsum_14
area53_prec_cumsum_13
Feature Types
Cumulative Prec
Lagged tavg
Lagged tmin
Lagged tmax
0.000
0.005
0.010
0.015
0.020
0.025
Feature Importance



Having completed step 1.1, where discharge data for the Bheri, Tilla, Seti, and Upper Karnali sub-basins was gathered, generated by the SPHY model (forced with snow and rain-dominated scenario datasets), the next step involves analyzing this data using the RandomForestClassifier. The SPHY model was run with various snow and rain scenarios to simulate high precipitation events. This provided a diverse set of discharge data under different climatic conditions. With this dataset, the RandomForestClassifier was used to identify the lag periods (ranging from 5 to 20 days) that best correlate with patterns observed when discharge at Chisapani exceeds 8000 m³/s.

The RandomForestClassifier is reported to be effective in determining which lagged discharges are most significant in predicting high flow events. By analyzing the discharge data with lag periods between 5 and 20 days, the classifier identified the specific lags that follow the discharge patterns leading to extreme events at Chisapani. This analysis revealed the temporal patterns and lag days that show the strongest correlation with high discharge events. The feature importance analysis, illustrated in figure 2.15 , highlights the top 10 lagged discharge features. Notably, the Tilla sub-basin shows the highest feature importance with lags of 15 and 14 days being the most significant. Other important features include lagged discharges from the Upper Karnali and Bheri sub-basins, with lag periods of 16, 15, and 14 days demonstrating substantial importance. The Seti sub-basin also features significantly with lags of 16 and 15 days.

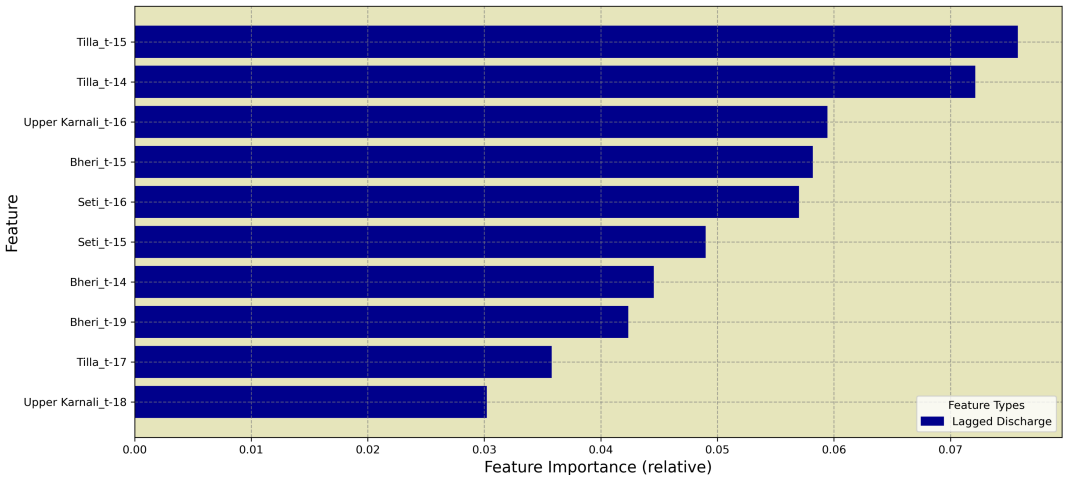


Figure 8.5: The most importance feature predictors used by the trained RandomForestClassifier machine learning model.

Feature importance in a RandomForestClassifier shows how useful each feature is in making accurate predictions. The values on the x-axis indicate how much each feature helps the model. Higher

values mean that the feature is very important and greatly improves the model's predictions. In simpler terms, the more a feature helps the model make correct predictions, the higher its importance score will be.

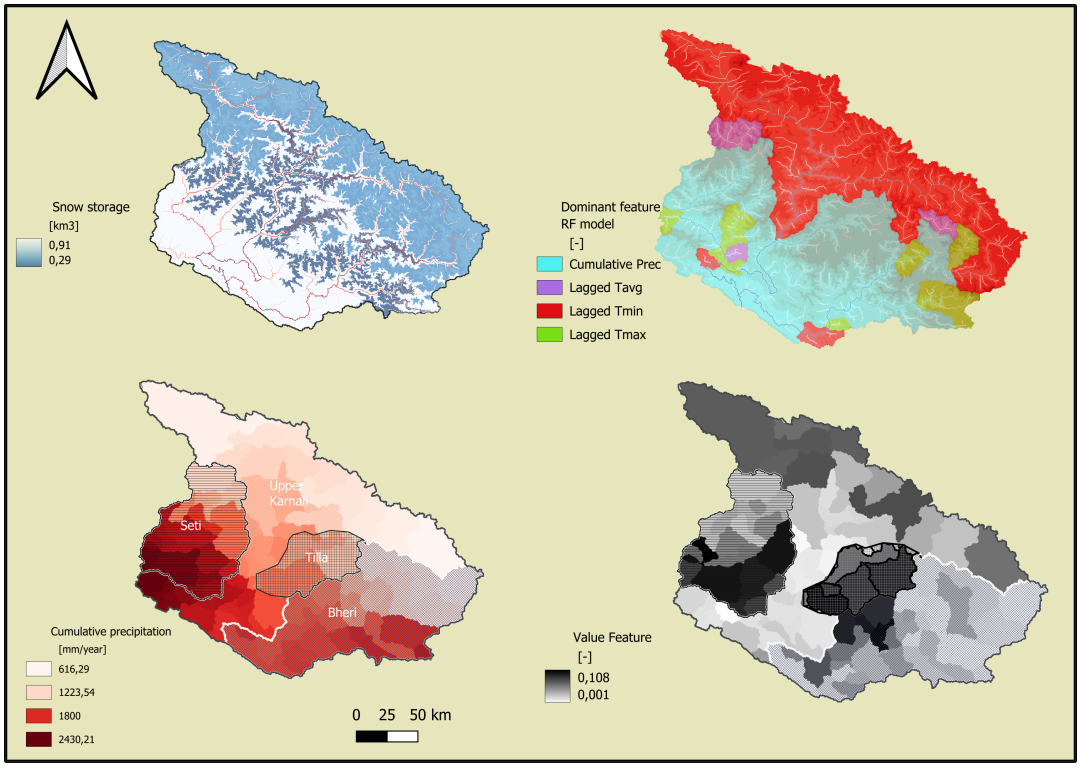
This importance is often measured by the decrease in impurity. A decrease in impurity refers to how well a feature splits the data into correct categories. In decision trees, impurity measures how mixed the data is in each node, with common measures being entropy and Gini impurity. When a feature is used to split the data, it reduces impurity by grouping similar data points together, making the nodes purer. The more a feature reduces impurity, the more important it is. Even though a feature might be very important overall, it doesn't necessarily appear at the top of any single decision tree in the forest. Instead, it can contribute to the prediction in many smaller ways across lots of trees. In a random forest, hundreds of decision trees are used, and each one might use the feature differently. A feature might not be the top splitter in any tree, but if it helps improve predictions a little bit in many trees, it ends up being very important overall. This means the feature is valuable because it consistently helps make better decisions across the entire forest, even if it doesn't stand out in just one tree.

Table 8.1: The lowest mean and standard deviations from the different contributing sub-basins while predicting events at Chisapani above 8000 m^3/s

River	Lag Day	Mean (m^3/s)	Standard Deviation (m^3/s)
Upper Karnali	5	1773	242
Seti	18	1162	588
Bheri	7	1248	341
Tilla	5	405	87

For the lag days listed below, the mean and standard deviations will be further observed to determine if any visual patterns can be represented. Specifically, the focus will be on the 10-day lag, 5-day lag, and 1-day lag to see if consistent patterns emerge.

8.3. Appendix C: Supllement prediction model



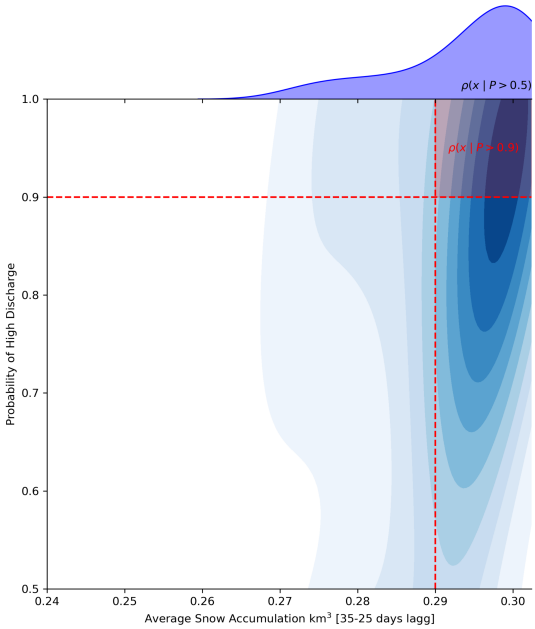


Figure 8.6: Area 2, 4000-5000 [m]: In the 4000-5000 m range, the probability of exceeding the 2000 m³/s discharge threshold surpasses 90% when snow accumulation reaches approximately 0.29 km³. The steep rise in probability and the peak in the top curve indicate a strong relationship between snow accumulation and high discharge events in this range.

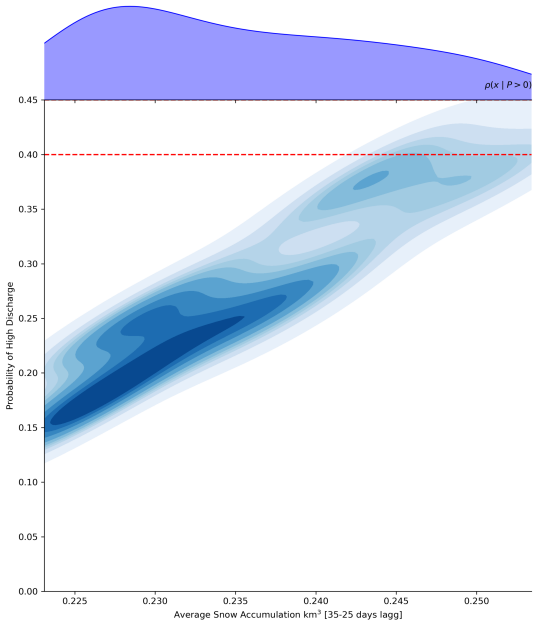


Figure 8.7: Area 2, 5000-6000 [m]: In the 5000-6000 m range, although there is a correlation between snow accumulation and discharge probability, the effect is less pronounced. The probability does not exceed 40%, as reflected in the flatter top curve, indicating a weaker predictive power for snow accumulation at these higher elevations..

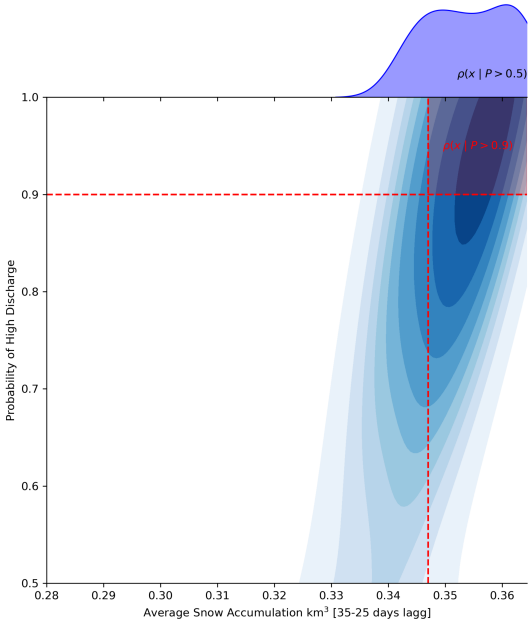


Figure 8.8: Area 5, 4000-5000 [m]: In the 4000-5000 m range, the figure reveals a strong link between snow accumulation and the likelihood of high discharge events. As snow accumulation reaches around 0.35 km³, the probability of surpassing the 2000 m³/s threshold approaches 90%. The sharp peak in the top curve emphasizes this robust predictive relationship.

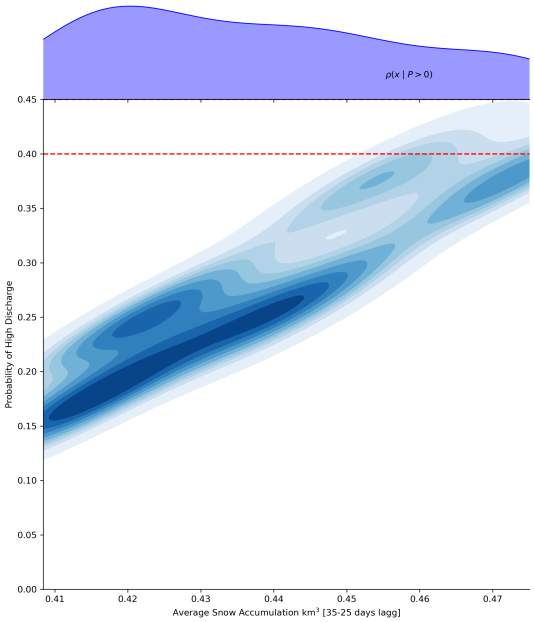


Figure 8.9: Area 5, 5000-6000 [m]: In contrast, for the 4000-5000 m range, snow accumulation shows a much weaker correlation with high discharge events. The probabilities remain low, with the top curve displaying a flat profile, indicating that snow accumulation at these higher elevations is not a reliable predictor of extreme discharge.

Enriching machine learning model with observed patterns

We can observe that the majority of the melt primarily occurs in the elevation range of 4000-5000 meters. Consequently, we further conducted the analysis to see if we could create a probability density function using the trained machine learning model. This involved running the model through a series of time sequences and evaluating its predictions for high discharge events.

The inputs for the model included:

- **Snow Accumulation**: The average snow accumulation from the five indicator regions 25-35 days prior to the event, in both the 4000-5000 meter and 5000-6000 meter ranges.
- **Lagged Temperature**: The average temperature 10-20 days prior to the event.

Based on these inputs, the model predicts the probability of exceeding the discharge threshold at outlet F. This approach allows us to understand and predict the likelihood of extreme discharge events based on historical and real-time data inputs.