VERIFICATION OF A GLOBAL STREAMFLOW FORECAST FOR THE PURPOSE OF RUN-OF-RIVER HYDROPOWER OPERATION IN NEPAL

by

Abhinab Kadel

B.Eng., Carl von Ossietzky University of Oldenburg, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Atmospheric Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

Verification of a Global Streamflow Forecast for the Purpose of Run-of-River Hydropower Operation in Nepal

submitted by Abhinab Kadel in partial fulfillment of the requirements for

the degree of Master of Science

in Atmospheric Science

Examining Committee:

Prof. Roland Stull, Atmospheric Sciences, UBC
Supervisor

Dr. Douglas McCollor, Atmospheric Sciences, UBC
Supervisory Committee Member

Prof. Werner Antweiler, Sauder School of Business, UBC
Supervisory Committee Member

Dr. Sean Fleming, National Water and Climate Center, US Department of Agriculture
External Examiner

Abstract

Most developing countries lack the resources required to develop and maintain their own nationwide fine-resolution streamflow forecasting systems, which has led to the emergence of global streamflow forecasting methods with coarser resolution. The GEOGloWS-ECMWF Streamflow Service (GESS) is a global forecasting model that uses runoff output from the European Center for Medium-range Weather Forecasts (ECMWF) and uses the vector-based routing tool called RAPID to produce 52-member ensemble forecasts 15 days into the future at smaller catchments worldwide.

This thesis examines the quality of GESS streamflow forecasts over the mountainous terrains of Nepal for two years, 2014 and 2015, focusing on run-of-river (ROR) hydropower operation. A reforecast dataset is used for the forecasts and is compared with streamflow observations at five different sites with existing hydropower facilities. The forecasts are verified using verification metrics such as bias, flow variability, correlation, Kling-Gupta efficiency, the Nash-Sutcliffe efficiency, and the Continuous Ranked Probability Score. The verification is performed across two flow seasons: wet and dry, distinguished by the 70^{th} percentile of climatological flow.

First, the raw forecasts are verified. The results show an overall poor performance of the forecasts. Second, a simple moving window multiplicative bias correction approach called the Degree of Mass Balance (DMB) is tested. The 2014 year is set aside as the calibration year to calculate the best bias correction approach, such as the window length and the best DMB formulation. The best DMB configuration for each site and forecast horizon is then tested in the independent verification year of 2015. The bias-corrected forecasts show much-improved performance in all metrics.

Finally, the bias-corrected GESS forecasts are evaluated for two use-cases commonly faced by ROR hydropower operators in Nepal: flood forecasting in the wet season and energy generation forecasting in the dry season. The GESS forecasts raised more false alarms and would not have predicted at least half of the flood events in the sites studied. Furthermore, the forecasts did not yield more revenue than a simple persistence forecast. Thus, there is a need to improve the forecasts before they can add real value to the ROR operators.

Lay Summary

Streamflow forecasts provide future information on water flow in a river, which is valuable for the operation of hydropower systems and flood-damage mitigation. This thesis examines the quality of a global streamflow forecasting model called the GeoGLOWS ECMWF Streamflow Service (GESS). GESS was developed to help forecast damaging floods worldwide, and it is the only system that provides forecasts for Nepal's relatively smaller streams.

The specific focus is on hydropower operations in Nepal, which are vulnerable to flooding events and could benefit from streamflow forecasts for their regular operation. First, I investigate the general quality of the forecasts for sites with hydropower installations in Nepal. Second, I investigate how well the forecasts predict damaging floods. Finally, I examine if the forecasts add value when used to predict future electricity generation. I found that the GESS streamflow forecasts were inaccurate for the streams studied here.

Preface

The thesis is the original, unpublished work of the author Abhinab Kadel. The idea for the thesis topic and the research questions were devised by Abhinab Kadel, on advice from Prof. Roland Stull. The reforecast dataset used in the thesis was created and made available by Jorge Luis Sanchez-Lozano, from Brigham Young University. The calibration options for the optimum bias correction approach were determined in consultation with Prof. Roland Stull. The idea of value assessment by comparing with the no future information and perfect future information was suggested by Prof. Werner Antweiler.

Table of Contents

List of Tables

List of Figures

List of Symbols

γ	:	flow variability
$f_k\ (o_k)$	:	k^{th} forecast (observation)
g	:	acceleration due to gravity
E_d	:	Bid energy
$\bar{o}$	:	climatological mean observation
r	:	correlation coefficient
ρ	:	density of water
E_g	:	generated energy
t_o	:	hours of operation
$\mu_f\ (\mu_o)$	:	mean forecast (observation)
h	:	net head of the ROR installation
R_{PPA}	:	PPA rate
$P_f(k)\ [P_o(k)]$	:	Probability distribution of forecasts [observation]
η_p	:	ROR plant efficiency
$\sigma_f\ (\sigma_o)$	:	standard deviation of forecast (observation)
β	:	total bias
N	:	total number of forecast-observation days
Q_{in}	:	volumetric flow of water entering the turbine
w_k	:	weights for bias correction

List of Abbreviations

AD	:	Availability Declaration
CRPS	:	Continuous Ranked Probability Score
DHM	:	Department of Hydrology and Meteorology
DMB	:	Degree of Mass Balance
ECMWF	:	European Centre for Medium-Range Weather Forecasts
e-flow	:	environmental flow
FAR	:	False Alarm Ratio
GESS	:	GEOGloWS-ECMWF Streamflow Service
GloFAS	:	Global Flood Awareness System
GWh	:	Gigawatt-hours
HTESSEL	:	Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land
IEX	:	Indian Energy Exchange
INR	:	Indian Rupees
IPP	:	Independent Power Producer
KGE	:	Kling-Gupta Efficiency
LDMB	:	Linearly-weighted Degree of Mass Balance
LSM	:	Land Surface Model
MHP	:	Marsyangdi Hydropower Project
MMHP	:	Middle Marsyangdi Hydropower Project
MSE	:	Mean Squared Error
MW	:	Megawatts
NEA	:	Nepal Electricity Authority
NPR	:	Nepali Rupees
NSE	:	Nash-Sutcliffe Efficiency
NWP	:	Numerical Weather Prediction
PPA	:	Power Purchase Agreements

RAPID	:	Routing Application for Parallel computation of Discharge
ROR	:	Run-of-River
UMHP	:	Upper Marsyangdi Hydropower Project
USD	:	United States Dollar
WRF	:	Weather Research and Forecasting model

Acknowledgements

First, I would like to express my endless gratitude to my supervisor Prof. Roland Stull for his continuous support and supervision throughout the thesis. He was always available to meet upon short notice and provided valuable advice to this thesis work. Prof. Stull was extremely helpful in helping me catch up in the field of meteorology, which was entirely new for me. Furthermore, he was very understanding and accommodating during the tough Covid-19 transitions.

My committee members: Prof. Werner Antweiler and Dr. Doug McCollor, deserve special thanks. My discussions with Prof. Antweiler were instrumental in finalizing the analysis method for the value assessment study done in the thesis.

Thank you to the different funding agencies, research grants, and scholarships that assisted my study: Mitacs, NSERC, BC Hydro, Chih-Chuang, and Yien-Ying Wang Hsieh Memorial Scholarship, and the UBC International Tuition Award. Also, thanks to the Hydroinformatics team at the Brigham Young University, the Department of Hydrology and Meteorology, Nepal, and the Nepal Hydropower Portal for providing the reforecast, observation, and the hydropower locations data, respectively.

My thanks to my team members at the Weather Forecast and Research Team, especially Tim Chui, for being available (literally) any hour of the day to answer my computing-related questions, and to Chris Rodell for being available for adventures outside of work.

I am grateful for the other friends I had within and outside the Earth, Ocean, and Atmospheric Sciences and to my regular badminton buddies, who were available for a good racquet session every once in a while.

Finally, my special thanks to my loving parents, sister, and dearest friends worldwide. Their support, success, and inspirational stories drive me to move forward.

Dedication

To my loving parents, warm, sunny weather, and the many more experiences that life has to offer…

Chapter 1: Introduction

Open-channel systems such as rivers and streams are the most widespread source of freshwater for humans. Freshwater is a vital component for the survival and proper functioning of the human body. Easy water access was a primary motivation for early human settlements' proximity to open-channel systems. While humans could enjoy the fertile flood plains and reliable drinking water sources, the proximity to rivers made these settlements prone to seasonal flooding events. The initial motivation for streamflow forecasting was to have skillful information on future states of water flow to help mitigate the disastrous effect of the flooding events (Roundy, Duan, & Schaake, 2019). As humans have evolved, they have discovered newer applications of water resources and open-channel systems: irrigation, transportation, water mills, and recently renewable energy extraction. Modern streamflow forecasting approaches are designed to aid efficient water management in these sectors and assist in disaster mitigation.

Generating high-quality streamflow forecasts is an intensive task that relies on a dense hydrometeorological observation network, computing resources, and a trained workforce to maintain the forecasting systems (Souffront Alcantara et al., 2019). Current regional or continental scale operational streamflow forecasting systems are primarily based in developed economies that satisfy these criteria. Examples of some of these models include: the United States National Water Model (NOAA, n.d.), Short-term Water Information Forecasting Tools (Australian Government Bureau of Meteorology, 2022), and the European Flood Awareness System (Copernicus EMS, n.d.). Developing countries have relied on donor support to maintain their hydrometeorological forecasting infrastructure. However, the maintenance costs of these systems continue past the initial project investment, resulting in their premature end of life (Nelson, 2021).

Over the past decade, there has been an emergence of global high-resolution forecasting models that have been built to fill this void. These include GloFAS: Global Flood Awareness System (Alfieri et al., 2013), PC-GLOBWB: PC Raster GLOBal Water Balance model (Sutanudjaja et al., 2018), and GFMS: Global Flood Monitoring System (Wu et al., 2014). Global-scale forecasts have been possible due to improved hydrological and meteorological models, increased computing power, and improved satellite observations (Emerton et al., 2016). These models calculate the streamflow through open-channel systems by dividing them into grids. While this method is effective in modeling streamflow in major river systems, implementing it for smaller tributaries requires significant computing resources. An alternative is to represent the river network as a set of lines and then perform the flow calculations for individual stream segments. This vector-based approach can represent smaller streams, thereby maximizing the application of global forecasting systems.

This thesis explores the quality of a global ensemble streamflow forecasting service over the mountainous terrains of Nepal. The GEOGloWS-ECMWF Streamflow Service (GeoGLOWS,

2022), hereafter called GESS, is a service from the Group on Earth Observations Global Water Sustainability (GEOGloWS). GESS uses the gridded surface runoff from the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble forecast system, which is then routed through a vector-based stream network using the Routing Application for Parallel computation of Discharge (RAPID) routing model. GESS provides 15-day forecasts for most major river segments in Nepal, most of which are not covered by other global forecasting systems. While these forecasts are mainly designed for flood forecasting applications, the thesis explores their utility for another potential end-user: Nepal's Run-of-River (ROR) hydropower sector.

1.1 Research questions

The main goal of the thesis is to explore the quality of the GESS streamflow forecasts and their potential value to the ROR hydropower operators in Nepal. This aim can be fulfilled by answering the following research questions:

1) What is the quality of the GESS streamflow forecasts in areas of interest for hydropower operation in Nepal? Can these GESS streamflow forecasts be used at all in these areas?
2) How useful are the GESS streamflow forecasts for different operational scenarios faced by a ROR hydropower operator?

Question 1 involves the overall technical skill of the streamflow forecasts, while Question 2 concerns the value added in terms of the requirements of a hydroelectric operator in Nepal. This thesis aims to provide more nuanced insight into the forecast performance in the mountainous watersheds of Nepal, which has always been a forecasting challenge.

1.2 Thesis Structure

The thesis is structured as follows:

- Chapter 2 provides a brief introduction to the fundamental concepts used in this thesis. These include ensemble streamflow forecasting, run-of-river hydropower, and previous forecast-value studies.
- Chapter 3 describes the data sources and the methodology used in the thesis.
- Chapter 4 presents the findings and a discussion of the results.
- Chapter 5 summarizes the thesis results, discusses some limitations, and provides recommendations for future work.

Chapter 2: Fundamentals

2.1 Run-of-river hydropower

Hydropower systems can be divided into: reservoir, run-of-river, and pumped-storage hydropower. Reservoir hydropower plants store water behind a dam in a reservoir, partially mitigating the seasonal resource variability of streamflow (IRENA, 2012). Pumped storage hydropower schemes use cheaper off-peak electricity to pump water from a lower reservoir to a nearby higher reservoir. They can then use this water to generate electricity at peak times and provide grid stability and flexibility of services (IRENA 2012, p. 8). Run-of-river (ROR) installations have limited or no storage capabilities. Hence, electricity generation is driven by the natural flow and elevation drop of a river (IRENA 2012, p. 8**)**.

ROR systems use smaller dams to divert water from a river. Hence, the natural flow of a river is disrupted only for a short distance between the dam and the tailrace, where the diverted water is re-introduced back to the river. Even for the disrupted portion, the environmental flow (e-flow) regulations dictate a certain amount of water to be left in the natural drainage path of the river. This limits ecological disturbance of the upstream and the downstream habitats (Energy BC, 2016).

Some ROR designs have limited pondage capabilities that allow short-term water storage in hourly or daily timescales. Due to the little or no storage capacity, the resulting energy generation varies as hydrological conditions fluctuate. Hence ROR systems are generally used for baseload generation (IEA, 2021).

Despite the lack of operational flexibility, an ROR system has lower construction costs and can be built at sites where the other two hydropower types would not be feasible. They have smaller pondage than other hydropower types, so they are less likely to disrupt existing human and animal habitats. These conditions have encouraged the adoption of ROR systems in lower-income countries such as Nepal.

2.2 Streamflow forecasting

Streamflow forecasting is a sub-unit within the broader field of hydrometeorology. The term hydrometeorology is a combination of hydrology and meteorology. Hydrology represents the study of the physical processes that drive and affect the movement of water on and under the earth's surface (Dingman, 2008). Meteorology deals with the study of atmospheric circulation, with the primary focus here being the transport of moisture. This involves the processes of evaporation from oceans and other water bodies that finally end with the water falling back to the earth's surface in the form of precipitation. (Roundy et al., 2019).

Streamflow forecasting aims to predict future discharge (explained in 2.2.1) values, enabling users to make informed decisions and prepare for upcoming weather events. Initial motivations for

streamflow prediction were to mitigate potential flooding damage. Modern applications have extended to water management in various fields such as drinking water supply, energy generation, and irrigation projects. While the denotation of “streams” and “rivers” depends on the open channel system's size and relative cultural importance, these terms will be used interchangeably in the rest of the thesis.

2.2.1 River-flow fundamentals

The drainage basin of a stream is the region upstream, where the water molecules falling anywhere within this area can eventually flow into the stream. A drainage basin is also referred to as a *watershed* or a *catchment area*. The watershed provides the domain where flow conservation rules can be applied.

Open-channel flow can be measured using the *discharge* and the channel *stage*. Discharge is the rate of the volumetric flow of water through a stream’s cross-section. Stage is the height of the river surface from the riverbed. It shows the water level in a river at any time, and this can be converted into a discharge measurement using a flow rating curve.

Open channel flow has three interconnected source processes (see Figure 1): (1) groundwater, (2) melt from glaciers and snowpack, and (3) precipitation. Discharge variability from groundwater and glaciers is gradual. Hence, the flow from these sources is also called *baseflow* (Sene, 2016). Snowpack melt and precipitation cause a faster response in discharge values downstream in an open channel.

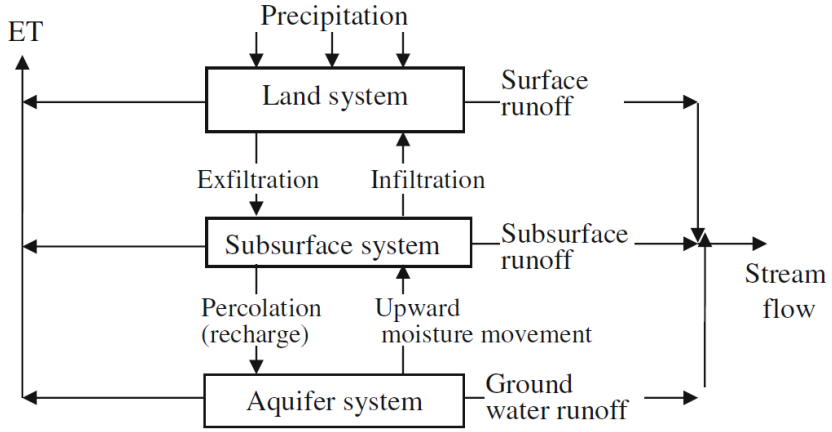


Figure 1: Processes in the hydrological cycle that lead to streamflow. ET stands for evapotranspiration.
Source: (Jain & Singh, 2019, p. 321)

The *catchment response* is the time between an event (such as precipitation) and changes in flow variables (stage/discharge) downstream. The catchment response for liquid precipitation events is usually on the scale of hours to days. Snowpack in mountainous regions could lead to a catchment response ranging from hours (due to diurnal pattern of insolation) to months (due to seasonal changes in insolation and surface temperature). While rain is the most common form of

precipitation, there are other forms such as snow and hail. Precipitation as snow generates a slower event response as the snow accumulates to form a snowpack.

Precipitation-induced catchment response can be through rain falling directly on the stream or anywhere in the drainage area (Dingman., 2008). Before reaching the bare ground, precipitation can be *intercepted* by vegetation or other structures.

The water then seeps into the ground through soil pores, in a process called *infiltration* (Jain & Singh, 2019). Infiltrated water is absorbed by surface vegetation through its roots. Vegetation has further contributions to the water balance as they lose water from their leaves in a process called *transpiration*. This process is often measured together with the evaporation of water particles from the surface. Hence, the term *evapotranspiration* is a combined term used to define these processes (Jain & Singh, 2019).

When water infiltration exceeds the soil's moisture retention capacity, excess water will flow overland towards lower elevation surfaces until reaching the *drainage line* carved by the open channel flow. The excess water flowing is called *surface runoff*. A similar process called *subsurface runoff* can happen below the surface, where groundwater flows laterally into the open channel.

2.2.2 Modelling streamflow

Streamflow modelling uses hydrological models, which are mathematical-computational models that represent the hydrological characteristics of a watershed and use hydrometeorological observations to estimate the future streamflow values at different temporal resolutions (Bourdin, Fleming, & Stull, 2012; Jain & Singh, 2019). The most used hydrometeorological input variable is precipitation. Input variables vary with the model type, and other commonly used input variables include air temperature, soil moisture, snow cover, incoming solar radiation, and land use (Sene, 2016).

Streamflow forecasts are issued for a certain period in the future, called the *forecast horizon*. Table 1 highlights the commonly used classification scheme for hydrometeorological forecasts based on the forecast horizon. The forecasts used in this thesis go up to 15 days in the future. The thesis uses the following breakdown for further sub-classification:

- Short: 1 – 3 days
- Medium: 4 – 10 days
- Extended: 11 – 15 days

The focus of this thesis will be on short- to medium-range forecasts.

Hydrological models are broadly divided into two classes: empirical (section 2.2.2.1) and conceptual (section 2.2.2.2). Raw forecasts using either of these approaches have inherent uncertainties, but post-processing can improve forecast utility. Post-processing involves further

operations done using raw outputs from forecast runs. Post-processing includes bias correction, verification, creating secondary forecast products such as hydroelectric energy yield using inflow values, and communicating the forecast values to the end-users via maps, diagrams, and graphs.

Table 1: Classification of streamflow forecasts based on the time horizon.

Term	Forecast horizon
Nowcasts	3 – 12 hours
Short	< 2-3 days
Medium	3 – 10/15 days
Long	weeks – months

2.2.2.1 Empirical models

Empirical models use statistical relationships, transfer functions, and artificial neural networks to determine the relationship between input rainfall data and output streamflow values (Sene, 2016). The mapping between the inputs and the outputs is done by calibrating the models at certain known conditions. The calibrated model is then used for forecasting streamflow. Hence, the physics is accounted for by using constants or parameterizations instead of representing explicit processes. Some commonly used empirical approaches include linear/non-linear regression and machine learning techniques such as Artificial Neural Networks.

Empirical methods are computationally efficient, parsimonious, and well optimized for any particular watershed (Sene, 2016). However, they are overly simplified and do not help explain the physical processes that lead to a particular streamflow event. If the models are not well trained, they might be useful only in limited physical and environmental conditions for which they were initially developed (Bourdin et al., 2012).

2.2.2.2 Process-oriented models

Process-oriented models attempt to simulate the physical hydrological processes operating in the watershed (Bourdin et al., 2012). These can be further divided into conceptual and physically-based models. Conceptual models are intermediate between empirical and physically-based models. They use storage cells to account for hydrological processes such as groundwater reserves, soil moisture, and reservoir levels (Sene, 2016).

Physically-based models use mass, momentum, and energy transfer equations related to hydrological variables to calculate streamflow values in future timesteps (Jain & Singh, 2019). A region of interest is divided into several vertical and horizontal grid cells, over which the solutions to these equations are solved. Runoff is calculated for the individual cells, then routed into the neighboring downstream cells (Sene, 2016). River routing is the process by which river flow rate is calculated as a volume of water moves downstream of the river segment (Liu, Wang, Xu, & Duan, 2017).

2.3 Ensemble Forecasting

Hydrological models are simplified representations of a complex physical system. As such, the modelling outputs come with uncertainty. The uncertainty can be due to any of the following reasons:

1) Inaccuracy and lack of input data, such as precipitation in different spatiotemporal scales
2) Uncertainty in the modelling parameters used
3) Incomplete knowledge of the physical processes (meteorological or hydrological) that are at play

Hydrometeorological processes have inherent nonlinear and chaotic nature, making the forecast outputs highly sensitive to the input model parameters and initial conditions (Sene, 2016). Even the slightest deviation from the "true" system state in the model input conditions will grow into significant differences in the forecasts. Since the initial input parameters and conditions will always have uncertainty, the forecast uncertainty will increase as the forecast horizon increases. Hence, the forecast skill of hydrological models reduces over longer lead times (Stull, 2017).

Accordingly, the result of a single *deterministic* forecast run cannot be trusted entirely. This limitation is improved by using the ensemble approach, where multiple forecast runs are performed for the same period by changing the initial conditions, model parameters, and/or grid resolution.

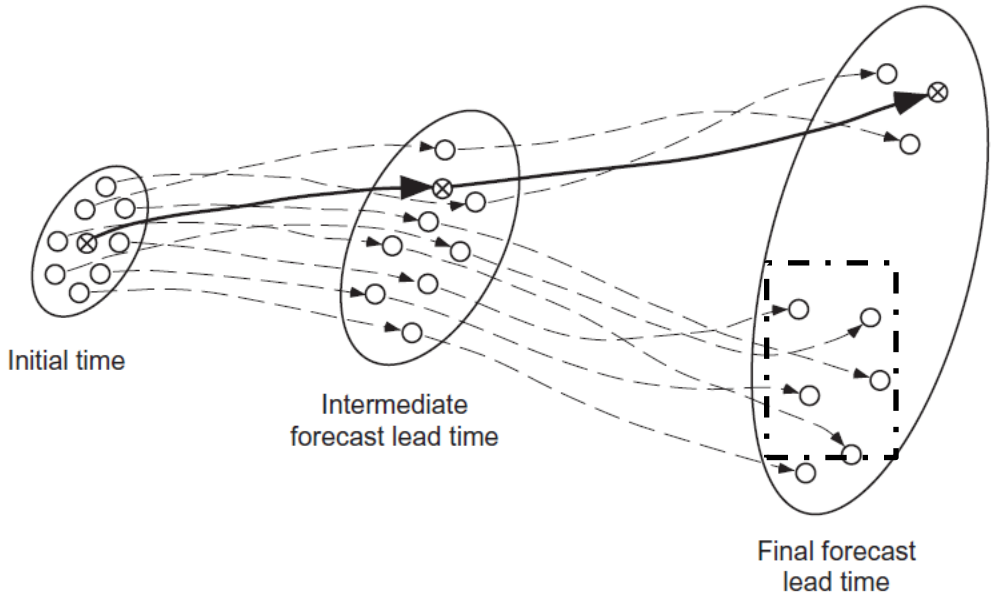


Figure 2: The concept of ensemble forecasting.
Source: (Wilks, 2019b)

Figure 2 shows an illustration of the ensemble forecasting approach. The circular regions denote all the possible hydrometeorological states of a region. The smaller circle size at the initial time illustrates that although the initial state is known with the highest certainty, it cannot be precisely

determined due to measurement errors and the inadequate distribution of measurement devices. The lines between the different lead times denote the different ways the hydrometeorological variables could evolve over time.

A single hydrological model run (solid black line), aka the *deterministic forecast*, would provide a single output, which may not necessarily be the most accurate forecast. An ensemble of the forecast runs (dotted lines) increases the likelihood that the spread in the possible states of the atmosphere is captured more accurately.

For the illustration in Figure 2, the final lead time forecast shows that most ensemble members congregate in the lower region (denoted by the dotted box). The forecast end-user can use this information to predict that an event that falls within the dotted box is of higher probability for that lead time. This allows the user to make risk-based decisions using the stochastic information provided by the spread of the ensemble members, which is in contrast to the deterministic forecast where the user could either trust or not trust the forecast.

When the ensemble of multiple forecasts is averaged to give a single deterministic forecast, the resulting ensemble mean (or the median) is usually better than any individual ensemble member. The difference between the ensemble members in the form of variance or spread denotes the magnitude of forecast uncertainty. The ensemble members could be further post-processed to yield probability distribution forecasts, allowing end-users to quantify the risk associated with a forecasted event.

2.4 Land Surface Model as Hydrological Model

Operational ensemble streamflow forecasting systems increasingly use ensemble Numerical Weather Prediction (NWP) models to provide the precipitation and temperature inputs, which are then used to drive the hydrological models. The broader implementation of this approach into hydrologic models was due to the success of the ensemble forecasting approach for meteorological forecasts. NWP forecasts rely on Land Surface Models (LSMs) that predict the surface fluxes of heat, moisture, and momentum (Warner, 2011). This is because the surface moisture and heat budgets impact the transfer of these components between the surface and the lower atmosphere, making them essential to model in an NWP forecast.

Modern NWP systems divide the land surface features into multiple layers to account for the hydrologically relevant fluxes within them (see Figure 3). Multiple soil layers are drawn to represent areas of vegetation water uptake and moisture fluxes between the layers. Individual soil layers have water retention capacity and contribute to the subsurface runoff. Vegetation can have multiple canopy layers to account for radiation attenuation by taller trees and different rates of evapotranspiration. Dynamic vegetation is implemented to account for seasonal changes in vegetation, such as leaf loss. Snow is also modelled as a dynamic layer, where snow volume in a grid cell changes over time, and a grid cell could also have fractional snow coverage. Similarly,

some models represent snow as multiple layers. Blyth et al. (2021) highlight the history and recent advances in representing the land surface processes in LSMs.

The physical processes modeled by LSMs are the important processes that lead to streamflow. As part of operational NWP configurations, they run across a domain with both vertical and horizontal grid points. Hence, they represent a part of the modeling chain for the physically-based models introduced in section 2.2.2.2. The configuration of using LSM to simulate the land surface processes and use a routing scheme to produce streamflow from the runoff values has been used in operational models.

The United States National Water Model uses Noah-MP as the LSM and implements separate routing schemes for the surface, subsurface, and channel flow routing (NOAA, n.d.). The Global Flood Awareness System (GloFAS), jointly developed by the European Commission and ECMWF, produces streamflow forecasts by routing the gridded runoff outputs from Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) LSM using the Lisflood routing model (Alfieri et al., 2013).

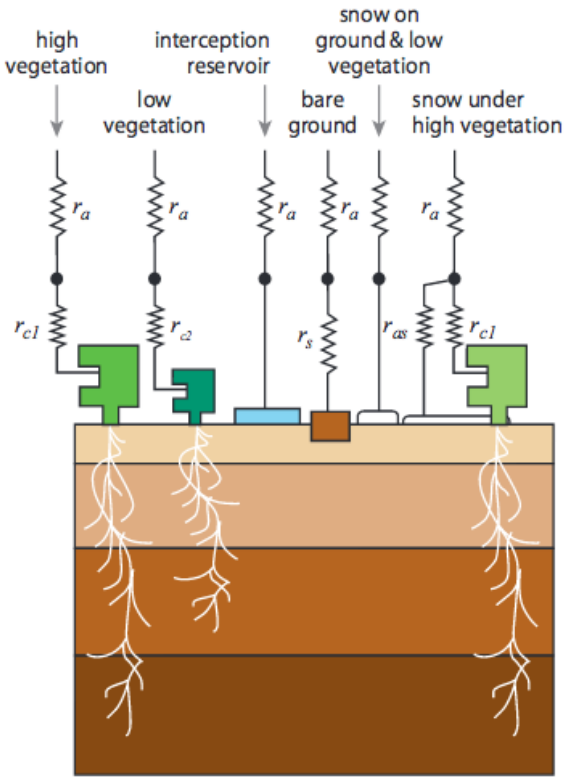


Figure 3: Schematic of the HTESSEL LSM used in ECMWF operational NWP systems. Two vegetation, four soil layers, and separate layers for reservoirs (lakes/ponds), snow, and bare ground are represented.
Source: (ECMWF, 2020)

The streamflow forecasts used in this thesis are created by using a modified approach to GloFAS. The gridded runoff is obtained from the operational ECMWF runs, which have HTESSEL

embedded within them to provide the lower boundary conditions for the NWP model. The routing of the gridded runoff to individual stream segments is then done by another model called RAPID (see section 3.1.2).

2.5 Streamflow Forecasts and Hydropower

Streamflow forecasts have been extensively used for the efficient operation of hydroelectric systems. M.-A. Boucher & Ramos (2018) provide an overview of the operational constraints different hydropower systems face and how ensemble streamflow forecasts can be used for optimal hydropower operation.

In reservoir-based hydropower systems, the primary motivation to implement streamflow forecasts is for effective water management. Reservoir-based hydropower systems can regulate the natural variability in streamflow such that the hydropower plant can meet the electricity demands throughout the year. The higher the reservoir water level, the higher the energy production possible. However, storing maximum water in the reservoir is risky — a heavy precipitation event could lead to dam overflow and destructive flooding downstream. Furthermore, reservoirs are often designed for multiple purposes and have additional compliance restrictions such as minimum water levels for recreational, fisheries, irrigation, or habitat conservation purposes (McCollor & Stull, 2008b).

As reservoir-based plants are more flexible in terms of water management options, they are the focus of most studies exploring the utility of streamflow forecasts in hydropower systems. Fan et al. (2016) use probabilistic streamflow forecasts to force a short-term optimization model to operate a reservoir-based hydropower system in Brazil, with the goal of maximizing energy production while ensuring flood control downstream. Kleiven & Steinsland (2019) explore a post-processing method to improve inflow forecasts for the Osali catchment in Norway. Bourdin et al. (2014) introduce a method to produce reliable probabilistic forecasts for the Daisy Lake Reservoir in Canada from an ensemble of streamflow forecasts.

Other studies introduce hydropower-scheduling models, which take stochastic inflow and price inputs to minimize risk in deregulated energy markets (Hongling, Chuanwen, & Yan, 2008). Ensemble streamflow forecasts can be used as stochastic inflow inputs in such modelling schemes. Several regional/national power producers such as EDF and CNR in France (M.-A. Boucher & Ramos, 2018), CEMIG in Brazil (Fan et al., 2016), and BC Hydro in western Canada rely on ensemble inflow forecasting along with stochastic decision-making models to assist in their operation.

2.5.1 Economic value added by streamflow forecasts

Certain studies have also included an economic value assessment while researching the streamflow forecasts for hydropower systems. McCollor & Stull (2008c) use the cost-loss and the decision

theory model to assess the economic value of probabilistic forecasts generated using the University of British Columbia ensemble system for several dams operated by BC Hydro. M. A. Boucher et al. (2012) designed an ensemble streamflow forecasting system for the Gatineau hydropower complex in Canada. They show that the ensemble system fed stochastic decision-making model leads to higher energy production by reducing spillage while maintaining low reservoir levels to account for future high-flow events. Cassagnole et al. (2020) explored how different aspects of forecast quality affect the economic value provided by the ensemble mean forecast for hydroelectric reservoir management. They found that revenue losses were more significant for over-estimating forecasts, followed by under-estimation and then the under-dispersed forecasts.

2.5.2 Studies relating to ROR systems

Studies that explore the use of short-term ensemble streamflow forecasting systems for ROR operation do not exist to the best of the author's knowledge. Limited generation flexibility and a comparably smaller share of ROR schemes in the generation portfolio of most countries could be possible reasons behind the lack of such studies. Stokelj et al. (2002) create a 2-day deterministic inflow forecast using Artificial Neural Networks for a ROR system in Greece. Contreras et al. (2020) explore the use of seasonal climate forecasts for ROR operation.

However, it has been identified that ensemble streamflow forecasts could be helpful for flood forecasting as flood events can damage ROR facilities and threaten the safety of its workers (M.-A. Boucher & Ramos, 2018). Ensemble streamflow forecasts could also play a role in deregulated markets with day-ahead trading or in markets with "carrot-and-stick" feed-in-tariffs, where a hydropower operator could be penalized for not meeting promised energy supply (see section 2.6.2).

2.6 Nepal

This section aims to provide a brief overview of Nepal, the geographical area where the research is focused. Nepal is a landlocked country between India and China in South Asia (see Figure 4). Almost 80% of the country is covered by mountains, which makes it a challenging region to create NWP forecasts. The Himalayas lie to the north and are the highest elevation regions in the country. Mt. Everest, on the Nepal China border, is 8849 m tall. Flat low-elevation lands called Terai lie on the southern part along the border with India. In between the Himalayas and the Terai lies the Mahabharat region, which consists of hilly regions ranging from 1000 m to roughly 4000 m in elevation.

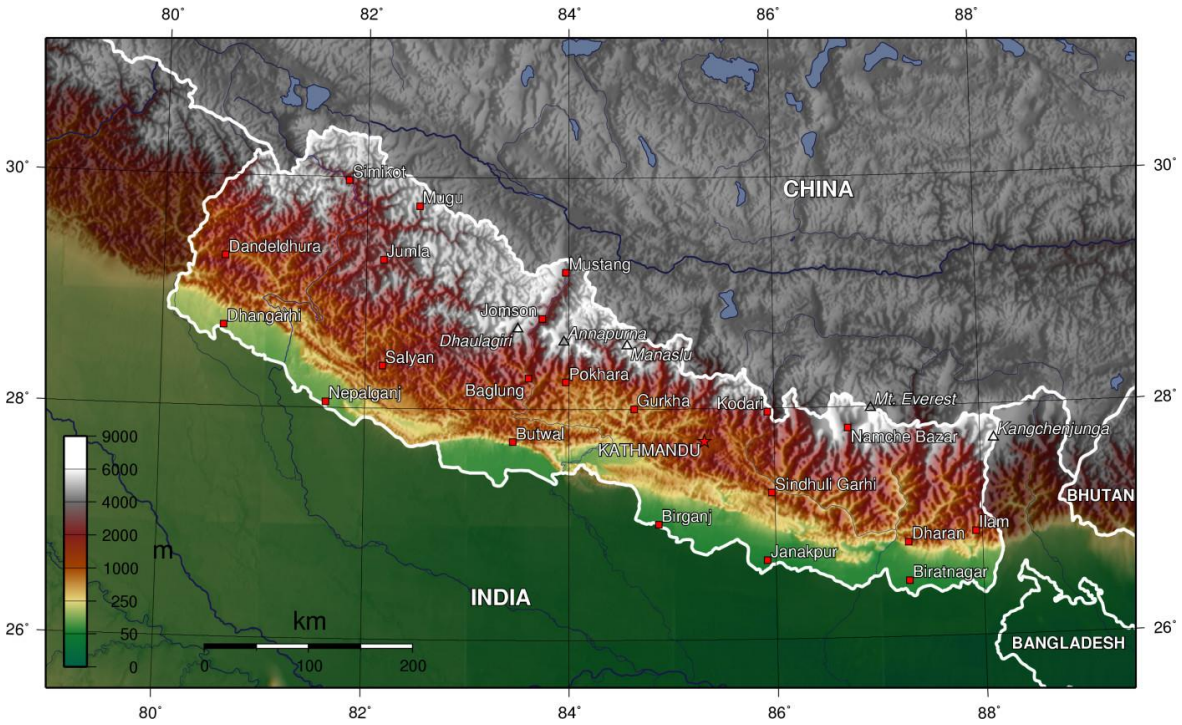


Figure 4: Elevation profile of Nepal.
Source**:**(Wikimedia Commons, 2006)

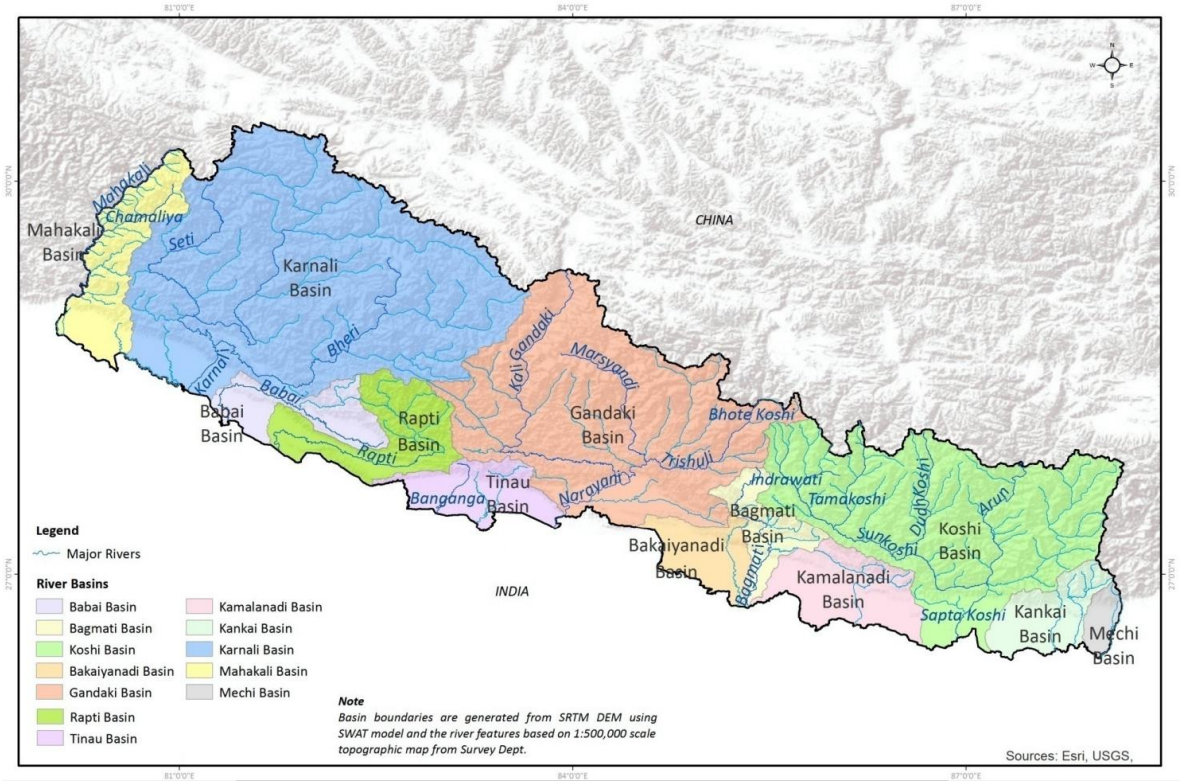


Figure 5: River basins of Nepal. The three biggest basins are Karnali (blue), Gandaki (orange), and Koshi (green).
Source: (WECS, 2019)

The major river systems in Nepal are tributaries of the Ganges basin (see Figure 5). The glaciers present in the Himalayas of Nepal and the trans-Himalayan region in Tibet are the sources of the three major river systems in Nepal: *Karnali*, *Gandaki*, and *Koshi*. Almost 80 percent of precipitation occurs due to the southwest monsoon winds, usually from June until September. Precipitation in other times is due to either western disturbances or due to high-intensity orographic thunderstorms observed during the pre-monsoon season (Department of Hydrology and Meteorology, 2018).

The local currency of Nepal is the Nepali Rupees (NPR), which is pegged to the Indian Rupee (INR) at a fixed rate of 1 INR = 1.6 NPR. The value of streamflow forecasts in this thesis will utilize these currencies.

2.6.1 Existing streamflow forecast

The Department of Hydrology and Meteorology (DHM) is the government entity responsible for producing hydrometeorological forecasts and maintaining the surface measurement stations in Nepal. They are also responsible for disseminating extreme weather and flooding/landslide warnings. The streamflow forecasting system currently employed by DHM mainly relies on measurements from upstream catchments to issue warnings to the downstream areas. There are limited rainfall-runoff models for 2-3 major basins, but they only provide streamflow forecasts for river segments in the flatlands of Terai, where flooding events have the most impact (*personal communication*). To the best of the author's knowledge, GESS is the only existing streamflow modelling framework that covers the entirety of Nepal, including the flashier hilly region watersheds.

2.6.2 Hydropower Market

Nepal has a vertically integrated energy market with the Nepal Electricity Authority (NEA) solely responsible for the transmission, distribution, and generation networks. The government-sanctioned generation projects are funded solely through NEA or Public-Private Partnership, with NEA as the majority stakeholder. The private sector can participate in the electricity market as a generator by negotiating Power Purchase Agreements (PPA) with the NEA.

Table 2: Annual total electricity consumption by source for 2020-21

Source	Amount (GWh)	Share (%)
NEA power stations	2808.4	31.5
IPPs	3308.3	37.1
From India	2805.8	31.4
Total	**8922.5**	**100**

Hydroelectric projects, mainly the ROR type, dominate Nepal's electricity generation portfolio (see Table 2), with a total installed capacity of 1464 MW. This portfolio is insufficient to meet the peak demand of 1482 MW (NEA, 2021). Hence, imports through the Indian grid also play an

essential role in grid stability. During the peak dry season months (January through April), the share of electricity imported from India can reach as high as 60% of the monthly demand. That number goes down to 5% during the wet season (August through November), when the monsoon rains drive up streamflow in Nepal (NEA, 2021).

IPPs that plan to build their own generation facilities will most likely sell electricity to the NEA as it is the sole owner of the distribution and transmission networks. The selling price available to an IPP depends on the type of hydropower installation and the hydrological season when the energy is being sold (ERC, 2019). ROR schemes are currently priced at NPR 8.40 per kWh for the dry season (December – May) and NPR 4.80 per kWh for the wet season (June – November). Most contracts have a take-or-pay provision, which guarantees project cashflows for the electricity produced except in force majeure conditions. The PPAs are long-term PPAs of 25 years with price escalation provisions for the first eight years after the commercial operation date.

Electricity trade between Nepal and India is conducted through bilateral government-to-government deals and the day-ahead market on the Indian Energy Exchange (IEX). Nepal's participation in the Indian spot market started in November 2021. Currently, there is one 400 kV transmission line between India and Nepal through which up to 1000 MW of electricity can be traded (NEA, 2021). The Power Trading Unit within the central Load Dispatch Centre of NEA is responsible for bidding in the day-ahead market. There are ongoing talks about the possibilities of IPPs directly participating in the IEX by paying certain wheeling charges for using NEA's grid (*personal communication*). However, that has not come into regulation.

Besides the day-ahead market, Nepal's electricity imports are mainly through PPAs with the neighbouring Indian states of Bihar and Uttar Pradesh or with NTPC Limited, India's largest power utility. The average PPA price through these agreements with Indian states is currently NRs. 8.15 per kWh (ERC, 2020). However, most of Nepal's imports are through PPA with the NTPC, which is at NRs. 6.70 per kWh of electricity imported (Rose, Duwadi, Palchak, & Joshi, 2021). The day-ahead market allows NEA to access competitive prices. In 2021, the average market clearing price in IEX was NRs. 6.32 per kWh (IEX, 2022), which is cheaper than the PPA prices above. The average energy prices in the bidding areas corresponding to Bihar and Uttar Pradesh were also similar.

Chapter 3: Methodology

3.1 Data Used

This section describes the data sources and how they are used in the thesis. The data analysis was done primarily using the *pandas* package on Python. Forecast datasets were in NetCDF format, and the *xarray* package was used to read the files. Most of the plotting was performed using the *Plotly* package.

3.1.1 Observations

Streamflow gauge data were retrieved from the hydrologic observation network of DHM, the central meteorological body of Nepal. DHM did not have automated measurement stations at the time of data acquisition, and most gauge readings were manually taken daily. DHM's local offices store these observation data. After each block of 4 to 5 years of observed data is available, a final bias correction is performed (which itself takes more time) before making the observation data available to the public[1]. The most recent block of available observation data at the start of this research includes years only through the end of 2015.

3.1.2 Streamflow forecasts

The primary focus of this thesis is the quality of the GESS streamflow forecasts. The hydrological modelling chain comprises three steps: 1) gridded runoff generation, 2) downscaling of gridded runoff to individual river segments, and 3) flow routing of the runoff to segments downstream.

The primary hydrological model used is the HTESSEL land surface model, which is embedded into the ECMWF Integrated Forecast System (ECMWF, 2021). HTESSEL simulates the land surface processes (see section 2.4) to provide ensemble gridded runoff. The 52-member ensemble system consists of a control forecast member, 50 perturbed forecasts, and one high-resolution member.

The perturbed forecasts are created by adding small-amplitude perturbations to the initial conditions. The perturbed initial conditions are constructed from singular vectors and an ensemble of 4D-Var data assimilations, which lead to the 50 perturbed forecast members (ECMWF, 2022a). Hence, the ensemble members are created by accounting only for the weather forecast uncertainty. The control forecast and the 50 perturbed members have an interpolated resolution of 0.2° (~18 km), while the high-resolution member is produced at 0.1° (~9 km) resolution.

The gridded runoff, $R_{i,j}$ is then assigned to individual river segments via area-weighted downscaling. The area-weighted method calculates the area overlap of each LSM grid to the

[1] personal communication with DHM Data Section in-charge Mr. Rudra Pariyar

intersecting catchment. This weight, $w_{i,j}$ is then used to calculate the total volume of lateral inflow, Q_e, available in the catchment using equation (1). Here, A_{LSM} is the geographical area of the grid for which the runoff value is valid.

$$Q_e = \sum R_{i,j} \times A_{\mathrm{LSM}} \times w_{i,j} \quad (1)$$

The routing is done using the RAPID model. RAPID is a vector-based routing method that simulates water flow propagation in a river network based on the Muskingum method (Cunge, 1969), adapted to matrix-vector notation. Muskingum routing depends on two parameters, *k* and *x*. *k* is related to wave celerity, which denotes the speed at which a water wave moves through a channel. *x* represents the diffusivity of the wave and takes a value between 0 and 0.5 (David et al., 2011).

In contrast to grid-based routing, where a set of grid cells represents a river channel, vector-based routing represents rivers as a set of lines and nodes (see Figure 6). The routing takes the upstream forecasts and the catchment-wide lateral flow as input and calculates flow at the mouth of the river segment.

$$Q_3 = Q_1 + Q_2 + Q_e \quad (2)$$

Equation (2) shows the volumetric balance used in the routing calculation for a river segment that receives inflows from two upstream segments, Q_1 and Q_2. The lateral flow component (Q_e) is calculated using the gridded runoff from the LSM and is assumed to enter at the top of the river segment.

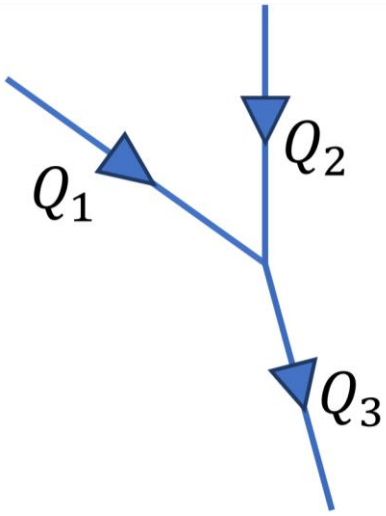


Figure 6: Example representation of a river in RAPID routing. *Q* represents the discharge outflow at the end of each river segment.

The operational archive of forecasts for Nepal was available beginning in March 2020. However, since Nepal has observational data only until 2015, this record is not as helpful. Instead, a

reforecast dataset was made available for the thesis from the Hydroinformatics team at the Brigham Young University (BYU), one of the contributing institutes to making GESS (Souffront Alcantara et al., 2019). The reforecast dataset has daily forecast re-runs from January 2014 until 2019. The dataset does not cover all the stream reaches as the operational archive; however, it provides forecasts for major river segments in Nepal (see Figure 7).

Thus, there are two years of overlap in the availability of the reforecasts and the observations: 2014 and 2015. Hence, these two years will be used for the analysis in this thesis.

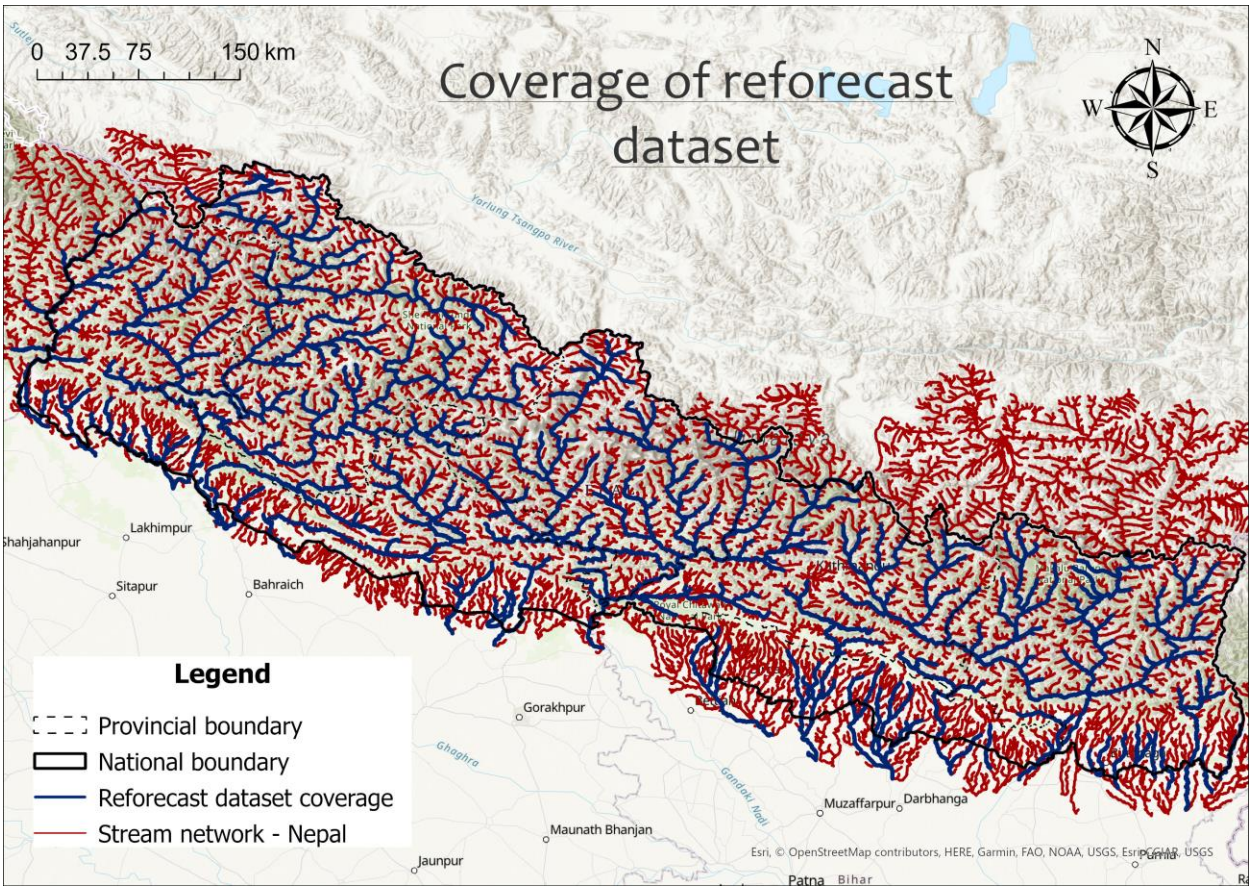


Figure 7: Coverage of the reforecast dataset (blue) vs. the operational archive (red)

3.1.3 Hydroelectric Projects

Data for the hydroelectric projects were retrieved from the Nepal Hydropower Portal (Niti Foundation, 2018). The portal contains a dataset compiled from IPPs and various governmental agencies involved in different stages of hydropower construction. Data can be filtered based on physical location (province, districts), operational status (under study, under construction, operational), and project capacity. The portal is used to retrieve data for hydropower projects that satisfy the following criteria:

- Project capacity greater than 5 MW
- Operation status of either under construction or operational

While the dataset provided rough coordinates of the project's location, it did not provide information on the exact location of the headworks (dam) site. The location information from the portal was used in conjunction with the project website to identify the dam locations on satellite imagery using Google Earth Pro (Google LLC, 2020). The dam sites for smaller hydroelectric projects were difficult to identify due to the lesser volume of water pooling. Hence, a 5 MW cutoff was applied. Nepal has one operational reservoir hydropower project and one under construction. Both these projects were excluded after retrieving the data.

3.2 Overall analysis workflow

The overall analysis workflow implemented in this thesis is as follows

1. Optimum sites that are ideal for this study are identified (see Section 3.3),
2. Relevant data is retrieved for each of these sites,
3. The quality of the raw forecast dataset is assessed using verification metrics (see Section 3.4),
4. A simple bias correction methodology is tested and verified (see Section 3.5),
5. Possible applications of the forecasts for ROR operation are identified (see Section 3.6),
6. Forecasts are verified in terms of potential value added for each site (see sections 3.8 and 3.7).

3.3 Site selection using Spatial Analysis

As the focus of the thesis is on ROR hydropower projects, the analysis is done for the river segments that match the following criteria:

- Forecasts using the reforecast dataset are available,
- Stream gauge data exists, and
- ROR project is operational or under construction.

ArcGIS Pro (ESRI Inc., 2021) is used to identify the sites that match these criteria through spatial data-analysis methods.

Figure 8 shows the result of the analysis. The river segments that match the criteria mentioned above are highlighted in green. The streamflow gauge was located in the stream section between the dam and the tailrace for some of these sites. These gauges would provide inaccurate information about the existing streamflow. Hence, these sites are excluded from the study. The sites were chosen to maximize diversity in the physio-geographic conditions where the forecast performance can be assessed. The step-by-step procedure to achieve the results is explained in Appendix A.

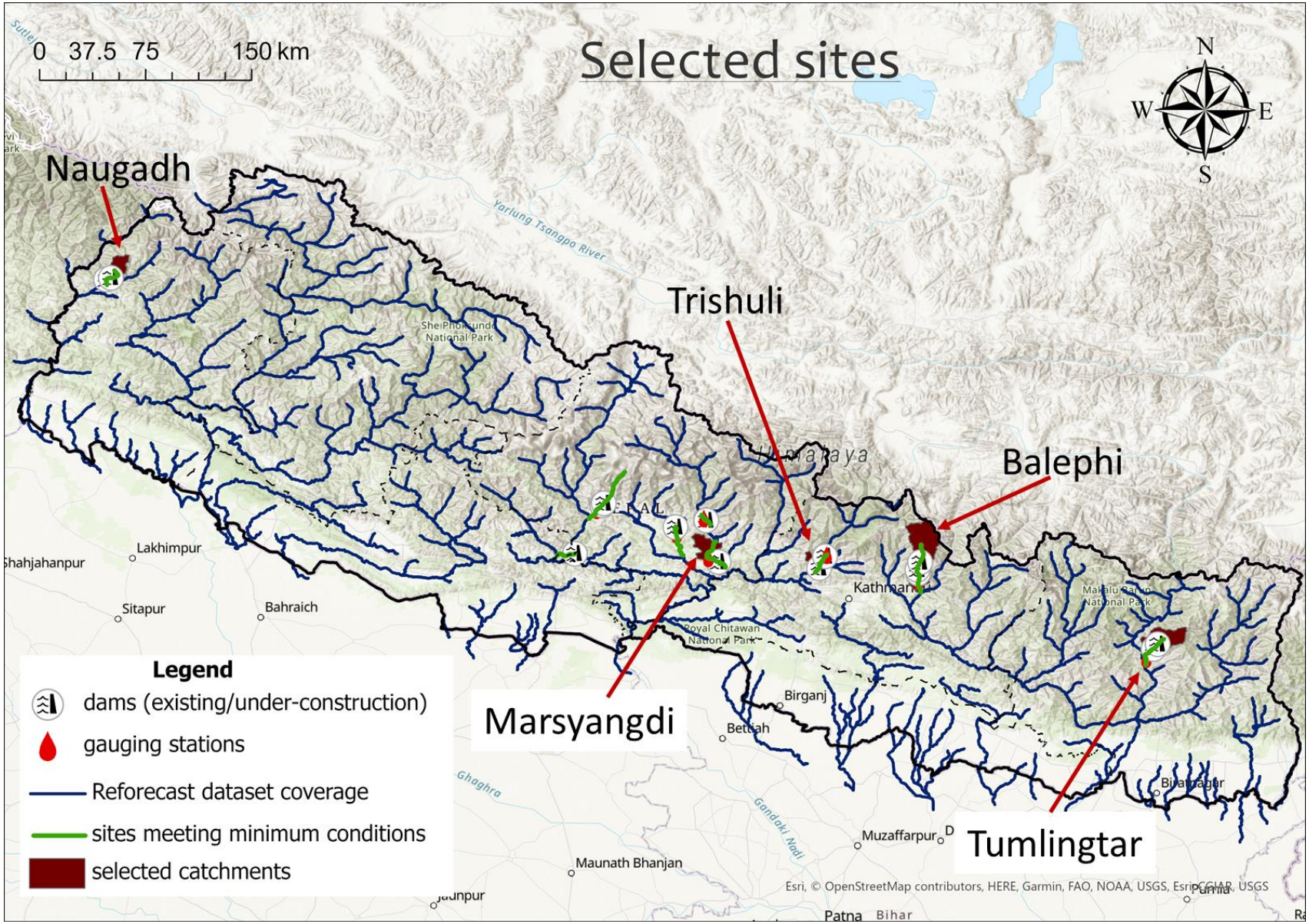


Figure 8: The nine catchments that met the minimum criteria are shown in green. The chosen sites for this study are shown in red.

3.4 Verification metrics

Forecast verification is the process of assessing the quality of the ensemble forecasts using different verification metrics. Forecast quality can be assessed by exploratory methods or numerical statistics (Potts, 2012). Exploratory methods are graphical techniques that allow visual assessment of the trends in the forecast accuracy and help narrow down areas where the forecast performance is lagging. In this thesis, the following exploratory methods are used:

- Time series plots of streamflow forecasts and observations for selected forecast horizons
- Scatter plots of overall forecasts vs. observations

Numerical statistics are metrics used to quantify different aspects of forecast quality. The use of statistical metrics, such as bias and correlation, is also referred to as the measures-oriented approach to verification (Murphy, 1997). Measures-oriented verification metrics are suited for deterministic forecasts.

For an ensemble of forecasts, the deterministic forecast could be the best-performing member, the ensemble mean/median, or any specific member (e.g., the high-resolution member in the GESS). A brief description of the measures-oriented verification metrics used in the thesis is described next.

Throughout the section, μ_f denotes the mean of the deterministic forecast f, and μ_o denotes the mean of the corresponding observations o. k denotes the time step (1 day in this case), with N being the total number of days over which the verification metric is computed.

1. Bias (β)
   Bias is the ratio between the forecast and the observation. It is a measure of the accuracy of forecasts. An unbiased forecast has a bias of 1. Greater (lower) values mean the forecasting system is consistently over (under) predicting the future flows. Bias is calculated using equation (3):

$$\beta = \frac{\mu_f}{\mu_o} \tag{3}$$

2. Pearson Correlation coefficient (r)
   Pearson correlation shows the association between the forecasts and the observations. Values of 1 indicate perfect linear correlation, -1 indicates a perfect negative linear correlation, and 0 indicates no correlation. Pearson correlation is calculated using equation (4). f_k and o_k represent the k^{th} streamflow forecast and observation, respectively.

$$r = \frac{\sum_{k=1}^{N}(f_k - \mu_f)\,(o_k - \mu_o)}{\left[\sum_{k=1}^{N}(f_k - \mu_f)^2 \sum_{k=1}^{N}(o_k - \mu_o)^2\right]^{1/2}} \quad (4)$$

3. Variability ratio (γ)
   The variability ratio is the ratio of the coefficient of variation between the forecasted and observed streamflow values. The coefficient of variation measures the dispersion of the data relative to its mean; hence, it uses both the first and second moments of the distribution. From a hydrological perspective, γ is a measure of how well the flow duration curve of the observed and the simulated inflows align (Gupta, Kling, Yilmaz, & Martinez, 2009). The variability ratio is calculated using equation (5), and its ideal value is 1.

$$\gamma = \frac{\sigma_f \,/\, \mu_f}{\sigma_o \,/\, \mu_o} \quad (5)$$

4. Nash-Sutcliffe Efficiency (NSE)
   NSE is a commonly used metric in the hydrology community (Krause, P.; Boyle, D. P.; Baese, 2005). It is a measure of the forecast performance with respect to the climatological mean forecast, $\bar{o}$. As such, it can be considered to be a form of skill score. NSE is calculated using equation (6), with the perfect value being 1. NSE = 0 means the forecast is no better than a climatological forecast. As the differences are squared, it penalizes errors in high-flow situations more than at low-flow times.

$$NSE = 1 - \frac{\sum_{k=1}^{N}(o_k - f_k)^2}{\sum_{k=1}^{N}(o_k - \bar{o})^2} \quad (6)$$

   Based on NSE, there are studies on what constitutes a good model. The thesis uses the benchmark suggested by Moriasi et al. (2015), presented in Table 3.

Table 3: Model performance evaluation based on NSE values

Range	Performance rating
$0.8 < NSE$	Very good
$0.7 < NSE \leq 0.8$	Good
$0.5 < NSE \leq 0.7$	Satisfactory
$NSE \leq 0.5$	Not satisfactory

   NSE can also be represented in terms of Mean Square Error (MSE), as shown in equation (7), with σ_o being the standard deviation of the observed flow.

$$NSE = 1 - \frac{MSE}{\sigma_o^2} \tag{7}$$

5. Kling-Gupta Efficiency (KGE)
   KGE is a simplified metric (equation 8) that combines the bias, Pearson correlation coefficient, and flow variability metrics described above. These metrics are used because they represent important factors that forecasts should have compared to the observations. Namely, better forecasts are the ones where the volumetric flow (shown by β), the spread of flows (shown by γ), and the shape of the hydrograph (shown by r) match best with the observed hydrograph. This metric was devised by Gupta et al. (2009) as an improvement to the NSE. A perfect model has a KGE of 1. While both NSE and KGE have perfect values of 1, other values of KGE and NSE are not interchangeable (Knoben, Freer, & Woods, 2019).

$$KGE = 1 - \sqrt{(r-1)^2 + (\gamma-1)^2 + (\beta-1)^2} \tag{8}$$

The measures-oriented verification metrics do not assess the forecasting system's spread or uncertainty aspect. For such information, a *distributions-oriented approach* is required, where the forecast verification is a task of assessing the joint distribution of both the forecasts and the observations. Several verification methods exist that assess the quality of the entire distribution, such as the Brier Score, Reliability diagrams, Rank histograms, and Probability Integral Transforms (PIT). Approaches such as the Brier score and reliability diagrams assess forecast quality for specific events, defined by setting a flow threshold.

In contrast, the Continuous Ranked Probability Score (CRPS) assesses the overall accuracy of the entire ensemble forecast over all possible events (threshold values in the Brier Score) (Anctil & Ramos, 2019). Hence, CRPS is used as the primary metric to assess the quality of the entire ensemble.

$$CRPS = \int_{-\infty}^{\infty} \left[P_f(k) - P_o(k)\right]^2 dk$$
$$P_o(k) = \begin{cases} 0, & k < 0 \\ 1, & k \geq 0 \end{cases} \tag{9}$$

CRPS is calculated using equation (9), where $P_f(k)$ and $P_o(k)$ are distributions of forecasts and observations, respectively. The observed streamflow is a single value. So, its distribution is represented by a Heaviside step function centered at the observation. CRPS rewards forecasts

values that are closer to the observations. Hence, a sharper, more accurate forecast will have lower CRPS (Wilks, 2019a).

3.5 Bias correction approach

The ensemble forecasting approach accounts for some of the random errors in the forecasts. The systematic errors present in the forecast outputs are corrected by bias correction approaches (Stull, 2017). Several statistical bias correction approaches exist, such as analog and regression-based methods (M.-A. Boucher, Roulin, & Fortin, 2019). However, such methods require an extensive database of forecast and observation values extending to several years. Other approaches, such as matching the cumulative distribution function taken by Wood & Schaake (2008), are more useful for longer timescale forecasts (seasonal/monthly).

This thesis tests a simple multiplicative bias correction approach called Degree of Mass Balance (DMB). This method has been tested successfully for precipitation (McCollor & Stull, 2008a) and streamflow forecasts (Bourdin & Stull, 2013) in British Columbia, which has complex mountainous terrain similar to the study sites in Nepal. Multiplicative bias correction has the further advantage of ensuring that the bias-corrected streamflow values stay non-negative.

$$DMB = \frac{\sum_{k=1}^{N} f_k}{\sum_{k=1}^{N} o_k} \tag{10}$$

$$LDMB = \sum_{k=1}^{N} w_k \frac{f_k}{o_k} \tag{11}$$

$$DMB_{var} = \sum_{k=1}^{N} \frac{f_k}{o_k} \tag{12}$$

$$LDMB_{var} = \frac{\sum_{k=1}^{N} w_k\, f_k}{\sum_{k=1}^{N} w_k\, o_k} \tag{13}$$

The DMB bias correction approach applies a correction factor (*DMB ratio*) to upcoming forecast outputs based on the forecast/observation ratio over the past time periods. Equations (10) − (13) show the several DMB ratio calculation approaches tested in this thesis. Here, *LDMB* stands for the Linearly-weighted DMB method.

The DMB ratio is calculated over a rolling window, which represents N days of the most recent forecast/observation pairs. Equation (10) is the implementation by (McCollor & Stull, 2008a), applying equal weights to all the forecast-observation pairs. (Bourdin & Stull, 2013) extended this

approach by introducing linear weighting in calculating the DMB ratio. The weight, w_k in equations (11) and (13) is calculated by:

$$w_k = \frac{N - k + 1}{\sum_{i=1}^{N} i} \tag{14}$$

Equations (12) and (13) are mathematical variations of the original DMB implementations tested in the aforementioned literature (hence the suffix *var*). Once the DMB ratio is available, the bias correction is performed using equation (15).

$$f_{\text{bc}} = \frac{f_{\text{raw}}}{DMB \mid LDMB} \tag{15}$$

The first year of the available dataset (i.e., 2014) is set aside as a calibration dataset. The forecast-observation pairs for this year are used to answer the following questions:

- What is the best DMB ratio variation among equations (10) − (13) to use?
- What is the optimum window length to use?
- Is it better to bias correct every member individually (**first approach**) or apply the same DMB ratio to all the 52 members (**second approach**)?

The value of KGE, NSE, and CRPS across different combinations of bias correction approaches is tested, and the combination that yields the best values for these three is implemented for the verification data year, i.e., 2015. Each forecast horizon is bias-corrected separately. The performance of the bias-corrected forecast for each horizon is re-evaluated using the verification metrics presented in section 3.4. Finally, the individual bias-corrected forecasts are then combined to produce 15-day forecasts for the end-user.

3.6 Value assessment of the 15-day bias-corrected forecasts

The first step in identifying the usefulness of a forecasting system is to identify areas where it could add value, which is done by identifying a list of possible operating constraints faced by an ROR operator in Nepal. For this purpose, I interviewed 2 IPP companies in Nepal, and the station chief of Marsyangdi hydropower, one of the hydroelectric stations used in this study. The informal interview identified two areas where the short-medium range streamflow forecasts could assist in ROR hydro operations:

- flood forecasting
- energy yield forecasting

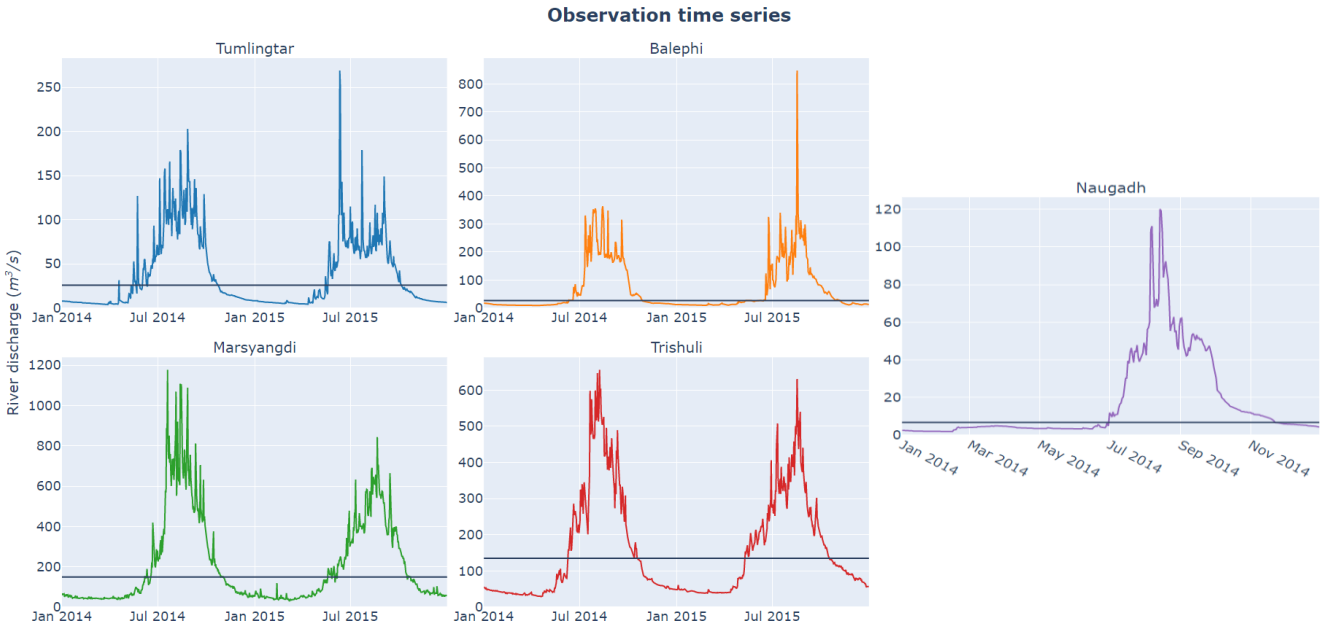


Figure 9: Time series of observed streamflow for the selected sites over the analysis period. The horizontal black line represents the 70th percentile flow value used as a cut-off between the high- and low-flow periods in this thesis.

Flood forecasting is of interest in the wet season, whereas energy yield forecasting is primarily in the dry season. As such, the verification of the forecast dataset and the potential value analysis is performed by distinguishing the forecasts into **low-flow** (i.e., dry season) and **high-flow** (i.e., wet season) periods. The Q70 value, which is the 70th percentile of flow values based on the climatological dataset, is used to distinguish between high- and low-flow periods. This value, chosen for each site, corresponded with the time when the rise and recession of the hydrograph were observed, as expected for the high-flow periods (see Figure 9).

3.7 Wet season – Forecast of flooding events

In the wet season, the river flow is high, and the ROR systems operate at maximum capacity as the average river flow is higher than the discharge required for the maximum electricity generation. However, monsoon-induced heavy rainfall can bring damaging floods. The high-volume inflow in flood events brings excessive sediment deposition to the dam site, which also demands faster sediment clearing. While the ROR hydro infrastructure itself does not have water management features, early flood warnings can assist save the lives of the ROR project staff and other humans around the project. Similarly, they can prepare for the timely mobilization of the sediment removal resources.

Flooding events are denoted by flows above the 95th percentile of climatological flows. First, the number of flooding events for each site is identified (see section 4.3). Two sites had the most flooding events. Case studies of the forecast performance for three flood events that occurred

across these two sites are presented (see sections 4.3.1 and 4.3.2). The case studies use visual assessment to answer the following questions:

- Did the forecast predict a high-flow event around the time period?
- What was the timing and magnitude error for the specific event(s)?
- Was the event identified multiple days in advance?

Finally, the hit and miss rates (Stull, 2017) of the forecasts across all the flood events are calculated by converting the ensemble forecasts into categorical forecasts (see section 4.3.3).

3.8 Dry season – Forecast of future energy yield

The river flow is lower during the dry season, and the ROR systems do not always produce their maximum (nameplate) energy. In Nepal, an IPP must declare the amount of energy it promises to deliver to the grid in the future period. A company is fined if the generated energy, E_g is less than 80% of the promised amount E_d. The fine amount is calculated by applying the per unit PPA rate to the electricity amount by which the generator is short, as shown in equation (16). Meanwhile, if the generation, E_g is higher than E_d, the IPP is paid only 50% of the PPA rate for the excess amount.

$$Fine = \left(0.8\, E_d - E_g\right) \times R_{\mathrm{PPA}} \tag{16}$$

A forecasting system adds value if it can provide accurate estimates of forecasted inflow, which helps the hydropower operator make accurate bids to the utility and hopefully avoid fines. As the generated electricity is the final variable of interest to the operators, a simple hydropower energy generation model can be used by the ROR operators to convert the forecasted inflows to forecasted energy generation.

3.8.1 Energy yield model

$$E_f = Q_{in} \,.\, g \,.\, \rho \,.\, h \,.\, \eta_p \,.\, (24 - t_o) \tag{17}$$

The daily hydropower generation is calculated using equation (17). The quantities represented in the equation are as follows:

g : acceleration due to gravity (m.s^{-2})
ρ : density of water (kg.m^{-3})
h : net head (m)
Q_{in} : flow diverted towards the turbines (m^3.s^{-1})
t_0: hours of operation (h)
η_p : plant efficiency

Q_{in} is calculated by the river flow minus the environmental flow (e-flow). ROR systems divert water at the dam site towards a powerhouse located further downstream. As a result, there is often a long stretch of river (~ 4−5 km) between the dam and the powerhouse where the river could run completely dry. The e-flow regulations ensure that a minimum volume of water is constantly flowing in the river for a stable aquatic habitat.

In Nepal, the e-flow regulation mandates ROR operators to release at least 10% of the monthly average climatological flows to the rivers from the dam sites. The maximum value of Q_{in} is decided by the rated discharge of the turbines installed in the ROR plant. Net head, h information is taken directly from the project specifications. The plant efficiency η_p is calculated by comparing the energy produced while operating at maximum capacity with the theoretical maximum energy.

3.8.2 Revenue assessment

The revenue assessment is calculated for a hydropower operator interested in making day-ahead bids bounded by the economic restrictions mentioned in section 3.8. The forecast is compared against *no information* and *perfect information*. No information, also called persistence forecast, is when the ROR operator does not have information on future inflows. So, the latest observation is considered to continue into the future. For the perfect information (also called hindsight) case, the observations recorded are taken as the forecast.

The revenue achieved across the dry season period using the three information sources is then calculated, and the relative value is assessed. Ideally, a forecasting system that adds value should provide revenue better than the persistence forecast. For this research, the best deterministic forecast is first used as the forecast. Then, we investigate whether the information about uncertainty provided by the spread information could have provided additional value.

Chapter 4: Results

For easy reference, the results for individual metrics are presented in the order of the easternmost site to the westernmost site.

4.1 Analysis of raw forecasts of river discharge

4.1.1 Scatter plots

Figures 10 – 12 show the scatter plots of river discharge for forecast-horizon days 1 – 3, 6, 10, and 15 for each site. These are useful to get a glimpse of the general tendency of the forecasts.

Outlier forecasts, where the forecasted streamflow is unusually high, were seen for all the sites. For minimal flow values (bottom left corner of the scatter plots), all the sites had more points that were over-forecasting than under-forecasting. The ensemble median (*cyan*) and the high-resolution (*blue*) ensemble-forecast members have similar distribution and follow similar trends as other ensemble members.

The sites corresponding to first-order streams are Tumlingtar (Figure 10), Balephi (Figure 11), and Naugadh (Appendix B). They mainly show an under-forecasting tendency, particularly Naugadh and Balephi, which show an overall under-forecasting tendency. For Tumlingtar, higher flow events are under-forecast, but the forecasts develop an over-forecasting tendency for lower flow events. The change in the spread of the data points, which indicates forecast precision, is minimal for all three sites as we move to a longer forecast horizon.

Sites located at higher-order streams: Trishuli (Appendix B) and Marsyangdi (Figure 12), show an over-forecasting tendency. The spread of the data points increases for longer horizons, which is distinctively seen in the scatter plots for day-1 and day-2 forecasts. The forecasts for Marsyangdi develop an under-forecasting trend for higher flow events at longer forecast horizons. For both these sites, the correlation improves from the short-range to the medium-range. This is inferred from the closer proximity of the ensemble median and the high-resolution member to the diagonal.

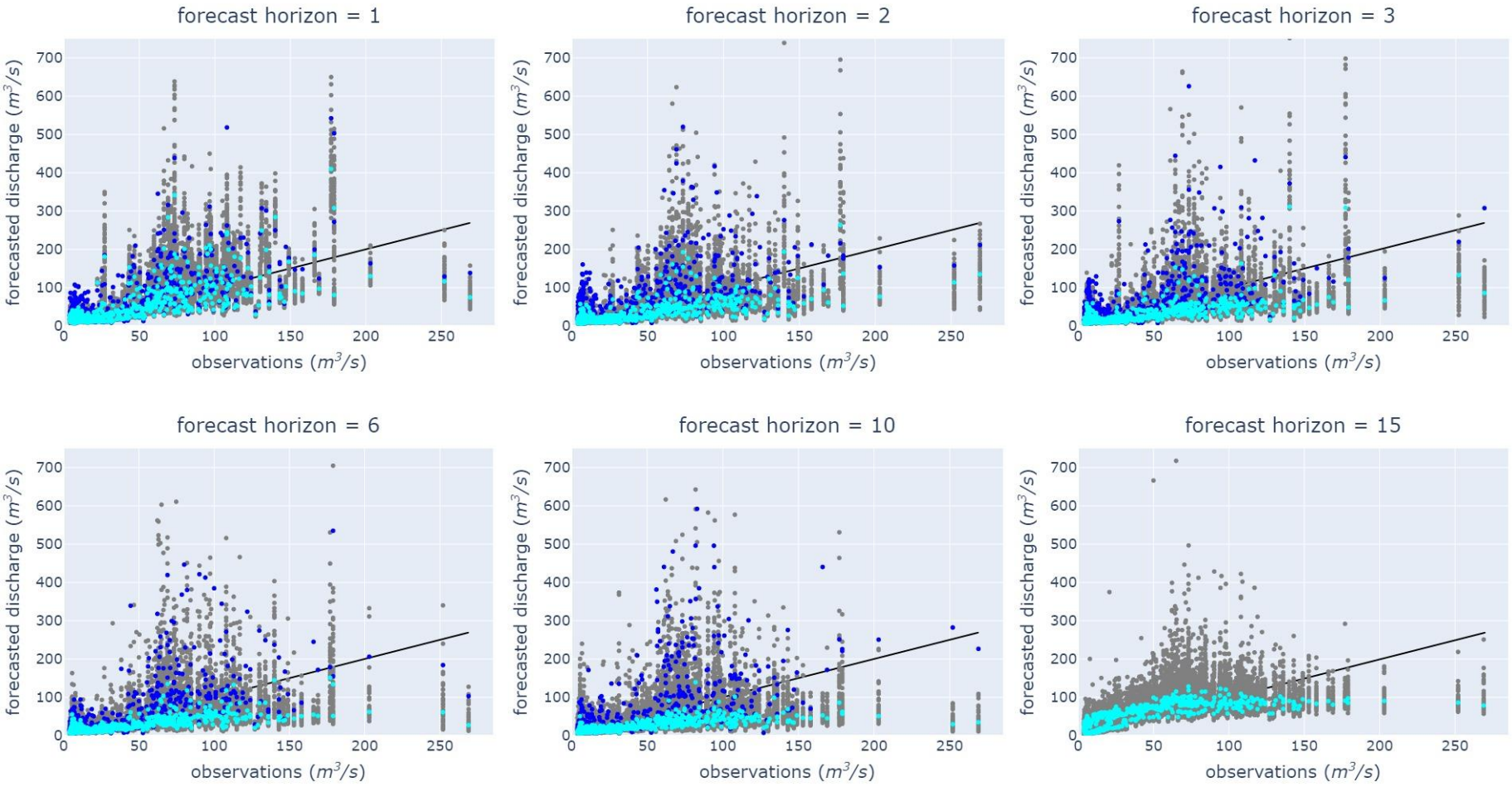


Figure 10: Scatter plot of river discharge for Tumlingtar, the far eastern site in this study. Grey dots represent the ensemble forecasts, cyan is the ensemble median, and blue is the high-resolution member. The solid black line represents the y = x line. Ideally, all the data points should fall along this diagonal line.

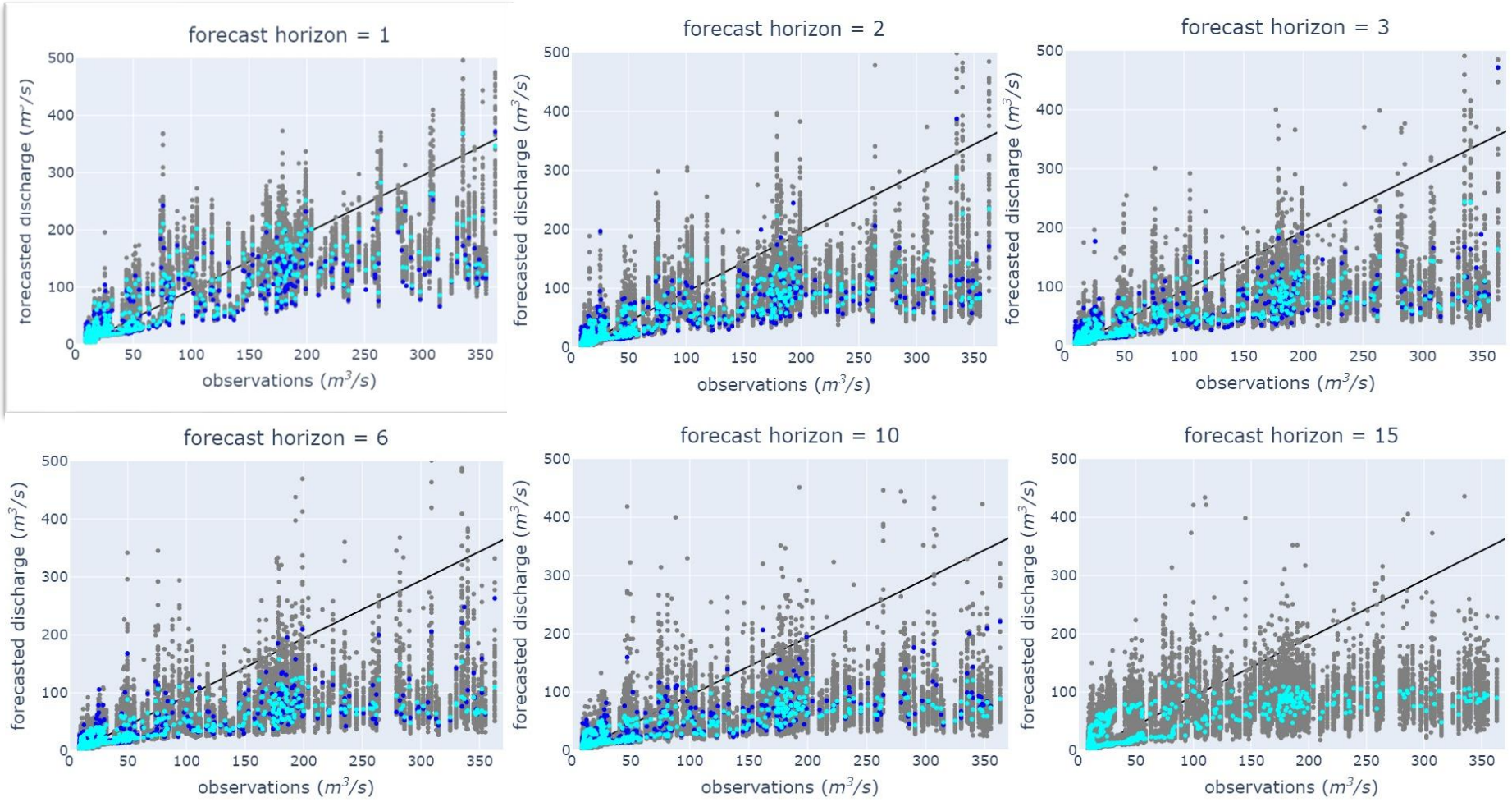


Figure 11: Scatter plot of river discharge for Balephi, the eastern site in this study. Grey dots represent the ensemble forecasts, cyan is the ensemble median, and blue is the high-resolution member. The solid black line represents the y = x line. Ideally, all the data points should fall along this diagonal line.

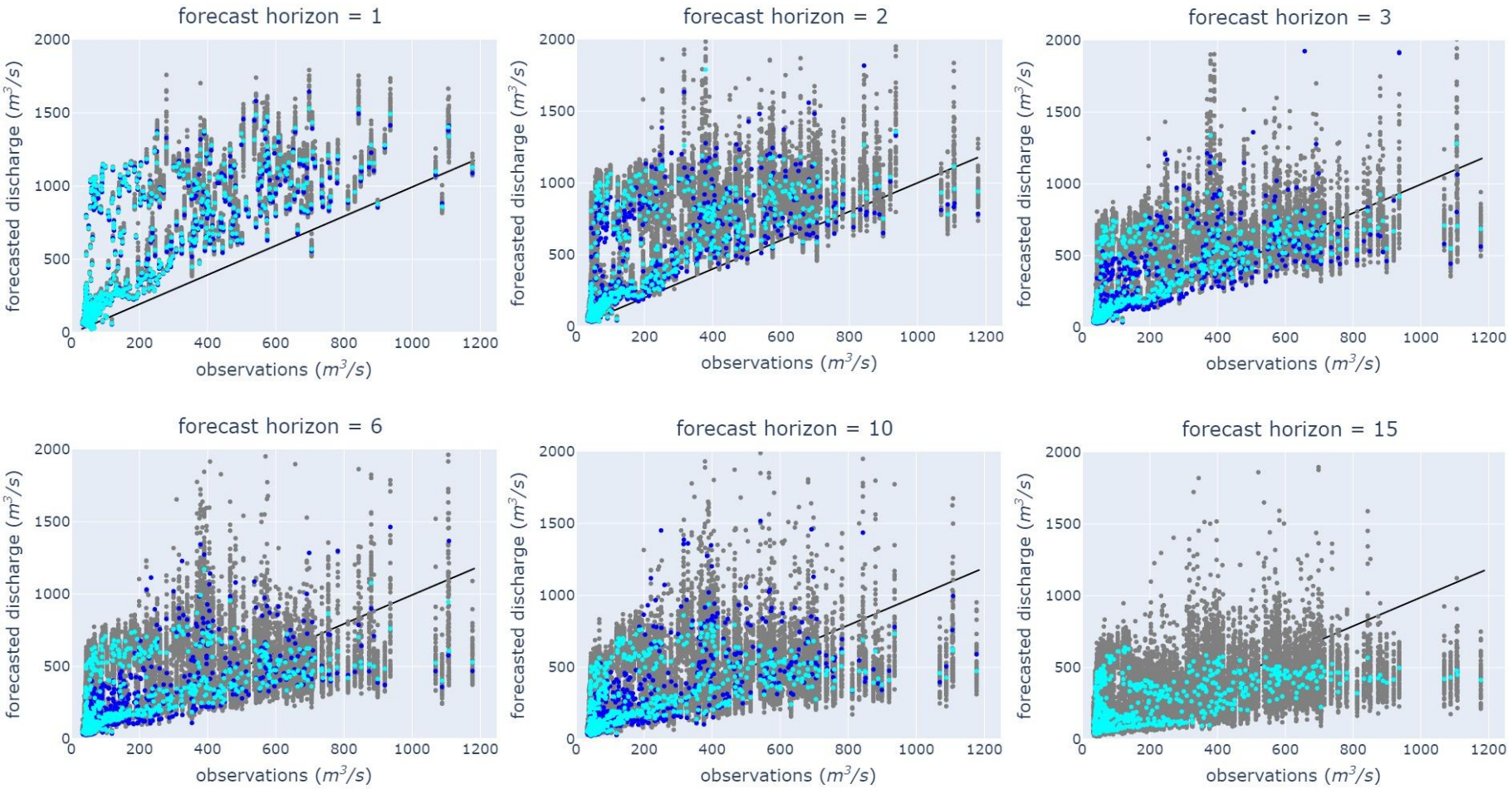


Figure 12: Scatter plot of river discharge for Marsyangdi, the western site in this study. Grey dots represent the ensemble forecasts, cyan is the ensemble median, and blue is the high-resolution member. The solid black line represents the y = x line. Ideally, all the data points should fall along this diagonal line.

4.1.2 Numerical verification

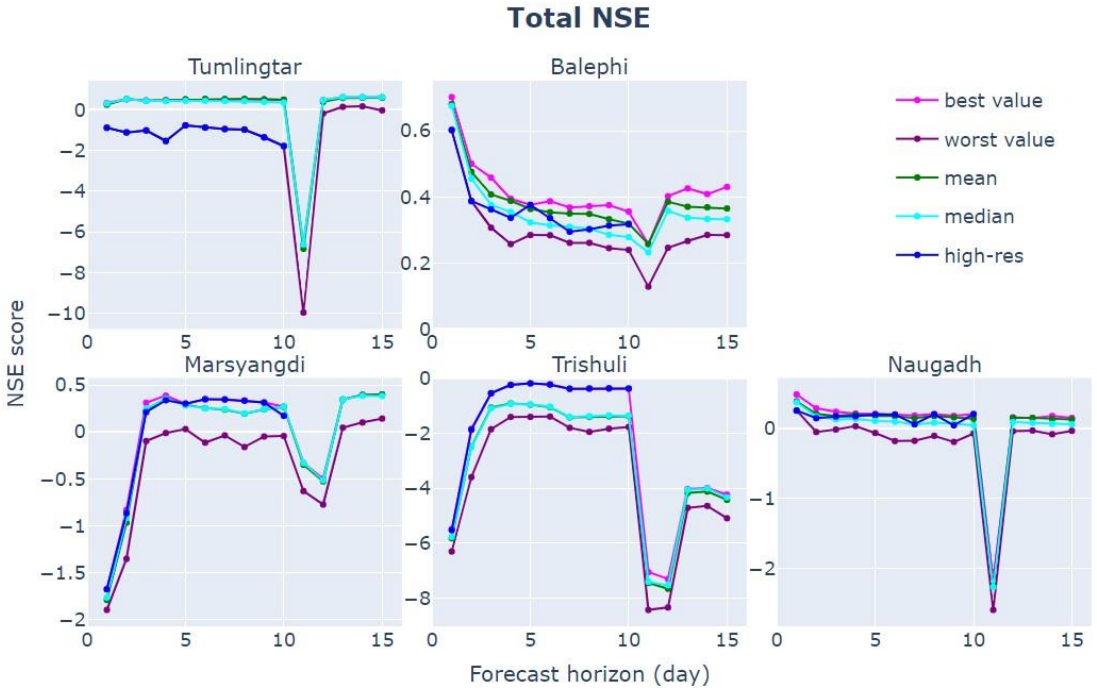


Figure 13: NSE for the entire river-discharge dataset. Values closer to 1 are preferred.

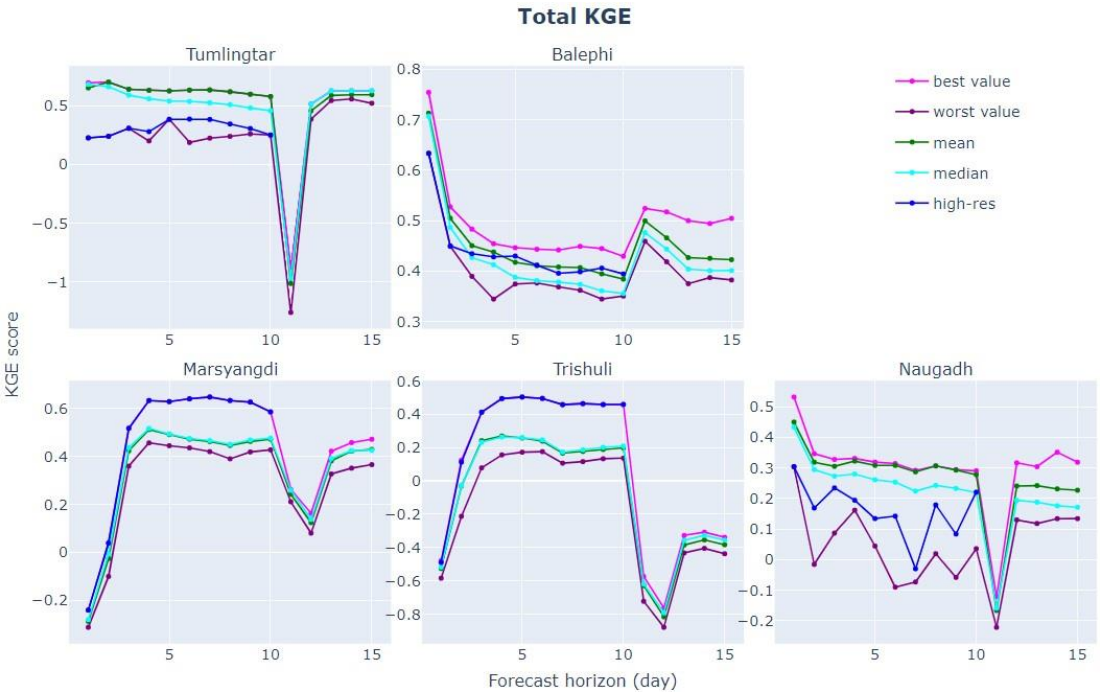


Figure 14: KGE for the entire river-discharge dataset. Values closer to 1 are preferred.

An analysis of the entire two years of available data is performed, which is shown by the NSE (Figure 13), the KGE (Figure 14), and the CRPS (Figure 15) plots. KGE already includes equal contributions from the flow variability, the bias, and the correlation metrics; hence, these three

individual metrics are not shown here. The KGE/NSE value of the best (worst) member for each forecast horizon is represented by magenta (purple). The best/worst value can correspond to separate ensemble members for different sites or forecast horizons.

The overall forecast quality of the GESS is poor. Balephi shows the highest NSE value for day-1 with an NSE value of 0.7. Based on the model ratings presented in Table 3, this corresponds to "Satisfactory" model performance. The performance of the best-valued ensemble member and the ensemble mean/median is in the "Not Satisfactory" range for most sites. For Trishuli, the NSE value is entirely on the negative scale, which means a forecast based entirely on climatology would have performed better than the GESS forecasts for the site. All sites show a decline in NSE on day-11 before climbing up again (see discussion in section 4.1.3.1). The KGE plots also follow similar trends as the NSE plots, except for the day-11 value for Balephi. This anomaly for Balephi is explained by assessing the time series plots in section 4.1.3.1.

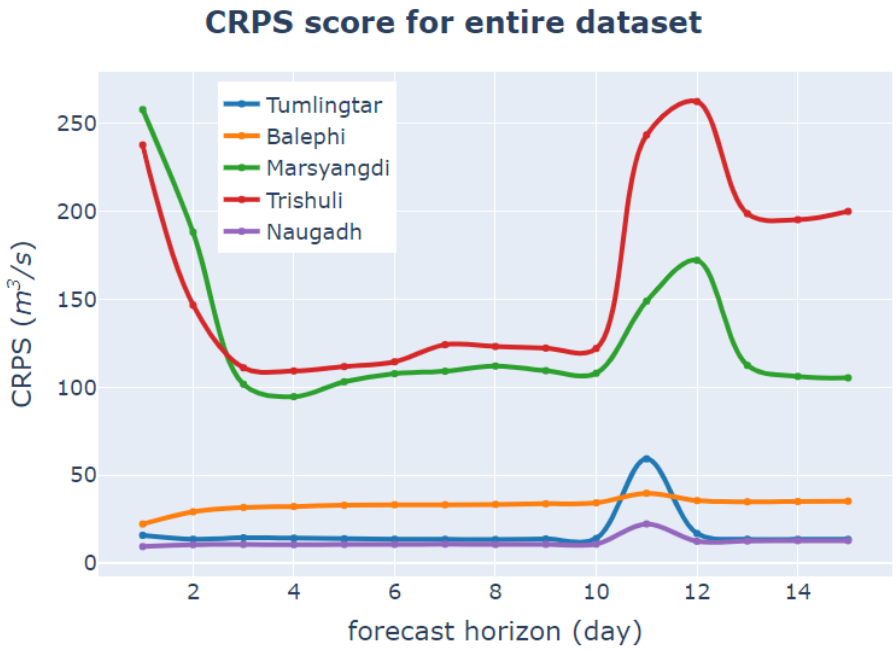


Figure 15: CRPS values of raw GESS forecasts for the entire dataset for the five sites.

Generally, the forecasts are expected to be most accurate in the shorter term, as it is closer to the present when we have the best possible knowledge of the state of the hydrometeorological variables. The first-order stream sites follow this expected trend. In contrast, the higher-order stream sites: Marsyangdi and Trishuli, show a poor forecast quality in the short term. The KGE and the NSE value are gradually improving until reaching a stable value starting on day-4. Furthermore, the CRPS plot starts with a very high error magnitude before stabilizing starting on day-3.

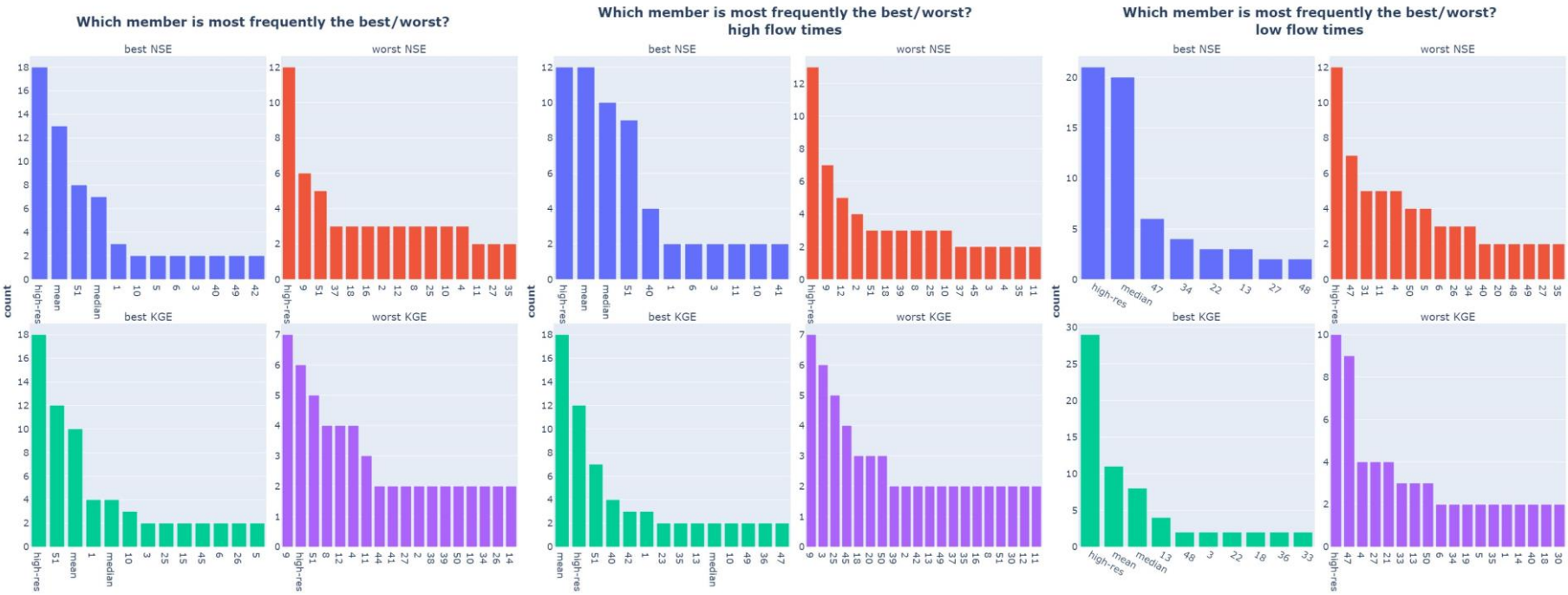


Figure 16: Which member occurred most frequently as the best/worst member for river discharge forecasts? From left to right: analysis across the entire dataset, at high-flow and low-flow times. Only members that were the best/worst member at least two times are included.

The overall shape of the NSE and the KGE plots for both sites is similar in the short-to-medium-term horizon. Both sites show a dip in the deterministic metrics on day-11 and 12 before recovering in the extended range. The reasoning behind the dip is explained in section 4.1.3.1. While the metrics return to positive for Marsyangdi, they stay on the negative values for Trishuli.

The high-resolution member was the best member for both Trishuli and Marsyangdi. However, it was often among the worst-performing members, especially based on the NSE metric. A closer inspection of the time series plots showed significant over forecasting of the flow volume. As NSE penalizes the flow difference, the high-resolution member is highly penalized in NSE compared to the KGE. The most frequent best/worst members across the two years can be plotted using a bar chart, as shown in Figure 16. The high-resolution member ranks among the most frequent best and the worst member, resulting from a stark difference in its performance for the different sites.

The ensemble mean outperformed the ensemble median in the first-order sites, while the ensemble median slightly outperformed the mean for both higher-order stream sites. This result is consistent with several literature findings that suggest that the ensemble mean outperforms individual forecast members over multiple verification intervals. While the difference is pronounced for the first-order sites, it is not as pronounced for the higher-order sites. Figure 16 also shows that both the ensemble mean and the ensemble median are among the high-performing deterministic forecasts for the entire dataset.

4.1.3 Time series plot

A time series analysis is used to investigate three main issues that were seen in the forecasts:

- Excessive streamflow values for day-11 forecasts
- Faster hydrograph rise in the pre-monsoon season
- High fluctuations in the river flow in the dry season

4.1.3.1 Excessive streamflow values for day-11 forecasts

The KGE and the NSE plots in section 4.1.2 show a marked decrease in the quality of day-11 forecasts. Figure 17 shows the time series for the forecast horizons of day-10 through 12 for Tumlingtar, one of the sites that had the aforementioned issue. There is a stark increase in the streamflow values on day-11 forecasts compared to day-10 forecasts, and the flow values decrease again on day-12. The streamflow value for a specific date can differ across different forecast horizons. However, the total volume of forecasted water across the entire dataset year should be similar. Here, the forecast horizons 12 – 15 have similar trend and flow ranges, making the increased forecast volumes seen for day-11 an aberration.

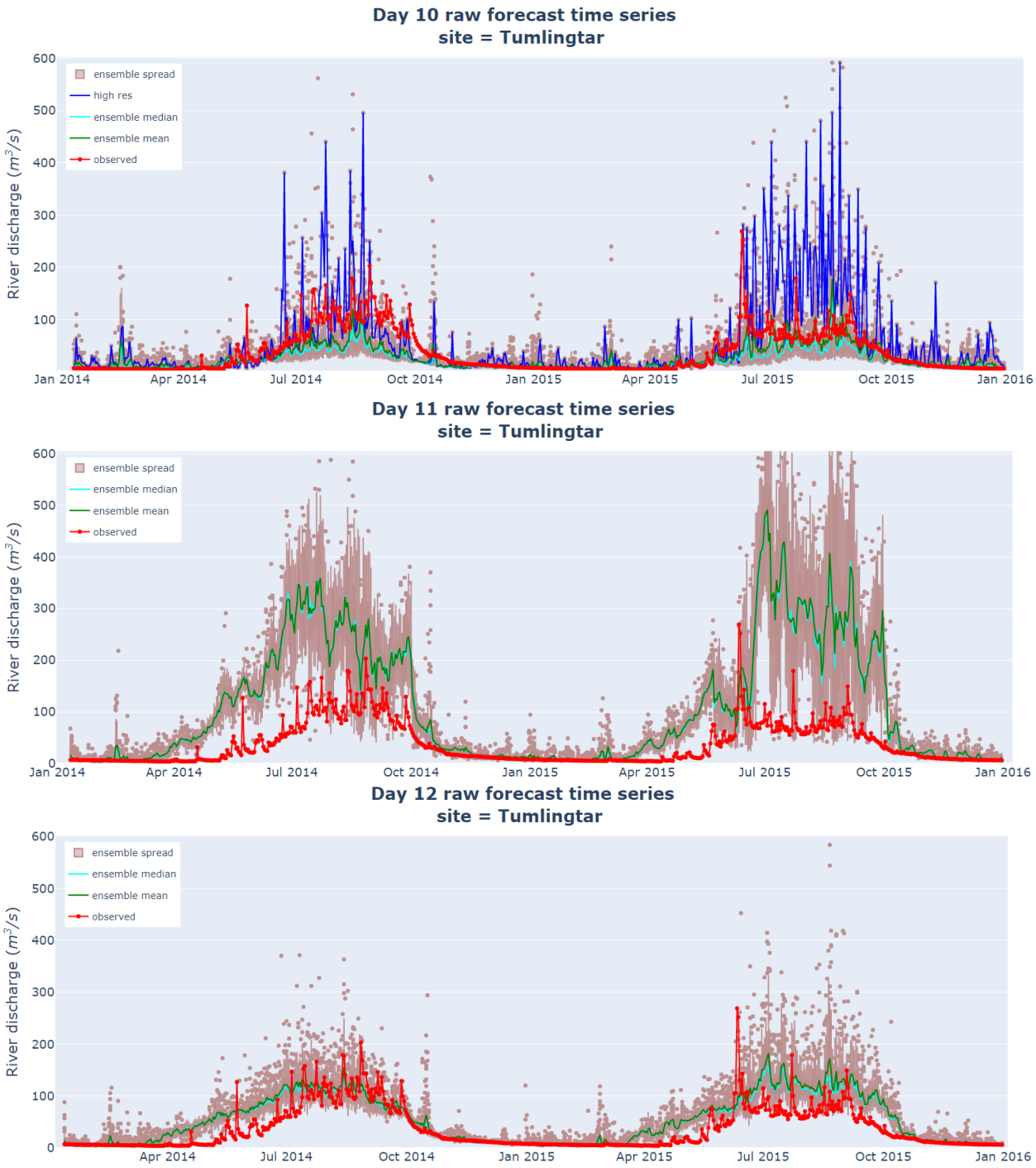


Figure 17: Time series plots for day-10 to 12 forecast horizons for Tumlingtar (far eastern site). A sudden increase in the forecast volume is seen for day-11, which is not seen for day-10 or day-12 forecasts.

Tumlingtar does not have any other upstream rivers, as it is a first-order stream segment. For first-order streams, the only input for the RAPID routing is the area-weighted runoff values from the ECMWF grids that overlap with the stream segment's catchment (see section 3.1.2). The RAPID

routing is a mass conservation process. Hence, the streamflow volume depends entirely on the magnitude of the catchment-wide runoff and the streamflow forecasts from the upstream segments.

Figure 18 compares the catchment-wide day-10 and day-11 total runoff forecasts for Tumlingtar. The catchment-wide runoff was calculated using equation (1), and the process is explained in 3.1.2. While there is an increase in the runoff volume from day-10 to day-11 forecasts, the increase is not almost two-fold, as seen in Figure 17. Hence, this could be an error in the reforecast creation process.

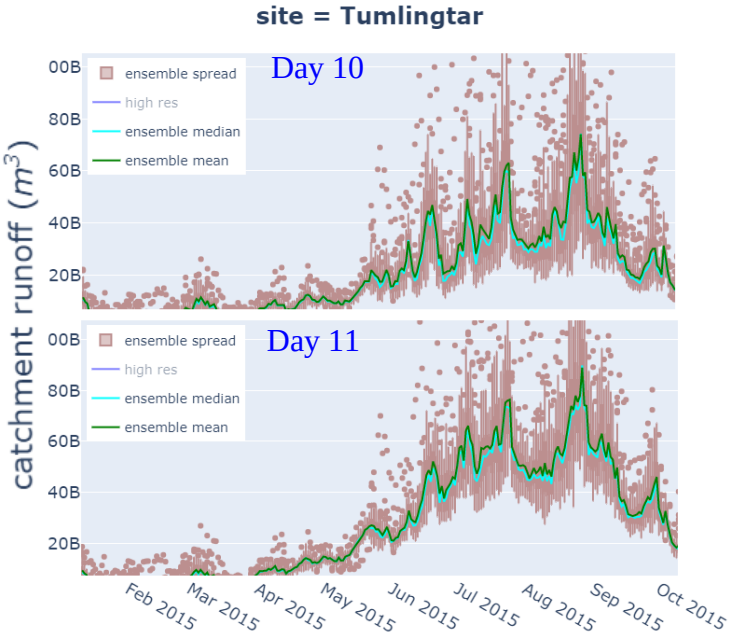


Figure 18: Catchment-wide total runoff time series for forecast horizons 10 and 11 for Tumlingtar.

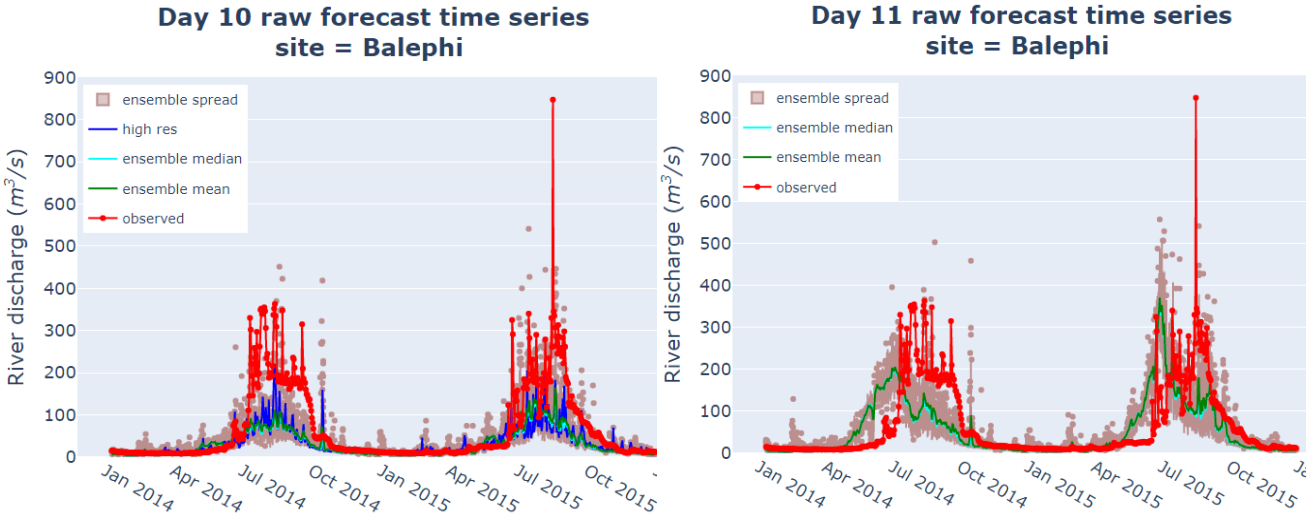


Figure 19: Comparison of time series for day 10 (left) and day 11 (right) GESS forecasts for Balephi.

In contrast to other sites, the GESS forecast for Balephi has an increase in the KGE score for day-11 forecasts. An analysis of the time series plots (Figure 19) shows an under-forecasting tendency for the site until day-10. The increase in the forecasted volume on day-11, especially in the summer, improved the bias and the flow variability. This early summer excess volume is not seen in the observations. Hence, the correlation component of the KGE was penalized. However, the gains in the former two components were adequate to improve the KGE score for the period. For Balephi, the effect of this change is seen in the CRPS plot (Figure 15) as it has the smallest degradation in the day-11 values compared to the other sites.

4.1.3.2 Faster hydrograph rise in the pre-monsoon season

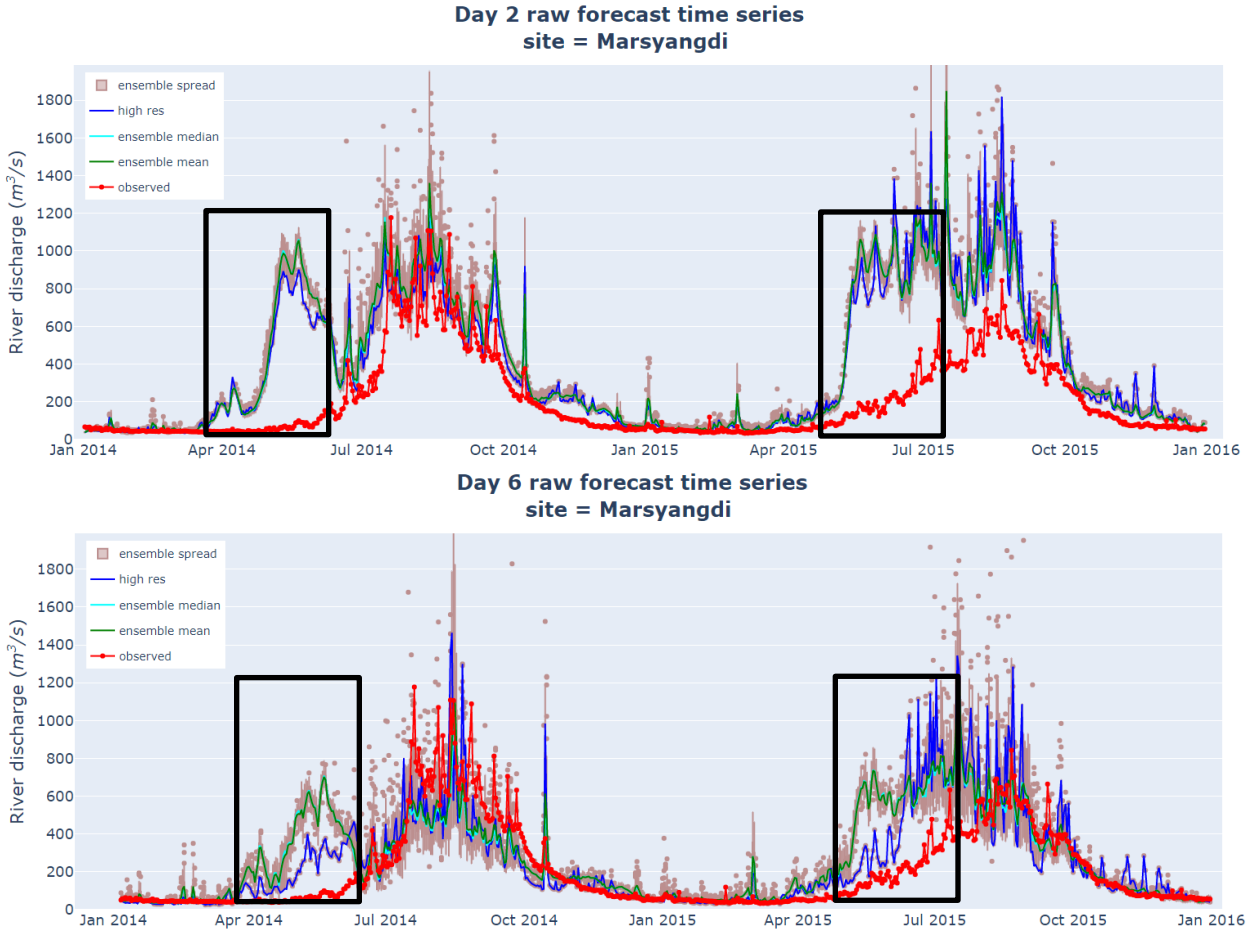


Figure 20: Time series plots for day-2 and day-6 forecasts for Marsyangdi. The forecast hydrograph shows a steep rise in the late summer/early spring season (April/May).

The forecast time series for Marsyangdi and Trishuli shows peculiarly high streamflow values in the pre-monsoon season (see Figure 20) compared to the observations. One of the hypotheses for this behavior was that the upstream hydropower dams might have stored the extra volume of water during the pre-monsoon season, which caused a lack of water downstream that was not accounted for in the forecast. The Marsyangdi Hydropower Project (MHP), which is the site of interest in this study, has two peaking ROR type projects (see Figure 21) upstream: the Middle Marsyangdi

Hydropower Project (MMHP) and the Upper Marsyangdi Hydropower Project (UMHP). The UMHP is a recently constructed project whose reservoir construction began in late 2013.

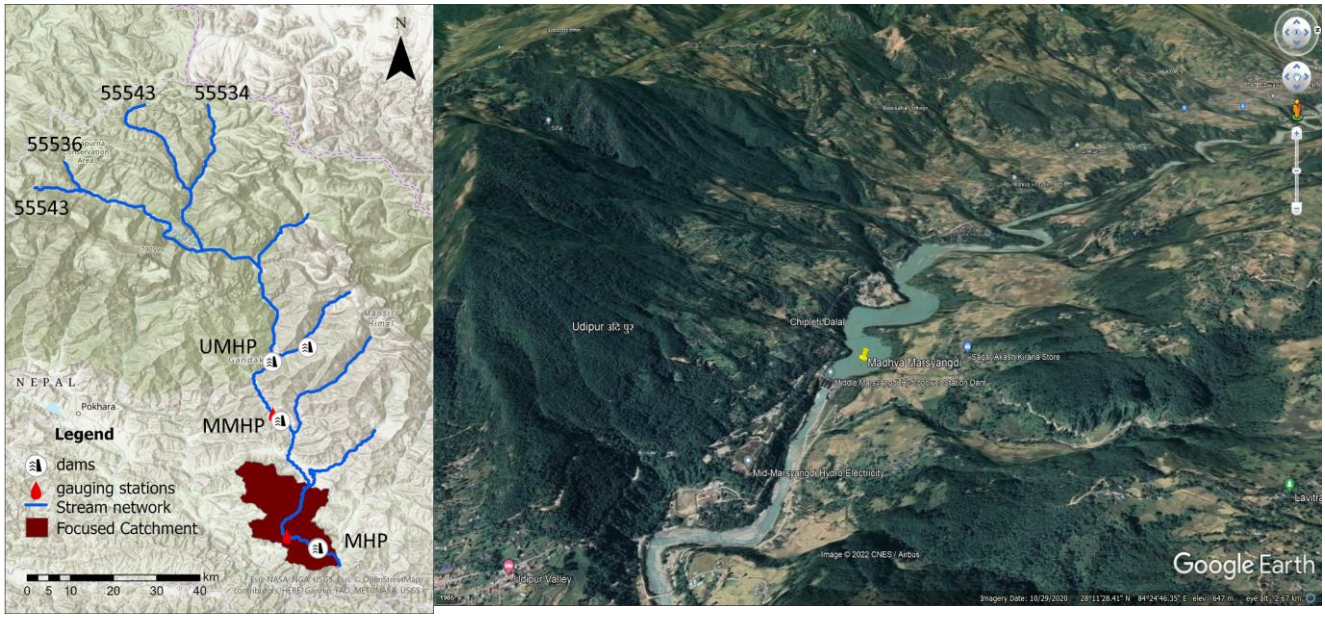


Figure 21: (left) The upstream network of the Marsyangdi basin. The GESS forecast was evaluated for the catchment marked in maroon. Red drops represent gauging stations.
(right) Satellite imagery from Google Earth showing some water pooling at the MMHP site.

The only other gauging station of interest in this network is just upstream of the MMHP dam site (see Figure 21). The construction of the UMHP dam had not started until late 2013. Hence, the observation recorded before this period represents unimpeded flow. Figure 22 (top) shows the upstream observation measurements (in blue) for 2012. The observation hydrograph for the upstream stations from 2011 until mid-2013 does not show the steep rise in the volume as seen in the forecasts. Hence, only the data for 2012 is shown as a representative of the unimpeded flow.

Figure 22 also compares the discharge measurements for the two gauging stations for 2012 and 2015. The general shape of the hydrographs for both stations is similar, with the primary difference being the increased volume, which is expected as the two stations are ~40 km apart with a ~300 m elevation difference. Furthermore, the pooling capacity of MMHP is 1.65 million m^3. With a conservative estimate of 100 m^3s^{-1} for the inflow, the reservoir would be at its capacity in 4.5 hours. Hence, in contrast to the forecasts, the reservoir holding capacity is not large enough to cause seasonal variation in downstream flow.

However, the upstream MMHP does influence the flow downstream. A closer inspection of Figure 22 shows spikes in the observation data for flows received by MHP, which are not present in the upstream gauge. This is possibly due to the evacuation of the MMHP reservoir for maintenance

purposes. As GESS does not account for human processes in the forecasting process, it is seen in upcoming sections that the model is penalized for missing these "events".

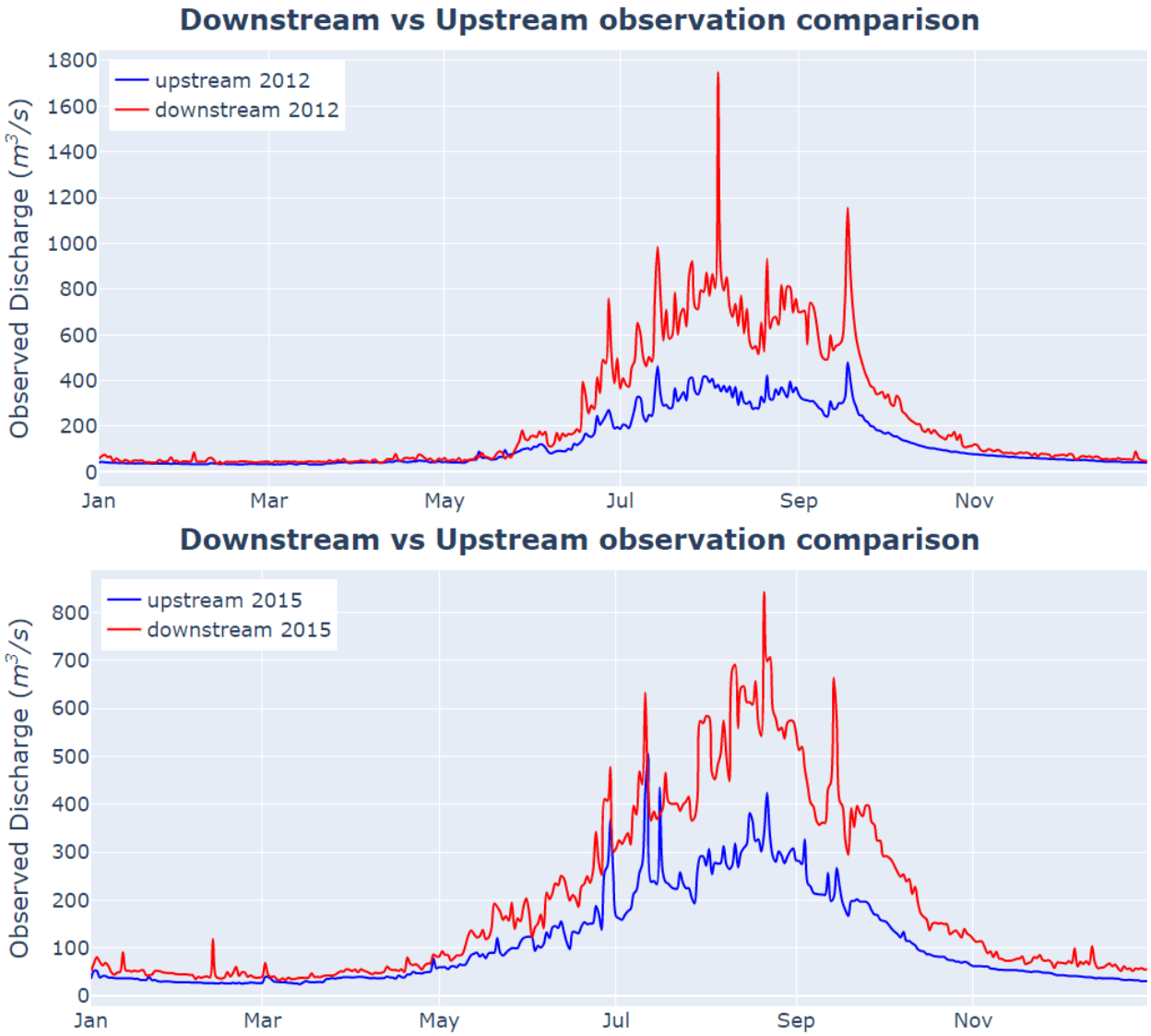


Figure 22: Observations recorded by gauging stations at the MMHP (blue) and the MHP (red) for the years 2012 (top) and 2015 (bottom).

Figure 23 shows the progression of the forecasted hydrograph from the uppermost four reaches (see Figure 21 *left*). It is seen that the uppermost reaches influence the overall shape of the hydrograph. These four reaches correspond to a heavily glaciated region, as seen by the aerial imagery taken on March 25, 2015, shown in Figure 23. Another site with similar results, Trishuli, also has a long waterway with glaciated sources (not shown). The smooth forecast hydrograph rise and the highly glaciated region suggest that the LSM is overestimating the runoff due to snowmelt at these locations.

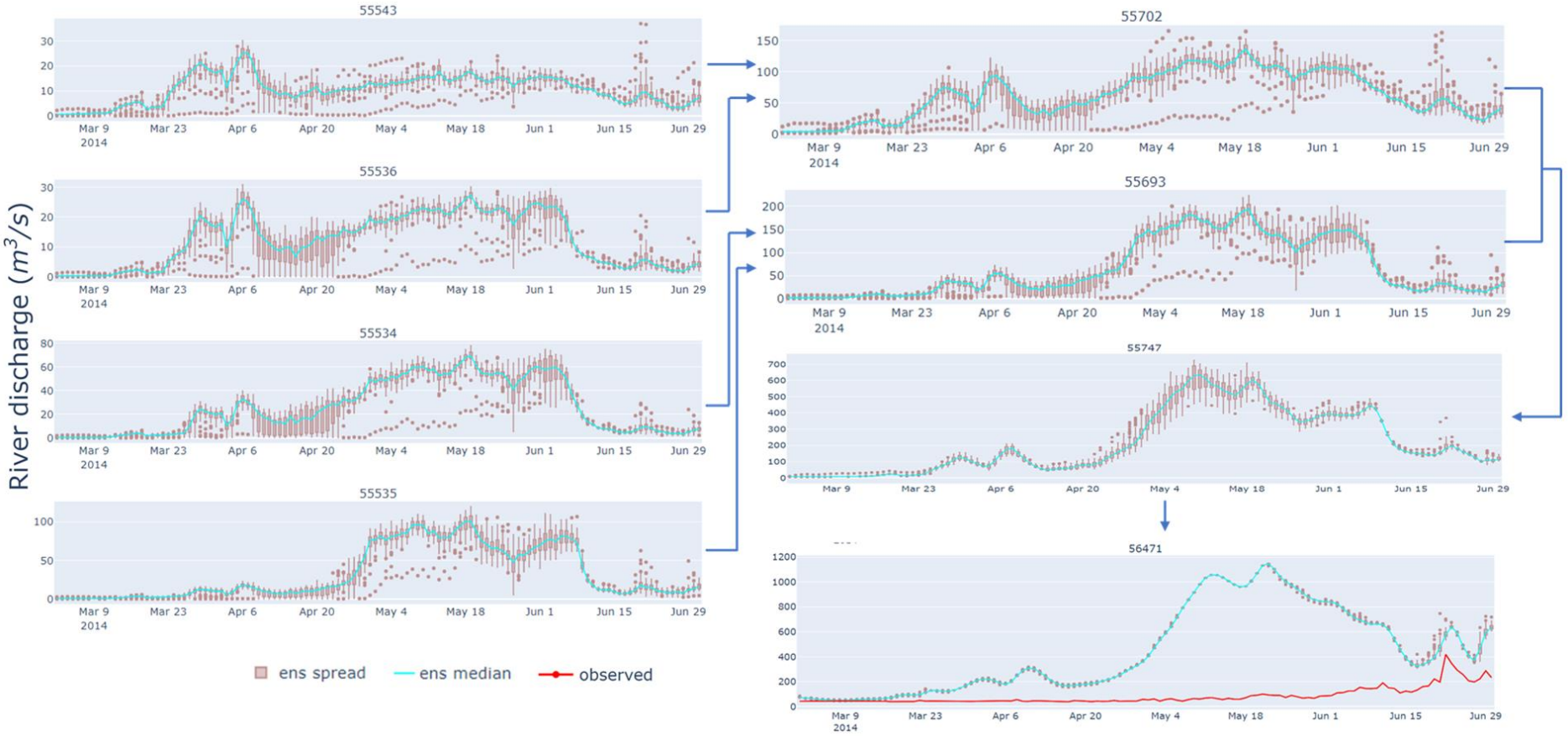


Figure 23: Progression of the forecast hydrograph from source (leftmost) to the site of interest. Note the difference in the y-axes scales.

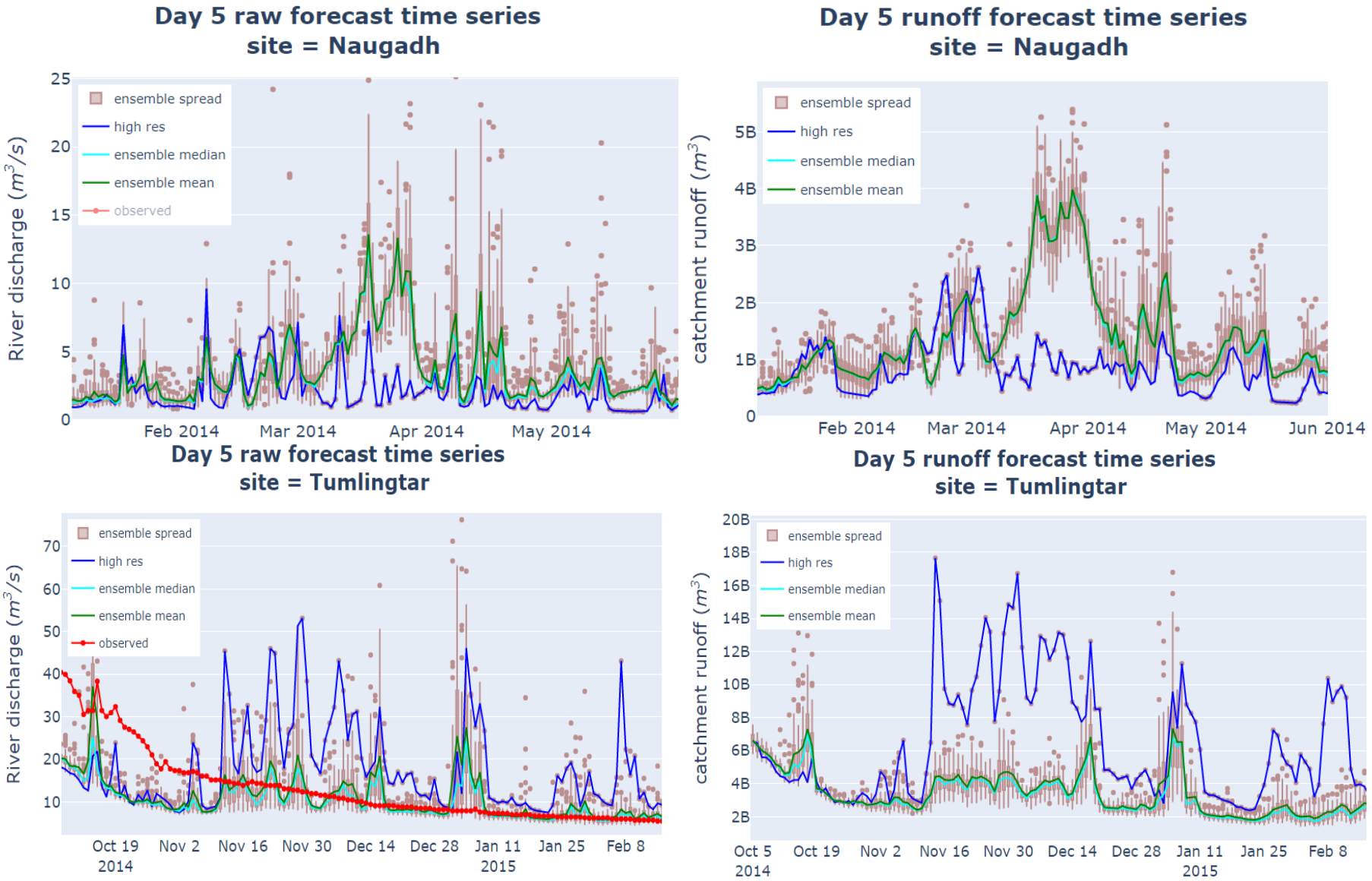


Figure 24: High fluctuations in the forecasted streamflow shown for two sites. The runoff forecast plots for the respective sites are shown on the right.

In section 4.1.2, it was noted that the forecast skill for the higher-order sites was poor in the short-term horizon before improving in the medium-term range. Figure 20 shows that the spring runoff generation is excessively high in the short-term range (day-2) compared to the medium-term forecasts (day-6). This trend was seen for Trishuli as well. The short-range forecasts are penalized more than the medium-range due to the higher degree of over-forecasting of the spring runoff.

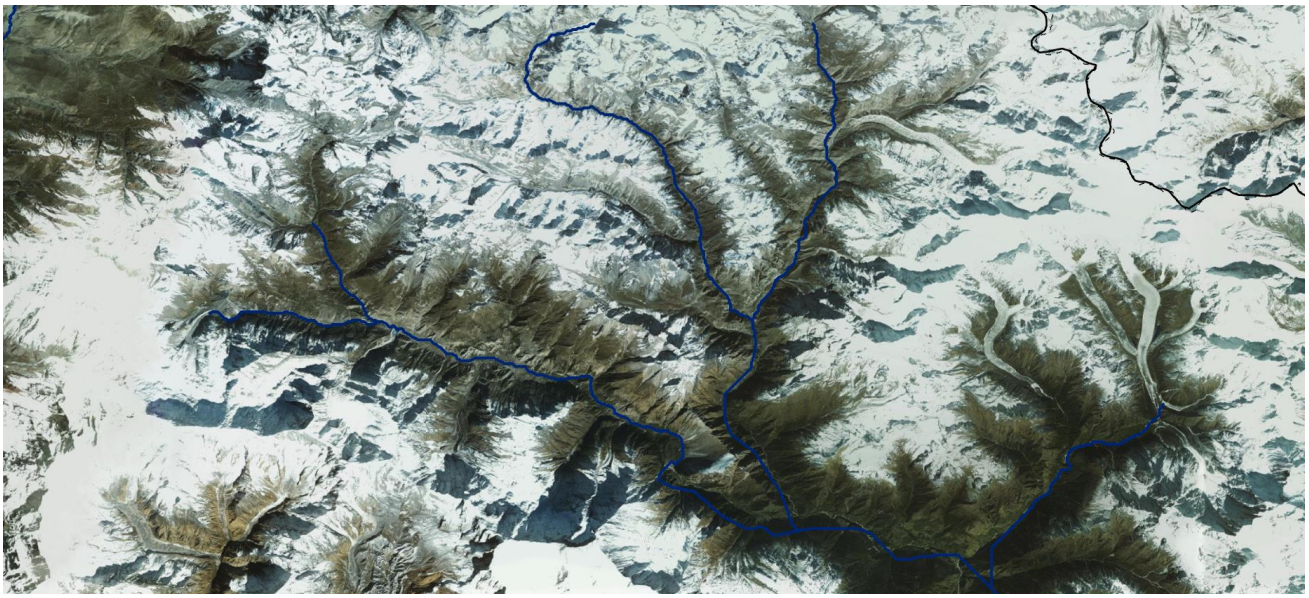


Figure 25: Aerial imagery of the uppermost stream segments (shown by blue lines) that feed into Marsyangdi. Source: ESRI, World Imagery

4.1.3.3 High fluctuations in the river flow in the dry season

The forecast showed high inter-day fluctuations during the dry season that were not seen in the respective observation hydrographs. The fluctuations were more pronounced in the flashy first-order streams, especially with the high-resolution member, as shown in Figure 24. Figure 24 also shows the total catchment-wide runoff fed into the RAPID model to produce the streamflow forecasts. The forecasted hydrograph and the runoff trends are similar, showing that this behavior can be attributed to the LSM's representation of the hydrological processes.

4.1.4 Limitations of the HTESSEL LSM

Balsamo et al. (2009), who introduced the HTESSEL LSM, noted that higher runoff generation was seen with the configuration due to the errors in the snow scheme. In particular, faster runoff generation in the spring season was noticed because there is no refreezing mechanism for the water in the snowpack. This is problematic because the diurnal temperature and insolation changes in the high mountains encourage patterns of snowmelt in the daytime and refreezing of a portion of the runoff in the night hours.

Furthermore, the resolution of ~18 km is still coarse for such complex terrain. The averaging of vegetation classes in that resolution causes a loss of information in a place like Nepal, where there

are significant climatic zone changes in that distance. While HTESSEL accounts for sub-grid variability in orography, it does not account for the aspect and the slope of the mountain faces (EMCWF, 2022). So, specific grid cells might receive greater insolation in the modeling process than in the real world, causing a more significant runoff generation through snowmelt.

4.2 Bias correction outputs

The results of the simple statistical bias correction are presented here. First, the results for the calibration questions presented in section 3.5 are shown in section 4.2.1. Verification of the bias correction for the independent 2015 dataset is presented in section 4.2.2.

4.2.1 Calibration phase

The combinations of four DMB formulae, two bias correction approaches, and eight window lengths are tested on the calibration dataset.

The ensemble mean was among the best performing raw forecast (see Figure 16), and its behavior is typical of all the ensemble members. Hence, the calibration results were first based on the ensemble mean. Based on the deterministic metrics, the results heavily favored the second bias correction approach, which is performed by applying the same DMB value to the 52 ensemble members. The common DMB value is the mean of the 52 DMB values, which is the same as calculating the DMB using the ensemble mean. So, this approach is expected to provide the best result for the ensemble mean. Naturally, this would mean that the calibration results rely on the deterministic metrics of the ensemble mean, which would favor the second bias correction approach. The deterministic metrics of the ensemble median are used instead of the ensemble mean for the calibration to counter this imbalance.

The main results from the calibration process are as follows:

- Shorter bias correction windows are favored, with 2-3 days the most frequent.
- The unweighted DMB, as mentioned in equation (10), performed the best overall.
- The first bias correction approach, where each member is bias-corrected individually, was favored more than the second, where the same DMB ratio is applied to all 52 members. However, there was no clear winner.

For cases where there is a disagreement between the deterministic and the probabilistic metrics, the comparative gain in the CRPS and the deterministic metrics for the two best combinations are compared, and the one that leads to the highest gain was chosen.

In many cases, the calibration yielded different combinations for the high- and the low-flow seasons. During the calibration phase, the forecast skill improvement after bias-correction was generally higher for the low-flow season than for the high-flow season. Hence, the metric that yielded the best result for the high-flow season was selected. Upon comparing the metric values

between the best high-flow and the best low-flow metrics, the differences were minimal, and both combinations could be interchangeable in most cases.

Table 4: Differences between the best combination for the high- and low-flow season for day 13 forecast at Trishuli.

	High-flow	Low-flow
NSE	0.838 (+0.007)	0.884 (+0.000)
KGE	0.911 (+0.001)	0.933 (-0.006)
CRPS	30.09 (-0.35)	4.12 (+0.04)

As an example, for Trishuli on day 13, the best result for high-flow is for the second bias correction approach, whereas the best results for low-flow are for the first bias correction approach. The second bias correction approach is selected for this forecast horizon based on the reasoning above. Table 4 highlights the final score values based on the chosen configuration. The number in the bracket shows the change in the final score due to the selection of the second bias correction approach over the first for each flow season. While the relative gain in CRPS is apparent, the gain/loss magnitude for both the metrics is still small enough that both combinations could be interchanged.

Table 5: Final bias correction configuration to be applied to the verification dataset.

Site	window length	DMB formulation	Bias correction
Tumlingtar	2 (3 on day 11)	DMB [eq. (10)]	Approach 1 (2 on day 11)
Balephi	2		Approach 1 (2 on day 6)
Trishuli	2		Approach 1
Marsyangdi	2		Approach 2

Based on these decisions, the final calibration configuration used with the verification year is shown in Table 5. The reader is directed to Appendix C for the raw calibration results that identify the best configuration for each forecast horizon and flow condition for all the sites. Results for Naugadh are not shown as the site did not have observation data for 2015, which is the verification year in this study.

4.2.2 Verification phase

Figure 26 shows the most frequent best/worst member throughout the verification year, the low-flow and the high-flow period. While the high-resolution member was consistently the best for the raw forecasts, the bias correction method was not as effective for this member. As such, it was frequently among the worst of the bias-corrected members but never among the best members. The ensemble median and the ensemble mean were the best members, with the ensemble median being more frequent than the mean.

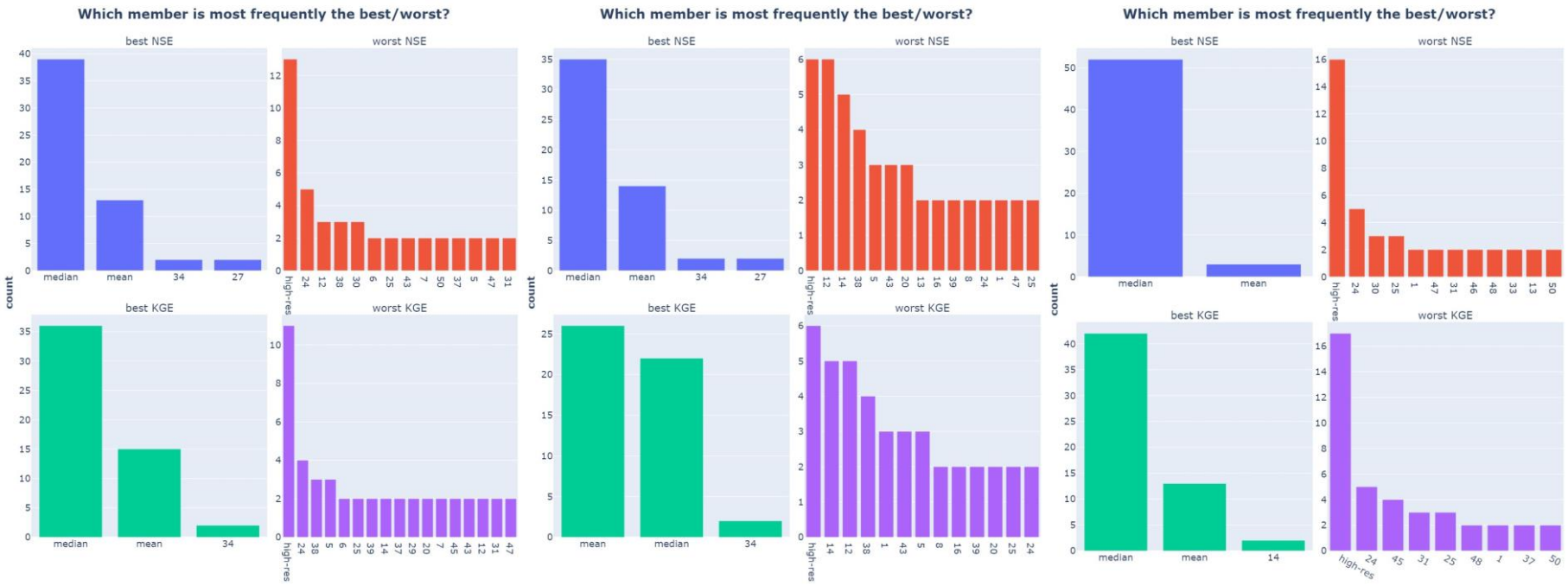


Figure 26: Most frequent best/worst member post bias correction based on the verification dataset (2015). From left to right: analysis across the entire dataset, at the high-flow times and the low-flow times. Only members that were at least two times the best/worst member are included. This chart is analogous to Figure 16, which showed the results for the entire two years of raw forecasts.

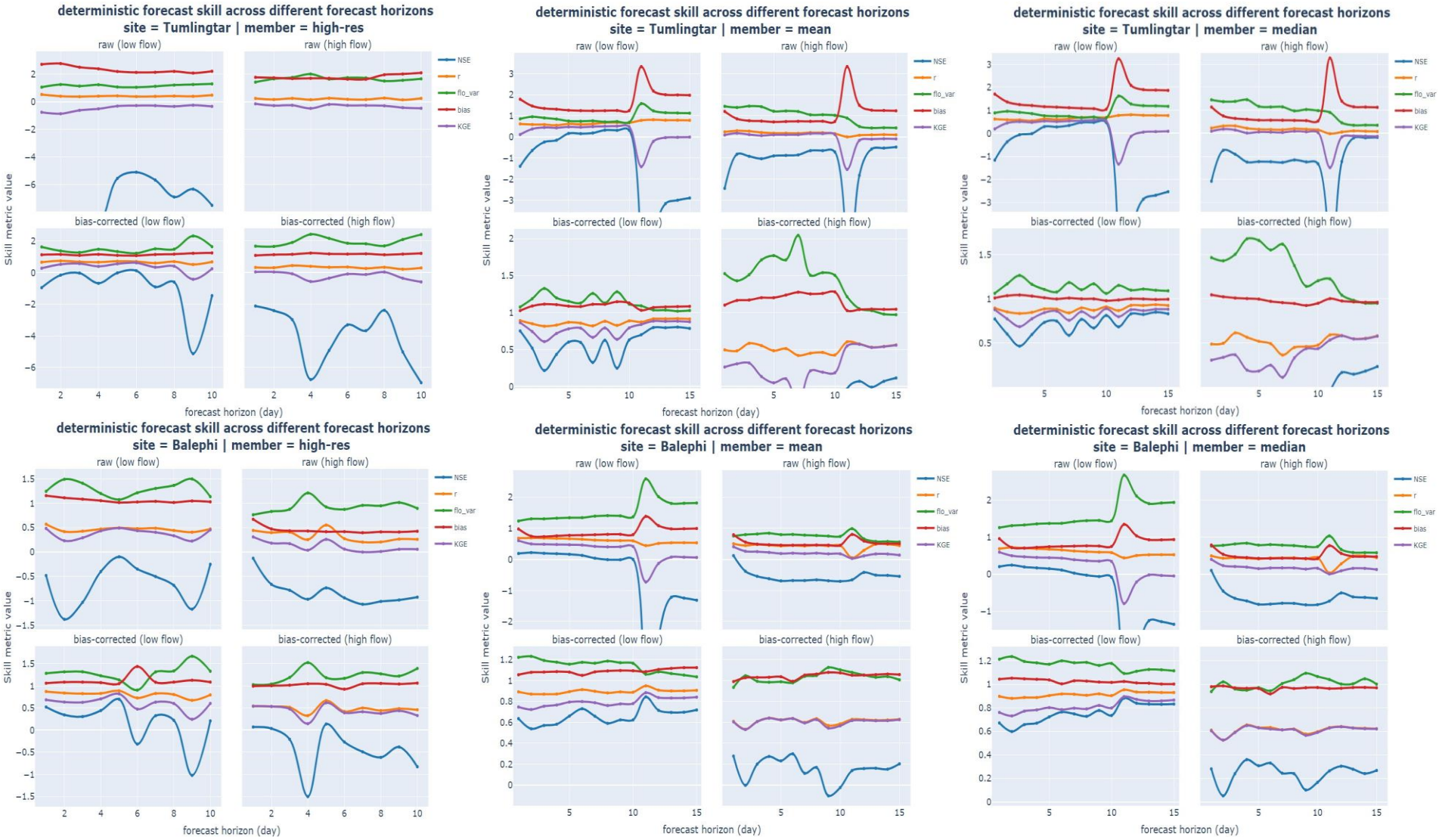


Figure 27: Deterministic metrics for the bias-corrected forecasts for first-order streams: Tumlingtar (top row) and Balephi (bottom row). The NSE curves have been cut-off for some sites to improve the visibility of other metrics. 1 indicates a perfect forecast across all the metrics.

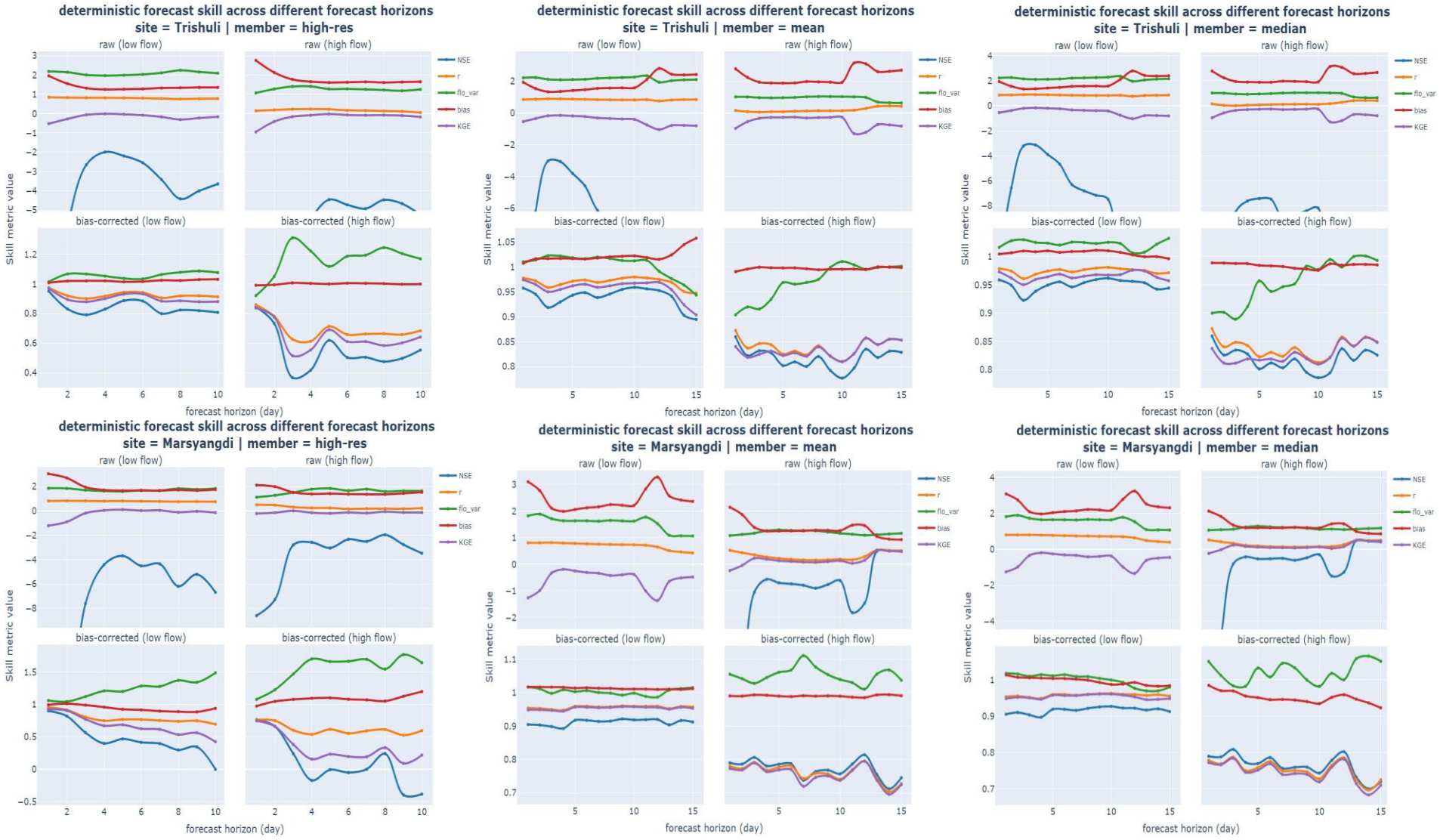


Figure 28: Deterministic metrics for the bias-corrected forecasts for higher-order streams: Trishuli (top row) and Marsyangdi (bottom row). The NSE curves have been cut-off for some sites to improve the visibility of other metrics. 1 indicates a perfect forecast across all the metrics

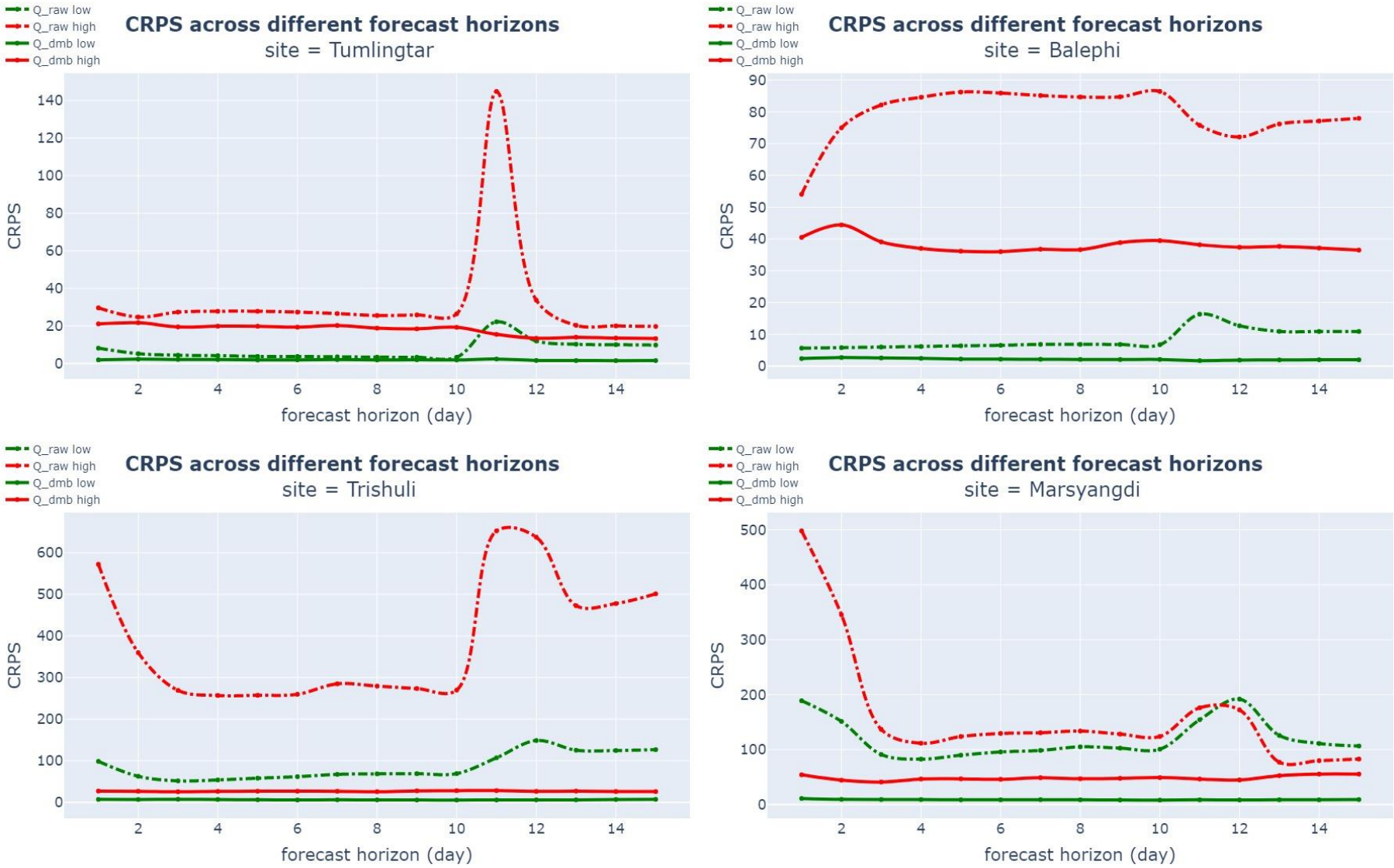


Figure 29: CRPS scores for the raw (dashed) and the bias-corrected (solid) forecasts for the four sites broken down by the high (red) and low (green) flow seasons. Lower CRPS indicates a better forecast.

Figure 27 – 28 show the deterministic skill metrics for the four sites. The three deterministic forecasts presented are: (1) the high-resolution member, which was one of the best members for raw forecasts, (2) the ensemble mean, and (3) the ensemble median, which was among the best performing for bias-corrected forecasts. These results are discussed next.

Forecasts show an improvement in most scenarios. The shape of the individual metric plots for the raw and bias-corrected forecasts does not align because the magnitude of improvement varies with the forecast horizon. The trends in the NSE and KGE curves for the ensemble mean and the ensemble median are similar for all the sites. However, the ensemble median had overall better values than the ensemble mean.

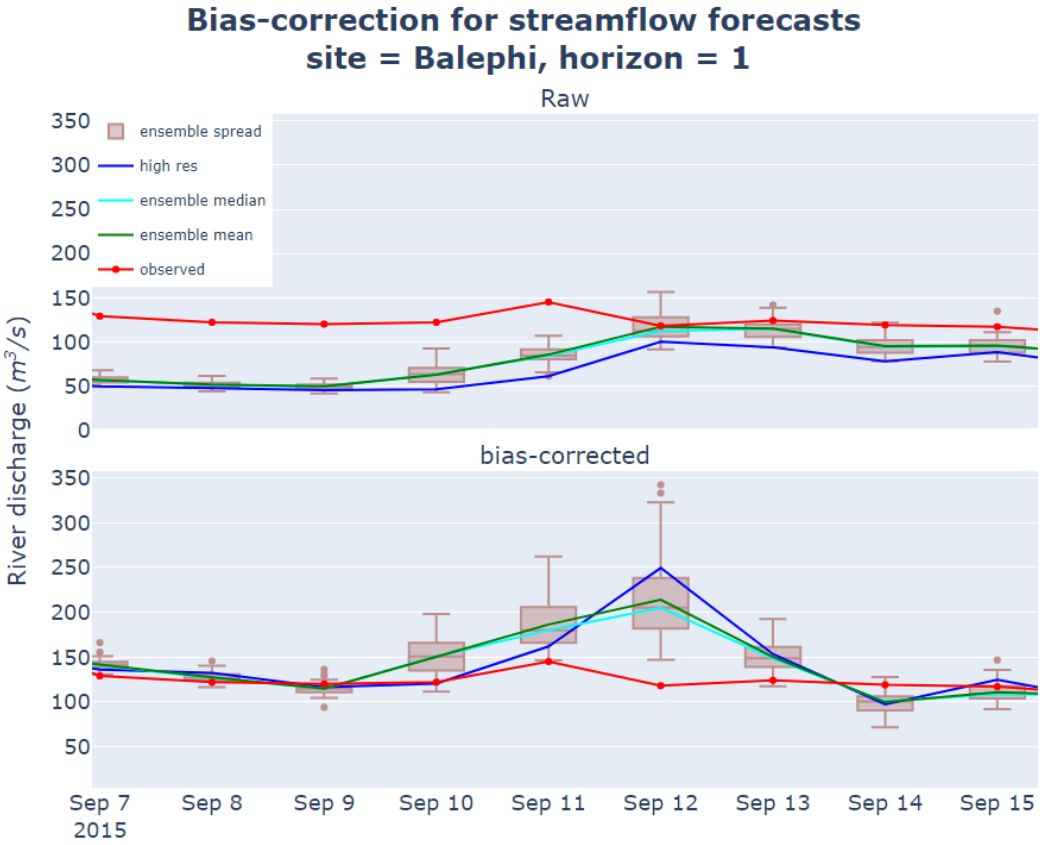


Figure 30: Pseudo peak-flow event introduced by the bias corrector. Compare the raw and bias-corrected results for Sept 11 and Sept 12.

The performance of the bias corrector is much better in the low-flow than in the high-flow season. The high-flow season has multiple peak-flow events, which do not occur during the low-flow season. If the forecast missed a peak-flow event, the bias corrector applies the difference to the next forecast. The forecast will then create a peak-flow event when the raw forecast did not predict one. See Figure 30 for an illustration of one such scenario. This shows the importance of the raw forecast events matching the observation trends, which the simple DMB correction methods cannot accurately correct. As the high-flow season had more of such “misses, " the bias corrector's relative improvement is also lower.

Table 6: NSE value range for bias-corrected forecasts for the study sites. Values closer to 1 are preferred.

	High-resolution		Mean		Median	
	high	low	high	low	high	Low
Tumlingtar	−6.97 → −2.12	−5.13 → 0.12	−4.96 → 0.12	0.22 → 0.80	−1.24 → 0.23	0.46 → 0.82
Balephi	−1.50 → 0.14	−1.02 → 0.70	−0.10 → 0.30	0.54 → 0.84	0.05 → 0.36	0.60 → 0.88
Trishuli	0.37 → 0.84	0.80 → 0.95	0.77 → 0.86	0.89 → 0.95	0.78 → 0.86	0.92 → 0.96
Marsyangdi	−0.38 → 0.76	0.00 → 0.90	0.71 → 0.81	0.89 → 0.92	0.70 → 0.81	0.90 → 0.93

The results in Figure 27 – 28 show the deterministic skill metrics for the four sites. The three deterministic forecasts presented are: (1) the high-resolution member, which was one of the best members for raw forecasts, (2) the ensemble mean, and (3) the ensemble median, which was among the best performing for bias-corrected forecasts. These results are discussed next.

also show that the bias correction resulted in much better forecasts for the higher-order streams than for the lower-order streams. The ensemble median for the higher-order stream sites of Marsyangdi and Trishuli had a forecast performance of "*very good*" in the low-flow times as per the criteria presented in Table 3. In the high-flow times, the forecasts were mainly in the "*good"* range for Marsyangdi while in the "*very good"* category for Trishuli.

However, the performance of the bias-corrected forecasts in the first-order stream sites of Balephi and Tumlingtar for high-flow events is "*unsatisfactory"*. The forecast quality at these sites improved with longer horizons for the low-flow times. The NSE values were in the "*very good"* range starting on day-11 for Balephi and day-12 for Tumlingtar.

The high-resolution member did not show as good forecast performance as the ensemble mean/median post bias correction. The performance in high-flow times is often unsatisfactory, with NSE values in the negative range. There was one instance (day-6 forecasts for Balephi) in the low-flow times where the forecast performance was in the "*satisfactory"* range.

The CRPS results for the bias-corrected forecasts are presented in Figure 29. The bias correction has significantly improved the overall inaccuracy of the forecasts. The high inaccuracy seen in the raw forecasts at different intervals has been corrected to yield a horizontal line. Compare that to the deterministic skill metric plots, such as for the ensemble median during high-flow events for Marsyangdi (Figure 28) or the ensemble mean during the low-flow events for Balephi (Figure 27). The deterministic skill plots had significant fluctuations from one forecast horizon to another. The horizontal nature of the CRPS plot shows that the overall accuracy of the ensemble is stable across different horizons, although individual members might perform better/worse with varying horizons.

The bias corrector makes significant improvements on the bias. As the bias corrector is solely based on bias over the past 2-3 days, the correction for the correlation and flow variability is purely

based on improving the forecast accuracy. Hence, there is minimal improvement in other quality aspects, such as the association (as shown by r). Furthermore, the DMB bias corrector may degrade another skill metric while improving upon the bias. This is seen in the ensemble-mean plots for Trishuli in Figure 28, where the flow variability after bias correction is worse than for the raw forecasts.

One of the issues noted with the raw forecasts in section 4.1.3 was that frequent fluctuations are seen during the dry season at some sites. Figure 31 (left) shows the results before and after bias correction for Tumlingtar, where this behavior was seen. The bias correction dampens the magnitude of the fluctuations and shifts the overall hydrograph to align better with the observations. Another issue was with faster hydrograph rise in the early summer for higher-order streams. Figure 31 (right) shows the raw and the bias-corrected forecasts for Trishuli. Again, the bias corrector is able to dampen the magnitude and align the forecasts better with the observations.

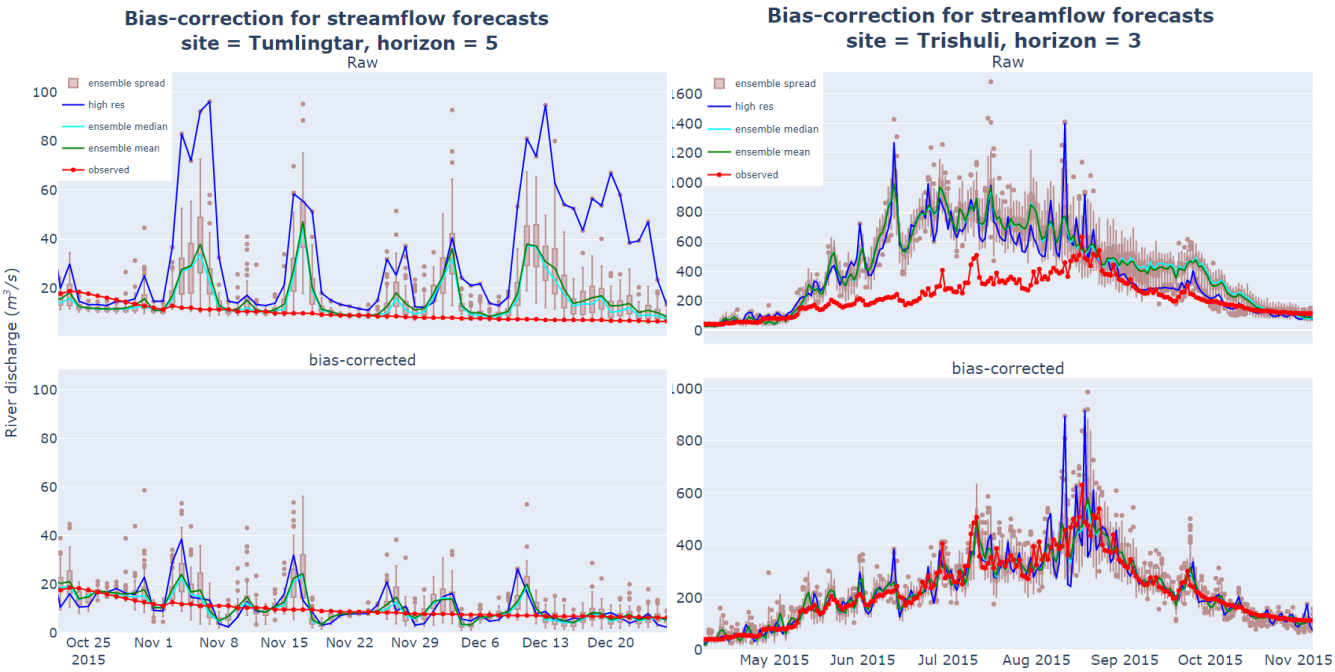


Figure 31: Improvements to the issues noted in section 4.1.3 after bias correction: high fluctuations in the dry season (left) and the faster hydrograph rise (right).

The results show that the bias correction can counter some of the flaws in the raw forecasts. However, this means that statistical correctors significantly contribute to forecast skill rather than the accurate representation of the physical processes in the modelling chain. In the following two sections, the thesis will explore how the **bias-corrected forecasts** would have performed for the two different scenarios faced by ROR hydropower systems in Nepal, as explained in section 3.6.

4.3 Wet season – flood forecasting

Table 7 highlights the number of flood events and the flood threshold value (95th percentile of climatological flow) for all the sites. Multiple days with flows higher than the flood threshold are considered a single event. Tumlingtar and Balephi had the most flood events in that year. The climatological time series for Trishuli and Marsyangdi showed that the flow in that basin was lower than in previous years. Hence, the flooding events were also fewer.

Table 7: Flood threshold and the number of flood events identified for the verification year.

Site	Flood threshold (m^3s^{-1})	Flood events
Tumlingtar	108.00	9
Balephi	252.75	21
Trishuli	533	2
Marsyangdi	786.5	1

First, the results of an exploratory analysis are presented using forecasted time series of 3 multi-day flooding events in Tumlingtar and Balephi. Then, a numerical analysis is presented in section 4.3.3.

4.3.1 Tumlingtar, June 10 – 16

This case study had major peak-flow events on June 11 and 12, followed by a smaller flow event on June 15 (see Figure 32 top row).

The raw ensemble forecast predicts an increase in flow around this time, as seen by a bump in the hydrograph. However, the ensemble predicts only a single high-flow event and does not predict the second event of a lower magnitude on June 15. The magnitude of the ensemble forecast is lower, with only the high-resolution member being similar. The high-resolution member was known a priori for high fluctuations in the dry season.

The DMB bias corrector increases the magnitude of the flow values. The ensemble detects the observed flow as it lies within the box and whisker plot. Note the difference in scale for the bias-corrected forecasts with a longer forecast horizon. The individual member bias correction approach used for the site has increased the uncertainty in the peak flow values at longer lead times. The whiskers for the day-7 forecast (Figure 32) span the values between 63 m^3s^{-1} and 1675 m^3s^{-1}. The huge forecast spread raises the question of the sharpness of the forecasts for longer-range forecast horizons and makes it challenging to use for decision-making.

The peak flows align on day-2 forecasts. However, they are late by one day for days 1, 3, and 4. The delay in the event detection increases at longer horizons, with the early medium-range forecasts showing two days delay that extends up to three days delay starting for the day-8 forecast horizon (not shown). The raw ensemble missed the actual flood peak, except on day-2 forecasts.

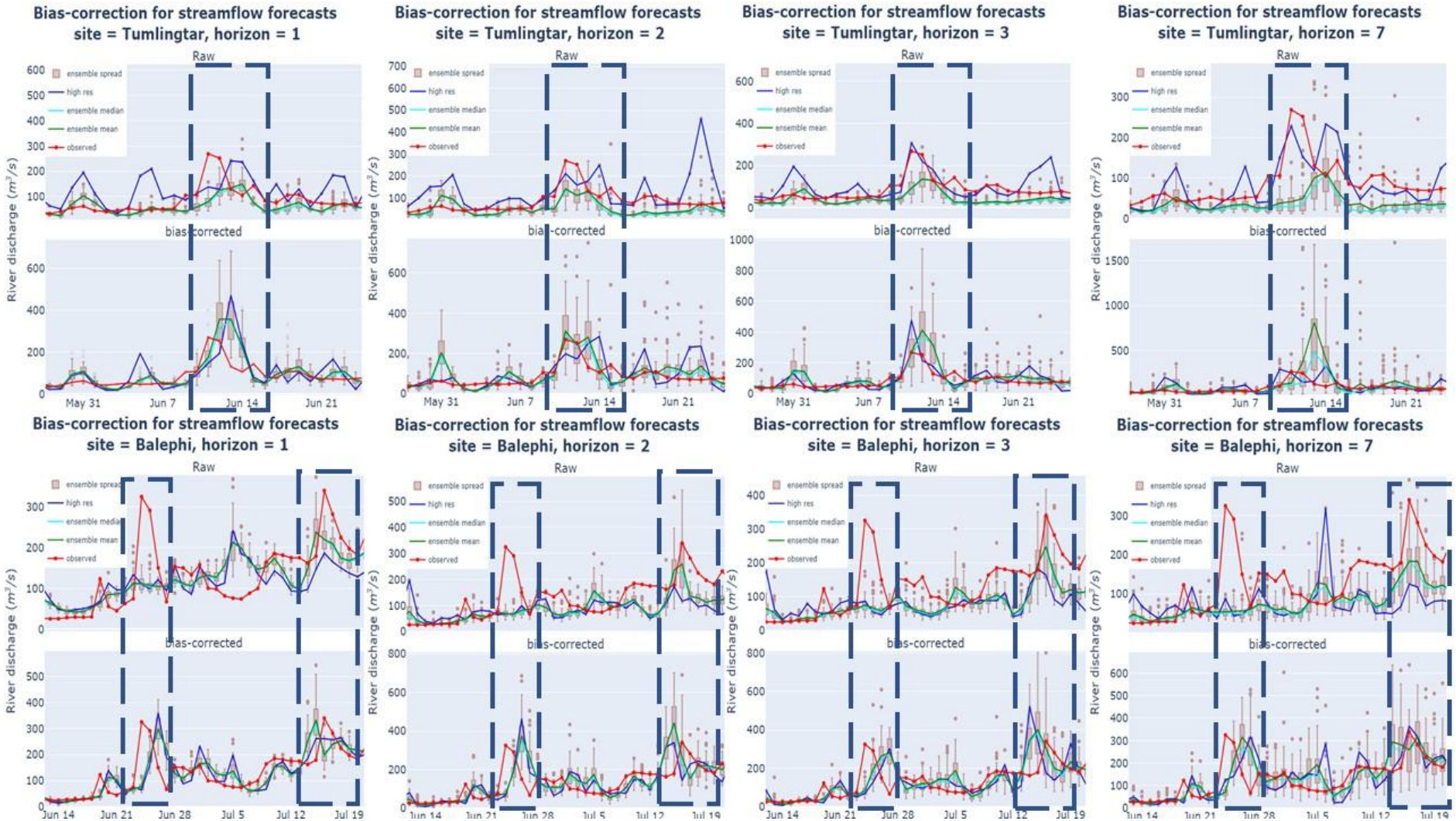


Figure 32: Forecasted and observed hydrographs for the three flooding events discussed in sections 4.3.1. and 4.3.2. Hydrographs for Tumlingtar (top) and (Balephi) bottom are shown. The dashed box highlights the flooding events investigated in the case study.

4.3.2 Balephi, June 23 – 27 and July 15 – 20

These two events occurred three weeks apart. A rapid rise of inflow is seen with peak flows on June 24 and July 16, respectively, followed by recession over the next 2-3 days for both events (see Figure 32 bottom row).

Figure 32 shows that the raw forecast has missed the June 23 event, whereas the forecast has predicted the July 16 event. However, the forecasted magnitude of flow is low. More than 50% of the ensemble members predict flow smaller than the flood threshold. Hence, an end-user would not have acted on it. Bias correction has improved the forecasted flood magnitude to cross the threshold, suggesting that an end-user could have taken the necessary action.

The day-1 and the day-2 forecasts predict the July peak-flow event to happen one day too early. As we move to a longer forecast horizon, the timing aligns better with the observed event. As expected, the bias correction has not corrected for the timing errors. However, the exceptionally high increase in ensemble spread seen for Tumlingtar in section 4.3.1 is not seen for these events.

4.3.3 Other flooding events

The results shown in 4.3.1 and 4.3.2 are for three specific case studies only. Next, the overall performance of the forecasts in giving flood warnings is identified by using contingency matrices and calculating the hit rates and false alarm ratio. A contingency matrix is created for each site and forecast horizon.

Table 8: An example of a contingency matrix. Letters h, m, f, n refer to hit, miss, false alarm, and no event.

		Forecasted flood	
		Yes	No
Observed flood	Yes	7 (*h*)	14 (*m*)
	No	3 (*f*)	341 (*n*)

An example contingency matrix for the day-1 forecast for Balephi is shown in Table 8. *Hits* (*h*) are the number of events correctly predicted by the forecasts. *Misses* (*m*) are the events that the forecasts missed, such as the June 24 peak flow discussed in section 4.3.2. *False alarms* (*f*) are the events that the forecasts predicted but were not observed. The remaining events that were not classified as floods by the forecasts and the observations make up the *no event* (*n*) cell.

$$H = \frac{h}{h + m} \tag{18}$$

$$FAR = \frac{f}{h + f} \tag{19}$$

The **hit rate** (H) and **false alarm ratios** (FAR) are then calculated using the values from the contingency tables and equations (18) and (19). The hit rate is the proportion of observed flood

events correctly forecasted, and a perfect forecast would have $H = 1$. The forecasting system could get all the observed events correct and raise warnings (i.e., false alarms) for events that did not occur. The false alarm ratio gives the frequency of the false alarms raised for forecasted flood events. A perfect forecast would have $FAR = 0$.

Table 9: Hit rates and False Alarm Ratio for verification period.

Site	Hit Rate (H)		False Alarm Ratio (FAR)	
	Range	Short-term average	Range	Short-term average
Tumlingtar	$0.33 \rightarrow 0.44$	0.44	$0.67 \rightarrow 0.81$	0.81
Balephi	$0.28 \rightarrow 0.52$	0.31	$0.18 \rightarrow 0.57$	0.46
Trishuli	0.00	0.00	1.00	1.00
Marsyangdi	0.00	0.00	1.00	1.00

Table 9 shows the results of the hit rates and the false alarm ratios. A forecast is considered to have forecasted floods if at least 50 % of the ensemble members predict flow higher than the flood threshold. The short-term averages of H and FAR are for forecast horizons of 1 – 3 days. The "Range" columns in the table show the minimum and the maximum values of H and FAR achieved for the specific sites across the forecast horizons.

H provides information on the accuracy of the flood warnings, while *FAR* provides an idea of the reliability of the forecast warnings. Table 9 shows that the GESS forecasts have low accuracy in forecasting floods. It was unable to predict more than half of the flood events. Similarly, it had a high proportion of false alarm ratios. Hence, it cannot be solely relied on for flood warning purposes.

Trishuli and Marsyangdi had only 2 and 1 flood events, respectively, and they were both missed by the forecasts. Hence, both these sites have hit rates of 0 and false alarm ratios of 1. In the case of Marsyangdi, there were only three instances (across all the horizons) where the bias-corrected forecasts issued a flood warning. Hence, the sample size for the contingency table analysis is small to get accurate results.

The higher false alarm ratio is a concern for flood warning purposes. If the flood warning is issued for a community near the ROR hydro systems, they would have to be displaced more frequently due to the false alarms. This might make the community individuals question the reliability of the forecasts issued. The following section discusses the results of the energy yield forecasting using the GESS for the dry season.

4.4 Dry season – energy yield

Figure 33 shows the revenue (in million Nepali Rupees) from the three different information sources. The "*forecast (lowest)"* and the "*forecast (highest)"* columns in the table are the minimum and the maximum possible revenues that would have been accrued if the decision was solely based on the user selecting one of the 52 GESS members.

The ensemble median was the best performing deterministic forecast. Furthermore, it represents the 50^{th} percentile of the ensemble forecasts, allowing the user to minimize the risk if the observation lies in either extreme of the ensemble distribution. Hence, the ensemble median is considered as the "go-to" information source used by the hydro operator to make future bids.

The results show that the persistence forecasts perform better than the ensemble median. However, if the user selected the best ensemble member for every decision, it would yield better revenue than the persistence forecast. On the flip side, reliance on the GESS-based energy-yield forecasts could also lead to worse outcomes than the persistence forecasts. For Tumlingtar, the revenue could be up to 28% worse than a persistence forecast, with a chance of being only 3% better if the best GESS forecast values were chosen.

The performance of the ensemble-median energy yield compared to the persistence forecast was better in the higher-order streams. This agrees with the results from section 4.2.2, where the deterministic metrics for these sites fared much better than the first-order streams during the low-flow period. However, even the highest possible revenue by using these forecasts is not a significant improvement over the revenue by using persistence forecasts, with there being no relative improvement for Trishuli and a meager 1% improvement for Marsyangdi.

Table 10 shows the number of instances where the end-user would be fined or paid only half the PPA price while using a particular information source. All the sites show that the use of ensemble median led to more instances of fines and half-price payments. The ensemble forecast has similar instances as the persistence forecasts of maximum revenue bids for Marsyangdi only. Even then, there are more instances of fines than for the no-future-information values.

Further investigation showed that the observations often did not fall within the uncertainty bounds of the ensemble forecasts. The fluctuating nature of the forecast hydrograph in the dry season is one of the reasons behind this inaccuracy. Figure 34 shows an example for day-1 forecasts in Trishuli. The overall bias across the low-flow season was 1.004, which is an almost perfect score in terms of bias. Due to the fluctuating nature, the combination of multiple under/over forecasting intervals can balance out to give perfect bias, which is what happened here.

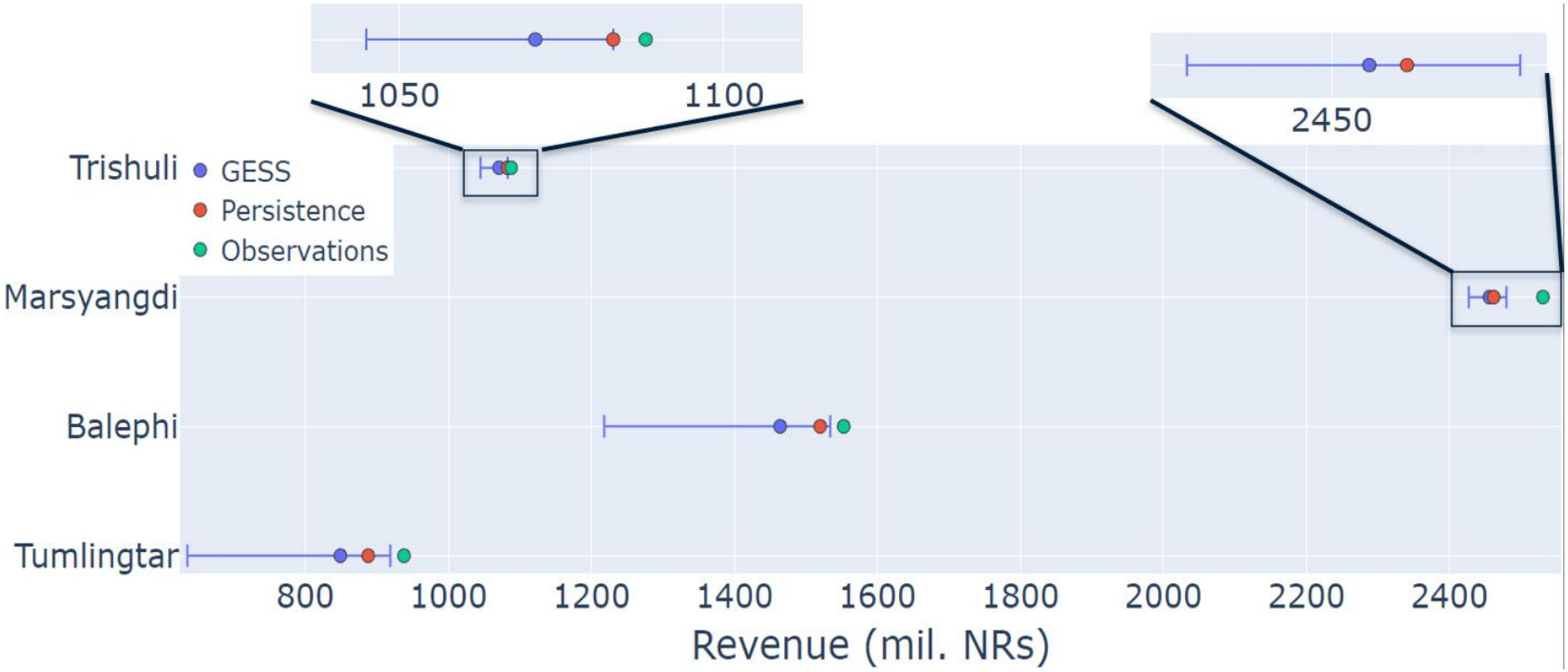


Figure 33: Revenue across the low flow season from different sources of information for making day-ahead electricity generation bids.

Table 10: Instances of fines and halved PPA rates if the bids are made using the ensemble median versus the persistence forecasts (no information).

Site	No information			GESS ensemble median		
	Fines	Halved rate	Max. revenue	Fines	Halved rate	Max. revenue
Tumlingtar	15 (6.15%)	45 (18.44%)	75.41%	38 (15.57%)	128 (52.46%)	31.97%
Balephi	1 (0.39%)	103 (40.55%)	59.06%	30 (11.81%)	126 (49.61%)	38.58%
Marsyangdi	17 (6.59%)	90 (34.88%)	58.53%	22 (8.53%)	84 (32.56%)	58.91%
Trishuli	0 (0.00%)	51 (19.69%)	80.31%	11 (4.25%)	65 (25.10%)	70.66%

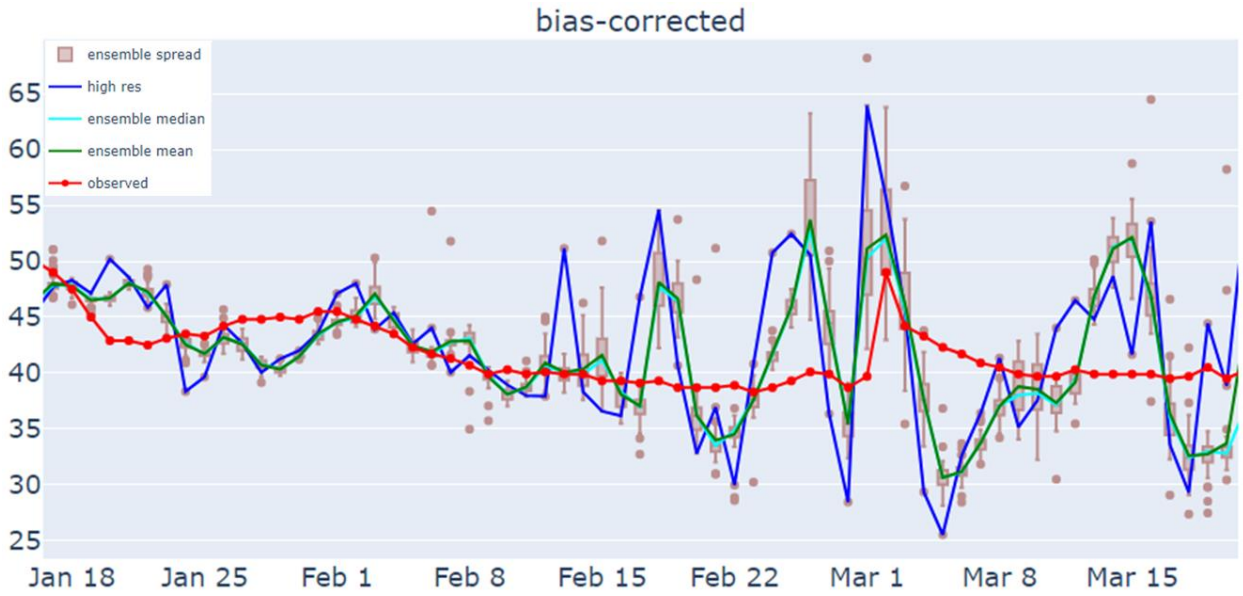


Figure 34: A zoomed-in view of the day-1 bias-corrected forecasts' hydrograph for Trishuli focusing on a period of the low-flow season. The forecasts show many fluctuating events while the ensemble spread is lower in the relatively calmer region to the left.

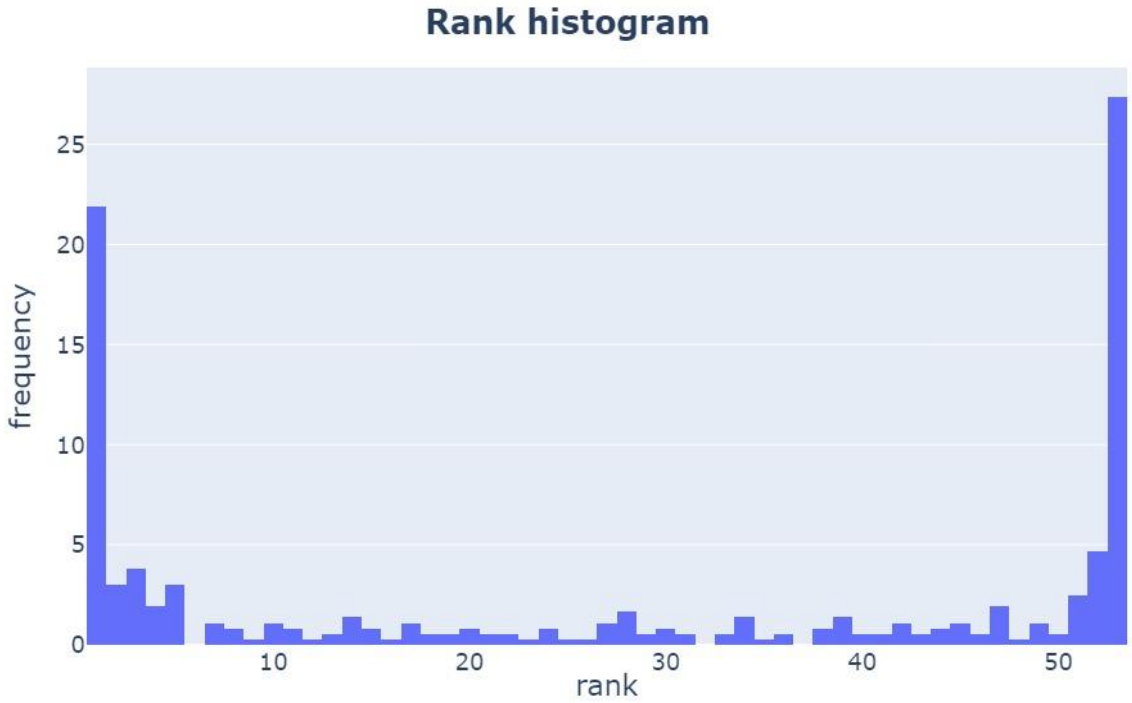


Figure 35: Rank histogram for Trishuli. The U-shaped hydrograph shows the under-dispersive nature of the ensemble forecasts.

The second reason behind the observations not falling within the ensemble spread is due to the poor dispersiveness of the ensemble forecasts. This can be seen from a rank histogram. The rank histogram tests how the observed outcomes rank with respect to the corresponding ensemble

members (Weigel, 2011, p. 146). Figure 35 shows the rank histogram for day-1 forecasts for Trishuli. Other sites showed a similar U-shaped histogram, which indicates that the observations are often greater or lower than the ensemble members without systematically inclining towards one end. So, the GESS ensemble is not dispersive enough to account for all the possible events.

The GESS ensemble is created by accounting for uncertainty in the initial conditions of the atmospheric conditions. They do not account for uncertainty in the meteorological processes post initialization, the current hydrological state, and the hydrological processes. Accounting for all these uncertainties requires a multi-model ensemble approach that combines variations in:

a. meteorological initial conditions,
b. grid size in the NWP model,
c. meteorological model processes,
d. hydrological initial conditions,
e. hydrological model processes, and
f. physics parameterizations in the hydrometeorological model (LSM in this case).

GESS fulfills only factors (a) and (b), which might explain its inability to forecast the possible scenarios and hence have wider dispersion.

Chapter 5: Conclusions

5.1 Summary

The main goal of the thesis was to evaluate the quality of the GESS streamflow forecasts and their potential value to the run-of-river hydropower operators in Nepal. The GESS streamflow forecasts are generated by routing gridded runoff outputs from the HTESSEL LSM used in the operational ECMWF NWP system. I compared the quality of the GESS forecasts against the observed river discharge values at five different sites with existing ROR infrastructure. Further, I tested a simple bias-correction method and evaluated the results faced by the ROR operators if they relied entirely on the bias-corrected streamflow forecasts for energy yield and flood early warning.

My results suggest that the raw GESS forecasts show poor performance, in some cases worse than relying on climatology. A tendency to over-forecast higher flow events was seen, which can be attributed to the runoff outputs from the LSM. Multiple flow peaks were seen in the dry season when the flow hydrograph should be smoother. Hence, it was shown that the HTESSEL LSM could not adequately capture the hydrological processes in the Himalayan-Mahabharat regions of Nepal.

I tested multiple bias correction configurations and found that the shorter window un-weighted bias correction performed the best for the GESS forecasts. The improvement through bias correction was more profound in the dry season than the wet season, and the forecasts at the higher-order streams performed much better than the forecasts at the first-order streams. For the low-flow season, the bias correction had near-perfect scores. However, the overall skill of the forecasts was primarily through the statistical post-processing rather than GESS's ability to accurately represent the underlying physical processes behind streamflow.

Finally, from the value assessment tests, I conclude that the forecasts cannot be solely relied on for predicting floods or energy-yield forecasting. The GESS forecasts had low hit rates, with an inability to predict even half of the flooding events in the sites studied. A large proportion of the predicted flooding events were false alarms, and timing errors of a few days were common, further questioning their reliability. Despite the high verification scores, the GESS forecasts did not result in higher revenue than one would get using a persistence forecast.

Addressing the first part of my first research question: "What is the quality of the GESS streamflow forecasts in areas of interest for hydropower operation in Nepal?"; my results suggest that the forecasts are of poor quality, with the bias correction being able to improve the forecast quality in terms of statistical metric considerably. The final bias-corrected forecasts were in the "satisfactory to good" performance range.

Addressing the second research question: “How useful are the GESS streamflow forecasts for different operational scenarios faced by a run-of-river hydropower operator?”; the forecasts do not add value to flood or energy-yield forecasts.

Finally, for the second half of the first research question: “Can these GESS streamflow forecasts be used at all in these areas?”; the answer from the results is “no” regarding ROR hydropower operation in the short-to-medium range. The bias-corrected GESS forecasts had good performance metrics for low-flow situations and reasonable performance for the extended forecast horizons, where the forecast skills are expected to be worse. So, there might be other applications where they could still add value.

5.2 Limitations

The results of the thesis are based on two years of forecast-observations data. While this is sufficient to investigate the general quality of the forecasting system, it is not adequate to investigate rare events. This was seen in the flood forecasting analysis for the two sites that had one or two flood events to verify against. This is a tiny sample size to make statistical conclusions.

Second, the sites in this study were chosen such that there would not be a strong spatial correlation between the forecast variables between them. This was necessary to ensure that a univariate post-processing method such as the simple DMB bias correction could be used. If the bias correction is to be applied to the entire streamflow domain, it is vital to ensure that the continuity of the flow downstream is conserved, which would further complicate the bias correction process.

Third, the GESS forecasts were developed to provide forecast information in countries like Nepal that do not have an operational ensemble forecasting system. While GESS is a welcome initiative, the 18 km resolution is inadequate to represent most of Nepal's steep mountainous topography correctly.

Next, there are flaws in the GESS forecast creation chain, such as using previous forecasts to initialize the future forecast runs. Hence, the existing forecast errors are carried into future forecast runs. With the lack of real-time streamflow measurement systems in Nepal, a reliable data assimilation system cannot be created to remedy this issue.

Finally, the results from the study further raise the question: How good is an LSM designed to provide lower-boundary conditions for operational NWP models for hydrological forecasting? The hydrological forecasting capability of LSMs needs to be continuously improved without degrading the computation speeds required to produce timely forecasts of atmospheric variables such as temperature or precipitation. Nevertheless, is there a maximum limit on the quality that one can gain from such configuration, beyond which it is better to use a specialized hydrological model?

5.3 Future Work

I identified three major issues with the forecasts from the GESS reforecast dataset, as explained in section 4.1.3.

For the first issue of the excessive streamflow on day-11 forecasts, I concluded that it is an error in the reforecast dataset creation chain. It is most likely a computation error because there is a change in the forecast configuration going from day-10 to day-11 forecasts due to the absence of the high-resolution member in the extended horizon. Further work would look into whether this issue persists in the operational forecasts by comparing hydrographs for forecast horizons of 10 through 12, as done in Figure 17.

For the second (excessive spring/early summer volume) and the third issue (high fluctuations in the dry season), I concluded that these issues result from the runoff outputs from the HTESSEL land surface model. I concluded based on the results of the runoff outputs, but I did not have any further input data to assess which process representation within the land surface model contributed to that error. The main reason behind the erroneous runoff forecasts could be the poor representation of any of the following processes:

1. The volume of precipitation in the catchment or upstream
2. Type of precipitation (rain or snow) in the catchment or upstream
3. Surface temperature forecasts that are fed into the LSM, which affect the snowmelt rate
4. Soil type and variation within the LSM grid cell
5. Soil moisture retention capacity
6. Evapotranspiration through vegetation
7. Snowmelt scheme
8. The distinction between runoff generation from snowmelt vs. glaciers

ECMWF produces precipitation, temperature, and soil moisture forecasts, all of which influence the runoff generated. Verification studies can be performed on individual forecast datasets as done for streamflow in this study. Precipitation and surface temperature stations are well-distributed across Nepal. So, the performance of forecasts for these variables can be well tested. Further, global precipitation datasets such as IMERG allow for a gridded comparison of observed and forecasted precipitation values. These verification steps allow us to isolate the influence of processes 1 – 3 from the list above.

There are several variables related to snow accumulation, such as snowfall rate and snow depth, for which ECMWF has separate forecast datasets available. Most precipitation sensors in high elevation regions can detect precipitation in terms of snowfall, but snow depth stations are not as widespread. An alternative is to use remote sensing and reanalysis-based datasets (e.g., Hu et al., 2022), but these are at a coarser resolution of ~25 km. The quality of the remote

sensing/reanalysis-based snow products could be compared with the surface observation data to check if they are good enough to verify the forecasts. If the snowfall rate and the snow depth forecasts are accurate, the error in runoff generation could be due to the thermal processes such as the temperature of the snowpack and heat fluxes between the snowpack and the surface/atmosphere.

The aforementioned steps help identify individual forecast variables' quality that affects the forecasted runoff volume. A further step would be to perform a sensitivity analysis where each of these variables is changed while keeping the others constant. This allows us to check if there are specific variables whose accuracy is most vital to the runoff formulation process.

The HTESSEL LSM does not account for glacial melt processes separately from the snowmelt processes, which could be a significant source of error because the runoff generation through glaciers is more gradual than through the snowpack melt. The performance of the GESS forecasts can be checked in other glaciated regions around the world to see if it is just a localized issue.

In section 5.2, I alluded that the 18 km resolution of the ECMWF ensemble forecasts might not be adequate for the mountainous terrain of Nepal. ECMWF is increasing the resolution of all its medium-range forecast members to 9 km resolution (ECMWF, 2022b). It would be interesting to see if this change improves the runoff estimates as the terrain complexity is represented better with the finer resolution.

This study needs to be expanded in terms of longer data years and greater location diversity for a better statistical analysis. The first step would be to expand the observation network for streamflow within Nepal so that the operational archive of the forecasts can be used instead of relying on the coarser reforecast dataset. This also allows us to check some surprising results, such as increasing skill with increasing forecast horizon, which is atypical of NWP-driven streamflow forecasts such as GESS. Further, verifying the operational archive should help confirm whether the issues presented in this thesis are glitches from the reforecast dataset or actual flaws in the modelling chain.

The thesis also focused entirely on the quality check of a process-based model. Nevertheless, how would an empirical catchment-specific hydrological model perform compared to the GESS in Nepal? It is worth noting that most catchments in Nepal are data-scarce. So, while the calibration of the model might not cover the diverse scenarios, it might still perform better than the global models.

Another option to explore would be using LSMs from other models and different spatial resolutions. The Noah MP, one of the LSMs used in the Weather Research and Forecasting (WRF) model, is already used for hydrological forecasting applications. The Weather Forecast Research

Team at the University of British Columbia has an operational suite of ensemble forecasts that go down to 1 km resolution and use the WRF model. It is worth exploring the performance of the gridded runoff output from these models in similar mountainous terrain such as British Columbia.

One interesting result from the thesis was that despite high values in terms of the verification metrics, the forecasts did not add value in the low-flow conditions. It raises the question of forecast quality vs. forecast value. It is known that the increment in forecast quality does not necessarily provide the same increment in forecast value.

However, *forecast value* is a more subjective variable than *forecast quality*. Depending on the user, a magnitude improvement in forecast quality can provide different degrees of improvement in forecast value. It would be interesting to investigate what forecast quality metric improves the forecast value the most. This could be done by creating synthetic forecasts and using them in sensitivity studies that are flawed in certain quality aspects[2]. The relative value gained by improving these quality aspects could provide information on where the post-processing should focus to maximize value for a specific end-user.

Finally, another avenue to research would be to explore multi-variate postprocessing and verification methods that can be implemented on the entire streamflow dataset. A river network of forecasts can be considered similar to forecasts of spatial fields. Hence, the post-processing techniques such as convolutional neural networks used in the latter scenario could be used for multi-variate bias correction purposes.

[2] See (Bradley, Demargne, & Franz, 2019, p. 864) for an example

Bibliography

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci*, *17*, 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Anctil, F., & Ramos, M.-H. (2019). Verification Metrics for Hydrological Ensemble Forecasts. In Q. Duan, F. Pappenberger, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 893–922). https://doi.org/10.1007/978-3-642-39925-1_3

Australian Government Bureau of Meteorology. (2022). 7-day Ensemble Streamflow Forecasts.

Balsamo, G., Viterbo, P., Beijaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., & Scipal, K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008JHM1068.1

Blyth, E. M., Arora, V. K., Clark, D. B., Dadson, S. J., De Kauwe, M. G., Lawrence, D. M., … Yuan, H. (2021). Advances in Land Surface Modelling. *Current Climate Change Reports 2021 7:2*, *7*(2), 45–71. https://doi.org/10.1007/S40641-021-00171-5

Boucher, M.-A., & Ramos, M.-H. (2018). Ensemble Streamflow Forecasts for Hydropower Systems. *Handbook of Hydrometeorological Ensemble Forecasting*, 1–19. https://doi.org/10.1007/978-3-642-40457-3_54-1

Boucher, M.-A., Roulin, E., & Fortin, V. (2019). Short-Range Ensemble Forecast Post-processing. In Q. Duan, F. Pappenberger, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 795–818). https://doi.org/10.1007/978-3-642-39925-1_71

Boucher, M. A., Tremblay, D., Delorme, L., Perreault, L., & Anctil, F. (2012). Hydro-economic assessment of hydrological forecasting systems. *Journal of Hydrology*, *416–417*, 133–144. https://doi.org/10.1016/j.jhydrol.2011.11.042

Bourdin, D. R., Fleming, S. W., & Stull, R. B. (2012). Streamflow modelling: A primer on applications, approaches and challenges. *Atmosphere - Ocean*, *50*(4), 507–536. https://doi.org/10.1080/07055900.2012.734276

Bourdin, D. R., Nipen, T. N., & Stull, R. B. (2014). Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system. *Water Resources Research*, *50*(4), 3108–3130. https://doi.org/10.1002/2014WR015462

Bourdin, D. R., & Stull, R. B. (2013). Bias-corrected short-range Member-to-Member ensemble forecasts of reservoir inflow. *Journal of Hydrology*, *502*, 77–88. https://doi.org/10.1016/j.jhydrol.2013.08.028

Bradley, A. A., Demargne, J., & Franz, K. J. (2019). Attributes of Forecast Quality. In Q. Duan, F. Pappenberger, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 849–892). https://doi.org/10.1007/978-3-642-39925-1_2

Cassagnole, M., Ramos, M.-H., Zalachori, I., Thirel, G., Garçon, R., Gailhard, J., & Ouillon, T. (2020). Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs – a conceptual approach. *Hydrology and Earth System Sciences Discussions*, 1–36. https://doi.org/10.5194/hess-2020-410

Contreras, E., Herrero, J., Crochemore, L., Aguilar, C., & Polo, M. J. (2020). Seasonal climate forecast skill assessment for the management of water resources in a run of river hydropower system in the Poqueira River (Southern Spain). *Water (Switzerland)*, *12*(8).

https://doi.org/10.3390/W12082119
Copernicus EMS. (n.d.). European Flood Awareness System. Retrieved January 31, 2022, from https://www.efas.eu/en
Cunge, J. A. (1969). On The Subject Of A Flood Propagation Computation Method (Musklngum Method). *Journal of Hydraulic Research*, *7*(2), 205–230. https://doi.org/10.1080/00221686909500264
David, C. H., Maidment, D. R., Niu, G. Y., Yang, Z. L., Habets, F., & Eijkhout, V. (2011). River Network Routing on the NHDPlus Dataset. *Journal of Hydrometeorology*, *12*(5), 913–934. https://doi.org/10.1175/2011JHM1345.1
Department of Hydrology and Meteorology. (2018). *Standard Operating Procedure (SOP) for flood early warning system in Nepal*. Kathmandu.
Dingman., S. L. (2008). *Physical Hydrology* (2nd ed.). Waveland Press, Inc.
ECMWF. (2020, October 24). Energy Fluxes - Land - HTESSEL - Forecast User Guide - ECMWF Confluence Wiki. Retrieved January 14, 2022, from https://confluence.ecmwf.int/display/FUG/Energy+Fluxes+-+Land+-+HTESSEL
ECMWF. (2021). IFS Documentation CY47R3 - Part V Ensemble prediction system. In *IFS Documentation*. *IFS Documentation CY47R3*. https://doi.org/10.21957/zw5j5zdz5
ECMWF. (2022a). Quantifying forecast uncertainty. Retrieved January 23, 2022, from https://www.ecmwf.int/en/research/modelling-and-prediction/quantifying-forecast-uncertainty
ECMWF. (2022b, April 7). Global numerical modelling at the heart of ECMWF's forecasts | ECMWF. Retrieved April 7, 2022, from https://www.ecmwf.int/en/about/media-centre/focus/2022/global-numerical-modelling-heart-ecmwfs-forecasts
EMCWF. (2022). Energy Fluxes - Land - HTESSEL. Retrieved March 9, 2022, from ECMWF Confluence Wiki website: https://confluence.ecmwf.int/display/FUG/Energy+Fluxes+-+Land+-+HTESSEL
Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., … Cloke, H. L. (2016). Continental and global scale flood forecasting systems. *WIREs Water*, *3*(3), 391–418. https://doi.org/10.1002/wat2.1137
Energy BC. (2016). Run of River Power. Retrieved January 9, 2022, from http://www.energybc.ca/runofriver.html
ERC. (2020). *विद्युत खरिद बिक्री तथा अनुमति प्राप्त व्यक्तिले पालना गर्नुपर्ने शर्त सम्बन्धि विनियमावली (Vidyut Kharid Bikri tatha Anumati praapta Byaktile Paalanaa garnuparne sarta sambandhi Biniyamaawali*. Nepal.
Fan, F. M., Schwanenberg, D., Alvarado, R., Assis dos Reis, A., Collischonn, W., & Naumman, S. (2016). Performance of Deterministic and Probabilistic Hydrological Forecasts for the Short-Term Optimization of a Tropical Hydropower Reservoir. *Water Resources Management*, *30*(10), 3609–3625. https://doi.org/10.1007/s11269-016-1377-8
GeoGLOWS. (2022). About the GEOGloWS ECMWF Streamflow Service. Retrieved February 1, 2022, from https://geoglows.ecmwf.int/about
Google LLC. (2020). *Google Earth Pro*. Google LLC.
Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/J.JHYDROL.2009.08.003
Hongling, L., Chuanwen, J., & Yan, Z. (2008). A review on risk-constrained hydropower scheduling in deregulated power market. *Renewable and Sustainable Energy Reviews*, *12*(5),

1465–1475. https://doi.org/10.1016/j.rser.2007.01.018
Hu, Y., Che, T., Dai, L., Zhu, Y., Xiao, L., Deng, J., & Li, X. (2022). A long-term daily gridded snow depth dataset for the Northern Hemisphere from 1980 to 2019 based on machine learning. *Earth System Science Data Discussions*. https://doi.org/https://doi.org/10.5194/essd-2022-63
IEA. (2021). *Hydropower Special Market Report*. Retrieved from www.iea.org
IEX. (2022). Area Prices. Retrieved January 3, 2022, from https://www.iexindia.com/marketdata/areaprice.aspx
IRENA. (2012). *RENEWABLE ENERGY TECHNOLOGIES: COST ANALYSIS SERIES*. Retrieved from www.irena.org/Publications
Jain, S. K., & Singh, V. P. (2019). Hydrological Cycles, Models, and Applications to Forecasting. In Q. Duan, F. Pappenberger, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 311–339). https://doi.org/10.1007/978-3-642-39925-1_20
Kleiven, A., & Steinsland, I. (2019). Inflow forecasting for hydropower operations: Bayesian model averaging for postprocessing hydrological ensembles. *Proceedings of the 6th International Workshop on Hydro Scheduling in Competitive Electricity Markets*, 33–40. https://doi.org/10.1007/978-3-030-03311-8_5
Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/HESS-23-4323-2019
Krause, P.; Boyle, D. P.; Baese, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *53*(5), 89–97. https://doi.org/10.1016/0021-9169(91)90041-5
Liu, Z., Wang, Y., Xu, Z., & Duan, Q. (2017). Conceptual Hydrological Models. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1–23). https://doi.org/10.1007/978-3-642-40457-3_22-1
McCollor, D., & Stull, R. (2008a). Hydrometeorological accuracy enhancement via postprocessing of numerical weather forecasts in complex terrain. *Weather and Forecasting*, *23*(1), 131–144. https://doi.org/10.1175/2007WAF2006107.1
McCollor, D., & Stull, R. (2008b). Hydrometeorological short-range ensemble forecasts in complex terrain. Part I: Meteorological evaluation. *Weather and Forecasting*, *23*(4), 533–556. https://doi.org/10.1175/2008WAF2007063.1
McCollor, D., & Stull, R. (2008c). Hydrometeorological short-range ensemble forecasts in complex terrain. Part II: Economic evaluation. *Weather and Forecasting*, *23*(4), 557–574. https://doi.org/10.1175/2007WAF2007064.1
Moriasi, D. N., Gitau, M. W., Pai, N., Daggupati, P., Gitau, M. W., Member, A., & Moriasi, D. N. (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Transactions of the ASABE*, *58*(6), 1763–1785. https://doi.org/10.13031/trans.58.10715
Murphy, A. H. (1997). Forecast Verification. In R. W. Katz & A. H. Murphy (Eds.), *Economic Value of Weather And Climate Forecasts* (pp. 19–74). https://doi.org/10.1017/cbo9780511608278
NEA. (2021). *A Year in Review Fiscal Year 2020/2021*.
Nelson, E. J. (2021, June 9). Predicting Disasters Using Global Water Intelligence. *Scientia*.

https://doi.org/10.33548/SCIENTIA682
Niti Foundation. (2018). Nepal Hydropower Portal. Retrieved February 12, 2022, from https://hydro.naxa.com.np/
NOAA. (n.d.). The National Water Model website. Retrieved January 14, 2022, from https://water.noaa.gov/about/nwm
Potts, J. M. (2012). Basic Concepts. In Ian T. Jolliffe; David B. Stephenson (Ed.), *Forecast Verification* (Second, pp. 11–29). https://doi.org/10.1002/9781119960003.ch2
Rose, A., Duwadi, K., Palchak, D., & Joshi, M. (2021). *Policy and Regulatory Environment for Utility-Scale Energy Storage: Nepal*. Retrieved from https://www.nrel.gov/docs/fy21osti/80591.pdf.
Roundy, J. K., Duan, Q., & Schaake, J. C. (2019). Hydrological Predictability, Scales, and Uncertainty Issues. In H. L. . Duan, Q.; Pappenberger, F.; Wood, A.; Cloke & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 3–31). https://doi.org/10.1007/978-3-642-39925-1
Sene, K. (2016). *Hydrometeorology*. Cham: Springer.
Souffront Alcantara, M. A., Nelson, E. J., Shakya, K., Edwards, C., Roberts, W., Krewson, C., … Gutierrez, A. (2019). Hydrologic Modeling as a Service (HMaaS): A New Approach to Address Hydroinformatic Challenges in Developing Countries. *Frontiers in Environmental Science*, *7*. https://doi.org/10.3389/fenvs.2019.00158
Stokelj, T., Paravan, D., & Golob, R. (2002). Enhanced Artificial Neural Network Inflow Forecasting Algorithm for Run-of-River Hydropower Plants. *Journal of Water Resources Planning and Management*, *128*(6), 415–423. https://doi.org/10.1061/(asce)0733-9496(2002)128:6(415)
Stull, R. (2017). *Practical Meteorology: An Algebra-based Survey of Atmospheric Science* (1.02b). Vancouver.
Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., … Bierkens, M. F. P. (2018). PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/GMD-11-2429-2018
Warner, T. T. (2011). Numerical weather and climate prediction. In *Numerical Weather and Climate Prediction* (First, Vol. 9780521513). https://doi.org/10.1017/CBO9780511763243
WECS. (2019). *Assessment of Hydropower Potential of Nepal: Final Report*.
Weigel, A. P. (2011). Ensemble Forecasts. In *Forecast Verification* (pp. 141–166). https://doi.org/https://doi.org/10.1002/9781119960003.ch8
Wikimedia Commons. (2006, November 27). File:Nepal topo en.jpg - Wikipedia. Retrieved March 10, 2022, from Wikipedia website: https://en.wikipedia.org/wiki/File:Nepal_topo_en.jpg
Wilks, D. S. (2019a). Chapter 9 - Forecast Verification. In D. S. Wilks (Ed.), *Statistical Methods in the Atmospheric Sciences (Fourth Edition)* (Fourth Edi, pp. 369–483). https://doi.org/https://doi.org/10.1016/B978-0-12-815823-4.00009-2
Wilks, D. S. (2019b). *Statistical Methods in the Atmospheric Sciences* (Fourth). https://doi.org/10.1016/B978-0-12-815823-4.09987-9
Wood, A. W., & Schaake, J. C. (2008). Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. *Journal of Hydrometeorology*, 132–148. https://doi.org/10.1175/2007JHM862.1
Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., & Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resources Research*, *50*(3), 2693–2717. https://doi.org/10.1002/2013WR014710

Appendix A: Optimum Site Selection using ArcGIS Pro

The analysis steps followed by the author to identify the research sites are explained in this section. **Bold** terms are the ArcGIS Pro tools used, while ***bold-italics*** are the relevant layer names. The layer names have been shortened for simplicity and do not necessarily match the exact layer names used by the author in his analysis.

1. Load all the datasets into ArcGIS Pro:
   a. The following steps create the ***dams*** layer:
      - Import the operational and the under-construction dam sites as separate layers
      - Merge the two layers
      - Apply a filter to only select projects above 5 MW
   b. The ***stations*** layer contains the location of the gauging stations, which was taken from the stations list published by DHM.
   c. The ***river network*** layer for Nepal was available from Jorge Luis Sanchez-Lozano, from BYU.
   d. The GESS forecasts were available in NetCDF format. A random NetCDF file was chosen, and the data points were converted to an ArcGIS Point layer. The conversion process is as follows:
      - Convert the NetCDF file into a table view using the **Make NetCDF Table View** The latitude and the longitude are set as variables, and river id is set as row dimensions
      - Convert the table into a point layer using **XY Table To Point**
   e. The runoff data points are also in NetCDF format. The same steps as for d) are followed. The only difference is that the row dimensions are now latitude and longitude, whereas the variables include runoff. Separate data points are loaded into the
   f. A fine resolution map of catchments was available through Jorge Luis Sanchez-Lozano, from BYU.

2. Snap the ***dams*** and the ***stations*** layers to the ***river network*** layer. This is done to ensure that the points lie precisely on the lines.

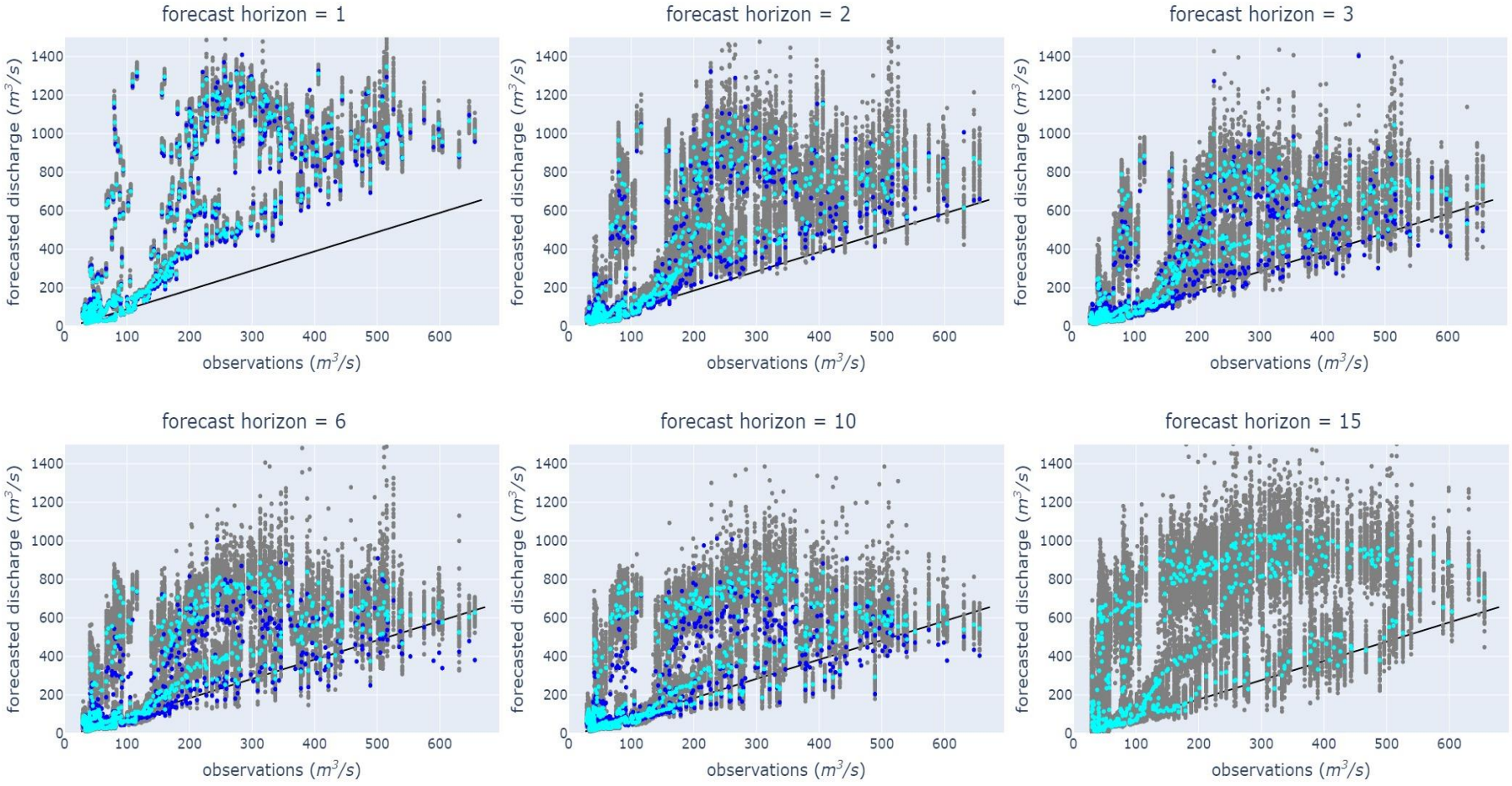
3. Merge the **dams** and the **stations** layers into ***dams-stations***:
   a. First, create a new field in both layers called “type”, which identifies whether the layer entries are “stations” or “dams”
   b. Merge the two layers using the **Merge** tool

4. Perform a **one-to-one spatial join** of the ***dams-stations*** layer to the ***river network*** layer using the match option **Intersect**. This will create a new layer ***rivers-joined***.

5. Export the relevant stream segments as ***interested segments*** layer:
    a. Use **Select Layer by Attribute** to filter for stream segments that have all of the following attributes:
        - Stations data is available for at least 2014, and this also filters all the segments that contain at least one station
        - Join count is greater than two because there should be at least two point features: ideally, one for the dams and another for the stations within one river segment
        - Data type as **dams**. This provides a filter for segments that have at least one dam
    b. Export the selected segments using **Feature Class To Feature Class.**

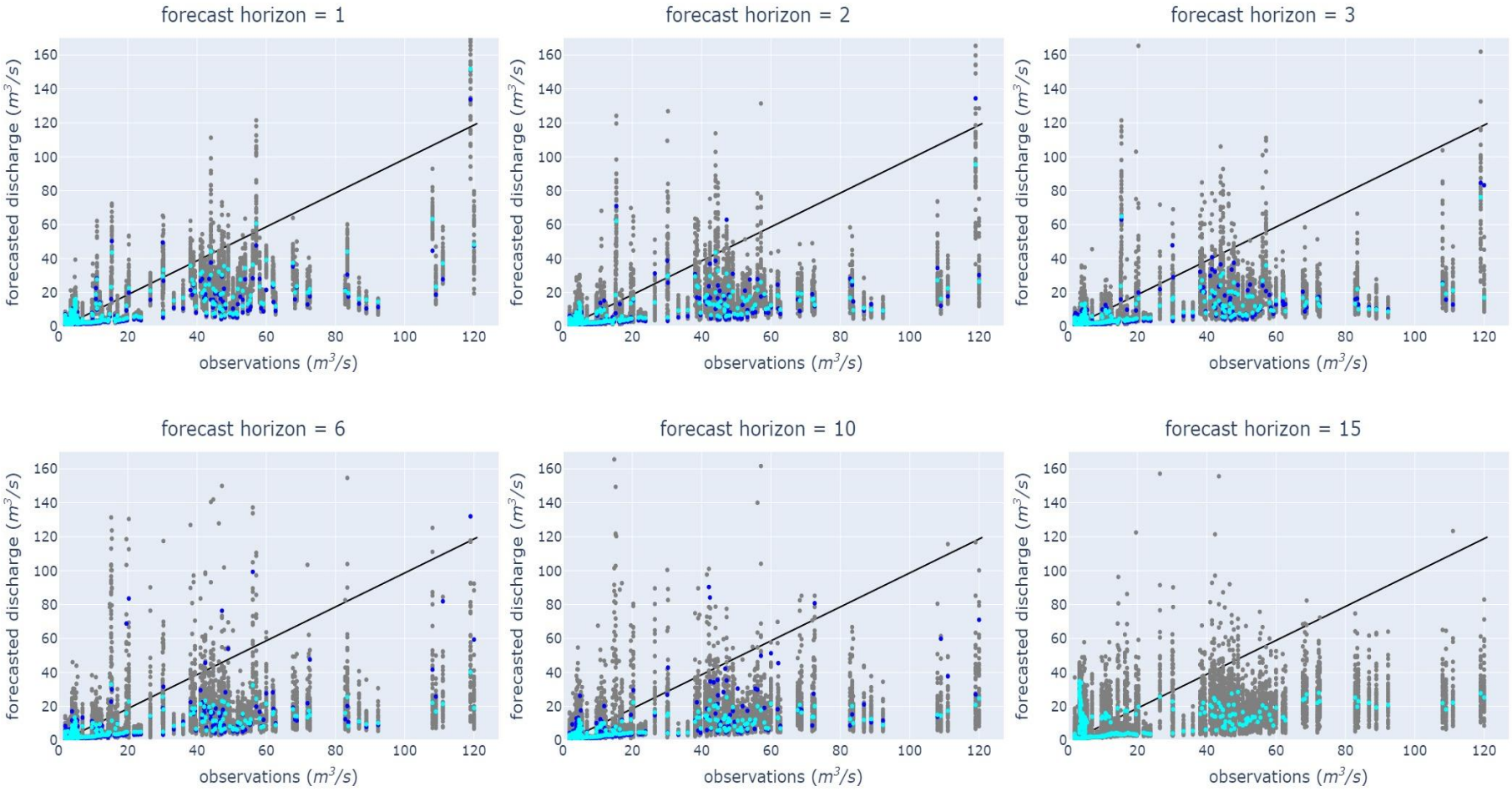
6. Display only the interested sites using the **Clip** geoprocessing tool.

Appendix B: Raw Scatter plots



Scatter plot for Trishuli, the central site in this study. Grey dots represent the ensemble forecasts, cyan is the ensemble median, and blue is the high-resolution member. The solid black line represents the y = x line. Ideally, all the data points should fall along this line.

Forecasts vs Observations | Site: Naugadh



Scatter plot for Naugadh, the far western site in this study. The site has fewer points as the observation data was only available for 2014. Grey dots represent the ensemble forecasts, cyan is the ensemble median, and blue is the high-resolution member. The solid black line represents the y = x line. Ideally, all the data points should fall along this line.

Appendix C: Calibration Results

		Tumlingtar			Balephi			Trishuli			Marsyangdi			Naugadh		
day	flow	NSE	KGE	CRPS	NSE	KGE	CRPS	NSE	KGE	CRPS	NSE	KGE	CRPS	NSE	KGE	CRPS
1	high	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_ldmb')	(2, 'Q_dmb-var')	(15, 'Q_ldmb_2')	(2, 'Q_dmb')
2		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb-var')	(15, 'Q_ldmb_2')	(2, 'Q_dmb')
3		(2, 'Q_dmb_2')	(2, 'Q_ldmb-var')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(5, 'Q_dmb_2')	(3, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb-var')	(7, 'Q_ldmb')	(2, 'Q_dmb')
4		(2, 'Q_dmb_2')	(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(5, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_ldmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
5		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')
6		(3, 'Q_dmb')	(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
7		(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
8		(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
9		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
10		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
11		(5, 'Q_dmb_2')	(3, 'Q_dmb_2')	(3, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
12		(3, 'Q_dmb')	(2, 'Q_dmb')	(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
13		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(3, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
14		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
15		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_ldmb-var')	(2, 'Q_dmb_2')
1	low	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')
2		(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')
3		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')
4		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')
5		(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')
6		(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')
7		(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')
8		(2, 'Q_ldmb')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')
9		(2, 'Q_ldmb_2')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')
10		(2, 'Q_ldmb')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')
11		(2, 'Q_dmb')	(2, 'Q_ldmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
12		(2, 'Q_ldmb')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
13		(2, 'Q_ldmb')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
14		(2, 'Q_ldmb')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')
15		(2, 'Q_ldmb')	(2, 'Q_ldmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')	(2, 'Q_dmb')	(2, 'Q_dmb')	(2, 'Q_dmb_2')

Note: Highlighted cells (yellow/red/green/blue/purple) represent the most frequently occurring best combinations. The first element of the ordered pair is window length, and the second element is the DMB method, followed by the bias correction approach used. So, (2, Q_dmb_2) means the best metric value was obtained for a window length of 2 days, using the equal weight DMB of equation (10) and the second bias correction approach where all members are applied same DMB corrector.