

Assessing the Shyft Modelling Framework in Nepal: Impact of Snow Routines and Terrain Representation on Simulated Water Balance Components

Jacob Qvam Skavang

Hydrology and Glaciology
60 ECTS study points

Department of Geosciences
Faculty of Mathematics and Natural Sciences

Spring 2023



Jacob Qvam Skavang

Assessing the Shyft Modelling
Framework in Nepal: Impact of
Snow Routines and Terrain
Representation on Simulated Water
Balance Components

Supervisors:
Olga Silantyeva
Lena Merete Tallaksen
Kristoffer Aalstad

Abstract

This study focuses on the evaluation of the Shyft modeling framework's effectiveness in simulating primary water balance components in the Budhi Gandaki catchment in Nepal. The region, originating from the Himalayan and Tibetan Plateau, holds significant importance for millions of people due to its water supply. The focus is particularly on key variables such as discharge and snow-related factors like snow-covered area, snow-cover duration, snow water equivalent, and glacier melt. To achieve this, the study utilizes bias-adjusted ERA5 reanalysis data (WFDE5) as the forcing data. Temperature is interpolated using the Bayesian Temperature Kriging method, while precipitation, relative humidity, wind speed, and shortwave radiation employ the Inverse Distance Weighting (IDW) method. Two different snow routines, namely Gamma Snow (a simplified surface energy balance model for snow melt) and Snow Tiles (a temperature-index-based model), are employed. In terms of terrain representation, a regular grid and four different Triangular Irregular Networks (TINs) with varying resolutions are used. The calibration and evaluation process involves ten different model configurations calibrated daily over a five-year period. The evaluation metrics used include Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), Root Mean Square Error (RMSE), and the natural logarithm of the Nash-Sutcliffe Efficiency (LnNSE). The simulated snow-covered area is also compared against observed data from MODIS. The findings of the study reveal pronounced seasonal variation in the forcing data. Additionally, it is observed that the Snow Tiles routine requires considerably less time for calibration compared to the Gamma Snow routine, especially when higher resolution TINs are used. In terms of simulating discharge, both snow routines perform satisfactorily for both the calibration (2000-2004) and validation (2005-2015) periods, with the Snow Tiles routine exhibiting slightly better overall performance. However, when evaluating snow-covered area against MODIS observations, both models demonstrate low performance, overestimating the snow-covered area and snow-cover duration, particularly in the northern part of the catchment. Moreover, the Snow Tiles routine exhibits an unexpected linearly increasing trend in snow water equivalent, which may be attributed to its limitations in capturing snow dynamics at higher elevations, where incoming shortwave radiation is the primary energy input and surface temperature has a weaker correlation with snow melt compared to lower elevations. To enhance future modeling accuracy, the study suggests exploring additional tools such as downscaling and in-situ precipitation data. These approaches could potentially address the limitations observed in simulating snow dynamics and improve the overall performance of the modeling framework.

Abstract

Acknowledgement

I would like to express my deep appreciation to my three supervisors who have played an essential role in completion of this thesis.

First of all, I am grateful to my supervisor Olga Silantyeva at the University of Oslo for her guidance and support throughout the research process. Her work has provided a basis upon which this thesis is built on, and her expertise, invaluable feedback and insightful suggestions has helped me refine my ideas and improved my writing. The resources she has provided me have been crucial to this thesis. She has given me the necessary tools and resources for this study, and assisted me in using Shyft. Furthermore, she has also provided the Triangular Irregular Networks (TINs) data which are used in this study. She has also helped me learn and grow throughout the process. Her dedication and commitment to open science principles has been a source of inspiration for me. Without her contributions, this achievement would not be possible.

Secondly, I would like to thank my two co-supervisors Lena Merete Tallaksen and Kristoffer Aalstad at the University of Oslo, for their help and contributions throughout this journey. I am thankful for them taking time to meet with me and provide constructive criticism. Their support and sharing of knowledge has helped me to plan and organize my thesis. Kristoffer has also provided me with the MODIS data containing fractional snow-covered area (fSCA) and snow-cover duration (fSCD). He has also helped me on how to use this data.

Thridly, I would like to thank my colleague Felix Matt at Statkraft for guiding me on the Snow Tiles routine he has developed, and the Department of Hydrology and Meteorology in Nepal for providing me the observed discharge series. I would also like to thank Bikas Chandra Bhattarai for the regular grid cell data and catchment boundary data which is used in this study.

Lastly, I would like to thank my friends and family who has been a been a great source of encouragement, love and support. They have been a motivating force during challenging moments of this journey.

Abstract

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Aim and objectives	2
1.3	Thesis outline	3
2	Background	5
2.1	The importance of mountains as water towers	5
2.2	Climate change in High Mountain Asia and its effect on hydrology	5
2.3	The hydrology of High Mountain Asia	6
2.4	Hydrological modelling in High Mountain Asia	7
2.5	Overcoming challenges in hydrological modelling in High Mountain Asia	11
3	Study area	13
3.1	Geography	15
3.2	Climate	16
3.3	Hydropower	18
4	Data and pre-processing	19
4.1	Meteorological forcing data	19
4.1.1	The bias-adjusted ERA5 reanalysis data for impact studies (WFDE5)	20
4.2	Topographical Data	23
4.3	Land Cover Data Sets	23
4.4	Catchment shapefile	23
4.5	River network data	24
4.6	Cell data and TINs	24
4.7	Country polygons	25
4.8	Glacier polygons	25
4.9	Data used for evaluation (target data)	25
4.9.1	Observed river discharge	25
4.9.2	NASA MODIS Snow Cover Products	25
4.10	Summary of forcing and validation data	26
4.11	Pre-processing of the data	27
4.11.1	Spatial data	27
4.11.2	WFDE5: Data inspection and selection	27
4.11.3	Conversion of units	28
4.11.4	Obtaining elevation for the cells	29
4.11.5	Coordinate conversion	29
4.11.6	Creating netCDF files for Shyft	29

4.12	Catchment Discretization Technique	29
4.12.1	Regular grid method	30
4.12.2	Triangular irregular network (TIN) method	32
4.12.3	Summary of the discretization techniques	34
4.13	Data management and FAIR principles	34
5	Methods	37
5.1	Hydrological Modeling Framework	37
5.1.1	Model stacks: PTGSK and PTSTK	38
5.2	Model setup	45
5.2.1	Configuration of a Shyft simulation	46
5.2.2	Defining the simulation	46
5.2.3	Defining the region domain	46
5.2.4	Datasets and set up of repositories	47
5.2.5	Defining the model stacks	48
5.2.6	Defining interpolation methods	48
5.2.7	Running the first simulation	50
5.2.8	Calibration	50
5.2.9	Summary of the simulation and calibration runs	54
5.2.10	Post-calibration run and data extraction	54
5.2.11	Model evaluation	54
6	Results	59
6.1	Spatial and temporal variation in the forcing data	59
6.1.1	Analysis of seasonality and trends in the High Mountain Asia region	59
6.1.2	Temporal and spatial variation in the Budhi Gandaki region	62
6.1.3	Analysis of Near-Surface Variables at Selected Elevations	69
6.1.4	Observed discharge at Arughat gauging station	73
6.1.5	Interpolation results from Inverse Distance Weighting and Bayesian Temperature Kriging	75
6.2	Impacts of Terrain Representation and Snow Routines on Calibration and Model Performance: Discharge	78
6.2.1	Optimized Parameters Using Min BOBYQA Method	78
6.2.2	Impact of Snow Routines and Terrain Representations on Simulated Discharge	81
6.3	Impact of Terrain Representations and Snow Routines on Model Performance: SCA, SCD, and SWE and Glacier Melt	91
6.3.1	Evaluating Simulated SCA using the Critical Success Index and Observed SCA from MODIS	91
6.3.2	Comparison between Simulated and Observed Snow-Cover Duration	95
6.3.3	Comparing Simulated SWE and Glacier Melt between Models	97
7	Discussion	99
7.1	Analysis of Spatial and Temporal Variation in Forcing Data and Evaluation of Interpolation Methods	99
7.2	Impacts of Terrain Representations and Snow Routines on Calibration and Model Performance: Discharge	102
7.3	Impacts of Snow Routines and Terrain Representations on Model Performance: Snow-Covered Area, Snow-Cover Duration, Snow Water Equivalent and Glacier Melt	104

8	Conclusions and outlook	107
A	Shyft modelling framework	109
A.1	The architecture of Shyft	109
A.1.1	C++ core	110
A.1.2	Python API	110
A.1.3	Repositories	113
A.2	Orchestration	113
A.2.1	Conceptual model	113
A.2.2	Hydrologic modelling	115
A.3	Shyft requirements	116
A.4	Calibration	117
B	Net radiation in Shyft	119
C	Gamma Snow routine	121
D	Processing of spatial data	123
E	Forcing data	125
F	Spatial statistics of regular grid cells and TINs	127
G	YAML-configuration files	131

Contents

List of Figures

2.1	Classification of hydrological models. Hydrological models can broadly be categorized as stochastic or deterministic. They can also be characterized by how the hydrological process is described as conceptual, empirical or physical. The figure is made using Canva	8
2.2	The hydrologic cycle. The figure illustrates the major components of the hydrologic cycle. The figure is made using Canva	9
3.1	View over the Budhi Gandaki valley. Image captured on the Manaslu Circuit Trek with Manaslu (8163 m.a.s.l.) to the right and Samdo to the middle (5177 m.a.s.l.). Photo taken by Hendrikz (2019)	13
3.2	Map of the Budhi Gandaki catchment. The figure showcases an overview map to the top left, a snow cover map to the bottom left, a land cover type map in the bottom center and an elevation map to the right. The overview map displays the national boundaries of Nepal, providing context for the location of the Budhi Gandaki catchment within the country. The catchment is displayed in red and cid10 is the catchment id used in this study. The fractional snow cover area (fSCA) map depicts the mean distribution and extent of snow cover during the winter season 2005-2006 (DJF). Values range from 0 (no snow cover) to 1 (completely snow covered). The snow cover data is derived from MODIS (NSIDC 2023) . The land cover map identifies different land cover types, such as forest, water bodies, urban areas, and grass and wetland, offering insights into the composition and distribution of the landscape. The data is derived from the MODIS Land Cover Type product (Sulla-Menashe and Friedl 2018) . The elevation map visualizes the varying topography and elevations within the catchment. Green color shows low elevations and brown/white shows high elevation. The river network is displayed in blue, where the thickness of the line represents the stream size (using Strahler order). Arughat gauging station is marked with a green point.	14
3.3	Landscape diversity in the Budhi Gandaki catchment. The climate in the catchment ranges from temperate to polar tundra.	15
3.4	Elevation profile of the Budhi Gandaki catchment	16
3.5	a) Climate zones in the Budhi Gandaki catchment. The figure shows the elevation plotted against cumulative area [%]. Elevation and area is based upon the values found in the TIN with the highest resolution (4681 cells). The climate zones are based upon (Khatri et al. 2018). b) Land cover type distribution in Budhi Gandaki catchment. The figure shows aggregated values based on the GlobCov 2009 land cover type data (Bontemps et al. 2011).	17

List of Figures

4.1	Selected WFDE5 grid cells for the Budhi Gandaki catchment. The midpoint of the grid cells are colored in red.	28
4.2	Examples of terrain representations: a) Triangular Irregular Network (TIN), b) regular grid cells and c) elevation zones. The red dots show the midpoints of the triangles in the TIN and the regular grid. Figures are made in Canva	30
4.3	The figure shows the mid-points in the regular grid (1077 cells). The colors represent the slope value of each cell. The coordinates are in UTM (EPSG:32645). The figure here shows a resolution of 1 km ² . In the Shyft simulations, each grid cell has a resolution of 2 × 2 km ²	31
4.4	TIN with lower resolution (911 cells) and a TIN with higher resolution (4681 cells). The colors represent the elevation at each triangle. Larger triangle area represent flatter terrain.	33
5.1	Multiple configurations in Shyft as of April 2023 (Shyft version v5.x (4.26.24 in Python). The figure is an adaptation of material provided by Postdoctoral Fellow Olga Silantyeva at the University of Oslo, and is made using Canva	39
5.2	The PTGSK model stack in Shyft as of April 2023 on keilhau.uio.no (Shyft version v5.x (4.26.24 in Python). The input data are shown to the left in the figure. The light blue color represents the routines in the PTGSK model stack. The figure is made using Canva	40
5.3	The PTSTK model stack in Shyft as of April 2023 on keilhau.uio.no (Shyft version v5.x (4.26.24 in Python). The input data are shown to the left in the figure. The light blue color represents the routines in the PTSTK model stack. The figure is made using Canva	43
5.4	Illustration of the workflow in Shyft through an iterative process of configuration, simulation, calibration, and post-calibration run and data extraction. The figure is made using Canva	45
5.5	Region domain in Shyft. The grid is 75 x 105 cells (7875 cells). Projection is EPSG:32645.	47
6.1	Mean seasonal temperature High Mountain Asia (1990-2019) from the WFDE5 data. DJF is December, January and February, MAM is March, April and May, JJA is June, July and August, while SON is September, October and November.	60
6.2	Mean seasonal accumulated precipitation in High Mountain Asia (1990-2019) from the WFDE5 data.	61
6.3	Time series decomposition (1990-2019). a) Daily mean temperature in HMA, b) daily mean precipitation, c) relative humidity and d) radiation. Original time series are shown at the top, and then trend, season and residual below respectively.	62
6.4	Mean seasonal temperature in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data.	63
6.5	Mean seasonal accumulated precipitation [mm] in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.	64

6.6	Mean seasonal accumulated snowfall [mm] in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.	65
6.7	Mean seasonal wind speed in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.	66
6.8	Mean seasonal incoming shortwave radiation in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.	67
6.9	Mean seasonal relative humidity in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data.	68
6.10	Temperature against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily mean temperature [°C] at different elevations. The black line represents a 365-day moving average. e) Shows mean annual temperature plotted against elevation.	69
6.11	Precipitation against elevation. a) Elevation [m] map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily mean accumulated (acc. precip.) [mm] at different elevations. The black line represents a 365-day moving average. e) Shows mean daily accumulated precipitation plotted against elevation.	70
6.12	Wind speed against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily mean wind speed [m/s] at different elevations. The black line represents a 365-day moving average. e) Shows mean annual wind speed plotted against elevation.	71
6.13	Relative humidity against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily relative humidity [°C] at different elevations. The black line represents a 365-day moving average. e) Shows mean annual relative humidity plotted against elevation.	72
6.14	Incoming shortwave radiation against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The catchment is shown in grey. b)-e) shows daily mean incoming shortwave radiation [W/m ²] at different elevations. The black line represents a 365-day rolling mean. e) Shows mean annual SW radiation plotted against elevation.	73

List of Figures

6.15 Total annual volume and monthly observed river discharge measured at Arughat gauging station (2000-2015). The box plot in the left sub-figure show the locality, spread and skewness of the total annual volume through its quartiles. Each box extend from the lower quartile to the upper quartile. The lines outside the boxes indicates variability outside the quartiles. The dots (outliers) are values significantly different from the rest of the data set. The median is represented by a horizontal line inside each box. The sub-figure to the right shows the monthly median river discharge (2000-2015) as a black solid line. The dotted lines represent the 25 and 75 % percentiles of the monthly discharge. The grey dashed lines represent the observed discharge for each year in the period. 74

6.16 Monthly mean interpolated input variables from Shyft (2000-2015). a) Shows the mean monthly accumulated precipitation (mm), b) mean monthly temperature ($^{\circ}\text{C}$), c) mean monthly relative humidity (-), d) mean monthly wind speed (m/s) and e) mean monthly radiation (W/m^2). Regular grid and TINs have different radiation slope factor, and has therefore different interpolated values. 75

6.17 Interpolated temperature and precipitation for JJA and DJF in the PTSTK simulation using TIN-1071 (2000-2015). The scatter points are mid-points in the TIN-1071. 76

6.18 Mean interpolated precipitation (mm/hr) and temperature ($^{\circ}\text{C}$) against elevation (m), and elevation area against cumulative area (-). The values from the PTGSK model stack using TIN-1071 (2000-2015). 77

6.19 Optimized parameter ranges for the Gamma Snow and Snow tiles models using regular grid and TINs. The boxplots show the locality, spread and skewness of the CSI for each month through their quartiles. Each box extend from the lower quartile to the upper quartile. The lines outside the boxes indicates variability outside the quartiles. The dots (outliers) are values significantly different from the rest of the dataset. The median is represented by a horizontal line inside each box. 80

6.20 Figure shows the calibration run time [s] for all model configurations. Blue color shows the PTGSK model, and red shows the PTSTK mode. 81

6.21 NSE and LnNSE for each year during the simulation period (2000-2015). Figure a) shows the NSE, b) shows the LnNSE, c) shows the KGE and d) shows the RMSE. The colorscale for the NSE, LnNSE and KGE goes from -1 in red to +1 in green. RMSE values goes from the minimum RMSE value to the maximum value. The vertical dashed lines separate the to snow routines and the thick dash dotted lines divide calibration period from validation period. 83

6.22 Daily discharge during calibration (2000-01-01 to 2004-12-31) and validation period (2005-01-01 to 2015-12-01). The calibration period is shown in red and the validation period in blue. Observed discharge is in grey. PTGSK models are shown to the left, PTSTK to the right. A 365-day moving average is shown as a grey solid line. 85

6.23 Daily observed and simulated discharge in the year with a low NSE (2011) and a high NSE (2013). Solid lines represent simulated discharge, while dashed line represent the observed discharge. 86

6.24	QQ-plots over daily observed and simulated discharge during calibration (2000-2004) and validation (2005-2015). The two upper QQ-plots show the a) calibration period and b) the validation period. The two lower QQ-plots are shown on a log-scale. The 1%, 5%, 10%, 90%, 95% and 99% quantiles for the observed discharge are represented by the dashed vertical lines.	88
6.25	Observed discharge, mean daily precipitation, mean daily temperature and SWE in year 2005. Mean daily temperature [°C] is shown in red, daily precipitation [mm/day] in blue, SWE [mm] in green and observed discharge [m ³ /s] in black. The temperature and precipitation values are the interpolated values in Shyft for the region. The SWE value is calculated in Shyft using the calibrated PTGSK model for regular grid cells.	90
6.26	Comparison of mean fractional snow-covered area (SCA) between simulated (PTSTK) and observed (MODIS) during the winter season (DJF) and summer season (JJA) in the years 2000-2015. The polygons represent the glacier outlines from RGI 2017. Blue color represents no snow-cover and white represents full snow-cover.	92
6.27	Monthly Critical Success Index (CSI) for PTGSK. The CSI is calculated for each month using mean fSCA from MODIS as reference data. The CSI is shown using boxplots. Boxplots is a standardised way of showing the distribution of data using a five number summary (minimum, maximum, median, first quartile and third quartile (Galarnyk 2022)). The boxes, called inter-quartile range (IQR), range from the 25th to 75th percentiles of the data. The horizontal line inside the boxes show the median value. The extended lines shows the "minimum" (25th percentile minus 1.5-IQR) and "maximum" (75th percentile plus 1.5-IQR). The dots (outliers) are values different from the rest of the data set.	93
6.28	As in Figure 6.27, but for PTSTK.	94
6.29	Simulated mean fractional snow-cover duration (fSCD) plotted against mean observed fSCD from MODIS. The diagonal line is the identity line (1:1 line). Points following the identity line represents a perfect fit. Each scatter point represents the mean simulated fSCD within a given water year plotted against the mean fSCD from MODIS in that water year.	95
6.30	Snow water equivalent (SWE) [mm] and glacier melt [m ³ /s] for all simulations during validation period (2005-01-01 to 2015-12-31).	96
6.31	Difference in fractional snow-covered area between 2014 and 2001. Red values show an increase in fSCA values and blue values show a decrease. The figure is showing Reg-1077 PTSTK.	97
A.1	Shyft arcitechture. The yellow part of the figure represents Shyft components written in the Python language, while the blue represents the C++ code base. The green API is in Python. The figure is made using Miro , and is an adaptation of the figure in (Burkhart et al. 2021).	110
A.2	Overview of the main Shyft API types. API types used for simulation are shown to the left of the blue dashed line, while calibration API types are found to the right of the blue dashed line. The figure shows an example of the PTSSK model stack (Priestly-Taylor-Skaugen-Snow-Kirchner). The figure is an adaptation of the figure in (Burkhart et al. 2021), and was made using Miro	112

List of Figures

A.3 Illustration of the Inverse Distance Weighting (IDW) method. Illustration made in **Canva** 115

F.1 Regular cells and TINs. The upper plots show the elevation at each cell. The middle plots illustrate the amount of cells at each elevation bin. The bottom plots show how many cells that have a similar area (km²). 129

F.2 Land cover type fractions for a) Reg-1077 (regular grid cells), b) TIN-911 (low resolution TIN), c) TIN-1071, d) TIN-1382 and e) TIN-4681 (high resolution TIN). The plots at the top show the glacier fraction. White color represents a glacier and dark blue color means no glacier. The polygons in the glacier fraction plots represent glaciers registered in the Randolph Glacier Inventory 6.0 (RGI 2017). The bottom plots show the forest fraction where green is forest and white is no forest. 129

List of Tables

4.1	Summary of the forcing datasets	19
4.2	WFDE5 variables	21
4.3	SRTM 1 Arc-Second Global	23
4.4	Description of the regular cell data	24
4.5	MODIS dataset description	26
4.6	Forcing data and validation data	26
4.7	Coordinates used for Budhi Gandaki	27
4.8	Regular cell data set. The data set contains 9 variables representing the orography of a grid cell. There are 1077 cells with a float64 datatype.	31
4.9	Topography representations being used.	34
5.1	Parameters used to set up a Shyft simulation.	46
5.2	Parameters used to set up the region domain	47
5.3	Parameters used in the IDW routine	48
5.4	Parameters used in the BTK routine for temperature	50
5.5	Optimization method, time period, target and goal function used in calibration.	51
5.6	Calibration parameters for the PTGSK model stack	52
5.7	Calibration parameters for the PTSTK model stack	53
5.8	Multiple simulation and calibration configurations	54
5.9	Contingency table. Hit (H) indicates that both the simulated and observed snow-cover agree. Miss (M) means that snow-cover is detected by MODIS, but is not in simulation. False (F) is false alarm, and means that snow-cover is simulated, but not observed. True Null Event (Q) is when there is no snow-cover in both simulated and observed. The values for H, F, M and Q are summed values.	57
6.1	Central tendency and variation Shyft input data (WFDE5) covering the Budhi Gandaki region (2000-01-01 to 2015-12-31).	63
6.2	Statistics for the observed discharge (2000-2015) at Arughat gauging station	74
6.3	Results calibration using the Min BOBYQA optimization.	79
6.4	Model evaluation metrics. The first column shows the configuration type, and the rest of the columns shows the results for the NSE, LnNSE, KGE and RMSE for the calibration period and the validation period. The calibration is done daily for 5 year. The evaluation is done hourly for a 5 year calibration period and the validation is done hourly for the 10 years after the calibration period.	82
E.1	Description of the forcing data	125

List of Tables

F.1 Statistics of mesh geometries utilised in the Shyft simulations. 128

Chapter 1

Introduction

1.1 Motivation

The region encompassing the Himalayan and Tibetan Plateau is of immense significance as it provides water resources for millions of people residing in the surrounding areas (Bookhagen and Burbank 2010). The water from this region provides drinking water, supports agricultural activities, is used for generating hydro power, and catering to other agro-economic requirements (Ménégoz et al. 2013). The area holds tremendous potential for developing hydro power and thus making the transition towards a greener economy more feasible.

The Budhi Gandaki catchment, located in Nepal's Gorkha district, receives a high mean annual precipitation of 1495 mm and has an extremely diverse topography, with elevations varying from 479 to 8163 meters above sea level (R. P. Devkota et al. 2017). Its unique characteristics make it an attractive location for hydro power development, with an installed capacity of 1200 megawatts (MW) and an average energy generation of 3383 gigawatt-hours (GWh). However, the hydro power potential is subject to the prevailing climatic conditions in precipitation, evaporation, temperature, and snow/ice within the catchment, as noted in (Edenhofer et al. 2011). Climate change can have serious implications to the hydro power production (Dandekhya et al. 2017). Changing rainfall pattern and increased temperatures will affect power generation. The retreat of glaciers, expansion of glacial lakes, and alterations in the seasonality and intensity of precipitation are some of the factors that will impact power generation in the future (Dandekhya et al. 2017). In addition, the region is also susceptible to Glacial Lake Outburst Floods (GLOFs), which can cause devastating floods downstream (Dandekhya et al. 2017).

The complex hydrology of the High Mountain Asia poses challenges to hydrological modelling due to highly variable distribution of hydro-meteorological variables, steep elevation gradients and pronounced seasonality by the Indian Monsoon, as well as differing moisture regimes between the Tibetan Plateau and the Indian Ocean region (Bhattarai, Burkhart, Tallaksen, et al. 2020; Fan et al. 2019). Additionally, climate change can likely alter the regional water balance components (Koch et al. 2011). Accurate measurement of meteorological variables is challenging due to the region's heterogeneous topography (Pellicciotti et al. 2012). It is nearly impossible to make observations at all elevations, especially for snow. Insufficient data can lead to errors in discharge predictions, making satellite observations and reanalysis data sets vital for the decision-making processes. Furthermore, the input data and choice of models can have a considerable impact the accuracy of the outcome (Kauffeldt et al. 2016). In the

past, several hydrological models have been developed for various applications. Some of these models include SRM (Martinec et al. 1983), HEC-HMS (Halwatura and Najim 2013), J2000 (Nepal et al. 2014) and Shyft (Burkhart et al. 2021). Unlike other many other models, Shyft offers the advantage of integrating various well-known hydrological routines (each defined for processes such as evapotranspiration, snow accumulation and melt) into multiple models. This allows for testing multiple different scientific hypotheses using only this framework. Furthermore, it is important that a model can accurately represent terrain features since the terrain in High Mountain Asia is complex. Most distributed models discretize the catchment and derive terrain features from Digital Elevation Models (DEMs) (Bhattarai, Silantyeva, et al. 2020). Shyft is able to use several different catchment discretization methods, such as regular grid and Triangular Irregular Networks (TINs). TIN-based models have a better accuracy in representing complex terrain than grids, while also being computationally efficient. Few studies have used TINs in Himalayan catchments, but Bhattarai, Silantyeva, et al. 2020 has shown that TINs can improve simulation results compared with grids. Shyft is therefore considered a particularly valuable tool for its ability to evaluate different models, and to consider terrain heterogeneity and generate distributed output variables. Shyft also aims to become a FAIR framework for learning and research activities promoting open-science efforts (Wilkinson et al. 2016), which makes it ideal for scientific purposes.

1.2 Aim and objectives

The *primary aim* of this study to evaluate the Shyft modelling framework in its ability to simulate primary water balance components in the Budhi Gandaki catchment, with its complex terrain and limited availability of data. To achieve this goal the study will be divided into the several objectives.

The *first objective* will be to investigate spatial and temporal variation in the forcing data (temperature, precipitation, global radiation, relative humidity and wind speed). The recently added bias-adjusted ERA5 reanalysis data is chosen for this study because it has shown to have a lower mean absolute error and higher correlation with observed precipitation than the WDFEI reanalysis. The analysis of temporal variation is done to see if there are any trends or seasonality that might affect the hydrology of the High Mountain Asia and the Budhi Gandaki catchment. This is important because potential changes in the climate may affect the model performance. The spatial variation will be investigated through seasonal maps of the forcing variables and by statistical methods. The analysis of spatial variation may help explain if the forcing data is able to capture leeward and windward precipitation patterns that might affect the model performance. Furthermore, two different interpolation methods in Shyft (Bayesian Temperature Kriging and Inverse Distance Weighting) will be discussed.

The *second objective* is to evaluate the effect of different terrain representations and snow routines on model calibration and performance using observed river discharge as target. To achieve this, the catchment is discretized using two different discretization methods. A regular grid representation will be compared with four different Triangular Irregular Networks (TINs) of different resolutions. The two snow routines used include the Gamma Snow routine (simplified surface energy balance model for snow melt) and the recently introduced Snow Tiles routine. The Gamma Snow routine is a simplified surface energy balance model that has been tested in the region before with satisfactory results. The Snow Tiles routine is a temperature-index based model which has never been tested in the region. The Snow Tiles routine is chosen because it has considerably

fewer parameters than the Gamma Snow routine, potentially making calibration faster and thus lower uncertainty stemming from equifinality. The model performance will be evaluated by comparing simulated and observed discharge using visualisation and evaluation criteria such as the Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE) and Root Mean Square Error (RMSE).

The *third objective* is to evaluate the impact of the different terrain representations and snow routines on simulated snow-covered area, snow water equivalent, snow cover duration and glacier melt. Snow melt can be a significant contributor to runoff, and it is therefore important that the accuracy of modelled snow is assessed. Simulated snow-covered area (SCA) is compared with observed snow-covered area and snow-cover duration (SCD) derived using observed SCA by MODIS. This method provides an additional way of evaluating model performance. The evaluation criteria used for this comparison is the Critical Success Index (CSI).

1.3 Thesis outline

The Thesis is divided into several chapters. The introduction ([chapter 1](#)) outlines the motivation for this study. It also explains the aim of this study and its significance. The background chapter ([chapter 2](#)) introduces hydrological modelling in general terms and summarizes the current state of knowledge and knowledge gaps of the hydrology of High Mountain Asia. This chapter also identifies some key challenges of hydrological modelling in the region. The study area chapter ([chapter 3](#)) provides a description of the geography and climate for the study area. The data and pre-processing chapter ([chapter 4](#)) describes the data being used in the study, how it was acquired, and the pre-processing that has been done. The Shyft modelling framework will be introduced in the methods chapter ([chapter 5](#)). The methods chapter describes how the different simulations and calibrations are carried out, including the parameterization of the models and the methods used for calibration. The chapter also presents the methods that are being used for the evaluation of the model results, including statistical and graphical analyses. The results chapter ([chapter 6](#)) presents the results from the pre-processing of data and the simulation results from Shyft. It includes a presentation of the model outputs, focusing on the water balance components and the impact of different snow routines and terrain representations. The chapter also provides an evaluation of the model performance, comparing the simulated and observed data. The discussion part ([chapter 7](#)) discusses the results of the study and their implications for hydrological modelling in the study area. It provides an assessment of the Shyft modelling framework in its ability to simulate primary water balance components in a catchment with complex terrain using different snow routines and terrain representations. The chapter also identifies the key uncertainties and research needs arising from the study. The conclusions and outlook chapter ([chapter 8](#)) provides concluding remarks and suggestions for further work. It summarizes the main findings of the study and their significance for hydrological modelling in the region. The chapter also discusses the broader implications of the research for the understanding of the hydrological processes in complex mountainous terrain.

Chapter 1. Introduction

Chapter 2

Background

2.1 The importance of mountains as water towers

The world's mountains act as "water towers", meeting a considerable portion of both natural and human water demands (Immerzeel, Beek, et al. 2010; Immerzeel, Lutz, et al. 2020; Viviroli, Dürr, et al. 2007). These "water towers" are acutely sensitive and susceptible to changes in the climatic conditions (Hock et al. 2019; Viviroli, Archer, et al. 2011). A water tower generates seasonally a greater amount of runoff from precipitation than downstream due to orographic enhancement, and it stores water in the form of snow, glaciers and lake reservoirs resulting in a delayed water release caused by lower temperatures at high altitudes (Immerzeel, Lutz, et al. 2020). This buffering capacity provides a consistent supply of water to downstream areas, for example by supplying glacial meltwater during hot and dry periods. In their study, Immerzeel, Lutz, et al. 2020 ranked all water towers in terms of their water supplying role and their downstream dependence of ecosystems and society. They found that the dependency on the Asian river basins was higher than other basins across the world due to intense irrigation and dense population. In these basins, a large gap between demand and supply downstream is often then closed by unsustainable groundwater pumping (Immerzeel, Lutz, et al. 2020). The rivers originating from the high mountains in Asia are also among the most meltwater dependent river systems in the world (Lutz et al. 2014).

2.2 Climate change in High Mountain Asia and its effect on hydrology

Evaluations of the effects of climate change in High Mountain Asia are characterized by uncertainties stemming from discrepancies in climate change forecasts among various general circulation models (GCMs), regional disparities in climate projections, and the consequent response of the cryosphere (Lutz et al. 2014). Moreover, a lack of understanding of the present-day hydrological regime constitutes a noteworthy source of uncertainty when evaluating the impact of climate change on the hydrology of the region (Lutz et al. 2014).

In recent decades, observations indicate a decline in low-elevation snow-cover, glaciers, and permafrost due to effects of climate change (Hock et al. 2019). In High Mountain Asia, glaciers lost $-150 \pm 110 \text{ kg m}^2 \text{ yr}^{-1}$ in the period 2006-2015 (Hock et al. 2019). Snow-cover, glaciers and permafrost are projected to decline in almost all regions throughout the 21st century, and low level snow-cover will likely decrease by

10-40% for 2031-2050 compared to 1986-2005 (Hock et al. 2019). The decline in glaciers, permafrost and snow has already altered the frequency, magnitude and location of most related natural hazards, and changed high mountain ecosystems in recent decades.

Until at least 2050, the quantity of glacier melt water that contributes to the total flow in all basins in High Mountain Asia is expected to remain relatively stable, as the reduction in glacier area is balanced out by an increase in the melt rate (Lutz et al. 2014). The contribution from rainfall to runoff in the Upper Ganges Basin will largely depend on precipitation projections which are highly uncertain. In general, the RCP4.5 and RCP8.5 both indicate an increase in precipitation in the Upper Ganges Basin during the monsoon season, resulting in an increase in total annual runoff by 10% for RCP4.5 and 27% for RCP8.5. Since absolute amounts of glacier and snow melt do not experience considerable changes, their relative contribution to runoff is likely to decrease in the Upper Ganges Basin (Lutz et al. 2014).

The projected snow-cover and glacier decline will further change the amount and seasonality of river runoff in snow dominated and glacier-fed river basins with negative impacts on agriculture, hydro power, and water quality in some regions (Hock et al. 2019). The average winter snow melt runoff is expected to increase, and spring peaks are projected to occur earlier. While projected trends in annual runoff vary among regions, there is high confidence that the average annual runoff from glaciers will have reached a peak that will decline at the latest by the end of the 21st century (Hock et al. 2019). This declining runoff is expected to reduce the productivity of irrigated agriculture and impact hydro power operations.

Even though hydro power comprises only 2.5% of the total primary energy supply in the world (in 2019) (IAE 2023), it is close to 100% in many mountainous regions (Hamududu and Killingtveit 2012). There is robust evidence that hydro power facilities will experience changes in their water input due to cryosphere-related impacts of runoff (Hock et al. 2019). Increased runoff, especially during winter and spring, can increase the risk of overflows, and increased volumes of glacier lakes can damage hydro power infrastructure if they suddenly empty (Hock et al. 2019). Nepal has been among the countries where Glacier Lake Outburst Floods (GLOFs) have had greatest impact on the nationwide economy (Carrivick and Tweed 2016). One example of this was the Dig Tsho flood in Khumbu Himal of Nepal in 1985 that damaged a hydro power plant and caused economic losses estimated to 500 million USD (A. B. Shrestha and Joshi 2011). In February 2021, Uttarakhand state in India experienced a devastating landslide that resulted in the destruction of two hydropower plants and the loss of over 200 lives, either through fatalities or missing persons (Shugar et al. 2021).

2.3 The hydrology of High Mountain Asia

Khanal et al. 2021 identifies 15 main river basins in the High Mountain Asia. These include Amu Darya (AMU), Balkash (BAL), Brahmaputra (BRA), Ganges (GAN), Helmand (HEL), Indus (IND), Irrawaddy (IRR), Mekong (MEK), Plateau of Tibet Interior (TP), Salween (SAL), Syr Darya (SYR), Tarim interior east, (TIE) Tarim interior west (TIW), Yangtze (YAN) and Yellow (YEL). The Indus and Tarim interior river basins have the highest glaciated area with 97 590 glaciers covering over 29 700 km² and 27 700 km². Khanal et al. 2021 find that during the years 2001-2004, the IND basin has similar contributions from glacier melt, rainfall and snow melt. The AMU basin has similar contributions from snow and glacier melt. The Ganges basin is mainly fed by rainfall and SYR by snow melt. BRA, MEK and SAL have relatively similar

contributions from rainfall, baseflow and snow melt, where rainfall is the dominant contributor.

The climate in the eastern part of the Himalaya is characterized by the East-Asian and Indian monsoon, resulting in most precipitation occurring during June-September with a strong north-south gradient, while the precipitation the west (Hindu-Kush and Karakoram ranges) is distributed more evenly over the year due to westerly and southwesterly flows (Lutz et al. 2014). The precipitation in the Nepalese Himalays is in large parts controlled by a large-scale circulation in combination with orographical and thermal induced circulation (Ueno and Yamada 1990). The precipitation may also vary with elevation depending on the precipitation mechanism (Immerzeel, Petersen, et al. 2014). The monsoon is the dominant mechanism during June-September, and the precipitation decreases from Kathmandu (1400 m) and higher (Immerzeel, Petersen, et al. 2014). The precipitation during winter is produced by synoptic disturbances from the west (Immerzeel, Petersen, et al. 2014). The ratio of solid and liquid precipitation within a basin is determined by basin hypsometry, and the contribution of each runoff component to the total runoff determines a basin's response to climate variability and change (Lutz et al. 2014). The runoff components of a basin consist of solid precipitation, which can be stored long-term as perennial snow and ice or short-term as seasonal snow before melting, and liquid precipitation, which runs off directly. Infiltration into the soil and recharge to groundwater can further delay each of these components. The contribution of these components to the total runoff determines a basin's runoff composition, and influences its response to climate variability and change. There is a significant difference in the amount of glacier and snow melt contributing to the total runoff in High Mountain Asia's river basins, and this is poorly quantified. The lack of understanding about the hydrological regimes in the area is a significant factor in the uncertainty of assessing the impacts of climate change on the region's water systems (Lutz et al. 2014).

2.4 Hydrological modelling in High Mountain Asia

A hydrological model can be defined as the characterization of real hydrologic features and systems by the use of small-scale physical models, mathematical analogues and computer simulations (Allaby 2008). Hydrological models are used to address environmental transport questions where water excess, scarcity or dissolved or solid content is of primary importance (Burgess 1986). There may exist many plausible solutions to such questions, and therefore there is not a single best model (Alderton and Elias 2021). Hydrological modelling aims to analyze the separate flows within a hydrological system, with various parameters influencing the intensity of these flows, as well as the characteristics of the soil, land, climate, and rivers (Anees et al. 2016). The accuracy of these models is largely dependent on the number of parameters considered and their precise implementation within the modeling techniques, along with specific boundary conditions like atmospheric input data.

Hydrological models can be categorized based on the presence of random variables, their spatial distribution and temporal variation as shown in [Figure 2.1](#) (Dwarakish and Ganasri 2015). Broadly speaking hydrological models are either stochastic or deterministic. In addition, hydrological models can be classified on whether the hydrological process is described empirically, conceptually or physically. A stochastic model involves a random variable (i.e. explicitly uncertain). This entails that a stochastic model that is executed repeatedly with identical inputs, will generate unique outcomes of the hydrological system, which are distinguishable from each other. In a deterministic model randomness is not considered, meaning a given input will always produce the same outcome (Dwarakish and Ganasri 2015). Both stochastic and deterministic models can be used for predictions. A hydrological model can be further categorized on a spatial basis as lumped, semi-distributed or distributed.

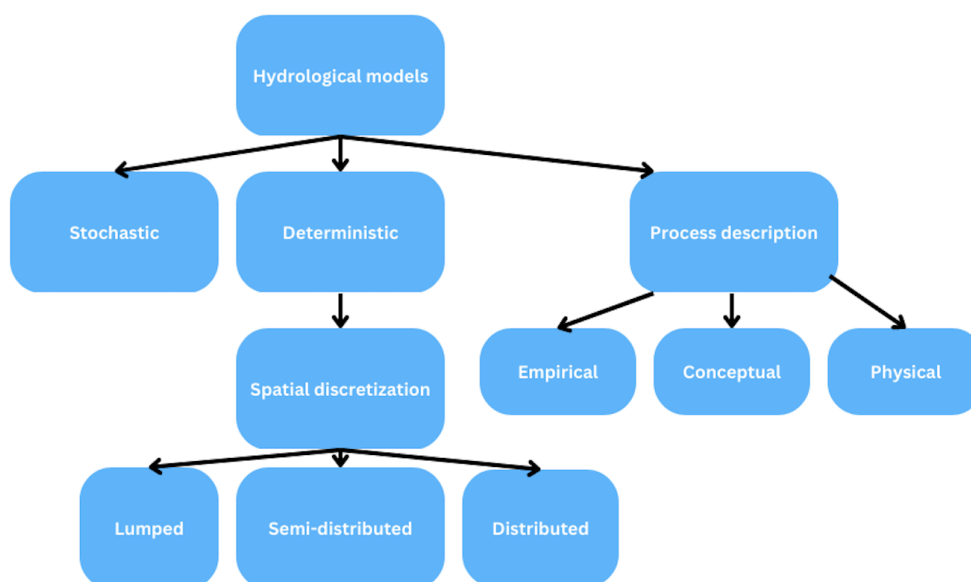


Figure 2.1: Classification of hydrological models. Hydrological models can broadly be categorized as stochastic or deterministic. They can also be characterized by how the hydrological process is described as conceptual, empirical or physical. The figure is made using [Canva](#).

Deterministic lumped models treat the river catchment as a single unit when the parameters are computed and the variables are averaged over this unit (Dwarakish and Ganasri 2015). Deterministic semi-distributed models partition the catchment into smaller sub-catchments or Hydrologic Response Units (HRUs) based on variables such as land use, slope or soil type (Dwarakish and Ganasri 2015). Deterministic distributed model defines the hydrological processes occurring on a grid and expresses the model variables as functions of the spatial dimensions.

Figure 2.2 shows the main components of the hydrological cycle. The hydrological cycle is the continuous circulation of water in the Earth-Atmosphere system (NOAA 2023). Generally speaking it is the flow of water from the ground to the atmosphere and back again. Evaporation is the processes when a substance, here liquid water, changes state to gas (NOAA 2023). This process requires energy which can come from the Sun, the atmosphere, the Earth or objects on the surface (NOAA 2023). Evaporation takes up latent heat, in contrast to condensation that releases latent heat. Transpiration is when water evaporates from plants through small openings underneath leaves connected to vascular tissue called stomata. The process where water vapour in the atmosphere changes state to liquid state is called condensation (NOAA 2023). This process releases latent heat. Precipitation is when small condensation particles grow too large, through coalescence or collision, for the rising air to support such that the particles fall to the earth in the form of rain, snow, hail or sleet (NOAA 2023). Runoff occurs when there is excessive water from precipitation and the ground is saturated (NOAA 2023).

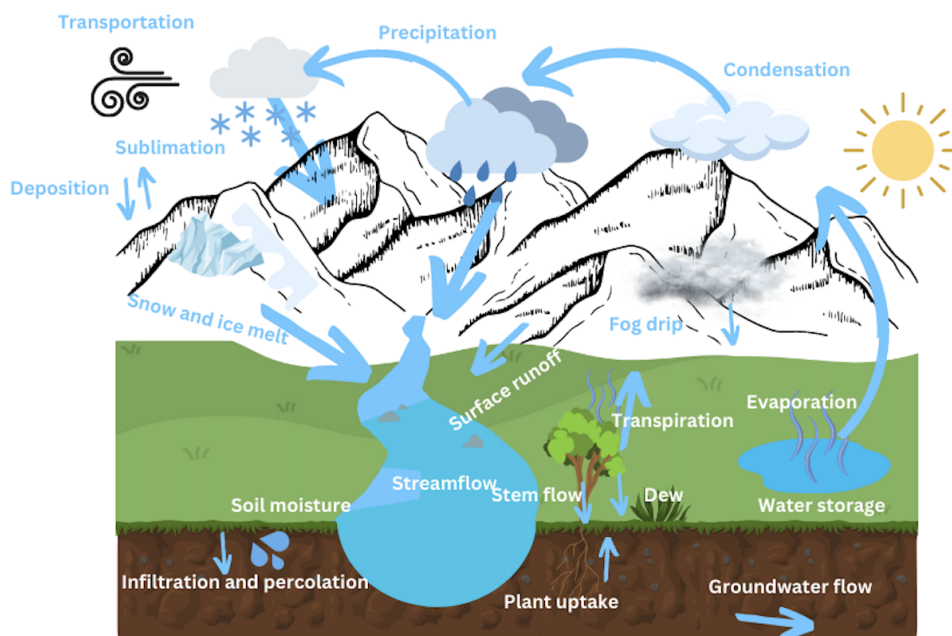


Figure 2.2: The hydrologic cycle. The figure illustrates the major components of the hydrologic cycle. The figure is made using [Canva](#).

There are several challenges associated with hydrological modelling in High Mountain Asia. First of all, the complex topography and land use cover poses challenges to hydrological modelling (Bhattarai, Burkhart, Tallaksen, et al. 2020). The distribution of hydrometeorological variables is very heterogeneous, varying both in space and time (Bhattarai, Burkhart, Tallaksen, et al. 2020). The rugged and remote terrain of the region makes it also difficult to have a complete coverage of ground observations (S. K. Mishra et al. 2021), and many in-situ observations records are too short for assessing trends and variability in a robust way (Yatagai et al. 2012). Insufficient amount of reliable data sources can then lead to errors in discharge predictions (Kauffeldt et al. 2016). Even though remotely sensed products are available, studies suggest that they can suffer from retrieval relative uncertainties by as much as 100-140%, where these biases can be enhanced over mountainous and snowy regions (Y. Liu and S. A. Margulis 2019; Tian and Peters-Lidard 2010).

While the large-scale patterns in precipitation are fairly well known in the region, the details that are needed for understanding precipitation at scales relevant for water resource management are less defined (S. K. Mishra et al. 2021). There is pronounced seasonality by the Indian Monsoon, as well as differing moisture regimes between the Tibetan Plateau and the Indian Ocean region (Bhattarai, Burkhart, Tallaksen, et al. 2020; Fan et al. 2019). The general precipitation uncertainty in the region is high due to a lack of reliable in-situ observations (S. K. Mishra et al. 2021). There is also heterogeneity in the snowpacks and challenges in measuring the snow-water equivalent (SWE) remotely (Bormann et al. 2018). Furthermore, the parametrization of the snow albedo is associated with a lot of uncertainty. Therefore the amount of solar radiation being absorbed and the rate of melt will also be uncertain (S. K. Mishra et al. 2021). The presence of light-absorbing particles in snow can lead to a reduction in the snow's ability to reflect sunlight, and thus increase snowmelt. Such snow impurities can contribute a notable amount to annual discharge (Matt and Burkhart 2018).

Land use, such as groundwater extraction, can modulate the local hydrology, but is studied at either a very coarse scale or with a focus on a small regional scale, making regional studies of groundwater challenging and reliant on coarse resolution gravimetry observations (S. K. Mishra et al. 2021). GRACE is an example of a gravimetry technique, which estimates terrestrial water storage anomalies at a coarse resolution of ~ 300 km (Giroto et al. 2017). The impacts of permafrost degradation on downstream hydrology is also poorly quantified (Sun et al. 2019).

There is a lack of understanding of the spatial and temporal variations in climate, surface impurities of snow and ice such as black carbon and dust which alter the surface albedo, and the glacier mass balance and dynamics (Matt and Burkhart 2018; S. K. Mishra et al. 2021). The knowledge gap presents a considerable challenge in determining the precise locations and timing of economically and ecologically changes in river flow (S. K. Mishra et al. 2021).

2.5 Overcoming challenges in hydrological modelling in High Mountain Asia

Reanalysis data and satellite observations are useful tools that can be used to address some the challenges mentioned previously. The region has complex terrain and data sources are scarce, but reanalysis data can provide a long-term record of weather and climate conditions that are essential to capture the variability and trends in the region's long-term weather. It is important capture variability in mountainous regions because the weather conditions can vary over short distances. Reanalysis data can also provide valuable information on near-surface hydro-meteorological variables as there are very few ground stations in the area. Remote sensing data may serve as validation data, since it is not possible to measure the snow-cover area in any other way on a regional scale. However, it is worth noting that reanalysis data is derived from a combination of observed and observational data (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020), which may have biases or errors in their simulations. Thus, errors in the observational data may propagate into the assimilated models. The spatial resolution may also be too coarse for impact studies depending on the data set (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020).

Frequent cloud cover is a major challenge to satellite observations of snow-cover (Gurung et al. 2011). Estimating snow-cover in areas where there is a steep elevation gradient is also challenging since there may be large spatial variation within the horizontal resolution (e.g. MODIS has 500 meter resolution) that a satellite measure (Khatri et al. 2018). However, results from various studies have shown that daily MODIS snow maps have an overall accuracy of about 93%, but that the accuracy is lower in forested and complex areas (D. K. Hall and Riggs 2007). Spatial heterogeneity can lead to mixed pixel problems, meaning a pixel is composed of more than one end member. Aalstad et al. 2020 show that the mixed pixel problems can be accounted for in MODIS using a regression-based retrieval of fractional snow-covered area (fSCA). This method has given satisfactory results for MODIS in high-arctic, Svalbard (Norway).

In this study, reanalysis data and satellite observations of snow-covered area is used. Despite their limitations, they provide a good temporal and spatial coverage for the study area. The ERA5 reanalysis data used in this study has been bias corrected using well documented methods and is tailored for impact studies (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020). The satellite observational data of snow-cover by MODIS has shown satisfactory performance for the region (P. Mishra et al. 2016), and serves as validation data for the modelled snow-covered area.

Chapter 2. Background

Chapter 3

Study area



Figure 3.1: View over the Budhi Gandaki valley. Image captured on the Manaslu Circuit Trek with Manaslu (8163 m.a.s.l.) to the right and Samdo to the middle (5177 m.a.s.l.). Photo taken by [Hendrikz \(2019\)](#).

The Gandaki river basin is a transboundary basin located in the central Himalayan region, and a tributary to the Ganges river. The Ganges river basin is mainly fed by rainfall (Khanal et al. [2021](#)). It extends from the Tibetan Plateau in China to the north, through Nepal to India in the south (Dandekhya et al. [2017](#)). The Gandaki river basin has a total drainage area of 46 300 km² with 72% in Nepal, 18% in India and 10% in China. (Dandekhya et al. [2017](#)). The river has seven major tributaries, including Kali Gandaki, Seti Gandaki, Madi Marsyangdi, Daraudi, Trishuli and Budhi Gandaki.

Figure 3.2 shows a map of the Budhi Gandaki catchment. The catchment lies between the borders of China and Nepal as can be seen in the top left figure. The bottom left figure shows the mean fractional snow cover area (fSCA) in the winter of 2005-2006. Most of the snow can be found in the northern parts of the catchment. The land cover type figure shows that most parts of the catchment is grassland, barren or glaciated. The figure to the right shows the extreme variation in elevations in the catchment. Manaslu, the eighth highest mountain in the world (8163 m.a.s.l.), is located to the western edge of the map.

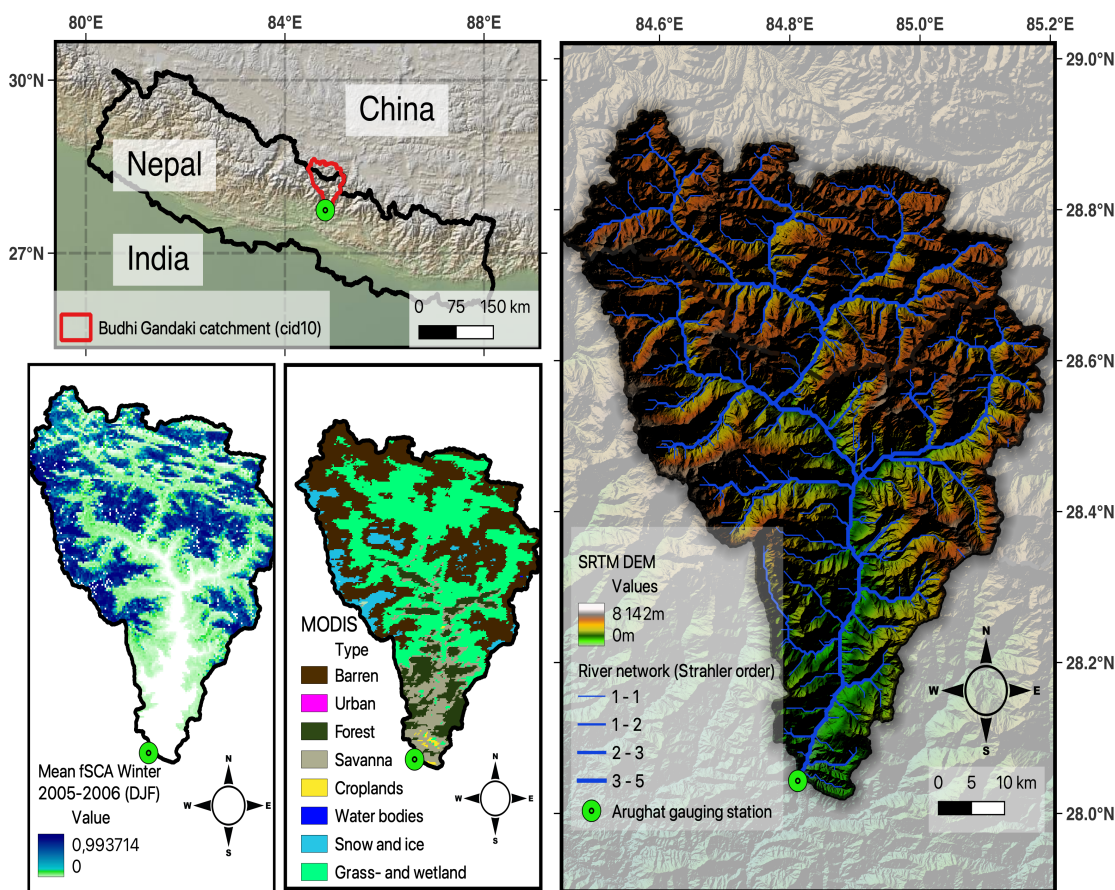


Figure 3.2: Map of the Budhi Gandaki catchment. The figure showcases an overview map to the top left, a snow cover map to the bottom left, a land cover type map in the bottom center and an elevation map to the right. The overview map displays the national boundaries of Nepal, providing context for the location of the Budhi Gandaki catchment within the country. The catchment is displayed in red and cid10 is the catchment id used in this study. The fractional snow cover area (fSCA) map depicts the mean distribution and extent of snow cover during the winter season 2005-2006 (DJF). Values range from 0 (no snow cover) to 1 (completely snow covered). The snow cover data is derived from **MODIS** (NSIDC 2023). The land cover map identifies different land cover types, such as forest, water bodies, urban areas, and grass and wetland, offering insights into the composition and distribution of the landscape. The data is derived from the **MODIS Land Cover Type product** (Sulla-Menashe and Friedl 2018). The elevation map visualizes the varying topography and elevations within the catchment. Green color shows low elevations and brown/white shows high elevation. The river network is displayed in blue, where the thickness of the line represents the stream size (using **Strahler order**). Arughat gauging station is marked with a green point.

3.1 Geography



(a) Picture of the Ribum Monastery captured on the Manaslu Circuit Trek. Photo by [Gaiusmarius181/Wikimedia Commons](#) (2017).



(b) Budhi Gandaki river view from Arughat bazaar located between Dhading and Gorkha district of Nepal. Photo by [Mzhushain/Wikimedia commons](#) (2019).

Figure 3.3: Landscape diversity in the Budhi Gandaki catchment. The climate in the catchment ranges from temperate to polar tundra.

The Budhi Gandaki catchment is located in central Nepal (between 27°50' and 29°00' N latitudes, and 84°30' and 85°10' E longitudes (Marahatta, L. Devkota, et al. 2021)). The catchment is part of the larger Narayani River Catchment originating in Tibet to the north, and is encompassing parts of the Gorkha and Dhading districts of Nepal (R. P. Devkota et al. 2017). The Budhi Gandaki catchment is one of the main tributaries of the Trishuli River in Nepal which in turn is a major tributary to the Gandaki River that meets the Ganges in India (Khatri et al. 2018). The Ganges catchment has the third-largest glaciated area (1.1 %), following Brahmaputra (3.1%) and Indus (2.2%), which all originate from the Hindu Kush region (Immerzeel, Beek, et al. 2010).

The Budhi Gandaki catchment is about 113 km long and between 15 and 30 km wide (Marahatta, L. Devkota, et al. 2021). The catchment area upstream of the gauging station at Arughat bazaar is 3871.38 km^2 (Khatri et al. 2018). About 56 % of the catchment area is in Nepal and 44 % lies in China. The catchment elevation ranges from green hills (479 m.a.s.l.) at the gauging station in the south to the snow covered Manaslu mountain peak (8163 m.a.s.l.) (Khatri et al. 2018).

Figure 3.4 shows a south-north transect of the Budhi Gandaki catchment using the SRTM 1 Arc-Second Global DEM file provided by [USGS Earth Explorer](#) in QGIS ([EROS 2017](#); [JPL 2013](#)). Figure 3.4a shows a map of the transect and Figure 3.4b shows the elevation profile of this transect plotted in Python. The figure shows that the catchment can be divided into roughly three physiological zones (elevation and topography). The Siwalik hills is a zone characterized by dry unconsolidated soil materials which are prone to erosion and landslides ([Dandekhya et al. 2017](#)). Lesser Himalaya is characterized by hills with large areas of dense broad-leaved and mixed forest with extensive agriculture ([Dandekhya et al. 2017](#)).

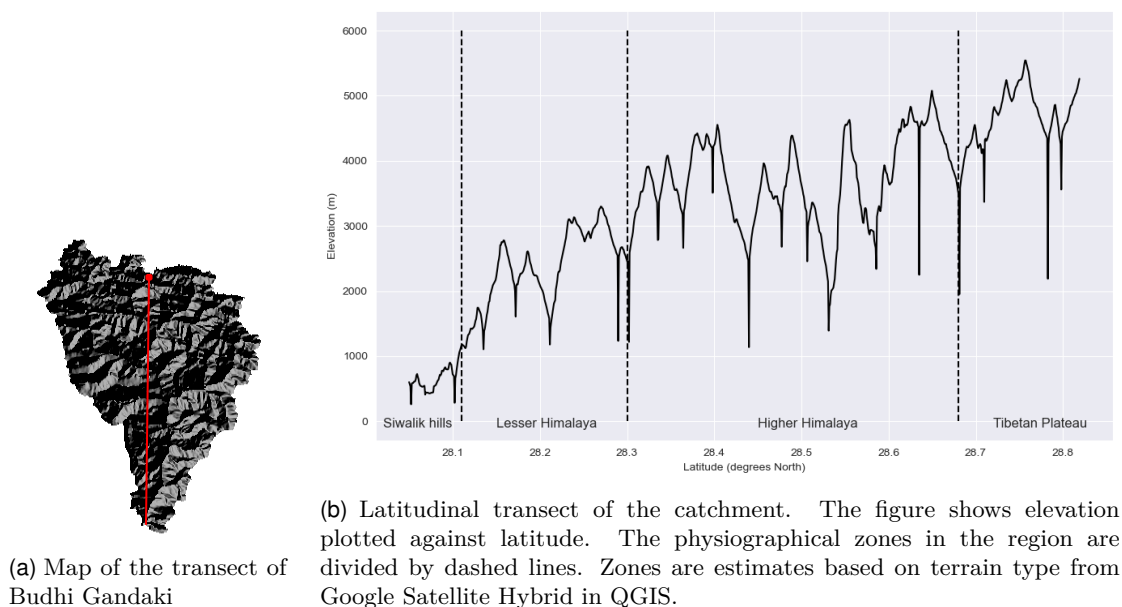


Figure 3.4: Elevation profile of the Budhi Gandaki catchment

3.2 Climate

The Himalayan region has a tundra climate characterized by covered by snow- and ice-covered areas ([A. B. Shrestha and Joshi 2011](#)). The snow line elevations are higher in the east than in the west, and the snow cover extent is highest during late winter and spring ([A. B. Shrestha and Joshi 2011](#)). The permanent snow line in Himalaya is about 5000 meters above sea level, where precipitation occurs in the form of rain and snow ([Pangali Sharma et al. 2020](#)).

According to [Pangali Sharma et al. 2020](#), the Budhi Gandaki catchment contains three climatic zones. The climate changes from temperate (between 479-3500 m) to polar tundra (3500 – 6500 m) and polar frost (above 6500 meters).

Figure 3.5 shows the elevation plotted against cumulative area (%) in sub-figure a) and the total percentage of land cover types in the catchment in sub-figure b). In sub-figure a), 24.4% of the area is located in the temperate zone which is characterized by dry winters and warm summers ([Karki et al. 2016](#); [Pangali Sharma et al. 2020](#)). 74.6% of the catchment is located in the polar tundra zone with bare landscape and permafrost. About 1% of the catchment area is characterized by polar frost climate where there is permanent snow/ice. Sub-figure b) shows that forest is the most dominant land cover type in the catchment (about 37%). 8% of the catchment is cropland and 17 % is

herbaceous vegetation (savannas, grassland or lichens/mosses). The remaining area is sparsely vegetated, where vegetation is 19%, water bodies is 0.04%, and permanent snow/ice is 18%. About 80% of the catchment area is either polar tundra or polar frost, but less than 20% of the catchment is covered with permanent snow/ice. Therefore, most of the polar tundra is not permanent snow/ice.

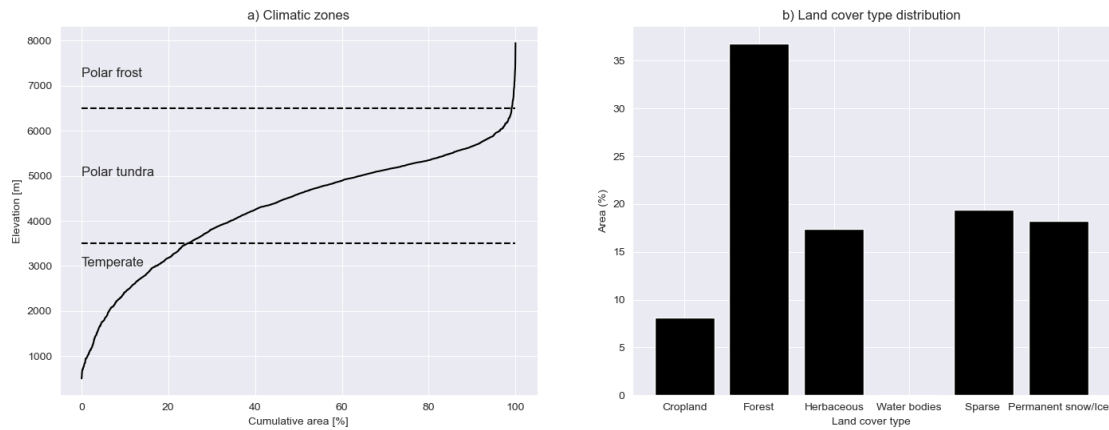


Figure 3.5: a) Climatic zones in the Budhi Gandaki catchment. The figure shows the elevation plotted against cumulative area [%]. Elevation and area is based upon the values found in the TIN with the highest resolution (4681 cells). The climatic zones are based upon (Khatri et al. 2018). b) Land cover type distribution in Budhi Gandaki catchment. The figure shows aggregated values based on the GlobCov 2009 land cover type data (Bontemps et al. 2011).

Based on historical trends, the upstream part of the Budhi Gandaki catchment is getting wetter while the central and down-stream area is getting slightly drier (Siwakoti et al. 2016). The temperature increase in Nepal is predicted to be 0.6 °C per decade, slightly less than the global average of 0.74 °C (B. Shrestha et al. 2007; Stocker 2014). These historical trends are expected to continue. The annual precipitation is expected to increase from +1% to +35% compared to the 1981-2005 baseline under three RCP scenarios (RCP2.6, RCP4.5 and RCP8.5) and three future time windows (near future 2011-2041, mid-future 2042-2072 and far future 2073-2100). Similarly, the average monthly minimum and maximum temperatures are expected to increase by +1% to +6% and +2% to +7% respectively (Siwakoti et al. 2016). The frequency of extreme events, such as high floods and extremely high floods, are projected to increase, especially in the monsoon season. The water availability is projected to increase until the end of the century.

The mean temperature ranges from 6 °C during winter to 35 °C in summer measured near the Budhi Gandaki Hydropower Project proposed dam site (R. P. Devkota et al. 2017). A study conducted by Baidya et al. 2008 examined daily temperature data from eight stations and rainfall data from 26 stations across Nepal's climate zones. Of these stations, three for temperature and five for rainfall were located in the Gandaki basin. The study revealed an increasing trend in the number of warm days and a decreasing trend in the number of cool days, which was more pronounced at higher elevations. The warming trend seem to be more significant at higher elevations (Bajracharya et al. 2011; Chaulagain 2009; A. B. Shrestha, Wake, et al. 1999).

There is substantial spatial variability in long-term annual precipitation within the catchment (Bhattarai, Silantyeva, et al. 2020). The long-term annual catchment precipitation (1983-2012) is 1301 mm (maximum of 278 mm in July and minimum of 11 mm in November (Marahatta, L. Devkota, et al. 2021)). The annual precipitation

varies from greater than 2500 mm at Arughat gauging station to less than 700 mm in the Tibetan part of the catchment (Marahatta, L. Devkota, et al. 2021). The hydrological regime is greatly influenced by the seasons (Bhattarai, Burkhart, Tallaksen, et al. 2020). The four seasons are defined as winter (December to February), pre-monsoon (March to May), monsoon (June to September) and post-monsoon (October to November) (Bhattarai, Burkhart, Stordal, et al. 2019; A. B. Shrestha, Wake, et al. 1999). The catchment receives about 80 % of the annual precipitation during the monsoon season (Khatri et al. 2018). The maximum relative humidity is also reached during the monsoon season. The precipitation decreases from west to east during this period, while the greatest intensity in precipitation is occurring on the south facing slopes of the mountains. The catchment is among the tributaries to the Narayani river that are monsoon fed (those in the middle and high mountain regions), in contrast to the glacier and snow melt fed (those originating in a higher Himalayan region) (Bhattarai, Burkhart, Tallaksen, et al. 2020). The precipitation in the winter is largely due to the western disturbances (Dimri 2004; Dimri, Niyogi, et al. 2015). These are cyclonic storms associated with the mid-latitude Sub-tropical Westerly Jet (Dimri, Niyogi, et al. 2015). The winter precipitation provides input to the glacier mass balance and the snowpack, and modulates the albedo (Dimri, Niyogi, et al. 2015).

The orography (topographical relief) controls and/or modulates the precipitation patterns in the region and its elevation dependent distribution (Dimri, Palazzi, et al. 2022). The amount and intensity of precipitation are significantly influenced by various factors, including topography, strength and moisture content of the wind, and the relief of the range. These factors result in more precipitation on the windward side compared to the leeward side of the Himalayan mountain range (Anders et al. 2006; Singh and Kumar 1997). The topography acts as a physical barrier which alters the flow of the wind and disturbs the vertical stratification of the atmosphere.

The river discharge in the Gandaki river basin varies throughout the year, affected by both precipitation and snow melt (Dandekhya et al. 2017). The observed stream flow at Arughat seems to closely follow the precipitation pattern (Marahatta, L. P. Devkota, et al. 2021), with a mean annual flow near the Budhi Gandaki Hydroelectric Project (BGHEP) dam site that is estimated to be 222 m/s (Dordé and L'Hostis 2015).

3.3 Hydropower

During the period of interest in this study (2000-2015), there are no registered power plants with an installed capacity above 1 MW in the Budhi Gandaki catchment (DOED 2023). However, a hydroelectric power plant located on the Budhi Gandaki river has been proposed (Ministry of Energy 2023). The project was identified during a feasibility study at the end of the 1970's, and a feasibility project was done for the project in 1984. The proposed power plant is to be located at Benighat, which is 2 km upstream of the confluence between Budhi Gandaki river and Trishuli river (55 km from Kathmandu). The recommended capacity of 1200 MW will make it the largest hydro power plant in Nepal (Bhushal 2021). The project was meant to be finished by 2022, but has gone through numerous changes of changes since 2011 and its future is uncertain due to the massive costs and uncertainty related to relocation of people (Bhushal 2021).

Chapter 4

Data and pre-processing

In this chapter the data sources used in the study will be described. The chapter also includes how the data is being handled and pre-processed. All the data used in this study is available in the [Zenodo repository](#) for this study (Jacobqs 2022). The code can be found at the [GitHub repository](#) for this study (J. Q. Skavang 2023).

4.1 Meteorological forcing data

The bias-adjusted ERA5 reanalys data for impact studies (WFDE5) is used as forcing data (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020). Forcing data are used to drive the hydrological model. Shyft needs precipitation [mm/hr], relative humidity [-], temperature [°C], wind speed [m/s] and incoming shortwave radiation [W/m^2] as input data. Note that Shyft does not use incoming longwave radiation, but calculates (see [subsubsection 5.1.1](#)) net radiation using the temperature, T [K], global radiation [W/m^2], relative humidity [-], and the vapor pressure [kPa] (Statkraft 2023). [Table 4.1](#) gives a summary of the forcing data used in this study.

Table 4.1: Summary of the forcing datasets

Forcing dataset	Data period	Spatial resolution	Temporal resolution
WFDE5	1990-2019	$0.5^\circ \times 0.5^\circ$	Hourly

4.1.1 The bias-adjusted ERA5 reanalysis data for impact studies (WFDE5)

The bias-adjusted ERA5 reanalysis data for impact studies (WFDE5) is generated by applying the WATCH Forcing Data methodology to surface meteorological variables from the ERA5 reanalysis (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020). Reanalysis data is a combination of model data and worldwide observations into a globally complete and consistent data set using data assimilation (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020; Hersbach et al. 2020). The WFDE5 data set contains 11 variables with an hourly temporal resolution on a regular longitude-latitude 0.5 degree grid. The data set has a global coverage, but is only defined at land and lake points. The WFDE5 data set is spanning from January 1979 to the end of 2019, and has been adjusted using an elevation correction and monthly-scale bias based on Climatic Research Unit (CRU) data (for temperature, diurnal temperature range, cloud-cover, wet days number and precipitation fields). CRU is a widely used climate data set with monthly climate variables interpolated from an extensive network of weather station observations (Harris et al. 2020). The WFDE5 data set used in this study is bias-corrected using CRU TS4.03 from CRU (Harris et al. 2020) from 1979 to 2019 for all variables, and the GPCCv2018 full data product (Schneider et al. 2018) for rainfall.

The WFDE5 dataset is distributed by the Copernicus Climate Change Service (C3S) through the Climate Data Store (CDS) as monthly netCDF files that are downloaded as zip-files or compressed tar files (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, Cagnazzo, et al. 2022). NetCDF files are self-describing data that are machine-independent. The netCDF format supports creation, access, and sharing of array-oriented scientific data. The data can be downloaded via the CDS API or through the [C3S Climate Data Store](#). In this study the WFDE5 data was downloaded using the [CDS API](#). The code for downloading the data is included in the [GitHub Repository](#).

The near-surface air temperature, specific humidity, pressure, snowfall flux, rainfall flux and surface downwelling shortwave radiation data variables are downloaded for the period 1990-2019. The near-surface wind speed is downloaded only for the years 1999 to 2016 since it is only used for the simulation period (2000-2015), and to save space. The data is downloaded for the entire High Mountain Asia (HMA) for regional analysis purposes, and then later a subset of this data set is selected for the Budhi Gandaki region which is used in Shyft simulations (see [subsection 4.11.2](#)). The selected coordinates for the HMA region are 60.85°-105.04° East and 15.95°-39.31° North based on the coordinates used by Bajracharya et al. 2011.

Table 4.2 describes the near-surface variables from WFDE5 used in this study. The tables provides a description, units and the monthly bias correction methods used on the ERA5 variables.

Table 4.2: WFDE5 variables

Variable name	Description	Units	Monthly bias correction method
Wind	10 m wind speed	m s^{-1}	-
Tair	2 m air temperature	K	1)
PSurf	Pressure at the surface	Pa	-
SWdown	Downward shortwave radiation flux	W m^{-2}	2)
Rainf	Rainfall flux	$\text{kg m}^{-2} \text{s}^{-1}$	3)
Snowf	Snowfall flux	$\text{kg m}^{-2} \text{s}^{-1}$	3)
Qair	2 m specific humidity	kg kg^{-1}	-

1) CRU TS4.03 temperature and diurnal temperature range

2) CRU TS4.03 cloud and effects of inter-annual changes in atmospheric loading

3) ERA5 ratio of rainfall/precipitation and snowfall gauge correction

Bias correction in WFDE5

Reanalyses have been used for a long time in impact studies because it provides a physically consistent global reconstruction of historical weather without any gaps in space or time (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020). Although reanalyses, such as the ERA-Interim global reanalysis, are designed to represent the most plausible state of the atmosphere, land surface and ocean waves (Dee et al. 2011), the assumptions being made in sub-grid parametrizations and the coarse resolutions of models are known to produce biases relative to ground-based observations and radisondes (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020).

The ERA5 reanalysis has recently superseded the ERA-Interim reanalysis (Hersbach et al. 2020). The ERA5 reanalysis is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and provides a detailed record of the global atmosphere, land surface and ocean waves from 1950 onwards (Hersbach et al. 2020). It is updated with a 5 day latency and quality checks 2-3 months later. The move from ERA-Interim to ERA5 prove to be a step change in overall quality and level of detail (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020).

The WATCH Forcing Data methodology applied to surface meteorological variables from ERA5 reanalysis (WFDE5) is made to have a common framework for land surface models and global hydrological models to assess the hydrological cycle in the 20th and 21th centuries (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020). Before elevation and bias correction are performed on ERA5 from the Climate Data Store (CDS), the ERA5 reanalysis is regridded to a 0.5° by 0.5° resolution. Then the rate variables are shifted 1 hour backwards to fit the WATCH Forcing Datasets (Weedon, Gomes, et al. 2011). Once this is done, sequential elevation and monthly bias correction is applied to the regridded data following the methods by Weedon, Balsamo, et al. 2014; Weedon, Gomes, et al. 2011 for all variables except specific humidity. Since the near-surface specific humidity and relative humidity are not directly available in the ERA5 from the CDS (indirectly from dewpoint temperature), a slightly different method is used.

The relative humidity is found by calculating the ERA5 vapour pressure and saturation vapour pressure, e and e_{sat} , at the surface using methods by Buck 1981. The relative humidity is found using $100 \cdot e/e_{sat}$. The downward shortwave radiation is adjusted at a monthly timescale using the Climate Research Unit (CRU) cloud cover and the local linear correlation between monthly average (aggregated) ERA5 cloud cover and downward shortwave radiation (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020). The snowfall and rainfall fluxes in WFDE5 include gauge catch corrections using either CRU TS4.03 or GPCCv2018 precipitation totals. Since the GPCCv2018 database includes 3-4 times as many precipitation stations as CRU, it is being used in this study. Since errors related to the precipitation phase can arise locally where there are large elevation differences between ERA5 grids and CRU grids, an additional correction is done using maximum and minimum temperatures (Cucchi, Weedon, Amici, Bellouin, Lange, Müller Schmied, Hersbach, and Buontempo 2020).

4.2 Topographical Data

To produce elevation maps and to get elevations for the forcing data, a Digital Elevation Model (DEM) was used. The SRTM 1 Arc-Second Global elevation data from the NASA's Shuttle Radar Tomography Mission (NASA-SRTM) provides a worldwide coverage of void filled data filled with a resolution of 1 arc-second (~ 30 meters) (EROS 2017). The Digital Elevation Model (DEM) data can be downloaded from the coverage map in the [USGS Earth Explorer](#). Each GeoTIFF file covers 1 degree tiles, and is about 25 MB each. 36 tiles are downloaded from USGS Earth Explorer and merged into one DEM in QGIS (see [Appendix D](#) for more detail).

[Table 4.3](#) shows the DEM data that was downloaded for this study.

Table 4.3: SRTM 1 Arc-Second Global

Projection	Geographic
Horizontal datum	WGS84
Vertical Datum	EGM96 (Earth Gravitational Mode)
Vertical Units	Meters
Spatial resolution	1 Arc-second (~ 30 m)
Raster size	1 degree tiles
Tiles	N25-31 E82-88
C-band wavelength	5.6 cm

The DEM files are merged into one DEM file in QGIS. The DEM file are then opened in Python using the [Rasterio package](#) which reads and writes GeoTIFF files, and provides a Python API based upon Numpy N-dimensional arrays and GeoJSON. The coordinate points from the forcing data were then sampled in the DEM file to get the elevations needed for the Shyft input data.

4.3 Land Cover Data Sets

The Land Cover Classification System (LCCS) Land Cover Map Fine Resolution V2.3 Global data set from the GlobCover Portal provides global composites and land cover maps (Bontemps et al. 2011). The GlobCover products have been processed by ESA and by the Université Catholique de Louvain using input observations from the 300 m MERIS sensor on board on the ENVISAT satellite mission. The land cover maps covers December 2004 - June 2006 and January - December 2009. The surface reflectance mosaic products are projected in a Platé-Carré projection (WGS84 ellipsoid) with a $1/360^\circ$ pixel resolution. The land cover classes as defined by a set of classifiers.

4.4 Catchment shapefile

A shapefile for the Budhi Gandaki catchment is used to make maps. This shapefile contains the catchment outline, catchment id and the catchment area. The file is the same the data set used in Bhattarai 2020.

4.5 River network data

River network data is used for the catchment map seen in [Figure 3.2](#). The river network data is a vectorized line network of all the global rivers that have a catchment area of at least 10 km² of a river flow of more than 0.1 m³/s ([Lehner and Grill 2013](#)). HydroRIVERS is a free and open-source database for scientific, educational and commercial use. The downloaded zip data contains a documentation file (.pdf) and a shapefile that can be visualized in QGIS. Version 1.0 of [HydroRIVERS](#) is downloaded in this study downloaded.

4.6 Cell data and TINs

The regular grid cell data produced by [Bhattarai 2020](#), and the TINs for the Budhi Gandaki catchment are made using the [Rasputin](#) software described in [subsection 4.12.2](#). The regular grid cell data is in a netCDF format and contains 9120 cells covering the entire Narayani river basin. Only the cells within the Budhi Gandaki catchment are used in this study (catchment ID 10). To select only the relevant cells the [Xarray](#) package in Python was used to read the netcdf data and then select the cells using the "where(catchment_id == 10)" method on the data set.

[Table 4.8](#) shows the description of the regular grid cell data. The cells have three dimensions: x [m], y [m] and z [m]. The data variables include land cover type fractions [-] (such as forest, lake and glacier), area [m²], catchment id [-], CRS [-], aspect [-] and slope [-]. The coordinate reference system is WGS84 / UTM Zone 45N (EPSG:32645), which is used in areas between 84° E and 90° N on the northern hemisphere between equator and 84°N (Bangladesh, Bhutan, China, India, Kazakhstan, Mongolia, Nepal and the Russian Federation) ([Maptiler 2023](#)).

Table 4.4: Description of the regular cell data

Description	Unit
Dimensions	Cells: 9120 (only 1077 cells are used)
Coordinates	X, Y, Z [m]
Data variables	Forest fraction [-], Lake fraction [-], Glacier fraction [-] Area [m ²], Catchment id [-], CRS [-], Aspect [-] and Slope [-]
Grid mapping name	Universal transverse mercator
Coordinate reference system	WGS84/ UTM Zone 45N (EPSG:32645)

4.7 Country polygons

GeoJSON polygons of countries are downloaded from datahub.io. These polygons are used in figures where country boundaries are illustrated. The data is in the GeoJSON format and is licensed under the Open Data Commons Public Domain Dedication and License (PDDL) v1.0.

4.8 Glacier polygons

Glacier polygons are downloaded from the [NASA Earth Data Explorer](https://earthdata.nasa.gov/earth-data-explorer) in a shapefile format (Li et al. 2021). The glacier polygons are globally glacier polygons based on the Ralph Glacier Inventory v6.0

4.9 Data used for evaluation (target data)

4.9.1 Observed river discharge

Daily stream flow data from the period 2000-2015 is obtained from The Department of Hydrology and Meteorology, DHM (DHM 2022). The flow gauging station is located at Arguhat bazaar (485 m.a.s.l.) at 28.043611° N and 84.816389° E. It has a drainage area of 4270 km^2 (DHM 2022). The station has been operating since 28 November 1963.

4.9.2 NASA MODIS Snow Cover Products

Snow-cover data from the MODerate Resolution Imaging Spectroradiometer (MODIS) is used for evaluation of model performance and for the snow-cover map seen in [chapter 3](#). The MODIS instrument onboard the Terra (2000-present) and Aqua (2002-present) satellites provides multispectral imagery in the visible shortwave infrared (VSWIR) wavelength spectrum (Aalstad et al. 2020). The optical sensors measure the reflected shortwave radiation from the upper 5-10 cm in the snowpack in multiple bands of the VSWIR spectrum. These bands can be combined to give a spectral signature of the snow so that snow-covered area, albedo, snow grain size and impurity concentration can be measured (Aalstad et al. 2020). MODIS has a daily revisit period with a ground sampling distance (GSD) of 500 meters (Aalstad et al. 2020). The NASA MODIS snow cover products compare favourably with other products in terms of quality and resolution (D. K. Hall and Riggs 2007).

The data set contains daily, gridded, snow-cover and albedo from radiance data for the period 2000-02-24 to 2023-02-17 acquired by MODIS. The snow-cover is found using the Normalized Difference Snow Index (NDSI) and a series of screens designed to alleviate errors and flag uncertain snow-cover detections (NSIDC 2023). Only data from the Terra satellite is being used (MOD10A1). The Terra MODIS snow products have been validated under both ideal and non-ideal conditions (Aalstad et al. 2020; D. K. Hall and Riggs 2007). The NASA snow-cover product is distributed by the [National Snow and Ice Data Center \(NSIDC\)](#) (D. Hall et al. 2006; D. K. Hall, Riggs G., et al. 2015). [Table 4.5](#) summarizes the key features of the MODIS snow-cover dataset.

The fSCA and fSCD used in this study, is retrieved from MODIS using methods described in (Aalstad et al. 2020). The pre-processed data set contains fractional snow-covered area for each pixel in the domain with a daily resolution, as well as statistics such as mean fSCA and snow-cover duration (fSCD). The data set includes the water

Table 4.5: MODIS dataset description

Description	Value
Name	MOD10A1
Parameters	Snow-covered area and albedo
Platform	Terra
Sensor	MODIS
Data format	HDF-EOS2
Temporal coverage	Daily
Spatial resolution	500 m x 500 m
Tile name	h25v06
Projection	Sinusoidal
Version	v6.1
Spatial coverage	N: 90, S: -90, E: 180, W:-180

years 2000-2014 (2000-10-01 to 2015-09-30). The fSCA is calculated from the MODIS snow product, and the values range from 0 to 1. The fSCD is the sum of daily fSCA within a given water year. A fSCD equal to 365 indicates that the pixel is snow-covered for the entire water year (i.e. a glacier), while a fSCD equal to 0 indicates that a pixel is never snow-covered. The projection of the pre-processed and native MODIS data is sinusoidal.

4.10 Summary of forcing and validation data

Table 4.6 summarizes the forcing data and validation data being used in this study. Both the input data and the observed discharge data has an hourly resolution. The snow-cover data from MODIS has a daily resolution.

Table 4.6: Forcing data and validation data

Input	Units	Source	Temporal resolution
<i>Input for model simulations in Shyft</i>			
Precipitation	mm h ⁻¹	WFDE5	Hourly
Temperature	°C	WFDE5	Hourly
Wind speed	m s ⁻¹	WFDE5	Hourly
Relative humidity	-	WFDE5	Hourly
Global radiation	W m ⁻²	WFDE5	Hourly
<i>Data used for validation</i>			
Discharge	m ³ s ⁻¹	DHM	Hourly
Snow-cover	per pixel	MODIS	Daily

4.11 Pre-processing of the data

The pre-processing of the data is done in Python and QGIS. In this study the forcing data is first inspected and then made available for Shyft in a netCDF format following the same structure as the demo data that can be found at the [Shyft GitLab repository](#). The raw data can also be read by Shyft setting up a repository with instructions on how to read the data. In this study, this was not done, but rather the raw data was pre-processed and made available in a netCDF format.

4.11.1 Spatial data

Spatial data is processed using [QGIS](#). QGIS is a free and open-source Geographic Information System (QGIS 2023). The QGIS software allows the user to create, edit and visualize geospatial information on Windows, Linux, macOS and mobile devices. In this study, QGIS is mostly used to visualize data and to make maps. A more detailed description on how spatial data is processed in QGIS is given in [Appendix D](#).

4.11.2 WFDE5: Data inspection and selection

After the WFDE5 was downloaded, the data is read using the [Xarray](#) package in Python that supports netCDF files (Xarray 2023). Xarray is an open-source Python package that is built upon the widely adopted Numpy. Xarray makes working with labelled multi-dimensional arrays in Python simple and efficient.

The data was inspected to check for missing values. This was done by plotting a binary map of the entire High Mountain Asia (HMA) region where missing values (0) were showing in a different color than the locations that had values (1). This inspection revealed that the data set was complete for the period 1990-2019. Next a subset of the entire data set (HMA) was selected for Budhi Gandaki only. To figure out the maximum and minimum longitudes and latitudes to use when selecting from the entire data set, the WFDE5 netCDF-file for temperature (year 2000) was plotted in QGIS in combination with the shapefile of the Budhi Gandaki catchment. The maximum and minimum coordinates were selected in such a way that all of the catchment was covered by the forcing data. To obtain this, the coordinates were extended by outside the catchment boundary as well. [Figure 4.1](#) shows the selected grid cells from the WFDE5 data set for the Budhi Gandaki catchment. The red dots in the figure are the midpoints of the WFDE5 grid cells.

The inspection in QGIS results in the following coordinates (World Geodetic System 1984 (WGS84), used in GPS) for the Budhi Gandaki catchment:

Table 4.7: Coordinates used for Budhi Gandaki

	South-West	South-East	North-East	North-West
East [°], North [°]	84.25, 27.75	85.75, 27.75	85.75, 29.25	84.25, 29.25

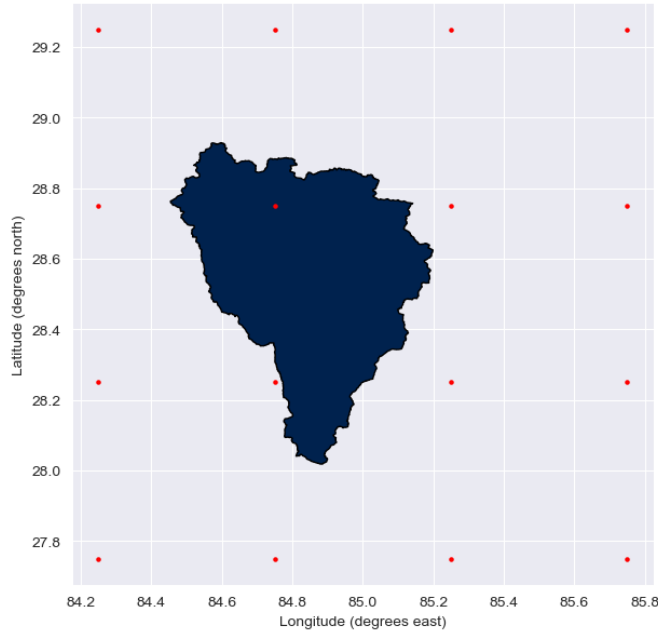


Figure 4.1: Selected WFDE5 grid cells for the Budhi Gandaki catchment. The midpoint of the grid cells are colored in red.

4.11.3 Conversion of units

Shyft requires the units of the forcing data to be converted into standard input units. Temperature is converted from Kelvin (K) to Celsius degrees ($^{\circ}\text{C}$) by simply subtracting 273.15 from the absolute temperature. Similarly, the surface pressure was converted from Pa to hPa by multiplying with 0.01. Precipitation is converted to mm/hr by summing the snowfall flux ($\text{kg m}^{-2} \text{s}^{-1}$) and the rainfall flux ($\text{g m}^{-2} \text{s}^{-1}$), and then multiplying the sum with 3600 (s h^{-1}). Since 1 kg of rain over 1 m^2 is equivalent to 1 mm, the precipitation unit then becomes mm/hr.

The relative humidity is calculated using the MetPy package in Python (May et al. 2022). The relative humidity function in this package calculates relative humidity from total atmospheric pressure (hPa), air temperature ($^{\circ}\text{C}$) and specific humidity (kg kg^{-1}). The formula is based upon (Wallace 1977) and (Salby 1996).

$$RH = \frac{q}{(1 - q)w_s} \quad (4.1)$$

where RH is relative humidity as unit less ratio, q is the specific humidity and w_s is the saturation mixing ratio. The saturation mixing ratio is the ratio between the density of water vapor and the density of dry air (Kasahara and Kanamitsu 2003).

The units of wind (m s^{-1}) and incoming shortwave radiation (W m^{-2}) are not converted.

4.11.4 Obtaining elevation for the cells

The DEM mentioned in [section 4.2](#) is used to obtain the elevation for each grid cell. The raw DEM products arrive as a set of GeoTIFF files, which first has to be merged into one DEM file for the region. This is done in QGIS using the Raster merge tool. The merged DEM file is then opened in Python using the open-source [Rasterio](#) package in Python. Rasterio can use GeoTIFF and other formats to organize and store gridded raster data (Rasterio [2023](#)).

The elevations at specific locations are found by sampling elevations from the DEM at specific points. The locations (longitude [°E] and latitude [°N]) are the mid-points of the 16 WFDE5 grid cells. These points are stored as a [Shapely](#) geometry object called "Point" in a Geopandas Data Frame (GDF) (GeoPandas [2022](#)). [Geopandas](#) is an open-source project that extends the datatypes used by [Pandas](#) (Pandas [2023](#)) to allow spatial operations on geometric types. The elevations at each point are obtained by providing a tupled list of the points to the "sample" method contained in the Rasterio object (containing the DEM data). The elevations are stored in the same Geopandas Data Frame as the coordinate points.

4.11.5 Coordinate conversion

The region domain in Shyft needs UTM coordinates (x [m], y [m]), but the WFDE5 data is referenced in a different coordinate system. The WFDE5 data contains EPSG:4326 - WGS84 coordinates (lon [°], lat [°]). To convert the coordinates to UTM coordinates, the [Pyproj](#) package is used (Whitaker [2023](#)). The conversion is done providing the source CRS (WGS84 - EPSG:4326) and the target CRS (WGS84 / 45N UTM - EPSG:32645) to a transformer object. The transformer object is then used to transform the coordinates from WGS84 to 45N UTM. The resulting coordinates are values for x and y .

4.11.6 Creating netCDF files for Shyft

After completing all the aforementioned steps, it is necessary to store the data in a file format that can be read by Shyft. The forcing data is stored in a netCDF format using the [Xarray](#) package in Python. A description of this data set is given in [Appendix E](#).

4.12 Catchment Discretization Technique

Five different terrain representations are being used in Shyft. One of these is a regular grid and the remaining four are what is called Triangular Irregular Networks (TINs). This section will first explain how topography can be represented in a simulation using discretization techniques. Then regular grids and Triangular Irregular Networks are described more in detail. Lastly, a summary of the five terrain representations being used is provided. Statistics and figures of the terrain representations is presented in [Appendix F](#).

Catchment discretization techniques are used to divide a continuous land surface the catchment into sub-units (Bhattarai, Silantyeva, et al. [2020](#)). These sub-units allow for more accurate representation of complex land features such as slopes or land cover. The amount of sub-units will affect the computational efficiency of the modelling. A commonly used discretization technique is to divide the catchment into a regular grid of cells of fixed sizes. This method is a relatively simple method that makes it easy to calculate spatially averaged properties. It can also allow for more computational efficient

calculations in flat homogeneous terrain where the grid resolution can be coarser without losing accuracy. However, a regular grid structure can oversimplify complex terrain resulting in less accurate results (Bhattarai, Silantyeva, et al. 2020). In complex terrain other discretization techniques such as the triangular irregular network (TIN) method or dividing into elevation zones (hypsograph) can be used to more accurately represent the terrain features of the catchment. In this study both regular grid and triangular irregular network were used as discretization techniques.

Figure 4.2 illustrates three different terrain representations: a) Triangular Irregular Network (TIN), b) regular grid cells and c) elevation zones. A Triangular Irregular Network (Figure 4.2a) consists of multiple triangles. Triangles with smaller area represent steep terrain and triangles with largest area represent flatter area. The area of each grid cell in a regular grid (Figure 4.2b) is the same. The contour lines in the elevation zones figure are curves where elevation has a constant value. The elevation gradient is perpendicular to each curve. Shorter distance between curves represent steep terrain.

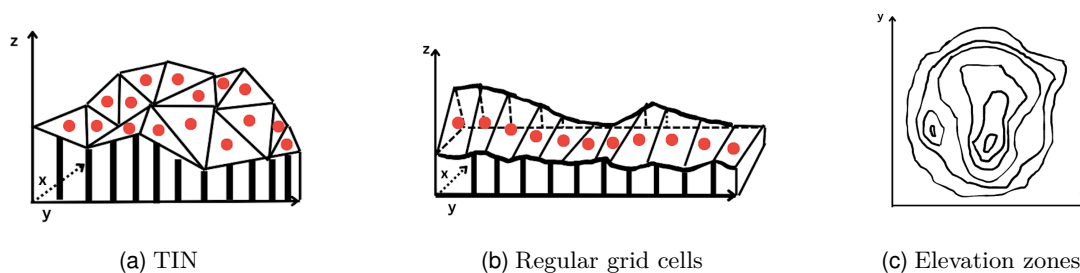


Figure 4.2: Examples of terrain representations: a) Triangular Irregular Network (TIN), b) regular grid cells and c) elevation zones. The red dots show the midpoints of the triangles in the TIN and the regular grid. Figures are made in [Canva](#)

4.12.1 Regular grid method

A regular grid typically consists of two or more finite sets of parallel lines evenly spaced at particular angles to each other in a plane, or the points where these lines intersect (Bern et al. 1999; Uznanski 2023). Orthogonal grids, characterized by two sets of lines perpendicular to each other (e.g., square grid), and isometric grids, with three sets of lines at 60-degree angles to each other (e.g., triangular grid), are the most common types (Bern et al. 1999; Uznanski 2023). The term regular grid cell in this study is referring to a square grid cell containing certain key physical characteristics of the terrain it is meant to represent. This include land cover type fractions, aspect, slope and area of the grid cell.

The regular grid cells used in this study are the same as used in Bhattarai 2020, and is the same as used in (Bhattarai 2020). The regular cells are created in QGIS using the vector creation tools. The forest fractions, glacier fractions, reservoir fractions and the lake fractions are obtained using the land cover data described in section 4.3. The slope and aspect is retrieved from a Digital Elevation Model (DEM). The area of each grid cell is calculated in QGIS using the calculator in attributes table. The regular grid cells have a resolution of $2 \text{ km} \times 2 \text{ km}$ (Silantyeva et al. 2023). Some of the cells at the

edges have an area of less than 4 km² due to the polygon clipping process described in (Burkhart et al. 2021).

Figure 4.3 shows the 1077 regular grid cells used in this study. The midpoints of each grid cell in the data set are plotted on a grid using the WGS84 - 45N UTM coordinate system (EPSG:32645). Note that the grid in this figure is not the same as the one being used to obtain the grid cells, but rather the midpoints of each grid cell. Each cell contains spatial information which is summarized in Figure 4.3.

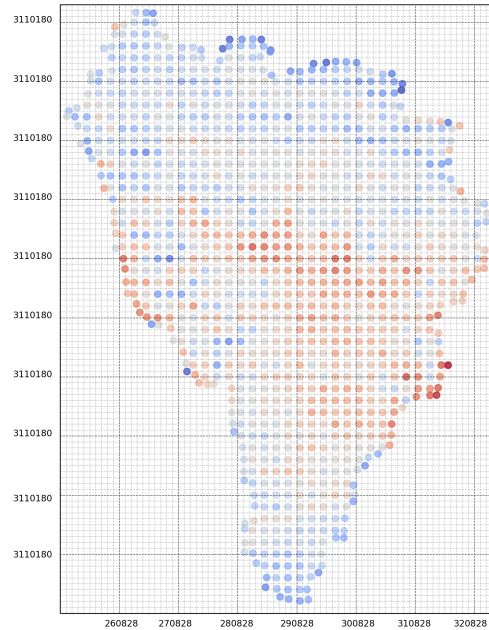


Figure 4.3: The figure shows the mid-points in the regular grid (1077 cells). The colors represent the slope value of each cell. The coordinates are in UTM (EPSG:32645). The figure here shows a resolution of 1 km². In the Shyft simulations, each grid cell has a resolution of 2 × 2 km².

Table 4.8 shows the variables, dimensions and data types contained in the provided netCDF file. This data is used in Shyft to set up a region model that is used to run a Shyft simulation.

Table 4.8: Regular cell data set. The data set contains 9 variables representing the orography of a grid cell. There are 1077 cells with a float64 datatype.

Variable	Dimension	Data type
Forest fraction [-]	Cell: 1077	<i>float64</i>
Reservoir fraction [-]	Cell: 1077	<i>float64</i>
Lake fraction [-]	Cell: 1077	<i>float64</i>
Glacier fraction [-]	Cell: 1077	<i>float64</i>
Area [m ²]	Cell: 1077	<i>float64</i>
Catchment id [-]	Cell: 1077	<i>float64</i>
Aspect [°]	Cell: 1077	<i>float64</i>
Slope [°]	Cell: 1077	<i>float64</i>
CRS [EPSG:32645]	Cell: 1077, crs_dim: 0	<i>float64</i>

4.12.2 Triangular irregular network (TIN) method

Structured meshes, commonly referred to as rasters and grids, offer a method of discretizing landscapes. This approach involves dividing the landscape into cells of uniform size, providing a structured framework for organizing and analyzing spatial data. (Marsh, Pomeroy, et al. 2020). Nevertheless, rasters come with notable limitations, the most significant of which is their fixed spatial resolution that remains constant throughout the entire catchment (Tucker et al. 2001). Consequently, this can lead to significant computational inefficiencies as the topography may be overrepresented in the raster data, resulting in redundant calculations and unnecessary processing. This issue arises from the need to use small raster cells or elements to accurately capture the spatial variability in regions with high topographic or subsurface variations, such as vegetation or soil properties. As a consequence, areas with limited spatial variability are overrepresented in the raster data.

There are several ways of reducing the dimensionality of data, while also retaining meaningful properties of the original data. Some of these techniques include clustering (Joel Fiddes et al. 2019), multi-resolution approach (Baldo and Steven A. Margulis 2017), and unstructured triangular meshes (Marsh, Spiteri, et al. 2018). The triangular irregular network (TIN) method is a method of discretizing a continuous surface into a network of non-overlapping triangles (Bhattacharai, Silantyeva, et al. 2020). Each triangle is defined by three points, known as vertices, with their associated coordinates. These points are typically chosen from a set of points on a land surface or from high resolution Digital Elevation Models (DEMs). The vertices of each triangle define its shape and size. The geometry of the TINs can then be used to calculate various properties on the surface such as slope, aspect and curvature. Additionally, the distribution of the triangles can be used to identify areas of the catchment that are likely to contribute more to runoff, or areas that are more prone to erosion or sediment transport. In order to capture the spatial variability, areas with considerable topographic variations often require a higher density of small triangles, while regions with relatively homogeneous topography can have fewer large triangles (Marsh, Pomeroy, et al. 2020).

TINs can be a more efficient way of representing surface topography as it requires less computational elements than regular raster based solutions (Silantyeva et al. 2023). Physically based models have also been shown to directly benefit from taking terrain structures into account. However, the application of TINs for operational hydrology purposes has been limited (Silantyeva et al. 2023). Furthermore, the complexity of implementation TINs must also be taken into account. Bhattacharai, Silantyeva, et al. 2020 show promising results for the use of TINs in complex terrain, but that there is a limit to how many cells will improve the model performance. Despite the computational advantages of TINs, there is still a need for a conversion to unstructured meshes. This conversion results in increased uncertainty due to aggregation of the landscape into control volumes (Marsh, Pomeroy, et al. 2020).

The TINs are created using the **Rasputin** software (Silantyeva et al. 2023). The Rasputin software can convert a point set of coordinates to a TIN using a Digital Elevation Model (DEM) and an outline of the catchment in a wkt-file (Bhattacharai, Silantyeva, et al. 2020) using Delaunay triangulation methods (Silantyeva et al. 2023). This is done by converting a DEM into simplified triangulated meshes. The triangulation routine in Rasputin is based on the **CGAL Delaunay method** (Silantyeva et al. 2023; Yvinec 2023). The 2D Delaunay triangulation in Rasputin is non constrained, meaning that the Delaunay triangulation is purely based on the position of a given set of vertices disregarding how they ought to be connected by edges (Silantyeva et al. 2023). The land

types are designated to each TIN by determining the position of the middle point in the land cover data set. In this study, Rasputin uses the [GlobCov 2009](#) data set with a 300 m resolution for the land types. Only one land type is assigned to each triangle, which differs from the regular grid cells that have fractional land cover types (Silantyeva et al. 2023). In addition to creating the TINs, the software also calculates physical parameters for each TIN facet (such as slope, aspect and area) from the geometry (Bhattarai, Silantyeva, et al. 2020). The TIN generated from Rasputin is in a [XDMF format](#) and can be visualized using various visualization tools or in a [H5 format](#) that can be parsed and visualised using Python. In this study, the H5 format will be used to parse the TINs for Shyft and to visualise the TINs. The Rasputin software is freely available under [GNU GPL v3.0 license](#).

Figure 4.4 shows two TINs with different resolutions. The left figure shows a TIN with low resolution (911 cells) and figure to the right shows a TIN with higher resolution (4681 cells). In addition to these two TIN-meshes, a TIN-mesh with 1071 cells (similar amount of cells as the regular grid) and one with 1382 cells are being used. Greater sized triangles in Figure 4.4 represent flatter area. This is as a result of the triangulation process in Rasputin (Silantyeva et al. 2023).

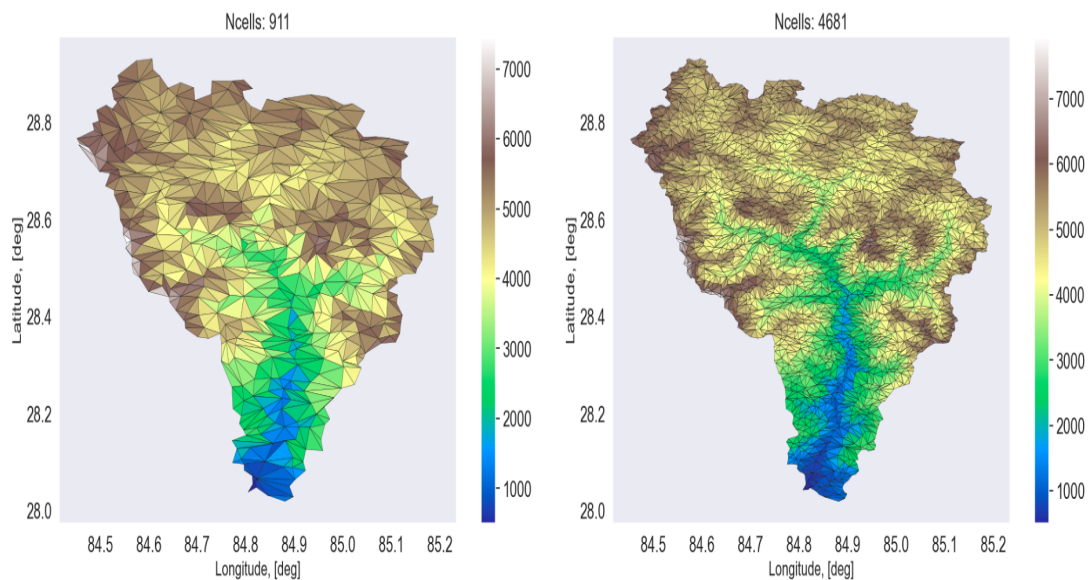


Figure 4.4: TIN with lower resolution (911 cells) and a TIN with higher resolution (4681 cells). The colors represent the elevation at each triangle. Larger triangle area represent flatter terrain.

4.12.3 Summary of the discretization techniques

In this study, five different terrain representations are being used. The first is a regular grid which contains 1077 cells, and which will be referred to as Reg-1077. The four remaining terrain representations are TINs with different resolutions (higher resolution has more cells). These TIN-meshes will be referred to as TIN-911, TIN-1071, TIN-1382 and TIN-4681. The numbers refer to the amount of cells in the TIN. Statistics and figures for these terrain representations are given in [Appendix F](#). [Table 4.9](#) gives a summary of the regular grid cells and TINs being used.

Table 4.9: Topography representations being used.

Name	Type	Number of cells
Reg-1077	Regular grid	1077
TIN-911	TIN	911
TIN-1071	TIN	1071
TIN-1382	TIN	1382
TIN-4681	TIN	4681

4.13 Data management and FAIR principles

This study aims to manage data according to the FAIR principles, meaning it has to be findable, accessible, interoperable and reusable (Wilkinson et al. 2016).

The WFDE5 data, MODIS data and the DEM data are described by rich metadata with a unique identifier of the data they describe. The data sources used here are also open and free to use including documentation on how to download the data. The WFDE5 data and the preprocessed MODIS data is also interoperable using a Network Common Data Form (NetCDF) format. The shapefile format is a geospatial vector data format for geographic information system (GIS) software (ESRI 2023b). The coordinate data assumes a Cartesian coordinate system, using (X, Y) or (Easting, Northing). Each file usually has an attribute that describe it such as name.

All the processing of data is done using open-source software such as [Python](#) and [QGIS](#). Both softwares are open-source, user friendly and easy to learn (Python 2023; QGIS 2023). The code for processing the data and doing simulations in Shyft is documented at the [GitHub](#) repository for this study. The raw and processed data is made available in netcdf format and is made available at the [Zenodo repository](#). Zenodo, an OpenAIRE project managed by the European Organization for Nuclear Research (CERN), provides a digital library for storing data. It follows the principles of being open-source, open-access, and open data, allowing researchers to openly share and access the stored information. It also makes the data being citable with a Digital Object Identifier (DOI). The study also has Reasearch Object at [RoHub](#) (Q. J. Skavang et al. 2022). The Reasearch Object (RO) which will function as a framework for packaging all the research products into FAIR objects (RoHub 2023). A RO is a multidimensional digital object that includes essential information about experiments and investigations to facilitate reusability, reproducibility and better understanding. The RO can aggregate an arbitrary number of heterogeneous resources, internal or external (linked by reference), such as the data used or results produced, methods that are used, and people that

are involved. The research object can include any number of metadata associated to the resources (or the RO itself), enabling the understanding and interpretation of the scientific work.

In this study a majority of the data is stored in a netCDF format. NetCDF, or Network Common Data Form, is a self-describing, portable, scalable, appendable, sharable and archivable (Rew et al. 1989). The global metadata stored within the netCDF files contain information about when the file was created, name of the institution and/or model used to generate the file, links to peer-reviewed papers and technical documentation describing the climate model and software used to generate the file. Moreover, the netCDF format contain metadata about variable dimensions (time, latitude, longitude and height) and variable metadata (e.g. units, averaging period (if relevant) and additional descriptive data).

During the study, the data is being stored on unique location at [OneDrive UiO](#). OneDrive UiO is a cloud based storage platform and is a standard storage platform for all PCs at UiO with the Intune operating system. OneDrive UiO provides easy access from Windows and Mac, including mobile phones, at uio-my.sharepoint.com. The OneDrive UiO is also backed up every day. The data is also archived on Zenodo.

Chapter 5

Methods

5.1 Hydrological Modeling Framework

Shyft is an open–source cross–platform hydrological toolbox built to provide a computation framework for spatially distributed hydrologic models (Burkhart et al. 2021). The platform is developed by Statkraft AS, the largest generator of renewable energy in Europe, in cooperation with the research community at the University of Oslo. The Shyft software is developed for operational, regional-scale hydro power inflow forecasting, and it aims to i) provide a flexible forecasting toolbox for operational environments, ii) facilitate computationally efficient calculations of hydrologic response at the regional scale, iii) enable the use of multiple hypotheses to quantify forecast uncertainties, iv) allow for multiple model forcing configurations, and v) promote expeditious implementations of the findings from research into operational modeling.

One of the strengths of Shyft is that multiple models may be built through the creation of hydrologic algorithms from a library of well known routines or by the construction of new routines (Burkhart et al. 2021). The user of Shyft can get access to all of the components of the framework via Python through an application programming interface (API), while still having high computational performance as the algorithms are implemented in C++. This API set-up enables rapid use of different model configurations and selection of an optimal forecast model.

Since the Shyft toolbox is built for operational, region-scale hydro power inflow forecasting, some of its key design principles vary from what may be prioritized in a pure research environment (Burkhart et al. 2021). Shyft has well-documented open source code and follows the latest code standards, due to the strict requirements regarding software security, testing and code quality in an operational environment. One of the key concepts in the Shyft philosophy is that "data should live at the source". This implies that users of Shyft are encouraged to write their own repositories that connect directly to the source data. This is done to ensure that the data is always up-to-date and that the data is not stored in multiple places. Another key design principle of Shyft facilitates flexible entry points for novel algorithms, while also testing these in parallel with operational runs. This also allows for flexible and exploratory research.

In this study, all Shyft simulations and calibrations are done using the Keilhau server at the University of Oslo. This server has Shyft version v5.x installed. The version of the Shyft package in Python that is used is version 4.26.24. Keilhau is an experimental, virtual machine using Fedora 36 (Linux) as operating system. The machine has 8 cores (section of Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50 GHz) and a 32GB RAM.

5.1.1 Model stacks: PTGSK and PTSTK

A model stack in Shyft is a sequence of hydrologic methods (Burkhart et al. 2021). Each method ("routine") included in a model stack represents a distinct hydrologic process. When combined, these methods create a complete rainfall-runoff model. The model stacks in Shyft are designed for hydro power production environments in Nordic countries, with focus on snow and melt processes. Since Shyft allows for multiple model configurations, the user is able to test multiple hypotheses.

Figure 5.1 provide an overview on input data can be processed through interpolation, and how distinct routines can be combined into model stacks. The first step is to define which input data to use. In this study, this include hourly temperature [$^{\circ}\text{C}$], precipitation [mm/hr], global radiation [W/m^2], relative humidity [-] and wind speed [m/s]. Next, the user defines which interpolation method(s) to use. If the data is downscaled, the user may choose to skip this step. There is also an option to translate the shortwave radiation to inclined surfaces. In this study, the Bayesian Temperature Kriging (BTK) for temperature and Inverse Distance Weighting (IDW) for precipitation, radiation, wind speed and relative humidity are being used. The third step is to specify routines for calculating potential evapotranspiration (PET), actual evapotranspiration (AET), snow response, glacier melt, runoff and soil moisture (only for the HBV model). The potential evapotranspiration can be calculated using the Priestley-Taylor (PT) method (Priestley and Taylor 1972) or the Penman-Monteith (PM) method (Monteith 1965). The actual evapotranspiration can then be calculated using either linear or exponential scaling. The various snow algorithms in Shyft include the HBV snow routine (Bergström 1976; Berström 1992), Gamma Snow routine (Statkraft 2023) or the Snow Tiles routine (Statkraft 2023). The routing can be either calculated from the cells to the river or through a river network (Statkraft 2023). The runoff response routines in Shyft include the HBVtanks (Statkraft 2023) and the Kirchner routine (Kirchner 2009).

The model stacks in Shyft v5.x include the PTGSK, PTSSK, PTHSK, and the HBV (lumped model) which are described in Burkhart et al. 2021, and the PTSTK, RPTGSK, RPMGSK model stacks that have been added later (Statkraft 2023). In this study the Priestly-Taylor Gamma-Snow Kirchner (PTGSK) and the Priestley-Taylor Snow Tiles Kirchner (PTSTK) model stacks are chosen. The Gamma Snow routine, which is a simplified surface energy balance model, has been previously tested in the region. On the other hand, the Snow Tiles routine is a temperature-index based model that has not yet been tested in the region. These two model stacks differ in the way they calculate snow water equivalent (SWE), snow-covered area (SCA) and melt from snow, and the difference is highlighted in bold text. Both of these model stacks use the same methods for potential evapotranspiration (Priestly-Taylor), glacier melt (temperature-index model) and runoff (Kirchner). The PTGSK and PTSTK model stacks are described in more detail in the sections below.

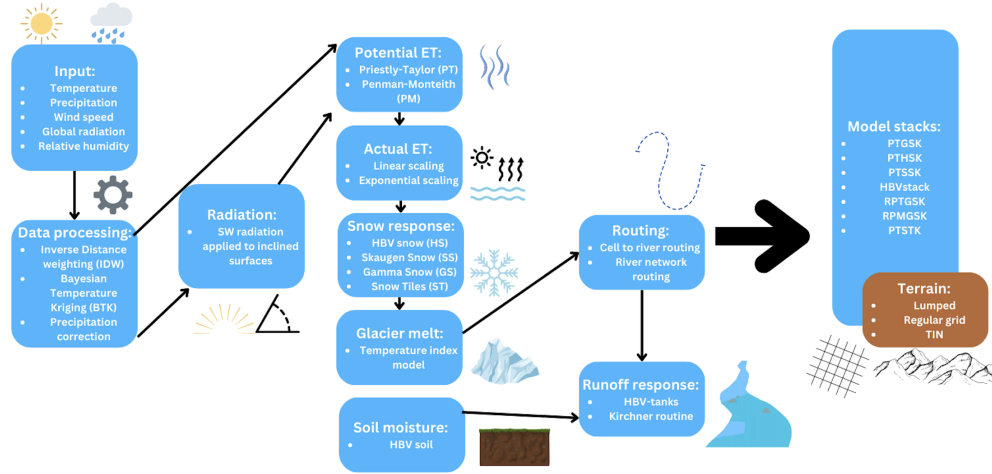


Figure 5.1: Multiple configurations in Shyft as of April 2023 (Shyft version v5.x (4.26.24 in Python)). The figure is an adaptation of material provided by Postdoctoral Fellow Olga Silantyeva at the University of Oslo, and is made using [Canva](#)

The Priestly-Taylor Gamma Snow Kirchner (PTGSK) model stack

Figure 5.2 shows how the Priestly-Taylor Gamma Snow Kirchner (PTGSK) model stack is built up through several distinct routines. These routines will be described below.

Systematic bias in the input precipitation [mm/hr] can be corrected for using a a *precipitation scaling factor*, $scale_factor$ [-]. The precipitation factor is used for both models and uses the following equation:

$$P_{corr} = P_{forcing} * Scale_factor \quad (5.1)$$

where P_{corr} is the adjusted precipitation [mm/hr], $P_{forcing}$ is the precipitation [mm/hr] given to Shyft prior to correction, and $Scale_factor$ is the precipitation scaling factor [-].

The PTGSK model stack utilizes the *Priestley-Taylor (PT)* method to calculate potential evapotranspiration, and a simple scaling approach is used to estimate actual evapotranspiration (Burkhart et al. 2021; Priestley and Taylor 1972). The PT model is based on evaporation from a saturated surface and involves a parameter that corresponds to a multiple of the equilibrium evapotranspiration (Nakayama and Nakamura 1982; Priestley and Taylor 1972). Temperature, radiation, and relative humidity data are used to calculate the potential evapotranspiration (Burkhart et al. 2021). The Priestly-Taylor constant, α [-], and the land albedo, $albedo$ are used to calculate the potential evapotranspiration [mm/s] (Statkraft 2023). The Priestley-Taylor constant has a typical range of $1.26 \pm 0.XX$, and is usually not used in calibration, but rather treated as a constant (Statkraft 2023). The land albedo is used to estimate the net radiation, R_n [W/m^2] (see Equation B.1). The equation for the potential evapotranspiration using the PT method is shown in Equation 5.2.

$$e_{pot} = \alpha \frac{\Delta}{\Delta + \gamma} (R_n - G) \quad (5.2)$$

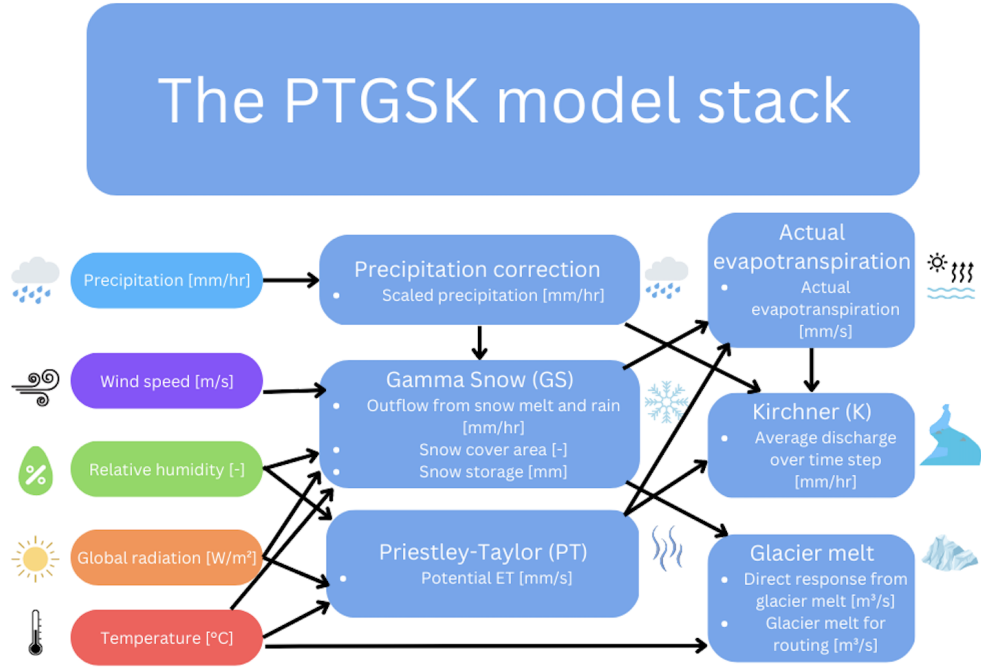


Figure 5.2: The PTGSK model stack in Shyft as of April 2023 on keilhau.uio.no (Shyft version v5.x (4.26.24 in Python)). The input data are shown to the left in the figure. The light blue color represents the routines in the PTGSK model stack. The figure is made using [Canva](https://www.canva.com/).

where α is the Priestly-Taylor constant (1.26), Δ is the slope of the saturation water vapor pressure curve at 2 meter air temperature, γ is the psychrometric constant (0.67 hPa K^{-1}), R_n is the net radiation, and G is the soil heat flux based on available observations of ground heat flux found to be a near-constant fraction of 0.12 of R_n . The net radiation in Shyft is estimated using the temperature, T [K], incoming shortwave radiation [W/m^2], relative humidity [-], and the vapor pressure [kPa] (Statkraft 2023). The calculation of the net radiation is given by Equation B.1 found in Appendix B.

The *actual evapotranspiration*, e_a [mm/hr], is calculated using Equation 5.3.

$$e_a = e_{pt}(1 - e^{\frac{3}{ae_{sf}}q}) \quad (5.3)$$

where e_{pt} is the potential evapotranspiration [mm/hr], the ae_{sf} is the actual evapotranspiration scaling factor [-], and q is the water level.

The *Gamma Snow (GS)* routine is used to calculate snow accumulation and melt-adjusted runoff using time series data for precipitation, wind speed [m/s], temperature [°C], relative humidity [-] and radiation [W/m^2] (Burkhart et al. 2021). This routine is an energy-balance based snow routine that uses a gamma function to represent the snow distribution at sub-cell level (Kolberg, Rue, et al. 2006). The Gamma snow routine adopts an energy balance (Essery 2009; Hegdahl et al. 2016), which is given by Equation 5.4.

$$\Delta E = S \cdot (1 - \alpha) + L_{in} - L_{out} + H_{SE} + H_L + E_G \quad (5.4)$$

where ΔE is the net energy flux at the surface available for snow melting [W/m^2], S is the incoming shortwave radiation [W/m^2], L_{in} is incoming longwave radiation [W/m^2], L_{out} is outgoing longwave radiation [W/m^2], H_{SE} and H_L are sensible [W/m^2] and latent heat [W/m^2] fluxes and E_G is the subsurface energy flux [W/m^2].

The GS routine has 18 parameters what can be set by the user of Shyft. These parameters are listed in [Table 5.6](#). A more detailed description of the Gamma Snow routine is provided in [Appendix C](#).

The simple temperature index approach by Hock [2003](#) is used to account for *glacier melt*, m_g . The temperature index approach uses temperature, T [$^{\circ}\text{C}$] and a degree time step factor for ice melt, dtf [$\text{mm}/\text{day}/^{\circ}\text{C}$], to calculate glacier melt via [Equation 5.5](#) (Statkraft [2023](#)).

$$m_g = dtf \cdot T (a_g - a_{sc}) c \quad (5.5)$$

for $T > 0$, a_g is the glacier area [m^2], a_{sc} is the snow-covered area [m^2] and c is a conversion factor used to calculate to mm, equal to $0.001\text{m}/86400\text{s}$. The PTGSK model stack is the only one in Shyft that uses an *energy-balance approach* for snow accumulation and melt processes (Burkhart et al. [2021](#)).

The *Kirchner's* approach is used to calculate the catchment response (Burkhart et al. [2021](#)). The Kirchner's approach to modelling water fluxes in catchments treats them as first-order nonlinear dynamical systems (Kirchner [2009](#)). The approach yields a first-order nonlinear differential equation that can be used directly to simulate streamflow hydrograph from precipitation and evapotranspiration time series. The Kirchner response uses potential evapotranspiration and precipitation over the reservoirs as input.

The foundation of the Kirchner response algorithm lies in the logarithmic transformation of the ordinary differential equation (ODE), which describes the change in discharge over time as a result of precipitation, evapotranspiration and discharge (equation 19 in Kirchner [2009](#)).

$$\frac{d(\log Q)}{dt} = \frac{1}{Q} \frac{dQ}{dt} = \frac{g(Q)}{Q} (P - E - Q) = g(Q) \left(\frac{P - E}{Q} - 1 \right) \quad (5.6)$$

where Q [$\text{m}^3/\text{unit of time}$] is the discharge, P [$\text{mm}/\text{unit of time}$] is the precipitation, E is the evapotranspiration [$\text{mm}/\text{unit of time}$], $g(Q)$ is a non-linear function called the "sensitivity-function" (expresses the sensitivity of the discharge to changes in the storage, dS/dt) and t is time.

The non-linear function, $g(Q)$, is defined by [Equation 5.7](#) (equation 6 and 9 in [Kirchner 2009](#)).

$$g(Q) = \frac{dQ}{dS} = \frac{dQ/dt}{dS/dt} = \frac{dQ/dt}{P - E - Q} \simeq c_1 + c_2 \log Q + c_3 (\log Q)^2 \quad (5.7)$$

where c_1 , c_2 and c_3 are the catchment specific outlet parameters, called the Kirchner parameters ([Bhattarai, Silantyeva, et al. 2020](#)). In Shyft the non-linear function is approximated by a quadratic function defined by [Equation 5.7](#) using the three parameters c_1 [-], c_2 [-] and c_3 [-].

PTSTK

The methods, input data, and responses from the routines in the PTSTK model are shown in [Figure 5.3](#). The input variables are shown to the left in the figure, the methods are noted in bold text and the responses in bullet points. The primary equations behind the Snow Tiles routine will be described below.

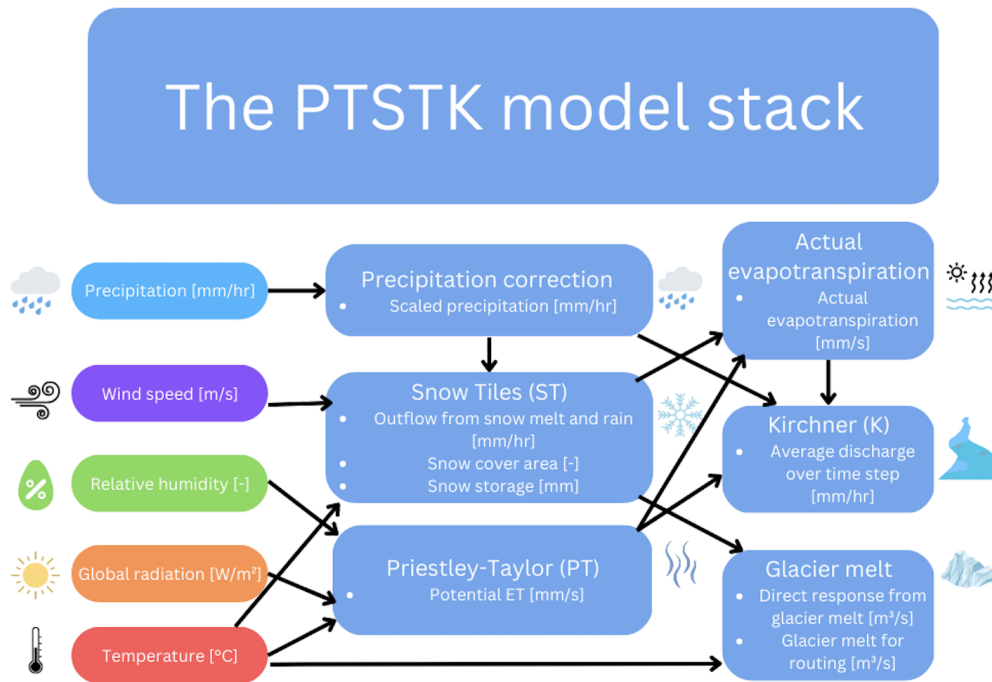


Figure 5.3: The PTSTK model stack in Shyft as of April 2023 on keilhau.uio.no (Shyft version v5.x (4.26.24 in Python)). The input data are shown to the left in the figure. The light blue color represents the routines in the PTSTK model stack. The figure is made using [Canva](#).

The PTSTK model stack is similar to the PTGSK model stack since it also utilizes the same algorithms for precipitation correction, potential evapotranspiration (Priestley-Taylor), glacier melt, actual evapotranspiration (linear scaling) and runoff (Kirchner). However, the PTSTK model has a different approach for calculating snow accumulation and melt processes. The PTSTK model stack uses a "Snow Tiles" approach to simulate the distribution of snow within each cell and to calculate the outflow from snow melt and rain, and the snow-covered area (SCA) (Statkraft 2023). This Snow Tiles routine has been tested for Nordic conditions, but never for the Himalayan region.

The Snow Tiles routine differs from the Gamma Snow routine since it uses a temperature-index approach (conceptual model), rather than a simplified surface energy balance model. A great advantage of the simplified surface energy balance model used in the Gamma Snow routine is that it is more physically based. The Gamma Snow considers energy fluxes between the snowpack and the surrounding environment, which provides a more comprehensive understanding of complex interactions that influence snowmelt rates. However, the Gamma Snow routine is more complex having more parameters. The Gamma Snow routine has 18 parameters, while the Snow Tiles routine has only 6. The complexity of the Gamma Snow routine can make calibration more time consuming and

increase uncertainty stemming from equifinality. Equifinality refers to the concept that different model structures and/or sets of parameters can produce similar hydrological outcomes, even though the pathways or mechanisms leading to those outcomes may vary. In other words, different combinations of model structures and parameters can result in comparable hydrological simulations, highlighting that there can be multiple valid ways to represent and reproduce a given hydrological behavior or outcome (Khatami et al. 2019). Temperature-index models have been shown to have high model efficiency compared with energy budget models, and found to sufficient when net radiation, not turbulent heat flux, dominate snowmelt dynamics (Debele et al. 2010).

In the Snow Tiles routine, the user has the ability to define the size of each tile, represented as a fraction of the total cell area. The amount of snowfall within each tile is determined by a multiplier, which follows a Gamma distribution. It is important for the mean of these multipliers to be set to 1 in order to ensure mass conservation.

The shape parameter of the Gamma distribution is considered a calibration parameter, allowing for adjustments to the snowpack simulation. The development of the snowpack within each tile is simulated using a standard temperature index approach. Additionally, potential melt is assumed to be evenly distributed over the tiles, ensuring a uniform distribution of melting processes throughout the modeled area. By default, 10 percentile snow tiles are used, each representing 10 % of the cell area.

Environmental processes often exhibit threshold-type behavior, where certain fluxes like snowmelt, recharge, and quick flow have distinct onset points (Kavetski and Kuczera 2007). However, incorporating these thresholds into hydrological models can introduce discontinuities into the objective functions used for model calibration, making parameter estimation more challenging. In contrast, smoothing the model concerning its parameters and inputs can help eliminating spurious secondary optima in objective functions, and enhance the macro scale characteristics. Consequently, model calibration and sensitivity analysis become simplified as the complexity of the objective functions is reduced. The Snow Tiles routine therefore introduces a smoothing strategy when splitting the precipitation into solid and liquid phase in order to remove microscale discontinuities and spurious secondary optima in the objective function (Statkraft 2023).

The Snow Tiles routine splits the precipitation into a solid and liquid phase. The precipitation phase is given as a fraction of snow in precipitation, s_{frac} [-] as a function of temperature T [°C] via the threshold temperature, tx [°C], determining the snow/rain, and a smoothing controlled by $m_{frac} = 0.5$. The s_{frac} is calculated using Equation 5.8 (Statkraft 2023).

$$s_{frac} = \frac{1}{1 + e^{\frac{T-tx}{m_{frac}}}} \quad (5.8)$$

The potential melt, m [mm], is calculated using the temperature index model with the threshold temperature for melt onset, ts [°C], and the degree-day melt factor, cx [mm/day]. The smoothing controlled by $m_m = 0.5$. The potential melt is calculated using Equation 5.9

$$m = cx * t_{days} * m_m * (T_m + \log(1 + e^{-T_m})) \quad (5.9)$$

where

$$T_m = \frac{T - ts}{m_m} \quad (5.10)$$

The refreezing of liquid water content, f [mm], is given via the refreezing coefficient, cfr [-], by a linear relationship with T when the temperature is below the melting onset threshold, ts , and the smoothing controlled by m_{rf} (0.5) is applied to the surrounding ts . The refreezing liquid water content is calculated using [Equation 5.11](#)

$$f = cfr * cx * t_{days} * m_{rf} * (T_{rf} + \log(1 + e^{-T_{rf}})) \quad (5.11)$$

where

$$T_{rf} = \frac{T - ts}{m_{rf}} \quad (5.12)$$

The max holding capacity for liquid water given as a fraction of ice in the snowpack, $lwmax$ [-], is used as an additional parameter.

5.2 Model setup

In this section, the different steps used to run simulations and calibrations in Shyft will be described. These steps include configuring the simulation, running the first simulation, calibrate the model and run a new simulation with the optimized parameters from the calibration. After this is done the data is extracted and exported in a netCDF file that can be accessed outside the Keilhau server. This workflow is an iterative process where the same procedure is done for each simulation case and is illustrated in [Figure 5.4](#). Each of the steps shown in [Figure 5.4](#) is described below.

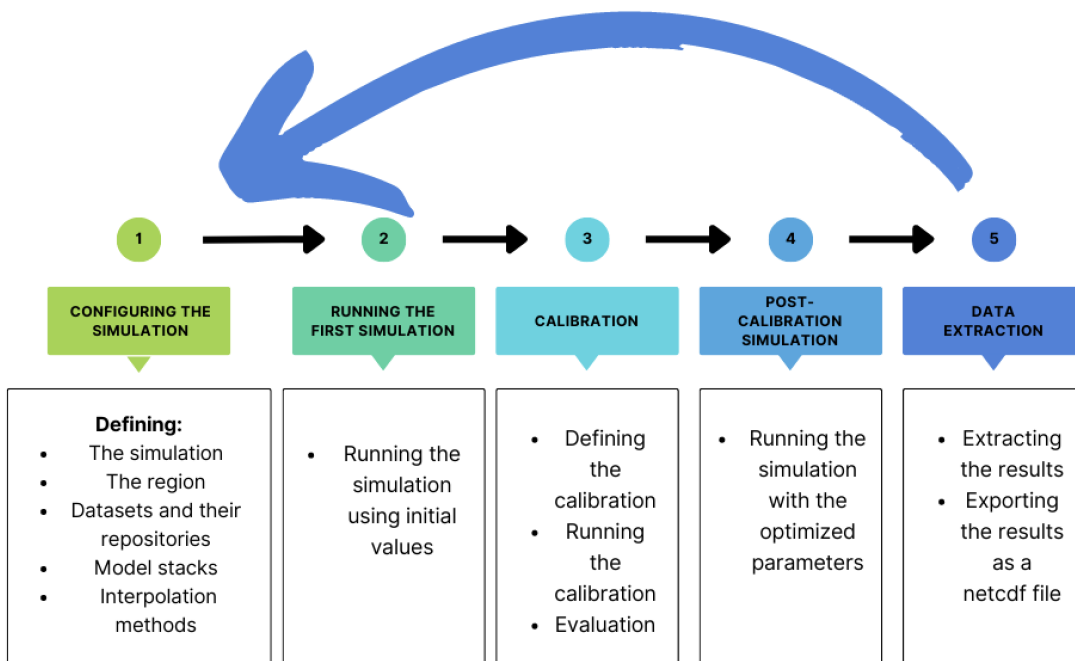


Figure 5.4: Illustration of the workflow in Shyft through an iterative process of configuration, simulation, calibration, and post-calibration run and data extraction. The figure is made using [Canva](#).

5.2.1 Configuration of a Shyft simulation

The configurations for each Shyft simulation are defined in YAML files provided in [Appendix G](#) and the [GitHub repository](#). These YAML files are used by the `YAMLSimConfig` class in the Shyft orchestration to set up a Shyft simulation. In this study, the configuration files are contained within a folder structure that is used to organize the simulations. There is one folder for each simulation and calibration case investigated in this study. The use of YAML-configuration files is considered a high-level approach, while it is also possible to use the Shyft API directly to set up simulations (Statkraft [2023](#)).

5.2.2 Defining the simulation

In this study, 10 different simulations are conducted using 2 distinct model stacks (PTGSK and PTGSK) and 5 simulation units (regular cells and TINs). Each simulation begins on January 1st, 2000, which is the start date of the observed discharge series, and runs for a period of 140256 time steps with a time resolution of 3600 seconds (1 hour). The simulation period thus ranges from 1 January 2000 to 31 December 2015. The time steps used for simulation and calibration are different since the model is being calibrated on a daily basis from 1 January 2000 to 31 December 2004. The region model id is unique for each simulation case and the initial state of the simulation is provided by the "budhi_gandaki_model.yaml" file (see [Appendix G](#)). The model utilized in each simulation is defined in this file and can be found at the "shyft.hydrology" repository (Statkraft [2023](#)). The observed discharge is used as a reference time series, and this is stored in a netCDF format. The parameters used for setting up the simulations are listed in [Table 5.1](#).

Table 5.1: Parameters used to set up a Shyft simulation.

Parameter	Value
Start datetime	2000-01-01T00:00:00
Run time step [s]	3600
Number of steps	140256

5.2.3 Defining the region domain

The region is defined by a domain and a repository to the simulation unit. The region defines the extent of the Shyft simulation and thus how many cells there are. The values for the region domain in this study are shown in [Table 5.2](#). The EPSG stands for European Petroleum Survey Group, and is a public registry of geodetic parameters with standard codes for coordinate systems, datums, spheroids and units (Surveyor [2023](#)). nx and ny are the number of cells in x-direction and y-direction. $step_x$ and $step_y$ are the distances (in meters) in each direction. The $lower_x$ and $lower_y$ defines the lower, western corner of the grid. The cells are located in the center of each grid cell.

The region extent is determined in QGIS by using the "Vector" → "Research tools" → "Create grid". The Grid type is rectangle (polygon), the grid extent is the Budhi Gandaki shapefile, and the horizontal spacing is 1000. The grid CRS is "EPSG:32645 - WGS84/UTM Zone 45N". The grid extent is extended by 500 meters in x and y direction to make sure all of the catchment is covered.

Table 5.2: Parameters used to set up the region domain

Parameter	Value
EPSG code	32645
Number of cells in x-direction (nx)	38
Number of cells in y-direction (ny)	53
Steps in x direction [m]	2000
Steps in y direction [m]	2000
Lower left x	250828
Lower left y	3100180

Figure 5.5 shows a map the region in Shyft with the Budhi Gandaki catchment shown inside. The catchment is discretized into a 38 by 53 grid where each grid cell is 4 km² (2000 m × 2000 m).

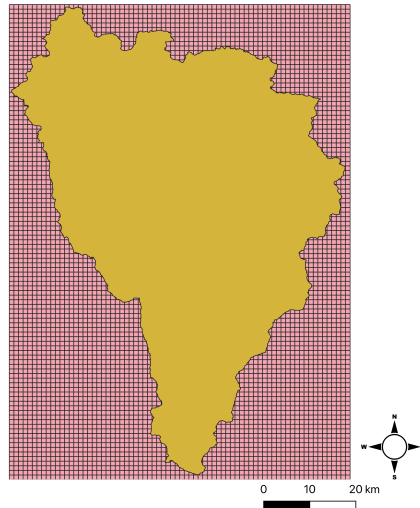


Figure 5.5: Region domain in Shyft. The grid is 75 x 105 cells (7875 cells). Projection is EPSG:32645.

5.2.4 Datasets and set up of repositories

In order for Shyft to read the forcing data correctly, repositories are made for the forcing data and for the TIN cell data. The repository for the forcing data is based upon the "cf_region_model_repository" from "[Shyft.hydrology.repository.netcdf](#)". The only change that needed to be made here was to replace the "data_cs.proj4" with "data_cs.proj" in line 77. This was done because since there had been changes to the Python package Pyproj. The repository is named "wfde5_cf_geo_ts_repository.py", and stored in a new folder called "mst_himalaya" inside the "python3.10/site-packages" folder inside the Keilhau server (see section [section A.3](#)).

The forcing data is defined the "budhi_gandaki_datasets.yaml" file. This file includes the repositories for data handling, the file path and the type of the sources. Here all the datasets are in netcdf format.

5.2.5 Defining the model stacks

The model stack used for each simulation is defined in the "budhi_gandaki_model.yaml" file. This file contains the initial values for the parameters and the repository for the model stack methods. The model configuration file used in this study can be found in [Appendix G](#). This file contains the calibrated values, and not initial values, and therefore only serves as an example.

5.2.6 Defining interpolation methods

The forcing data has much coarser resolution than the simulation domain, and it is therefore preferred that the input data is interpolated to the cells in Shyft before a simulation is run. Shyft is able to interpolate gridded forcing data (or station data) to the Shyft cells using different interpolation routines. In this study, the Bayesian Temperature Kriging (BTK) and Inverse Distance Weighting (IDW), described in [subsubsection A.2.2](#), are used. These interpolation methods are chosen based on previous studies using Shyft (Bhattarai, Burkhart, Tallaksen, et al. [2020](#); Lombr a 2017; Sapkota [2016](#)).

IDW is a deterministic interpolation since it is directly based on the surrounding measured values (ESRI [2023a](#)). A second type of interpolation methods consists of geostatistical methods, such as kriging, which are based on statistical models that include autocorrelation (relationship between the measured points). This means that kriging can provide a measure of the certainty or accuracy of predictions.

The Inverse Distance Weighting (IDW) (Shepard [1968](#)) is the main interpolation method for most variables (Burkhart et al. [2021](#)). The IDW method in Shyft is adopted from Enki method idwtemp by Kolberg and Bruland [2012](#). It is able to transform a set of time-series data from source locations (x,y,z) to another set of time-series data at some target locations (Statkraft [2023](#)). The approach involves using a weighted average of the maximum nearest neighbours that have available data to the target location. It excludes neighbour locations that exceed the maximum allowable distance. The interpolation method does the interpolation for each cell. This method is used for precipitation, global radiation, wind speed and relative humidity. The parameters used for IDW can be seen in [Table 5.3](#). The maximum distance is the maximum distance of temperature sources being used for on interpolation. The distance measure factor is the square distance. It is computed as $1/X^c$, where X is the euclidean distance and c is the distance measure factor (Statkraft [2023](#)). Maximum member is the maximum number of sources used for one interpolation (Statkraft [2023](#)). The scaling factor is used for weighing neighbours.

Table 5.3: Parameters used in the IDW routine

Parameter	Value			
	Precipitation	Radiation	Wind speed	Rel. hum.
Maximum distance	400000.0 [m]	600000.0 [m]	600000.0 [m]	600000.0 [m]
Maximum members	10	10	10	10
Distance measure factor	1	1	1	1
Scale factor	0.93	1.0	1.0	1.0

Kriging assumes that the spatial variation of a continuous climatic variable is too irregular to be modeled by a mathematical function, and its spatial variation could be better predicted by a probabilistic surface (Vicente-Serrano et al. 2003). Kriging involves generating an estimated surface based on a collection of scattered points with corresponding z -values (ESRI 2023a). The technique operates under the assumption that the spatial relationship or correlation between the sample points can be utilized to explain the variation observed in the surface. By considering the distance or direction between the points, this method aims to create a coherent representation of the surface that captures its underlying spatial characteristics.

The Bayesian Temperature Kriging (BTK) method in Shyft is similar to the IDW in the way that it weights the surrounding measured values to derive a prediction for an unmeasured location. The kriging formula is given by (ESRI 2023a).

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (5.13)$$

where $Z(s_i)$ is the measured value at the i th location, λ_i is an unknown weight for the measured value at the i th location, s_0 is the prediction location and N is the number of measured values. The IDW depends only on λ_i , while the BTK also considers the overall spatial arrangement of measured points.

The BTK method includes a method of uncovering dependency rules and then making prediction. To do this it first creates the semivariograms (a measure of the degree of spatial autocorrelation) and covariance functions to estimate the statistical dependence (spatial autocorrelation) values that depend on the model of autocorrelation (ESRI 2023a). After this is done it predicts the unknown values.

The Bayesian Temperature Kriging (BTK) method in Shyft first calculates the elevation difference between two points, p_1 and p_2 , and multiplies the difference with a elevation scale factor, $Z - scale$ (Statkraft 2023). Thereafter, the BTK algorithm builds covariance matrices from sources and destinations. This is done by defining 3D-points, where $p_1 = (x_1, y_1, z_1)$ is the source destination and $p_2 = (x_2, y_2, z_2)$ is the target destination. The distance, d , is defined by Equation 5.14.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + l(z_1 - z_2)^2} \quad (5.14)$$

where l is the height scaling factor.

The covariance of p_1 and p_2 is then calculated by Equation 5.15

$$cov(p_1, p_2) = (s - n) \cdot e^{-\frac{d}{r}} \quad (5.15)$$

where s is the sill, n is the nugget magnitude and r is the range of the semivariogram (Statkraft 2023). The BTK algorithm in Shyft has a default temperature gradient [$^{\circ}\text{C}/\text{m}$]. In this study the temperature gradient is set to -0.6 $^{\circ}\text{C}/100\text{m}$ based on a previous study (Bhattarai, Burkhart, Tallaksen, et al. 2020). The Z -scale factor is used to discriminate neighbours at different elevations than target point (Statkraft 2023). The sill is the value of semivariogram at range and the nug is the nugget magnitude.

Table 5.4 shows the parameters used in this study for the BTK method.

Table 5.4: Parameters used in the BTK routine for temperature

Parameter	Value
Temperature gradient [$^{\circ}\text{C}/100\text{ m}$]	-0.6
Temperature gradient standard deviation	10
Nugget	0.5
Sill [$^{\circ}\text{C}$]	25.0
Z-scale	20.0

5.2.7 Running the first simulation

A Shyft simulation is done through several steps. Once the configuration files are set, the input files are ready and Shyft is installed, a simulation can be run. The first step is to provide the path to the "simulation.yaml" file to the `YAMLSimConfig` class in Python and give it a name. This object is then given to the `ConfigSimulator` class and stored as an object in Python. This object, for example named "simulator", provides a wrapping of the underlying API functionality that Shyft provides and can be used to run simulations (Statkraft 2023). The core functionality of the simulator object is stored in the "simulator.region_model". A simulation needs a time axis and an initial state. This is stored inside the simulator, and a simulation is then simply done by calling the "simulator.run()" method. The simulator uses a time axis in a Shyft time series format and forward the simulation over this time axis. The time axis (length of the simulation) defines the simulation period and the step sizes.

5.2.8 Calibration

The methods for calibrating each model configuration are given below. Calibration is done automatically using an optimization method named Min BOBYQA. A more detailed description of the optimization methods in Shyft is provided in section A.4. The YAML-files used for calibration runs are given in Appendix G.

To calibrate a model in Shyft, one must specify an optimization method, its parameters, the time period, an objective function, and a target. In this study this is defined in a YAML file (Appendix G). The `YAMLSimConfig` class in Shyft (see Statkraft 2023) utilizes this file by providing it the path to the calibration configuration file and a name. The `ConfigCalibrator` class can then be used to run a calibration.

The Min BOBYQA optimization is used for this study (see section A.4). The Min BOBYQA solves the optimization problem using the Bounded Optimization by Quadratic Approximation method by Powell 2009 specified by model M (Statkraft 2023). The Min BOBYQA method searches for optimal parameters such that a goal function is minimized. The maximum number of evaluations is set to 1500, and the optimization method will stop if it is not converging after n evaluations. The initial trust region radius, `tr_start`, is set to 0.1 and the stopping trust region radius, `tr_stop`, is set to 10^{-5} . The target is set to observed river discharge. Start time is set to 2000-01-01 and run time step is set to 86400 (1 day). The number of steps is set to 1826 (1 January 2000 to 31 December 2004). The Nash-Sutcliffe Efficiency (NSE) is used as goal function, and is specified in the target specification. After several iterations the best parameter set based on the NSE is reached, either due to convergence or when maximum number of iterations are met (Bhattarai, Burkhart, Tallaksen, et al. 2020). The target weight and

the scaling factors (s_{corr} , s_{var} and s_{bias}) used in the objective function, are set to 1.0. A summary of the calibration is given in [Table 5.5](#).

Table 5.5: Optimization method, time period, target and goal function used in calibration.

Description	Input
Optimization method	Min BOBYQA
Maximum number of evaluations	1500
Start date time	2000-01-01T00:00:00
Run time step	86400
Number of time steps	1827
Target	Observed river discharge
Weight	1.0
Goal function	NSE
s_{corr}	1.0
s_{var}	1.0
s_{bias}	1.0

[Table 5.6](#) shows the parameters in the PTGSK model stack, in addition to the upper and lower limits of the parameters. The limits are based upon previous studies listed in the table. Several of the parameters are set as constants since they are found to be less sensitive than the others (Teweldebrhan et al. 2018). Therefore, 11 out of 31 parameters are searched for in the Min BOBYQA optimization method.

Table 5.6: Calibration parameters for the PTGSK model stack

Calibrated parameters				
Name	Description and unit	Lower limit	Upper limit	Source
kirchner.c1	Outlet empirical coefficient 1, [-]	-8.0	0.0	1)
kirchner.c2	Outlet empirical coefficient 2, [-]	-1.0	1.2	1)
kirchner.c3	Outlet empirical coefficient 3, [-]	-0.15	-0.05	1)
ae.ae_scale_factor	Actual evapotranspiration scale factor, [-]	1.0	1.0	1)
gs.tx	Threshold temperature determining precipitation phase, [°C]	-3.0	2.0	1)
gs.wind_scale	Slope in turbulent wind function [m/s]	1.0	6.0	1)
gs.fast_albedo_decay_rate	Albedo decay rate during melt [days]	5.0	15.0	1)
gs.slow_albedo_decay_rate	Albedo decay rate in cold conditions [days]	20.0	40.0	1)
gs.surface_magnitude	Surface layer magnitude [-]	30.0	30.0	1)
p_corr.scale_factor	Precipitation scaling factor, [-]	0.4	2.0	1)
gm.direct_response	Direct response of glacier melt, [-]	0.0	1.0	4)
Prescribed parameters				
Name	Description and unit	Value		
gs.max_water	Maximum liquid water content	0.1		1)
gs.wind_const	Intercept in turbulent wind function	1.0		1)
gs.max_albedo	Maximum albedo value [-]	0.9		1)
gs.min_albedo	Minimum albedo value [-]	0.6		1)
gs.snowfall_reset_depth	Snowfall required to reset albedo [mm]	5.0		2)
gs.snow_cv	Spatial coefficient variation of fresh snowfall [-]	0.4		2)
gs.snow_cv_forest_factor	Additional factor for the effective snow-cover in forest (geo.forest_fraction()*snow_cv_forest_factor) [-]	0.0		2)
gs.snow_cv_altitude_factor	Additional factor for the effective snow-cover due to altitude (altitude [m] *snow_cv_forest_factor [-])	0.0		2)
gs.initial_bare_ground_factor	Bare ground fraction at melt onset [-]	0.04		3)
gs.winter_end_day_of_year	Last day of winter	100.0		3)
gs.calculate_iso_pot_energy	Whether or not to calculate the potential energy flux [bool (0/1)]	0.0		3)
gs.n_winter_days	Number of winter days	100.0		
gs.glacier_albedo	Glacier ice fixed albedo [-]	0.4		2)
gm.dtf	Degree timestep factor for glacier melt, [mm/day/°C]	6.0		3)
pt.albedo	Priestly-Taylor albedo, [-]	0.2		2)
pt.alpha	Priestly-Taylor constant, [-]	1.26		2)
routing.velocity	Routing velocity, [m/s]	0.0		3)
routing.alpha	Routing alpha, [-]	0.9		3)
routing.beta	Routing beta coefficient, [-]	3.0		3)
msp.reservoir_direct_response_fraction	Direct response from reservoir	1.0		4)

1) Bhattarai, Burkhart, Tallaksen, et al. 2020

2) Sapkota 2016

3) Lombrana 2017

4) Statkraft 2023

Similarly to the PTGSK calibrations, the PTSTK calibration configuration includes prescribed parameters based on previous studies. The parameters in PTSTK model stack are also calibrated using the Min BOBYQA method. Some of the parameters are the same for both PTGSK and PTSTK (Kirchner parameters, actual evapotranspiration scaling factor, precipitation correction factor, Priestley-Taylor parameters and the reservoir direct response parameter). The range for the PTSTK are listed in [Table 5.7](#).

Table 5.7: Calibration parameters for the PTSTK model stack

Calibrated parameters				
Name	Description and unit	Lower limit	Upper limit	Source
kirchner.c1	Outlet empirical coefficient 1, [-]	-8.00	0.00	1)
kirchner.c2	Outlet empirical coefficient 2, [-]	-1.00	1.20	1)
kirchner.c3	Outlet empirical coefficient 3, [-]	-0.15	-0.05	1)
st.tx	Threshold temperature determining precipitation phase, [C]	-3.0	2.0	1)
st.cx	Degree-day melt factor, [mm/C/day]	0.0	30.0	4)
st.ts	Threshold temperature for melt onset, [C]	-4.0	4.0	4)
st.lwmax	Max liquid water content given as a fraction of ice in the snowpack, [-]	0.0	1.0	4)
st.cfr	Refreeze coefficient, [-]	0.0	1.0	4)
p_corr.scale_factor	Precipitation scaling factor, [-]	0.4	2.0	1)
gm.direct_response	Direct response of glacier melt, [-]	0.0	1.0	4)
Prescribed parameters				
Name	Description and unit	Value		
ae.ae_scale_factor	Actual evapotranspiration scale factor, [-]	1.0		1)
pt.albedo	Priestly-Taylor albedo, [-]	0.20		2)
pt.alpha	Priestly-Taylor constant, [-]	1.26		2)
st.shape	Shape parameter of the Gamma distribution defining the multiplication factors for snowfall, [-]	2.00		4)
routing.velocity	Routing velocity, [m/s]	0.00		3)
routing.alpha	Routing alpha, [-]	0.90		3)
routing.beta	Routing beta coefficient, [-]	3.00		3)
gm.dtf	Degree timestep factor for glacier melt, [mm/day/C]	6.00		3)
msp.reservoir_direct_response_fraction	Direct response from reservoir	1.0		4)

1) Bhattarai, Burkhart, Tallaksen, et al. [2020](#)

2) Sapkota [2016](#)

3) Lombrana [2017](#)

4) Statkraft [2023](#)

5.2.9 Summary of the simulation and calibration runs

Table 5.8 shows the different configurations that are used in this study. Case 1 and 2 tests two different snow routines for regular cells. Cases 3-10 tests snow routines for TINs of different sizes.

Table 5.8: Multiple simulation and calibration configurations

Configuration	Input data	Input processing	Model stack	Simulation unit	Calibration
1-2:	WFDE5	IDW and BTK	PTGSK/PTSTK	Reg-1077 (regular grid)	Min BOBQYA
3-4:	WFDE5	IDW and BTK	PTGSK/PTSTK	TIN-911 (low resolution TIN)	Min BOBQYA
5-6:	WFDE5	IDW and BTK	PTGSK/PTSTK	TIN-1071	Min BOBQYA
7-8:	WFDE5	IDW and BTK	PTGSK/PTSTK	TIN-1382	Min BOBQYA
9-10:	WFDE5	IDW and BTK	PTGSK/PTSTK	TIN-4681 (high resolution TIN)	Min BOBQYA

5.2.10 Post-calibration run and data extraction

Once the calibration is complete, a new file named "calibrated_model.yaml" will be generated in the directory specified in the "budhi_gandaki_calibration.yaml" file. This file includes the optimized parameters that can be used to conduct a new simulation. To do so, the "budhi_gandaki_model.yaml" file is replaced with the newly generated file in the configuration file for the simulation (see Appendix G). Once the changes are saved these changes and the Python kernel is restarted, the post-calibration simulation can be executed.

After the new simulation has finished running, the results from the simulation are extracted. Catchment-wide timeseries for interpolated forcing data, simulated discharge, SWE and SCA is extracted accessing the *region_model.statistics* class. Furthermore, mean values from each cell in the regular PTGSK and PTSTK, and average PTGSK/PTSTK, are calculated by accessing the time series in the cells (*region_model.cell*) and then looping over each *env_ts* and taking the mean of each time series. This is done so that it is possible to make maps from the Shyft simulations afterwards. The same is done for actual evapotranspiration, *ae*, snow water equivalent, *swe*, snow-cover area, *sca*, glacier melt, *gm*, and average discharge of each cell, *avg_discharge* using the *rc* method instead of the *env_ts*.

Once all the extraction of the relevant data is finished, the data is assembled into a Xarray data set and saved as a netCDF file using the "to_netcdf" method on the data set.

5.2.11 Model evaluation

Once the calibration is done and the simulation has run for a second time using optimized parameters, the results are extracted and stored in a netCDF file. The model performance is evaluated using different evaluation metrics and through plots, such as QQ-plots and time series plot for both observed and simulated discharge.

The evaluation metrics that have been used include the Nash-Sutcliffe Efficiency (NSE), the log-transform NSE, the Kling-Gupta Efficiency (KGE), the Root Mean Square Error (RMSE), Critical Success (CSI). These are described more in detail in the sections below. Before calculating the NSE, log-transform NSE, KGE and the RMSE, the simulated and observed discharge series are split into a calibration period of five years (2000-01-01 to 2004-12-31), and a validation period of 11 years (2005-01-01 to 2015-12-31). The observed discharge has a daily resolution, so the simulated discharge

series are resampled from an hourly to a daily resolution for comparison. Furthermore, only the time points that are present in both the observed and simulated discharge series are used in the calculation of the evaluation metrics. This results in a slightly shorter simulation period as the observed discharge series is missing some time points in the year 2014. All of the evaluation metrics are calculated for each of the 10 simulation cases (see Table 5.8) using a Python package called **Hydroeval** (Hallouin 2021).

Nash-Sutcliffe Efficiency (NSE)

The Nash-Sutcliffe Efficiency (NSE) is one of the model performance criteria used in this study. The NSE is a commonly used metric for assessing the predictive power of hydrological models and ranges from $-\infty$ to 1, with a value of 1 indicating a perfect match between simulated and observed discharge values (Bhattarai, Silantyeva, et al. 2020). The NSE is often used since it normalizes the model performance into an interpretable scale (Knoben et al. 2019). The NSE is also widely used due to its flexibility to be applied to various types of mathematical models. However, the NSE tends to overestimate the model performance during peak stream flow, and underestimate it during low flow periods (Bhattarai, Silantyeva, et al. 2020). It is also sensitive to bias in model predictions (Ritter and Muñoz-Carpena 2013). Therefore, various other indicators have been proposed, such as the log-transform NSE. Even though the use of benchmark series can help mitigating some of the limitations of the NSE, these are not always available for all regions of the world. The NSE provides an intuitive interpretability and has a fundamental meaning at zero. In combination with other indicators and visualization it is therefore found useful in this study. The NSE is calculated using Equation 5.16 (Ritter and Muñoz-Carpena 2013)

$$E_{NSE} = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} = 1 - \left(\frac{RMSE}{SD}\right)^2 \quad (5.16)$$

where O_i and P_i represent the sample (of size N) containing the observations and the model estimates, respectively. \bar{O} is the mean of the observed values, SD represents the standard deviation of the observations and $RMSE$ is the Root Mean Square Error. The log-transform NSE is calculated ?? with logarithmic values instead.

The NSE and the log-transformed NSE values are calculated by splitting up the observed and simulated discharge series into a calibration period of five years (2000-01-01 to 2004-12-31), and a validation period of 11 years (2005-01-01 to 2015-12-31). Once the discharge series are split into a calibration period and a validation period, the simulated discharge series is resampled from an hourly to a daily resolution. Then only the timepoints that are present in both the observed and simulated discharge series are used for the NSE calculation. This results in a slightly shorter simulation period as the observed discharge series is missing some timepoints. The NSE is calculated for each of the 10 simulation cases (see Table 5.8) using the Python package called **Hydroeval** (Hallouin 2021).

Kling-Gupta Efficiency (KGE)

The Kling-Gupta Efficiency (KGE) is another efficiency measure that is used to assess the predictive power of a hydrological model (Bhattarai, Burkhart, Tallaksen, et al. 2020). The KGE is different from the NSE since it is not derived from the Mean Square Error (MSE), but rather by using the Euclidean distance computed using the coordinates of bias, standard deviation and correlation (Clark et al. 2021; Gupta et al. 2009). The

KGE range from $-\infty$ to 1, where 1 is a perfect fit between observed and simulated values (Knoben et al. 2019). A KGE less than 0 indicates that the mean of observations provides better estimates than simulations (Knoben et al. 2019). The NSE and KGE values cannot be directly compared since they do not have a unique relationship. In this study, a positive KGE value is interpreted as an indications of a "good model" performance (using 0 as a threshold). The original KGE has three components: r , α , and β as per Gupta et al. 2009. r is the linear correlation between observations and simulations, α is a measure of the flow variability error and β is the bias term (Knoben et al. 2019). The formulas for these components are given in Equation 5.18, Equation 5.19, and Equation 5.20.

$$E_{KGE} = 1 - \sqrt{[r - 1]^2 + [\alpha - 1]^2 + [\beta - 1]^2} \quad (5.17)$$

$$r = \frac{\text{cov}(e, s)}{\sigma(e) * \sigma(s)} \quad (5.18)$$

$$\alpha = \frac{\sigma(s)}{\sigma(e)} \quad (5.19)$$

$$\beta = \frac{\mu(s)}{\mu(e)} \quad (5.20)$$

where e is the evaluation series, s is (one of) the simulation series, cov is the covariance, σ is the standard deviation, and μ is the arithmetic mean.

Root Mean Square Error (RMSE)

Another indicator frequently used is the Root Mean Square Error (RMSE) (Ritter and Muñoz-Carpena 2013). The RMSE quantifies the error in terms of the units of the variable (Legates and McCabe 1999). It is calculated by Equation 5.21 (Ritter and Muñoz-Carpena 2013).

$$E_{RMSE} = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (5.21)$$

where O_i and P_i represent the sample (of size N) containing the observations and the model estimates, respectively. It ranges from 0 to ∞ , and a value of 0 indicates a perfect match between simulated and observed discharge values.

Critical Success Index (CSI)

The agreement between observed and simulated snow is evaluated using the Critical Success Index (CSI). This method is also called the Threat Score (Schaefer 1990). The CSI score ranges from 0 to 1, with 1 being the best score. This method includes comparing the observed mean fractional snow-covered area fraction (fSCA) from MODIS with the simulated fSCA from each Shyft simulation. The first step in this method is to make two binary series. If the mean fSCA is above or equal to a threshold value, a value of 1 is assigned meaning that the catchment is snow-covered. If the fSCA is below the threshold value it is considered not snow-covered. A threshold value of 0.4 based on work by Bhattarai, Silantjeva, et al. 2020 is used to determine if an event is snow-covered or not.

Once the two binary time series are made, a two-dimensional contingency table is used to calculate the CSI from Figure 1 in Agha Kouchak and Mehran 2013. The contingency table is made by summing the events where the observed and simulated values agree, where they both disagree and where they differ. Table 5.9 shows the contingency table used in this study.

Table 5.9: Contingency table. Hit (H) indicates that both the simulated and observed snow-cover agree. Miss (M) means that snow-cover is detected by MODIS, but is not in simulation. False (F) is false alarm, and means that snow-cover is simulated, but not observed. True Null Event (Q) is when there is no snow-cover in both simulated and observed. The values for H, F, M and Q are summed values.

		<i>Snow-covered?</i>	
		MODIS observations	
		Yes [1]	No [0]
<i>Snow-covered?</i> Simulations	Yes [1]	Hit (H)	False (F)
	No [0]	Miss (M)	True Null Event (Q)

The CSI is calculated using Equation 5.22 (Agha Kouchak and Mehran 2013).

$$CSI = \frac{H}{(H + M + F)} \quad (5.22)$$

Chapter 6

Results

6.1 Spatial and temporal variation in the forcing data

This section includes an analysis on the spatial and temporal variation of the forcing data, including temperature, precipitation, global radiation, relative humidity, and wind speed. The objective of this analysis is to understand the characteristics and patterns of these variables in the High Mountain Asia region and specifically within the Budhi Gandaki catchment. By examining the spatial and temporal variations, the aim is to assess the suitability of the forcing data for hydrological modeling and evaluate its potential impact on the model performance. The analysis of temporal variation can help identify trends, seasonality, and other patterns that might influence the hydrology of the study area. Understanding these temporal dynamics is crucial as changes in climate conditions can have significant implications for hydrological processes. Furthermore, the investigation of spatial variation provides valuable information about the ability of the forcing data to capture spatial patterns and heterogeneity within the study area. By analyzing the seasonal maps of the forcing variables and applying statistical methods, the aim is to assess if the data adequately represent the leeward and windward precipitation patterns, which are important factors affecting the hydrological response in mountainous regions. Additionally, this section includes an examination of two different interpolation methods implemented in Shyft: Bayesian Temperature Kriging (BTK) and Inverse Distance Weighting (IDW).

6.1.1 Analysis of seasonality and trends in the High Mountain Asia region

Figure 6.1 illustrates the seasonal variation of mean temperature ($^{\circ}\text{C}$) in the High Mountain Asia (HMA) region, based on the WFDE5 input data. The plot clearly exhibits pronounced seasonal patterns in temperature across the region. Notably, a distinct temperature gradient is observed in the north-south direction during all seasons. This gradient appears to be most prominent during December, January and February (DJF), March, April and May (MAM), and September, October, November (SON). During DJF, the Tibetan Plateau registers mean temperatures around or below zero degrees, while regions situated south of the Himalayas exhibit positive temperatures. The highest mean temperatures (30 degrees or above) are observed during spring and summer in areas located to the south of the Himalayas, including India, Nepal, Pakistan, Bangladesh, and Myanmar. Conversely, the lowest mean temperatures (below -10 degrees Celsius) are found during winter (DJF) within the Tibetan Plateau.

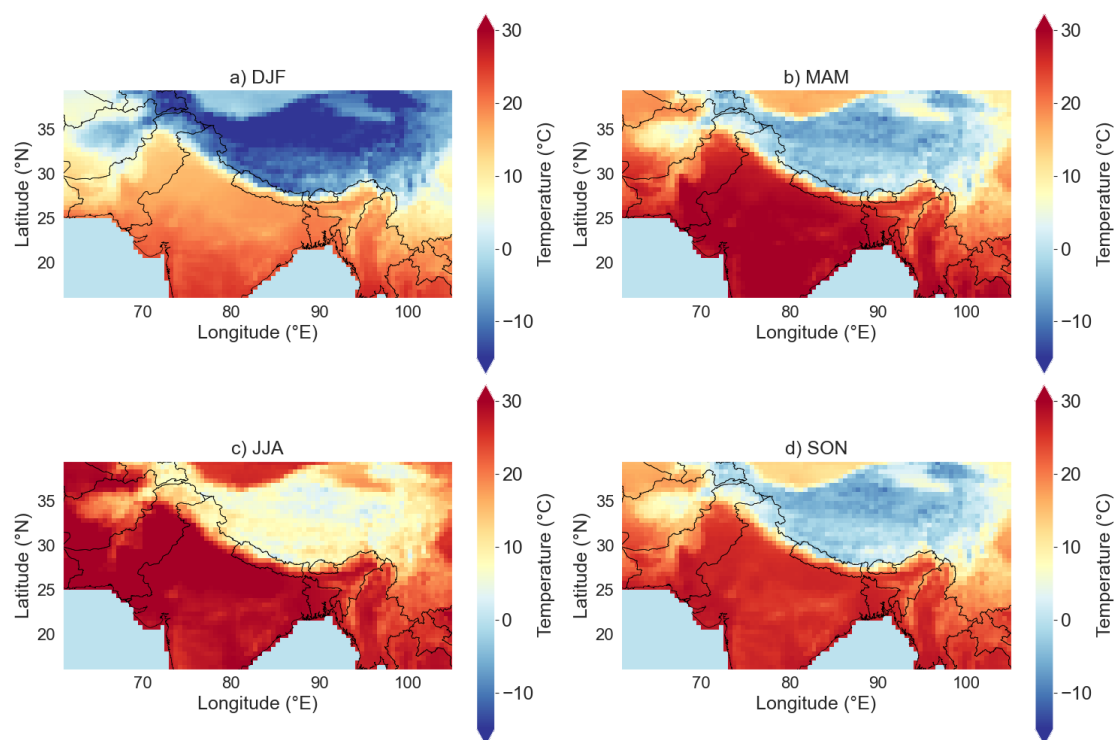


Figure 6.1: Mean seasonal temperature High Mountain Asia (1990-2019) from the WFDE5 data. DJF is December, January and February, MAM is March, April and May, JJA is June, July and August, while SON is September, October and November.

Figure 6.2 shows the mean accumulated seasonal precipitation (mm) in HMA during the period (1990-2019) from the WFDE5 data. The color bar is scaled such that it goes from the lowest mean value to the highest mean seasonal value. This makes comparison between seasons easier. The sub-figure a) shows that there is little precipitation during the winter season (DJF) for the entire region. The areas to the North-East (the Hindu-Kush and Karakoram areas) experience relatively more precipitation during winter than other areas. During spring there is an increase in precipitation in the south eastern part of the region (Bangladesh, North Eastern India and Myanmar). During JJA the precipitation increases considerably in India, Nepal, Bangladesh, Bhutan, Myanmar. The most intense mean seasonal precipitation is found on the Western coast of India and in the areas surrounding the Bay of Bengal (Bangladesh, North Eastern India and Myanmar). The increase in precipitation from spring to summer in the Tibetan Plateau (north of the Himalayas) is much less pronounced than for the areas south of the Himalayas. The accumulated precipitation decrease during fall, where most of the precipitation still occurs south of the Himalayas. The precipitation in the Tibetan Plateau is very during all seasons, but there is higher precipitation in the Eastern part during JJA than for the other seasons.

6.1. Spatial and temporal variation in the forcing data

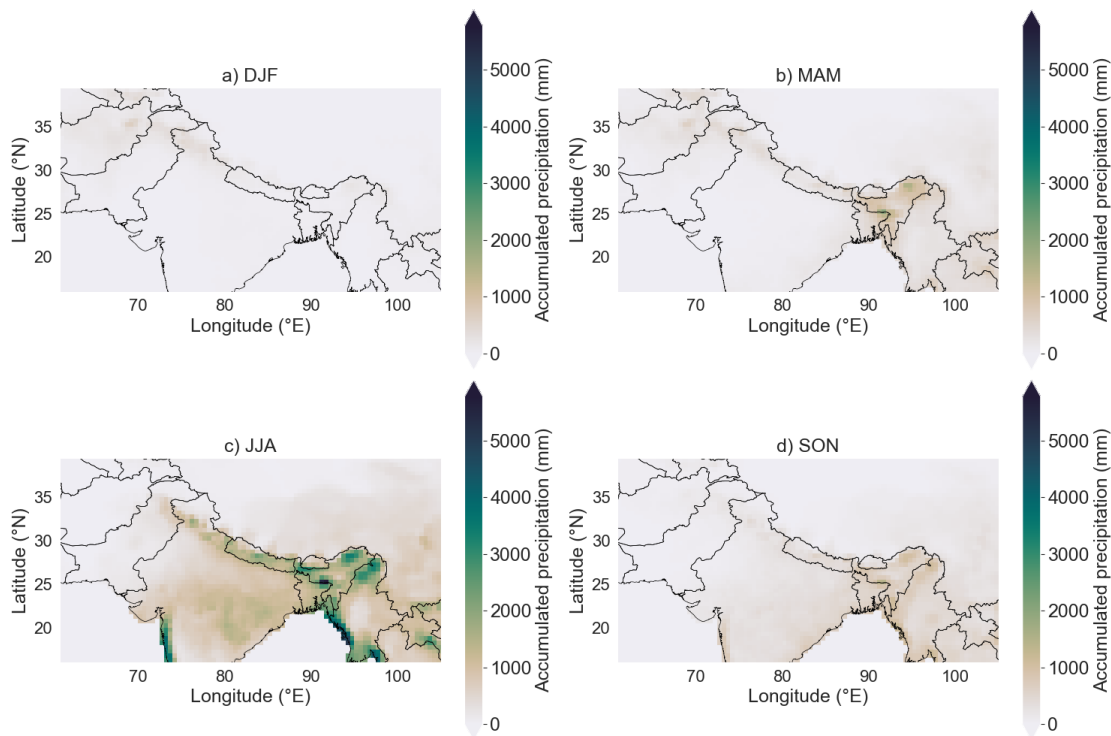


Figure 6.2: Mean seasonal accumulated precipitation in High Mountain Asia (1990-2019) from the WFDE5 data.

Figure 6.3 shows a time series decomposition for a) daily mean temperature [$^{\circ}\text{C}$], b) daily mean precipitation [mm/hr], c) daily mean radiation [W/m] and d) daily mean relative humidity [-] in HMA during the period 1990-2019 from the WFDE5 data. A time series decomposition breaks down the original time series into three components: trend, seasonality and residuals. Trend is a pattern found in the data, and used to describe upward or downward progression for part, or all, of the time series (Statista 2023). Seasonality shows if there is any seasonal variation and residuals (weather) are what is left when trend and seasonality is taken away. The residuals statistical noise which is random and irregular. The original time series from WFDE5 are shown in the top figures, and the components are illustrated below. The figure shows an upward linear trend for the temperature and a downward linear trend for radiation. All variables exhibit seasonal variation (third row in the figure).

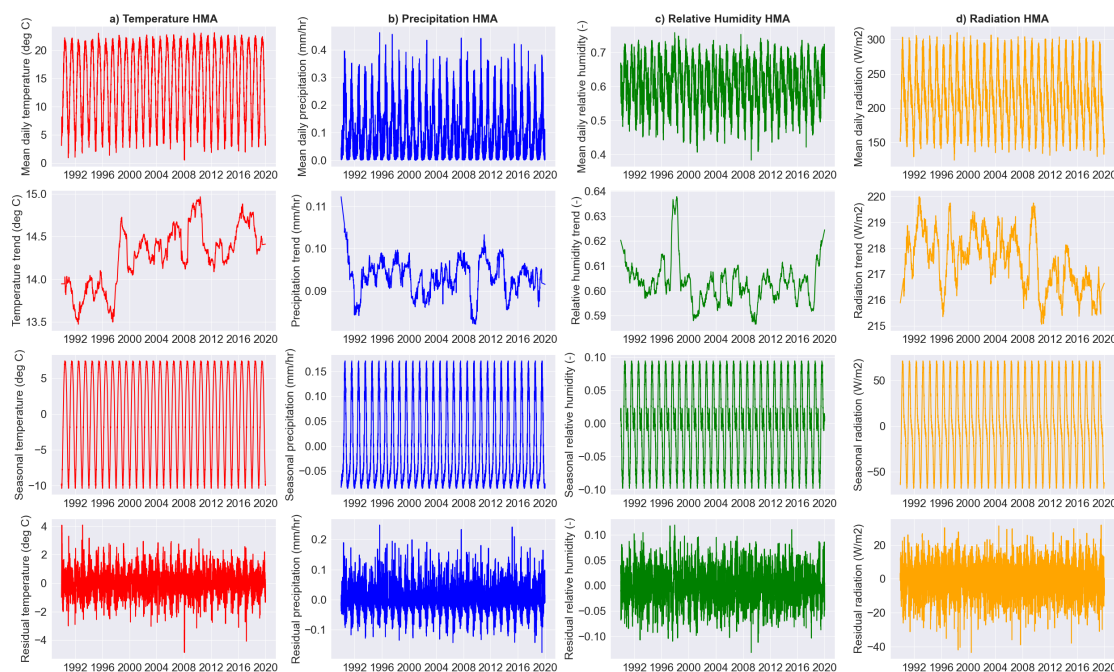


Figure 6.3: Time series decomposition (1990-2019). a) Daily mean temperature in HMA, b) daily mean precipitation, c) relative humidity and d) radiation. Original time series are shown at the top, and then trend, season and residual below respectively.

6.1.2 Temporal and spatial variation in the Budhi Gandaki region

In this section, an analysis of the forcing data used in Shyft is provided. This includes an analysis of temperature [$^{\circ}\text{C}$], precipitation [mm/hr], relative humidity [-], wind speed [m/s], and radiation [W/m^2]. It is important to note that this is an analysis of the forcing data before it is interpolated. Furthermore, an analysis of the observed discharge [m^3/s] measured at Arughat gauging station, which is used as target data.

Table 6.1 provides a summary of the central tendency (mean) and variation (minimum, maximum, and standard deviation) of the WFDE5 data from the period 2000-01-01 to 2015-12-31. The mean annual temperature of 5.00°C suggests a relatively cool climate overall. However, the standard deviation of 12.93°C indicates a wide variation in temperatures throughout the year, with the minimum and maximum values ranging from -36.69°C to 41.74°C . This suggests that the region experiences both extremely cold and hot temperatures, leading to significant seasonal variations. The mean annual precipitation is 0.16 mm/hr . However, the standard deviation of 0.62 mm/hr indicates a considerable variation in precipitation levels. The minimum value of 0.00 mm/hr and the maximum value of 24.20 mm/hr highlight the potential for both dry and wet periods within the region. The mean wind speed of 1.30 m/s suggests generally calm wind conditions in the Budhi Gandaki region. However, the standard deviation of 0.80 m/s and the maximum value of 10.73 m/s indicate the possibility of occasional strong winds. The mean relative humidity is 0.72 . The standard deviation is 0.19 , making the min value of 0.22 is well below 1 standard deviation from the mean. The mean radiation of 223.78 W/m^2 reflects the average amount of solar radiation received in the region. The standard deviation of 301.75 W/m^2 and the maximum value of 1196.41 W/m^2 indicate substantial variability in solar radiation levels.

6.1. Spatial and temporal variation in the forcing data

Table 6.1: Central tendency and variation Shyft input data (WFDE5) covering the Budhi Gandaki region (2000-01-01 to 2015-12-31).

	Temperature [°C]	Precipitation [mm/hr]	Wind speed [m/s]	Relative humidity [-]	Radiation [W/m ²]
Mean	5.00	0.16	1.30	0.72	223.78
Standard deviation	12.93	0.62	0.80	0.19	301.75
Min	-36.69	0.00	0.12	0.22	0.00
Max	41.74	24.20	10.73	1.00	1196.41

Figure 6.4 shows the mean seasonal temperature (°C) from the WFDE5 data in the Budhi Gandaki region for the period 2000-2015. The figure reveals a pronounced temperature gradient from the northeast to the southwest throughout all seasons. Specifically, the months of December, January, and February (DJF) exhibit the coldest temperatures. Conversely, the southern region of the catchment experiences the highest temperatures during summer. Notably, the mean temperatures across the four grid cells encompassing the catchment remain consistently low throughout all seasons, with values below 5 degrees Celsius.

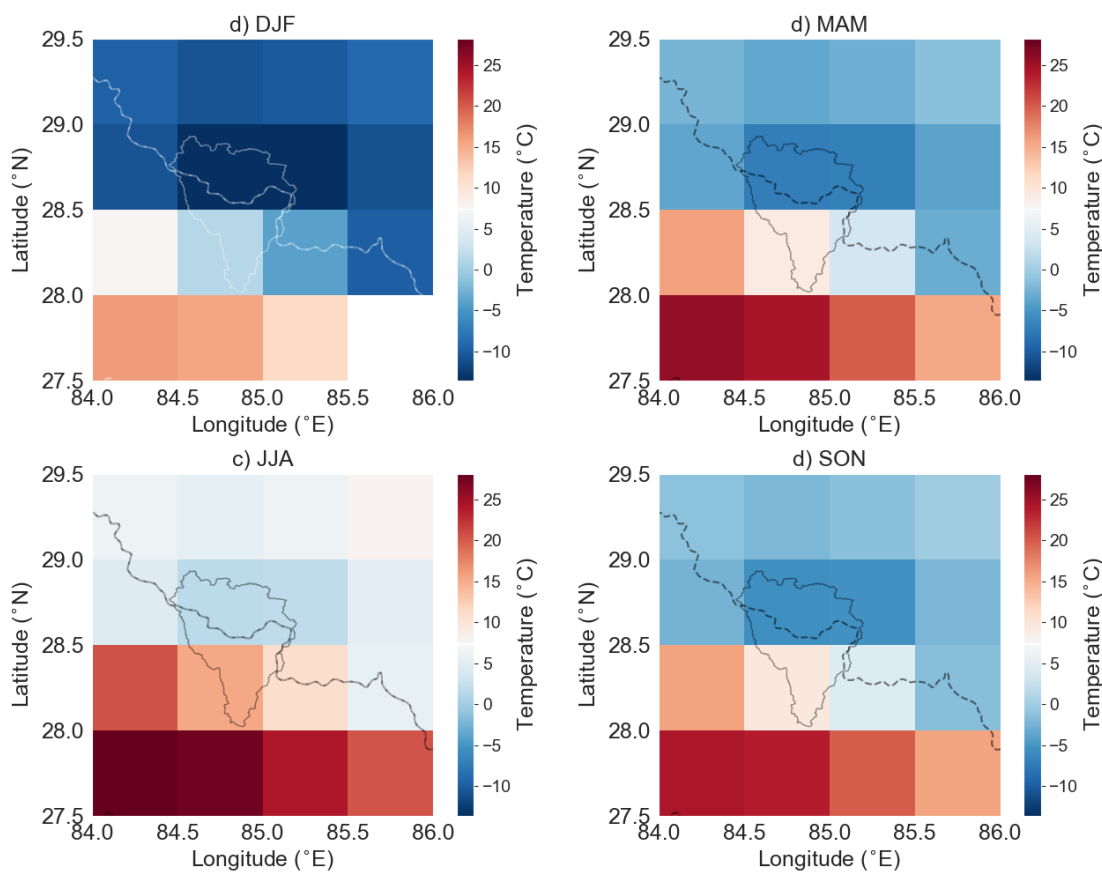


Figure 6.4: Mean seasonal temperature in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data.

The seasonal accumulated precipitation (mm/hr) from the WFDE5 data in the Budhi Gandaki region (2000-2015), as depicted in Figure 6.5, exhibits pronounced variations throughout the year. The winter months (DJF) experience relatively low precipitation levels, while the summer months (JJA) receive the highest amount of rainfall. Interestingly, the shoulder seasons (MAM and SON) demonstrate a similar pattern, with lower precipitation in the upper grid cells (Tibetan part) and higher precipitation in the lower grid cells (Nepal). This observation highlights the distinct precipitation patterns in different regions of the Budhi Gandaki catchment. However, the resolution of the WFDE5 data prohibits

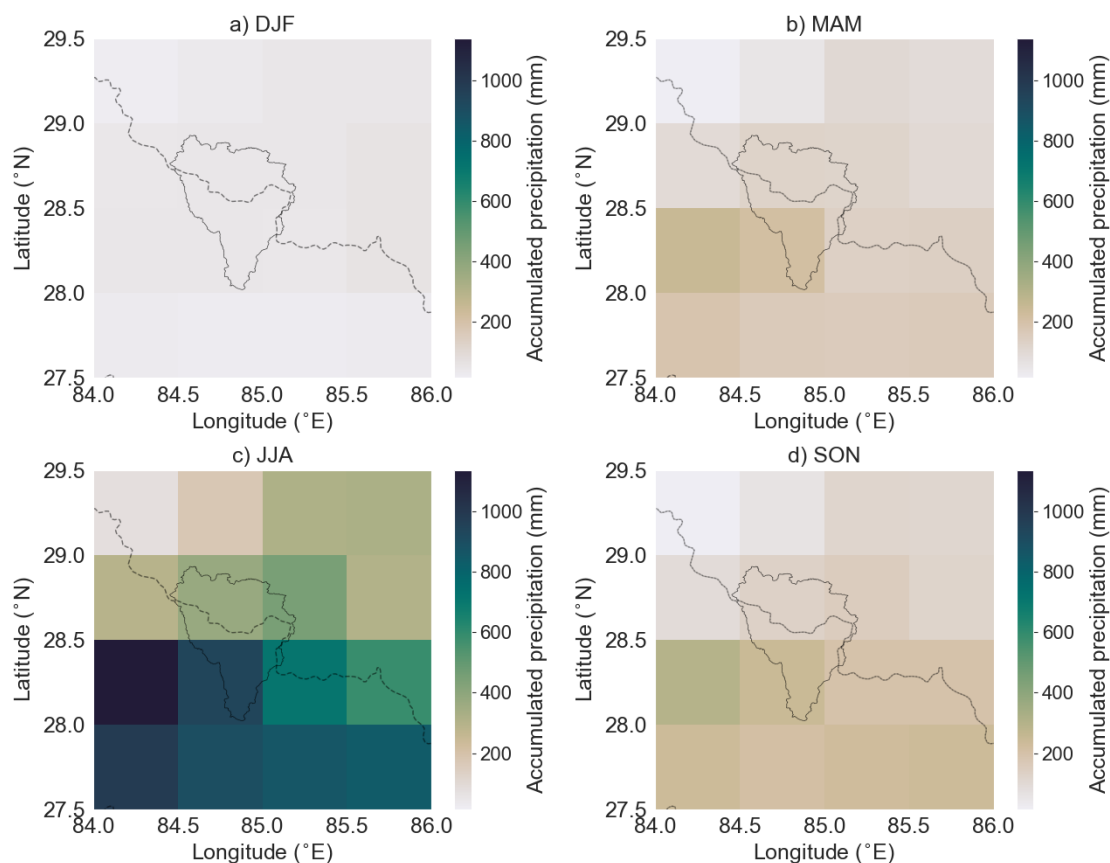


Figure 6.5: Mean seasonal accumulated precipitation [mm] in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.

6.1. Spatial and temporal variation in the forcing data

Figure 6.6 shows the mean seasonal accumulated snowfall (mm) from the WFDE5 data in the Budhi Gandaki region (2000-2015). The figure shows that most snow occur in the northern parts of the catchment. The mean snowfall is highest during spring (2662.5 mm) in the grid cell covering the northern part of the catchment, while the lowest values is found during summer and fall in the southern parts of the catchment.

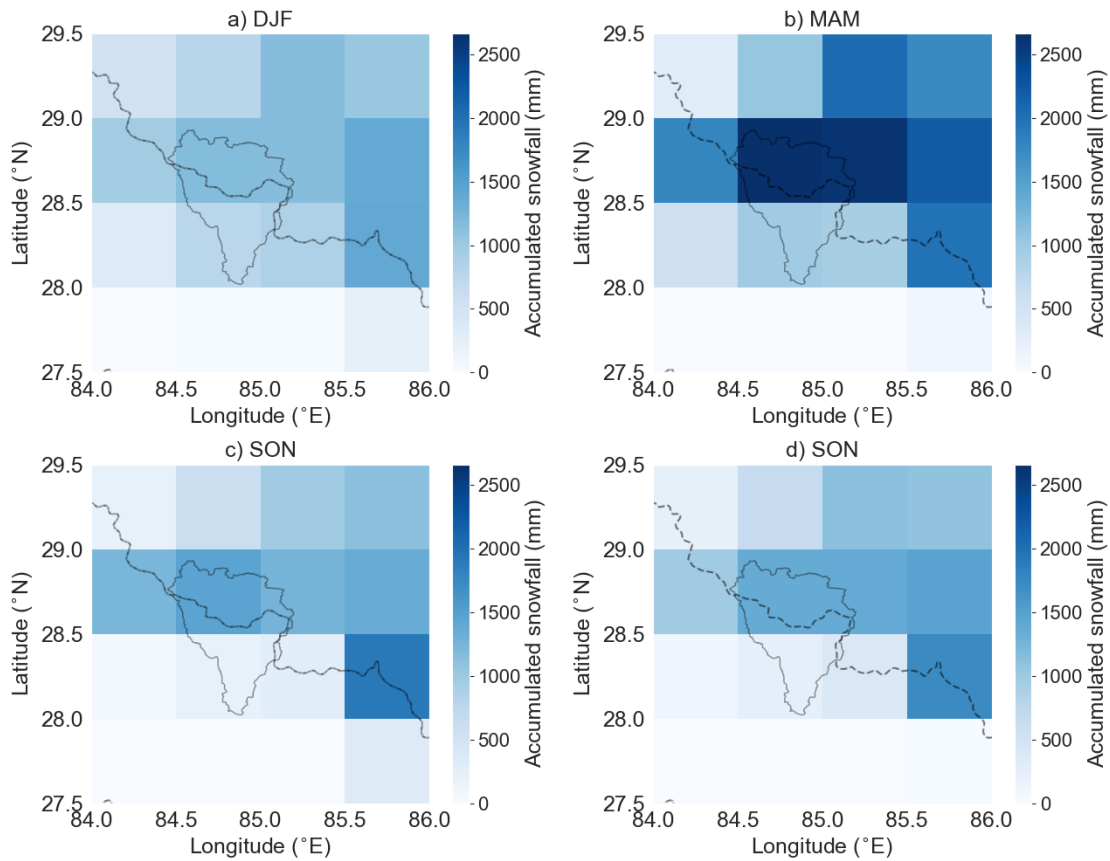


Figure 6.6: Mean seasonal accumulated snowfall [mm] in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.

The mean seasonal wind speed [m/s], as shown in Figure 6.7, reveals relatively consistent wind patterns throughout the seasons. There is minimal seasonal variation in wind speed. The North-East region experiences the strongest wind speeds, while the South-West region shows relatively weaker winds.

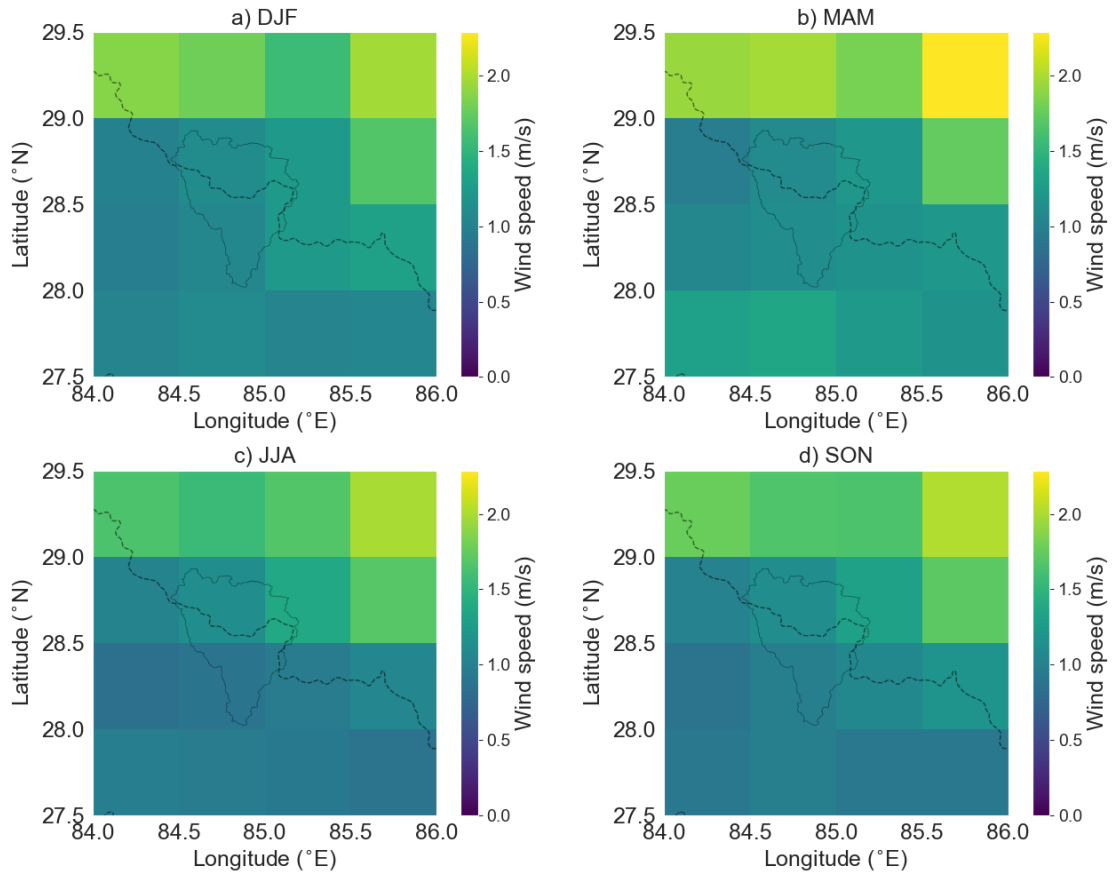


Figure 6.7: Mean seasonal wind speed in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.

6.1. Spatial and temporal variation in the forcing data

Figure 6.8 depicts pronounced seasonal variations in incoming shortwave radiation [W/m^2] levels within the Budhi Gandaki region. The figure reveals that spring exhibits considerably higher radiation levels compared to other seasons. In contrast, radiation levels are relatively lower during winter and fall across all grid cells. Moreover, the radiation distribution appears to be more uniform during fall and winter seasons.

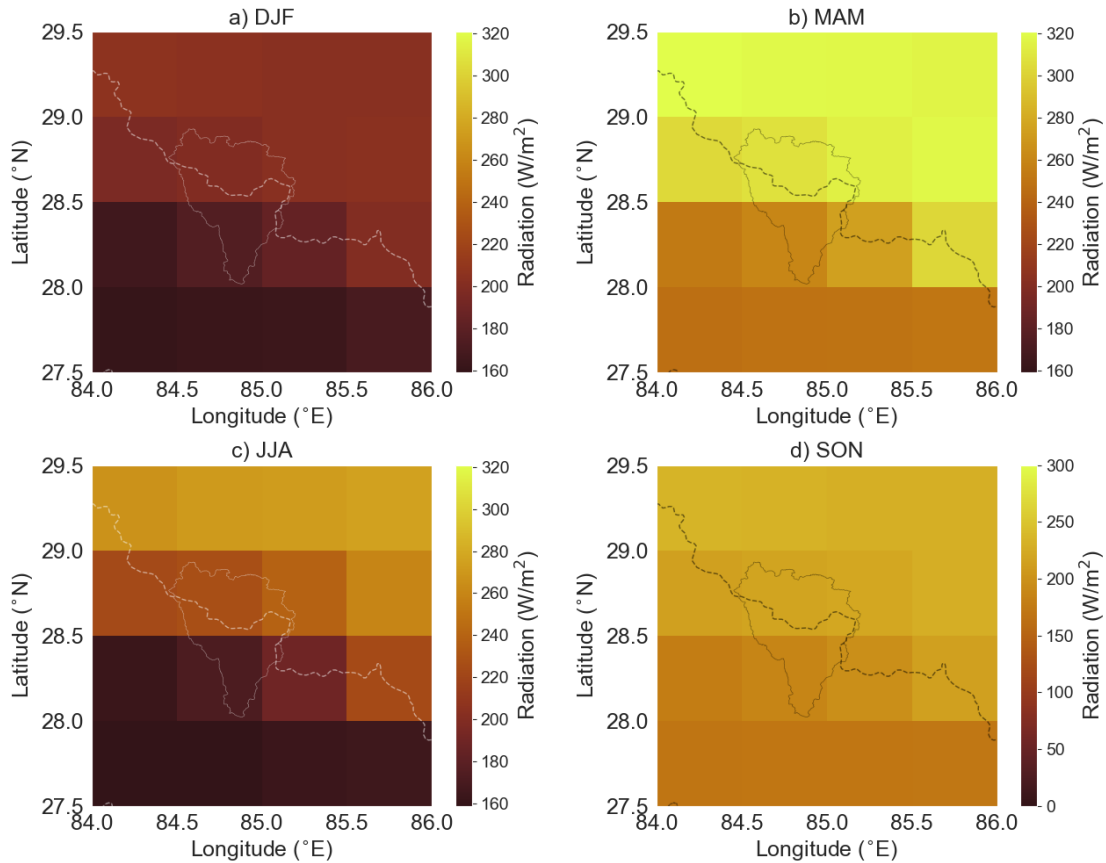


Figure 6.8: Mean seasonal incoming shortwave radiation in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data. The catchment boundary is shown as a solid line, while the Nepal-China border is depicted as a dashed line.

The mean seasonal relative humidity in the Budhi Gandaki region is illustrated in Figure 6.9, based on the WFDE5 data from 2000 to 2015. The analysis reveals distinct patterns in relative humidity across the seasons. The summer season exhibits the highest humidity levels, with mean relative humidity above 0.7 for the grid cells covering the catchment. In contrast, during winter, the relative humidity ranges between 0.6 and 0.8.

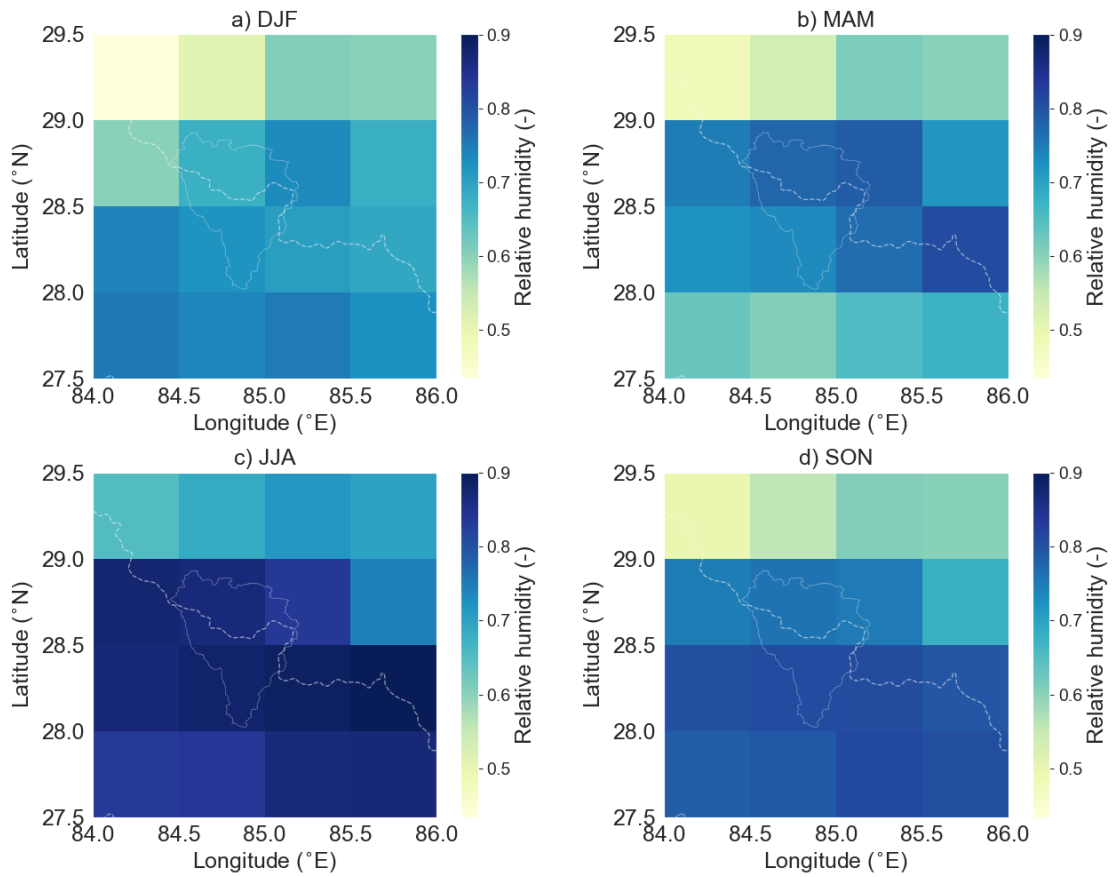


Figure 6.9: Mean seasonal relative humidity in the Budhi Gandaki catchment (2000-2015) from the WFDE5 data.

6.1.3 Analysis of Near-Surface Variables at Selected Elevations

In this section, near-surface forcing variables from the WFDE5 data set is investigated against elevation. Figure 6.10 shows a map of the Budhi Gandaki catchment, where the elevations of 4 WFDE5 grid cells are plotted for the period 2000-2015. Sub-figure a) shows that elevation is increasing from south to north. The daily mean temperature is lowest in at the Tibetan Plateau (4532 m), with a 365-day moving average just below zero. The temperatures here range from below -20 to around 8 degrees Celsius. At Higher Himalaya (sub-figure c), with an elevation of 4381 m, the temperatures have a similar profile as the Tibetan Plateau. The temperature increases considerably in the Lesser Himalaya (2163 m), which is also about 2200 m lower, with a moving average around 10 degrees Celsius. The highest temperatures are found in the Siwalik (1258 m) to the south (blue point). Here the temperature ranges between 15 to 30 degrees, with a moving average above 20 degrees. Sub-figure e) shows mean annual temperature plotted against elevation. It can be seen that the temperature decreases with height.

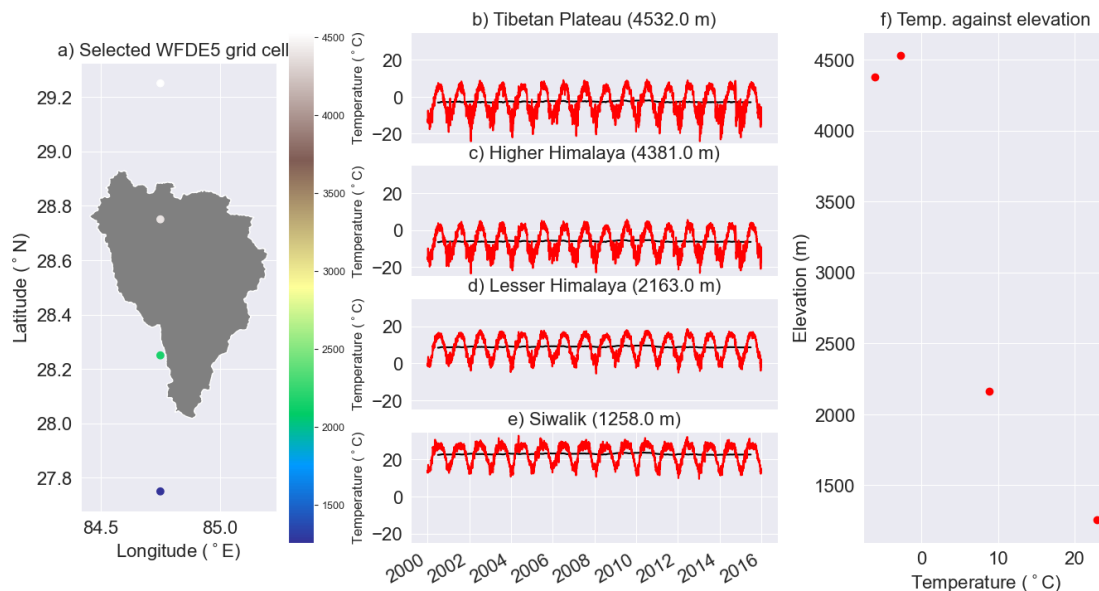


Figure 6.10: Temperature against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily mean temperature [°C] at different elevations. The black line represents a 365-day moving average. e) Shows mean annual temperature plotted against elevation.

Figure 6.11 shows daily mean accumulated precipitation [mm] against elevation. More precipitation is arriving in the Siwalik and the Lesser Himalaya, than in the higher Himalaya and the Tibetan Plateau. The most intense precipitation is occurring in the Siwalik, reaching values above 200 mm in a day in the most extreme cases. The daily accumulated precipitation is rarely above 50 mm in the Higher Himalaya and 30 mm in the Tibetan Plateau. The mean daily accumulated precipitation is increasing from with height from 1258 m to 2163 m, but decreasing for higher elevations. There are too few data points to see if this is a general pattern.

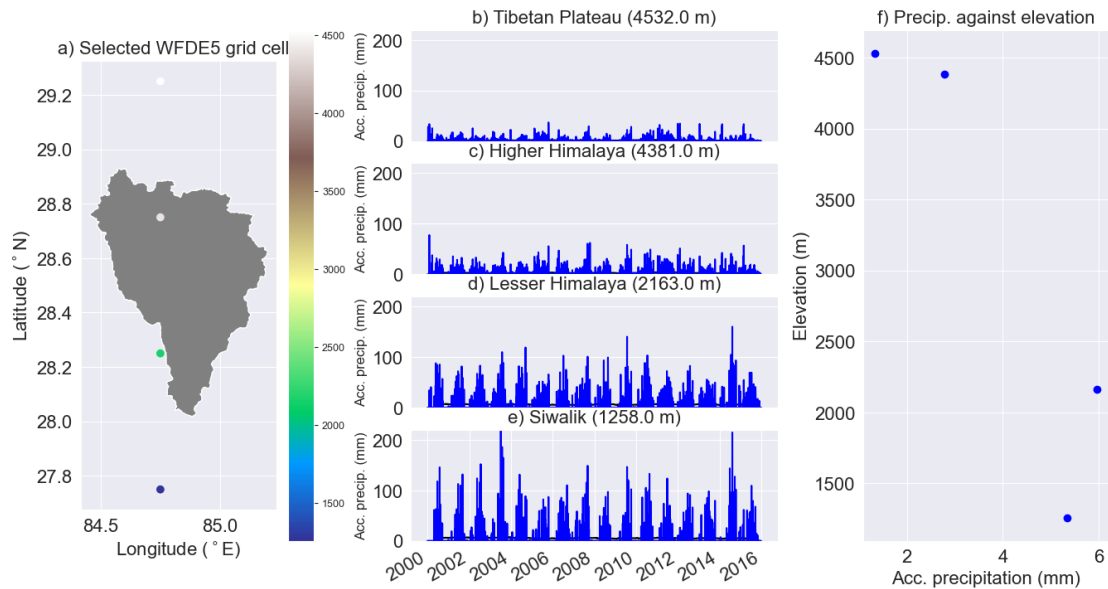


Figure 6.11: Precipitation against elevation. a) Elevation [m] map of selected grid cells in the WFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily mean accumulated (acc. precip.) [mm] at different elevations. The black line represents a 365-day moving average. e) Shows mean daily accumulated precipitation plotted against elevation.

6.1. Spatial and temporal variation in the forcing data

Figure 6.12 shows daily mean wind speed [m/s] at different elevations. The daily mean wind speed is relatively similar in the Siwalik and the Lesser Himalaya, ranging between 0.5 and 2.0 m/s. The wind speed in the Higher Himalaya has a similar range as the two more southern regions, but there is more inter-daily variation. The wind speed increases at the Tibetan Plateau, ranging between 0.5-4.0 m/s. The variation in wind speed here is comparatively much greater. The mean annual wind speed decreases with height from the Siwalik to the Lesser Himalaya. The mean annual wind speed is slightly higher at 4381 m, than at 2163 m. The mean annual wind speed is considerably higher at the highest elevation.

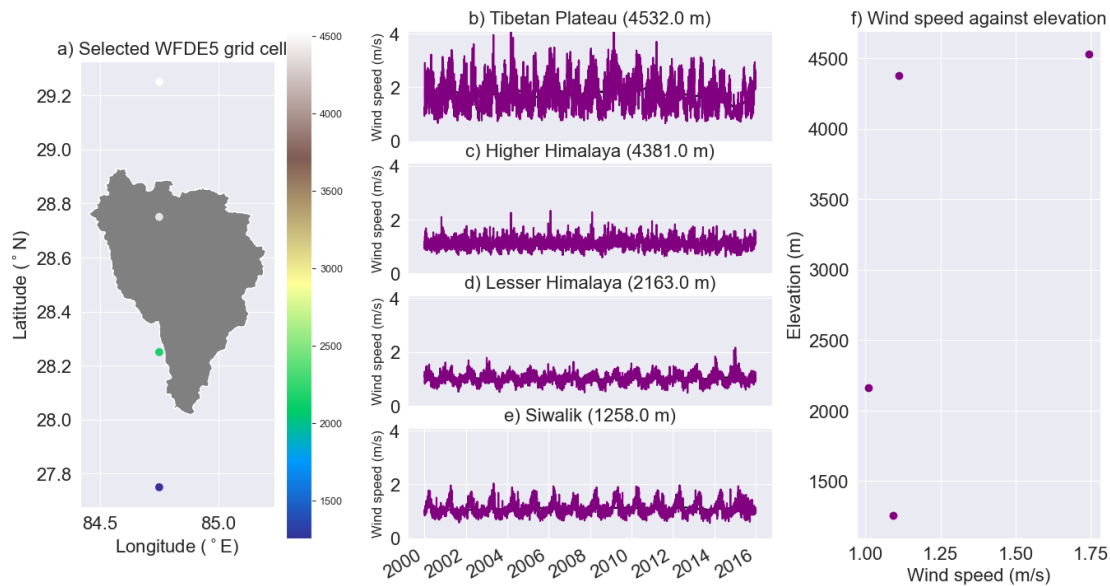


Figure 6.12: Wind speed against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily mean wind speed [m/s] at different elevations. The black line represents a 365-day moving average. e) Shows mean annual wind speed plotted against elevation.

Figure 6.13 shows the vertical relative humidity profile. There is greater variation in relative humidity in the Tibetan Plateau (sub-figure b), than in the Siwalik. The Relative humidity reaches below 0.5 in the Tibeatan Platau, which happens rarely at the lowest elevation. The relative humidity decreases with high (seen in sub-figure f).

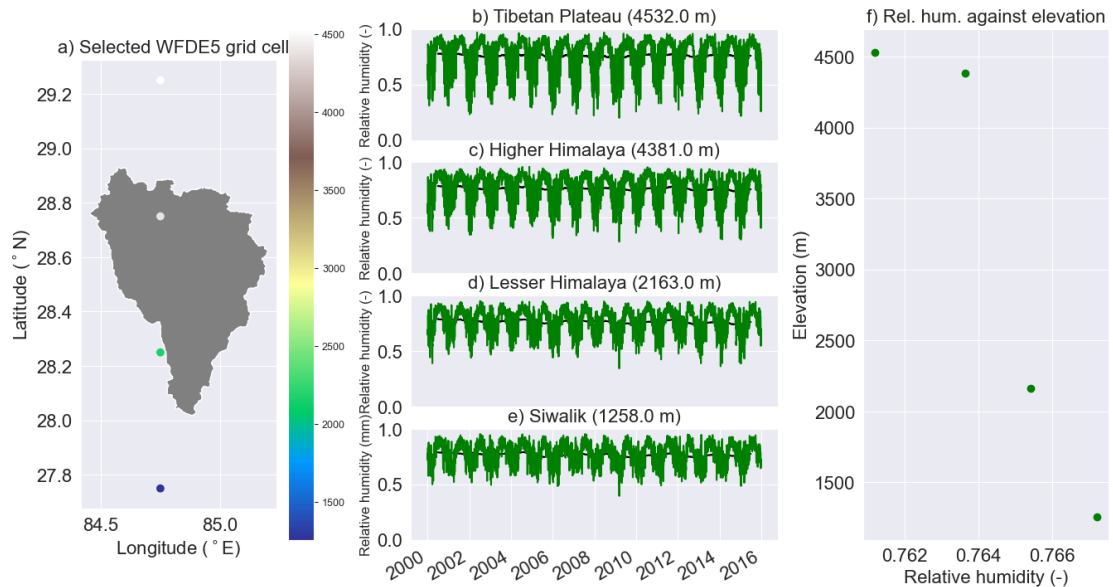


Figure 6.13: Relative humidity against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The points represent the mid-points of each grid cell. The catchment is shown in grey. b)-e) shows daily relative humidity [°C] at different elevations. The black line represents a 365-day moving average. e) Shows mean annual relative humidity plotted against elevation.

Figure 6.14 shows the incoming shortwave radiation [W/m^2] plotted against elevation. In general, the radiation increases with height. At the lowest elevation (1258 m), the daily mean radiation ranges from about $20 \text{ W}/\text{m}^2$ to $300 \text{ W}/\text{m}^2$, while from about 100 to $400 \text{ W}/\text{m}^2$ at the highest elevation (4532 m).

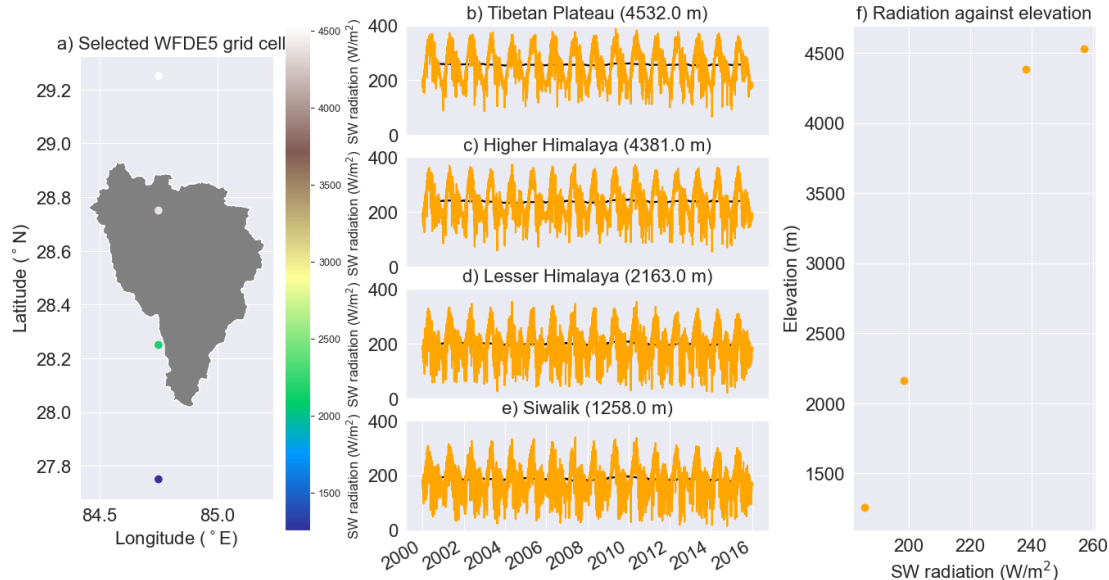


Figure 6.14: Incoming shortwave radiation against elevation. a) Elevation map of selected grid cells in the WDFDE5 data. The catchment is shown in grey. b)-e) shows daily mean incoming shortwave radiation [W/m^2] at different elevations. The black line represents a 365-day rolling mean. e) Shows mean annual SW radiation plotted against elevation.

6.1.4 Observed discharge at Arughat gauging station

The statistics for the observed discharge at the Arughat gauging station from 2000 to 2015 are presented in Table 6.2. The analysis reveals several key observations about the discharge data.

The mean discharge of $158.6 \text{ m}^3/\text{s}$ exceeds the median discharge of $71.8 \text{ m}^3/\text{s}$, indicating a right-skewed distribution. This suggests that there are a few high discharge values that contribute to the higher mean. The observed discharge exhibits a wide range, ranging from a minimum flow of $19.6 \text{ m}^3/\text{s}$ to a maximum of $906.0 \text{ m}^3/\text{s}$. Additionally, the mean observed discharge ($158.6 \text{ m}^3/\text{s}$ at the Arughat gauging station) represents approximately 71% of the observed discharge estimated at the Budhi Gandaki Hydropower plant site ($220 \text{ m}^3/\text{s}$), located 60 km downstream in Bahuntar, Nepal. This comparison provides a relative measure of the discharge changes along the river course.

Table 6.2: Statistics for the observed discharge (2000-2015) at Arughat gauging station

Statistic	Value	Unit
Mean	158.6	m ³ /s
Median	71.8	m ³ /s
Min	19.6	m ³ /s
Max	906.0	m ³ /s
Standard deviation	166.5	m ³ /s
Mean annual volume	$2.0 \cdot 10^9$	m ³

Figure 6.15 presents two sub-figures depicting the total annual volume and the median monthly observed discharge at the Arughat gauging station.

The left sub-figure illustrates the median annual volume per year, which is approximately 2.02×10^8 m³. The data exhibit significant variability, with values ranging from 1.6 to 2.4×10^8 m³. Outliers in the data set exhibit even more extreme values, indicating the presence of exceptional events or conditions.

The right sub-figure shows on the median discharge values across different months. It reveals that the median discharge reaches its peak during August. In contrast, the median discharge is lowest during the winter months. The observed discharge plotted for each year shows a great inter-annual variation, and is most visible during the pre-monsoon (March-May) and monsoon period (June-September).

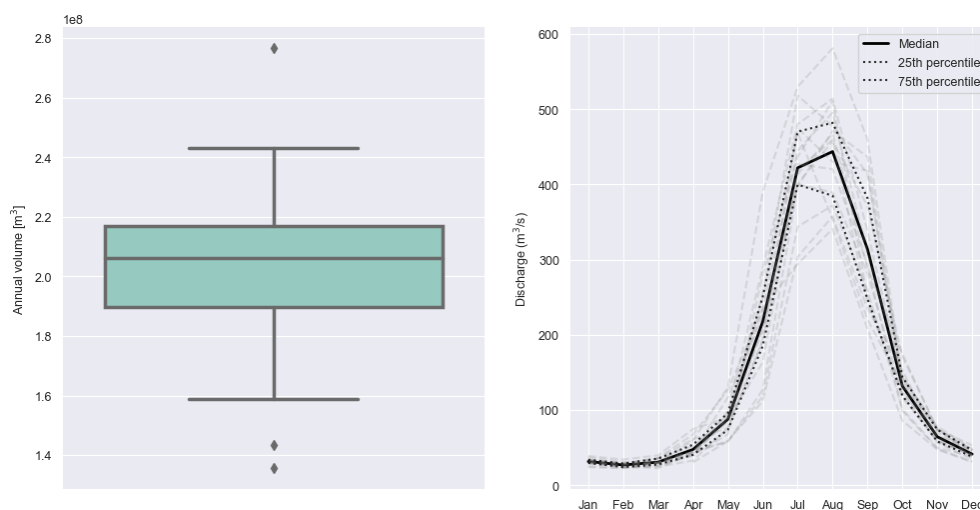


Figure 6.15: Total annual volume and monthly observed river discharge measured at Arughat gauging station (2000-2015). The box plot in the left sub-figure show the locality, spread and skewness of the total annual volume through its quartiles. Each box extend from the lower quartile to the upper quartile. The lines outside the boxes indicates variability outside the quartiles. The dots (outliers) are values significantly different from the rest of the data set. The median is represented by a horizontal line inside each box. The sub-figure to the right shows the monthly median river discharge (2000-2015) as a black solid line. The dotted lines represent the 25 and 75 % percentiles of the monthly discharge. The grey dashed lines represent the observed discharge for each year in the period.

6.1.5 Interpolation results from Inverse Distance Weighting and Bayesian Temperature Kriging

Figure 6.16 shows the catchment-wide mean monthly interpolated values for the input variables that goes into the Shyft simulations. Sub-figure a) shows that the mean monthly accumulated precipitation is highest (just over 4000 mm) in July and lowest in November. The precipitation increases from November-July and then decreases again. The temperature (sub-figure b) increases from about -12°C in January to 3 degrees in July. The mean monthly temperatures are below zero from September-May. The relative humidity follows the same pattern where it increases from 0.65 in January to 0.90 in July, before decreasing again. There is a substantial decrease in the mean monthly wind speed from about 1.09 m/s (June) to about 0.95 m/s in the monsoon period (August). Then the wind speed increases equally much from August to October (1.1 m/s). The mean monthly wind speed is relatively stable from December to June. The radiation is systematically higher for the TINs than for the regular grid cells. The radiation seem to have two peaks during the year. The highest peak is in April (300 W/m^2 for the TINs and about 270 W/m^2 for the regular cells). The second peak can be seen in October, when the radiation reaches just above (below) 200 W/m^2 for the TINs (regular cells). The radiation is at its lowest during January and February (about $160\text{-}180 \text{ W/m}^2$) and July.

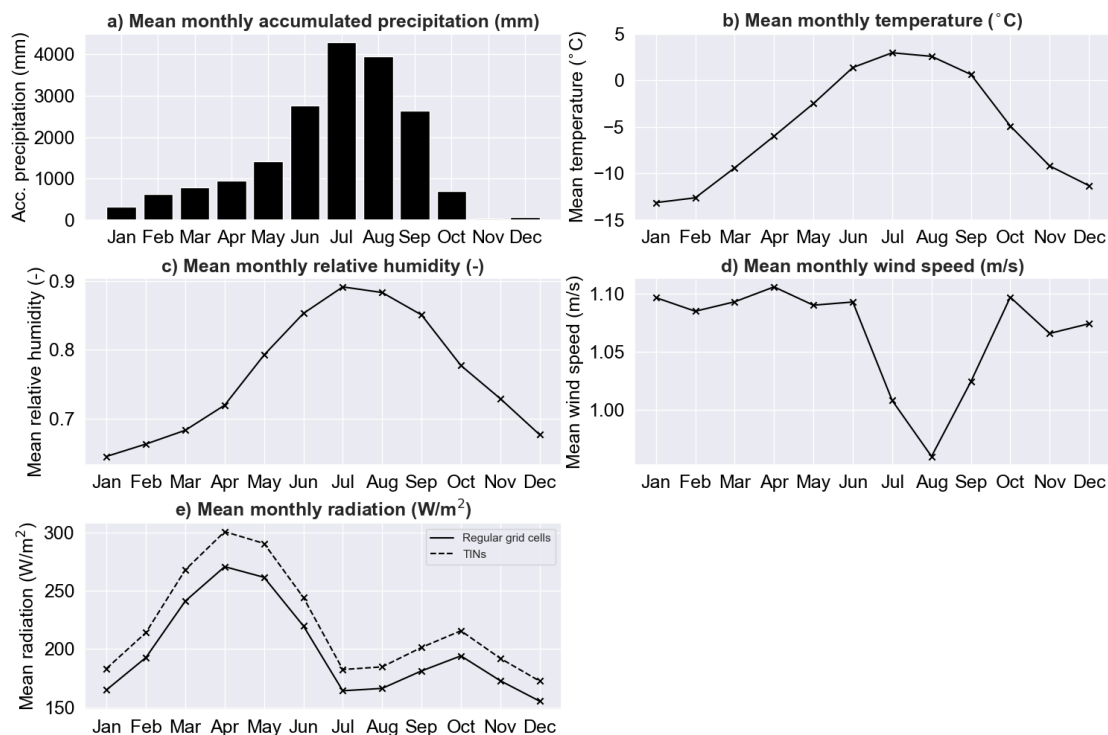


Figure 6.16: Monthly mean interpolated input variables from Shyft (2000-2015). a) Shows the mean monthly accumulated precipitation (mm), b) mean monthly temperature ($^{\circ}\text{C}$), c) mean monthly relative humidity (-), d) mean monthly wind speed (m/s) and e) mean monthly radiation (W/m^2). Regular grid and TINs have different radiation slope factor, and has therefore different interpolated values.

Figure 6.17 shows the mean winter and summer temperature in the PTSTK model using the TIN-1071 terrain representation. The mean DJF temperature (sub-figure a) is below zero for most of the northern parts of the catchment, while between 0-10 degrees Celsius in the southern parts. The lowest temperatures (around -25°C) are found to the north-west, north and north-east. The same applies to the summer temperature. The summer temperature is ranging from -25 in the West to above 20 degrees to the South. As the mean monthly temperature is below -10 degrees (sub-figure b) in Figure 6.16), it must imply that most of the catchment area is experiencing negative mean monthly temperature during winter. The same applies for the mean monthly summer temperatures, where the catchment mean is between 0-5 degrees (sub-figure in Figure 6.16).

The mean winter precipitation is very low (0.00-0.05 mm/hr) for most areas in the catchment, except for lowest parts of the catchment to the south (0.075-0.200 mm/hr). In summer, large parts of the catchment has precipitation of 0.2 mm/hr or above. The precipitation is less at the highest elevations during this season. The increase in accumulated precipitation during DJF seen Figure 6.16, is likely to have a considerable contribution from the precipitation seen in the southern part of the catchment (sub-figure c) in Figure 6.17.

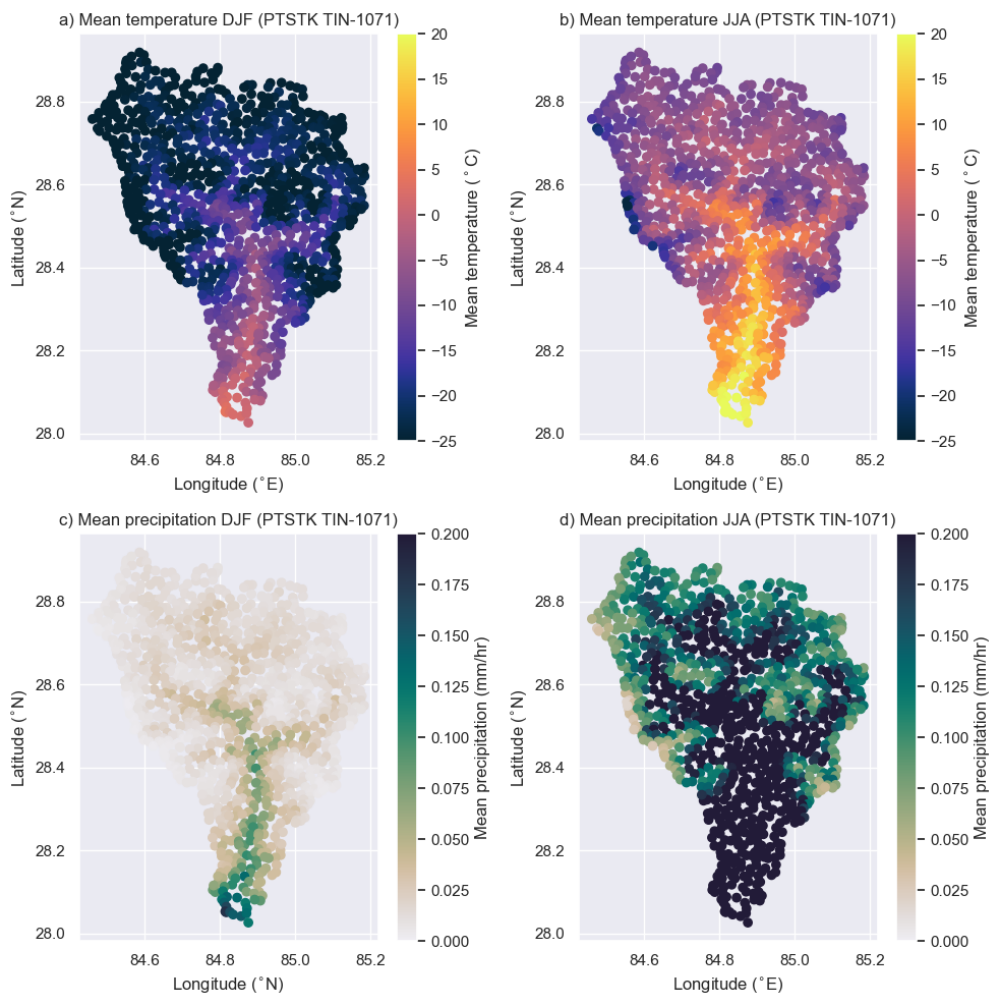


Figure 6.17: Interpolated temperature and precipitation for JJA and DJF in the PTSTK simulation using TIN-1071 (2000-2015). The scatter points are mid-points in the TIN-1071.

6.1. Spatial and temporal variation in the forcing data

Figure 6.16 shows the mean interpolated precipitation (mm/hr) and temperature ($^{\circ}\text{C}$) against elevation (m), and the cumulative area (-) against elevation (m). The figure shows that the temperature decreases linearly with height, while the precipitation decreases exponentially with height. About half of the cumulative area is above 4700 m.a.s.l. and only a few percent is above 6000 m.a.s.l. (less than 5%). Around 60% of the catchment area has a mean temperature below 0 degrees Celsius. The interpolation of both temperature and precipitation is very similar for both models. Most of the catchment area ($> 50\%$) is between 4000-5500 m.a.s.l.

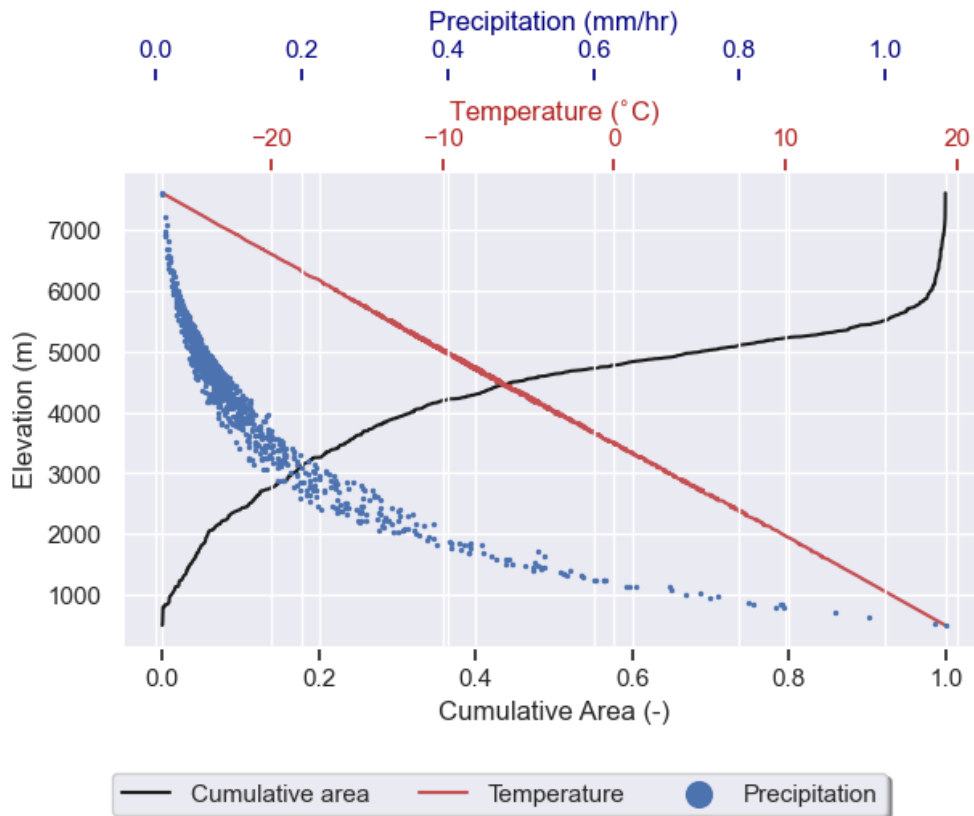


Figure 6.18: Mean interpolated precipitation (mm/hr) and temperature ($^{\circ}\text{C}$) against elevation (m), and elevation area against cumulative area (-). The values from the PTGSK model stack using TIN-1071 (2000-2015).

6.2 Impacts of Terrain Representation and Snow Routines on Calibration and Model Performance: Discharge

This section presents the results of the calibration and model performance evaluation. Optimized parameters will be presented and compared using normalized values. Thereafter, simulated discharge is compared with observed discharge using different evaluation criteria. Lastly the effects of using snow routines and terrain representations on simulated water balance components will be presented.

The calibration of the simulated discharge is done daily from 1 January 2000 to 31 December 2004. The simulation is run again hourly using optimized parameters after the calibration, and resampled to daily resolution to match the time steps in the observed discharge. Both simulated (using optimized parameters) and observed discharge series range from 1 January 2000 to 31 December 2015.

6.2.1 Optimized Parameters Using Min BOBYQA Method

Table 6.3 shows the optimized parameters for each calibration run using the Min BOBYQA method. The table shows that the *kirchner.c1* parameter (empirical outlet coefficient) is slightly less negative for the PTGSK models than the PTSTK models using the same simulation unit (e.g. regular cells/ TINs). The PTSTK with the regular PTSTK has the lowest *kirchner.c1* value (-6.02), and the PTGSK with a TIN-1071 has the highest value (-5.69). The *kirchner.c2* parameters is only negative for the PTSTK using Reg-1077 (-0.11), and positive for the other models (0.01 to 0.22). *kirchner.c3* ranges between -0.15 and -0.06. The *tx* parameter (threshold for determining precipitation phase) is between -3.00 and 0.97. The precipitation correction scale factor, *p_corr.scale_factor* is lowest for the PTSTK model using a low resolution TIN (TIN-911) (1.25), and highest for the regular PTSTK (1.48). The *gm.direct_response* (direct response from glacier melt) ranges from 0.49 for the PTGSK using TIN-991 and PTSTK with TIN-1071, to 1.0 for regular cells using either PTGSK or PTSTK.

Looking only at the PTGSK model, TIN-1382 has the lowest value for the *gs.wind_scale* (slope in turbulent wind function) parameter (3.32) and the PTGSK model with TIN-1071 has the highest value (6.00). The *gs.fast_alebdo_decay_rate* (albedo decay rate during melt days) is ranging from 6.00 (regular grid) to 11.74 (TIN-911). The *gs.slow_alebdo_decay_rate* (albedo decay rate in cold conditions) is ranging from 21.96 to 40.0.

For the PTSTK model, the *cx* parameter (degree-day melt factor) is highest for TIN-1071 (27.70), and lowest for TIN-1382 (13.65). The *ts* (threshold temperature for melt onset) parameter goes from -2.79 (TIN-4681) to -0.17 (TIN-1071), while the *lwmax* (max liquid water as a fraction of ice in the snowpack) parameter goes from 0.47 (TIN-1071) to 0.99 (TIN-1382). The *cfr* (refreeze coefficient) ranges between 0.02 (TIN-1071) to 0.74 for TIN-4681.

6.2. Impacts of Terrain Representation and Snow Routines on Calibration and Model Performance: Discharge

Table 6.3: Results calibration using the Min BOBYQA optimization.

Parameter	Reg-1077		TIN-911		TIN-1071		TIN-1382		TIN-4681	
	PTGSK	PTSTK	PTGSK	PTSTK	PTGSK	PTSTK	PTGSK	PTSTK	PTGSK	PTSTK
<i>Used in both</i>										
kirchner.c1	-5.78	-6.02	-5.70	-5.85	-5.69	-5.84	-5.73	-5.87	-5.75	-5.91
kirchner.c2	0.11	-0.11	0.13	0.13	0.17	0.11	0.22	0.12	0.01	0.03
kirchner.c3	-0.11	-0.15	-0.10	-0.09	-0.10	-0.10	-0.06	-0.09	-0.15	-0.12
tx	-2.17	-0.53	-2.09	-0.62	-2.97	-0.23	-2.66	-0.43	-3.00	0.97
p_corr.scale_factor	1.43	1.48	1.32	1.25	1.31	1.26	1.31	1.30	1.28	1.30
gm.direct_response	1.0	1.0	0.49	0.51	0.52	0.49	0.52	0.55	0.52	0.52
<i>Model specific</i>										
gs.wind_scale	6.00	-	4.53	-	5.30	-	3.32	-	4.20	-
gs.fast_albedo_decay_rate	6.60	-	11.32	-	11.74	-	11.38	-	9.07	-
gs.slow_albedo_decay_rate	21.96	-	28.88	-	34.65	-	39.10	-	40.0	-
st.cx	-	14.79	-	19.19	-	27.70	-	13.65	-	18.91
st.ts	-	-0.88	-	-0.62	-	-0.58	-	-2.01	-	-2.71
st.lwmax	-	0.58	-	0.58	-	0.47	-	0.99	-	0.76
st.cfr	-	0.13	-	0.21	-	0.02	-	0.28	-	0.74

Figure 6.19 shows the optimized parameter ranges for the the Gamma Snow and Snow Tiles models. Sub-figure a) shows the the *kirchnerc1* [-] parameter has a narrow range both both models. The values are lower in the Gamma Snow routine than in the Snow Tiles routine. There is an overlap over optimized parameter values for the *kirchnerc2* [-] between the models. There is a great range in optimized values for the *kirchnerc3* [-] parameter the PTSTK model, compared with the PTGSK model (except to outliers). The threshold for determining precipitation phase parameter, *tx* [°C], has lower values in the PTGSK than in the PTSTK model. The precipitation correction factor is lower in the PTSTK model, but there are outliers far off the lower and upper quartiles for both models.

The model specific parameters (f-i) for the PTGSK show that the optimized parameter values vary quite a lot for the *wind_scale* [m/s], *fast_albedo_decay_rate* [days] and *slow_albedo_decay_rate* [days]. The outliers seen for the *glacier_melt* are the regular grid models. For the PTSTK model, the *cfr* and *cx* parameters have comparatively lower variation than *ts* and *lwmax*. Especially the threshold temperature for snow melt onset, *ts*, varies quite a lot.

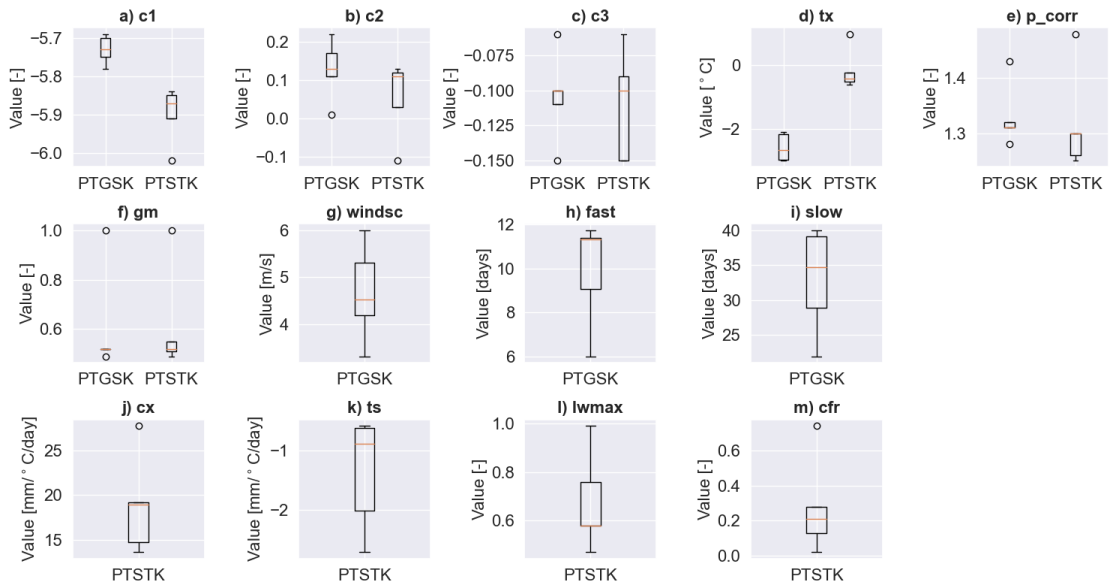


Figure 6.19: Optimized parameter ranges for the Gamma Snow and Snow tiles models using regular grid and TINs. The boxplots show the locality, spread and skewness of the CSI for each month through their quartiles. Each box extend from the lower quartile to the upper quartile. The lines outside the boxes indicates variability outside the quartiles. The dots (outliers) are values significantly different from the rest of the dataset. The median is represented by a horizontal line inside each box.

The effect of the two different snow routines and terrain representation on calibration can also be measured by how long time it takes to calibrate each model. [Figure 6.20](#) shows how calibration run time [s] for each model configuration. The PTSTK with a regular grid takes the least amount of time to calibrate. It has lower resolution (1077 cells) than some of the high resolution TINs, but it also has much fewer parameters than the PTGSK. The figure shows that the impact of the two snow routines on calibration run time, is greater than the impact of increasing resolution (cells). For the regular grid, the difference in calibration run time between the Gamma Snow and Snow Tiles is about 47 min (or about -46% time reduction using the Snow Tiles routine). However, the time increases with about 27 min (26 %) using a TIN with similar resolution when comparing Reg-1077 with TIN-1071 in the Gamma Snow routine, and with 8 min (15.4%) when comparing Reg-1077 with TIN-1071 for the Snow Tiles routine. There is an -14.8 % reduction in calibration run time using the lower resolution TIN (TIN-911) than regular grid in the PTGSK model. This percentage reduction is similar to the reduction in simulation elements (cells) when going from 1077 cells to 911 cells (-15.4 %). However, the opposite is true for the Snow Tiles routine where there is an increase in time of 18 min (34.0 % increase). The fact that the PTSTK model with TIN-1071 takes slightly less time to calibrate than TIN-911, indicate that other factors such as server memory usage and connection speed when calibrating may be affecting the result. TIN-4681 has about 334% more cells than the regular grid, but takes about 17 more hours (1008% increase in time) for the Gamma Snow and 8 more hours for the Snow Tiles routine (879%), which is quite substantial. Thus, when using higher resolution TINs, there is a clear time advantage to use the Snow Tiles routine.

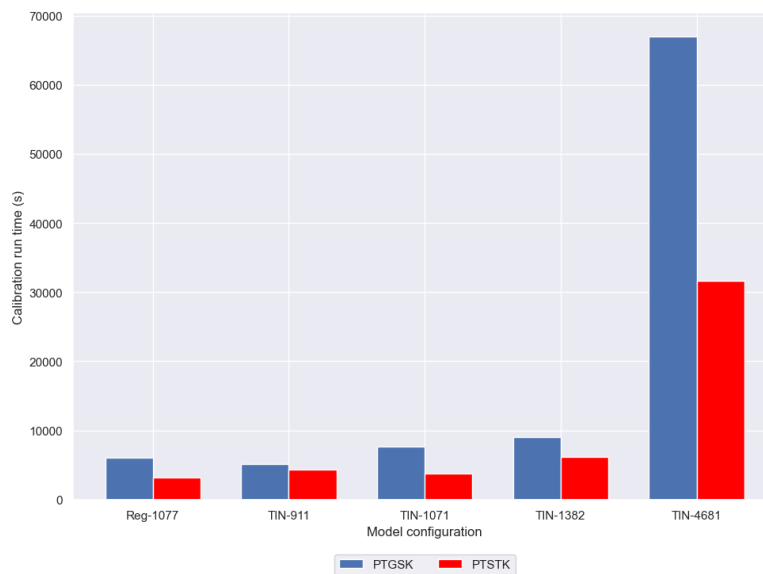


Figure 6.20: Figure shows the calibration run time [s] for all model configurations. Blue color shows the PTGSK model, and red shows the PTSTK mode.

6.2.2 Impact of Snow Routines and Terrain Representations on Simulated Discharge

To evaluate the of the two snow routines and terrain representations, the simulated discharge series and the observed discharge mentioned in the beginning of subsection 6.2.1, are first split into two series. The first series, here called calibration, is the period for which the simulation is calibrated (2000-01-01 to 2004-12-31). The calibration is done daily for this period. The period for which the model is not calibrated on, but there is available observed discharge data, is here called validation period (2005-01-01 to 2015-12-31). Once the discharge series are split into two, the model performance using the observed discharge as target, can be evaluated using different evaluation criteria. This is done both for the calibration period and the validation period.

Table 6.4 compares the 10 different simulations for the calibration period (2000-2004) and the validation period (2005-2015). The NSE, LnNSE, KGE and RMSE is calculated for both period and the values show the total scores for each period.

All of the configurations have a NSE value of 0.88 or above for the calibration period. This indicate that there is a very good match between the observed and simulated data for all simulations, and that the models are able to capture a large part of the variability in the observed data. The highest NSE during calibration (0.90) is found for all the PTSTK models using TINs. The regular grid (Reg-1077) with PTGSK has the lowest NSE (0.88). The LnNSE is lower than the NSE (between 0.83-0.87) for all calibration runs, indicating that the model perform worse during low flows. TIN-911 with PTSTK and TIN-1071 with PTSTK, have the highest LnNSE (0.87). The regular grid (Reg-1077) with PTSTK has the lowest LnNSE. The highest KGE is found for all of the PTSTK configurations with a value of 0.93 during the calibration. This suggests that there is a very good match between the observed and simulated PTSTK data, and that the models are able to capture the mean, variability and the correlation in the data very well. The KGE for the PTGSK configurations during the calibration period is between 0.88-0.90. PTSTK with TIN-1382 has the lowest RMSE value (59.51 m³/s), while the regular grid with PTGSK has the highest RMSE (64.44 m³/s) for the calibration runs.

Looking at the validation period (2005-2015), the NSE is ranging between 0.78-0.80, indicating a satisfactory match between simulated and observed values. The highest NSE (0.8) is found for the PTSTK model with regular grid. Looking at the LnNSE the values, the PTSTK with TINs have value of 0.84, while the PTGSK model with TINs have lower values ranging from 0.73-0.74. The highest KGE value is found for the PTSTK model with regular grid (0.75). The KGE is higher for regular grid in both models, compared with the TIN with similar resolution (TIN-1071) and lower (TIN-911). For the other models using TINs, there is a slight improvement in using the PTSTK model. The RMSE is about 10-12 m³/s higher for the validation period than for the calibration period for all configurations, ranging between 70.18-74.72 m³/s.

Table 6.4: Model evaluation metrics. The first column shows the configuration type, and the rest of the columns shows the results for the NSE, LnNSE, KGE and RMSE for the calibration period and the validation period. The calibration is done daily for 5 year. The evaluation is done hourly for a 5 year calibration period and the validation is done hourly for the 10 years after the calibration period.

Configuration	Calibration				Validation			
	NSE	LnNSE	KGE	RMSE [m ³ /s]	NSE	LnNSE	KGE	RMSE [m ³ /s]
Reg-1077 + PTGSK	0.88	0.84	0.88	64.44	0.77	0.71	0.71	74.72
Reg-1077 + PTSTK	0.89	0.83	0.93	60.73	0.80	0.81	0.75	70.18
TIN-911 + PTGSK	0.89	0.86	0.89	61.96	0.78	0.73	0.70	72.78
TIN-911 + PTSTK	0.90	0.87	0.93	59.58	0.78	0.84	0.70	72.07
TIN-1071 + PTGSK	0.89	0.86	0.90	62.17	0.78	0.75	0.70	72.84
TIN-1071 + PTSTK	0.90	0.87	0.93	59.80	0.79	0.84	0.71	71.69
TIN-1382 + PTGSK	0.89	0.86	0.89	62.13	0.78	0.73	0.70	72.78
TIN-1382 + PTSTK	0.90	0.86	0.93	59.40	0.79	0.84	0.72	70.48
TIN-4681 + PTGSK	0.89	0.86	0.90	61.87	0.78	0.74	0.70	72.79
TIN-4681 + PTSTK	0.90	0.86	0.93	59.42	0.79	0.84	0.72	70.76

Figure 6.21 shows the NSE, LnNSE, KGE and the RMSE calculated for each year during the simulation period. The NSE value is high (0.84 or above) for all models in the years 2002, 2003, 2004 and 2014 (see sub-figure a). The models perform worse in the years 2005 and 2009-2011 where the value is between 0.58-0.74. Between the models, the PTSTK model with Reg-1077 performs better in this period most of the years, except for year 2011. The NSE is generally high during the calibration period, compared with the validation period. The LnNSE in sub-figure b), shows that the PTSTK models perform better for low flows in general. The Snow Tiles routine have a higher LnNSE in the worst years for the NSE. The highest LnNSE values for all models are found in the calibration period and the first 5 years of the validation period, except for year 2005. The regular PTGSK has the lowest LnNSE value (0.46 in year 2011), and the PTSTK with TIN-1071 has the highest LnNSE in the year 2008. The KGE, illustrated in sub-figure c), is highest during the calibration period and the years 2006-2008, in addition to years 2013, for all the models. The models perform poorly during the years 2005 and 2009-2011. Sub-figure d) shows that the RMSE is never above 1 standard deviation in the observed discharge (166.5 m³/s). The RMSE is highest in year 2011 and lowest in 2004 and 2013.

The difference in NSE between the two snow routines and between terrain representation is not that apparent. For example, in the year with the lowest NSE (2011), the difference between the regular grid and the TIN with similar resolution (TIN-1071) is 0.06 for PTGSK, but 0.03 in the Snow Tiles routine for Reg-1077 and

6.2. Impacts of Terrain Representation and Snow Routines on Calibration and Model Performance: Discharge

TIN-1071. The difference between PTGSK and PTSTK for Reg-1077 is 0.02, and 0.01 for TIN-1071. In the year 2009, which is also a year where the NSE is generally low, the difference between Reg-1077 and TIN-1071 is 0.06 in the PTGSK, while 0.1 using PTSTK. The difference between the Gamma Snow and Snow Tiles routines is 0.01 for Reg-1077 and 0.01 for TIN-1071. In the year 2010, the difference between Reg-1077 and TIN-1071 is 0.01 in PTGSK, while 0.03 in PTSTK. The difference between the two snow routines in 2010, is 0.02 for Reg-1077 and 0.01 for TIN-1071. These results indicate a greater variation within each snow routine simulation, at least for the worst years, than between the two snow routines. However, using the LnNSE it can be seen from the figure that the PTSTK in general perform better for the worst years.

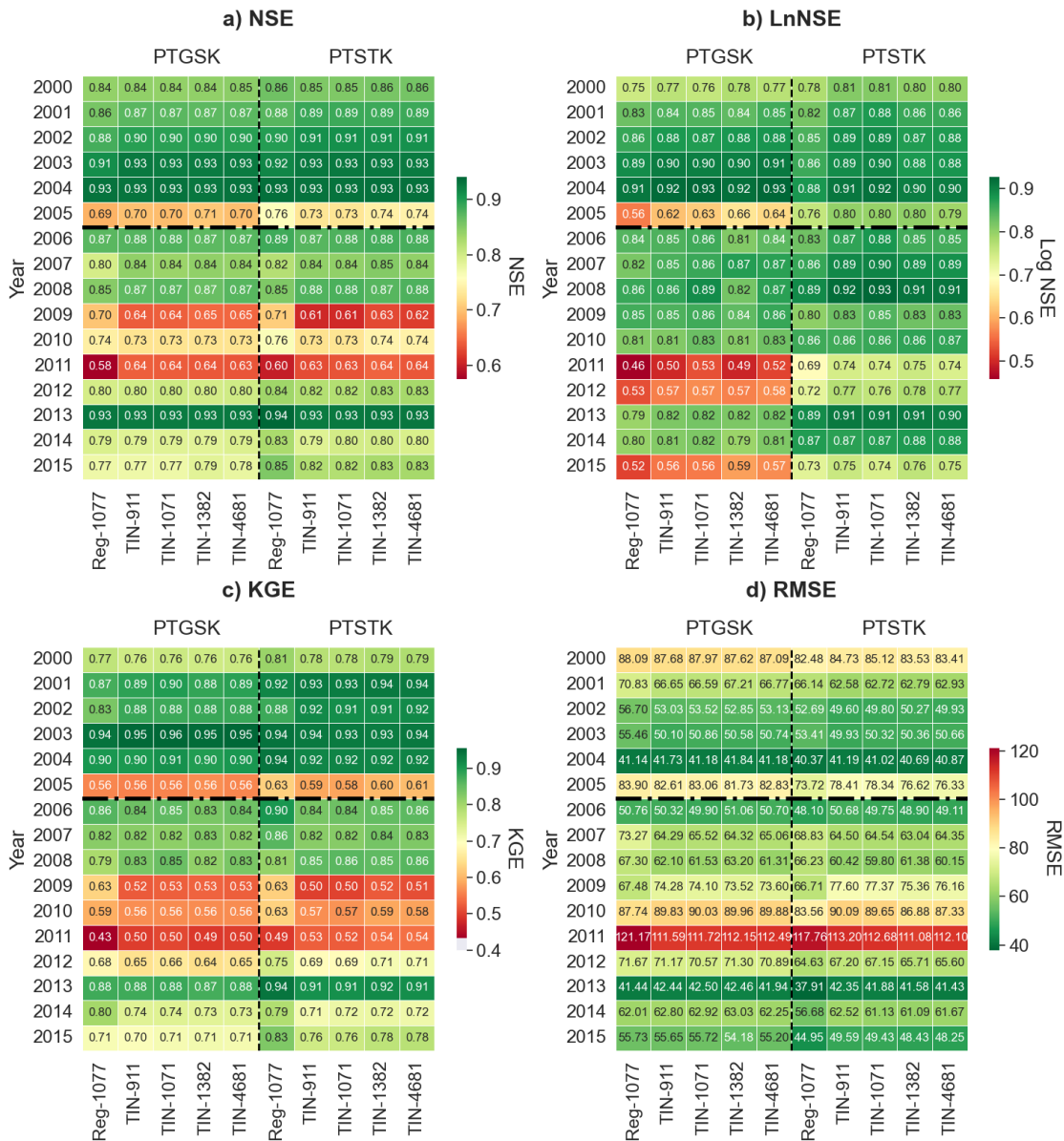


Figure 6.21: NSE and LnNSE for each year during the simulation period (2000-2015). Figure a) shows the NSE, b) shows the LnNSE, c) shows the KGE and d) shows the RMSE. The colorscale for the NSE, LnNSE and KGE goes from -1 in red to +1 in green. RMSE values goes from the minimum RMSE value to the maximum value. The vertical dashed lines separate the to snow routines and the thick dash dotted lines divide calibration period from validation period.

Figure 6.22 shows the comparison of the daily simulated discharge and the observed discharge during the calibration period (2000-2004) and validation period for all configurations. The figures show that there is generally a good agreement between the observed and simulated discharge, but that the peak flow is underestimated by all the models, especially during the calibration period. The PTSTK models capture the low flow better than the PTGSK models. This is most clear during the low flows in years 2012 and 2014, when observed discharge is much lower than the simulated during low flow. The variability in the models during low flow is greater than in the observed data. Both the Gamma Snow and Snow Tiles routines capture the inter-annual variability quite well. The inter-annual variation in observed discharge is also captured by the models. The timing of the peaks and lows match quite well for all models, but there is a weaker match between the simulated peaks and lows for the models. The models underestimate the peaks in the calibration period. The degree of overestimation or underestimation of peaks is less visible for the validation period. There are now clear trends for the entire period (2000-2015), but 365-day moving average (MA) from the observed discharge indicate that the mean annual discharge decreased from 2000 to 2007, before some high peak years (2008-2012) increased the moving average, before the MA remains relatively stable until 2014. The models also captures this in the sense that the observed and simulated peaks tend to be lower from 2000 to 2007.

6.2. Impacts of Terrain Representation and Snow Routines on Calibration and Model Performance: Discharge

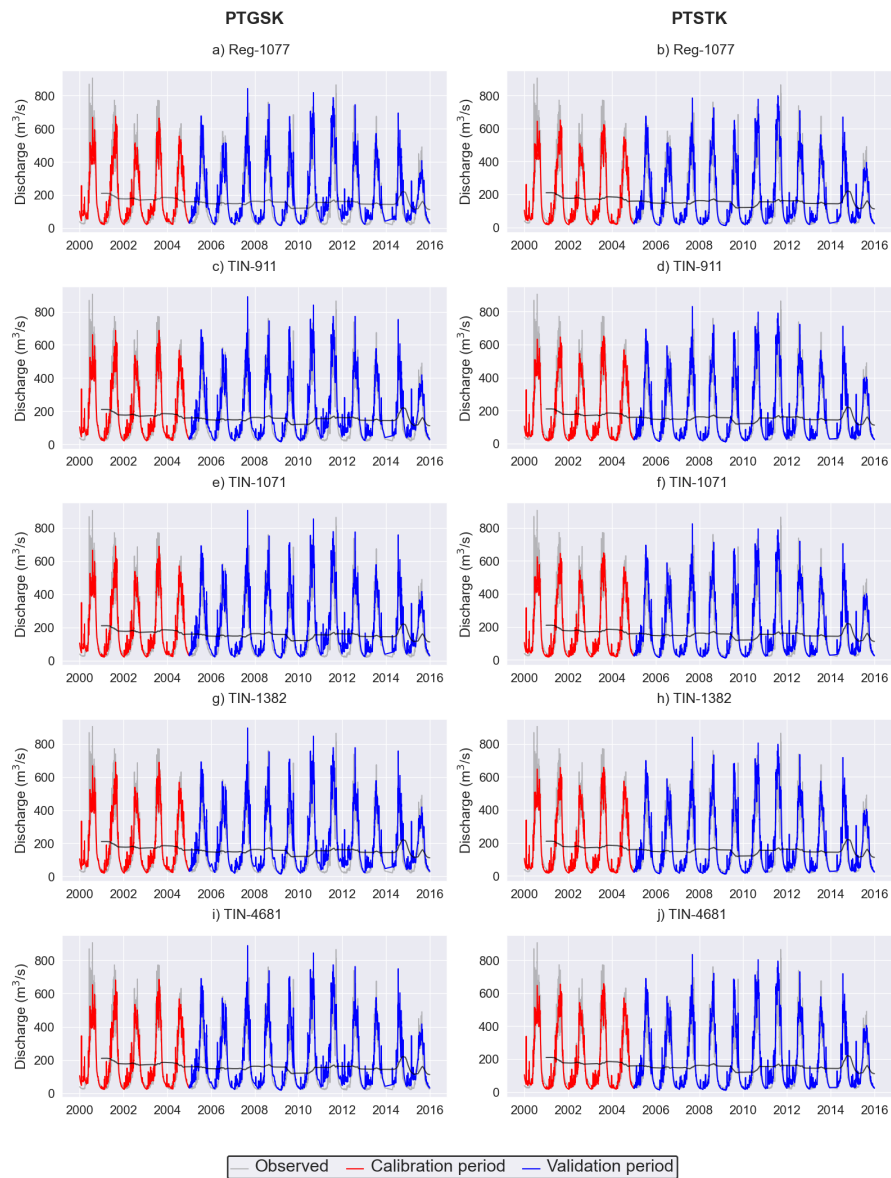


Figure 6.22: Daily discharge during calibration (2000-01-01 to 2004-12-31) and validation period (2005-01-01 to 2015-12-01). The calibration period is shown in red and the validation period in blue. Observed discharge is in grey. PTGSK models are shown to the left, PTSTK to the right. A 365-day moving average is shown as a grey solid line.

Figure 6.23 shows the daily observed and simulated discharge in the years 2011 and 2013. The year 2011 is a year where the NSE is low (NSE between 0.58 and 0.64), and the year 2013 is a year where the NSE is high (0.93-0.94). These years are chosen for a comparison of simulated discharge in years with very different performance. Sub-figures a) and b) show the low performance year 2011. It can be seen the both snow routines tend to overestimate the discharge for most of this year. The overestimation is greater in the PTGSK model during the pre-monsoon season. Both models capture the rapid increase from June to August, but the two models lags behind the rapid increase seen in the observed discharge in early June by about a week. Also the simulated peak discharge lags behind the observed discharge peak in July by a similar amount of time as for the rapid rise in June. The observed discharge is also decreasing earlier than the simulated discharge in the period August-September. However, there is a rapid increase in observed discharge observed from the start of September to mid September in this year, where the observed discharge is reached its highest value for that year. The same discharge peak values during September-October are not seen in neither of the models, but rather a much smaller uptick in discharge.

In the year 2013, the NSE is high (0.93-0.94). It can be seen from the figure that both models perform better during this year. The Gamma Snow routine overestimates the discharge more than the Snow Tiles during February to May. The models perform better in the transition periods between low and high flow, but doing worse during low and peak flow. Both models also lags behind the rapid rise seen in May-June.

The differences in performance are greater between the snow routines, than for the terrain representations in both years. Both models are also more responsive than what is observed (especially during winter and spring).

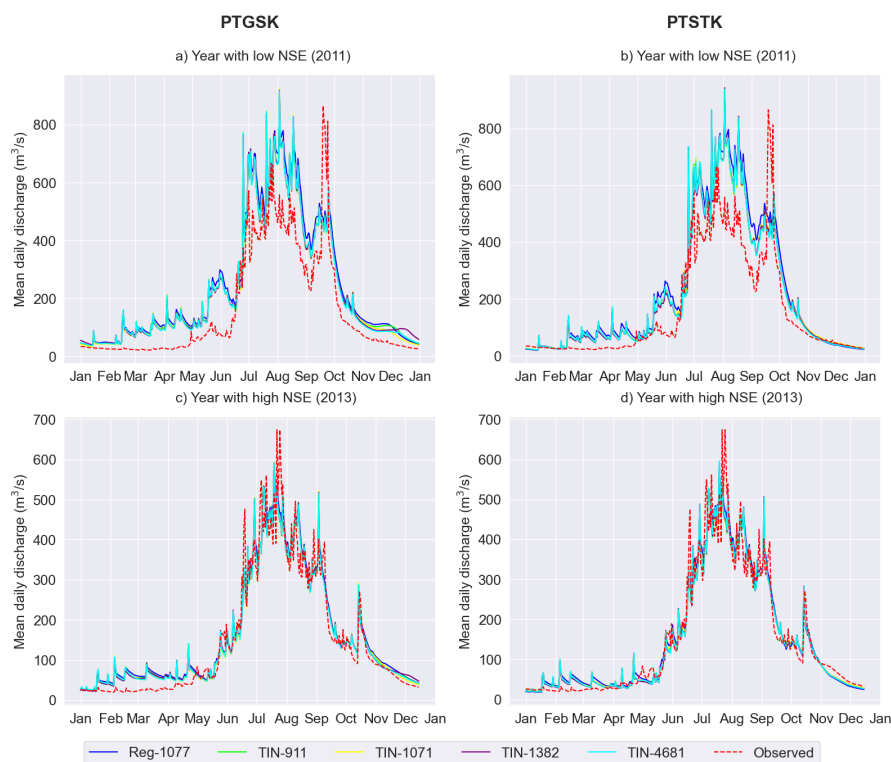


Figure 6.23: Daily observed and simulated discharge in the year with a low NSE (2011) and a high NSE (2013). Solid lines represent simulated discharge, while dashed line represent the observed discharge.

Figure 6.24 shows Quantile-Quantile (QQ) plots for the different simulations during the calibration and validation period using a daily time step. QQ-plots can be used to determine if two data sets come from the same distribution (Renard et al. 2010). If two distributions are the same, the points in the QQ-plot lie exactly on the identity line (1:1 line in the figure). The QQ-plot for the calibration period shown in sub-figure a), shows that the models to some degree are overestimating the discharge up until about 430 m³/s. For flow greater than 430 m³/s, the models are increasingly underestimating the flow. It can be said that the models do relatively well until the 90% quantile of the observed discharge, but perform poorly for more extreme values.

The QQ-plot for the validation period, sub-figure b), show that all the models systematically overestimate the flow. Furthermore, the models also perform worse for the values higher than the 90% quantile than below, except for the most extreme values (higher than 770 m³/s) where there is a relatively good fit. There is a great range between the models for flow above the 99% quantile in the validation period.

The QQ-plots are plotted on a log-scale in the two lower sub-figures to highlight low-flows. All of the models seem to underestimate low flow below the 1% quantile during the calibration period as can be seen in sub-figure c). The regular PTGSK overestimates low flow between the 5-10% quantile, while the TINs with PTGSK underestimate in the same interval. All of the PTGSK models overestimate for values between the 10% quantile up until right before the 90% quantile. The PTSTK models underestimate the flow up until about 4.5 log-observed discharge (m³/s). However, the PTSTK models are closer to the identity line between about 3.7-4.5 log-observed discharge (m³/s), than the PTGSK models. The two snow routines compare similarly for log-observed discharge above 5.0 m³/s.

The log QQ-plots for the validation period in sub-figure d) show that the PTGSK models match quite well for flow below the 1% quantile of the observed discharge, but overestimate quite substantially between the 1% quantile and the most extreme cases (>> 99% quantile). The PTSTK models underestimate the low flow (below 10% quantile) more than the PTGSK models, and overestimate flow until the 90% quantile. The overestimation in the PTSTK models is less than for the PTGSK up until a log observed discharge of about 4.5 m³/s where they even out for higher values. The models perform better in this period for the most extreme peak flows in the validation period than in the calibration period by comparing figure d) with c).

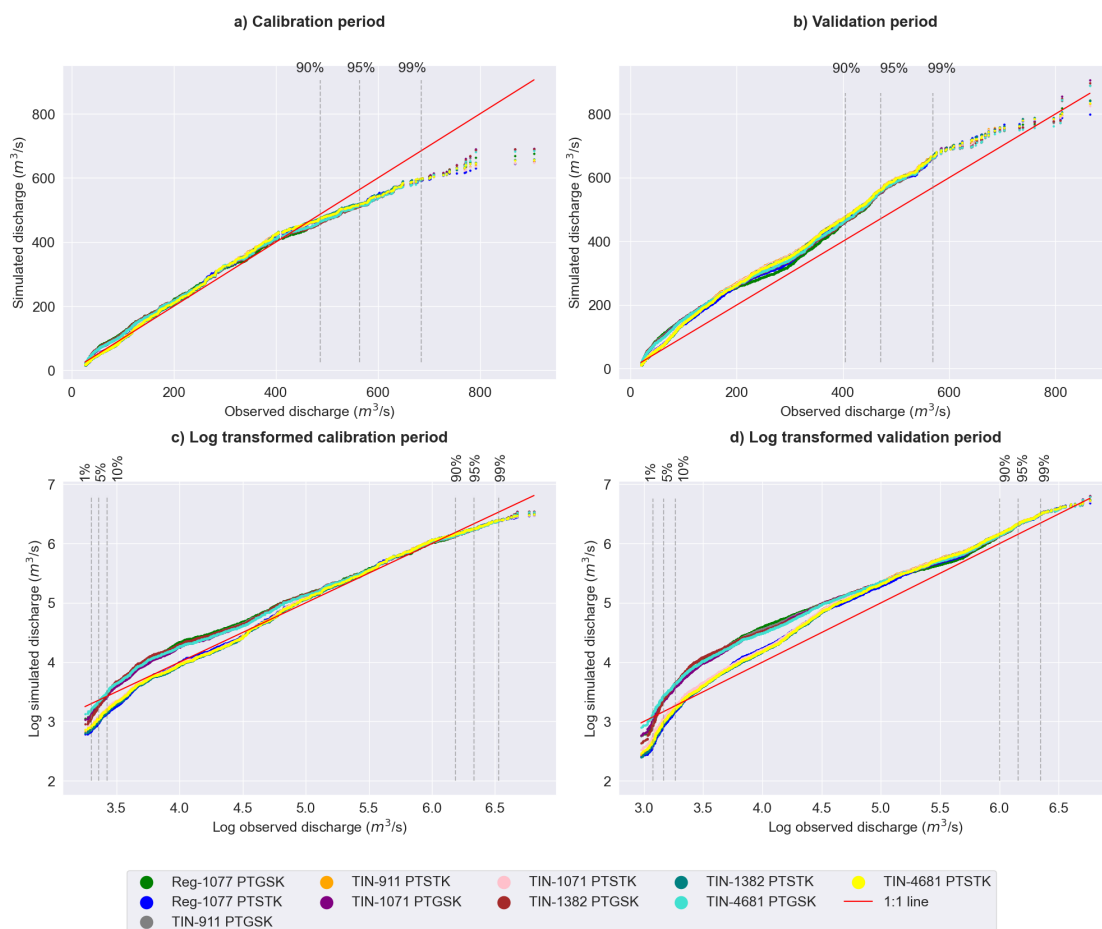


Figure 6.24: QQ-plots over daily observed and simulated discharge during calibration (2000-2004) and validation (2005-2015). The two upper QQ-plots show the a) calibration period and b) the validation period. The two lower QQ-plots are shown on a log-scale. The 1%, 5%, 10%, 90%, 95% and 99% quantiles for the observed discharge are represented by the dashed vertical lines.

Figure 6.25 shows observed discharge [m^3/s], simulated mean daily precipitation [mm/day], simulated mean daily temperature [$^{\circ}\text{C}$], simulated discharge [m^3/s] and simulated snow water equivalent (SWE) [mm] in the year 2011 and 2013. The year 2011 is a year where the NSE is low (0.58 for PTGSK and 0.60 for PTSTK using Reg-1077) and 2013 where it is high (0.93 for PTGSK and 0.94 for PTSTK using Reg-1077). The temperature, precipitation and SWE is from the PTGSK model only.

Sub-figure a) shows the low performance year 2011. In the winter months, there are two prominent spikes in precipitation, indicating periods of intense precipitation events. These spikes likely correspond to snowfall events or heavy rainstorms during colder periods. The first spike produces an immediate response in the SWE, but the SWE is decreasing to its starting levels before the next precipitation event. This indicates that the snow here must have been located at lower elevations where the temperature is warm enough to melt out the snow. The snow after the second event does not melt out, but is contributing to the accumulation seen in early spring. The mean temperatures are well below zero in this time of year, indicating that most of the precipitation is snow. This is reflected in the simulated SWE, which is increasing from January to early March. There is also little response in the observed discharge indicating that this is the case. From March to May there is little response in the observed discharge to the precipitation events, but both simulated discharge series do respond. The simulated

6.2. Impacts of Terrain Representation and Snow Routines on Calibration and Model Performance: Discharge

discharge in the PTGSK model is higher than the PTSTK model during this period. Both models respond to the precipitation events occurring in this period, indicating that the simulated precipitation during spring is likely too high in this year. The simulated snowpack is decreasing from March to July, but the observed discharge is not rising notably before May, and with a rapid rise in June and July. The mean temperatures increase from January to August. The temperatures are only positive from July to end of September in this year. During the first half of the monsoon period, the PTGSK model overestimates more than the PTSTK, while the opposite is the case for the months August-October. During the monsoon period the simulated discharge is generally higher than the observed discharge. The precipitation is characterized by intense precipitation events. The peak discharge in the simulations also lag behind the peak discharge in the observed data. The simulated snow pack reaches a peak in August and decreases rapidly from October to January. The SWE is never zero, indicating glaciers in the catchment. The observed peak discharge in 2011 is not during the the summer months, but occurs in late September. This peak is much higher than the increase in simulated discharge seen about 1-2 weeks earlier. Since the simulated precipitation is decreasing during this period, and the temperature and SWE is relatively stable, it is difficult to see what is the main contributing factor(s) to this discrepancy without observational data.

Sub-figure b) shows the evolution of observed discharge [m^3/s], simulated mean daily precipitation [mm/day], simulated mean daily temperature [$^{\circ}\text{C}$], simulated discharge [m^3/s] and simulated snow water equivalent (SWE) [mm] in 2013 when the NSE is high for both models (0.93-0.94). During January the mean daily temperature drops substantially (from about -6 to -20 degrees Celsius). At the same there is a big precipitation event occurring at the same time. There is an immediate response in the SWE, which is increasing until late February after several smaller precipitation events. Most of the precipitation is likely snow as there is almost no response in the observed discharge during this period and the mean temperatures remain below zero. The SWE is decreasing quite substantially from late February to June, and reaches a level lower than in 2011 for the same period. There are some bigger precipitation events that the simulations respond to, but that the observed discharge barely responds to. The simulations times the rising discharge seen from late May to late July much better than in 2011, but the peak happens slightly earlier in the simulations than in the observed discharge in July. The precipitations events during the monsoon have generally lower values in 2013, than in 2011. The positive mean temperatures also last longer in 2013 than in 2011. There is considerable less snow accumulated in the high performance year, than in the low performance year. The peak discharge seen in September in 2011, is not seen in 2013 (happens in July), but there are to periods during August and September where the simulated and observed discharge are increasing before decreasing again. There is a big precipitation event in October which is also reflected in observed and simulated discharge (and SWE).

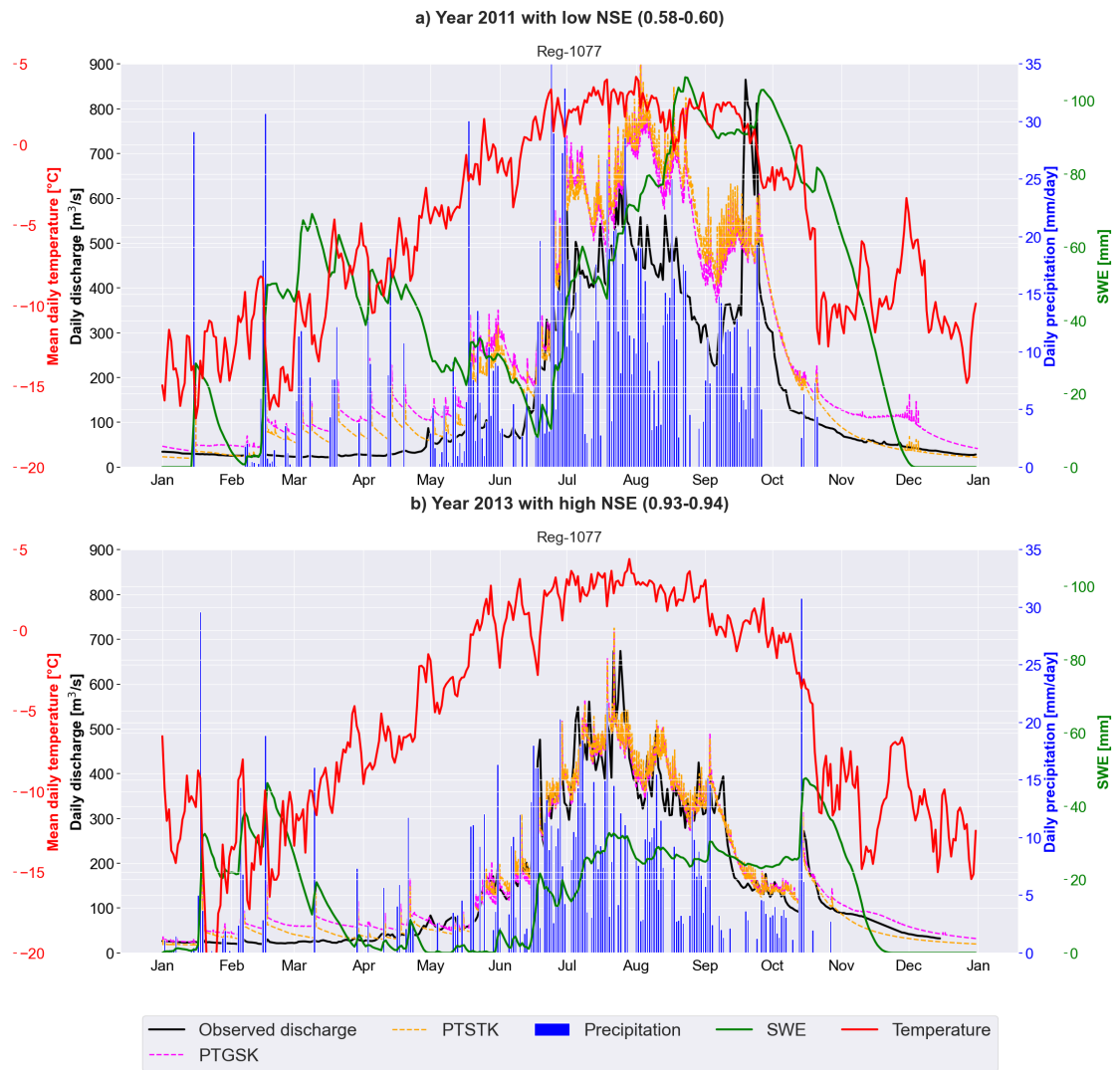


Figure 6.25: Observed discharge, mean daily precipitation, mean daily temperature and SWE in year 2005. Mean daily temperature [°C] is shown in red, daily precipitation [mm/day] in blue, SWE [mm] in green and observed discharge [m³/s] in black. The temperature and precipitation values are the interpolated values in Shyft for the region. The SWE value is calculated in Shyft using the calibrated PTGSK model for regular grid cells.

6.3 Impact of Terrain Representations and Snow Routines on Model Performance: SCA, SCD, and SWE and Glacier Melt

The aim of this section is to assess the impact of different terrain representations and snow routines on fractional snow-covered area (fSCA), snow-cover duration (fSCD), snow water equivalent (SWE) and glacier melt. The accurate representation of snow and its subsequent melt is important for understanding its contribution to runoff, making it essential to evaluate the performance of snow models.

Simulated fSCA and observed fSCA from MODIS are compared through visualisation, and evaluated using the Critical Success Index (CSI) (or Threat Score). Observed snow-covered area provides an additional means of assessing model performance. The Critical Success Index (CSI), which allows a quantification of the agreement between simulated and observed snow characteristics. The observed fSCA and fSCD are from the water years in the period 2000-10-01 to 2015-09-30. The simulated fSCA is therefore only selected for this period. It should be noted that the CSI is calculated by dividing the number of times both simulated and observed snow cover are both observed (hits) by the sum of hits + false alarms + misses. Therefore, the CSI is not affected by the number of non-events, meaning the times no snow is detected in neither the simulated or observed snow-covered area data. Therefore CSI is a biased score that is dependent upon the frequency of an event.

6.3.1 Evaluating Simulated SCA using the Critical Success Index and Observed SCA from MODIS

Figure 6.26 compares the modelled mean fractional snow-cover area (fSCA) [-] with the observed mean SCA [-] from MODIS during winter (DJF) and summer (JJA) in the period 2000-2015. The simulated values are from the PTSTK model using TIN-1071. The mean simulated fSCA is greater during both seasons than what is observed by MODIS. This is most clear to the North. During winter, the northern parts are almost completely covered with snow in the model (sub-figure a), while most of the same area is only partially snow-covered for MODIS (sub-figure b). This difference is also seen in the frequency plots, where most cells are snow-covered in the model, while most cells are not snow-covered in MODIS. Both simulated and observed mean fSCA values for DJF, match for the areas to the South where elevations are lower.

The frequency of cells with a high snow-covered area value, is even higher in the spring in the model. A majority of the cells in the model has an fSCA above 0.75. However, the observed data shows that most pixels are not snow-covered during spring, and that there are similar amounts of cells with the same fSCA values.

During summer (JJA) the differences are clearly visible. There are still large areas in the north that is snow-covered during summer in the model (sub-figure c), but very few areas in the MODIS data (sub-figure d). Even though most cells in the model are not snow-covered during summer, the frequency figure below c) shows that there many cells have a mean JJA fSCA value above 0.5. This is not what is observed by MODIS during summer. The majority of the MODIS pixels have a mean fSCA value for JJA below 0.25. The discrepancy between simulated and observed is most visible to the North.

In fall, slightly more cells with an fSCA value of 1.0, and slightly less have a value of 0, than in summer. The frequency distribution is also slightly more even than in summer. MODIS shows that most cells are not snow-covered during fall. The discrepancy is most visible to the North.

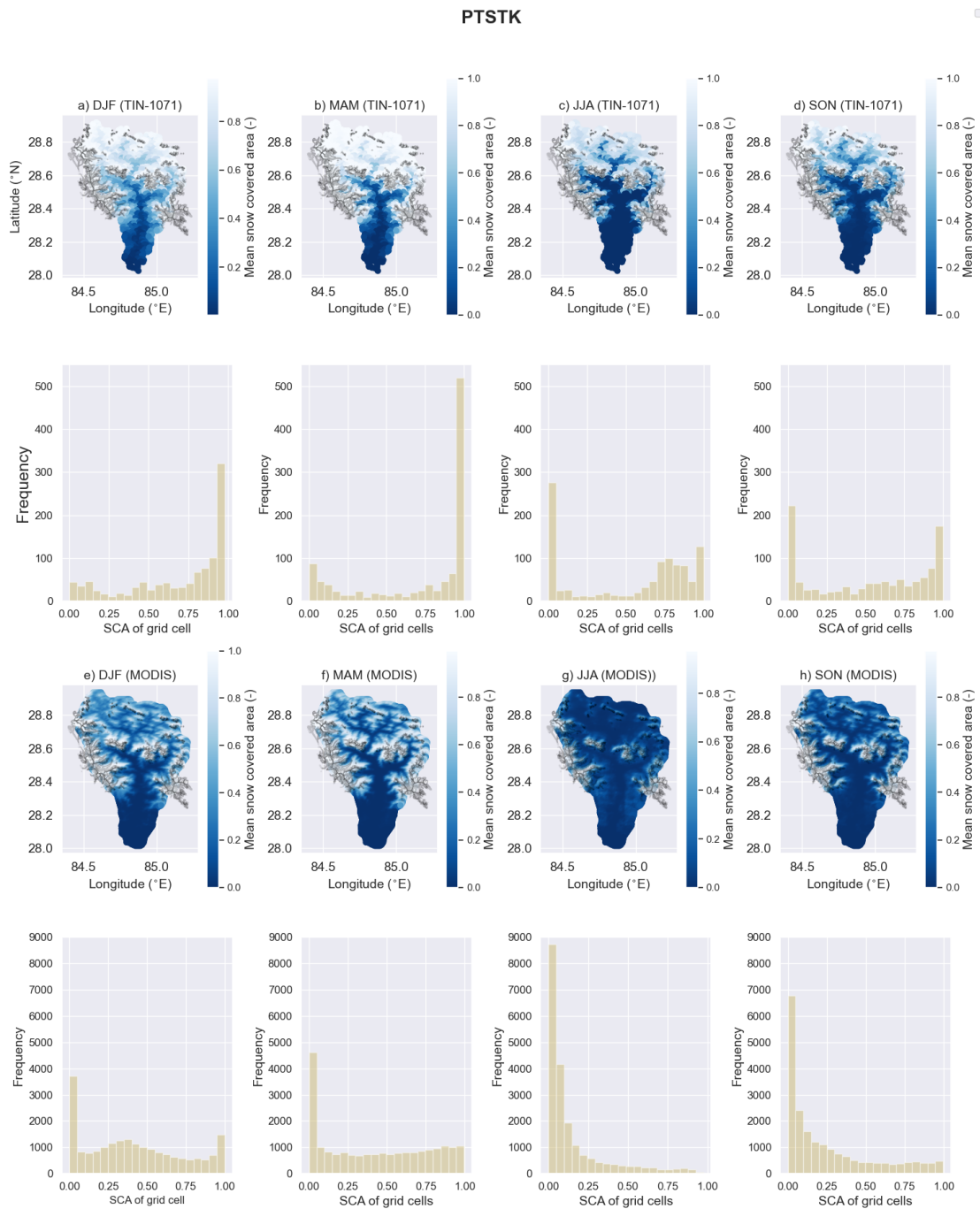


Figure 6.26: Comparison of mean fractional snow-covered area (SCA) between simulated (PTSTK) and observed (MODIS) during the winter season (DJF) and summer season (JJA) in the years 2000-2015. The polygons represent the glacier outlines from RGI 2017. Blue color represents no snow-cover and white represents full snow-cover.

6.3. Impact of Terrain Representations and Snow Routines on Model Performance: SCA, SCD, and SWE and Glacier Melt

Figure 6.27 shows the monthly Critical Success Index (CSI) for the PTGSK models evaluated for the water years 2000-2014 (2000-10-01 to 2015-09-30). The figure shows that there is great variation between the different months and within each month. The highest CSI is found for the the first 4 months of the year (sub-figure a) where the median CSI is between 0.6-0.8. However, there is a great variability for these months. The variability is also great in October, when the CSI is is very low (median below 0.3). The median CSI is below 0.2 for all months, except the 4 first months of the year. These results show that the PTGSK models perform poorly in simulating snow-covered area. This is also seen in the previous figure. The CSI is zero during summer indicating that the fractional snow-covered area is below the threshold value of 0.4 in both observed and simulated data, and therefore the number of hits are zero. Since true null events are not accounted for, the values are therefore low. The regular grid (Reg-1077) and the TIN with similar resolution (TIN-1071), have almost identical monthly CSI values. The effect of terrain representation on the fSCA is negligible.

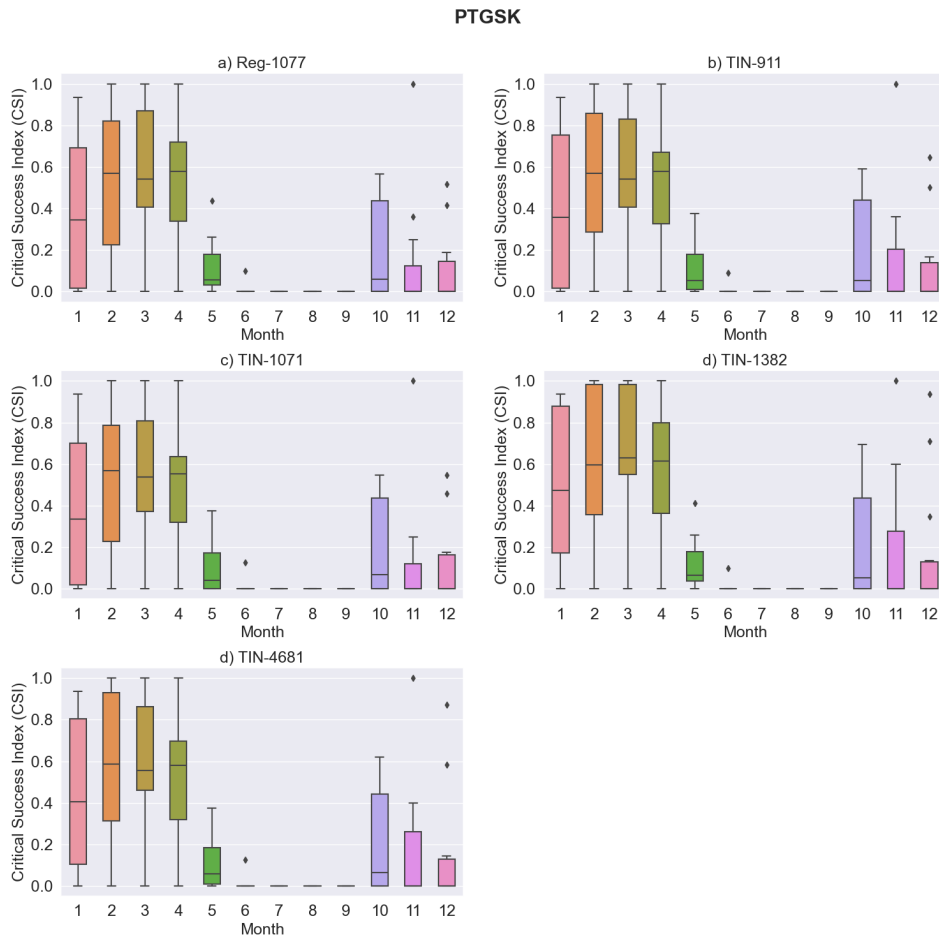


Figure 6.27: Monthly Critical Success Index (CSI) for PTGSK. The CSI is calculated for each month using mean fSCA from MODIS as reference data. The CSI is shown using boxplots. Boxplots is a standardised way of showing the distribution of data using a five number summary (minimum, maximum, median, first quartile and third quartile (Galarnyk 2022)). The boxes, called inter-quartile range (IQR), range from the 25th to 75th percentiles of the data. The horizontal line inside the boxes show the median value. The extended lines shows the "minimum" (25th percentile minus $1.5 \cdot \text{IQR}$) and "maximum" (75th percentile plus $1.5 \cdot \text{IQR}$). The dots (outliers) are values different from the rest of the data set.

Figure 6.28 shows the monthly CSI results for the PTSTK models. The monthly CSI for the PTSTK model follow a similar pattern as the PTGSK models, where the median CSI is highest for the first four months. The median CSI is slightly higher for the Snow Tiles models than the Gamma Snow models during these four months. The highest median CSI is found for TIN-1382 in March and April. The CSI is zero during summer indicating that the fractional snow-covered area is below the threshold value of 0.4 in both observed and simulated data. The CSI during October to December is similar to that of the PTGSK models. There is generally high variability during winter and spring, except for May (which is also much lower). The different terrain representations exhibit very similar patterns, and the effect of terrain representation is negligible.

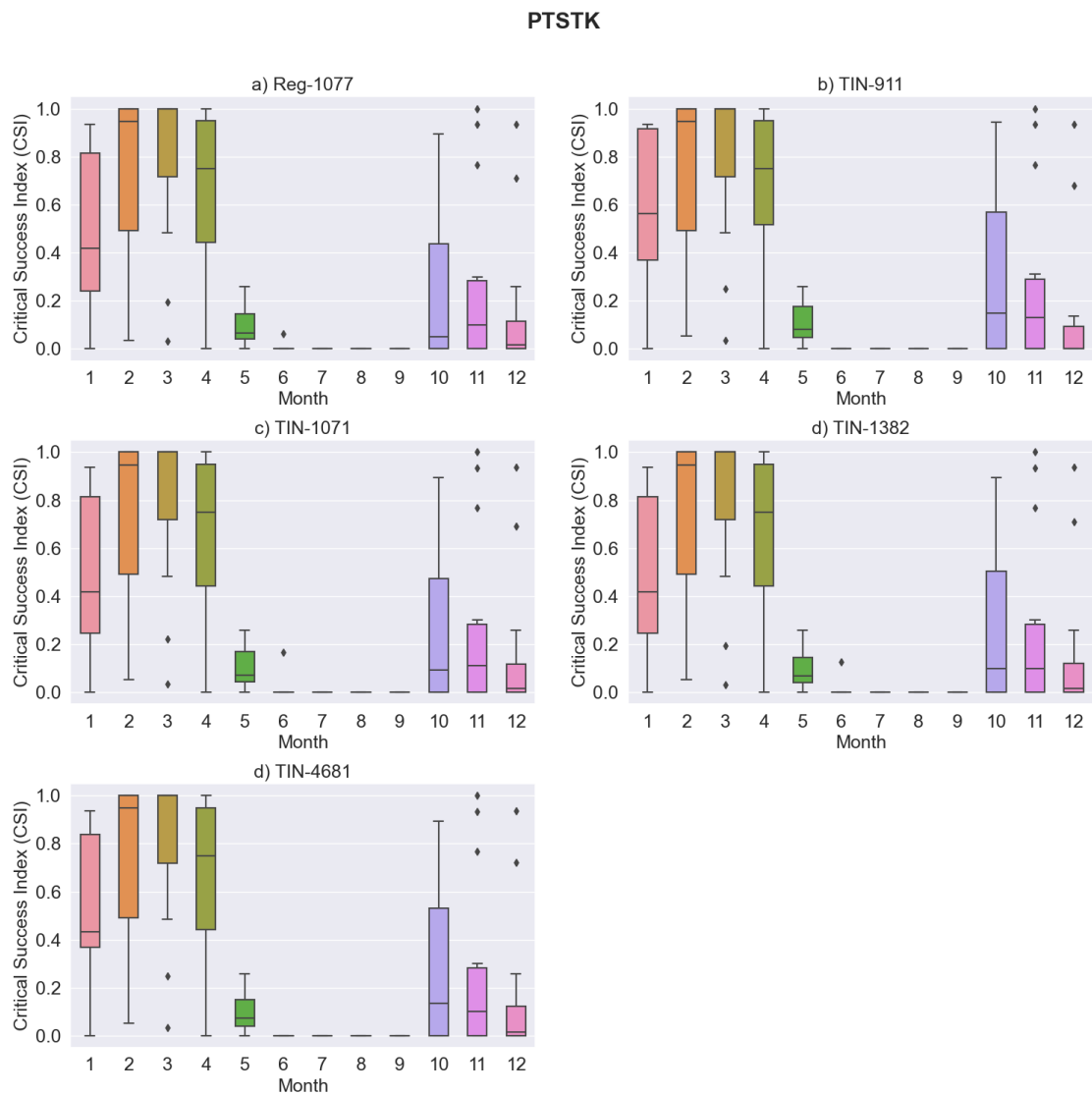


Figure 6.28: As in Figure 6.27, but for PTSTK.

6.3.2 Comparison between Simulated and Observed Snow-Cover Duration

Figure 6.29 shows the mean simulated fractional snow-cover duration (fSCD) plotted against mean observed fSCD by MODIS. The mean fSCD is the sum of daily mean fSCA of the catchment within a given water year. A mean fSCD value of 0 means that the catchment is snow-free during the entire year and a value of 1 means that the entire catchment is a glacier the whole water year. The mean fSCD from the MODIS data is given in water years (2000-10-01 to 2015-09-30). The figure shows that all of the models overestimate the snow cover duration. When comparing regular the PTGSK model with the PTSTK model using a regular grid (Reg-1077), it can be seen that the simulated fSCD derived from the Gamma Snow routine (PTGSK) model, has a closer fit to the observed fSCD from MODIS. When comparing regular grid (Reg-1077) with a TIN of similar resolution (TIN-1071) using the same model, it can be seen that the TINs only has a slight improvement compared with the regular grid in both models. Having higher resolution TINs do not improve the overall fit between simulated and observed values.

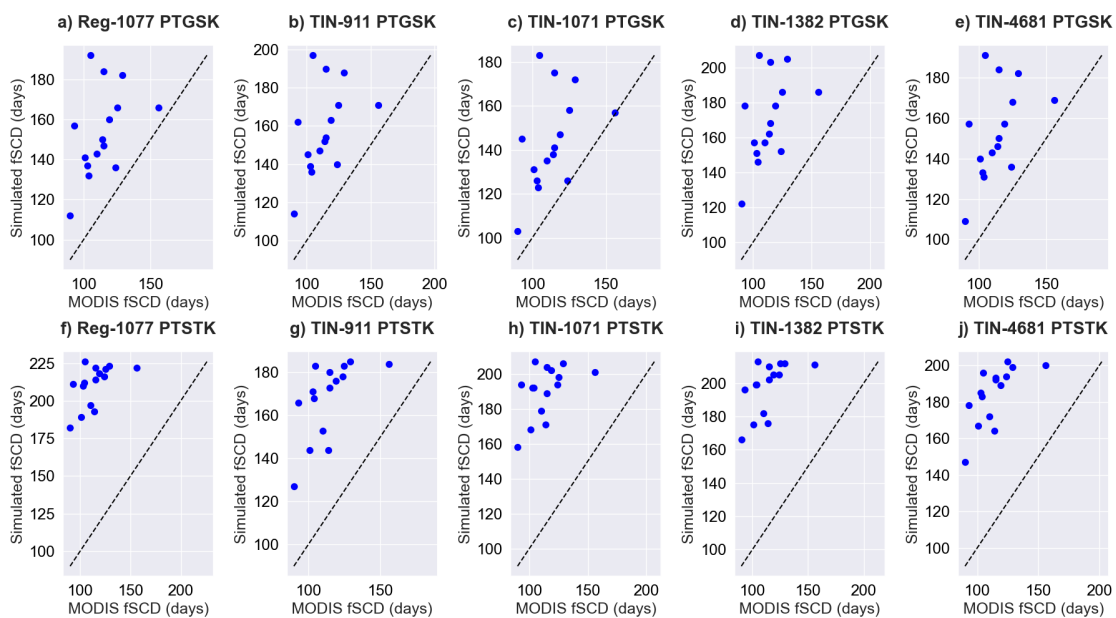


Figure 6.29: Simulated mean fractional snow-cover duration (fSCD) plotted against mean observed fSCD from MODIS. The diagonal line is the identity line (1:1 line). Points following the identity line represents a perfect fit. Each scatter point represents the mean simulated fSCD within a given water year plotted against the mean fSCD from MODIS in that water year.

Figure 6.30 shows the simulated snow water equivalent (mm) and glacier melt for all the simulations during the validation period (2005-2015). The snow water equivalent is shown in sub-figures a-b) and glacier melt in c-d). All of the PTGSK models in sub-figure a) (PTGSK model), follow an expected pattern where the SWE increases during periods when precipitation coincides with negative temperatures, before decreasing again during warm periods. Furthermore, there is considerable inter-annual variability in the SWE. The TIN with the highest resolution (TIN-4681) and TIN-1071 have the lowest SWE values throughout the period for the PTGSK models. However, the PTSTK models have a linearly increasing trend over the years, indicating that the catchment is becoming more glaciated. This is not what is expected, and the trend can be seen for all terrain representations. The regular PTSTK has the highest SWE values, while TIN-1071 with PTSTK has the lowest values. TIN-911 and TIN-1071 values are quite similar and considerable lower than TIN-1382, TIN-4681 and Reg-1077.

The peak glacier melt is higher in the PTSTK models, than in the PTGSK models. The regular grid has lower glacier melt than the TINs, but is has also much lower glacier fraction (0.16 vs. 0.34-0.35) (see Appendix F). The highest glacier melt is seen for the the TINs with the lowest resolution. The difference is largest between the models, than between the TINs (not accounting for the regular grid which has much lower glacier fraction).

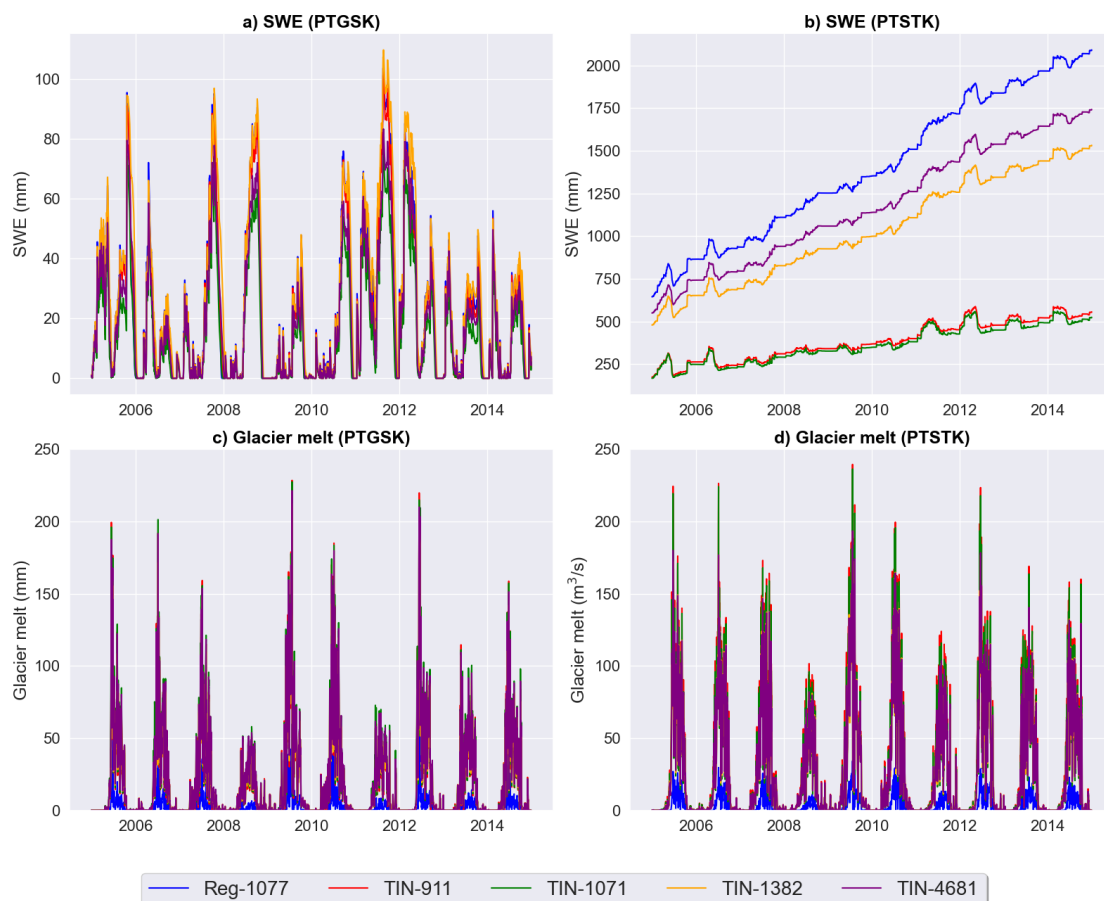


Figure 6.30: Snow water equivalent (SWE) [mm] and glacier melt [m³/s] for all simulations during validation period (2005-01-01 to 2015-12-31).

6.3.3 Comparing Simulated SWE and Glacier Melt between Models

To investigate further the unexpected finding of the increasing SWE in the PTSTK model, the difference in mean annual fSCA between year 2014 and 2001 is shown in Figure 6.31. The figure shows that the fSCA is increasing in for large portions of the northern part of the catchment. However, the mean fSCA is decreasing in many of the areas to the north along the borders of the catchment where elevations are higher. This is also where existing glacier are located (especially to the West). There is little difference in the lower elevations to the south. Since both PTGSK and PTSTK models have the same interpolated data, this suggests that the Snow Tiles routine performs poorly when simulating fSCA using the current configurations in the Budhi Gandaki catchment.

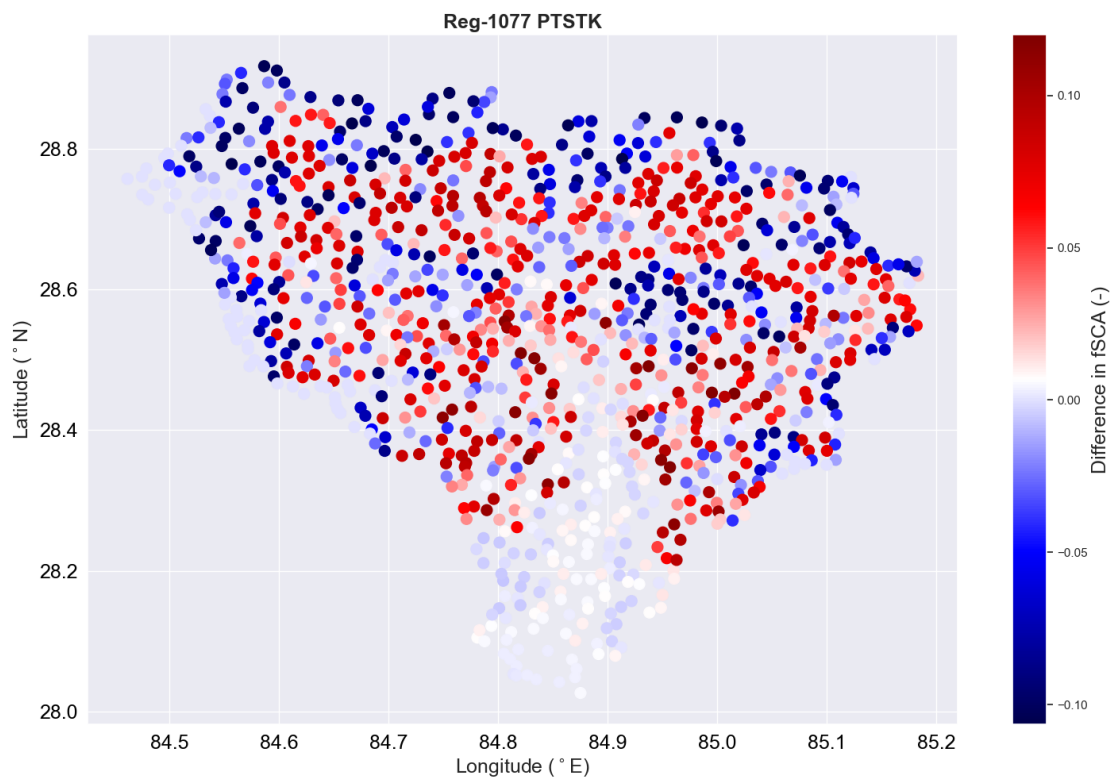


Figure 6.31: Difference in fractional snow-covered area between 2014 and 2001. Red values show an increase in fSCA values and blue values show a decrease. The figure is showing Reg-1077 PTSTK.

Chapter 7

Discussion

In the discussion section, the focus is on the findings and implications of the study, specifically regarding the evaluation of the Shyft modelling framework in simulating primary water balance components in the Budhi Gandaki catchment. This chapter will address each of the study's objectives and analyze their outcomes. Firstly, this includes a discussion of the spatial and temporal variation in the forcing data (temperature, precipitation, global radiation, relative humidity, and wind speed) which can have implications for the hydrology of the region and model performance. Secondly, the effect of different terrain representations and snow routines on model calibration and performance is evaluated using observed discharge as a target. Key evaluation criteria such as NSE, KGE, and RMSE are utilized to compare simulated and observed discharge. Lastly, the impact of different terrain representations and snow routines on simulated snow variables, including snow-covered area, snow-covered duration, snow water equivalent and glacier melt, is assessed. MODIS-derived observations and the Critical Success Index (CSI) are employed as evaluation criteria. These objectives provide valuable insights into the performance and capabilities of the Shyft modelling framework in simulating primary water balance components in a complex and data-limited catchment.

7.1 Analysis of Spatial and Temporal Variation in Forcing Data and Evaluation of Interpolation Methods

The analysis of the spatial variation in temperature [$^{\circ}\text{C}$] and precipitation [mm/hr] from the WFDE5 data in the period 1990-2019 reveals pronounced seasonal patterns. Firstly, there is a strong temperature gradient over land from the Tibetan Plateau in the North to India in the South. This is expected as temperature tends to decrease with height, and the Tibetan Plateau is located at higher elevations. There is also more land mass to the north than in the south, which cools off more quickly than liquid water. The Himalayas also acts as a topographical barrier, thus affecting the transport of warm moist air over the mountain range since much of the moisture will precipitate on the southern side before reaching the Tibetan Plateau. Furthermore, the results suggest that the temperature gradient is weakened during summer. This is also expected, since the Indian Monsoon causes widespread precipitation and cloud cover on the southern side of the Himalayas. The presence of clouds and increased moisture tends to limit daytime heating and temperature rise, resulting in a less pronounced temperature gradient.

The seasonal variation can also be seen for precipitation in the region. The study finds that there is little precipitation in HMA during DJF, MAM and SON compared with JJA. This is also what is found by (Lutz et al. 2014), where most precipitation occurs during the June-September with a strong north-south gradient, while the precipitation in the West is distributed more evenly during the year due to westerly and southwesterly flows.

The time series decomposition suggests that there is a linearly increasing trend in High Mountain Asia in the period 2000-2019. This is a global trend that is supported by various studies (Allen et al. 2022; Hock et al. 2019). The results also indicate a linearly decreasing trend in incoming shortwave radiation. This is supported by Padma Kumari et al. 2007, which found an average solar dimming over India in the period 1984-2004 of $\sim -0.86 \text{ W/m}^2$ per year.

The analysis for WFDE5 data for the Budhi Gandaki region reveals a cool climate for the region, with a mean of 5°C . The minimum and maximum values suggest however that there are extreme variations in temperature (from -36.69°C to 41.74°C). The mean precipitation is 0.16 mm/hr with a standard deviation of 0.62 mm/hr . However, the maximum precipitation of 24.20 mm/hr , which is far off 1 standard deviation, indicates the presence of dry and wet periods. This emphasizes the potential for drought and heavy precipitation events in the region which is increasing the risk of flooding and landslides. Similarly, the variation in wind speed, relative humidity and radiation is also great, illustrating the extreme variations in weather conditions in the region.

Temperatures are found to be lower in the North and higher in the South of the catchment. Since the elevation is increasing from south to north, this is to be expected. The weakening of temperature gradient seen in the High Mountain Asia during JJA, is also found in the catchment. There is little precipitation in the catchment during DJF, MAM and SON compared with JJA. The WFDE5 data is also able to capture a gradient in precipitation during the Monsoon, where most of the precipitation occurs to the South and least to the North. However, there are only four grid cells covering the catchment, which is a coarse resolution for doing an impact study where the elevations within short distances as seen in chapter 3. Most of the snow occurs to the North in Budhi Gandaki, where the mean accumulated snowfall is highest during spring. This is likely due to several factors coinciding factors such as orographic lifting (moist air rises, cools, condenses and then precipitates) (Hunt et al. 2022), atmospheric instability happening during the transition from winter to summer with development of convective clouds and precipitation (Lai et al. 2021) and the influence of Western Disturbances (Lai et al. 2021). The wind speed does not change notably across the seasons in the catchment, and the mean seasonal wind speed (2000-2015) is similar for all four WFDE5 grid cells covering the catchment. The reanalysis incoming shortwave (SW) radiation reveals marked seasonal variations. The winter months experience the lowest SW radiation, while the highest values are found during MAM at the Tibetan Plateau. This is to be expected, as the SW radiation is lower during DJF on the Northern Hemisphere, and with higher latitudes ($^\circ\text{N}$). During spring the high radiation seen to the North can be explained by factors such as lower cloud coverage than to the South. During JJA there is lower SW radiation to the South than to the North, which is also when cloud coverage is increasing as the Monsoon occurs. The WFDE5 reanalysis reveals that there are great variation in relative humidity across the WFDE5 grid cells. The summer has considerably higher relative humidity than the other seasons. Relative humidity is affected by several factors, such as temperature which determines the holding capacity for moist air, water vapor content, proximity to water bodies, atmospheric circulation

and vegetation.

The results suggest that temperature is increasing with height in the catchment. However, there are too few data points to determine a robust lapse rate. Based on a 20 years record of surface air temperature from 56 stations in on the southern slope of Nepal at elevations from 72 to 3920 m Kattel et al. 2013, finds that the temperature lapse rate exhibits bi-modal pattern, where it has two maxima in the pre- and post-monsoon season, and two minima in the winter and summer seasons. In this study, a constant temperature lapse rate of $-0.6^{\circ}\text{C}/100\text{ m}$ is used, but this is likely too low value in summer when a temperature lapse rate of $-5.1\text{ }^{\circ}\text{C}/100$ is found in this study. Precipitation is increasing with height in the catchment from elevations between 1258 m and 2163 m. This is expected as precipitation generally increases with height as moist air rises, condenses and then eventually precipitates out of the cloud. However, for higher elevations the precipitation decreases. Immerzeel, Petersen, et al. 2014 finds that there is great variability in precipitation gradients over short horizontal distances, and that an uniform valley scale precipitation gradient cannot be established. Bookhagen and Burbank 2006 finds a two rainfall peaks along the Himalaya, one at around 1200 m and a second at 2100 m, at higher elevations the precipitation decreases. The mean annual wind speed [m/s] is found the be quite similar for heights between 1258-4532 m, and to be substantially higher at the Tibetan Plateau. Relative humidity is decreasing with height, which can be explained by factors lower temperatures. The increase in SW radiation with height can be explained by less cloud coverage to the North. The great inter-annual variation found in observed discharge can be explained by inter-annual variation in the intensity of the Monsoon and variation in temperature and precipitation that snow accumulation and ablation.

Temperature was interpolated using Bayesian Temperature Kriging (BTK) and the other variables with Inverse Distance Weighting (IDW) (subsubsection A.2.2). The Inverse Distance Weighting (IDW) method has several advantages since it is a simplistic method that doesn't require complex calculations. It is also flexible, since the parameters can be set by the user of Shyft, and intuitive since the interpolated values are directly influenced by distances and values of neighbouring points. The weights in the IDW do also not change over time. However, it is sensitive to data distribution. There are only 16 WFDE5 grid cells, but these are spread across the leeward and windward sides of the Himalayas, but it remains uncertain how representative these few data points are for the Budhi Gandaki catchment. For example, as seen in chapter 3, most of the catchment area has an elevation higher than 4500 m, but the four WFDE5 grid cells closest to the catchment in south-north direction has an elevation much lower than this. The IDW method can also be influenced by outliers which can distort the interpolation results. Furthermore, the different parameter sets has not been investigated for IDW in this study. The parameter values have been chosen based on a previous in the region using IDW and Shyft, but preferably other parameter sets should be investigated. Since IDW is deterministic, there is no estimation of uncertainty using this method.

Alternatively, to the deterministic IDW method kriging can be used. This is what is done for temperature with the BTK method. The advantage of kriging is that the weights are not only a function of the distances between measured and prediction points, but also the overall spatial arrangement between the points. The method quantifies the auto correlation between between points to use in the weights. The kriging method is therefore able to minimize prediction errors by estimating optimal weights based on spatial auto correlation. However, kriging assumes stationarity. From the analysis of the forcing data, the temperatures are increasing during the period 1990-2019, and the

temperature is therefore not stationary. The kriging method can be reliable within the range of the sample data, but has the same issue as the IDW method of whether the data points it uses to interpolate are representative or not for the region.

The interpolated precipitation in the Shyft models is exponentially decreasing with height, but as discussed above, this should not be the case, as the precipitation should be increasing with height until a certain height, before decreasing with higher elevations. This suggests that the IDW method used here is not reflecting the expected precipitation pattern with height. The interpolated temperature is linearly decreasing with height, which may be a fair assumption on an annual basis, but it does not capture the bimodality in the temperature lapse rate mentioned previously. During summer the atmosphere is more moist, releasing more latent heat, and hence making the lapse rate less steep. The onset of the monsoon control in large parts the variability and the marked seasonality in air temperature, while the spatial variation can be influenced by the factors such as topography, valley and katabatic winds and debris on glacier tongues (Immerzeel, Petersen, et al. 2014).

7.2 Impacts of Terrain Representations and Snow Routines on Calibration and Model Performance: Discharge

The results suggest that the most of the parameters have great range of optimized values. The Min BOBYQA searches for an optimal value within a neighbouring set of candidate solutions. However, there is an uncertainty to whether the optimized values are global minimum values. There is also an uncertainty to whether there are more than one plausible solutions to optimized parameters. Since the Snow Tiles routine has less parameters than the Gamma Snow routine, this uncertainty should be lower for the Snow Tiles routine.

Earlier studies has shown that the PTGSK model is particularly sensitive to the Kirchner coefficients, precipitation correction factor and the threshold temperature (Teweldebrhan et al. 2018). The snow routine in the PTSTK model has not been tested before in this region. The calibration results suggest that there is variation between models and discretization methods for the temperature threshold, kirchner c1 and precipitation correction factor parameters. There is also variation between the regular grid cell and the TINs in the direct response of glacier melt. It has been pointed out that model simulations are in particular sensitive to the precipitation scaling factor as it may account for under catch and lack of representative stations (Lakew et al. 2017). The Kirchner response parameters are more dependent on the physical characteristics of the catchment and thereby less sensitive to the forcing data (Bhattarai, Burkhart, Tallaksen, et al. 2020), but that they can highly influence a given simulation (Engeland et al. 2016; Teweldebrhan et al. 2018).

The results show that the models using a terrain representation with lower resolution, have a lower calibration run time. This is to be expected as there are fewer simulation elements. The Snow Tiles routine is faster than the Gamma Snow routine for all terrain representations. This is also expected since there are less parameters to calibrate. There is no time advantage in using TIN-911 compared with the regular grid (1077 cells) on calibration run time, as the lowest resolution TIN has a similar calibration run time as the regular grid for the Snow Tiles routine and slightly less for the Gamma Snow routine. However, a TIN with an even lower resolution should be tested. The results suggest that the calibration run time increases relatively more than the increase in resolution for TINs

with greater resolution. For the highest resolution TIN, there is a clear advantage in run time for the Snow Tiles routine in absolute numbers.

All of the simulations show good a match between observed and simulated discharge (NSE between 0.88-0.90) during the calibration period, but a lower value for the validation period (between 0.77-0.80). The PTSTK model stack has a slightly higher NSE for the validation period overall. The difference in performance in the validation period between the models is even greater using the LnNSE. When comparing two simulations that use the same discretization. The results indicate that the PTSTK models perform better for low flows than the PTGSK . Furthermore, the difference between snow routines is greater than between terrain representations. The KGE is high for the calibration period (0.88-0.93), but drops to about 0.70-0.75 for the validation period.

The notably lower performance results for the validation period than calibration period could be due to the way the calibration and evaluation has been performed. The models are calibrated daily from 2000-01-01 to 2004-12-31 using discharge as target. The parameters are either prescribed or calibrated within physical limits. Min BOBYQA as optimization method and NSE as objective function. After this is done, new simulations are run hourly using optimized parameters. Next, the simulations are resampled to daily resolution. The observed and simulated series are then split into a calibration period (2000-01-01 to 2004-12-31) and validation period (2004-12-31), before using the NSE, LnNSE, KGE and RMSE as evaluation criteria. The fact that the calibration is done daily, but then run on an hourly resolution after calibration and re sampled to daily, could imply that the hourly variation in simulated discharge propagates as noise into the resampled daily simulated discharge. This will affect the model performance. Furthermore, the calibration is done for 5 years, validated for 10 years. If there are any trends during this period, this may affect model performance. As seen from the daily observed and simulated discharge figure ([Figure 6.22](#)), the 365-day moving average did have a slight decrease from 2000 to 2007, which covers the entire calibration period.

The analysis of yearly NSE, indicate that there is greater variation in performance within each snow routine, than between snow routines during years were the performance is low. On the other side, the LnNSE shows that the difference in model performance is greater between the snow routines. The overall performance using NSE and LnNSE is greater for the Snow Tiles routine, especially during low flow. In the year with the lowest NSE (2011), the Gamma Snow routine overestimates the discharge more than the Snow Tiles routine during the pre-monsoon season and lags behind the rising limb of the observed discharge during June-August by about 1-2 weeks. Furthermore, the model is not able to capture the peak observed discharge well during this year (which arrives in mid September). The fact that the observed discharge has a peak late into the monsoon season this year, but that none of models simulate a similar peak, could imply that a heavy local rainfall event occurred during this year which is not reflected in the interpolated data. The interpolated temperature suggest there was no rapid increase in temperature during this period could contribute to this rapid rise through snow melt and glacier melt. During the year with a high NSE, PTGSK still overestimates the discharge more than the PTSTK model during winter and spring.

The analysis of years with low and high NSE, shows that models respond quickly to precipitation events during winter and spring. However, the same response is not seen in the observed discharge. The SWE is also increasing during these precipitation events. The rapid response in simulated discharge could mean that interpolated precipitation is too high during winter, or that temperatures at lower elevations are high enough

to melt more snow than it is. However, the decrease in SWE during late February to April this year has a moderate effect on the simulated discharge. This suggests that the precipitation is driving these responses in discharge, not snow melt.

The daily discharge plots shows that a general good agreement between the observed and simulated discharge, but that the models tend to perform worse for peak discharge. The PTSTK models perform better for low flow. The models simulate higher discharge during spring than what is observed. This could be due to many reasons such as a higher estimation of snow than it is, that the threshold for onset of snow melt is too high, that the temperatures are too high, that the precipitation is too high, a too low surface albedo or that the degree day factor is too high. No registered hydro power is registered in the catchment during the study period, which can affect the results. However, smaller scale community driven hydro power can affect the river discharge.

7.3 Impacts of Snow Routines and Terrain Representations on Model Performance: Snow-Covered Area, Snow-Cover Duration, Snow Water Equivalent and Glacier Melt

Snow can strongly modulate the terrestrial surface energy balance given its unique properties (high albedo, large water storing capacity and low thermal conductivity). The PTSTK models to perform poorly when snow-covered area (fSCA) is evaluated against observed fSCA by MODIS. The simulated fractional snow-covered area in the PTSTK models is higher than what is observed by MODIS for all seasons. The comparison between simulated and observed fSCA show that the overestimation is greatest to the North. The Critical Success Index (CSI) indicate that both models perform poorly, but that the PTSTK models perform slightly better during spring. However, the CSI is not taking into account True Null Event (where both simulations and MODIS show no snow), which can explain the low values during summer. The CSI is also sensitive to threshold, so the performance should be evaluated against other metrics as well such as the RMSE (Aalstad et al. 2020). The results compare poorly with the monthly CSI calculated for the Marsyangdi-2 river catchment, which is a tributary of the Narayani river catchment (Bhattarai, Silantyeva, et al. 2020). According to Orsolini et al. 2019, many reanalyses tend to overestimate snow cover and snow depth over the Tibetan Plateau due to excessive snowfall. The ERA5 do not consider Interactive Multisensor Snow and Ice Mapping System (IMS) snow-cover for high altitudes, and the study suggests using satellite snow-cover and in-situ snow depth observations in data assimilation for future reanalyses in the region.

The PTSTK model is a temperature-index model, which implicitly assumes that temperature is driving ablation. It assumes that snow melt is a linear function of the difference between daily mean temperature and air threshold temperature (Litt et al. 2019). This is a well established relationship for lower elevations, but at high elevation glaciers incoming shortwave radiation is the dominant energy input and the surface energy balance only partly relates to daily mean temperature (Litt et al. 2019). The Gamma Snow routine is therefore better suited for modeling the net surface energy balance at high elevations, since it accounts for the albedo decay rate which modulate the energy that is absorbed and reflected, and takes into account turbulent wind. Furthermore, the shape parameter in the Snow Tiles model is fixed at the default value (2.0) in Shyft, but this parameter should also be included in calibration.

7.3. Impacts of Snow Routines and Terrain Representations on Model Performance: Snow-Covered Area, Snow-Cover Duration, Snow Water Equivalent and Glacier Melt

The modelled snow water equivalent (SWE) is found to be increasing over the years in the PTSTK, indicating that the catchment becomes more glaciated. This is not what is expected. A study done in the neighbouring Langtang catchment (Nepal), found that the seasonal evolution of snowpacks using observed ground data between May 2016 and June 2018 had a longevity of 5-6 months before SWE fully disappearing (Kirkham et al. 2019). They found that the SWE at Ganja La at 5000 m.a.s.l. in the Langtang catchment reached a peak SWE value of 200 mm w.e. in March which is more in line with some of the peaks of the PTGSK models, but where the PTSTK models are far off. The inter-annual variability found for both models agree with agree with previous findings (Tiwari et al. 2016).

Non-melt processes of ablation such as sublimation is not well accounted for in neither of the models. The sublimation in the region has so far remained unquantified, making it harder to understand the water balance in the region (Stigter et al. 2018). In a study by Stigter et al. 2018 on the Yala Glacier in Langtang Nepal, the fraction of the annual snowfall that returns to the atmosphere by sublimation and evaporation was found to be 21%, which is high compared to other regions in the world.

Chapter 8

Conclusions and outlook

The first objective of this study was to analyze the spatial and temporal variation in forcing data, including temperature, precipitation, global radiation, relative humidity, and wind speed. The analysis reveals pronounced seasonal patterns, such as a strong temperature gradient over land from the Tibetan Plateau in the North to India in the South. This gradient is weakened during summer due to the Indian Monsoon causing widespread precipitation and cloud cover on the southern side of the Himalayas. The precipitation in the region also shows seasonal variation, with little precipitation during DJF, MAM, and SON compared to JJA. The analysis also indicates a linearly increasing trend in temperature and a linearly decreasing trend in incoming shortwave radiation in High Mountain Asia. The results suggest that the interpolated precipitation does not reflect the expected pattern with height, and the interpolated temperature captures a linear decrease with height but does not capture the bi-modality in the temperature lapse rate. The calibration run time is lower for models with lower-resolution terrain representations, and the Snow Tiles routine is faster than the Gamma Snow routine, especially for higher resolutions.

The second objective was to evaluate the effect of different terrain representations and snow routines on model calibration and performance using observed discharge as a target. The evaluation criteria used are NSE, KGE, and RMSE to compare simulated and observed discharge. The results show that most parameters have a wide range of optimized values, and there is uncertainty about whether the optimized values are global minimum values. The calibration results suggest variation between models and discretization methods for certain parameters. The performance of the models is sensitive to parameters such as precipitation correction factor, Kirchner coefficients, and temperature threshold. The models using a terrain representation with lower resolution have a lower calibration run time, and the Snow Tiles routine is faster than the Gamma Snow routine. Overall, all simulations show a good match between observed and simulated discharge during the calibration period, but with lower values during the validation period. One explanation for the lower performance during the validation period is that the models are calibrated daily, simulated daily with optimized parameters, and then resampled and evaluated. The PTSTK model stack performs slightly better than the PTGSK model stack with discharge as target, and the difference between snow routines is greater than between terrain representations.

The third objective was to assess the impact of different terrain representations and snow routines on simulated snow variables, including snow-covered area, snow-covered duration, snow water equivalent, and glacier melt. MODIS-derived observations and the Critical Success Index (CSI) are used as evaluation criteria. The simulations show poor performance using CSI as evaluation criteria for both models. Other metrics should be used as well for assessing the performance. The CSI is sensitive to the threshold and does not take into account true null events. The models overestimate the snow cover duration. The simulated SWE in the Gamma Snow routine show an expected pattern, where seasonal snow melt out. The Snow Tiles routine show a linearly increasing trend, indicating that the catchment becomes more glaciated. This is not expected, and is likely due to its inability to take into account for snow dynamics at higher elevation where incoming shortwave radiation is the dominant energy input. The simulated glacier melt is lower in the regular grid, than for the TINs. This can be explained by the relatively lower glacier fraction in the regular grid.

Overall, the objectives provide valuable insights into the performance and capabilities of the Shyft modelling framework in simulating primary water balance components in a complex and data-limited catchment. The forcing data used in this study includes 16 grid cells, where 4 of these cover the catchment. This is a coarse resolution, and future studies should look into using downscaling tools such as [TopoScale](#), which uses the well-resolved description of the atmospheric column provided by climate models, together with high-resolution digital elevation models (DEMs), to downscale coarse-grid climate variables to a fine-scale subgrid (J. Fiddes and Gruber 2014).

Appendix A

Shyft modelling framework

A.1 The architecture of Shyft

The Shyft platform is distributed through three GitLab repositories: [shyft](#), [shyft-data](#) and [shyft-docs](#). The shyft repository contains the source code for the Shyft platform, while the shyft-data repository contains the test-data needed for the full runs and demo. The shyft-docs repository contains the end-user documentation for the Shyft platform. Additionally, there is also a [Dockers](#) repository that contains the Docker build recipes for the Shyft eco-system. The source code of Shyft is published under version 3 of the GNU General Public License (GPLv3). The end-user documentation on installation (Linux/Windows), how to use the Shyft API to construct a hydrological model, how to use Shyft repositories and the use of Shyft orchestration can be found at [Shyft: Read the Docs](#) (Statkraft 2022).

The Shyft platform is built on top of C++ and Python (Burkhart et al. 2021). Basic data structures, hydrologic algorithms and models in the Shyft core is written in C++ to provide high computational efficiency (Zehra et al. 2020). The C++ code is then compiled into a Python extension module, which is then imported into Python. The Shyft API which is coded in Python is the main interface for the user to interact with the Shyft platform. [Figure A.1](#) summarizes the architecture of the Shyft platform.

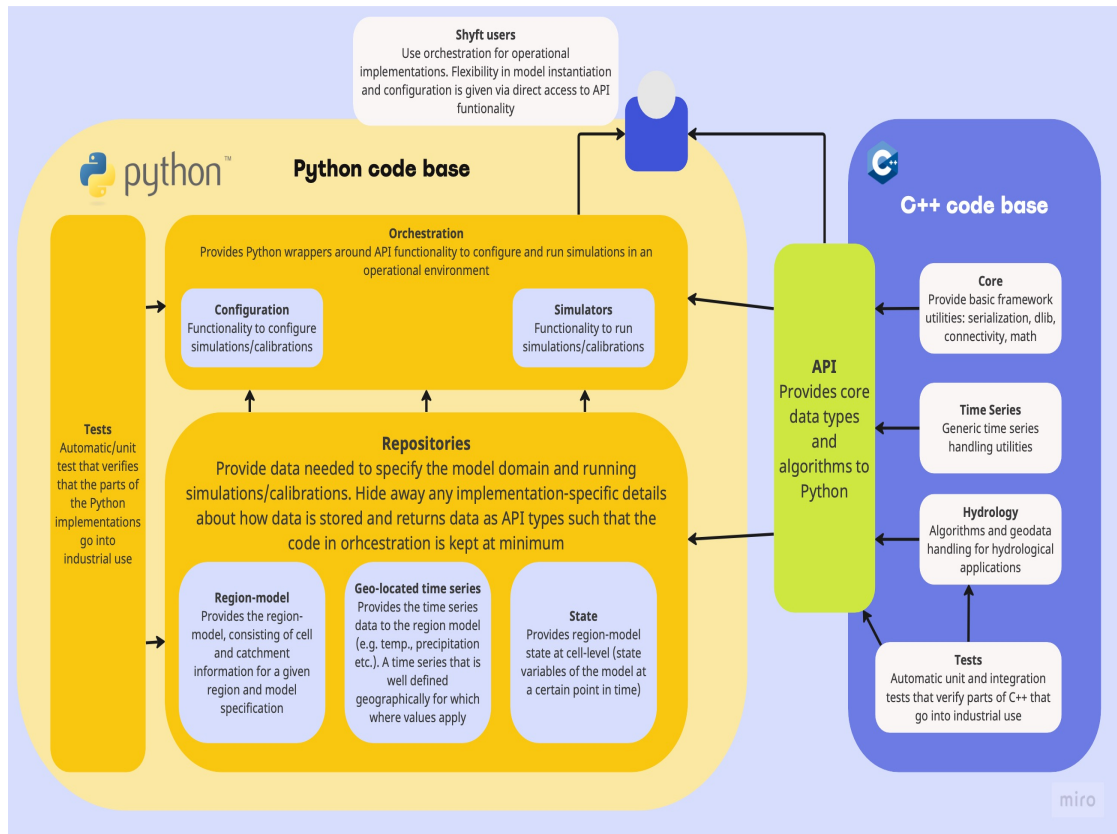


Figure A.1: Shyft architecture. The yellow part of the figure represents Shyft components written in the Python language, while the blue represents the C++ code base. The green API is in Python. The figure is made using [Miro](#), and is an adaptation of the figure in (Burkhart et al. 2021).

A.1.1 C++ core

The C++ code base contains several separate code folders (Burkhart et al. 2021). The core folder handles framework-related functionality such as serialization; the hydrology folder includes hydrological algorithms and the time series folder handles generic time series.

A.1.2 Python API

The Shyft API is providing the relevant Shyft Core implementations that are needed to configure and utilize models in Python (Burkhart et al. 2021). Figure A.2 gives an overview of the main Shyft API types. The idea behind the Python API is that users can write pure Python code without being exposed to the C++ code base. The API allows the user to configure and run a model and access data at various levels. Once a model is initiated, the data is stored at a random access memory of the computer, and the user can access the data using interactive shell commands like Jupyter Notebook or IPython. Thus, the Shyft API allows the user to interactively configure the model, provide forcing data to it, run the model, extract and plot variables. Model parameters can easily be updated once the model is instantiated in this way.

shyft.time_series is an API type that provides mathematical and statistical operations and functionality for time series (Burkhart et al. 2021). Time series can be

either expressions or concrete time series. Expressions are time series that are defined by a mathematical expression, while concrete time series are time series that are defined by a set of points in time. All time series have a time axis (set of non-overlapping periods), values and a point interpretation policy. The point interpretation policy defines how the value of a time series is interpreted at a given point in time, and can be either an instant value, average value or sum value.

api.GeoCellData represents common constant cell properties, regardless of model and cell assemblies. The *GeoCellData* object contains the following properties: elevation, latitude, longitude, slope, aspect, land type, soil type, land cover fraction, soil moisture, snow cover fraction, and the cell area.

Cell contains *GeoCellData* and *TimeSeries* of model forcings (*api.GeoPointSource*) (Burkhart et al. 2021). The *Cell* is also specific to the *Model* selected. The *CellVector* is a container for all the cells.

RegionModel is a container for all the *Cells* and the *ModelParameters* at region and catchment level. Since everything is vectorized, the *ModelState* collects all the states for each cell (Burkhart et al. 2021). The region model provides functionality like: initialization (*Model.initialize_cell_env(...)*), interpolation (*Model.interpolate(...)*), simulation (*Model.run_cells()*), calibration (*Optimizer.optimize(...)*). The *GoalFunction* that is being minimized is based upon the *TargetSpecification* in the optimizer.

Regionenvironment contains all the *Source* vectors of a certain type (e.g. precipitation, temperature, etc.) (Burkhart et al. 2021).

Appendix A. Shyft modelling framework

Figure A.2 shows an overview of the main Shyft API types. Broadly speaking these include the *api.GeoCellData* in blue, the *shyft_time_series* in green and the *cell* in yellow, and other support classes that are used to configure and run a model.

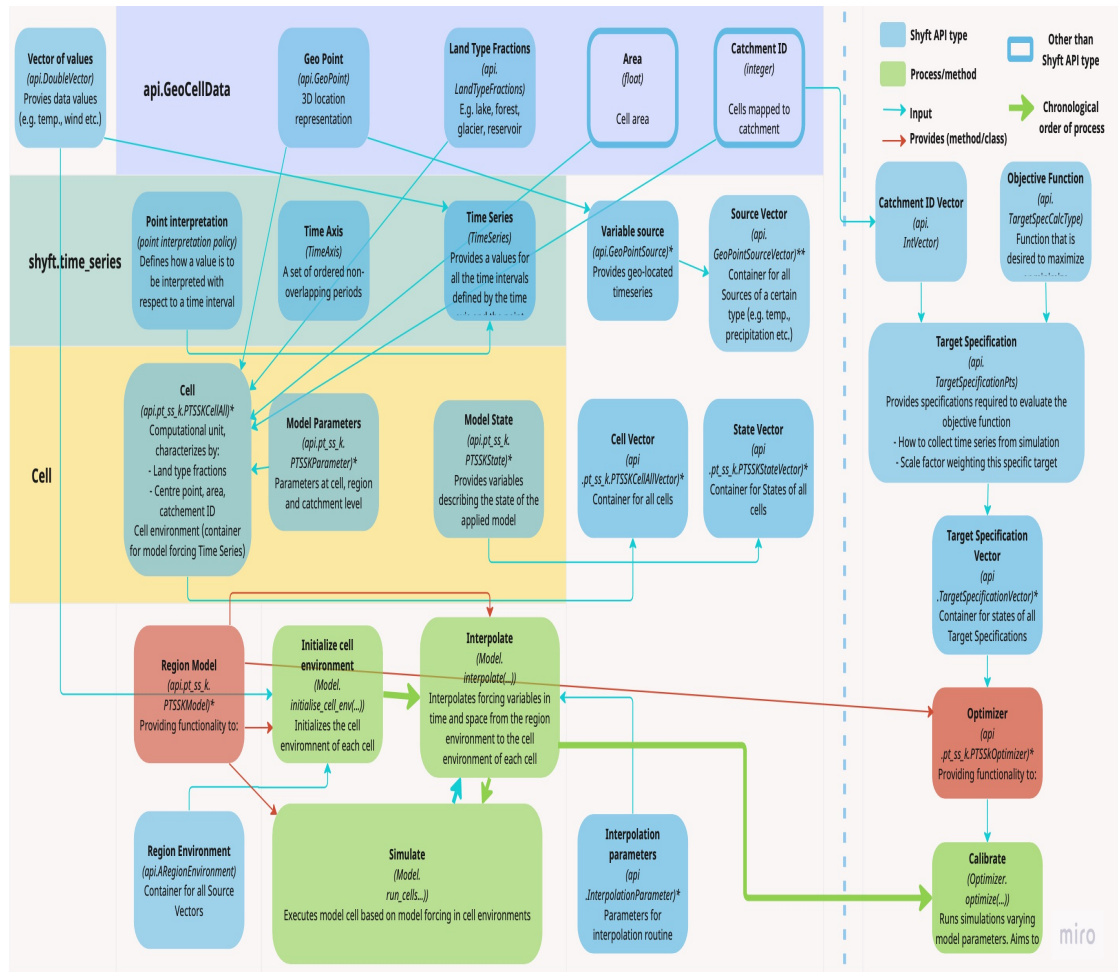


Figure A.2: Overview of the main Shyft API types. API types used for simulation are shown to the left of the blue dashed line, while calibration API types are found to the right of the blue dashed line. The figure shows an example of the PTSSK model stack (Priestly-Taylor-Skaugen-Snow-Kirchner). The figure is an adaptation of the figure in (Burkhart et al. 2021), and was made using Miro.

A.1.3 Repositories

The necessary data for running simulations depends on the chosen hydrological model. At a minimum, most models require temperature and precipitation, but some models also require wind speed, relative humidity and shortwave radiation.

The data required to run simulations in Shyft are managed through repositories (Fowler 2003). A repository is a Python based interface to data (Masnun 2023). Each repository in Shyft has a specific function, a defined interface and multiple implementations of these interfaces. The repositories return data in Shyft API types in order to keep the code in Shyft orchestration at a minimum. The following interfaces are available in Shyft:

Region model repository: Presents a configured region model, without revealing underlying implementation specific details regarding the storage of model configuration and data (such as netCDF files).

Geo-located time series repository: Provides geo-located time series for for meteorology- and hydrology-relevant types.

Interpolation parameter repository: Gives parameters for the interpolation method utilized in the simulation.

State repository: Stores the model states and provides them to the region model.

A.2 Orchestration

The orchestration layer is the composition of the simulation configuration (Burkhart et al. 2021). This includes defining the model domain, selecting datasets and model algorithms, and presenting the results. The Shyft orchestration is built on top of the API functionalities and offers utilities for configuring, running and post-processing of simulations. The two main objectives of the orchestration are to offer easy entry points for users of Shyft and to utilize all the functionalities of Shyft in operational environments.

A.2.1 Conceptual model

Shyft is designed to separate the model domain (region) from the model forcing data (region environment) and the model algorithms (Burkhart et al. 2021). This allows for more flexibility. In this section these three components are described in more detail. Together with the set of meteorological forcing data they form what is called the region model, which is central to Shyft.

Region: Model domain

Shyft defines a group of geo-located sub-units called cells (Burkhart et al. 2021). The cells contain different characteristics such as land type fractions, area, geographic location and a unique identifier (catchment id). The cells with the same catchment id belong to the same catchment. The region in Shyft is set up by the user where the number of cells and catchments are defined. The Shyft model domain doesn't use a grid to represent cells, but instead utilizes vectorization, which increases flexibility and efficiency.

Region environment

The model forcing data is stored as geo-located time series in what is called a region environment (Burkhart et al. 2021). The primary model forcing variables in Shyft are precipitation [mm/hr], temperature [$^{\circ}$], wind speed [m/s], shortwave radiation [W/m^2] and relative humidity [-]. To prepare for a simulation, the forcing data must be provided the region model which is described below (Statkraft 2023). This involves creating geo-located time series data, also known as "sources". Sources are categorized by type, such as temperature or precipitation, and are organized into SourceVectors before being added to the region model.

To generate these time series, a time axis is established and it is paired with data values (Statkraft 2023). A time axis consists of a collection of non-overlapping periods, with the time points indicating the start and end of each period. To ensure that there are as many periods as data values, an additional time point is added to mark the end of the final period in the time axis. The time series are assigned geo-locations. The coordinates must be based on an orthogonal coordinate system, such as UTM.

The SourceVectors can be filled with geo-located time series from various sources, such as station observations or numerical weather forecasts (Burkhart et al. 2021). The region environment provides the raw data without any assumptions about the spatial properties.

Model

Similarly to the region environment, the user defines the model algorithms independent of the region or the region environment (Burkhart et al. 2021). The model consists of a sequence. The model consists of a sequence of algorithms that represent various processes of the hydrological cycle, such as evapotranspiration, snow accumulation and melt processes, and soil response. These algorithms are compiled into model stacks, with different stacks differing in at least one method.

Region-model

After the region, region environment and the model is defined, these can be combined to create the region-model (Burkhart et al. 2021). The region model is the object that is used to run simulations. The region model includes functionalities such as interpolation of meteorological forcing data, running the model forward in time through cell by cell execution of the model stack, and provide access to all data, model parameters, and time series of model state result variables.

Targets

Shyft provides functionality to estimate model parameters using optimization algorithms such as the Bound Optimization BY Quadratic Approximation (BOBYQA) which is a derivative-free optimization algorithm (Burkhart et al. 2021; Powell 2009). The BOBYQA method aims to minimize a multi-variable function using a trust region approach and quadratic models formed by interpolation (Dlib 2018). The model results can be evaluated against target specifications such as observed discharge, mean catchment snow water equivalent (SWE), or snow-covered area (SCA). The calibration is single-objective, meaning it calibrates against one of these target specifications. The model performance can be evaluated by specifying the evaluation metric such as

Nash-Sutcliffe, Kling-Gupta, Absolute Difference or Root Mean Square Error (RMSE) functions, and by defining the search range.

A.2.2 Hydrologic modelling

Hydrologic modelling in Shyft generally is done by interpolating the model forcing data from the source locations to the cell locations in Shyft and calculating the model stack for each cell (Burkhart et al. 2021). Shyft provides several interpolation methods, including linear, inverse distance weighting, and kriging. The choice of interpolation method depends on the characteristics of the data, the spatial distribution of the data, and the desired accuracy.

Interpolation

Shyft offers several interpolation methods, but the user can also skip the interpolation step and input the data directly into the cells. This gives the user the option to use a preferred interpolation method or downscaling methods to transform the input data from source locations to the cell locations.

The inverse distance weighting (IDW) is a relatively fast and easy to compute spatial interpolation method (Lu and Wong 2008; Shepard 1968). The idea behind this method is that a value of an unsampled point is estimated as a weighted average of the known values in the surrounding neighbourhood. The weights are being inversely related to the distances between the prediction location and the sampled locations. The IDW method can be used for many different models for different variables (Burkhart et al. 2021). A great advantage of using the IDW method is that the weights do not change over time, if the spatial relationship between the sampled points and the target points is fixed. However, this can also be considered a weakness since the IDW is purely geometric and therefore it is difficult to combine with physical knowledge which can be done using Bayesian Temperature Kriging (BTK).

Figure A.3 illustrates the IDW method where the value of an unknown point is calculated using the weighted average of the available values at known points.

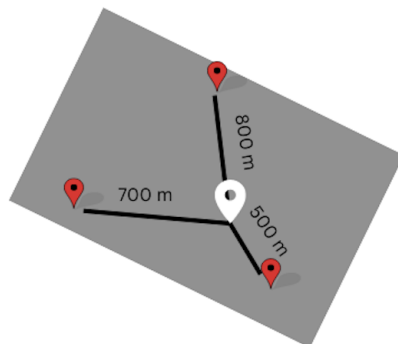


Figure A.3: Illustration of the Inverse Distance Weighting (IDW) method. Illustration made in Canva

A lapse rate (vertical gradient) method can be used to reduce systematic error when estimating local air temperature based on regional observations (Burkhart et al. 2021). This gradient value can be calculated in Shyft or set by the user. In many cases this simplistic approach is suitable for the purposes of optimizing the calibration. However, this approach may not always be suitable, especially when the temperature gradient is complicated and varies seasonally or with local weather. In such cases, Shyft provides an alternative approach based on Bayesian inference (a statistical inference method in which Bayes' theorem is used to update the probability of an event based on more evidence) to estimate a regional height gradient to model surface temperature. This approach buffers the most severe local effects and weights measurement data against prior information.

The IDW method utilized in Shyft consists of two main steps. Firstly, neighbors are selected for each destination point (Burkhart et al. 2021), based on user-defined parameters such as Z-scale (excluding neighbors of different heights), number of neighbors, and a distance limit to the neighbours. The second step involves applying a transformation method to project the signal type from its source location to the destination position. Shyft offers two options for computing temperature lapse rate: using nearest neighbors with sufficient and/or maximum vertical distance, or deriving a full 3D temperature flux from selected points and utilizing its vertical component (Burkhart et al. 2021). For precipitation, a user-specified scale factor and z distance is used as the source-destination destination. Radiation, on the other hand, allows for adjustment of slope and weight scaling factor at the destination cell.

Routing

Routing is a technique that is used to predict changes in the hydrograph as water moves through a river or reservoir. Shyft's routing process consists of two phases: cell-to-river routing and river network routing (Burkhart et al. 2021). During cell-to-river routing, water is directed to the nearest river object, which then receives lateral inflow from surrounding cells. While routing velocity and distance can vary among individual cells, catchment-specific unit hydrograph (UHG) shape parameters are utilized. In river network routing, water is directed from one river object to the next downstream river object, taking into account lateral inflow from cells as determined in the first phase. The combined upstream river discharge and lateral inflow are then passed on to the next downstream river object. To define UHG shape parameters and discretized time length, a UHG parameter structure is employed, with a gamma function currently used to determine UHG shape. The approach of (Skaugen and Onof 2014) involves summing up all cell responses at a river routing point and using a distance distribution profile to define the UHG for that routing point. The UHG, along with convolution, is used to determine the response from cells to the routing point.

A.3 Shyft requirements

The following requirements are needed to be fulfilled to compile and run Shyft (Statkraft 2023):

- A C++ compiler (gcc-11 or higher)
- The BLAS and LAPACK libraries (development packages)
- A Python (3.8 or higher) interpreter

- NumPy package ($\geq 1.15.0$)
- netCDF4 package ($\geq 1.2.1$)
- CMake building tool (3.19 or higher)
- 3rd party dependencies for c++ extensions and test boosts (≥ 1.78), dlib (≥ 1.78), armadillo ($\geq 9.300.2$), doctest, openssl, leveldb

In addition other Python packages are needed for data analysis and visualization of the results. The ones used in this study can be found in the "requirements.txt" file in the [GitHub repository](#).

A.4 Calibration

Shyft supports the following three optimization methods:

- **Min BOBYQA**: A local search algorithm using a quadratic approximation (Sapkota 2016). It can handle a lot of calibration parameters, but does not know whether a global optimum is found.
- **DREAM**: Differential evolution adaptive Metropolis, runs multiple different Markov chains in parallel using discrete proposal evolution to evolve the sampler to the posterior distribution (Laloy and Vrugt 2012).
- **Shuffled Complex Evolution Method (SCE-UA)**: General purpose global optimization program (Duan et al. 1992). It searches via a competitive evolutionary process for the global optimum of a function by evolving clusters of samples drawn from the parameter space (Naeini et al. 2019).

Appendix B

Net radiation in Shyft

The net radiation in Shyft is estimated using [Equation B.1](#) (Statkraft [2023](#)).

$$R_n = \sigma T^4 \cdot (\epsilon_{atm} - 0.98) + SW_{in} * (1.0 - \alpha_{land}) \quad (\text{B.1})$$

where σ is the Stephan-Boltzmann constant ($5.67 \cdot 10^{-8} \text{ W}/(\text{m}^2 \text{ K}^4)$), T is surface temperature in Kelvin and ϵ_{atm} is clear-sky emissivity [-] (dependent on surface temperature and/or vapor pressure, or both) calculated using emissivity model by Brutsaert [1975](#) with a cloud factor from Table 1 in Sicart et al. [2006](#). SW_{in} is the incoming shortwave radiation [W/m^2] and α_{land} is the surface albedo. ϵ_{atm} is calculated using [Equation B.2](#) (Statkraft [2023](#)).

$$\epsilon_{atm} = 1.24 \cdot \left(\frac{10 \cdot P_{vapor}}{T} \right)^{0.143} \cdot (0.85 + 0.5 \cdot RH) \quad (\text{B.2})$$

where P_{vapor} is the actual vapour pressure [kPa] calculated using [Equation B.3](#) and RH is the relative humidity [-] (Statkraft [2023](#)).

$$P_{vapor} = P_{sat} \cdot RH \quad (\text{B.3})$$

where P_{sat} is the saturated water vapor pressure [kPa] calculated by [Equation B.4](#) (Statkraft [2023](#)).

$$P_{sat} = ck1 \cdot e^{ck2[i] \cdot T \cdot ctt_{inv}} \quad (\text{B.4})$$

where $ck1$ is a constant equal to 0.610780, $ck2[i]$ is a negative or positive set of $ck[]$ constants given by Statkraft [2023](#), and ctt_{inv} is calculated by [Equation B.5](#) (Statkraft [2023](#)).

$$ckk_{inv} = \frac{1}{(ck3[i] + T)} \quad (\text{B.5})$$

Appendix C

Gamma Snow routine

This section give a more detailed look at the Gamma Snow routine. As mentioned in [subsubsection 5.1.1](#) the Gamma snow routine adopts an energy balance (Essery 2009; Hegdahl et al. 2016), which given by [Equation C.1](#).

$$\Delta E = S \cdot (1 - \alpha) + L_{in} - L_{out} + H_{SE} + H_L + E_G \quad (C.1)$$

where ΔE is the net energy flux at the surface available for snow melting [W/m^2], S is the incoming shortwave radiation [W/m^2], L_{in} is incoming longwave radiation [W/m^2], L_{out} is outgoing longwave radiation [W/m^2], H_{SE} and H_L are sensible [W/m^2] and latent heat [W/m^2] fluxes and E_G is the subsurface energy flux [W/m^2].

The snow albedo (α_t) quantifies the proportion of shortwave radiation reflected by the surface at a given time-step (t) and grid cell (Hegdahl et al. 2016). It is determined based on calculations conducted for each time-step and grid cell within the study area. The range of snow albedo is defined by the parameters α_{max} and α_{min} , representing its maximum and minimum values, respectively.

When the air temperature exceeds 0 °C, the snow albedo experiences a rapid exponential decay, as described by [Equation C.2](#). This implies that as temperatures increase, the amount of shortwave radiation reflected by the snow surface diminishes quickly. Conversely, when the air temperature falls below 0 °C, the snow albedo exhibits a slower linear decay, following the pattern outlined in [Equation C.3](#). This suggests that as temperatures decrease, the reduction in shortwave radiation reflected by the snow surface occurs at a slower rate.

$$\alpha_t = \alpha_{min} + (\alpha_{t-1} - \alpha_{min}) \times \left(\frac{1}{2^{1/\text{FastDecayRate}}} \right), \quad T_a > 0^\circ\text{C} \quad (C.2)$$

$$\alpha_t = \alpha_{t-1} - (\alpha_{max} - \alpha_{min}) \times \left(\frac{1}{2 \cdot \text{FastDecayRate}} \right), \quad T_a \leq 0^\circ\text{C} \quad (C.3)$$

The parameters *FastDecayRate* and *SlowDecayRate* represent the time it takes for the snow albedo to decrease by 95% of the difference between its maximum and minimum values (Hegdahl et al. 2016). When snowfall exceeds a certain depth (*ResetSnowDepth*), the albedo is reset to its maximum value (*MaxAlbedo*). The albedo of the glacier remains constant. The incoming and outgoing longwave radiations, L_{in} and L_{out} , are calculated using the Stephan-Boltzmann law. L_{in} depends on the air temperature (T_a), while the snow surface temperature (T_{ss}) is used to estimate the outgoing longwave radiation. T_{ss} is derived from T_a using the equation $T_{ss} = 1.16 \cdot T_a - 2.09$, with a maximum value of 0 °C.

Appendix C. Gamma Snow routine

The latent and sensible heat fluxes, H_L and H_{SE} , are calculated using a bulk-transfer method that considers factors such as wind speed, temperature, and air humidity (Hegdahl et al. 2016). The wind profile parameters, intercept (*wind_constant*) and slope (*wind_scale*), are determined through calibration. H_L and H_{SE} are always assumed to be positive and are set to zero when temperatures are below 0 °C.

The subsurface flux (E_G) is calculated assuming a linear temperature profile within the snow surface layer, with the temperature at the bottom of the layer set to 0 °C (Hegdahl et al. 2016). Snow melt in millimeters is derived by multiplying the available energy (ΔE) with the latent heat of fusion for water.

Appendix D

Processing of spatial data

This section describes how the spatial data is processed using QGIS. The catchment map (see [Figure 3.2](#)) is produced by downloading SRTM 1 Arc-Second Global DEM data from the [USGS Earth Explorer](#) which requires making an account first. The first step for downloading the DEM data is to select the coordinates of the study area in the search criteria. The coordinates used here is longitude 82-88 ° E latitude 26-32 ° N. This is much greater extent than the catchment, but is needed to cover the extent of the 16 WFDE5 grid cells used in Shyft. The search result will give several files called tiles. Each tile covers 1 degree in both north and east direction. All of the DEM files are opened in QGIS once they are downloaded as GeoTIFF files. In QGIS, the map projection is set to EPSG:43645 (45N UTM) and the DEM files are merged into 1 DEM file using the "Raster" → "Miscellaneous" → "Merge..." tool. Once the DEMs are merged into 1 DEM file, the plugin "SAGA" is installed and sinks (missing values or zero values) are corrected for by using the "fill sinks (wang & liu)" tool under "Processing Toolbox" → "SAGA". The fill sinks module in SAGA is based upon an algorithm by Wang and H. Liu [2006](#) that identifies and fills surface depressions. Once the DEM is pre-processed it is being clipped so that only the DEM relevant for the catchment is extracted. This is done using the "Raster" → "Extraction" → "Clip raster by Mask Layer...". The input layer is the merged, filled DEM and the mask layer is the catchment polygon (.shp file). Source and target CRS is EPSG:4326 (longitude/latitude).

The GlobCover 2009 and MODIS data is clipped in QGIS using the same clipping method as described above to match the catchment only. The colors of the clipped raster files are changes by selecting "Properties" → "Symbology" → and applying "Paletted/Unique values" under "Render type". MODIS data uses "Singleband pseudo-color" and "Blues" color ramp. The river network shapefile is clipped using "Vector" → "Geoprocessing Tools" → "Clip.." with the river network shapefile as input layer and the catchment outline as overlay layer. The color is changed to blue under "Properties" → "Line" → "Color".

The transect (latitudinal elevation profile) is made by installing "Profile Tool" module. In the Profile Tool the merged DEM is selected and a line is drawn from south to north on the DEM. This displays the profile in QGIS, but the table with the values is downloaded instead and visualized in Python.

To make statistics from the GlobCov 2009 data, the raster file is polygonized by selecting "Raster" → "Conversion" → "Polygonize (raster to vector)". The the attribute table values are exported to a ".csv" file that can be read in Python. The values in the attribute table correspond to the different land types in GlobCover. There are 23 different land cover types in GlobCover, but many of these include different nuances of

Appendix D. Processing of spatial data

similar land cover types. Similar land cover types are therefore aggregated based on similarities. In [Figure 3.5](#), the values 11, 14, 20 and 30 are classified as cropland, 40, 50, 60, 70, 100, 110, 120 and 130 are classified as forest, 140 as herbaceous vegetation, 150 and 200 as sparse area, 210 as water bodies and 220 as permanent snow/ice. See GlobCover 2009 legends documentation in the downloaded zip folder for more information about the different land types in the data set.

The maps are made in "New Print Layout" under "Project". Here the maps are added by selecting "Add map", scale bars are added by selecting "Add scale bar", north arrow is added by selecting "Add North Arrow" and legend is added by selecting "Add legend". The coordinates and grids are added by selecting "Grids" → "Plus sign" → "Modify Grid..." and selecting the grid intervals under "Interval". The borders are selected by selecting "Exterior ticks" under "Frame Style". To have several maps in one figure, one has to select "Lock layers" under "Layers" before adding another map.

Appendix E

Forcing data

The forcing data that is used in Shyft is stored in separate file netCDF files, one for each variable. The netCDF files are created using the Xarray package. An Xarray Dataset set is a multi-dimensional, in memory, array database (Xarray 2023). It consists of data variables, coordinates and attributes which combined form a self-describing data set. The dimensions of the output netCDF files are *stations* and *time*. The coordinates are *time*, *x*, *y* and *z*. The data variables are temperature, precipitation, radiation, wind speed and relative humidity. This dataset can then be saved as a netCDF file using the "to_netcdf" function. The files are saved into five separate files, one for each variable.

The result is a netCDF file containing the information that can be seen in Table E.1. As can be seen from the table the Budhi Gandaki has 16 stations, or midpoints in the WFDE5 dataset, and 262968 time steps (hours).

Table E.1: Description of the forcing data

Description	Name (Dimension)	Data type
Dimensions	Station: 16, Time: 262968	
Coordinates	Time (time)	<i>datetime64</i>
	X (station) Y (station) and Z (station)	<i>float64</i>
Data variables	Series name (station)	<i>int64</i>
	Series name (station)	<i>int64</i>
	CRS ()	<i>int32</i>
	Variable (e.g. temperature) (station)	<i>float64</i>

The forcing data can be found at the [Zenodo repository](#)

Appendix E. Forcing data

Appendix F

Spatial statistics of regular grid cells and TINs

Table F.1 provides a summary of geometric statistics for the different discretization methods being used. It can be seen that the TINs with higher resolutions have a lower mean area. All the regular grid cells have an area of 4 km², except the ones that lie at the edges of the catchment boundary due to the polygon clipping process. The median area for the TINs is lower than the mean area. The TIN with the lowest resolution (TIN-911) has the greatest maximum area, while TIN-4681 has the lowest maximum area. The mean cell elevation is greatest for the regular grid cells (Reg-1077) with a mean elevation of 4348.25 m, and lowest for TIN-1382 with a mean elevation of 4312.02 m. The median cell elevation is higher than the mean elevation for all meshes. The regular grid has the highest minimum cell elevation (546.0), while TIN-911 and TIN-1071 have the lowest min elevation (499.1 m). The TIN with the highest resolution (TIN-4681) has the highest max elevation, while TIN-911 has the lowest max elevation. The regular grid cells have a lower mean forest fraction (0.17) than the TINs (0.61). The glacier fraction is also lower for regular cells (0.16), than the TINs (0.34-0.35).

Table F.1: Statistics of mesh geometries utilised in the Shyft simulations.

Statistic	Reg-1077	TIN-911	TIN-1071	TIN-1382	TIN-4681
Number of cells	1077	911	1071	1382	4681
Mean area [km ²]	4.00	4.26	3.62	2.81	0.83
Median area [km ²]	4.00	3.87	3.30	2.55	0.73
Min area [km ²]	0.00*	0.02	0.02	0.01	0.00*
Max area [km ²]	4.00	17.06	12.78	12.86	3.64
Mean cell elevation [m]	4348.25	4320.21	4320.67	4314.02	4346.10
Median cell elevation [m]	4673.90	4659.35	4644.07	4628.91	4654.95
Min cell elevation [m]	546.0	499.1	499.1	509.1	501.1
Max cell elevation [m]	7353.3	7458.1	7606.0	7799.4	7949.6
Mean forest fraction [-]	0.17	0.61	0.61	0.61	0.61
Mean glacier fraction [-]	0.16	0.35	0.35	0.34	0.34
Mean lake fraction [-]	0.00*	0.00	0.00	0.00	0.00
Mean reservoir fraction [-]	0.00	0.00	0.00	0.00	0.00

* Value less than 10^{-2}

Figure F.1 gives a spatial overview of the regular cells and the TINs used in this study. It can be seen that adding more cells gives greater resolution. It becomes clear that the cells with highest elevations are located along the northern borders of the catchment, and the lowest elevations are located to the south. Most cells are found in the elevation range between 3500-6000 meters above sea level. The area plots at the bottom of the figure show that all of the regular cells have the same area of 4 km² (except a few that have a value of close to zero as a result of the clipping process in QGIS near the edges of the catchment). Most of the cells in TIN-911 and TIN-1071 cells have an area of between 2.0-6.0 km². The most of the cells in TIN-1382 have an area of 1.5-6.0 km². Most of the cells in TIN-4681 have an area of 0.2-1.5 km².

Figure F.2 shows the glacier fraction [-] and the forest fraction [-] represented by the different discretization methods. The glacier fraction plots (at the top) show great variation. Comparing with the glacier outlines provided by (RGI 2017), it can be seen that the regular cells have a better match with the glacier outlines than the TINs. The forest fraction plots (in the middle) also show a large variation between the regular cells and the TINs. There is very little forest represented in the regular cells around the northern parts of the catchment. Most of the forest is in the southern part. The TINs have much more forest covering the south-north axis of the catchment.

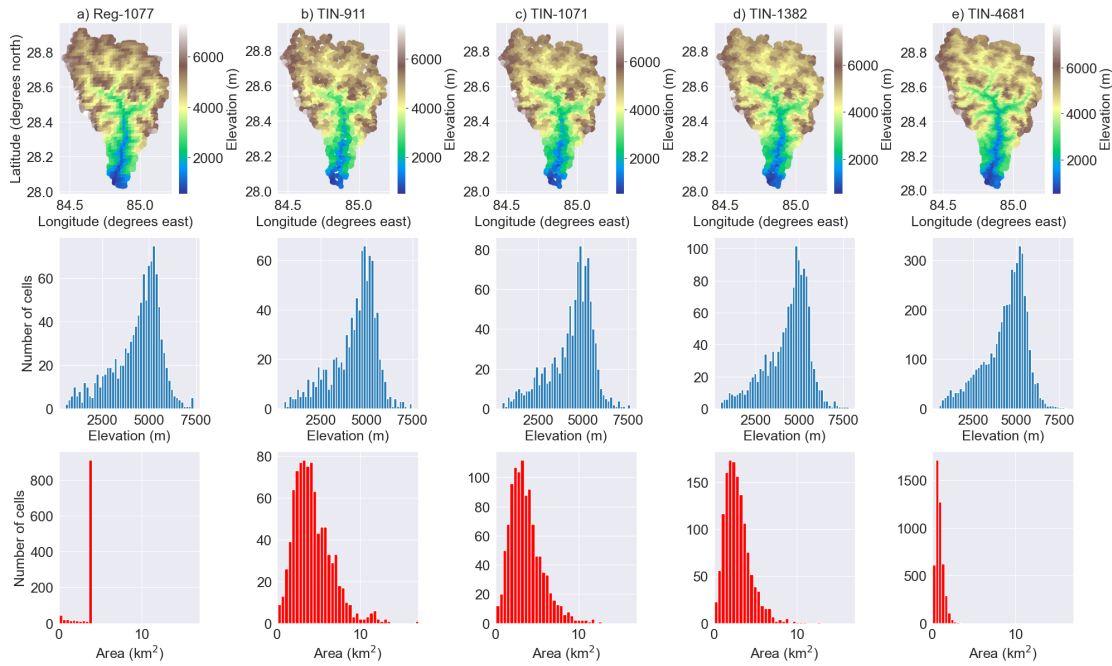


Figure F.1: Regular cells and TINs. The upper plots show the elevation at each cell. The middle plots illustrate the amount of cells at each elevation bin. The bottom plots show how many cells that have a similar area (km^2).

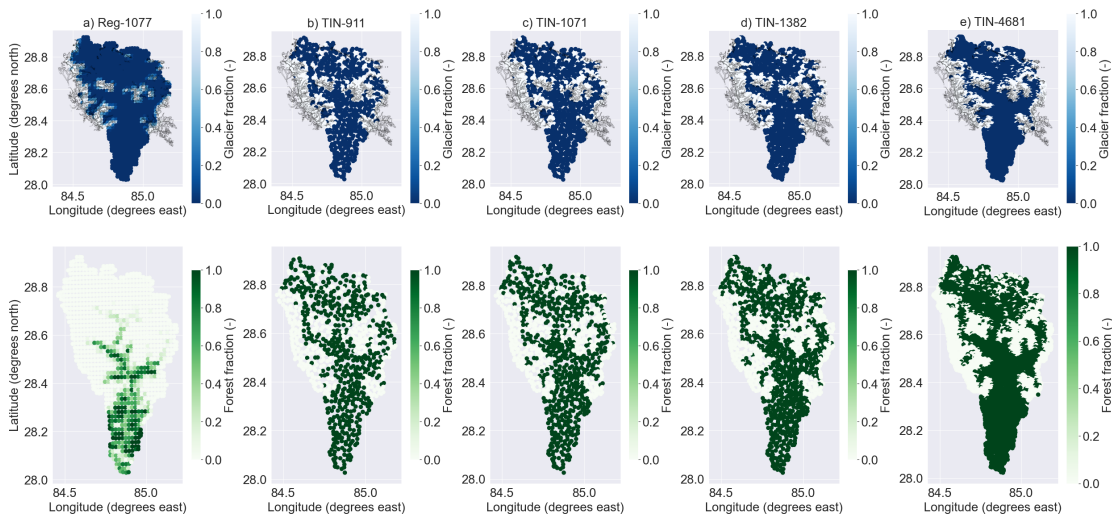


Figure F.2: Land cover type fractions for a) Reg-1077 (regular grid cells), b) TIN-911 (low resolution TIN), c) TIN-1071, d) TIN-1382 and e) TIN-4681 (high resolution TIN). The plots at the top show the glacier fraction. White color represents a glacier and dark blue color means no glacier. The polygons in the glacier fraction plots represent glaciers registered in the Randolph Glacier Inventory 6.0 (RGI 2017). The bottom plots show the forest fraction where green is forest and white is no forest.

Appendix G

YAML-configuration files

The YAML-configuration files used for the Shyft simulations are listed below. The first configuration file is called "budhi_gandaki_simulation.yaml" (listing 1). It contains the name of the simulation, "stanzas" or blocks of simulation configurations and the repository for the observed discharge data (Statkraft 2023). It is the file where Shyft finds pointers to the other configuration files. It also contains information about when the simulation starts ("YYYY-MM-DDThh:mm:ss"), the time step for each run in seconds (86400 is then 1 hour), number of time steps and pointers to the initial state of the simulation. The stanzas include YAML files for configuration of: the region, the model, the data sets, interpolation and the calibration.

The region configuration file (listing 2) called "budhi_gandaki_region.yaml" contains information about the domain of the simulation, the catchment ID and repository. The domain includes the EPSG code, the number of steps in x and y direction, and the lower corner x and y coordinates for the region. The repository describes which Python class to use for the region and where the regular cell data or TINs are located. The catchment ID is the ID of the catchment that is simulated, which is only relevant if the cell data file contains cells from multiple catchments.

The dataset configuration file called "budhi_gandaki_datasets.yaml" (listing 3) includes information about the input datasets and the repositories they are located in (Statkraft 2023).

The model configuration named "budhi_gandaki_model.yaml" (listing 4 and listing ??) includes the model stack class in Shyft and the initial model parameters. The initial model parameters are chosen based on previous literature (Bhattarai, Burkhart, Tallaksen, et al. 2020; Bhattarai, Silantyeva, et al. 2020; Lombraña 2017; Sapkota 2016).

The "budhi_gandaki_interpolation.yaml" file (listing 6) provides information on how the input data will be interpolated in the domain of the simulation. The interpolation method is chosen based on the type of input data. For example, the precipitation data is interpolated using the inverse distance weighted (IDW) method, while the temperature data is interpolated using the Bayesian Temperature Kriging (BTK) method. The interpolated values are what the cells in Shyft will use for the simulation.

The "budhi_gandaki_calibration.yaml" file have several components. The first component contains information about where to find the model configuration and where to output the calibrated model parameters (the "calibrated_model.yaml" file). It also describes which optimization method to use (e.g. Min BOBQYA) and the parameters that go into this method, and a repository for the target time series. This repository includes the path to the observed time series, the class to use, and definitions of the starting time of the calibration, the number of time steps, time steps in each run, catchment id, and the objective function with its scaling factors (e.g. the NSE). Since both model models use the same time information and optimization methods for calibration, this part separated out and shown in listing 7.

The calibration files also contains information about which model class is being used and and a component describing the ranges for each parameter used in the chosen model. This information is specific for each model. The calibration parameters for the PTGSK model are shown in listing 8, 9, 10 and 11 for the PTSTK.

A "calibrated_model.yaml" file is created after the calibration process is complete. This file contains the optimized parameters. Since these files are specific to all of the 10 calibration runs, these files are not included here, but summarized in [subsection 6.2.1](#) and the [GitHub Repository](#).

```

1  ---
2  budhi_gandaki:
3    region_config_file: budhi_gandaki_region_tin.yaml
4    model_config_file: calibrated_model.yaml
5    datasets_config_file: budhi_gandaki_datasets.yaml
6    interpolation_config_file: budhi_gandaki_interpolation.yaml
7    start_datetime: 2000-01-01T00:00:00
8    run_time_step: 3600 # 1 hour time step
9    number_of_steps: 140256 # hours from Jan 1 2000 to 31 Dec 2015 (16 years)
10   region_model_id: 'budhi_gandaki_tin911_ptstk_daily'
11   #interpolation_id: 2 # this is optional (default 0)
12   initial_state:
13     repository:
14       class: !!python/name:shyft.hydrology.repository.generated_state_repository.GeneratedStateRepository
15       params:
16         model: !!python/name:shyft.hydrology.pt_st_k.PTSTKModel
17       tags: []
18   references:
19   - repository: !!python/name:shyft.hydrology.repository.netcdf.cf_ts_repository.CFTsRepository
20     params:
21       file: netcdf/orchestration-testdata/discharge.nc
22       var_type: discharge
23     1D_timeseries:
24     - catch_id: [10]
25       type: discharge
26       uid: Arughat
27       run_time_step: 86400
28
29   ...

```

Listing 1: Configuration of the simulation

```
1 ---
2 repository:
3   class: !!python/name:mst_himalaya.cf_region_model_repository_tin.CFRegionModelRepository
4   params:
5     data_file: tin/narayani_cid10_small_tin.nc
6     get_model_from_tin_repo: True
7     tin_uid: ['narayani-cid-10-small']
8     tin_data_folder: budhi_gandaki/tin_archive
9
10  domain:
11    EPSG: 32645
12    nx: 38
13    ny: 53
14    step_x: 2000
15    step_y: 2000
16    lower_left_x: 250828
17    lower_left_y: 3100180
18
19  catchment_indices:
20    - 10
```

Listing 2: Configuration of the region model

```

1  ---
2  sources:
3  - repository: !!python/name:mst_himalaya.wfde5_cf_geo_ts_repository.CFDataRepos
4    types:
5      - precipitation
6    params:
7      filename: netcdf/orchestration-testdata/precipitation.nc
8      padding: 0.0
9  - repository: !!python/name:mst_himalaya.wfde5_cf_geo_ts_repository.CFDataRepos
10   types:
11     - temperature
12   params:
13     filename: netcdf/orchestration-testdata/temperature.nc
14     padding: 0.0
15  - repository: !!python/name:mst_himalaya.wfde5_cf_geo_ts_repository.CFDataRepos
16   types:
17     - wind_speed
18   params:
19     filename: netcdf/orchestration-testdata/wind_speed.nc
20     padding: 0.0
21  - repository: !!python/name:mst_himalaya.wfde5_cf_geo_ts_repository.CFDataRepos
22   types:
23     - relative_humidity
24   params:
25     filename: netcdf/orchestration-testdata/relative_humidity.nc
26     padding: 0.0
27  - repository: !!python/name:mst_himalaya.wfde5_cf_geo_ts_repository.CFDataRepos
28   types:
29     - radiation
30   params:
31     filename: netcdf/orchestration-testdata/radiation.nc
32     padding: 0.0
33  ...
34

```

Listing 3: Configuration of datasets

```
1 model_t: !!python/name:shyft.hydrology.pt_gs_k.PTGSKModel # model to construct
2 model_parameters:
3   ae: # actual_evapotranspiration
4     ae_scale_factor: 1.0
5   gs: # gamma_snow
6     calculate_iso_pot_energy: false
7     fast_albedo_decay_rate: 6.0
8     glacier_albedo: 0.4
9     initial_bare_ground_fraction: 0.04
10    max_albedo: 0.9
11    max_water: 0.1
12    min_albedo: 0.6
13    slow_albedo_decay_rate: 30.0
14    snow_cv: 0.4
15    tx: -1.68
16    snowfall_reset_depth: 5.0
17    surface_magnitude: 30.0
18    wind_const: 1.0
19    wind_scale: 2.01
20    winter_end_day_of_year: 100
21  #hbv_snow:
22  #   foo: 10.0
23  kirchner:
24    c1: -5.51
25    c2: 0.44
26    c3: -0.05
27  p_corr: # precipitation_correction
28    scale_factor: 1.0
29  pt: # priestley_taylor
30    albedo: 0.2
31    alpha: 1.26
32  routing:
33    alpha: 0.9
34    beta: 3.0
35    velocity: 0.0
36  msp:
37    reservoir_direct_response_fraction: 1.0
```

Listing 4: Configuration of the PTGSK model stack

```
1 model_t: !!python/name:shyft.hydrology.pt_st_k.PTSTKModel # model to construct
2 model_parameters:
3   ae: # actual_evapotranspiration
4     ae_scale_factor: 1.0
5   st: # gamma_snow
6     shape: 2.0
7     tx: -1.68
8     cx: 1.0
9     ts: 0.0
10    lwmax: 0.1
11    cfr: 0.5
12  gm:
13    dtf: 6.0
14    direct_response: 0.0
15  kirchner:
16    c1: -5.51
17    c2: 0.44
18    c3: -0.05
19  p_corr: # precipitation_correction
20    scale_factor: 1.0
21  pt: # priestley_taylor
22    albedo: 0.2
23    alpha: 1.26
24  routing:
25    alpha: 0.9
26    beta: 3.0
27    velocity: 0.0
28  msp:
29    reservoir_direct_response_fraction: 1.0
```

Listing 5: Configuration of the PTSTK model stack

```
1 interpolation_parameters:
2   temperature:
3     method: btk
4     params:
5       temperature_gradient: -0.6
6       temperature_gradient_sd: 0.25
7       nug: 0.5
8       range: 200000.0
9       sill: 25.0
10      zscale: 20.0
11  precipitation:
12    method: idw
13    params:
14      max_distance: 400000.0
15      max_members: 10
16      distance_measure_factor: 1
17      scale_factor: 0.93
18  radiation:
19    method: idw
20    params:
21      max_distance: 600000.0
22      max_members: 10
23      distance_measure_factor: 1.0
24  wind_speed:
25    method: idw
26    params:
27      max_distance: 600000.0
28      max_members: 10
29      distance_measure_factor: 1.0
30  relative_humidity:
31    method: idw
32    params:
33      max_distance: 600000.0
34      max_members: 10
35      distance_measure_factor: 1.0
```

Listing 6: Configuration of interpolation methods

```

1  budhi_gandaki:
2    model_config_file: budhi_gandaki_simulation.yaml
3    calibrated_model_file: calibrated_model.yaml # file where the calibrated params will go
4    optimization_method:
5      name: min_bobyqa # can be 'min_bobyqa', 'dread' or 'sceua'
6      params:
7        max_n_evaluations: 1500
8        tr_start: 0.1
9        tr_stop: 1.0e-5
10   target:
11   - repository: !!python/name:mst_himalaya wfde5_cf_ts_repository.CFTsRepository
12     params:
13       file: netcdf/orchestration-testdata/discharge.nc
14       var_type: discharge
15     1D_timeseries:
16     - catch_id: [10]
17       uid: Arughat
18       start_datetime: 2000-01-01T00:00:00
19       run_time_step: 86400 # Seconds in 1 day
20       number_of_steps: 1827 # Days from jan 1 2000 to Dec 31 2004
21       weight: 1.0
22       obj_func:
23         name: NSE # Nash-Sutcliffe efficiency (NSE) or Kling-Gupta efficiency (KGE)
24         scaling_factors:
25           s_corr: 1.0
26           s_var: 1.0
27           s_bias: 1.0
28

```

Listing 7: Configuration of the calibration for both models. Includes defining the optimization method and the target (observed discharge) along with the time information. A daily calibration is being used.

```
1 overrides:
2   model:
3     model_t: !!python/name:shyft.hydrology.pt_gs_k.PTGSKOptModel
4   calibration_parameters:
5     kirchner.c1:
6       min: -8.0
7       max: 0.0
8     kirchner.c2:
9       min: -1.0
10      max: 1.2
11     kirchner.c3:
12       min: -0.15
13       max: -0.05
14     ae.ae_scale_factor:
15       min: 1.0
16       max: 1.0
17     gs.tx:
18       min: -3.0
19       max: 2.0
20     gs.wind_scale:
21       min: 1.0
22       max: 6.0
23     gs.max_water:
24       min: 0.1
25       max: 0.1
26     gs.wind_const:
27       min: 1.0
28       max: 1.0
29     gs.fast_albedo_decay_rate:
30       min: 5.0 # 5.0
31       max: 15.0 # 15.0
32     gs.slow_albedo_decay_rate:
33       min: 20.0 # 20.0
34       max: 40.0 # 40.0
35     gs.surface_magnitude:
36       min: 30.0
37       max: 30.0
38     gs.max_albedo:
39       min: 0.9
40       max: 0.9
41     gs.min_albedo:
42       min: 0.6
43       max: 0.6
44     gs.snowfall_reset_depth:
45       min: 5.0
46       max: 5.0
47     gs.snow_cv:
48       min: 0.4
49       max: 0.4
50
```

```
1 gs.snow_cv_forest_factor:
2   min: 0.0
3   max: 0.0
4 gs.snow_cv_altitude_factor:
5   min: 0.0
6   max: 0.0
7 gs.glacier_albedo:
8   min: 0.4
9   max: 0.4
10 gs.initial_bare_ground_fraction:
11   min: 0.04
12   max: 0.04
13 gs.winter_end_day_of_year:
14   min: 100
15   max: 100
16 gs.calculate_iso_pot_energy:
17   min: 0
18   max: 0
19 gs.n_winter_days:
20   min: 100
21   max: 100
22 gm.dtf:
23   min: 6
24   max: 6
25 gm.direct_response:
26   min: 0.0
27   max: 1.0
28 routing.velocity:
29   min: 0
30   max: 0
31 routing.alpha:
32   min: 0.9
33   max: 0.9
34 routing.beta:
35   min: 3.0
36   max: 3.0
37 p_corr.scale_factor:
38   min: 0.4
39   max: 2.0
40 pt.albedo:
41   min: 0.2
42   max: 0.2
43 pt.alpha:
44   min: 1.26
45   max: 1.26
46 msp.reservoir_direct_response_fraction:
47   min: 1.0
48   max: 1.0
```

Listing 9: Configuration of the PTGSK model stack calibration: Part 2

```
1  overrides:
2    model:
3      model_t: !!python/name:shyft.hydrology.pt_st_k.PTSTKOptModel
4  calibration_parameters:
5    kirchner.c1:
6      min: -8.0
7      max: 0.0
8    kirchner.c2:
9      min: -1.0
10     max: 1.2
11   kirchner.c3:
12     min: -0.15
13     max: -0.05
14   ae.ae_scale_factor:
15     min: 1.0
16     max: 1.0
17   st.shape:
18     min: 2.0
19     max: 2.0
20   st.tx:
21     min: -3.0
22     max: 2.0
23   st.cx:
24     min: 0.0
25     max: 30.0
26   st.ts:
27     min: -4.0
28     max: 4.0
29   st.lwmax:
30     min: 0.0
31     max: 1.0
32   st.cfr:
33     min: 0.0
34     max: 1.0
35   gm.dtf:
36     min: 6
37     max: 6
38   p_corr.scale_factor:
39     min: 0.4
40     max: 2.0
41   pt.albedo:
42     min: 0.2
43     max: 0.2
44   pt.alpha:
45     min: 1.26
46     max: 1.26
47   routing.velocity:
48     min: 0
49     max: 0
```

Listing 10: Configuration of the PTSTK model stack calibration: Part 1

```
1 routing.alpha:
2   min: 0.9
3   max: 0.9
4 routing.beta:
5   min: 3.0
6   max: 3.0
7 gm.direct_response:
8   min: 0.0
9   max: 1.0
10 msp.reservoir_direct_response_fraction:
11   min: 1.0
12   max: 1.0
```

Listing 11: Configuration of the PTSTK model stack calibration: Part 2

Bibliography

- Aalstad, K., S. Westermann, and L. Bertino (Mar. 2020). “Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography.” In: *Remote Sensing of Environment* 239, p. 111618. ISSN: 00344257. DOI: [10.1016/j.rse.2019.111618](https://doi.org/10.1016/j.rse.2019.111618). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425719306388> (visited on 05/04/2023).
- Agha Kouchak, A. and A. Mehran (Oct. 2013). “Extended contingency table: Performance metrics for satellite observations and climate model simulations: Technical Note.” In: *Water Resources Research* 49.10, pp. 7144–7149. ISSN: 00431397. DOI: [10.1002/wrcr.20498](https://doi.org/10.1002/wrcr.20498). URL: <http://doi.wiley.com/10.1002/wrcr.20498> (visited on 05/08/2023).
- Alderton, D. and S. Elias (2021). *Encyclopedia of Geology*. 2nd ed. Elsevier Academic Press. ISBN: 978-0-08-102909-1.
- Allaby, M. (2008). *Dictionary of earth sciences*. Oxford university press.
- Allen, M., B. Mustafa, and P. Shukla (June 9, 2022). *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*. 1st ed. Cambridge University Press. ISBN: 978-1-00-915794-0 978-1-00-915795-7. DOI: [10.1017/9781009157940](https://doi.org/10.1017/9781009157940). URL: <https://www.cambridge.org/core/product/identifier/9781009157940/type/book> (visited on 04/28/2023).
- Anders, A. M. et al. (2006). “Spatial patterns of precipitation and topography in the Himalaya.” In: Willett, S. D. et al. *Tectonics, Climate, and Landscape Evolution*. Geological Society of America. ISBN: 978-0-8137-2398-3. DOI: [10.1130/2006.2398\(03\)](https://doi.org/10.1130/2006.2398(03)). URL: <https://pubs.geoscienceworld.org/books/book/569/chapter/3802950> (visited on 04/27/2023).
- Anees, M. T. et al. (Dec. 2016). “Numerical modeling techniques for flood analysis.” In: *Journal of African Earth Sciences* 124, pp. 478–486. ISSN: 1464343X. DOI: [10.1016/j.jafrearsci.2016.10.001](https://doi.org/10.1016/j.jafrearsci.2016.10.001). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1464343X1630334X> (visited on 04/29/2023).
- Baidya, S. K., M. L. Shrestha, and M. M. Sheikh (2008). “Trends in daily climatic extremes of temperature and precipitation in Nepal.” In: *Journal of Hydrology and Meteorology* 5.1, pp. 38–51.
- Bajracharya, S. R. et al., eds. (2011). *The status of glaciers in the Hindu Kush-Himalayan region*. Kathmandu: International Centre for Integrated Mountain Development. 127 pp. ISBN: 978-92-9115-215-5 978-92-9115-217-9.
- Baldo, Elisabeth and Steven A. Margulis (May 2017). “Implementation of a physiographic complexity-based multiresolution snow modeling scheme.” In: *Water Resources Research* 53.5, pp. 3680–3694. ISSN: 0043-1397, 1944-7973. DOI: [10.1002/](https://doi.org/10.1002/)

Bibliography

- 2016WR020021. URL: <https://onlinelibrary.wiley.com/doi/10.1002/2016WR020021> (visited on 05/12/2023).
- Bergström, S. (1976). “Development and Application of a Conceptual Runoff Model for Scandinavian Catchments.” In: *SMHI*.
- Bern, M. W., J. E. Flaherty, and M. Luskin (1999). “Grid Generation and Adaptive Algorithms.” In: *Springer-Verlag*.
- Berström, S. (1992). “The HBV-Model—Its Structure and Applications.” In: *SMHI* 4.
- Bhattarai, B. C. (2020). *Hydrologic model forcing over the Himalaya*. 2267. Faculty of Mathematics and Natural Sciences, University of Oslo. ISBN: 1501-7710.
- Bhattarai, B. C., J. F. Burkhart, F. Stordal, et al. (July 5, 2019). “Aerosol Optical Depth Over the Nepalese Cryosphere Derived From an Empirical Model.” In: *Frontiers in Earth Science* 7, p. 178. ISSN: 2296-6463. DOI: [10.3389/feart.2019.00178](https://doi.org/10.3389/feart.2019.00178). URL: <https://www.frontiersin.org/article/10.3389/feart.2019.00178/full> (visited on 10/17/2022).
- Bhattarai, B. C., J. F. Burkhart, L. M. Tallaksen, et al. (Apr. 1, 2020). “Evaluation of global forcing datasets for hydropower inflow simulation in Nepal.” In: *Hydrology Research* 51.2, pp. 202–225. ISSN: 0029-1277, 2224-7955. DOI: [10.2166/nh.2020.079](https://doi.org/10.2166/nh.2020.079). URL: <https://iwaponline.com/hr/article/51/2/202/72386/Evaluation-of-global-forcing-datasets-for> (visited on 10/17/2022).
- Bhattarai, B. C., O. Silantyeva, et al. (Aug. 20, 2020). “Impact of Catchment Discretization and Imputed Radiation on Model Response: A Case Study from Central Himalayan Catchment.” In: *Water* 12.9, p. 2339. ISSN: 2073-4441. DOI: [10.3390/w12092339](https://doi.org/10.3390/w12092339). URL: <https://www.mdpi.com/2073-4441/12/9/2339> (visited on 09/29/2022).
- Bhushal, R. (2021). *Budhi Gandaki: Nepal’s mega-dam remains a mirage*. URL: <https://www.thethirdpole.net/en/livelihoods/budhi-gandaki-nepals-mega-dam-remains-a-mirage/> (visited on 05/04/2023).
- Bontemps, S. et al. (2011). *GLOBCOVER 2009 - Products Description and Validation Report*. UCLouvain & ESA Team.
- Bookhagen, B. and D. W. Burbank (2006). “Topography, relief, and TRMM-derived rainfall variations along the Himalaya.” In: *Geophysical Research Letters* 33.8, p. L08405. ISSN: 0094-8276. DOI: [10.1029/2006GL026037](https://doi.org/10.1029/2006GL026037). URL: <http://doi.wiley.com/10.1029/2006GL026037> (visited on 04/30/2023).
- (Aug. 10, 2010). “Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge.” In: *Journal of Geophysical Research* 115 (F3), F03019. ISSN: 0148-0227. DOI: [10.1029/2009JF001426](https://doi.org/10.1029/2009JF001426). URL: <http://doi.wiley.com/10.1029/2009JF001426> (visited on 09/29/2022).
- Bormann, K. J. et al. (Nov. 2018). “Estimating snow-cover trends from space.” In: *Nature Climate Change* 8.11, pp. 924–928. ISSN: 1758-678X, 1758-6798. DOI: [10.1038/s41558-018-0318-3](https://doi.org/10.1038/s41558-018-0318-3). URL: <http://www.nature.com/articles/s41558-018-0318-3> (visited on 04/29/2023).
- Brutsaert, Wilfried (Oct. 1975). “On a derivable formula for long-wave radiation from clear skies.” In: *Water Resources Research* 11.5, pp. 742–744. ISSN: 00431397. DOI: [10.1029/WR011i005p00742](https://doi.org/10.1029/WR011i005p00742). URL: <http://doi.wiley.com/10.1029/WR011i005p00742> (visited on 05/12/2023).
- Buck, A. L. (Dec. 1981). “New Equations for Computing Vapor Pressure and Enhancement Factor.” In: *Journal of Applied Meteorology* 20.12, pp. 1527–1532. ISSN: 0021-8952. DOI: [10.1175/1520-0450\(1981\)020<1527:NEFCVP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1981)020<1527:NEFCVP>2.0.CO;2).

- URL: [http://journals.ametsoc.org/doi/10.1175/1520-0450\(1981\)020%3C1527:NEFCVP%3E2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0450(1981)020%3C1527:NEFCVP%3E2.0.CO;2) (visited on 04/23/2023).
- Burges, S. J. (Aug. 1986). “Trends and directions in hydrology.” In: *Water Resources Research* 22.9, 1S–5S. ISSN: 00431397. DOI: [10.1029/WR022i09Sp0001S](https://doi.org/10.1029/WR022i09Sp0001S). URL: <http://doi.wiley.com/10.1029/WR022i09Sp0001S> (visited on 04/29/2023).
- Burkhart, J. F. et al. (Feb. 5, 2021). “Shyft v4.8: a framework for uncertainty assessment and distributed hydrologic modeling for operational hydrology.” In: *Geoscientific Model Development* 14.2, pp. 821–842. ISSN: 1991-9603. DOI: [10.5194/gmd-14-821-2021](https://doi.org/10.5194/gmd-14-821-2021). URL: <https://gmd.copernicus.org/articles/14/821/2021/> (visited on 09/29/2022).
- Carrivick, J. L. and F. S. Tweed (Sept. 2016). “A global assessment of the societal impacts of glacier outburst floods.” In: *Global and Planetary Change* 144, pp. 1–16. ISSN: 09218181. DOI: [10.1016/j.gloplacha.2016.07.001](https://doi.org/10.1016/j.gloplacha.2016.07.001). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0921818116301023> (visited on 04/18/2023).
- Chaulagain, N. P. (Feb. 1, 2009). “Impacts of climate change on water resources of Nepal: The physical and socio-economic dimensions.” In: *IOP Conference Series: Earth and Environmental Science* 6.29, p. 292029. ISSN: 1755-1315. DOI: [10.1088/1755-1307/6/29/292029](https://doi.org/10.1088/1755-1307/6/29/292029). URL: <https://iopscience.iop.org/article/10.1088/1755-1307/6/29/292029> (visited on 04/18/2023).
- Clark, M. P. et al. (Sept. 2021). “The Abuse of Popular Performance Metrics in Hydrologic Modeling.” In: *Water Resources Research* 57.9, e2020WR029001. ISSN: 0043-1397, 1944-7973. DOI: [10.1029/2020WR029001](https://doi.org/10.1029/2020WR029001). URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020WR029001> (visited on 05/08/2023).
- Cucchi, M., G. P. Weedon, A. Amici, N. Bellouin, S. Lange, H. Müller Schmied, H. Hersbach, and C. Buontempo (Sept. 8, 2020). “WFDE5: bias-adjusted ERA5 reanalysis data for impact studies.” In: *Earth System Science Data* 12.3, pp. 2097–2120. ISSN: 1866-3516. DOI: [10.5194/essd-12-2097-2020](https://doi.org/10.5194/essd-12-2097-2020). URL: <https://essd.copernicus.org/articles/12/2097/2020/> (visited on 09/29/2022).
- Cucchi, M., G. P. Weedon, A. Amici, N. Bellouin, S. Lange, H. Müller Schmied, H. Hersbach, C. Cagnazzo, et al. (2022). *Near surface meteorological variables from 1979 to 2018 derived from bias-corrected reanalysis*. DOI: [10.24381/CDS.20D54E34](https://doi.org/10.24381/CDS.20D54E34). URL: <https://cds.climate.copernicus.eu/doi/10.24381/cds.20d54e34> (visited on 05/12/2022).
- Dandekhya, S. et al. (2017). *The Gandaki Basin- Maintaining Livelihoods in the Face of Landslides, Floods, and Drought; HI-AWARE Working Paper 9*. Himalayan Adaptation, Water and Resilience (HI-AWARE).
- Debele, Bekele, Raghavan Srinivasan, and A. K. Gosain (Apr. 2010). “Comparison of Process-Based and Temperature-Index Snowmelt Modeling in SWAT.” In: *Water Resources Management* 24.6, pp. 1065–1088. ISSN: 0920-4741, 1573-1650. DOI: [10.1007/s11269-009-9486-2](https://doi.org/10.1007/s11269-009-9486-2). URL: <http://link.springer.com/10.1007/s11269-009-9486-2> (visited on 05/12/2023).
- Dee, D. P. et al. (Apr. 2011). “The ERA-Interim reanalysis: configuration and performance of the data assimilation system.” In: *Quarterly Journal of the Royal Meteorological Society* 137.656, pp. 553–597. ISSN: 00359009. DOI: [10.1002/qj.828](https://doi.org/10.1002/qj.828). URL: <https://onlinelibrary.wiley.com/doi/10.1002/qj.828> (visited on 04/23/2023).
- Devkota, R. P. et al. (Jan. 2017). “Climate change and adaptation strategies in Budhi Gandaki River Basin, Nepal: a perception-based analysis.” In: *Climatic Change* 140.2, pp. 195–208. ISSN: 0165-0009, 1573-1480. DOI: [10.1007/s10584-016-1836-5](https://doi.org/10.1007/s10584-016-1836-5). URL: <http://link.springer.com/10.1007/s10584-016-1836-5> (visited on 10/03/2022).

Bibliography

- DHM (2022). *Hydrological Station*. URL: <https://www.dhm.gov.np>.
- Dimri, A. P. (June 2004). “Impact of horizontal model resolution and orography on the simulation of a western disturbance and its associated precipitation.” In: *Meteorological Applications* 11.2, pp. 115–127. ISSN: 1350-4827, 1469-8080. DOI: [10.1017/S1350482704001227](https://doi.org/10.1017/S1350482704001227). URL: <http://doi.wiley.com/10.1017/S1350482704001227> (visited on 04/27/2023).
- Dimri, A. P., D. Niyogi, et al. (June 2015). “Western Disturbances: A review.” In: *Reviews of Geophysics* 53.2, pp. 225–246. ISSN: 87551209. DOI: [10.1002/2014RG000460](https://doi.org/10.1002/2014RG000460). URL: <http://doi.wiley.com/10.1002/2014RG000460> (visited on 04/27/2023).
- Dimri, A. P., E. Palazzi, and A. S. Daloz (July 2022). “Elevation dependent precipitation and temperature changes over Indian Himalayan region.” In: *Climate Dynamics* 59.1, pp. 1–21. ISSN: 0930-7575, 1432-0894. DOI: [10.1007/s00382-021-06113-z](https://doi.org/10.1007/s00382-021-06113-z). URL: <https://link.springer.com/10.1007/s00382-021-06113-z> (visited on 04/27/2023).
- Dlib (2018). *Optimization - find_min_bobqya*. URL: http://www.dlib.net/optimization.html#find_min_bobyqa.
- DOED (2023). *Power Plants*. URL: <https://www.doed.gov.np/license/54> (visited on 05/04/2023).
- Dordé, F. and M. L’Hostis (2015). *Feasibility Study and Detailed Design of Budhi Gandaki HPP*. Main Report 1. Tractebel Engineering & JADE Consult.
- Duan, Q., S. Sorooshian, and V. Gupta (Apr. 1992). “Effective and efficient global optimization for conceptual rainfall-runoff models.” In: *Water Resources Research* 28.4, pp. 1015–1031. ISSN: 00431397. DOI: [10.1029/91WR02985](https://doi.org/10.1029/91WR02985). URL: <http://doi.wiley.com/10.1029/91WR02985> (visited on 04/19/2023).
- Dwarakish, G. S. and B. P. Ganasri (Dec. 31, 2015). “Impact of land use change on hydrological systems: A review of current modeling approaches.” In: *Cogent Geoscience* 1.1. Ed. by L. De Stefano, p. 1115691. ISSN: 2331-2041. DOI: [10.1080/23312041.2015.1115691](https://doi.org/10.1080/23312041.2015.1115691). URL: <https://www.tandfonline.com/doi/full/10.1080/23312041.2015.1115691> (visited on 04/29/2023).
- Edenhofer, O. et al. (2011). *Renewable energy sources and climate change mitigation: Special report of the intergovernmental panel on climate change*. Cambridge University Press.
- Engeland, K. et al. (May 2016). “Effects of uncertainties in hydrological modelling. A case study of a mountainous catchment in Southern Norway.” In: *Journal of Hydrology* 536, pp. 147–160. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2016.02.036](https://doi.org/10.1016/j.jhydrol.2016.02.036). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022169416300701> (visited on 04/23/2023).
- EROS (2017). *Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global*. Medium: tiff. DOI: [10.5066/F7PR7TFT](https://doi.org/10.5066/F7PR7TFT). URL: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects (visited on 01/23/2023).
- ESRI (2023a). *How Kriging works*. URL: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-kriging-works.htm> (visited on 05/10/2023).
- (2023b). *Shapefiles*. URL: <https://doc.arcgis.com/en/arcgis-online/reference/shapefiles.htm> (visited on 04/23/2023).
- Essery, Richard (2009). “D.R. DeWalle and A. Rango. 2008. Principles of snow hydrology.” In: *Journal of Glaciology* 55.189, pp. 179–180. ISSN: 0022-1430, 1727-5652. DOI: [10.3189/002214309788608994](https://doi.org/10.3189/002214309788608994). URL: https://www.cambridge.org/core/product/identifier/S0022143000200178/type/journal_article (visited on 05/12/2023).
- Fan, Y. et al. (Feb. 2019). “Hillslope Hydrology in Global Change Research and Earth System Modeling.” In: *Water Resources Research* 55.2, pp. 1737–1772. ISSN: 0043-

- 1397, 1944-7973. DOI: [10.1029/2018WR023903](https://doi.org/10.1029/2018WR023903). URL: <https://onlinelibrary.wiley.com/doi/10.1029/2018WR023903> (visited on 04/23/2023).
- Fiddes, J. and S. Gruber (Feb. 21, 2014). “TopoSCALE v.1.0: downscaling gridded climate data in complex terrain.” In: *Geoscientific Model Development* 7.1, pp. 387–405. ISSN: 1991-9603. DOI: [10.5194/gmd-7-387-2014](https://doi.org/10.5194/gmd-7-387-2014). URL: <https://gmd.copernicus.org/articles/7/387/2014/> (visited on 05/15/2023).
- Fiddes, Joel, Kristoffer Aalstad, and Sebastian Westermann (Nov. 19, 2019). “Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering.” In: *Hydrology and Earth System Sciences* 23.11, pp. 4717–4736. ISSN: 1607-7938. DOI: [10.5194/hess-23-4717-2019](https://doi.org/10.5194/hess-23-4717-2019). URL: <https://hess.copernicus.org/articles/23/4717/2019/> (visited on 05/12/2023).
- Fowler, M. (2003). *Patterns of enterprise application architecture*. The Addison-Wesley signature series. Boston: Addison-Wesley. 533 pp. ISBN: 978-0-321-12742-6.
- Galarnyk, M. (2022). *Understanding Boxplots*. URL: <https://builtin.com/data-science/boxplot> (visited on 05/12/2023).
- GeoPandas (2022). *GeoPandas*. URL: <https://geopandas.org/en/stable/>.
- Giroto, Manuela et al. (May 16, 2017). “Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India: India Grace Data Assimilation.” In: *Geophysical Research Letters* 44.9, pp. 4107–4115. ISSN: 00948276. DOI: [10.1002/2017GL072994](https://doi.org/10.1002/2017GL072994). URL: <http://doi.wiley.com/10.1002/2017GL072994> (visited on 05/12/2023).
- Gupta, H. V. et al. (Oct. 2009). “Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling.” In: *Journal of Hydrology* 377.1, pp. 80–91. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2009.08.003](https://doi.org/10.1016/j.jhydrol.2009.08.003). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022169409004843> (visited on 04/18/2023).
- Gurung, D. R. et al. (2011). *Snow-Cover Mapping and Monitoring in the Hindu Kush-Himalayas*. Edition: 0. Kathmandu, Nepal: ICIMOD. DOI: [10.53055/ICIMOD.550](https://doi.org/10.53055/ICIMOD.550). URL: <https://lib.icimod.org/record/9418> (visited on 04/23/2023).
- Hall, D., V. V. Salomonson, and K. George (2006). *MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 5*. DOI: [10.5067/63NQASRDPDB0](https://doi.org/10.5067/63NQASRDPDB0). URL: <http://nsidc.org/data/MOD10A1/versions/5> (visited on 04/19/2023).
- Hall, D. K. and G. A. Riggs (June 15, 2007). “Accuracy assessment of the MODIS snow products.” In: *Hydrological Processes* 21.12, pp. 1534–1547. ISSN: 08856087, 10991085. DOI: [10.1002/hyp.6715](https://doi.org/10.1002/hyp.6715). URL: <https://onlinelibrary.wiley.com/doi/10.1002/hyp.6715> (visited on 04/23/2023).
- Hall, D. K., A. Riggs G., et al. (2015). *MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid*. DOI: [10.5067/MODIS/MOD10A1.006](https://doi.org/10.5067/MODIS/MOD10A1.006). URL: <http://nsidc.org/data/MOD10A1/versions/6> (visited on 04/19/2023).
- Hallouin, T. (Apr. 22, 2021). *hydroeval: an evaluator for streamflow time series in Python*. Version v0.1.0. Language: en. DOI: [10.5281/ZENODO.2591217](https://doi.org/10.5281/ZENODO.2591217). URL: <https://zenodo.org/record/2591217> (visited on 04/18/2023).
- Halwatura, D. and M.M.M. Najim (Aug. 2013). “Application of the HEC-HMS model for runoff simulation in a tropical catchment.” In: *Environmental Modelling & Software* 46, pp. 155–162. ISSN: 13648152. DOI: [10.1016/j.envsoft.2013.03.006](https://doi.org/10.1016/j.envsoft.2013.03.006). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1364815213000698> (visited on 05/11/2023).
- Hamududu, B. and A. Killingtveit (Feb. 14, 2012). “Assessing Climate Change Impacts on Global Hydropower.” In: *Energies* 5.2, pp. 305–322. ISSN: 1996-1073. DOI: [10.3390/energies5020305](https://doi.org/10.3390/energies5020305).

Bibliography

- 3390/en5020305. URL: <http://www.mdpi.com/1996-1073/5/2/305> (visited on 04/18/2023).
- Harris, I. et al. (Dec. 2020). “Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset.” In: *Scientific Data* 7.1, p. 109. ISSN: 2052-4463. DOI: [10.1038/s41597-020-0453-3](https://doi.org/10.1038/s41597-020-0453-3). URL: <http://www.nature.com/articles/s41597-020-0453-3> (visited on 10/17/2022).
- Hegdahl, Trine J. et al. (Aug. 1, 2016). “Discharge sensitivity to snowmelt parameterization: a case study for Upper Beas basin in Himachal Pradesh, India.” In: *Hydrology Research* 47.4, pp. 683–700. ISSN: 0029-1277, 2224-7955. DOI: [10.2166/nh.2016.047](https://doi.org/10.2166/nh.2016.047). URL: <https://iwaponline.com/hr/article/47/4/683/930/Discharge-sensitivity-to-snowmelt-parameterization> (visited on 05/12/2023).
- Hersbach, H. et al. (July 2020). “The ERA5 global reanalysis.” In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049. ISSN: 0035-9009, 1477-870X. DOI: [10.1002/qj.3803](https://doi.org/10.1002/qj.3803). URL: <https://onlinelibrary.wiley.com/doi/10.1002/qj.3803> (visited on 04/23/2023).
- Hock, R. (Nov. 2003). “Temperature index melt modelling in mountain areas.” In: *Journal of Hydrology* 282.1, pp. 104–115. ISSN: 00221694. DOI: [10.1016/S0022-1694\(03\)00257-9](https://doi.org/10.1016/S0022-1694(03)00257-9). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022169403002579> (visited on 04/16/2023).
- Hock, R. et al. (2019). *High Mountain Areas*. In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. 1st ed. Cambridge University Press. ISBN: 978-1-00-915796-4 978-1-00-915797-1. DOI: [10.1017/9781009157964](https://doi.org/10.1017/9781009157964). URL: <https://www.cambridge.org/core/product/identifier/9781009157964/type/book> (visited on 04/18/2023).
- Hunt, Kieran M. R., Andrew G. Turner, and Reinhard K. H. Schiemann (Apr. 2022). “Katabatic and convective processes drive two preferred peaks in the precipitation diurnal cycle over the Central Himalaya.” In: *Quarterly Journal of the Royal Meteorological Society* 148.745, pp. 1731–1751. ISSN: 0035-9009, 1477-870X. DOI: [10.1002/qj.4275](https://doi.org/10.1002/qj.4275). URL: <https://onlinelibrary.wiley.com/doi/10.1002/qj.4275> (visited on 05/14/2023).
- IAE (2023). *World Energy Balances: Overview*. URL: <https://www.iea.org/reports/world-energy-balances-overview/world> (visited on 04/18/2023).
- Immerzeel, W. W., L. P. H. van Beek, and M. F. P. Bierkens (June 11, 2010). “Climate Change Will Affect the Asian Water Towers.” In: *Science* 328.5984, pp. 1382–1385. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1183188](https://doi.org/10.1126/science.1183188). URL: <https://www.science.org/doi/10.1126/science.1183188> (visited on 09/29/2022).
- Immerzeel, W. W., A. F. Lutz, et al. (Jan. 16, 2020). “Importance and vulnerability of the world’s water towers.” In: *Nature* 577.7790, pp. 364–369. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/s41586-019-1822-y](https://doi.org/10.1038/s41586-019-1822-y). URL: <http://www.nature.com/articles/s41586-019-1822-y> (visited on 04/18/2023).
- Immerzeel, W. W., L. Petersen, et al. (Mar. 2014). “The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas.” In: *Water Resources Research* 50.3, pp. 2212–2226. ISSN: 0043-1397, 1944-7973. DOI: [10.1002/2013WR014506](https://doi.org/10.1002/2013WR014506). URL: <https://onlinelibrary.wiley.com/doi/10.1002/2013WR014506> (visited on 04/27/2023).
- Jacobqs (Apr. 30, 2022). *jacobqs/MST-Himalaya: Initial release*. Version v0.0.0. DOI: [10.5281/ZENODO.6510027](https://doi.org/10.5281/ZENODO.6510027). URL: <https://zenodo.org/record/6510027> (visited on 04/19/2023).

- JPL, NASA (2013). *NASA Shuttle Radar Topography Mission Global 1 arc second*. DOI: [10.5067/MEASURES/SRTM/SRTMGL1.003](https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003). URL: <https://lpdaac.usgs.gov/products/srtmgl1v003/> (visited on 10/13/2022).
- Karki, R. et al. (Aug. 2016). “New climatic classification of Nepal.” In: *Theoretical and Applied Climatology* 125.3, pp. 799–808. ISSN: 0177-798X, 1434-4483. DOI: [10.1007/s00704-015-1549-0](https://doi.org/10.1007/s00704-015-1549-0). URL: <http://link.springer.com/10.1007/s00704-015-1549-0> (visited on 09/29/2022).
- Kasahara, A. and M. Kanamitsu (2003). “Weather Prediction, Numerical.” In: *Encyclopedia of Physical Science and Technology*. Elsevier, pp. 805–835. ISBN: 978-0-12-227410-7. DOI: [10.1016/B0-12-227410-5/00824-3](https://doi.org/10.1016/B0-12-227410-5/00824-3). URL: <https://linkinghub.elsevier.com/retrieve/pii/B0122274105008243> (visited on 01/23/2023).
- Kattel, D. B. et al. (Aug. 2013). “Temperature lapse rate in complex mountain terrain on the southern slope of the central Himalayas.” In: *Theoretical and Applied Climatology* 113.3, pp. 671–682. ISSN: 0177-798X, 1434-4483. DOI: [10.1007/s00704-012-0816-6](https://doi.org/10.1007/s00704-012-0816-6). URL: <http://link.springer.com/10.1007/s00704-012-0816-6> (visited on 05/15/2023).
- Kauffeldt, A. et al. (Jan. 2016). “Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level.” In: *Environmental Modelling & Software* 75, pp. 68–76. ISSN: 13648152. DOI: [10.1016/j.envsoft.2015.09.009](https://doi.org/10.1016/j.envsoft.2015.09.009). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1364815215300529> (visited on 04/11/2023).
- Kavetski, Dmitri and George Kuczera (Mar. 2007). “Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration: Model Smoothing in Hydrology.” In: *Water Resources Research* 43.3. ISSN: 00431397. DOI: [10.1029/2006WR005195](https://doi.org/10.1029/2006WR005195). URL: <http://doi.wiley.com/10.1029/2006WR005195> (visited on 05/12/2023).
- Khanal, S. et al. (May 2021). “Variable 21st Century Climate Change Response for Rivers in High Mountain Asia at Seasonal to Decadal Time Scales.” In: *Water Resources Research* 57.5. ISSN: 0043-1397, 1944-7973. DOI: [10.1029/2020WR029266](https://doi.org/10.1029/2020WR029266). URL: <https://onlinelibrary.wiley.com/doi/10.1029/2020WR029266> (visited on 05/15/2023).
- Khatami, Sina et al. (Nov. 2019). “Equifinality and Flux Mapping: A New Approach to Model Evaluation and Process Representation Under Uncertainty.” In: *Water Resources Research* 55.11, pp. 8922–8941. ISSN: 0043-1397, 1944-7973. DOI: [10.1029/2018WR023750](https://doi.org/10.1029/2018WR023750). URL: <https://onlinelibrary.wiley.com/doi/10.1029/2018WR023750> (visited on 05/12/2023).
- Khatri, H. B., M. K. Jain, and S. K. Jain (Oct. 2018). “Modelling of streamflow in snow dominated Budhigandaki catchment in Nepal.” In: *Journal of Earth System Science* 127.7, p. 100. ISSN: 0253-4126, 0973-774X. DOI: [10.1007/s12040-018-1005-5](https://doi.org/10.1007/s12040-018-1005-5). URL: <http://link.springer.com/10.1007/s12040-018-1005-5> (visited on 09/29/2022).
- Kirchner, J. W. (Feb. 2009). “Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward: Catchments as Simple Dynamical Systems.” In: *Water Resources Research* 45.2. ISSN: 00431397. DOI: [10.1029/2008WR006912](https://doi.org/10.1029/2008WR006912). URL: <http://doi.wiley.com/10.1029/2008WR006912> (visited on 04/16/2023).
- Kirkham, J. D. et al. (July 16, 2019). “Near Real-Time Measurement of Snow Water Equivalent in the Nepal Himalayas.” In: *Frontiers in Earth Science* 7, p. 177. ISSN: 2296-6463. DOI: [10.3389/feart.2019.00177](https://doi.org/10.3389/feart.2019.00177). URL: <https://www.frontiersin.org/article/10.3389/feart.2019.00177/full> (visited on 04/27/2023).
- Knoben, W. J. M., J. E. Freer, and R. A. Woods (July 1, 2019). *Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency*

Bibliography

- scores. preprint. Catchment hydrology/Modelling approaches. DOI: [10.5194/hess-2019-327](https://doi.org/10.5194/hess-2019-327). URL: <https://hess.copernicus.org/preprints/hess-2019-327/hess-2019-327.pdf> (visited on 05/08/2023).
- Koch, F. et al. (Sept. 30, 2011). “How Will Hydroelectric Power Generation Develop under Climate Change Scenarios? A Case Study in the Upper Danube Basin.” In: *Energies* 4.10, pp. 1508–1541. ISSN: 1996-1073. DOI: [10.3390/en4101508](https://doi.org/10.3390/en4101508). URL: <http://www.mdpi.com/1996-1073/4/10/1508> (visited on 04/23/2023).
- Kolberg, S. and O. Bruland (2012). “ENKI - An Open Source environmental modelling platform.” In.
- Kolberg, S., H. Rue, and L. Gottschalk (June 1, 2006). “A Bayesian spatial assimilation scheme for snow coverage observations in a gridded snow model.” In: *Hydrology and Earth System Sciences* 10.3, pp. 369–381. ISSN: 1607-7938. DOI: [10.5194/hess-10-369-2006](https://doi.org/10.5194/hess-10-369-2006). URL: <https://hess.copernicus.org/articles/10/369/2006/> (visited on 04/16/2023).
- Lai, Yue et al. (Feb. 16, 2021). “Impacts of the Westerlies on Planetary Boundary Layer Growth Over a Valley on the North Side of the Central Himalayas.” In: *Journal of Geophysical Research: Atmospheres* 126.3. ISSN: 2169-897X, 2169-8996. DOI: [10.1029/2020JD033928](https://doi.org/10.1029/2020JD033928). URL: <https://onlinelibrary.wiley.com/doi/10.1029/2020JD033928> (visited on 05/14/2023).
- Lakew, H. B., A. M. Semu, and H. A. Dereje (2017). “Hydrological evaluation of satellite and reanalysis precipitation products in the Upper Blue Nile Basin: A case study of Gilgel Abbay.” In: *Hydrology* 4.3.
- Laloy, E. and J. A. Vrugt (Jan. 2012). “High-dimensional posterior exploration of hydrologic models using multiple-try DREAM and high-performance computing: Efficient MCMS for High-Dimensional Problems.” In: *Water Resources Research* 48.1. ISSN: 00431397. DOI: [10.1029/2011WR010608](https://doi.org/10.1029/2011WR010608). URL: <http://doi.wiley.com/10.1029/2011WR010608> (visited on 04/19/2023).
- Legates, D. R. and G. J. McCabe (Jan. 1999). “Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation.” In: *Water Resources Research* 35.1, pp. 233–241. ISSN: 00431397. DOI: [10.1029/1998WR900018](https://doi.org/10.1029/1998WR900018). URL: <http://doi.wiley.com/10.1029/1998WR900018> (visited on 04/19/2023).
- Lehner, B. and G. Grill (July 15, 2013). “Global river hydrography and network routing: baseline data and new approaches to study the world’s large river systems: Global River Hydrography and Network Routing.” In: *Hydrological Processes* 27.15, pp. 2171–2186. ISSN: 08856087. DOI: [10.1002/hyp.9740](https://doi.org/10.1002/hyp.9740). URL: <https://onlinelibrary.wiley.com/doi/10.1002/hyp.9740> (visited on 05/03/2023).
- Li, Yaojun et al. (Aug. 2021). “A new global gridded glacier dataset based on the Randolph Glacier Inventory version 6.0.” In: *Journal of Glaciology* 67.264, pp. 773–776. ISSN: 0022-1430, 1727-5652. DOI: [10.1017/jog.2021.28](https://doi.org/10.1017/jog.2021.28). URL: https://www.cambridge.org/core/product/identifier/S0022143021000289/type/journal_article (visited on 05/15/2023).
- Litt, Maxime et al. (Mar. 27, 2019). “Glacier ablation and temperature indexed melt models in the Nepalese Himalaya.” In: *Scientific Reports* 9.1, p. 5264. ISSN: 2045-2322. DOI: [10.1038/s41598-019-41657-5](https://doi.org/10.1038/s41598-019-41657-5). URL: <https://www.nature.com/articles/s41598-019-41657-5> (visited on 05/15/2023).
- Liu, Y. and S. A. Margulis (Nov. 1, 2019). “Deriving Bias and Uncertainty in MERRA-2 Snowfall Precipitation Over High Mountain Asia.” In: *Frontiers in Earth Science* 7, p. 280. ISSN: 2296-6463. DOI: [10.3389/feart.2019.00280](https://doi.org/10.3389/feart.2019.00280). URL: <https://www.frontiersin.org/article/10.3389/feart.2019.00280/full> (visited on 04/29/2023).

- Lombraña, J. U. (2017). “Evaluations of Snow Simulations in SHyFT.” PhD thesis. Trondheim: Norwegian University of Science and Technology (NTNU).
- Lu, G. Y. and D. W. Wong (Sept. 2008). “An adaptive inverse-distance weighting spatial interpolation technique.” In: *Computers & Geosciences* 34.9, pp. 1044–1055. ISSN: 00983004. DOI: [10.1016/j.cageo.2007.07.010](https://doi.org/10.1016/j.cageo.2007.07.010). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0098300408000721> (visited on 04/16/2023).
- Lutz, A. F. et al. (July 2014). “Consistent increase in High Asia’s runoff due to increasing glacier melt and precipitation.” In: *Nature Climate Change* 4.7, pp. 587–592. ISSN: 1758-678X, 1758-6798. DOI: [10.1038/nclimate2237](https://doi.org/10.1038/nclimate2237). URL: <http://www.nature.com/articles/nclimate2237> (visited on 10/03/2022).
- Maptiler (2023). *WGS 84 / UTM zone 45N*. URL: <https://epsg.io/?q=32645> (visited on 05/12/2023).
- Marahatta, S., L. Devkota, and D. Aryal (2021). “Hydrological Modeling: A Better Alternative to Empirical Methods for Monthly Flow Estimation in Ungauged Basins.” In: *Journal of Water Resource and Protection* 13.3, pp. 254–270. ISSN: 1945-3094, 1945-3108. DOI: [10.4236/jwarp.2021.133015](https://doi.org/10.4236/jwarp.2021.133015). URL: <https://www.scirp.org/journal/doi.aspx?doi=10.4236/jwarp.2021.133015> (visited on 09/29/2022).
- Marahatta, S., L. P. Devkota, and D. Aryal (May 31, 2021). “Application of SWAT in Hydrological Simulation of Complex Mountainous River Basin (Part I: Model Development).” In: *Water* 13.11, p. 1546. ISSN: 2073-4441. DOI: [10.3390/w13111546](https://doi.org/10.3390/w13111546). URL: <https://www.mdpi.com/2073-4441/13/11/1546> (visited on 09/29/2022).
- Marsh, Christopher B., John W. Pomeroy, and Howard S. Wheeler (Jan. 29, 2020). “The Canadian Hydrological Model (CHM) v1.0: a multi-scale, multi-extent, variable-complexity hydrological model – design and overview.” In: *Geoscientific Model Development* 13.1, pp. 225–247. ISSN: 1991-9603. DOI: [10.5194/gmd-13-225-2020](https://doi.org/10.5194/gmd-13-225-2020). URL: <https://gmd.copernicus.org/articles/13/225/2020/> (visited on 05/12/2023).
- Marsh, Christopher B., Raymond J. Spiteri, et al. (Oct. 2018). “Multi-objective unstructured triangular mesh generation for use in hydrological and land surface models.” In: *Computers & Geosciences* 119, pp. 49–67. ISSN: 00983004. DOI: [10.1016/j.cageo.2018.06.009](https://doi.org/10.1016/j.cageo.2018.06.009). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0098300417309676> (visited on 05/12/2023).
- Martinec, J., A. Rango, and E. Major (1983). “The Snowmelt-Runoff Model (SRM) User’s Manual.” In: *NASA Scientific and Technical Information Branch*.
- Masnun, A. A. (2023). *Interfaces in Python: Protocols and ABCs*. URL: <http://masnun.rocks/2017/04/15/interfaces-in-python-protocols-and-abc/> (visited on 02/02/2023).
- Matt, F. N. and J. F. Burkhart (Apr. 28, 2018). “Assessing Satellite-Derived Radiative Forcing From Snow Impurities Through Inverse Hydrologic Modeling.” In: *Geophysical Research Letters* 45.8, pp. 3531–3541. ISSN: 00948276. DOI: [10.1002/2018GL077133](https://doi.org/10.1002/2018GL077133). URL: <http://doi.wiley.com/10.1002/2018GL077133> (visited on 05/12/2023).
- May, R. M. et al. (Oct. 2022). “MetPy: A Meteorological Python Library for Data Analysis and Visualization.” In: *Bulletin of the American Meteorological Society* 103.10, E2273–E2284. ISSN: 0003-0007, 1520-0477. DOI: [10.1175/BAMS-D-21-0125.1](https://doi.org/10.1175/BAMS-D-21-0125.1). URL: <https://journals.ametsoc.org/view/journals/bams/103/10/BAMS-D-21-0125.1.xml> (visited on 01/30/2023).
- Ménégoz, M., H. Gallée, and H. W. Jacobi (Oct. 15, 2013). “Precipitation and snow cover in the Himalaya: from reanalysis to regional climate simulations.” In: *Hydrology and Earth System Sciences* 17.10, pp. 3921–3936. ISSN: 1607-7938. DOI: [10.5194/hess-](https://doi.org/10.5194/hess-17-3921-2013)

Bibliography

- 17-3921-2013. URL: <https://hess.copernicus.org/articles/17/3921/2013/> (visited on 10/13/2022).
- Ministry of Energy, Water Resources {and} Irrigation (2023). *Budhigandaki Hydropower Project*. URL: <https://bghep.gov.np/pages/about-the-project?lan=> (visited on 05/04/2023).
- Mishra, P. et al. (Dec. 2016). “Accuracy Assessment of MODIS Fractional Snow Cover Product for Eastern Himalayan Catchment.” In: *Journal of the Indian Society of Remote Sensing* 44.6, pp. 977–985. ISSN: 0255-660X, 0974-3006. DOI: [10.1007/s12524-016-0548-7](https://doi.org/10.1007/s12524-016-0548-7). URL: <http://link.springer.com/10.1007/s12524-016-0548-7> (visited on 04/29/2023).
- Mishra, S. K. et al. (Oct. 5, 2021). “Grand Challenges of Hydrologic Modeling for Food-Energy-Water Nexus Security in High Mountain Asia.” In: *Frontiers in Water* 3, p. 728156. ISSN: 2624-9375. DOI: [10.3389/frwa.2021.728156](https://doi.org/10.3389/frwa.2021.728156). URL: <https://www.frontiersin.org/articles/10.3389/frwa.2021.728156/full> (visited on 04/29/2023).
- Monteith, P. (1965). “Evaporation and environment.” In: *Cambridge University Press (CUP)* 19.
- Naeini, M. R. et al. (2019). “Three decades of the Shuffled Complex Evolution (SCE-UA) optimization algorithm: Review and applications.” In: *Scientia Iranica* 26.4. ISSN: 1026-3098. DOI: [10.24200/SCI.2019.21500](https://doi.org/10.24200/SCI.2019.21500).
- Nakayama, K. and A. Nakamura (1982). “Estimating Potential Evapotranspiration by the Priestley-Taylor Model.” In: *Journal of Agricultural Meteorology* 37.4, pp. 297–302. ISSN: 0021-8588, 1881-0136. DOI: [10.2480/agrmet.37.297](https://doi.org/10.2480/agrmet.37.297). URL: http://www.jstage.jst.go.jp/article/agrmet1943/37/4/37_4_297/_article (visited on 04/16/2023).
- Nepal, S. et al. (Jan. 30, 2014). “Understanding the hydrological system dynamics of a glaciated alpine catchment in the Himalayan region using the J2000 hydrological model: Hydrological System Dynamics of Himalaya Rivers.” In: *Hydrological Processes* 28.3, pp. 1329–1344. ISSN: 08856087. DOI: [10.1002/hyp.9627](https://doi.org/10.1002/hyp.9627). URL: <https://onlinelibrary.wiley.com/doi/10.1002/hyp.9627> (visited on 05/11/2023).
- NOAA (2023). *The Hydrologic cycle*. URL: <https://www.weather.gov/jetstream/hydro> (visited on 04/29/2023).
- NSIDC (2023). *MOD10A1*. URL: <https://nsidc.org/data/mod10a1/versions/6> (visited on 04/19/2023).
- Orsolini, Yvan et al. (Aug. 27, 2019). “Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations.” In: *The Cryosphere* 13.8, pp. 2221–2239. ISSN: 1994-0424. DOI: [10.5194/tc-13-2221-2019](https://doi.org/10.5194/tc-13-2221-2019). URL: <https://tc.copernicus.org/articles/13/2221/2019/> (visited on 05/15/2023).
- Padma Kumari, B. et al. (Nov. 10, 2007). “Observational evidence of solar dimming: Offsetting surface warming over India.” In: *Geophysical Research Letters* 34.21, p. L21810. ISSN: 0094-8276. DOI: [10.1029/2007GL031133](https://doi.org/10.1029/2007GL031133). URL: <http://doi.wiley.com/10.1029/2007GL031133> (visited on 05/14/2023).
- Pandas (2023). *Pandas*. URL: <https://pandas.pydata.org>.
- Pangali Sharma, T. P. et al. (June 17, 2020). “Assimilation of Snowmelt Runoff Model (SRM) Using Satellite Remote Sensing Data in Budhi Gandaki River Basin, Nepal.” In: *Remote Sensing* 12.12, p. 1951. ISSN: 2072-4292. DOI: [10.3390/rs12121951](https://doi.org/10.3390/rs12121951). URL: <https://www.mdpi.com/2072-4292/12/12/1951> (visited on 09/29/2022).
- Pellicciotti, F. et al. (Feb. 2012). “Challenges and Uncertainties in Hydrological Modeling of Remote Hindu Kush–Karakoram–Himalayan (HKH) Basins: Suggestions for

- Calibration Strategies.” In: *Mountain Research and Development* 32.1, pp. 39–50. ISSN: 0276-4741, 1994-7151. DOI: [10.1659/MRD-JOURNAL-D-11-00092.1](https://doi.org/10.1659/MRD-JOURNAL-D-11-00092.1). URL: <http://www.bioone.org/doi/10.1659/MRD-JOURNAL-D-11-00092.1> (visited on 10/13/2022).
- Powell, M. (2009). “The BOBYQA algorithm for bound constrained optimization without derivatives.” In.
- Priestley, C. H. B. and R. J. Taylor (Feb. 1972). “On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters.” In: *Monthly Weather Review* 100.2, pp. 81–92. ISSN: 0027-0644, 1520-0493. DOI: [10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2). URL: [http://journals.ametsoc.org/doi/10.1175/1520-0493\(1972\)100%3C0081:OTAOSH%3E2.3.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0493(1972)100%3C0081:OTAOSH%3E2.3.CO;2) (visited on 04/16/2023).
- Python (2023). *About Python*. URL: <https://www.python.org/about/> (visited on 04/23/2023).
- QGIS (2023). *QGIS Documentation*. URL: <https://www.qgis.org/en/docs/index.html> (visited on 04/23/2023).
- Rasterio (2023). *Rasterio: access to geospatial raster data*. URL: <https://rasterio.readthedocs.io/en/stable/#rasterio-access-to-geospatial-raster-data> (visited on 10/11/2022).
- Renard, B. et al. (May 2010). “Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors: Identifiability of Input and Structural Errors.” In: *Water Resources Research* 46.5. ISSN: 00431397. DOI: [10.1029/2009WR008328](https://doi.org/10.1029/2009WR008328). URL: <http://doi.wiley.com/10.1029/2009WR008328> (visited on 04/26/2023).
- Rew, R. et al. (1989). *Unidata NetCDF*. In collab. with W. Fisher et al. Language: en Medium: application/java-archive,application/gzip,application/tar. DOI: [10.5065/D6H70CW6](https://doi.org/10.5065/D6H70CW6). URL: <http://www.unidata.ucar.edu/software/netcdf/> (visited on 04/23/2023).
- RGI (2017). *Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 6*. DOI: [10.7265/4M1F-GD79](https://doi.org/10.7265/4M1F-GD79). URL: <https://nsidc.org/data/nsidc-0770/versions/6> (visited on 04/25/2023).
- Ritter, A. and R. Muñoz-Carpena (Feb. 2013). “Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments.” In: *Journal of Hydrology* 480, pp. 33–45. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2012.12.004](https://doi.org/10.1016/j.jhydrol.2012.12.004). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022169412010608> (visited on 04/19/2023).
- RoHub (2023). *What is RoHub?* URL: https://reliance.rohub.org/about?what_is_rohub (visited on 04/23/2023).
- Salby, M. L. (1996). *Fundamentals of atmospheric physics*. Elsevier.
- Sapkota, A. (2016). “Regional Modelling of the Narayani Basin in Nepal.” PhD thesis. Trondheim: Norwegian University of Science and Technology.
- Schaefer, J. T. (Dec. 1990). “The Critical Success Index as an Indicator of Warning Skill.” In: *Weather and Forecasting* 5.4, pp. 570–575. ISSN: 0882-8156, 1520-0434. DOI: [10.1175/1520-0434\(1990\)005<0570:TCSIAA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2). URL: [http://journals.ametsoc.org/doi/10.1175/1520-0434\(1990\)005%3C0570:TCSIAA%3E2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0434(1990)005%3C0570:TCSIAA%3E2.0.CO;2) (visited on 04/22/2023).
- Schneider, U. et al. (2018). “GPCC Full Data Monthly Version 2018.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data: Gridded Monthly Totals.” In: DOI: [10.5676/DWD_GPCC/FD_M_V2018_050](https://doi.org/10.5676/DWD_GPCC/FD_M_V2018_050).

Bibliography

- URL: https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-monthly_v2018_doi_download.html (visited on 10/17/2022).
- Shepard, D. (1968). "A two-dimensional interpolation function for irregularly-spaced data." In: *Proceedings of the 1968 23rd ACM national conference on -*. the 1968 23rd ACM national conference. Not Known: ACM Press, pp. 517–524. DOI: [10.1145/800186.810616](https://doi.org/10.1145/800186.810616). URL: <http://portal.acm.org/citation.cfm?doid=800186.810616> (visited on 04/16/2023).
- Shrestha, A. B. and S. P. Joshi (2011). "Snow Cover and Glacier Change Study in Nepalese Himalaya Using Remote Sensing and Geographic Information System." In: *Journal of Hydrology and Meteorology* 6.1, pp. 26–36. ISSN: 1818-2518. DOI: [10.3126/jhm.v6i1.5481](https://doi.org/10.3126/jhm.v6i1.5481). URL: <https://www.nepjol.info/index.php/JHM/article/view/5481> (visited on 04/11/2023).
- Shrestha, A. B., C. P. Wake, et al. (Sept. 1999). "Maximum Temperature Trends in the Himalaya and Its Vicinity: An Analysis Based on Temperature Records from Nepal for the Period 1971–94." In: *Journal of Climate* 12.9, pp. 2775–2786. ISSN: 0894-8755, 1520-0442. DOI: [10.1175/1520-0442\(1999\)012<2775:MTTITH>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2775:MTTITH>2.0.CO;2). URL: [http://journals.ametsoc.org/doi/10.1175/1520-0442\(1999\)012%3C2775:MTTITH%3E2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0442(1999)012%3C2775:MTTITH%3E2.0.CO;2) (visited on 10/17/2022).
- Shrestha, B., P.K. Mool, and S. R. Bajracharya (2007). *Impact of Climate Change on Himalayan Glaciers and Glacial Lakes: Case Studies on GLOF and Associated Hazards in Nepal and Bhutan*. Edition: 0. Kathmandu, Nepal: International Centre for Integrated Mountain Development (ICIMOD). DOI: [10.53055/ICIMOD.470](https://doi.org/10.53055/ICIMOD.470). URL: <https://lib.icimod.org/record/22442> (visited on 09/29/2022).
- Shugar, D. H. et al. (July 16, 2021). "A massive rock and ice avalanche caused the 2021 disaster at Chamoli, Indian Himalaya." In: *Science* 373.6552, pp. 300–306. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.abh4455](https://doi.org/10.1126/science.abh4455). URL: <https://www.science.org/doi/10.1126/science.abh4455> (visited on 05/12/2023).
- Sicart, J. E. et al. (Nov. 15, 2006). "Incoming longwave radiation to melting snow: observations, sensitivity and estimation in Northern environments." In: *Hydrological Processes* 20.17, pp. 3697–3708. ISSN: 08856087, 10991085. DOI: [10.1002/hyp.6383](https://doi.org/10.1002/hyp.6383). URL: <https://onlinelibrary.wiley.com/doi/10.1002/hyp.6383> (visited on 05/12/2023).
- Silantyeva, O. et al. (Jan. 13, 2023). *Shyft and Rasputin: a toolbox for hydrologic simulations on triangular irregular networks*. preprint. Hydrology. DOI: [10.31223/X5CS95](https://doi.org/10.31223/X5CS95). URL: <https://eartharxiv.org/repository/view/4875/> (visited on 05/03/2023).
- Singh, P. and N. Kumar (Dec. 1997). "Effect of orography on precipitation in the western Himalayan region." In: *Journal of Hydrology* 199.1, pp. 183–206. ISSN: 00221694. DOI: [10.1016/S0022-1694\(96\)03222-2](https://doi.org/10.1016/S0022-1694(96)03222-2). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022169496032222> (visited on 04/27/2023).
- Siwakoti, M. et al. (2016). "Assessment of the effects of climate change on the distribution of invasive alien plant species in Nepal." In: *Building knowledge for climate resilience in Nepal*, pp. 5–8.
- Skaugen, T. and C. Onof (July 15, 2014). "A rainfall-runoff model parameterized from GIS and runoff data: Moddel Parameterized from GIS and Runodd Data." In: *Hydrological Processes* 28.15, pp. 4529–4542. ISSN: 08856087. DOI: [10.1002/hyp.9968](https://doi.org/10.1002/hyp.9968). URL: <https://onlinelibrary.wiley.com/doi/10.1002/hyp.9968> (visited on 04/16/2023).
- Skavang, J. Q. (2023). *Forecasting discharge for hydropower production in Central Himalaya, Nepal*. URL: <https://github.com/jacobqs/MST-Himalaya> (visited on 04/19/2023).

- Skavang, Q. J. et al. (2022). “Forecasting discharge for hydropower production in Central Himalaya, Nepal.” In: *RoHub*. DOI: <https://w3id.org/ro-id/648630a1-c6a3-4f64-8a39-e4cd29a55221>.
- Statista (2023). *Definition Trend*. URL: <https://www.statista.com/statistics-glossary/definition/425/trend/> (visited on 05/11/2023).
- Statkraft (2022). *Shyft: The Open Source Toolbox for Energy Markets*. In: URL: <https://shyft.readthedocs.io/en/latest/> (visited on 04/16/2023).
- (2023). *Shyft*. URL: <https://gitlab.com/shyft-os/shyft>.
- Stigter, E. E. et al. (Aug. 24, 2018). “The Importance of Snow Sublimation on a Himalayan Glacier.” In: *Frontiers in Earth Science* 6, p. 108. ISSN: 2296-6463. DOI: [10.3389/feart.2018.00108](https://www.frontiersin.org/article/10.3389/feart.2018.00108). URL: <https://www.frontiersin.org/article/10.3389/feart.2018.00108/full> (visited on 04/27/2023).
- Stocker, T., ed. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. New York: Cambridge University Press. 1535 pp. ISBN: 978-1-107-05799-9 978-1-107-66182-0.
- Sulla-Menashe, D. and M. A. Friedl (2018). “User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product.” In: NASA. URL: <https://modis.gsfc.nasa.gov/data/dataproduct/mod12.php>.
- Sun, A. et al. (May 2019). “Three-dimensional distribution of permafrost and responses to increasing air temperatures in the head waters of the Yellow River in High Asia.” In: *Science of The Total Environment* 666, pp. 321–336. ISSN: 00489697. DOI: [10.1016/j.scitotenv.2019.02.110](https://linkinghub.elsevier.com/retrieve/pii/S0048969719305923). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0048969719305923> (visited on 04/29/2023).
- Surveyor, Virtual (2023). *What is an EPSG Code?* URL: <https://support.virtual-surveyor.com/en/support/solutions/articles/1000261353-what-is-an-epsg-code-#:~:text=Modified%20on%3A%20Thu%2C%2020%20Apr,spheroids%2C%20units%20and%20such%20alike>. (visited on 04/23/2023).
- Teweldebrhan, A. T., J. F. Burkhart, and T. V. Schuler (Sept. 28, 2018). “Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches.” In: *Hydrology and Earth System Sciences* 22.9, pp. 5021–5039. ISSN: 1607-7938. DOI: [10.5194/hess-22-5021-2018](https://hess.copernicus.org/articles/22/5021/2018/). URL: <https://hess.copernicus.org/articles/22/5021/2018/> (visited on 04/19/2023).
- Tian, Y. and C. D. Peters-Lidard (Dec. 2010). “A global map of uncertainties in satellite-based precipitation measurements: Uncertainties in Precipitation Data.” In: *Geophysical Research Letters* 37.24, n/a–n/a. ISSN: 00948276. DOI: [10.1029/2010GL046008](http://doi.wiley.com/10.1029/2010GL046008). URL: <http://doi.wiley.com/10.1029/2010GL046008> (visited on 04/29/2023).
- Tiwari, S., S. C. Kar, and R. Bhatla (Apr. 2016). “Interannual Variability of Snow Water Equivalent (SWE) over Western Himalayas.” In: *Pure and Applied Geophysics* 173.4, pp. 1317–1335. ISSN: 0033-4553, 1420-9136. DOI: [10.1007/s00024-015-1163-1](http://link.springer.com/10.1007/s00024-015-1163-1). URL: <http://link.springer.com/10.1007/s00024-015-1163-1> (visited on 04/27/2023).
- Tucker, Gregory E. et al. (Oct. 2001). “An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks.” In: *Computers & Geosciences* 27.8, pp. 959–973. ISSN: 00983004. DOI: [10.1016/S0098-3004\(00\)00134-5](https://linkinghub.elsevier.com/retrieve/pii/S0098300400001345). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0098300400001345> (visited on 05/12/2023).
- Ueno, K. and T. Yamada (1990). “Diurnal variation of precipitation in Langtang valley, Nepal Himalayas.” In: *Bulletin of glacier research* 8, pp. 93–101.

Bibliography

- Uznanski, D. (2023). “Grid.” In: *MathWorld—A Wolfram Web Resource*, (). URL: <https://mathworld.wolfram.com/Grid.html> (visited on 04/29/2023).
- Vicente-Serrano, S. M., M. A. Saz-Sánchez, and J. Cuadrat (2003). “Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature.” In: *Climate Research* 24, pp. 161–180.
- Viviroli, D., D. R. Archer, et al. (Feb. 4, 2011). “Climate change and mountain water resources: overview and recommendations for research, management and policy.” In: *Hydrology and Earth System Sciences* 15.2, pp. 471–504. ISSN: 1607-7938. DOI: [10.5194/hess-15-471-2011](https://doi.org/10.5194/hess-15-471-2011). URL: <https://hess.copernicus.org/articles/15/471/2011/> (visited on 04/18/2023).
- Viviroli, D., H. H. Dürr, et al. (July 2007). “Mountains of the world, water towers for humanity: Typology, mapping, and global significance: Mountains as Water Towers for Humanity.” In: *Water Resources Research* 43.7. ISSN: 00431397. DOI: [10.1029/2006WR005653](https://doi.org/10.1029/2006WR005653). URL: <http://doi.wiley.com/10.1029/2006WR005653> (visited on 04/18/2023).
- Wallace, J. M. (1977). *Atmospheric science: an introductory survey*. Academic Press.
- Wang, L. and H. Liu (2006). “An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling.” In: *International Journal of Geographical Information Science* 20.2, pp. 193–213.
- Weedon, G. P., G. Balsamo, et al. (Sept. 2014). “The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data.” In: *Water Resources Research* 50.9, pp. 7505–7514. ISSN: 00431397. DOI: [10.1002/2014WR015638](https://doi.org/10.1002/2014WR015638). URL: <http://doi.wiley.com/10.1002/2014WR015638> (visited on 04/23/2023).
- Weedon, G. P., S. Gomes, et al. (Oct. 1, 2011). “Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century.” In: *Journal of Hydrometeorology* 12.5, pp. 823–848. ISSN: 1525-755X, 1525-7541. DOI: [10.1175/2011JHM1369.1](https://doi.org/10.1175/2011JHM1369.1). URL: <http://journals.ametsoc.org/doi/10.1175/2011JHM1369.1> (visited on 10/12/2022).
- Whitaker, J. (2023). *pyproj Documentation*. URL: <https://pyproj4.github.io/pyproj/stable/> (visited on 03/02/2023).
- Wilkinson, M. D. et al. (Mar. 15, 2016). “The FAIR Guiding Principles for scientific data management and stewardship.” In: *Scientific Data* 3.1, p. 160018. ISSN: 2052-4463. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18). URL: <https://www.nature.com/articles/sdata201618> (visited on 04/23/2023).
- Xarray (2023). *Xarray documentation*. URL: <https://docs.xarray.dev/en/stable/> (visited on 10/09/2022).
- Yatagai, A. et al. (Sept. 1, 2012). “APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges.” In: *Bulletin of the American Meteorological Society* 93.9, pp. 1401–1415. ISSN: 1520-0477. DOI: [10.1175/BAMS-D-11-00122.1](https://doi.org/10.1175/BAMS-D-11-00122.1). URL: <https://journals.ametsoc.org/doi/10.1175/BAMS-D-11-00122.1> (visited on 04/29/2023).
- Yvinec, M. (2023). “2D Triangulations.” In: *CGAL User and Reference Manual*. URL: https://doc.cgal.org/latest/Triangulation_2/index.html (visited on 05/03/2023).
- Zehra, F. et al. (Dec. 21, 2020). *Comparative Analysis of C++ and Python in Terms of Memory and Time*. preprint. MATHEMATICS & COMPUTER SCIENCE. DOI: [10.20944/preprints202012.0516.v1](https://doi.org/10.20944/preprints202012.0516.v1). URL: <https://www.preprints.org/manuscript/202012.0516/v1> (visited on 01/31/2023).