MASTER THESIS REPORT

Performance Evaluation of Global Hydrological Models for climate change projections in Pan-Arctic river basins

"Bewertung verschiedener globaler hydrologischer Modelle für die Auswirkung des Klimawandels in Pan-Arktischen Flußeinzugsgebieten"

Aashutosh Aryal
Matriculation Number: 3652872

Academic Tutor: apl. Prof. Dr.-Ing Frank Molkenthin
Brandenburg University of Technology (BTU) Cottbus-Senftenberg

Institutional Tutors: Dr. Anne Gädeke, Dr. Valentina Krysanova
Potsdam Institute for Climate Impact Research

15 August 2019

Declaration

I hereby declare that this work was written independently without any unauthorized assistance or sources not given credit within the work. All words, phrases, passages, and data taken from other sources have been properly cited. No parts of this work in the same form or a similar form have ever been previously handed in to fulfill an examination.

_____

Aashutosh Aryal

Cottbus, 15 August 2019

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my institutional tutors Dr. Anne Gädeke and Dr. Valentina Krysanova for accepting to supervise my master thesis study and guiding me throughout this process without which this study would have been arduous. I am grateful for their feedback and comments on my research study, which was very beneficial and a great learning experience for me. I am thankful to them for providing me with an opportunity to do my master's thesis in PIK Potsdam, Germany. I would especially like to thank Dr. Anne Gädeke for providing me with all the necessary data and explaining its complexity to me clearly. I am grateful to her for providing me with the necessary ArcGIS shape files of the study areas for obtaining geographic information to carry out data analysis works. I am also thankful to Dr. Anne for her assistance with R. She helped me with the R codes to solve some crucial parts of analyzing the data for the study. I would also like to acknowledge Ms. María del Rocío Rivas López for providing me with the snow water equivalent observation data for the study.

I want to express my sincere thanks to the EuroAquae program for accepting me as a student and providing me with such a prestigious opportunity to complete my master's degree in an international platform with internationally renowned institutions. I would not have gotten nearly this far without the immense help from my academic tutor, Professor Dr. Frank Molkenthin. I cannot thank him enough for his invaluable guidance, insight, knowledge, skills, and support imparted to me during 2 years of my master's study. I am especially grateful for all the time he took in making sure I had a pleasant and fruitful journey. Also, I thank all the professors and teachers who taught me and helped me understand and grow my knowledge in hydroinfomatics and water management field. I want to sincerely thank Ms. Nora Reichert for translating the Abstract of my study in the German language.

Finally, this journey would not have been possible without the dedicated support of my family. I want to thank my father Mr. Hemlal Aryal, my mother Mrs. Kiran Aryal and my sister Ms. Aayushree Aryal, for always believing in me and supporting me throughout this journey. Their words of encouragement consoled me through my most stressful times, and I shall forever be indebted to them.

Abstract

The Pan-Arctic region has become highly susceptible to the increased risks of climate warming. Climate warming has substantial implications in many biophysical states and processes that are strongly influenced by the threshold and phase change of the freezing point. While it may be challenging to anticipate the changes brought on by a changing climate, the potential effects on hydrological processes are profound and manifold. However, Global Hydrological Models (GHMs) make it possible to simulate the terrestrial water cycle on a global scale, so they are used in this study to analyze model simulation results of land surface hydrologic dynamics processes of six Pan-Arctic river basins. Here, GHMs simulated cold region processes, including river discharge and snow water equivalent. Here simulations from nine GHMs participating in the second phase of the Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP2a) were evaluated. The simulated discharges by each individual hydrological model were compared against the observed discharges for the period 1971-2000. In addition to a visual comparison, three efficiency criteria-NSE, PBIAS, and Bias in Standard Deviation were used to validate the model simulations for monthly hydrographs, seasonal dynamics, high flows, and low flows. Furthermore, the models were evaluated and rated on their performance using the threshold values assigned for each efficiency criteria, and then aggregated indices were estimated for each model and basin using rating scores of 1 (good performance), 0.5 (weak) and 0 (poor) for every criterion and gauging stations. The study revealed large uncertainty levels in simulated river discharges with uncalibrated models showing considerable bias when compared against observed discharge. Therefore, significant differences in model performance were identified. WaterGAP2, MATSIRO, and MPI-HM models performed better than the rest of the models. These models had a better ability to replicate observed discharge than the rest. Four to five models out of nine showed acceptable performance for high and low flows (aggregated index > 60%) in the Ob basin, in other basins their performance was much weaker. The models were able to reproduce seasonality of snow cover, but the bias was quite high in many cases (not estimated numerically). The large uncertainty and bias observed in the simulated values highlight the urgent need for model improvements of cold region hydrological processes in Global Hydrological Models.

ग्लोबल हाइड्रोलोजिकल मोडलको कार्य दक्षता मूल्यांकन बिधिबाट पान-आर्कटिक नदी बेसीनहरूमा जलवायु परिवर्तन प्रक्षेपण

कार्यकारी सारांश

पान-आर्कटिक क्षेत्र मौसम परिर्वतनको बढ्दो जोखिमहरूको लागि अति संवेदनशील भएको छ। मौसमको तापमान बृध्दीबाट धेरै वायफिजीकल अवस्था र प्रकृयाहरुमा भएका उल्लेँख्य असरले फ्रीजिङ प्वाइन्ट थ्रेसोल्ड र चरणमा ठूलो प्रभाव पारेको छ। जबकि यो मौसम परिर्वतनले ल्याइएको परिवर्तनहरूको पूर्वानुमान गर्न चुनौतीपूर्ण हुन सक्छ, हाइड्रोलोजिकल प्रक्रियाहरूमा सम्भावित प्रभावहरू गहन र धेरै गुणा छन्। यद्पि, ग्लोबल हाइड्रोलजिकल मोडेलहरू (GHMs) ले स्थलीय पानी चक्र ग्लोबल स्केलमा सिम्यूलेट (नक्कल) गर्न सम्भव बनाउँदछ, त्यसैले तिनीहरू यहाँ अवस्थित छ वटा पान-आर्कटिक नदी घाटीहरूमा जमिन सतहको जलवायु गतिशीलताका नतीजालाइ मोडेल सिमुलेशन विश्लेषण गर्न प्रयोग गरिएको छ। GHMs हिउँ समतहमा नदीको पानी बहाव सहित चिसो क्षेत्रका प्रक्रियाहरू सिम्यूलेट गरिएको छ। यहाँ दोस्रो चरणमा भाग लिइरहेका नौ वटा अन्तर क्षेत्रगत अन्तरिक तुलनाका परियोजनाहरुको (ISIMIP2a) GHMs सिमुलेसन प्रभाव मोडल मूल्याङकन गरिएका छन्। प्रत्येक हाइड्रोलोजिकल मोडलका सिम्यूलेटेड वहाव वा डिस्चार्जहरूको तुलना अवलोकनबाट प्राप्त सन् १९७१ देखि २००० सम्म अवधिका डिस्चार्ज तुलनात्मक सिमुलेसन गरियो। यसका अलावा, अवलोकनका आधारमा तुलना गर्न, तीन दक्षताका मापदण्डहरु- NSE, PBIAS, र Bias in Standard Deviation लाइ मासिक हाइड्रोग्राफहरू, मौसमी गतिशीलता, उच्च वहाब र कम बहावहरू सिमुलेसन मोडेलमा प्रयोग गरियो। यसका अलावा, तिनीहरूको कार्य दक्षता प्रत्येक दक्षता मापदण्डहरुको लागि तोकिएका थ्रेसोल्ड सूचकहरू प्रयोग, र त्यसपछि एकत्रित सूचकाङकहरू प्रत्येक मोडेल र बेसिनको मूल्याङकन गरिएको थियो। यसमा सूचकका आधार र गेजिंग स्टेशनहरूको लागि ० देखि १ को रेटिंग स्कोर प्रयोग गरी अनुमान गरिएको थियो, १-राम्रो प्रदर्शन, ०.५-कमजोर र ०-खराब प्रयोग गरिएको थियो। अध्ययनले सिमुलेटेड नदी डिस्चार्जमा ठूलो अनिश्चितताको स्तर पत्ता लगायो अव्यवस्थित मोडेलले अवलोकनको तुलनामा तुलनात्मक पूर्वाग्रह डिस्चार्ज देखाउँदछ, त्यसकारण, मोडेल प्रदर्शनमा उल्लेख्य भिन्नताहरू पहिचान गरियो। WaterGAP2, MATSIRO, र MPI-HM मोडेलहरूले बाँकी मोडेलहरू भन्दा राम्रो प्रदर्शन गरे। यी मोडेलहरूमा बाँकी भन्दा अवलोकन डिस्चार्जको प्रतिकृति (Replicate) गर्ने क्षमता राम्रो पाइयो। ओब बेसिनमा नौ वटा मोडलहरु मध्ये चार देखि पाँच वटाले उच्च र न्यून प्रवाहहरूमा एकत्रित इन्डेक्सम)ा ६० प्रतिशत भन्दा बढीको लागि स्वीकार्य प्रदर्शन देखायो (, अन्य बेसिनहरूमा उनीहरूको प्रदर्शन धेरै कमजोर देखियो। मौसम अनुसार हिउँले ढाकेको क्षेत्र प्रदर्शित गर्न मोडलहरु सक्षम देखिए, तर धेरै केसहरूमा पूर्वाग्रह धेरै उच्च देखियो संख्यात्मक रूपमा अनुमान गरिए)को छैन।(सिम्यूलेटेड मानहरूमा ठूलो अनिश्चितता र पूर्वाग्रह अवलोकन भएको प्रकाश पार्नुका साथै चिसो क्षेत्रमा ग्लोबल हाइड्रोलोजिकल मोडेलहरूमा हाइड्रोलोजिकल प्रक्रियाहरूमा सुधारहरूको तत्काल आवश्यकता जरुरी भएको पाइयो।

Abstrakt

Die Arktische Region ist in zu nehmendem Maß den Risiken der Klimaerwärmung ausgesetzt. Liegen die Temperaturen überwiegend unter null Grad Celsius oder nahe dem Gefrierpunkt in einer Region, hat die Klimaerwärmung dort besonders weitreichende Konsequenzen. Eine umfassende Prognose ist bisher kaum möglich und auch die Folgen für die sensitiven regionalen hydrologischen Zusammenhänge sind tiefgreifend und komplex. Globale Hydrologische Modelle (GHMs) wurden entwickelt um zumindest in diesem Aspekt präzisere Vorhersagen über die zukünftige Entwicklung treffen zu können. Diese Modelle ermöglichen die Simulation hydrologischer Prozesse auf globaler Ebene und werden in dieser Arbeit genutzt, um existierende Modelergebnisse von 6 arktischen Wassereinzugsgebieten zu überprüfen. GHMs eignen sich für die Simulation in arktischen Gebieten, da die verschieden Aggregatzuständen von Wasser und deren Auswirkung auf hydrologische Zusammenhänge berücksichtigt werden können. In dieser Arbeit wurden die Ergebnisse von 9 GMHs Simulationen bewertet, welche in der zweiten Phase des Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP2a) durchgeführt wurden. Die ermittelten Abflussmengen der einzelnen hydrologischen Modelle wurden mit den gemessenen Werten aus den Jahren 1971-2000 verglichen. Zusätzlich wurden drei statistische Auswertungen mit Hilfe der Standardabweichung, dem PBIAS und der NSE-Kriterien vorgenommen, um die Modellergebnisse zum monatlichen Abfluss, saisonale Unterschiede, Trockenphasen und Nassphasen zu validieren. Die Einschätzung der Modellkapazität Langzeittrends in den Messdaten, ohne systematische Über- oder Unterschätzung sowie saisonale Trends wiederzugeben, bildete ein zentrales Element dieser Arbeit. Darüber hinaus wurden die Modelle auch dahin bewertet, wie sie bei Standardwerten in der statistischen Auswertung abschneiden. Bei den Ergebnissen wurden dabei für jedes Kriterium und jede Messstation zwischen 3 Kategorien unterschieden: 1 (gute Übereinstimmung), 0.5 (mittlere Übereinstimmung) und 0 (schwache Übereinstimmung). Die Arbeit konnte aufzeigen, dass es große Unsicherheiten bei den Abflussmengen von unkalibrierten Modellen gibt. So konnten signifikante Unterschiede in der Qualität der Modellergebnisse bei den unterschiedlichen Modellen festgestellt werden. Die Modelle WaterGAP2, MATSIRO und MPI-HM schnitten am besten ab. Im Vergleich zu den anderen Modellen, waren diese drei eher in der Lage die Übereinstimmungen mit den Messwerten zu erzielen. Für die Wassereinzugsgebiete Ob und Mackenzie während sowohl Nassphasen, als auch während Trockenphasen haben die meisten Modelle gut abgeschnitten. Die meisten Modelle waren auch in der Lage die Dynamik der Schneedeckentiefe wiederzugeben. Eine schwache Übereinstimmung mit den gemessenen Daten bei zahlreichen Kriterien und Messstationen in den anderen Wassereinzugsgebieten unterstreicht jedoch die dringende Notwendigkeit für eine Verbesserung der existierenden Modelle für die kalten Regionen dieser Erde, um auch globale Trends in der Klimaerwärmung genauer bestimmen zu können.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Climate warming has become the greatest threat to the earth in recent decades with warming risks increasing at an unprecedented rate. The Pan-Arctic region has become highly susceptible to the increased risks of climate warming. In many biophysical conditions and processes, climate warming has significant consequences that are strongly influenced by the freezing point level and phase transition. Changes in the water cycle, particularly in freshwater flow, have potentially significant implications in lower latitudes for ocean circulation and climate. While it may be challenging to anticipate the changes brought on by a changing climate, the potential effects on global hydrological processes are profound and manifold. Impact of climate warming on the high latitude water cycle can be accessed via Global Hydrological Models (GHMs) which simulate cold region water cycle processes, including runoff, snow accumulation and melt, and soil freeze and thaw processes.

A Global Hydrological Model (GHM) is a simple representation of a real-world hydrological system at a global scale designed to understand various water and environmental processes, hydrological dynamics of land surfaces in continental river system [Gosling and Arnell, 2011], predict system behaviour and provide consistent impact assessment. GHMs are generally designed to assess water balance components (e.g., renewable water resources) from a hydrologic perspective, incorporating storage equations and subsequent flows between the storages. They usually calculate the vertical water balance (e.g., precipitation, runoff, canopy, snow, and soil storages) and contain a river routing scheme directly within the model.

There are several GHMs available today to simulate various hydrological parameters, and as such in this study, outputs from 9 GHMs were used to verify the accuracy of the parameters such as discharge and snow water equivalent. The 9 participating GHMs (DBH, H08, MATSIRO, MPI-HM, PCR-GLOBWB, WaterGAP2, LPJML, ORCHIDEE, and JULES-W1) from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2a) were driven by 4 climate forcing datasets (gswp3, princeton, watch, wfdei). The focus regions of this study were 6 Pan-Arctic river basins covering most of the Arctic region, namely Kolyma, Lena, Yenisei, Ob, Mackenzie, and Yukon. 18 stream gauge stations in total were selected to represent these river basins and to assess the performance of

models in those basins. The models in this study were evaluated and validated under historical climate conditions considering historical analysis period of 1971-2000 for the discharge output parameter and the historical analysis period of 1980-2000 for snow water equivalent variable.

The outputs from the model simulations were compared against the observations, and then the model performance was evaluated using subjective and objective assessment methods. The subjective assessment was done using visual comparison of the plots between simulated and observed data at gauging station under consideration and the objective assessment was carried out using different efficiency criteria like Nash-Sutcliffe Efficiency (NSE), Percent Bias (PBIAS) and Percent Bias in Standard Deviation ($\Delta\sigma$) to compare the simulated values against the observations. GHMs were evaluated using the process where monthly mean discharge (monthly dynamics) and long-term average monthly discharge (seasonal dynamics) were compared separately with the observed data.

The best and poor models were then chosen based on the analysis using threshold values of each model efficiency criteria with the rating score of 1 (good), 0.5 (weak) and 0 (poor), and then aggregated the scores of each model to calculate aggregate performance index of the models in percentage. Furthermore, GHMs' simulated discharge and observed discharge were evaluated using linear trend analysis. The models were also evaluated for other hydrological indicators like high flows, low flows, and extreme flows. Similarly, the models were evaluated subjectively in case of snow water equivalent using long-term average monthly simulated and observed time series. The performance of GHMs depend on the models themselves and on many different factors and conditions associated with the river basins under preview, indicating that in some basins, some GHMs may perform better and in others, they may perform poorly.

The proper evaluation of GHMs performance in the historical period and consideration of their results in climate impact studies may be helpful for increasing the confidence in projected climate change impacts and decreasing uncertainty of projections. The research study of climate change impact assessment for future projections could be considered as future pipeline work as a continuation of the current study.

1.1 Motivation for the Study

The primary motivation behind carrying out this study was because of the study areas considered. The Pan-Arctic region is considered as one of the isolated areas in the world where few people live, and the socio-economic scenarios of this region are minimal with few disturbances from the human-induced activities. This region, therefore, was not in highlights previously, but in recent years due to the increased hype of climate change and its impact, this region has suddenly come into focus in the scientific community as this region is highly susceptible to the climate change impacts. This increased vulnerability from climate warming risks motivated me towards doing this research study at first. Many scientific studies are being carried out in this region nowadays with new findings and discoveries made available from these studies, which motivated me further to work in this region and get new findings of this region from my research work. The evaluation of 6 big Pan-Arctic river basins using an ensemble of 9 Global Hydrological Models under one study was a unique concept, and there are very few researches carried out using similar scope of the study. So, I thought it would be a good idea to research something which has not been dealt with by other researchers working in the same field. This motivated me further to conduct the study. Mass data handling has become a very lucrative opportunity for fresh graduates nowadays, as many companies are seeking Data Analysts or Data Scientists. Therefore, I got motivated to do data analysis and visualization in this study using the R environment, thinking about my future career. Lastly, this study was carried out with an aim at increasing the model confidence in projecting the climate change impacts for the future.

1.2 Aims of the Study

The main aim of this study was to do a systematic evaluation of outputs of the ISIMIP global hydrological models (GHMs) for six Pan-Arctic river basins in the historical period 1971-2000 considering river discharge from 9 GHMs and snow water equivalent from 6 GHMs. For that, the following simulated GHMs indicators were evaluated and compared with indicators based on observed (water) or remote sensing (snow) data:

- Monthly and long-term average monthly discharge, percentiles of high and low flows at two to four gauging stations in every basin;
- Snow water equivalent in four to five points in every basin (only visual comparison)

In addition, trend analysis of the simulated and observed river discharge time series was done to compare the correspondence of the statistically significant trends.

1.3 Thesis Outline

This section aims to provide a brief outline of the works presented in the thesis report.

Chapter 2 is the "Theoretical Framework," where relevant topics related to the study are presented. This chapter mainly highlights the underlying theory and principles of the related topics that came across while carrying out research work. This chapter provides information on the key aspects of the study, such as Arctic Rivers, Pan-Arctic Watershed, Global Hydrological Models (GHMs), Performance of GHMs, etc.

Chapter 3 is the "Project Description," where the description of different components of the project is presented. All the essential components that are associated with the project under which this study is based on are discussed briefly in this chapter. The brief overview of the study areas along with a short description of ISIMIP, ISIMIP2a, river basins profile, GHMs profile, climate forcing datasets profile, observed datasets profile considered in this study are highlighted.

Chapter 4 is the "Technical Concepts" where the technicalities of the study are discussed. In this chapter, the working environment for the study, such as software, tools, and programming language, are briefly discussed. It provides details about ArcGIS, the NetCDF Interface and the R programming language. These tools provided technical support to this study, so relevant information about these tools are discussed in this thesis. Moreover, this chapter also discusses time series, time series analysis and visualization, model efficiency criteria, and hydrological indicators used in the study.

Chapter 5 is the "Data Processing," where information about the methods of data processing used in this study is highlighted. In this chapter, a short description of the historical analysis period considered for the data analysis along with ArcGIS processing works is covered. It also presents the working procedure for model output data acquisition. The description of various R packages used for the data processing and analysis are provided. It provides information about the data processing steps applied to analyze discharge and snow water equivalent time series used in this study. It covers the

working procedures used in the study on how to read, manipulate, aggregate, and plot time series data. It also highlights the methodology used for carrying out different analysis such as linear trend analysis, model evaluation, model performance aggregated index and extreme flows analysis using discharge.

Chapter 6 is the "Discharge Data Analysis: Results and Discussion" where it deals with the results and discussion parts of discharge time series analysis and visualization. In this chapter, the results of the different analysis carried out using discharge data are presented in detail. First, the results of model validation run by comparing simulated and observed data using both subjective and objective methods of assessment are presented, followed by model evaluation results highlighting best and poor models at each gauging station and finally presenting the main results of model performance evaluation in each basin using model performance aggregated index. This chapter also deals with the results from the observed and simulated time series linear trend analysis followed by the analysis of the extreme flows where high and low flows are also dealt with. The results of all these analyses are followed by a brief discussion on the obtained results.

Chapter 7 is the "Snow Water Equivalent Data Analysis: Results and Discussion" where it deals with the results and discussion parts of snow water equivalent time series analysis and visualization. In this chapter, the results of the analysis carried out using snowpack data are presented in detail. First, a brief description of the location of interests chosen in the basins is discussed, followed by the results and discussion of model validation run by comparing simulated and observed data using only subjective assessment are presented.

Chapter 8 is the "Conclusion and Recommendation" where the major findings of the study are presented along with some recommendations that could be applied to improve the study further in the future. The possibility of future pipeline works that could be carried out in line with the findings from this study is also discussed here. In the end, Appendices where all the supplementary works of this study are provided such as R scripts used for data processing, analysis and visualization, extra figures and tables and the research poster of the study presented at 2019 cross-sectoral ISIMIP workshop, which was held from 4 to 7 June 2019 in Paris, France.

2. THEORETICAL FRAMEWORK

2.1 Arctic River Basins

Arctic rivers include some of the world's largest river systems, linking a vast and diverse region to the much smaller area of the receiving water body, the Arctic Ocean. Annually, Arctic rivers bring approximately twice as much freshwater as net precipitation over the ocean [Haine et al., 2015; Carmack et al., 2016; Bring et al., 2017] and also serve as conveyors of nutrients, carbon and other elements from their numerous watersheds, some of which stretch as far south as the mid-latitudes [Bring et al., 2016; Bring et al., 2017]. The rivers are thus central to the circulation of freshwater in the Arctic, a process that is rapidly changing with potential implications for the global climate [Prowse et al., 2015a, 2015b; Newton et al., 2016; Bring et al., 2017].

The Arctic region has become essential to climate change impact considerations as this environment is important to the global climate system and is highly likely to be affected by global warming [Peterson et al., 2002]. Early warming signs can already be seen in the Arctic as this area is particularly sensitive to warming, and climate change has already had a dramatic impact on the Arctic ecosystem's terrestrial, aquatic and marine components. Arctic soils produce vast amounts of ancient organic matter that can be released as thaws of permafrost, fostering a positive feedback loop and exacerbating warming. Since the Arctic includes several of the largest rivers on Earth, these rivers have a disproportionate influence on the ocean as they bring more than 10% of the global river discharge into the Arctic Ocean, which carries just 1% of the global seawater amount [Wagner et. Al., 2011; Holmes et al., 2013]. Arctic rivers flood large watersheds and patterns in drainage and chemistry already provide credible insights into changes taking place across the Arctic landmass [Holmes et al., 2013].

The discharges from Arctic rivers contribute about 50% of the total influx of freshwater into the Arctic Ocean [Frappart et al., 2011]. Arctic hydrological systems experience significant temporal variability due to large changes in the circulation of the atmosphere [Proshutinsky et al., 1999; Frappart et al., 2011]. Discharge observations suggest that Arctic discharge has increased significantly since the mid-1930s, with an increasing trend reported in recent decades [Peterson et al., 2002; Frappart et al., 2011]. Winter snow-

mass accumulation and its resulting removal often affects the extent of northern river streamflow. In the northern river basins, snowmelt and related flooding during the spring / summer cycle are the most significant hydrological activity of the year [Yang et al., 2003; Frappart et al., 2011]. Many inevitable shifts in the pattern of snow cover at high latitudes, such as the early onset of snowmelt correlated with warming in the winter and spring seasons, may illustrate the instability of the hydrological regime in these areas in the perspective of global warming [Barnett et al., 2005; Frappart et al., 2011].

2.2 Pan-Arctic Watershed

Figure 1*: Pan-Arctic Watershed, showing its major river basins [*Arctic Great Rivers Observatory*]*

According to the definition used by the Arctic Great Rivers Observatory [www.arcticgreatrivers.org], "the Pan-Arctic watershed is the region draining into the Arctic Ocean plus the watersheds of the Yukon River and rivers entering the Bering Sea north of the Yukon River" (Figure 1). Based on this definition, the Pan-Arctic watershed

covers 16.8 x 10^6 km^2. This area is narrower than the most comprehensive descriptions of the Arctic because it does not include Alaska, the drainage of the Hudson Bay or the Canadian Archipelago, but contains the major rivers that join the Arctic Ocean, which in some instances has watersheds well below 60 degrees North. (Figure 1) [Holmes et al., 2013]. The red line indicates the boundary of the pan-Arctic watershed.

"There are 14 rivers in the Pan-Arctic watershed that have mean annual discharges exceeding 25 km^3y^{-1}. Remarkably, 12 of these rivers are in Russia. The Yenisei, Lena, and Ob are each among Earth's largest rivers, having mean annual discharges exceeding 400 km^3. Six rivers in the pan-Arctic watershed have basin areas exceeding 5,00,000 km^2 ("Big 6" - the Ob, Yenisei, Lena, Mackenzie, Yukon, and Kolyma). Combined, the watersheds of these "Big 6" Arctic rivers cover 11.2 x 10^6 km^2 or 67% of the pan-Arctic watershed. The next eight largest Arctic watersheds (the "Middle 8") together only cover an additional 1.9 x 10^6 km^2, much less than the basin area of the Ob, Yenisei, or Lena rivers alone" [Holmes et al., 2013]. It shows the importance of including the largest rivers to accomplish a Pan-Arctic convergence, but also demonstrates the difficulties of further scaling-up because each additional river beyond the "Big 6" achieves only incremental gains [Holmes et al., 2013]. This research study mainly focused on the Pan-Arctic watershed of the "Big 6" Arctic rivers located in Russia, Canada and Alaska (USA).

2.3 Key Processes in Arctic Terrestrial Hydrology

The Arctic hydrological cycle is a vital component of the climate system, both locally in the Arctic region through its incorporation of land, environment, atmosphere and human processes with the Arctic Ocean [Vörösmarty et al., 2001; Bring and Destouni, 2009], and globally through its links with worldwide ocean circulation and climate feedback mechanisms [Houghton et al. 2001; Bring and Destouni, 2009]. Arctic terrestrial hydrology also plays an important role in the Arctic system's freshwater circulation [Bring et al., 2016]. There are few main processes to be considered in the Arctic terrestrial freshwater system, primarily focusing on freshwater storages and fluxes such as precipitation, evapotranspiration, surface runoff, and channel flows, permafrost and groundwater hydrology, and river and lake ice. All these processes are equally important for maintaining the balance in the Arctic ecological systems.

Precipitation is the major component of the Arctic terrestrial hydrological system. A significant amount of annual Arctic precipitation is deposited as snow with heterogeneous spatial distribution and released into the river network during spring snowmelt in a relatively short window of time period [Mernild et al., 2014; Bring et al., 2016]. Regional rainfall, which is deposited as snow, often adds directly to runoff in summer months — the precipitation cycle influences several spatial levels of annual and seasonal water flow. The frequency and intensity of rainfall often influences the production of runoff, with the spring freshet often rivaled by summer thunderstorms for smaller basins [Kane et al., 2008; Bring et al., 2016].

Most of the precipitation returns to the atmosphere in the form of evapotranspiration (ET) over most of the Arctic basins. This process links the water and energy cycles, connecting the land to the atmosphere. The annual Arctic ET water flux is generally lower than annual precipitation, with the exception of a few southern inland areas [Serreze et al., 2006; Bring et al., 2016] and over lakes and wetlands where summer ET may surpass summer rainfall, or even annual precipitation [Bring et al., 2016]. Since the transpiration depends on the canopy of the vegetation, the ET varies considerably, both on local scales and in time. The key influences on the ET water flux are the duration of the snow cover and growing seasons along with landscape changes such as lake area change [Hinzman et al., 2005; Karlsson et al., 2015]. "Large-scale evapotranspiration values are difficult to estimate, but recent satellite-based assessments indicate pan-Arctic averages of $\sim$230 mm yr^{-1}" [Zhang et al., 2009].

Besides ET, river discharge is the other significant water flux out of Arctic basins. Freshwater flow through the main rivers in the Arctic that carries water, heat, sediments, carbon, and nutrients to the coasts and ultimately into the Arctic Ocean. Most of the water is transported to the Arctic Ocean during the spring snowmelt and in summer. "The peak flow rates (May–June) can exceed the mean annual flow rate as much as 40 times for the Yenisei and Lena Rivers; the corresponding ratio for the Mackenzie River is much less, about 5 times" [Bowling et al., 2000]. The seasonal flow of many Arctic rivers, including the Yenisei, Lena and Mackenzie, is strongly influenced by the dam construction [McClelland et al., 2004; Yang et al., 2014]. The major Arctic rivers sustain year-round flow activity, including the "winter flows under seasonal ice cover" [Prowse et al., 2011].

Nonetheless, smaller northern rivers sometimes freeze to the floor in the winter without any movement. “A portion of winter discharge is seasonally stored as river and lake ice and released in spring” [Prowse et al., 2011].

Permafrost often plays an influential role in both water chemistry and water flows in Arctic basins. Dynamics of effective layers rule a wide range of surface and subsurface processes through permafrost environments and runoff generation control mechanisms. Soil moisture is strongly influenced by the topographic relief that varies between mountains, slopes, and flat terrain. “Due to the vast extent of the area underlain by permafrost, the active layer thickness (ALT) and behaviour vary across the Arctic, which influences soil moisture and storage. The mechanical stability of soil is also influenced by the water and ice content of the active layer” [Instanes et al. 2016; Bring et al. 2016].

The seasonal changes in surface ice are prominent features of the Arctic river systems on shorter time scales. “During winter, lake and river ice grow to cover 1.7×10^6 km^2, an area approximately equal to the Greenland ice sheet. The peak volume of 1.6×10^3 km^3 roughly matches the Northern Hemisphere snowpack on land” [Brooks et al., 2013]. Such freshwater ice causes various effects in Arctic freshwater reservoirs and stream networks on physical systems, ecosystem services, and socio-economic systems [Instanes et al., 2016], although the hydrological influences of many effects arise well outside sub-Arctic latitudes, via the headwaters of the large north-flowing rivers [Bennett and Prowse, 2010].

2.4 Global Hydrological Models (GHMs)

A hydrological model is simply a representation of a real-world hydrological system used to understand various water and environmental processes, predict system behaviour, and provide consistent impact assessment. Moreover, Global Hydrological Models (GHMs) are intended to use for “modelling the hydrological dynamics of land surfaces, to the scale of continental river system” [Gosling and Arnell, 2011], thus bridging the gap between Global Climate Models (GCMs) and Land Surface Models (LSMs). “GHMs simulate the hydrological response to meteorological variations with run-off generation and river routing processes” [Sood and Smakhtin, 2015]. There is, however, a difference in the

system definition, along with the parameter estimation, the temporal & spatial data resolution and the simulation step [Haddeland et al., 2011].

“GHMs provide information about hydrological balance and various processes involved at the global and continental-scale” [Sood and Smakhtin, 2015]. They are mainly built for global-scale studies. The research work on these models and their performance to predict the projections of future outcomes has become essential in recent times with the increase in natural disasters caused by climate change. The GHMs are relatively new and have managed to effectively increase activity to become a separate research field in the last two decades and a concerted effort is evolving. The plethora of global data available from satellites in the last two decades has helped a lot in the development of GHMs. As such, GHMs have been used in a wide range of applications including “short to extended-range flood forecasting, climate assessment, hazard and risk-mapping, drought prediction and water resource assessment” [Sood and Smakhtin, 2015].

As stated by Sood and Smakhtin [2015], GHMs have “different spatial and temporal resolutions, parameter estimation approaches, number of parameters, calibration methods, input-output variables, and overall structure”. These have few calibratable criteria and are either calibrated based on eco-regions, climatic areas, or large drainage basins. The resolution of GHMs is determined by the resolution of available global climate input data. As more functionality is added to them, GHMs are becoming more complex and resolute and finer global spatial datasets are becoming available. Issues of models ' responsiveness to differing spatial and temporal levels of input and output information, modeling uncertainty estimation and GHM coupling with other systems have become prevalent [Sood and Smakhtin, 2015].

Some of the GHMs do not even calibrate their parameter as such but provide the so-called 'adjustment factor' for tuning/adaptation to the data observed [Vörösmarty et al., 1998]. Water-Global Analysis and Prognosis model (WaterGAP) [Alcamo et al., 2003; Döll et al., 2003] is known for “its calibration and tuning mechanism”. “It first calibrates a single parameter, and the remaining parameters are tuned accordingly, depending on their over or underestimation of observed data” [Döll et al., 2003]. GHMs are used efficiently in data-scarce regions of the world, and with satellite data being more subjective, their demand

increases dramatically, at least for the initial calculation of the water budget, and projection scenarios of shifts in those figures. Increased use of these models can be seen when used for application-oriented plans such as water demand and supply, diagnostic prediction of discharge (capturing change in infrastructure and land cover), and study of climate change impacts [Sood and Smakhtin, 2015].

As discussed in Sood and Smakhtin [2015], in basin-scale applications, GHMs would not be the preferred choice because of the coarser resolution of GHMs at present and the fact that there is a large family of hydrological models designed for this purpose. However, GHMs may provide valuable spatial and temporal estimates of global water resources and help to analyze possible projections/scenarios of changes of those estimates; GHMs have been built effectively for this purpose. Global projections generated by GHMs would be an improvement over those focused on quantitative review of actual ground-based data but remains constrained on a global level and thus retains a great deal of uncertainty. Improved use of GHMs is exposed when connected to other models, such as explaining the global economy, agriculture, commerce, ecosystems, energy balance, land-use transition, climate change, plant growth and other water-related sustainability issues/components. [Sood and Smakhtin, 2015].

The structure, capabilities, and shortcomings of individual models are studied, and researchers worldwide are thoroughly researching the aims of model analysis and their validity in climate change predictions [Sood and Smakhtin, 2015]. The findings from these GHMs are currently discussing issues such as model uncertainty, data scarcity, integration with emerging and new remote sensing datasets, and spatial resolution. In addition, expanded efforts to address water-related issues in developing countries, the advancement of flood prediction services, developments in the collection of weather input data and advancements in numerical modeling techniques have actually helped such models to evolve in recent years [Sood and Smakhtin, 2015].

As stated by Abramowitz et al. [2008], "Land Surface Models (LSMs) are the components of Global Climate Models (GCMs) that predominantly simulate land surface processes, for instance, the absorption rate and hence infiltration, proportions of incoming and outgoing radiations, carbon and moisture content in a land system." They are provided

with meteorological conditions, “usually based on satellite as inputs that would generally include precipitation, temperature, wind speed, etc.” and produce outputs that include “latent and sensible heat fluxes, CO_2 fluxes, reflected solar and emitted longwave radiation, surface runoff and deep soil drainage” [Abramowitz et al., 2008]. “LSMs contain internal state variables such as soil moisture and temperature, vegetation and soil carbon pools, and snow and ice volume and density” [Abramowitz et al. 2008]. LSMs provide time-invariant parameters for the simulation duration that represent the properties of soil and vegetation. The most important outputs for climate simulation on decadal to century time scales are the turbulent fluxes, particularly latent heat and net CO_2 exchange of ecosystems [Abramowitz et al. 2008].

These models are based on energy balance solutions and water balance equation solutions. When parameterizing dynamic components such as evapotranspiration, vegetation land use, etc., energy balance helps. LSMs typically use a one-dimensional column system to characterize the root area, specific vegetation and therefore evapotranspiration by their measured flux [Overgaard et al. 2006]. These models can provide information at one point and generally facilitate future projections for different climatic conditions on the other. LSM development began early in the 1980s, from basin level to global scale. With advancements in remote sensing and satellite data providing more accurate and qualitative information than ever before, LSMs are becoming increasingly popular, particularly in data-scarce regions.

On the other side, it is potentially easy to parameterize the physical process under a one-dimensional column at least in the hydrological process [Koster Randal D. et al. 2000]. The hydrological process seeks, where possible, the finest form of distributed modeling. Such form captures all its system variables in space, but with its 1-D column approach it is difficult to capture them in LSMs. Soil moisture, for example, as one of its parameters, depends on various factors such as precipitation, surface temperature, irrigation, groundwater, etc. [Overgaard et al. 2006], known for interchanging flux within the grid cell. Nonetheless, for their individual grid cell simulation, such spatial flux exchange (between grid cells) is generally ignored, rendering its parameter not spatially representative.

2.5 Performance of GHMs

As suggested by Krysanova et al. [2018], it is crucial to analyze the performance of the model at the particular location in the same period as found in the observation datasets for that location. It is necessary to realize that sometimes a non-calibrated model may perform well enough, and the calibration may lead to problems related to over-tuning [Krysanova et al. 2018]. Several indicators and evaluation criteria, describing particular hydrological signature[s] (e.g., river discharge, or extreme flows) are used to determine the performance of a model. To assume that a model performs well, those indicators and evaluation criteria must be within the acceptable limits. The judgment of a "good performance category" depends on the considered hydrological signature, its temporal resolution, spatial scale, evaluation criteria used and quality of observational data [Kauffeldt et al. 2013; Beven and Smith 2015; Krysanova et al. 2018].

According to Krysanova et al. [2018], the performance of GHMs depends on the location and catchment scale, indicating that in some places, GHMs may perform well and in others, they may perform poorly. There have been numerous studies carried out to evaluate the performance of GHMs by comparing different simulated aspects of runoff. An interesting study that was mentioned in Krysanova et al. [2018] was the comprehensive evaluation of monthly discharge for more than 1000 gauge stations globally carried out using the PCR-GLOBWB and WaterGAP2.2 models [Van Beek et al. 2011, Müller Schmied et al. 2014], where the quality of climate forcing data was also discussed in relation to the evaluation results.

Some observations such as systematic overestimation of runoff in arid and semi-arid areas and systematic underestimation of discharge in high-latitude river basins were regarded as the basis for poor model performance from some GHMs understudy in Krysanova et al. [2018] paper. From the above observations, it was indicated that systematic biases across models are prevalent when representing processes such as snowmelt in high latitudes and evaporation in drylands [Gerten et al. 2004; Krysanova et al. 2018] and it was partly due to the climate forcing datasets and their uncertainties. Therefore, it was realized that more comprehensive and systematic evaluation for all models with the inclusion of other variables other than discharge needs to be carried out using the same set of evaluation metrics and observational databases.

Often, GHMs are used for studying changes in mean seasonal dynamics and their performance is evaluated using river discharge in many large-scale catchments but impact analysis is carried out for other hydrological indicators on maps at all scales from far upstream to far downstream and for a number of internal model variables such as snow cover, evapotranspiration, soil moisture, etc. [Krysanova et al. 2018]. Nowadays, GHMs are being used to investigate changes in extreme runoff characteristics, such as magnitude and frequency of high/low flows, floods, hydrological droughts and water scarcity [Dankers et al. 2014, Prudhomme et al. 2014, Schewe et al. 2014, Arnell and Gosling 2016, Gosling et al. 2017; Krysanova et al. 2018].

As mentioned by Krysanova et al. [2018], several other studies have focused on evaluating the GHM performance in more detail, for instance, a set of three to nine uncalibrated GHMs in Europe using a database of discharge observations in very small, pristine catchments [mostly sub-grid scale] which were assumed to represent grid-scale runoff. The evaluations included spatially aggregated runoff percentiles [Gudmundsson et al. 2012a], seasonality of the runoff at the grid and spatially aggregated scales [Gudmundsson et al. 2012b], and indices describing extremes of runoff [Prudhomme et al. 2011]. These studies contributed to understanding model performance at the grid to sub-catchment scale. These studies highlighted large variations in model outputs and model performance when outputs are aggregated over larger areas, leading to biases in both mean and variability of simulated runoff, and increasing biases at the extreme ends of the flow duration curve. It was shown that in many cases, individual models perform extremely poorly [Krysanova et al. 2018].

Krysanova et al. [2018] discussed a study by Beck et al. [2016], where they presented a globally regionalized model evaluated in over 1787 catchments of various sizes by comparing with the performance of 9 GHMs using standard metrics. They showed that the median performance of many of these models across the range of evaluated catchments is rather poor, indicating that the observed mean is a better predictor than the GHMs.

2.6 Uncertainties associated with GHMs

GHMs tend to have uncertainties associated with them due to various reasons. The sources of uncertainties can be found in every step of the modelling process. The errors could be in the measurements for climate input variables that are assigned in the model, or could be found in the model structure and other model parameters, due to model simplifications and unknown or non-ascertainable relations within the model [Jin et al. 2010; Setegn et al. 2010; Rafiei Emam et al. 2018]. To improve the reliability of modelling results, identifying sources and quantifying uncertainty in hydrological modelling is important. "The uncertainty analysis will help minimize difficulties in the calibration of hydrological models" [Rafiei Emam et al. 2018].

"Uncertainties are also associated when representing various physical processes in global hydrological models (GHMs)" [Gosling et al. 2010]. There are uncertainties in measurement/estimation of every component of a water balance to varying degrees, depending on the component. For instance, low spatiotemporal resolution of rainfall data, river discharge, groundwater flow, and human water abstraction/return flows. Evapotranspiration, particularly in forests may be the most significant uncertainty in estimation. Then there is interannual variability, which in the absence of a long time series adds to further uncertainty. The use of land cover maps focused on satellite imagery used for water balance and hydro modeling at the watershed level often relies on the accuracy of classification. Full management of model uncertainty is important so that decision-makers are confronted with a spectrum of information regarding GHM effect projections. This uncertainty is often underexplored due to the number of simulations required for human and computer processing time [Gosling et al. 2010].

The properties of the errors depend on the origin of uncertainty; thus, different techniques must be followed to handle such inconsistencies in order to obtain viable results from the simulations of the hydrological model [Liu and Gupta, 2007]. The calibration of hydrological models is a challenging task due to various sources of uncertainties. Nonetheless, parameter uncertainty is easy to control between different sources of uncertainty by various calibration procedures [Vrugt et al. 2003; Yang et al. 2007, 2008; Wu and Chen 2015; Rafiei Emam et al. 2018].

2.7 Influence of Model Performance Evaluation

Model evaluation is an essential step in the modelling process to get credibility from the scientific community and acceptance of model results for future projections from the user community. Model evaluation should be specific to the scale, location, and indicator [Krysanova et al., 2018] for which the predictions are being simulated. It is a suitable method to use multiple indicators for evaluation, mainly when the model output is intended to be used for decision-making support [Krysanova et al., 2018]. The model simulation outputs obtained from GHMs are assessed for their accuracy to predict the actual observations found at a site in a given time. The simulation results derived from the models will help to determine whether the model performed well or not based on how accurately simulation results were able to reproduce the observed values.

As stated by Krysanova et al. [2018], good performance of hydrological models in the historical period increases confidence for future projections and decreases uncertainty related to hydrological models. Moreover, good performing hydrological models correctly predict the real-life condition and hence, increase the credibility of their simulation results for both scientists and users, leading to a decrease in uncertainty limits. However, it should be kept in mind that the good performance of a hydrological model in the historical period is necessary, but this is not enough to extrapolate the model's capabilities for the future. It is not enough, because good performance under historical conditions is not a guarantee per se for good performance under different climatic conditions (the model might not account for processes that could occur in a changing climate) [Krysanova et al., 2018].

Good performance of a hydrological model in today's climate does not guarantee robust results under different climates [Krysanova et al., 2018]. This could be, in principle, resolved by designing frameworks for comprehensive model evaluation that takes into account model responses to changing climate, and model responses to several key processes such as runoff, evapotranspiration and snow (as a focus only on streamflow may be too simplistic) as argued in Krysanova et al. [2018]. These requirements obviously, are quite challenging for the modelling community due to the complexities arising to meet them. However, it seems that this is the only way to achieving more robust

predictions and future climatic projections and lowering uncertainty related to hydrological models [Krysanova et al., 2018].

It is the characteristic nature of GHM modelling that model performance varies according to locations, as complete tuning and validation are usually not possible due to lack of high-resolution input data at the global scale, comparably high computation costs, and intentional focus on representation of large-scale patterns and a variety of processes [Krysanova et al., 2018]. However, there are few steps to improve the situation, for instance, moving to a finer resolution of gHMs and applying regionalized calibration, according to Krysanova et al. [2018]. All properly evaluated models with good performance are plausible enough, but to improve the robustness of results and to achieve higher credibility from these results, poor performing models that are giving implausible outcomes should be discarded. Hydrological model evaluation specifically for indicators is the only way to estimate ranges of model capabilities and, thereby, to safeguard against the model's use for tasks beyond its demonstrated (legitimate) capabilities [Krysanova et al., 2018].

3. PROJECT DESCRIPTION

3.1 The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP)

ISIMIP was initiated by the Potsdam Institute for Climate Impact Research (PIK) and the International Institute for Applied Systems Analysis (IIASA). "ISIMIP is a community-driven climate-impacts modelling initiative aimed at contributing to a quantitative and comprehensive cross-sectoral synthesis of the differential impacts of politically and scientifically relevant climate change scenarios, including the associated uncertainties" [ISIMIP, 2018]. The key features of mission and process of ISIMIP are shown in Figure 2.

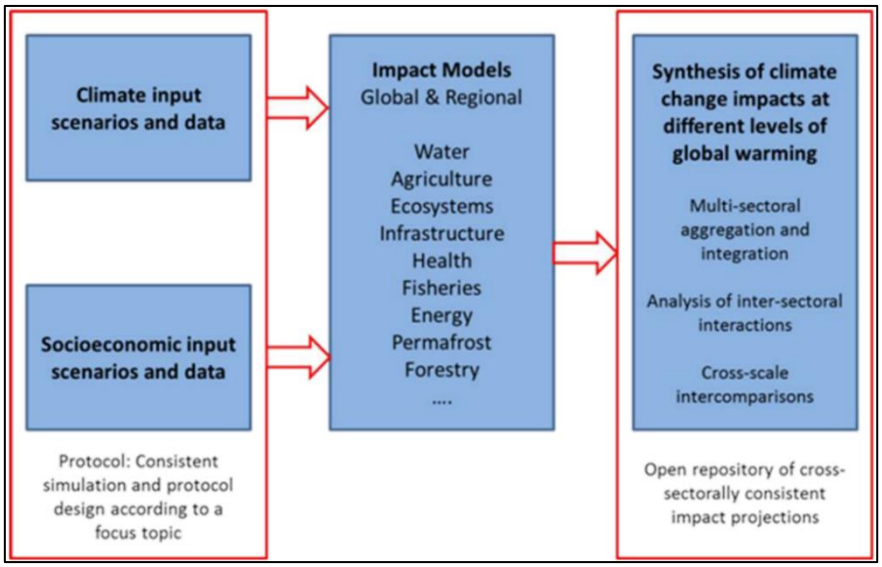


Figure 2: General structure of the ISIMIP process and mission [ISIMIP, 2018]

"ISIMIP aims to improve global and regional risk management from climate change risks through a multi-impact model framework that integrates climate-impact simulations across sectors, and scales" [ISIMIP, 2018]. ISIMIP make use of multi-model ensembles so that discrepancies can be quantitatively evaluated at the different modelling stages [ISIMIP, 2018]. Variations between the models can be used to define critical processes and parameterizations and to guide efforts to improve the model. Both the development of the system and the quantification of uncertainty were essential to the reliable and insightful analysis of climate threats. "ISIMIP is organized into simulation rounds, which are guided by a focus topic. For each round, a simulation protocol defines a set of common simulation scenarios based on the focus topic" [ISIMIP, 2018]. The simulation round that was considered in this study was ISIMIP2a, where models were evaluated with a focus on the representation of the impacts of extreme events.

3.2 ISIMIP2a Historical Simulation Runs

“ISIMIP2a is the second ISIMIP simulation round, focusing on historical simulations (1971-2010 approx.) of climate impacts on agriculture, fisheries, permafrost, biomes, regional and global water, and forests” [ISIMIP2a, 2018]. This can function as a framework for model analysis and improvement allowing accurate forecasts of climate change's biophysical and socio-economic impacts at different levels of global warming. The focus topic of ISIMIP2a is model evaluation and validation representing the extreme weather events impacts and climate variability threats. “It is explicitly designed to evaluate the models’ ability to reproduce observed historical variability, responses to extreme climatic events such as heatwaves, droughts, floods, heavy rains and storms, and representation of extreme impact events In this simulation run, four common global observational climate data sets were provided across all impact models and sectors. Also, appropriate observational data sets of impacts for each sector were collected, against which the models can be benchmarked” [ISIMIP2a, 2018].

Historical validation experiment's main objective is to test the capacity of existing impact models to replicate observed features of simulated variables, with an emphasis on (but not limited to) variability and extreme events. Simulations are structured to simulate historical conditions as closely as possible, within the constraints of, for instance, the availability of historical forcing data, range of model formulations and resources for model development [ISIMIP2a, 2018].

In order to compare different historical simulations, four observation-based historical climate datasets were selected to force the impact models. The set of data is unique in terms of the temporal scale, reliability of variables, prior implementations and typically reflects a realistic reconstruction of the earth's climate of the past 100 years [ISIMIP2a, 2018]. “The different historical simulations will allow a systematic quantification of the effect of the choice of forcing data on impact model results; allow comparison to previous studies using either of these datasets; and provide an extensive database for model evaluation and impact assessment, in particular with regard to the focus topic” [ISIMIP2a, 2018]. The common input data for the simulation runs are atmospheric data, oceanic data, land-use/land-cover, and soil data and socio-economic input.

3.3 Overview of Study Areas

Here in this section, a brief overview of the study areas under consideration for the research study is presented. In this research study, GHMs output data from ISIMIP2a simulation runs uploaded in the framework of the Inter-Sectoral Impact Model Intercomparison Project was used. The details about the ISIMIP and ISIMIP2a simulation runs are presented in the upcoming sections below. Figure 3 below shows the Pan-Arctic map with 6 river basins along with 18 gauging stations under study, which is highlighted by red dots. Similarly, Table 1 provides few more details about the river basins and the gauging stations under preview. The brief profile of all 6 river basins that are considered in this study is provided below. 9 Global Hydrological Models are being considered out of the available 13 GHMs in the ISIMIP2a data archive under global (water) sector. Table 2 provides the details of the participating models in this study with some of their main characteristics. Furthermore, a brief profile of all 9 participating models is presented below. All the historical simulation runs from 1971-2000 are carried out using 4 climate forcing datasets available in the ISIMIP data archive for all the participating models. Table 3 provides details of the climate forcing data sets used for the historical validation runs. A brief description of 4 climate forcing datasets is presented below.

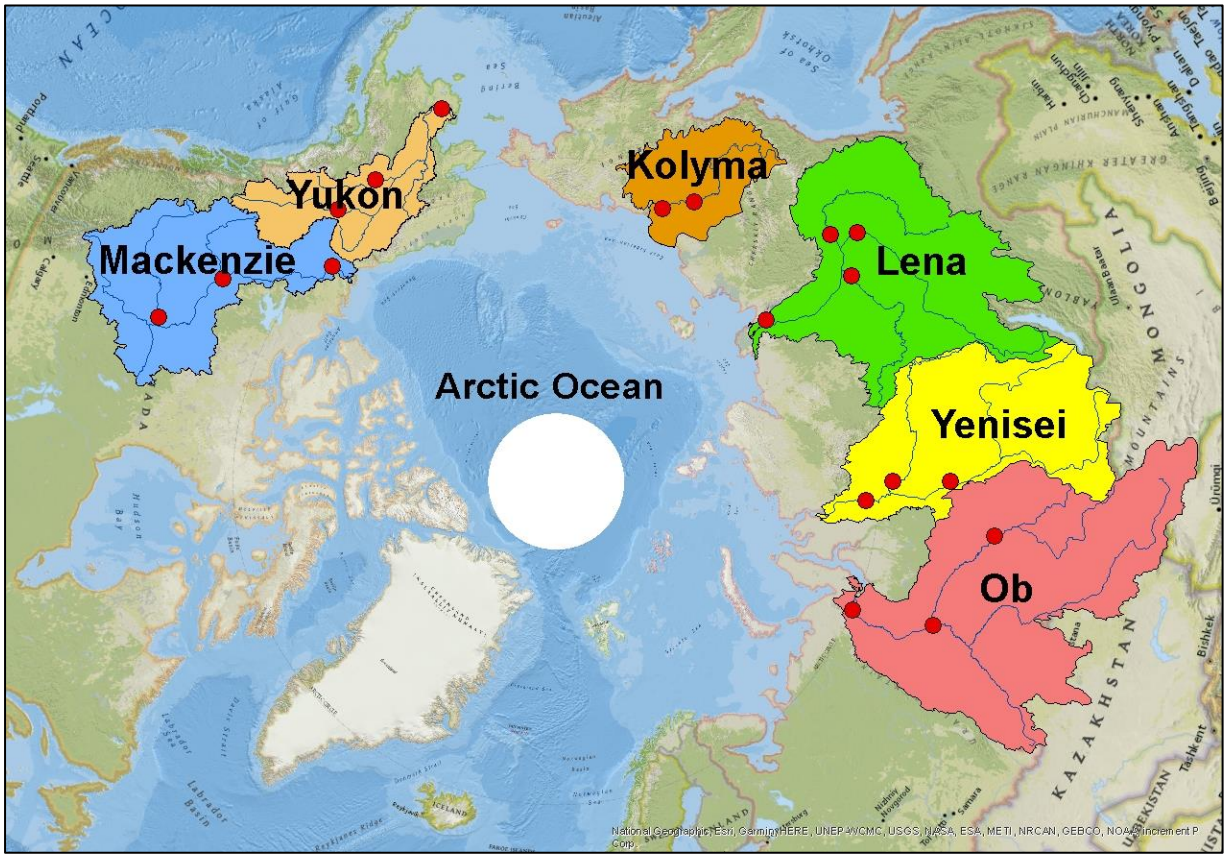


Figure 3: Pan-Arctic map showing the 6 river basins and 18 gauging stations understudy

Table 1: River Basins and Gauging Stations Details

River Basins	Countries	Basin Areas	Gauging Stations
Kolyma	Russia	0.64M km^2	**Kolymskaya** Sredne-Kolymsk
Lena	Russia	2.42M km^2	**Kusur** Verkhoyanski Perevoz Tabaga Hatyrik-Homo
Yenisei	Russia Mongolia	2.7M km^2	**Igarka** Bol. Porog Pod. Tunguska
Ob	Russia Kazakhstan China Mongolia	2.43M km^2	**Salekhard** Hanti-Mansisk Kolpashevo
Mackenzie	Canada	1.8M km^2	**Arctic Red River** Peace Point Alberta Fort Simpson
Yukon	Canada USA	0.83M km^2	**Pilot Point AK** Eagle AK Nenana AK

Table 2: Participating Models including their main characteristics

Model Name	Model Time Step	Spatial Resolution	Input Climate Variables	Runoff Routing/Routing Data	PET Method	Snow Scheme
WaterGAP 2	Daily	0.5° x 0.5°	P, T, LW, SW	Linear Reservoir (DDM30)	Priestley Taylor	Degree-day
DBH	Daily	0.5° x 0.5°	Tmax, T, Tmin, LW, Q, SW, SP, P	Linear Reservoir (DDM30)	Energy balance	Energy balance
H08	Daily	0.5° x 0.5°	T, LW, W, SW, S, SP, P	DDM30	Bulk Formula	Energy balance
MATSIRO	3-hourly	0.5° x 0.5°	T, LW, Q, SW, S, SP, P	DDM30	Penman-Monteith	Energy balance
MPI-HM	Daily	0.5° x 0.5°	T, P	Linear Reservoir (DDM30)	Penman-Monteith	Degree-day
PCR-GLOBWB	Daily	0.5° x 0.5°	T, P	Travel Time Routing	Hamon	Degree-day
LPJmL	Daily	0.5° x 0.5°	T, LWnet, SW, P	DDM30	Priestley Taylor	Degree-day
ORCHIDEE	Daily	0.5° x 0.5°	Tmax, Tmin, LW, W, Q, SW, SP, P	DDM30	Penman-Monteith	Energy balance

JULES-W1	Daily	0.5° x 0.5°	Tmax, T, Tmin, LW, W, Q, SW, SP, P	CaMaFlood Routing Model	Penman-Monteith	Energy balance

S = snowfall rate; P = total precipitation; T = daily mean air temperature; Tmax = daily maximum air temperature; Tmin = daily minimum air temperature; W = near-surface wind speed; Q = relative humidity; LW = longwave downward radiation; LWnet = longwave net radiation; SW = shortwave downward radiation; and SP = surface pressure

Table 3: Historical (Atmospheric) climate forcing datasets

Climate Forcing Datasets	Reanalysis Datasets	Years	Resolution	Coverage
GSWP3 [Kim n.d.]	20CR	1901-2012	0.5º	Land + Ocean
PGMFD v.2 (Princeton) [Sheffield et al., 2006]	NCEP/NCAR Reanalysis 1	1901-2010	0.5º	Land + Ocean
WATCH (WFD) [Weedon et al. 2011]	ERA-40	1901-2001	0.5º	Land
WFDEI.GPCC [Weedon et al. 2014]	ERA-Interim	1901-2012 (with 1901-1978 taken from WFD, WFDEI.GPCC data starting in 1979)	0.5º	Land

Finally, observation datasets are used to compare the output results from the model simulation runs using the climate drivers (climate forcing datasets). The output variables obtained from the historical simulation runs and which are to be considered in this study are streamflow (discharge) and snow cover (snow water equivalent). The observation data for streamflow at 18 gauging stations under consideration is obtained from the Global Streamflow Indices and Metadata Archive (GSIM) [Do et al. 2018]. The details about this dataset are presented briefly below. Similarly, the observation data for snow cover parameter, which in this case will be the data of snow water equivalent is obtained from the GLOBSNOW-2 SWE Product [Luojus et al. 2010; Pulliainen 2006; Takala et al. 2011]. The details about this remote sensing product are also presented briefly below.

3.4 River Basins Profile

Lena River Basin

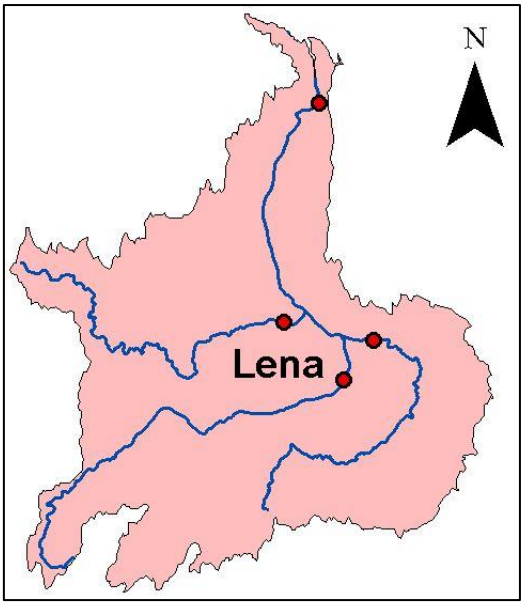


Figure 4: Lena River Basin

This brief profile of Lena River Basin is excerpted from Degtyarev [2016]: "The Lena (Figure 4) is one of the world's largest and longest rivers, renowned for its' distinguished length of 4,400 km with a catchment area of 2.5 million km^2. It drains one-fifth of Russia's territory and is entirely located within Russia. Its source starts in the Baikal Mountains and flows northeast to the Laptev Sea, and into the Arctic Ocean. The drainage system consists of 242,000 streams, 98% of which are up to 25 km in length, while 1.7% is over 1,000 km in length. A large delta extends 100 km into the Laptev Sea and is about 400 km wide. The water's maximum temperature is 19°C in the upper and middle reaches and 14°C in the lower reaches."

According to Degtyarev [2016], "The river remains iced over for approximately seven months: freezing happens in September – October, and spring break-up and ice melt begins in late April and early June. During the spring months, the river is prone to major flooding. During the ice melting season, ice jams accompanied by flooding (sometimes catastrophic) are very common when the water level can rise to 28 m. The river's primary source of water is precipitation and its spring peak flow is 530 times greater than its winter minimum. The annual water discharge ranges between 417 and 631 km^3. The Lena is one of the few undisturbed streams with no manmade structures along its watercourse today."

Kolyma River Basin

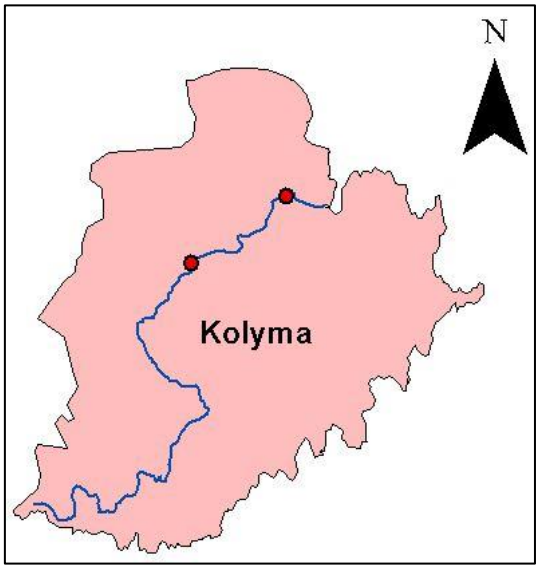


Figure 5: Kolyma River Basin

This brief profile of Kolyma River Basin is excerpted from the Encyclopaedia Britannica [2016]: "Kolyma River (Figure 5) lies in the north-eastern Siberia, far eastern Russia, rising in the Kolyma Mountains. It begins at the confluence of the Kulu River and the Ayan Yuryakh River. The river is 2,129 km long and drains an area of 647,000 km^2. In its upper course, it flows through narrow gorges, which makes the flow more rapid. Gradually its valley widens, and below Zyryanka it enters the wide, flat, and swampy Kolyma Lowland and flows north-eastward to discharge into the Kolyma Gulf of the East Siberian Sea, a division of the Arctic Ocean. The average discharge of the river is around 3,254 m^3/s, with a high flow of 26,201 m^3/s reported on June 1985, and low flow of 30.6 m^3/s reported on April 1979."

According to Encyclopaedia Britannica [2016], "The entire course is meandering and braided below the mountains with many channels stretching across a wide floodplain. The rim of the Yukaghir Plateau creates a steep, high-right bank in this lower reach. While hampered by a bar at its mouth, the river is navigable upstream to the Bakhapcha confluence, but the ice-free period is short: freezing conditions persist in the lower course from late September to early June, and as the southerly upper reaches start to thaw first, massive ice jams and extensive floods follow the breakdown of the ice. Nearly half of the annual discharge of the river comes in late spring and early summer. Gold mining is the only significant economic operation in the upper reach."

Ob River Basin

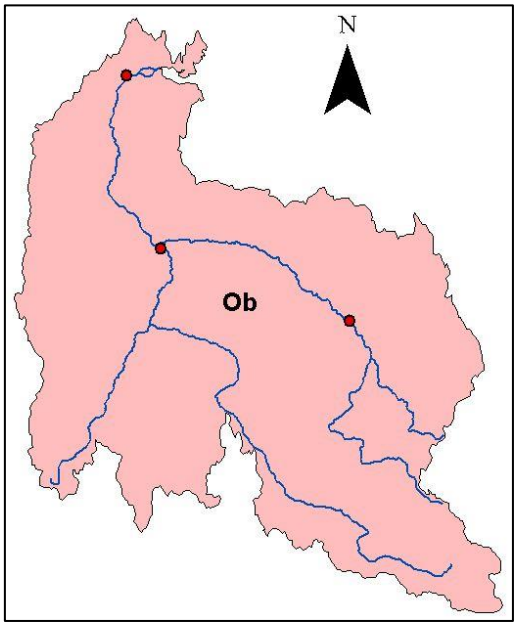


Figure 6: Ob River Basin

This brief profile of Ob River Basin is excerpted from Micklin [2018]: "Ob River (Figure 6) has a course of 3,650 km long, and if the Irtysh River is regarded as part of the main course rather than as the Ob's major tributary, then the length of the river is 5,410 km, making the Ob the seventh longest river in the world. The catchment area is approximately 2.98 million km^2. The Ob's catchment area is the sixth largest in the world. There are more than 1,900 rivers within the basin, with an aggregate length of about 180,000 km. The Irtysh, a left-bank tributary 4,250 km long, itself drains about 1.6 million km^2, that is 70 percent of the whole basin is drained by left-bank tributaries."

According to Micklin [2018], "Rainfall, which falls mainly in the summer, averages less than 400 mm per year in the north, 500–600 mm in the taiga zone, and 300–400 mm on the steppes. Snow cover lasts for 240 to 270 days in the north and 160 to 170 days in the south. The Ob has the third greatest discharge of Siberia's rivers, after the Yenisei and the Lena. On average, it pours some 400 km^3 of water annually into the Arctic Ocean—about 12 percent of that ocean's total intake from drainage. The volume of flow at Salekhard, just above the delta, is about 42,000 m^3/s at its maximum and 2,000 m^3/s at its minimum, while on the upper Ob, the corresponding figures are 9,600 and 200 m^3/s. The average annual discharge rate at the river's mouth is about 12,700 m^3/s. Most of the water comes from the melting of seasonal snow and rainfall; much less of it comes from groundwater, mountain snow, and glaciers."

Yenisei River Basin

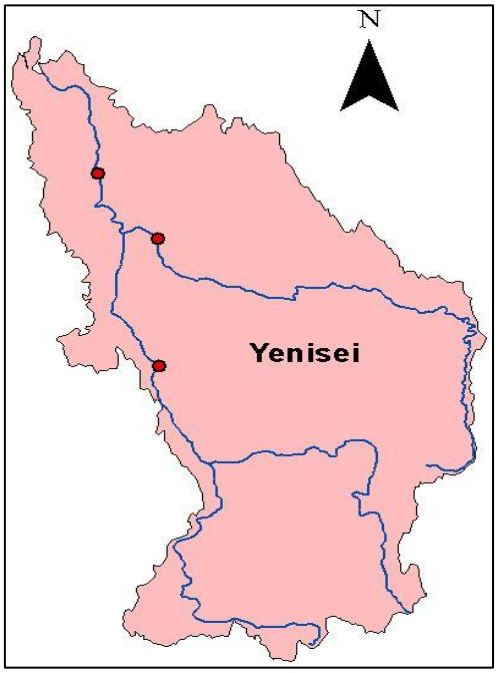


Figure 7: Yenisei River Basin

This brief profile of Yenisei River Basin is excerpted from Micklin [2016]: "Yenisei River (Figure 7) is one of the longest rivers in Asia and the world's sixth largest river in terms of discharge that lies in central Russia. The river begins at the confluence of its headstreams-the Great Yenisei and the Little Yenisei and flows along the border between eastern and western Siberia, before emptying into the icy Kara Sea. If the Great Yenisei is considered the source, then the river is 4,090 km long and if the headwaters of the Selenga River, which flow through Lake Baikal into the Yenisei, is considered as the source then the length reaches to 5,539 km long. The system within Siberia's boundaries comprises some 20,000 tributary or sub-tributary streams, with an aggregate length of approximately 885,000 km."

According to Micklin [2016], "About half of the Yenisei's water comes from snow, a little more than one-third from rainwater, and the remainder from groundwater. The river has violent spring floods followed by summer and autumn rain floods; in winter the runoff is reduced sharply, but levels remain high as ice jams are formed. In terms of runoff, the Yenisei is the largest river in Russia, with about 620 km^3 annually. Annual precipitation averages 400 to 500 mm in the north; 500 to more than 750 mm in the central portion, and up to 1,190 mm in the mountains south of the basin. Snow cover is light in most of the basin, averaging 40 cm in the south, and 60 cm in the north."

Mackenzie River Basin

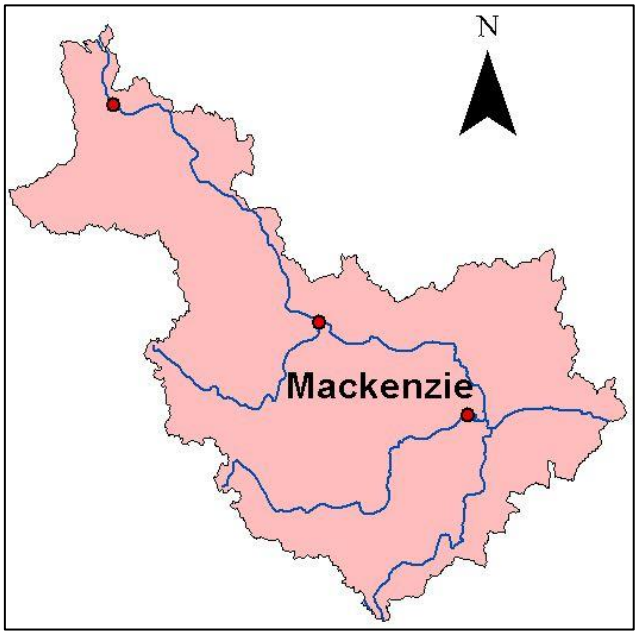


Figure 8: Mackenzie River Basin

This brief profile of Mackenzie River Basin is excerpted from Robinson [2019]: "Mackenzie River (Figure 8) is the main river system in the north-western North America. Its basin is the largest in Canada. The Mackenzie drains an area of around 1.8 million km^2. The entire river system runs for 4,241 km through the lake-strewn Canadian north to empty into the Beaufort Sea in the Arctic Ocean. The Mackenzie itself is 1,650 km long, according to the conventional measurement from Great Slave Lake. The river is generally wide, mostly from 1.6 to 3.2 km across, and in island-dotted sections, 4.8 to 6.4 km wide. It has a strong flow. The region is subject to a harsh winter climate, and its resources are few and less accessible than those of southern Canada."

According to Robinson [2019], "The ice on the Mackenzie River begins to break up in early to mid-May in its southern section. Tributary rivers are free of ice before the Mackenzie itself, and high water and flooding are common during the breakup period, mainly when ice dams form. The ice across the lower Mackenzie River breaks up in late May; the channels in the Mackenzie River delta are usually free of floating river ice by the end of May or early June. The average summer precipitation recorded at dispersed settlements is only about 175 to 200 mm; total annual precipitation throughout the Mackenzie valley is 250 to 355 mm. The mean annual discharge of the river on the lower course is 8,550 m^3/s but flows during the summer months usually will average 13,030 m^3/s."

Yukon River Basin

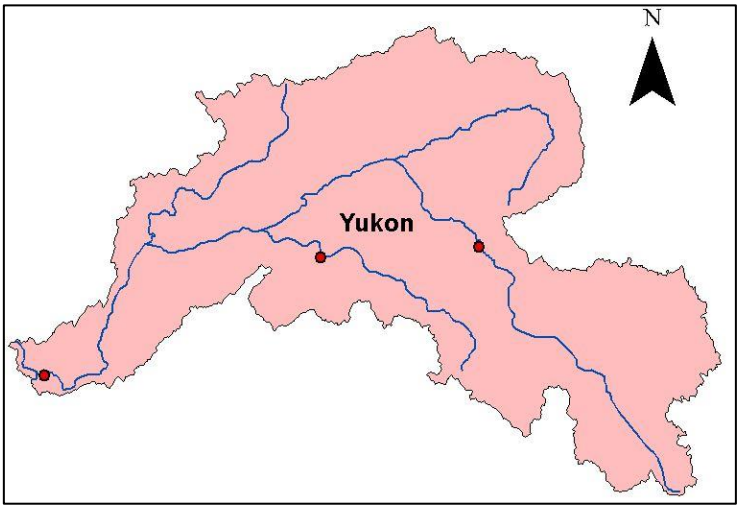


Figure 9: Yukon River Basin

This brief profile of Yukon River Basin is excerpted from Robinson [2016]: "Yukon River (Figure 9) is a major North American river that flows through north-western Canada and the central region of Alaska in the USA. It measures 3,190 km long from the headwaters. Its headwater tributaries drain an area of about 850,000 km^2. The river originates in the deep, narrow gorges in the Mountains with rapid water currents but on the lower reach it flows through broad, generally flat-bottomed valley. The snow-fed tributaries from the Mackenzie Mountains to the east reach their flood peak in June, but glacier-fed streams do not have their greatest runoff until July. Once these peak runoff flows have passed, the Yukon River becomes a shallow braided stream, as summer precipitation in the Yukon basin is low-about 150 mm. At Dawson, the Yukon has an average flow of 2,095 m^3/s."

According to Robinson [2019], "The valleys of the Yukon River basin have a subarctic climate with relatively warm, short summers; but the treeless upper mountain slopes are classified as having an Arctic climate. January mean temperatures are about -31°C at Dawson and about -21°C at Whitehorse. Average summer temperatures in the valleys of this high-latitude region, with its long hours of summer daylight, are moderately warm, with July mean temperatures of about 16°C at Dawson and slightly lower at Whitehorse. Annual precipitation is low, as it is in most of north-western Canada and central Alaska, with an average of about 260 mm recorded at Whitehorse and 305 mm at Dawson. Almost half of the precipitation falls as rain during the four summer months."

3.5 GHMs (Impact Models) Profile

"The ISIMIP2a water (global) outputs are based on simulations from global hydrology models according to the ISIMIP2a protocol. The models simulate hydrological processes and dynamics based on climate and physio-geographical information" [ISIMIP2a, 2018]. A brief profile of the 9 Global Hydrological Models (GHMs) used in this study is presented below. These hydrological models follow the ISIMIP2a protocol to form the base for the ISIMIP2a global water sector outputs.

WaterGAP2

The WaterGAP model was developed at the Centre for Environmental Systems Research of the University of Kassel, Germany, in cooperation with the National Institute of Public Health and the Environment of The Netherlands (RIVM). A global water assessment model called "WaterGAP2" [Alcamo et al. 2003; Schmied et al. 2016] has two main components—a Global Water Use model and a Global Hydrology model for the purpose of computing water use and availability on the river basin level. "The Global Water Use model consists of (a) domestic and industry sectors which consider the effect of structural and technological changes on water use, and (b) an agriculture sector which accounts especially for the effect of climate on irrigation water requirements. The Global Hydrology model calculates surface runoff and groundwater recharge based on the computation of daily water balances of the soil and canopy" [Alcamo et al. 2003]. WaterGAP2 uses daily time steps for calculation and the spatial resolution of this model is 0.5° by 0.5° [Schmied et al. 2016]. Each of the four climate forcings GSWP3, PGFv2, WFD and WFDEI_hom were calibrated against average annual discharge at 1319 discharge observation stations from the Global Runoff Data Center (GRDC) catalogue with the exception of WFD, where only 1312 stations could be used due to the earlier end of the forcing time series. [Schmied et al. 2016].

DBH

"The full form of DBH Model is Distributed Biosphere Hydrological Model. This model was developed to represent the effects of natural and anthropogenic heterogeneity on the water and energy balances of large river basins, mainly looking at the realistic estimates of evapotranspiration and runoff. The DBH model employs a flow intervals discretization scheme to represent geomorphologic properties and a simple biosphere

model to represent land cover characteristics to simulate integrated hydrological processes in a large river basin. The modelling framework embeds a biosphere model into a distributed hydrological scheme to represent topography and vegetation conditions in a mesoscale hydrological simulation. The model was applied to the Yellow River Basin to investigate its applicability to a region with significant variations in topography, land cover, and climate. The hydrological simulation has implemented by hourly time step with a spatial resolution of 10 km mesh" [Tang et al. 2006].

H08

"H08 is a grid-cell based global hydrological model, consisting of six sub-models, namely land surface hydrology, river routing, reservoir operation, crop growth, environmental flow, and water abstraction to assess global water availability. H08 spatially covers the whole globe at a resolution of 0.5°×0.5° to assess geographical heterogeneity of hydrology and simulate both natural and anthropogenic water use globally (excluding Antarctica). Simulation period is typically for several decades, and the calculation interval is a day" [Hanasaki et al., 2008]. The input meteorological forcing used by this model is the second Global Soil Wetness Project (GSWP2). The simulated runoff was compared and validated in 32 major river gauging stations worldwide with observation-based global runoff data sets and observed streamflow records [Hanasaki et al., 2008].

MATSIRO

The minimal advanced treatments of surface interaction and runoff (MATSIRO) has been developed at the Center for Climate System Research, the University of Tokyo, and the National Institute for Environmental Studies [Takata et al., 2003]. "MATSIRO is an LSM developed to compute biophysical exchanges in the GCM called MIROC (the Model for Interdisciplinary Research on Climate)" [Takata et al., 2003]. MATSIRO incorporated anthropogenic water control modules from the H08 model [Takata et al., 2003]. The model was developed with time scales ranging from a month to a few centuries for global and regional climate studies. "The model estimates the radiation transfer, the evaporation, the transpiration, the snow, the runoff, and so on considering the effects of vegetation, and solve the energy and water exchange between land and atmosphere, at that time and spatial scales" [Pokhrel et al., 2012]. "The spatial resolution of forcing data and simulations is 1° x 1° (longitude and latitude) global grids with a land-sea mask

defined by GSWP2 is used as climate data input in the modelling study. The atmospheric forcing data covers 29-years of simulation period from 1979 to 2007. The data include 6-hourly precipitation, temperature, radiations, surface pressure, specific humidity, and wind speed based on the atmospheric reanalysis data" [Pokhrel et al., 2012].

MPI-HM

"The MPI-HM, the Max Planck Institute – Hydrology Model computes the global water cycle only. The MPI-HM is a global hydrological model which solves the land surface water balance at a horizontal resolution of 0.5° with a time step of 1 day. It is restricted to the computation of water balance and does not consider any energy balance calculations. The MPI-HM consists of two formerly separated sub-components, the Simplified Land surface Scheme (SL-Scheme) and the Hydrological Discharge Model (HD-Model). The SL-Scheme includes a simple snow scheme based on the degree-day approach and uses a soil bucket scheme for the computation of the vertical water balance. The main outputs of the SL-Scheme are daily fields of runoff and drainage. These are given to the HD-Model, which is a state-of-the-art river routing model. It computes the retention time of water in overflow, baseflow, and river flow reservoirs using a linear reservoir cascade. The MPI-HM requires daily temperature and precipitation data as a climate forcing. Optionally, it is possible to use prescribed forcing for potential evapotranspiration (PET) instead of relying on the native PET calculation in the MPI-HM, which is based on the Thornthwaite formula. In this study, all simulations were conducted for the period 1958–1999" [Stacke et al. 2012].

PCR-GLOBWB

"The global hydrological model, PCR-GLOBWB simulates for each grid cell (0.5 degree globally over the land) and for each time step (daily) the water storage in two vertically stacked soil layers and an underlying groundwater layer, as well as the water exchange between the layers (infiltration, percolation, and capillary rise) and between the top layer and the atmosphere (rainfall, evapotranspiration, and snowmelt). The model also calculates canopy interception and snow storage. Water use for agriculture, industry, and households is dynamically linked to hydrological simulation at a daily time step. The simulated local direct runoff, interflow, and baseflow are routed through the river network that is also linked to water allocation and reservoir operation scheme. The

groundwater store is explicitly parameterized based on lithology and topography and represented as a linear reservoir model. The ensuing capillary rise is calculated as the upward moisture flux that can be sustained when an upward gradient exists, and the moisture content of the soil is below field capacity" [Wada et al. 2014].

LPJmL

"The dynamic global vegetation and water balance model LPJmL (Lund-Potsdam-Jena managed Land Model) is based on LPJ, a multi-sectorial dynamic global vegetation model that computes the establishment, growth and productivity of the world's major natural and agricultural plant types and the associated carbon and water fluxes as well as their spatiotemporal variations in response to climatic conditions and human interferences such as irrigation, typically on a 0.5° grid and at daily time steps. LPJmL model results have been extensively validated against small- and large-scale biophysical and biogeographical observations, including leaf phenology, CO_2 fluxes, and crop yields. In the model, precipitation and irrigation water is partitioned into soil moisture, transpiration, soil evaporation, interception, and runoff. Productive water consumption ET is calculated as a function of soil water supply and atmospheric demand. Unproductive water consumption consists of interception loss, and evaporation from soil, lakes, and canals" [Sitch et al. 2003].

ORCHIDEE

"The Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) is a process-based model for the interannual variability (IAV) of the fraction of absorbed active radiation, the gross primary productivity (GPP), soil moisture, and evapotranspiration (ET). ORCHIDEE is a land surface model that simulates carbon, water, and energy exchanges within ecosystems and between the land and the atmosphere" [Traore et al. 2014]. "ORCHIDEE includes two main modules: 1. SECHIBA (Schématisation des Echanges Hydriques à l'Interface entre la Biosphère et l'Atmosphère) that simulates energy and water exchanges between the atmosphere and the land surface. SECHIBA includes the two soil hydrology configurations. 2. STOMATE (Saclay Toulouse Orsay Model for the Analysis of Terrestrial Ecosystems), that simulates phenology and carbon dynamics" [Guimberteau et al. 2014].

JULES-W1

"A new community land surface model called the Joint UK Land Environment Simulator (JULES) was developed by the Met Office Surface Exchange Scheme (MOSES) at the UK Met Office for applications ranging from operational weather forecasting to Earth system modelling. It can be used as a standalone land surface model driven by observed forcing data or coupled to an atmospheric global circulation model. The JULES model has been coupled to the Met Office Unified Model (UM) and as such provides a unique opportunity for the research community to contribute their research to improve both, world-leading operational weather forecasting and climate change prediction systems" [Best et al. 2011]. JULES is developed as a modular structural dynamic modelling framework in which the connections between the modules illustrate the physical processes connecting these areas. This modular design enables the replacement of modules or the inclusion in the modelling process of new modules [Best et al. 2011].

3.6 Climate Forcing Datasets

In the ISIMIP2a, four state-of-the-art climate forcing datasets: GSWP3, PGFv2, WATCH (WFD), and WFD_WFDEI are available. For each forcing datasets, daily values of the variables such as precipitation, 2 m air temperature, shortwave downward radiation and longwave downward radiation at the surface level are used. In all data sets, "daily precipitation estimates were obtained by bias correcting the output of weather models by monthly precipitation data sets that had been derived from monthly precipitation observed at rain gages" [Schmied et al. 2016].

Global Soil Wetness Project 3 (GSWP3)

The third phase of the Global Soil Wetness Project (GSWP) integrates century-long (1901–2010) high-resolution global climate data [Kim 2014]. "The 20th Century Reanalysis (20CR) project done with the NCEP atmosphere land model which had a relatively low spatial resolution (~ 2.0°) and long-term availability (140 years) was dynamically downscaled into 0.5° resolution using Experimental Climate Prediction Center (ECPC) Global Spectral Model (GSM) by spectral nudging data assimilation technique. Also, Global Precipitation Climatology Centre (GPCC) version 6 (for P), Climate Research Unit (CRU) TS3.21 (for T), and Surface Radiation Budget project (SRB, for

SWD/LWD) were used for bias correction to reduce model-dependent uncertainty" [Schmied et al. 2016].

GSWP3 consists of three experiments: long-term retrospective (EXP1), long-term future climate (EXP2), and a short-term super-ensemble (EXP3) [Kim 2014]. Since the study focused on the historical simulations runs, EXP1 (long-term retrospective experiment) is highlighted here. "The long-term retrospective experiment investigates how interactions among energy-water- carbon cycles have changed through the past century (1901-2010) using a multi-model approach over a global 0.5° land grid. A standard product of the project will be an extensive set of land fluxes and state variables, which has the potential to serve as a long-term land-surface reanalysis. It also will serve as a reference set for the long-term variability of various processes in the terrestrial hydro-energy-eco system that respond to surrounding large-scale climate variability, such as changes in extremes (e.g., flood and drought), land carbon balances, and water and energy inputs to the atmosphere" [Kim 2014].

Princeton Global Meteorological Forcing Dataset (PGMFD v.2)

The Princeton Global Meteorological Forcing Dataset, version 2 (PGFv2) is an update of the forcing described by Sheffield et al. [2006]. "The dataset is constructed by combining a suite of global observation-based datasets with the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis and covers the period 1901–2012. This dataset provides near-surface meteorological data for driving land surface models and other terrestrial modelling systems. Daily-resolution observed climate data on a global (land and ocean) 0.5°x0.5° latitude-longitude grid are recorded by the Terrestrial Hydrology Group at Princeton University, based on the reanalysis data set (NCEP/NCAR Reanalysis) and use the bias targets Climate Research Unit (CRU), Surface Radiation Budget (SRB), Tropical Rainfall Measuring Mission (TRMM), Global Precipitation Climatology Project (GPCP) & World Meteorological Organization (WMO) and validated against the bias-corrected forcing dataset of the second Global Soil Wetness Project (GSWP2)" [Sheffield et al. 2006].

"Precipitation is disaggregated in space to 1.0 degree by statistical downscaling using GPCP daily product. Disaggregation in time from daily to 3 hourly is accomplished

similarly, using the TRMM 3-hourly real-time dataset. Similarly, other meteorological variables (downward short- and longwave, specific humidity, surface air pressure, and wind speed) are also downscaled in space with account for changes in elevation. The final product provides a long-term, globally consistent dataset of near-surface meteorological variables that can be used to drive models of the terrestrial hydrologic and ecological processes for the study of seasonal and inter-annual variability and the evaluation of coupled models and other land surface prediction schemes" [Sheffield et al. 2006].

WATCH (WFD)

"The WATCH Forcing Data (WFD) was developed under the scope of the European FP6-funded Water and Global Change (WATCH) project (https://www.eu-watch.org) to investigate the nature of the global water cycle on land" [Weedon et al., 2011]. "The data set is based on the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year reanalysis product (ERA-40) for the period 1958–2001 and on the reordered ERA-40 data for the period 1901–1957. The variables from ERA-40 are interpolated to half-degree resolution, and some are adjusted to monthly observation data, e.g., Precipitation is corrected using GPCC version 4 observations. Monthly Temperature is corrected to CRU TS2.1, and Shortwave Downward Radiation (SWD) is corrected to cloud cover of CRU TS2.1, whereas Longwave Downward Radiation (LWD) is not bias-corrected" [Weedon et al., 2011]. "The WFD is stored at 67,420 half-degree resolution land points only (excluding Antarctica) and, due to storage limitations, provided at 6-hourly time steps for five variables and 3-hourly time steps for three others" [Weedon et al., 2011].

Combined WFD and WFDEI (WFD_WFDEI)

WATCH Forcing Data methodology applied to ERA-Interim data (WFDEI) uses the same methodology as the WFD, but there are slight differences in the basic data, processing, and formatting [Weedon et al., 2014]. "The WFDEI data set was created by applying the WFD methodology to the newer ERA-Interim reanalysis data of ECMWF, which is improved compared to ERA-40, especially for SWD. The data set covers the period 1901-2012, where the data for 1901-1978 are taken from WFD, and 1979 onwards from WFDEI. Daily-resolution observed climate data on a global (land only) scale is 0.5°x0.5° latitude-longitude grid, based on the reanalysis data set ERA-Interim and using the biased target GPCC" [Weedon et al., 2014].

3.7 Observations Datasets

To validate the outputs from the GHMs used in this study, they had to be compared with the observations data. So, in this regard, for the discharge output variable, the Global Streamflow Indices and Metadata Archive (GSIM) was used to access the daily streamflow metadata of the 18 gauging stations under study. Similarly, to validate the snow cover data, the GLOBSNOW-2 Snow Water Equivalent (SWE) Remote Sensing Product was used to get access to snow water equivalent observations data for the basins under study. The brief details of both these data sets are presented below.

The Global Streamflow Indices and Metadata Archive (GSIM)

"The Global Streamflow Indices and Metadata Archive (GSIM) is a worldwide collection of metadata and indices derived from more than 35,000 daily streamflow time series" [Do et al. 2018]. The GSIM project was initiated to address the demand for a global streamflow database. GSIM can be used to advance large-scale hydrological research and to improve knowledge of the global water cycle using the information from an enormous number of stations and related metadata [Do et al., 2018]. "GSIM describes the development of three metadata products: (1) a GSIM catalogue collating necessary metadata associated with each time series, (2) catchment boundaries for the contributing area of each gauge, and (3) catchment metadata extracted from 12 gridded global data products representing essential properties such as land cover type, soil type, and climate and topographic characteristics" [Do et al., 2018].

"It compiled the daily streamflow time series from 12 databases (5 research databases and 7 national databases) that have either open access or restricted-access policies. The list of databases identified as part of GSIM is not exhaustive of all possible data sources, only of those that were known to the authors and readily accessible within the project time frame. The various data sources were classified as either a "research database" or a "national database." Research databases with daily streamflow data have been compiled on an ad hoc basis from a variety of sources by research organizations. National databases with daily streamflow data are made publicly available by national water authorities as part of water-related regulations" [Do et al., 2018].

GLOBSNOW-2 SWE Remote Sensing Product

"The GlobSnow project, funded by the European Space Agency (ESA), has resulted in new hemispherical records of snow parameters intended for climate research purposes. The snow water equivalent dataset is produced using a combination of passive microwave radiometer (satellite-retrieved information) and ground-based weather station data, spanning 34 years (1979-2013)" [Luojus et al. 2010]. As stated in Luojus et al. [2010], "the GlobSnow SWE record utilizes a data-assimilation based approach combining satellite microwave radiometer-based measurements (SMMR, SSM/I and SSMIS) with data from ground-based synoptic weather stations. The satellite sensors utilized provide data at K- and Ka-bands (19 GHz and 37 GHz respectively) at a spatial resolution of approximately 25 km." The SWE record is produced on a daily, weekly, and monthly basis. SWE information is provided for terrestrial non-mountainous regions of Northern Hemisphere, excluding glaciers and Greenland [Luojus et al. 2010].

"Due to the nature of the radiometer observations, the SWE product is reliably shown on areas with seasonal dry snow cover. Areas with occasional wet snow or a thin snow layer are not reliably detected and typically not present on the SWE product. The areas marked as snow-free may thus include areas with occasional wet snow cover" [Luojus et al. 2010]. "The snow water equivalent describes the amount of liquid water in the snowpack that would be formed if the snowpack was completely melted" [Luojus et al. 2010]. According to Luojus et al. [2010], "the SWE production system v2.0 utilizes SWE retrieval methodology [Pulliainen 2006] complemented with a time-series melt-detection algorithm [Takala et al. 2011]. The SWE retrieval and melt detection algorithms are combined to produce snow water equivalent maps incorporated with information on the extent of snow cover on coarse resolution (25x25km grid cells). The SWE estimates are complemented with uncertainty information on a grid cell level."

4. TECHNICAL CONCEPTS

4.1 ArcGIS

"A geographic information system (GIS) is a system to manage, analyze, and display of geographic information. This system is mainly used to create maps, compile geographic data, analyze mapped information, share and discover geographic information, and manage the information in a database" [ESRI, 2001]. A series of geographic datasets that model geography using simple, generic data structures represent geographic information and in order to manage this information a set of comprehensive tools is required that is fulfilled by ArcGIS [ESRI, 2001]. As stated in ESRI [2001], a geographic information system supports several views for working with geographic information:

1. The Geodatabase view: "A GIS is a spatial database containing datasets that represent geographic information in terms of a generic GIS data model (features, rasters, topologies, networks)" [ESRI, 2001].
2. The Geovisualization view: "A GIS is a set of intelligent maps and other views that show features and feature relationships on the earth's surface. Various map views of the underlying geographic information can be constructed and used as "windows into the database" to support queries, analysis, and editing of the information" [ESRI, 2001].
3. The Geoprocessing view: "A GIS is a set of information transformation tools that derive new geographic datasets from existing datasets. These geoprocessing functions take information from existing datasets, apply analytic functions, and write results, not new derived datasets" [ESRI, 2001].

"These three GIS views are represented in ArcGIS by the catalog (a GIS is a collection of geographic datasets), the map (A GIS is an intelligent map view), and the toolbox (a GIS is a set of geoprocessing tools)" [ESRI, 2001]. These three viewpoints are key components of the GIS and are used in these GIS implementations at different levels. In this study, all the preliminary processing of the watersheds and gauging stations was carried out using ArcGIS software. The details about the geographic data processing works carried using ArcGIS is provided in the 'Data Processing' chapter in the thesis report.

4.2 The NetCDF Interface

“The Network Common Data Format (NetCDF) is a self-describing machine-independent data format that was developed by Unidata. Unidata is a member of the University Corporation for Atmospheric Research (UCAR) and funded by the National Science Foundation (NSF). The NetCDF data has been around since 1989. The Network Common Data Format, or NetCDF, is an interface to a library of data access functions for storing and retrieving data in the form of arrays. It is a machine-independent data format that store big datasets in a compact format” [Uddameri 2017]. “The NetCDF is a self-describing, which means the file contains a header which includes information on the variables used and the structure of the data. Generally, the metadata is stored in a separate file. The data is also platform-independent, which means data can be ported between different operating systems without any problems” [Uddameri 2017]. The NetCDF file format is built on the traditional scientific data format and is gradually being made consistent with other data formats used for the processing of scientific data [Uddameri 2017]. The extensions ".nc and .nc4" are given to NetCDF files.

In Wikipedia, it has been stated that “NetCDF data is:

- *Self-Describing*. A NetCDF file includes information about the data it contains.
- *Portable*. A NetCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- *Scalable*. A small subset of a large dataset may be accessed efficiently.
- *Appendable*. Data may be appended to a properly structured NetCDF file without copying the dataset or redefining its structure.
- *Sharable*. One writer and multiple readers may simultaneously access the same NetCDF file.
- *Archivable*. Access to all earlier forms of NetCDF data will be supported by current and future versions of the software.”

4.2.1 The NetCDF Data Structure

The structure of the NetCDF file consists of 3 core components: variables, dimensions, and attributes. This structure was introduced with the very first NetCDF release and is still relevant today as they are considered the core of all NetCDF files.

Variables

Variables represent an array of the same type of values. The hydrometeorological phenomenon of interest such as temperature, precipitation, and runoff are represented by a variable. "A variable collected at a single location at various times is stored as a one-dimensional array. Similarly, a variable collected at several locations at a single point in time is stored as a two-dimensional array. Data collected at several locations at different times are stored in a three-dimensional array. A four-dimensional data storage is adopted when data are collected at multiple geographic locations and at various depths (heights) and times at each of these locations" [Uddameri 2017]. Six variables (char, byte, short, int, float, double), are included in the NetCDF files. The number of dimensions is used to determine a name, data type, and shape for variables. For the storage of bulk of data, variables are used. The coordinate variables are the kind of special variables that store data on dimension variables (typically - lat, lon, and time) [Uddameri 2017].

Dimensions

The axes of the data arrays are described by dimensions. "Dimension have names (usually lat, lon and time) and a length" [Uddameri 2017]. The infinite dimension has a length that can be extended to include more information at any time. At most one unlimited dimension is contained in NetCDF files. "The size is an integer indicating how many values are stored along each named dimension (e.g., data collected at 3 locations and 4 times would have dimensions- lat = 3, lon = 3, time= 4). Only one dimension can have UNLIMITED dimension. In gridded datasets, time is set to have an unlimited dimension to allow the addition of data as it becomes available" [Uddameri 2017].

Attributes

"Attributes are used to store ancillary data or metadata. Attributes associated with variables are identified using both variable and attribute name (e.g., discharge: units = m^3s^{-1}; stores the units for the discharge variable)" [Uddameri 2017]. A global attribute stores NetCDF file data. Attributes are always scalar values or 1D arrays to which a variable or a file can be connected. Although no defined limit is enforced, the user is expected to maintain minimal attributes.

4.2.2 Convention for file names and formats

File Names

The NetCDF files that were downloaded from the ESGF Portal through the ISIMIP website had a specific convention used for file names and formats, and here, this convention is explained in detail. As stated in ISIMIP2a [2018], one variable is reported per file, and the file names followed this convention for the historical validation runs:

<modelname>_<obs>_<bias-correction>_<clim_scenario>_<socio-econ-scenario>_<sens-scenario>_<variable>_<region>_<timestep>_<start-year>_<end-year>.nc

The parts in brackets were replaced with the appropriate specifier used as per the study (Table 4).

Table 4: File name specifiers for output data [ISIMIP2a, 2018]

Item	Possible Specifiers	Explanation
<modelname>	model name as registered in ISIMIP	Name of the impacts model
<obs>	gswp3, Princeton, watch, wfdei	Dataset providing the climate forcing data
<bias-correction>	nobc	Indicates that no bias correction was performed on the climate data
<clim_scenario>	hist	Historical climate information
<socio-econ-scenario>	nosoc	Naturalized runs (no human impact) assumed. No irrigation. No population and GDP data prescribed.
<sens-scenario>	co2	Transient CO2 concentration for CO2 fertilization effects
<variable>	variable names	Output variable od the impact model
<region>	global	For global simulations
<timestep>	daily, monthly	Daily (discharge) and Monthly (swe)
<start-year>_<end-year>	1971-2000 (historical time period)	1971-2000 (for discharge analysis) 1980-2000 (for swe analysis)

Two variables from the global (water) sector were extracted in the NetCDF format for this study. Table 5 and 6give the example of the NetCDF file name used for discharge and snow water equivalent (swe) output data analysis. Table 5 provides the list of all NetCDF files used in the analysis for discharge output data. The format for all 9 participating models is the same, so in Table 5, just replace the model name in the file name column to get the complete list of NetCDF files used in the study. Please note that JULES-W1 model does not have watch dataset simulation result. Therefore, in the case of discharge data, a total of 107 files were used for analysis. The combined total size of all 107 files is about 100 Gigabytes.

Table 5: NetCDF File Names for discharge output variable analysis

Model Name	NetCDF File Names (discharge output variable)
WaterGAP2	<modelname>_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1971_1980.nc
DBH	<modelname>_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1981_1990.nc
H08	<modelname>_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1991_2000.nc
MATSIRO	<modelname>_princeton_nobc_hist_nosoc_co2_dis_global_daily_1971_198
MPI-HM	0.nc
PCR-GLOBWB	<modelname>_princeton_nobc_hist_nosoc_co2_dis_global_daily_1981_199 0.nc
LPJmL	<modelname>_princeton_nobc_hist_nosoc_co2_dis_global_daily_1991_200
ORCHIDEE	0.nc
JULES-W1	<modelname>_watch_nobc_hist_nosoc_co2_dis_global_daily_1971_1980.nc
	<modelname>_watch_nobc_hist_nosoc_co2_dis_global_daily_1981_1990.nc
	<modelname>_watch_nobc_hist_nosoc_co2_dis_global_daily_1991_2000.nc
	<modelname>_wfdei_nobc_hist_nosoc_co2_dis_global_daily_1971_1980.nc
	<modelname>_wfdei_nobc_hist_nosoc_co2_dis_global_daily_1981_1990.nc
	<modelname>_wfdei_nobc_hist_nosoc_co2_dis_global_daily_1991_2000.nc

Table 6 provides the list of all NetCDF files used in the analysis for snow water equivalent (swe) output data. The format for all 6 participating models is the same, so in Table 6, just replace the model name in the file name column to get the complete list of NetCDF files used in the study. Therefore, in the case of swe data, a total of 24 files were used for analysis. The combined total size of all 24 files is about 60 Gigabytes.

Table 6: NetCDF File Names for swe output variable analysis

Model Name	NetCDF File Names (swe output variable)
WaterGAP2 DBH MATSIRO MPI-HM PCR-GLOBWB LPJmL	<modelname>_gswp3_nobc_hist_nosoc_co2_swe_global_monthly_1971_2010.nc <modelname>_princeton_nobc_hist_nosoc_co2_swe_global_monthly_1971_2012.nc <modelname>_watch_nobc_hist_nosoc_co2_swe_global_monthly_1971_2001.nc <modelname>_wfdei_nobc_hist_nosoc_co2_swe_global_monthly_1971_2010.nc

Format for Gridded Data

Gridded data is returned in NetCDF4 format. Global data for the above discussed NetCDF files is available for the ranges -89.75 to 89.75 degrees latitude, and -179.75 to 179.75 degrees longitude, i.e., 360 rows and 720 columns, or 259200 grid cells total. The output data is reported row-wise starting at 89.75 and -179.75 and ending at -89.75 and 179.75. The standard horizontal resolution is 0.5x0.5 degrees, corresponding to the resolution of the climate input data [ISIMIP2a, 2018]. Table 7 provides the naming and format conventions for NetCDF files.

Table 7: Naming and format conventions for NetCDF files [ISIMIP2a, 2018]

Dimension	Name	Unit
X	Lon	degrees east
Y	Lat	Degrees north
T	Time	<time steps> since 1901-01-01 00:00:00
Missing value	1.e+20f	

4.3 The R Environment

R is a programming language and software environment that focuses primarily on carrying out statistical analysis, representing various graphics and reporting and is mainly used for statistical inference, data analysis, and algorithms in machine learning. According to Wikipedia, "R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team." "R is a software language for carrying out a wide variety of complicated (and simple) statistical analyses techniques such as time-series analysis, geospatial analysis, classical statistical tests, linear and nonlinear modelling, classification, clustering, summarizing and exploring data, along with highly extensible graphical techniques. The main strength of R is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed" [R Project (n.d.)]. R can be built and run on many UNIX platforms and similar systems like Windows, MacOS, Linux, etc. as a free software.

The following are some of the essential features of R:

- R is a well-developed, simple and effective language which includes conditionals, loops, user-defined recursive functions and input, and output facilities,
- R has an effective data handling and storage facility,
- R provides a suite of operators for calculations on arrays and matrices in particular,
- R provides an extensive, coherent and integrated collection of tools for data analysis, and
- R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

R is designed as an actual computer language, where it allows its users to add additional functionality by defining new functions [R Project (n.d.)]. Most programs were themselves written in R, making it easy for users to obey their algorithmic choices. C, C++ and Fortran code can be linked and called on time for computationally intense activities [Wikipedia]. In this study, R has been extensively used in time series and geospatial data analysis and visualization. The short overview of these two analyses is presented in the sections below.

4.3.1 R Packages

"In R, the fundamental unit of shareable and reproducible R code is the package. A package bundles together code, functions, data, documentation, and tests, and is easy to share with others" [Wickham 2015]. As of July 2019, there are 14,579 packages available on the **C**omprehensive **R A**rchive **N**etwork, or CRAN, the public clearing house for R packages. A large range of packages is one of the reasons R is so successful: odds are someone has solved a problem you are working on, and by installing their package you will profit from their research. "The capabilities of R are extended through user-created packages, which allow specialised statistical techniques, graphical devices, import/export capabilities, reporting tools, etc. These packages are developed primarily in R, and sometimes in Java, C, C++, and Fortran. The R packages are also used by researchers to organise research data, code and report files in a systematic way for sharing and public archiving" [Wikipedia].

Within packages, both R functions and datasets are stored. It is only available when a package is enabled. This is achieved both for efficiency (which would require more space and take longer than a subset to search) and for protecting software creators, who are shielded from name collisions. Packages also rely on one another, and if loaded others may be loaded automatically. The standard packages (or base) are included in the R source code. Several authors have made thousands of contributed packages for R. In some cases, specific statistical methods are introduced, in others information or hardware is available, and others are designed to complement textbooks. Some packages (recommended packages) have all R binaries distributed. Most are available for download from CRAN (https://CRAN.R-project.org/ and its mirrors).

4.3.2 R Statistics and Graphics

The statistical R environment has become the de-facto norm for statistician code writing and is the most commonly employed statistical application in the world in a number of areas. According to Wikipedia, "R and its libraries implement a wide variety of statistical and graphical techniques, including linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, and others." R statistical and graphical environments, whether from the data analyst or the computer scientist, are very well designed [Fox and Anderson, 2005]. In the introduction to the R environment

statistics were not listed, but many people are using R as a statistical tool. R is now an environment that has implemented many classical and modern statistical techniques. Some of these are incorporated into the R core, but many are supplied as packages. A number of "recommended" packages and "contributed" packages were applied to the standard R system with many reflecting state of the art in different fields of statistical computing. “The basic R system augmented by the contributed packages is arguably the most extensive resource for statistical computing currently available” [Fox and Anderson, 2005].

Statistical software such as SAS and SPSS are primarily geared to combining instructions that can be entered via point/click interface to produce (often voluminous) output from a rectangular case-by-variable datasets. This software facilitates routine tasks in the analysis of data but makes doing innovative or non-standard things relatively difficult or extending the built-in capabilities. A good statistical environment, by contrast, facilitates routine data analysis and supports easy programming. R serves both needs, and consumers can compose programs that add to their already amazing facilities quickly. Within R, a statistical analysis is usually performed as several steps, and intermediate effects are preserved within objects. Although SAS and SPSS provide a copious output from a regression and classification test, R provides limited output and stores outcomes for further R functions in an artifact appropriate for eventual questioning.

R is also particularly powerful tool in the area of statistical graphics. The user uses a series of graphic functions to create a graphic output, which either develop a new plot or adds value to an existing plotR graphics follow an "painting model," meaning that the graphics output is produced in steps and later output is obscured by the overlap of any previous output. This makes quality graphs for printing, including mathematical symbols. Additional packages provide dynamic and interactive graphics.

4.4 Hydrological Indicators

The hydrological indicators have provided the means to manage large time series datasets into small series that are easy to use, manipulate, and visualize. These indicators were used to analyze the time series analysis and visualization of different variables, linear trend analysis of observed and simulated discharges and extreme flows analysis, including high-flows and low-flows conditions. The time series data were molded into different temporal resolutions in this study to make the analysis works easier. The annual mean discharge was regarded as a hydrological indicator for linear trend analysis of observed and simulated data. Similarly, other hydrological indicators such as mean monthly hydrographs (monthly dynamics) and long-term average monthly discharge (seasonal dynamics) were used for discharge and snow water equivalent time series analysis and visualization, and for validating and evaluating the models. The different percentile flows obtained from daily discharges was used as an indicator in this study to analyze the extreme flows, including high and low flows. The term 'hydrological indicators' to define all the above-mentioned scenarios is adopted by Gosling et al. [2018] and Zaherpour et al. [2018]. These hydrological indicators are briefly described below, but the details about them are provided in the upcoming chapters of this thesis.

Annual Mean Discharge for Linear Trend Analysis

To analyze the linear trend fit of the observed and simulated data, the annual mean discharge was used as an indicator in the study. The annual mean discharge was obtained from the monthly time series data. The simulated data obtained through ISIMIP2a simulation runs contain daily time series data, so the daily time series was converted first into monthly time series data. The observed data from GSIM metadata archive was already in the monthly time series, so no further manipulation was carried out. Then, the monthly times series were aggregated and manipulated to get annual mean discharges. The annual mean discharge was obtained by taking the mean value from 12 monthly time series data present in a year. This process was repeated for the entire analysis period considered for each gauging station. In the end, there was only one discharge value representing a year, and so, the total annual mean discharge time series data available for the trend analysis was 30.

Mean Monthly Discharge (Monthly Dynamics)

The daily time series data obtained from the model output simulation was aggregated to get the mean monthly time series data in this case. The mean monthly discharge time series was obtained by calculating the mean value from the available 30 or 31 days in a given month. Thus, 365 daily discharge time series available in a year got reduced to 12 monthly mean discharge series. This process continued for the entire 30 years of analysis period in most of the gauging stations. Therefore, the total amount of monthly dynamics data used in the study was 360. So, around 10,950 daily discharges accumulated from 30 years got reduced to 360 mean monthly discharges, which were regarded as monthly hydrographs or monthly dynamics in this study. Since the observed discharges already contained the monthly time series, so no further manipulation was necessary in this case.

Long-term Average Monthly Discharge (Seasonal Dynamics)

The simulated daily discharge was aggregated into monthly time series first. The process of obtaining a monthly time series is the same as mentioned above in the case of the monthly dynamics. The monthly discharge time series of 30 years is then used to get the long-term average monthly discharge series, also known as seasonal dynamics in the study. Here, the discharge values of a month are accumulated together from 30 years of data. Then, these accumulated 30 discharges of a month are used to calculate the mean discharge value for that month. The obtained discharge is, thus, called long-term average monthly discharge as it takes into account the monthly discharge data of 30 years. This process of getting the long-term average monthly value is carried out for all 12 months. Finally, 12 discharge values for each month is used for time series analysis and model performance evaluation.

Percentile Flows for Extreme Flows (High and Low Flows) Analysis

Using the daily discharge time series for both observed and simulated data, we calculate 10 different percentile flows ranging between $Q_{0.01}$ and $Q_{99.99}$ to represent high and low flow conditions. These percentile flows are used to analyze these extreme conditions and evaluate the models accordingly in the study. The percentile flow determines the magnitude of discharge that is exceeded given a percentage of the time in the time series. For example, Q10 indicates the magnitude of discharge that is exceeded 10% of the time in the time series.

4.5 What is a Time Series?

A time series is a continuous data set points, usually measured over successive times. "It is mathematically defined as a set of vectors $x(t)$, $t = 0,1,2,...$ where t represents the time elapsed. The variable $x(t)$ is treated as a random variable" [Hipel and McLeod 1994; Adhikari and Agrawal 2013]. The measurements taken in a time series during an occurrence were set in chronological order. "A time series containing records of a single variable is termed as univariate and records of more than one variable is termed as multivariate" [Adhikari and Agrawal 2013]. A time series can be continuous or discrete. Observations are measured in continuous time series at every instance of time, while discrete time series include observations measured at discrete points of time. For example, streamflow, temperature readings, etc. can be considered as a continuous time series, whereas, the population of a city, production of a company can be considered as a discrete time series. Generally, successive observations are reported at equally spaced time intervals in distinct time series, including separations between hours, days, weeks, months or annual cycles. Through merging data over a specified interval, a continuous time series can be easily converted into a discrete one [Adhikari and Agrawal 2013].

A time series that typically increases, decreases, or stagnates over a long period of time is considered a trend. Therefore, the trend in a time series can be said to be a long-term change. For example, the number of houses in a city, the population growth graph, etc. indicate an upward trend, while a downward trend can be seen in a series relating to mortality rates, epidemics, etc. In a time series, seasonal variations are differences within a year during the season. Climate and weather patterns, traditions, cultural practices, etc. are important factors affecting seasonal variations. In summer, for example, there are higher sales of ice cream and in winter, sales of woollen clothes increase. Seasonal variation is a key to predicting the correct projection for the future.

There can be different kinds of time series based on the complexity of the analysis and functional needs. To visualize the data's basic pattern, a time series is normally shown by a graph which displays the observations with the appropriate time. Below in Figure 10, a time series plot of monthly observed discharge at one of the gauging stations taken into consideration in this study is presented as an example.

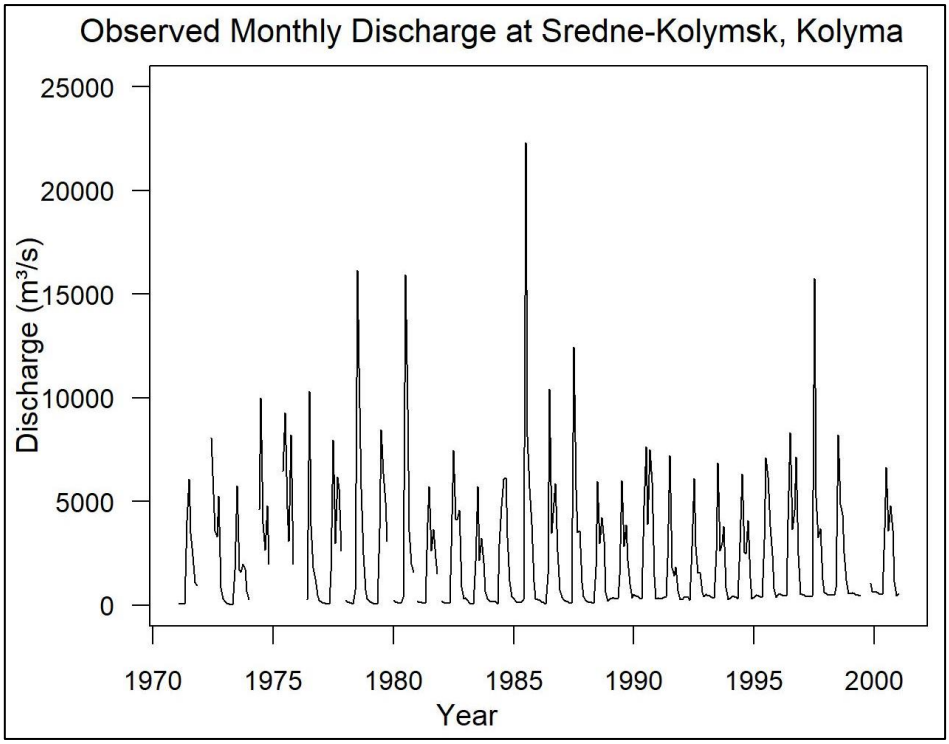


Figure 10: Observed monthly discharge series at Sredne-Kolymsk (Jan.1971- Dec.2000)

4.5.1 Time Series Analysis and Visualization

"In practice, a suitable model is simulated and fitted to a given time series obtained from the observations, and the corresponding parameters are estimated using the known data values. The procedure of fitting a time series to a proper model is termed as Time Series Analysis" [Hipel and McLeod 1994; Adhikari and Agrawal 2013]. It includes methods that try to understand the nature of the series and is often useful for future predictions and for the achievement of appropriate simulation results. "In time series forecasting, past observations are collected and analysed to develop a suitable mathematical model which captures the underlying data generating a process for the series" [Adhikari and Agrawal 2013]. Then the model is used to predict future events. Forecasts in time series have important applications in different fields and important decisions and precautionary measures to tackle the issues are made depending on the forecasted results. Thus it is very important to make a good forecast, that is, to fit an acceptable model into a time series. The time series forecasting is out of the scope for the current study, so it will not be discussed in detail here; however, this analysis method would be relevant for the future works.

For this study, time series analysis will be carried out to compare and validate simulated output results from the participating models with the observations data. The time series analysis for discharge and snow water equivalent (snow cover) variables would be carried out in detail in this study. The trend and seasonality analysis of the time series data will also be discussed in detail. A brief highlight of the procedures taken to carry out time series analysis for the related variables is presented below.

4.6 Model Efficiency Criteria

The analysis of the behavior and performance of the hydrological model is directly linked to modelled and observed data. Frequently, comparisons between simulated and measured streamflow are frequently carried out at the catchment outlet [Krause et al. 2005; Waseem et al. 2017]. However, in this study, in addition to choosing the outlet gauging station, other upstream gauging stations were also considered. The other upstream gauging stations were chosen to provide further support to the outlet station, analyze variables in a more evenly distributed manner in the basins, check the streamflow at those stations and check whether the model can capture and simulate the routing process in a river basin accurately. In this method, hydrologists utilize performance parameters generally to objectively assess the "closeness" of the modelled actions in the observed measurements [Krause et al. 2005; Waseem et al. 2017]. Therefore, in this study, model performance was evaluated based on the most common statistical model comparison tools, which are also referred to as model efficiency criteria.

According to Krause et al. [2005], “the model efficiency criteria is not only used to find the closeness of fit between simulated and observed data but also to improve the model for the future simulation.” Based on the performance, the model is assessed either subjectively or objectively. The subjective assessment includes visual inspection of the closeness of fit between the observed and simulated data where the model is evaluated based on systematic (underestimation/overestimation) and dynamic (periodic pattern) behavior. “The objective assessment involves mathematical analysis of the closeness of fit between the two datasets, and it is known as the efficiency criteria” [Krause et al. 2005]. It means that the smaller the deviation between observed and simulated data the greater the model's performance and the more effective it is to forecast past and future trends. “Most of the efficiency criteria are simply summations of the individual errors at

each time step of simulated data, which is then normalized by a measure of variability in the observations" [Krause et al. 2005]. "Most often the efficiency of a model is based solely on how well the predicted values fit the observed values, the assumption being that the observed data is error-free while this is not necessarily always the case" [Moriasi et al. 2007; Krause et al. 2005]. According to Legates & McCabe [1999] and Moriasi et al. [2007], "a good efficiency criterion should have at least three important components, i.e. one dimensionless statistic, one absolute error-index statistic, and one graphical technique."

The model optimization and the comparison of the accuracy of different models uses several statistical model performance evaluation criteria. These efficiency criteria help to validate the model's output with the observations. In this study, the Nash-Sutcliffe Efficiency (NSE) criterion, Percent Bias (PBIAS), Bias in Standard Deviation (Bias in SD) were selected. Such criteria are often used and reported in the literature in hydrological modelling research, but many other performance criteria can be used. Even the most accomplished water scientist will question the choice and use of particular performance metrics and analysis of the findings because each parameter can put a distinct focus on the various types of simulated or observable behaviour [Krause et al. 2005].

4.6.1 Nash-Sutcliffe Efficiency (NSE)

Nash and Sutcliffe in 1970 came up with an efficiency index to evaluate the hydrological simulations. This index aims to give a benchmark for a simulation in an objective manner. "The Nash-Sutcliffe Efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information")" [Nash and Sutcliffe, 1970; Moriasi et al. 2007]. NSE demonstrates how well the graph between the observed and the simulated data relates to the 1:1 line. NSE is computed, as shown in the equation below:

$$\text{NSE}=1 - \left[\frac{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^{n}(Y_i^{obs} - Y^{mean})^2}\right]$$

where Y_i^{obs} is the i^{th} observation for the constituent being evaluated, Y_i^{sim} is the i^{th} simulated value for the constituent being evaluated, Y^{mean} is the mean of observed data for the constituent being evaluated, and n is the total number of observations.

NSE ranges between –∞ and 1.0 (1 inclusive), with NSE = 1 being the perfect fit. Values between 0.5 and 1 are generally viewed as acceptable levels of performance, whereas values lower than 0 indicates that the mean value of observed times series would have been a better predictor than the simulated value from the model, which indicates unacceptable performance. "The normalization of the variance of the observation series results in relatively higher values of NSE in catchments with higher dynamics and lower values of NSE in catchments with lower dynamics. To obtain comparable values of NSE in a catchment with lower dynamics, the prediction must be better than in a basin with high dynamics" [Krause et al. 2005].

"The largest disadvantage of the Nash-Sutcliffe efficiency is the fact that the differences between the observed and predicted values are calculated as squared values. As a result, larger values in a time series are strongly overestimated, whereas lower values are neglected" [Legates and McCabe, 1999]. It leads to an overestimation of model performance in high flows and an underestimation in low flow conditions for the quantification of runoff predictions. "The Nash-Sutcliffe is not very sensitive to systematic model over- or underprediction, especially during low flow periods" [Krause et al. 2005].

NSE is recommended to be used for two major reasons: (1) it is recommended for use by many scientists working in hydrological modelling studies sector such as Sevat and Dezetter [1991]; ASCE [1993]; Legates & McCabe [1999]; Krause et al. [2005]; Moriasi et al. [2007]; Waseem et al. [2017] and (2) this dimensionless criterion is highly valuable since it is more popular to provide extensive information about the recorded values irrespective of the defects found in hydrological modelling [Gupta et al. 2009; Moriasi et al. 2007; Krause et al. 2005; Krysanova et al. 2018; Waseem et al. 2017]. The NSE is found to be the optimal objective function to represent a hydrograph's overall fit [Sevat and Dezetter, 1991]. Therefore, in this study, NSE has been used extensively to evaluate and validate the performance of the models in case of both monthly hydrographs and seasonal (long-term average monthly) dynamics of discharge.

4.6.2 Percent Bias (PBIAS)

"Percent bias (PBIAS) measures the tendency of the simulated data to be larger or smaller than their observed counterparts" [Gupta et al. 1999; Moriasi et al., 2007]. Bias lets one calculate the degree to which the observed value is far from a reference value (simulated value from the models). The optimal value of PBIAS is 0, with low magnitude values indicating accurate model simulation. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias [Gupta et al. 1999; Moriasi et al., 2007]. PBIAS is calculated with the equation provided below:

$$\text{PBIAS}=\left[\frac{\sum_{i=1}^{n}(Y_i^{obs}-Y_i^{sim})}{\sum_{i=1}^{n}(Y_i^{obs})}\right]*100$$

where PBIAS is the deviation of data being evaluated, expressed as a percentage, Y_i^{obs} is the i^{th} observation for the constituent being evaluated, Y_i^{sim} is the i^{th} simulated value for the constituent being evaluated.

PBIAS was selected for recommendation for several reasons: (1) it was recommended by ASCE [1993] and as stated in Huang et al. [2016], Moriasi et al. [2007] also recommended PBIAS to evaluate simulation of the monthly hydrograph, (2) PBIAS is used to measure errors in the water balance and can quickly be improved to load errors, and (3) PBIAS has the potential to clearly show poor performance of the model [Gupta et al., 1999]. This error-index criterion has been used by many hydrological scientists in their research study to evaluate different hydrological models and most of them have suggested that it be an efficient way to check the efficiency of the models. PBIAS is used just for evaluating monthly hydrographs in this study.

4.6.3 Bias in Standard Deviation (Δσ)

The bias in standard deviation (Δσ) of streamflow is calculated similarly as PBIAS. The bias in SD is used to calculate the measure of statistical dispersion between simulated and observed data for a period of analysis. In other words, σ denotes the standard deviation of the mean annual cycle for the years 1971-2000 for the observed and simulated values, and Δσ helps to measure the relative distance of the observed values from the simulated values. Bias in standard deviation is calculated with the equation provided below:

$$\Delta\sigma = \left[\frac{\sigma_{sim} - \sigma_{obs}}{\sigma_{obs}}\right] * 100$$

where $\Delta\sigma$ is the bias in the standard deviation of data being evaluated, expressed as a percentage, σ_{sim} is the simulated standard deviation for the constituent being evaluated and σ_{obs} is the observed standard deviation for the constituent being evaluated.

The standard deviation tests the precision of the experiment, namely the closeness of the individual measurements to each other, but it does not calculate the bias, which involves comparing the results with one of the observed values. The standard deviation measures how a test can be repeated at certain concentrations. Test repeatability can be consistent with low standard deviation, and low imprecision or inconsistent with high standard deviation, and high imprecision. It is optimal that the same specimen is repeatedly measured to get results as close as possible to each other. Test precision is important as it can lead to the loss of test reliability due to the lack of precision.

Bias, on the other hand, is the misjudgement in estimates due to systematic errors that lead to consistently high or low results when compared with actual values. The difference may also be due to random errors or any other inaccuracies if the estimate is not known to be biased. These inaccuracies can be either positive or negative, as opposed to bias that always acts in one direction. Find the errors by subtracting every calculation from the actual or observed value to measure the bias. Add up all the errors and divide by the number of estimates to get the bias. When the errors add up to zero, the estimates are unbiased, and the method delivers unbiased results. When the estimates are biased, the source of the bias can possibly be found and removed in order to improve the method. Bias in standard deviation is used just for evaluating seasonal dynamics in this study.

5. DATA PROCESSING

5.1 Historical Analysis Period

The historical analysis period considered for simulated discharge time series was 30 years (1971-2000). However, the observed discharge in some of the gauging stations did not have the complete time series of 30 years available so, the analysis period for some of these stations had to be adjusted as shown in Table 8. The information about the missing data of the observed discharge series from each station is also tabulated below. In case of snow water equivalent (swe), both observed and simulated data used the analysis period of 21 years (1980-2000) as they both had data available from 1979 to 2014. The observed swe series had some missing data of the same months in all basins. The number of missing months from observed snow water equivalent data was 48.

Table 8: Historical analysis period considered for observed discharge at gauging stations

River Basins	Gauging Stations	Analysis Period		Total Months	Data Missing (Months)
		Used	Available		
Kolyma	Kolymskaya	1978-1998	1978-2008	252	0
	Sredne-Kolymsk	1971-2000	1927-2002	360	31
Lena	Kusur	1971-2000	1935-2011	360	1
	Verkhoyanski Perevoz	1971-2000	1945-2002	360	1
	Tabaga	1971-1992	1950-1992	264	0
	Hatyrik-Homo	1971-1992	1967-1992	264	17
Mackenzie	Peace Point Alberta	1971-2000	1959-2014	360	5
	Arctic Red River	1972-2000	1972-2014	348	9
	Fort Simpson	1971-2000	1938-2014	360	0
Ob	Salekhard	1971-2000	1954-2010	360	12
	Hanti-Mansisk	1977-1997	1977-1997	252	29
	Kolpashevo	1971-1994	1936-1994	288	0
Yenisei	Igarka	1971-2000	1955-2011	360	47
	Bol Porog	1971-1990	1940-1990	240	1
	Pod Tunguska	1971-1993	1970-1993	276	0
Yukon	Eagle AK	1971-2000	1950-2016	360	0
	Pilot Point AK	1975-1996	1975-2016	264	12
	Nenana AK	1971-2000	1962-2016	360	0

5.2 ArcGIS Processing

ArcGIS was used to create maps of the study area, get geographic information such as location (longitude and latitude) of the gauging stations under preview (Table 9), create individual maps of the river basins considered, use information from the river networks to choose the relevant gauging stations and many other useful GIS applications. The latitude and longitude information using GIS was helpful to extract grid cells values of the individual gauging stations or any other points of interest from the NetCDF files. ArcGIS was a useful tool to process relevant information required in this study. All GIS data are in the WGS 1984 projection.

Table 9: Location and Grid cell values (Latitude and Longitude) of the gauging stations

Watersheds	Gauging Stations	Lat/Lon for 8 Models	Lat/Lon for ORCHIDEE Model
Kolyma	Kolymskaya	68.75/ 158.75 **43/678**	68.75/158.75 **43/678**
	Sredne-Kolymsk	67.25/ 153.25 **46/667**	67.75/154.25 **45/669**
Lena	Kusur	70.25/ 126.75 **40/ 614**	70.75/127.25 **39/615**
	Verkhoyanski Perevoz	63.25/ 132.25 **54/625**	63.25/132.25 **54/625**
	Tabaga	61.75/ 129.75 **57/620**	61.75/129.25 **57/619**
	Hatyrik-Homo	63.75/ 125.75 **53/612**	63.75/124.75 **53/610**
Yenisei	Igarka	67.25/ 86.75 **46/534**	67.75/86.25 **45/533**
	Bol. Porog	65.25/ 89.75 **50/540**	65.75/89.75 **49/540**
	Pod. Tunguska	61.75/ 90.25 **57/541**	61.75/89.75 **57/540**
Ob	Salekhard	66.75/ 66.75 **47/494**	66.75/66.75 **47/494**
	Hanti-Mansisk	60.75/ 69.25 **59/499**	61.25/69.25 **58/499**
	Kolpashevo	57.75/ 83.25 **65/527**	58.25/82.25 **64/525**
MacKenzie	Peace Point Alberta	59.25/ -112.75 **62/135**	59.25/-112.75 **62/135**
	Arctic Red River	67.25/ -133.75 **46/93**	67.25/-133.75 **46/93**
	Fort Simpson	61.75/ -121.75 **57/117**	61.75/-121.75 **57/117**

Yukon	Eagle AK	64.75/ -140.75 **51/79**	64.75/-140.75 **51/79**
	Pilot Point AK	61.75/ -162.75 **57/35**	62.25/-163.25 **56/34**
	Nenana AK	64.75/ -148.75 **51/63**	64.75/-149.25 **51/62**

5.3 Model Output Data Acquisition

For this study, the output data from the models running historical simulation runs through the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) (https://www.isimip.org/) using ISIMIP2a simulation run (https://www.isimip.org/about/#simulation-rounds) had to be used. The details about the project and the simulation runs are already covered in the previous sections. So, to access the output data required for the study, the ISIMIP website was viewed and a link to the 'Output Data Table' under the 'Outcomes' heading (https://www.isimip.org/outputdata/) was found. There, the information for external users was provided. The output data was available publicly and was downloaded via the ISIMIP ESGF server using the ESGF Portal. So, the ways to access the data via the ESGF Portal and some useful instructions were found on the website (https://www.isimip.org/outputdata/isimip-data-on-the-esgf-server/). In order to download the data, an account to get an ESGF OpenID had to be created and request to enrol in the ISIMIP Research group within the ESGF had to be sent. After completing these steps, I was finally able to login into the server and got access to the ISIMIP data archive (https://esg.pik-potsdam.de/search/isimip/) via the ESGF Portal (https://esg.pik-potsdam.de/projects/isimip/). Inside the data archive, the type of output data required according to the needs (required simulation runs, sectors, impact models, output variables, etc.) had to be chosen by the user and once the selection for the required dataset of interest was done, it went into Data Cart section. From the data cart, finally, the output data sets of choice were downloaded into the personal computer. All the downloaded output data files from the archive were in the NetCDF format. Keep in mind that one has to login first to the server to be able to download the required output data.

5.4 Read NetCDF files using R

R environment makes it convenient to manipulate NetCDF files since R offers various packages to deal with this data format efficiently. The R packages used for this study were ncdf4 and ncdf4.helpers. The ncdf4 package allows to read and write the NetCDF files. In this study, our primary focus will, however, be on reading NetCDF files. The ncdf4.helpers package provides some of the crucial functions to help manipulate the NetCDF files conveniently. A brief description of these two packages and some of their essential functions used in the study are provided below. The R script to read and manipulate the NetCDF files is provided in the Appendix.

ncdf4 Package

The "ncdf4" package is designed to work with the NetCDF library, version 4. It includes the ability to use compression and chunking, which seem to be some of the most anticipated benefits of the version 4 library. This package provides a high-level R interface to data files of Unidata's NetCDF library, which are binary data files that are portable across platforms and include metadata information in addition to the data sets. NetCDF files can be opened, and data sets can be read easily using this package. It is also easy to create new NetCDF dimensions, variables, and files, in either version 3 or 4 format, and manipulate existing NetCDF files. Some important functions used in this study from ncdf4 package are 'nc_open,' 'ncvar_get' and 'nc_close.' The function 'nc_open' helps to open the NetCDF file and provide access to all the contents/data stored in the given file. 'ncvar_get' function helps to read values from a variable from the NetCDF file. Returned values will be in ordinary R double precision if the NetCDF variable type is float or double. Returned values will be in R's integer storage mode if the NetCDF variable type is short or int. Returned values will be of character type if the NetCDF variable is of character type. Data can be subset specifying the necessary cells in the 'ncvar_get' to specify the starting point of the extraction in the array using start (in our case, lon, lat and time) and count (in our case, the number of lon, lat and time values to be extracted). Finally, 'nc_close' function closes an open NetCDF file, which flushes any unwritten data to disk.

ncdf4.helpers Package

This package provides several helper functions for NetCDF files opened using the ncdf4 package. Dealing with NetCDF format data is unnecessarily difficult. The ncdf4 package

does a good job of making many lower-level operations easier. The ncdf4.helpers package aims to build higher-level functions upon the foundation of ncdf4. Especially 'nc.get.time.series' function was used from this package to receive time information. Retrieving time data from a NetCDF file in an intelligible format is a non-trivial problem. The PCICt package solves part of this problem by allowing for 365- and 360-day calendars in addition to the Gregorian calendar. This function complements it by returns time data for a file as PCICt, doing all necessary conversions.

5.5 R packages used in Time series Analysis and Visualization

Several R packages were used in this study to analyze the data efficiently. Different R packages were used while dealing with the NetCDF files, time series analysis, geospatial data analysis, hydrological model evaluation using several various efficiency criteria, extreme flows analysis, and many other tasks. The description of some of the important packages used to carry out the above-mentioned tasks is presented below in brief. The brief description of these R packages is taken from the Comprehensive R Archive Network (CRAN).

'PCICt' package

Since the NetCDF files did not contain the actual dates in their time attribute. These files used time steps to store the data, which made it challenging to analyze the data. For this reason, I had to use this package to provide a work-alike to R's POSIXct class, which implements 360- and 365-day calendars in addition to the Gregorian calendar. Class 'POSIXct' represents the (signed) number of seconds since the beginning of 1970 (in the UTC time zone) as a numeric vector.

'lubridate' package

Date-time data can be frustrating to work in R. R commands for date-times are generally unintuitive and change depending on the type of date-time object being used. Moreover, the methods we use with date-times must be robust to time zones, leap days, daylight savings times, and other time-related quirks, and R lacks these capabilities in some situations. Lubridate makes it easier to do the things R does with date-times and possible to do the things R does not. In other words, the lubridate package functions to work with

date-times and time-spans for fast and user-friendly parsing of date-time data, extraction and updating of components of a date-time (years, months, days, hours, minutes, and seconds), algebraic manipulation on date-time and time-span objects. The 'lubridate' package has a consistent and memorable syntax that makes working with dates easy and fun.

'hydroGOF' package

The 'hydroGOF' package mainly deals with the Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. It functions to implement both statistical and graphical goodness-of-fit measures between observed and simulated values, mainly oriented to be used during the calibration, validation, and application of hydrological models. Missing values in observed and/or simulated values can be removed before computations. The model efficiency criteria (quantitative statistics): Ratio of Standard Deviations (rSD), Nash-Sutcliffe efficiency (NSE), Percent Bias (PBIAS), used in the model performance evaluation in this study were calculated using this package.

'hydroTSM' package

The 'hydroTSM' package stands for Time Series Management, Analysis, and Interpolation for Hydrological Modelling. This package is mainly used for management, analysis, interpolation, and plotting of time series used in hydrology and related environmental sciences. This package is highly oriented to hydrological modelling tasks. The focus of this package has been put in providing a collection of tools useful for the daily work of hydrologists.

'hydrostats' package

The 'hydrostats' package stands for Hydrologic Indices for Daily Time Series Data. This package helps to calculate a suite of hydrologic indices (streamflow indices) for daily time series data that are widely used in hydrology such as: basic summary statistics, high spells (magnitude, frequency, duration, timing etc.), low spells, cease-to-flow spells, Colwell's indices, partial series, baseflow components, etc. In the study, this package was mainly used to analyze the extreme flow conditions such as high flows (high spells) and low flows (low spells).

5.6 Discharge Time Series Analysis and Visualization

Reading and Manipulating Time Series Data

First, to read, manipulate, analyze, and visualize the time series data, a suitable software environment that can fulfil all these requirements was required. Therefore, the R programming language environment was decided to be used for the research study. The R statistical environment was used as a means of carrying out time series analysis and visualization. R is well-suited for the analysis due to the availability of extensive time series libraries, statistical methods, and straightforward plotting capabilities. The R scripts to carry out the analysis and to visualize the outcome of the analysis were written. The R time series analysis and visualization scripts can be viewed in the Appendix.

Basically, for the analysis, the R codes were written to read the NetCDF files (the simulation output data format) as discussed previously and then, as an output, the time series data present in the NetCDF file were converted into comma-separated (.csv) file for easy access to look at the data and pre-process them accordingly. The .csv file could be opened easily in Microsoft Excel, which made the work convenient. Then, the list of missing data (missing months) from the observed time series data were compiled. There was no problem of having any issue with missing data for the simulated results as the NetCDF files had the complete set of time series present. Thus, after collating the information on the missing data from the observation data, necessary changes to exclude those missing data from the simulated results were made so that both data were consistent and coherent.

This pre-processing step was done for all 18 gauging stations and 9 participating models with 4 climate forcing datasets each under consideration. More time was spent to finish this process as there was a large amount of data to be analysed. It took about 2-3 minutes to extract the daily time series data for 30 years from the NetCDF files for each climate, forcing dataset of a model. This step had to be repeated for 630 times as the mass data contained 630 (18 gauging stations x 8 models x 4 climate forcing datasets + 18 gauging stations x 1 remaining model x 3 climate forcing datasets) different iterations to be dealt with. Note that JULES-W1 model had simulation results using only 3 climate forcing data. Fortunately, a lot of time was saved because of the R script that was written to carry out

the pre-processing works. Few adjustments had to made in the script to make it reusable for processing other iterations.

Aggregating Time Series Data into different Temporal Resolutions

Once the daily time series simulated data were extracted from all 9 models for 18 gauging stations using the grid cell information from the NetCDF file with the help from the ArcGIS processing (Table 9), the daily time series data were aggregated into mean monthly discharge in order to make it coherent with the observed discharge data available. The monthly observations data of the gauging stations were only available from the GSIM metadata archive, so the simulated time series data had to be aggregated into mean monthly data. The way of aggregating the daily data into mean monthly is provided through the R script available in the Appendix. About 10,950 daily discharge series were aggregated into 360 monthly discharge series. The mean monthly time series (monthly dynamics) obtained was then, used to validate and evaluate the models later. The monthly time series data of 30 years had to be manipulated to get long term average discharge (seasonal dynamics) for each month. It was easier to evaluate the dynamics and seasonality behaviour of the time series when seasonal dynamics was used. It made the analysis work easier and quicker as 360 monthly data was aggregated further to just 12 long term average monthly data for further analysis.

Plotting Time Series

After completing the steps of reading, manipulating and aggregating the time series data into the desired format, the R script (Appendix) was written to visualize the time series data for both simulated and observed data. The R scripts helped to produce the plots of discharge time series in different temporal resolutions such as daily, monthly, and seasonal. The mean monthly simulated discharge series plot after aggregating the daily time series data and the long-term average monthly (seasonal) discharge series plot obtained from manipulating the monthly time series data are presented below in the chapter "Discharge Data Analysis: Results and Discussion."

5.6.1 Model Validation Runs for Discharge

One of the challenging tasks for the data analyst is to validate the model. Validation is the process of comparing two results, the representation of a conceptual model (simulated data) to the real system (observed data), and determining the degree to which the conceptual model corresponds to the real system, or at least accurately represents the model specification document, is referred to as model validation. If the comparison is accurate, then it is valid, else invalid. Simulation models are imitation of real-world systems, and they never precisely imitate the real-world system. Due to that, a model should be validated to the degree needed for the model's intended purpose or application. There is no simple test to establish the validity of a model. Validation is an inductive process through which the modeler concludes the accuracy of the model based on the evidence available. Gathering evidence to determine model validity is primarily accomplished by examining the model structure (i.e., the algorithms and relationships) to see how closely it corresponds to the actual system definition. For models having complex control logic, graphic animation can be used effectively as a validation tool. Finally, the output results should be analysed to see if the results appear reasonable. If circumstances permit, the model may even be compared to the actual system to see how they correspond. If these procedures are performed without encountering a discrepancy between the real system and the model, the model is said to have face validity.

5.6.2 Comparison of Simulated and Observed Discharge

The participating Global Hydrological Models are not calibrated, except for the WaterGAP2. WaterGap2 model was calibrated against long-term average monthly discharge for 1312 gauging stations worldwide [Hattermann et al. 2017; Krysanova et al. 2018; Alcamo et al. 2003; Schmied et al. 2016]. The monthly mean discharge (monthly dynamics/monthly hydrographs) and long-term average monthly discharge (seasonal dynamics) were used to validate and evaluate the models because of the large scale of the river basins, complex hydrological and anthropogenic conditions of the region under study [Huang et al. 2017], and poor availability of observational data. The plots were produced by comparing simulated and observed discharges at selected 18 gauging stations from 6 river basins. As it was difficult to compare the monthly hydrographs visually, the monthly dynamics plots were not produced, but they were used in case of objective assessment. Figure 11 shows just an example of mean monthly discharge time

series plot for 1971-2000 simulated by WaterGAP2 using four reanalysis climate forcing datasets at the outlet station, Kusur of the Lena basin. It was difficult to inspect the plot visually and draw conclusions from it due to a large amount of time series values available in monthly hydrographs.

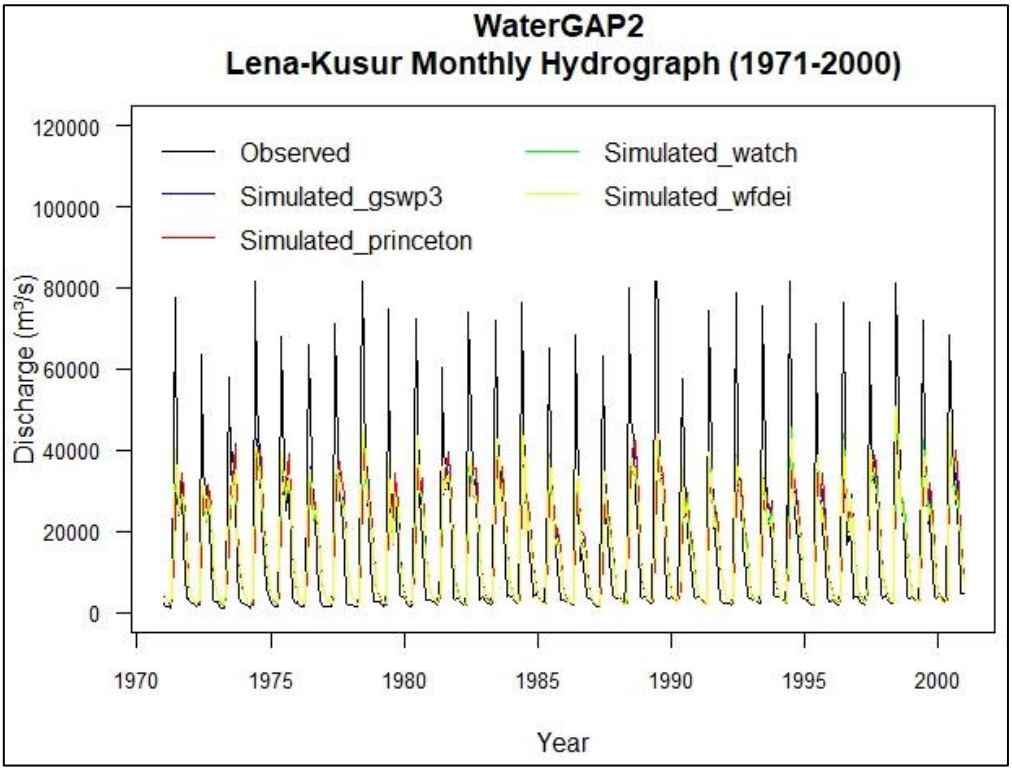


Figure 11: Comparison of observed and 4 different simulated monthly hydrographs as modelled by WaterGAP2 using 4 different climate forcing datasets at Kusur, Lena

Due to the difficulty in visually comparing and analyzing the dynamics and seasonality of the monthly hydrographs, the plots of long-term average monthly discharge (seasonal dynamics) were produced. The plots for seasonal dynamics were more comfortable to visualize and draw conclusions from as they had fewer time series data than the monthly dynamics. Figure 12 shows the seasonal dynamics plots simulated by WaterGAP2 using 4 climate forcing datasets at the outlet station of all basins: Lena (Kusur), Kolyma (Kolymskaya), Mackenzie (Arctic Red River), Ob (Salekhard), Yenisei (Igarka) and Yukon (Pilot Point AK). As seen in the figure below, the plots are readable and easy to analyze the dynamics and seasonality of the long-term average monthly discharge. Therefore, the seasonal dynamics plots were only produced for comparing simulated and observed data visually for this study. The seasonal dynamics values were also used in the objective assessment. The R scripts related to monthly hydrographs and seasonal dynamics analysis and visualization are provided in the Appendix. All the other seasonal dynamics

plots of the remaining 8 models are provided in the Appendix as the total number of plots is large in numbers, and all of them cannot be presented in the main report.

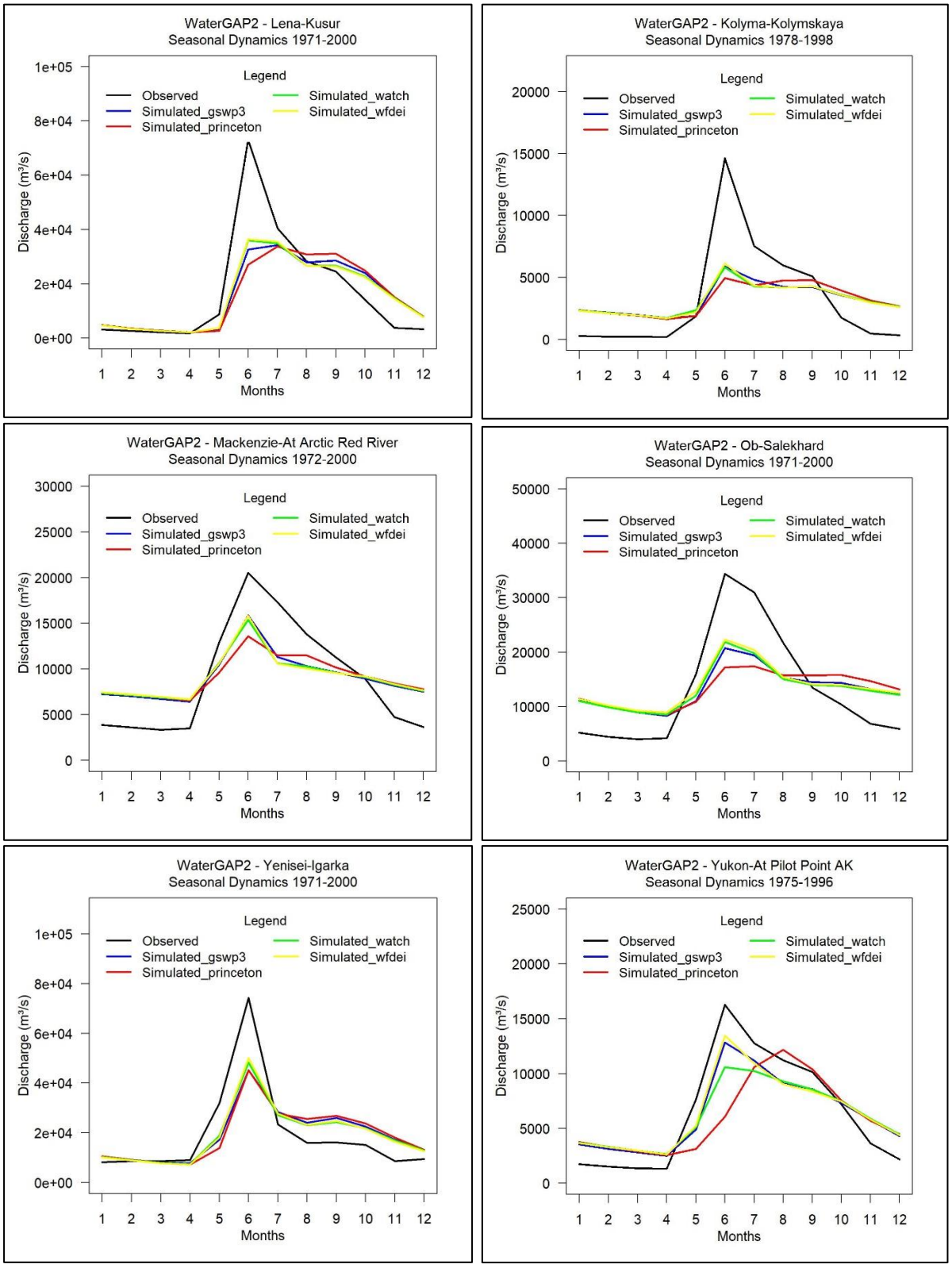
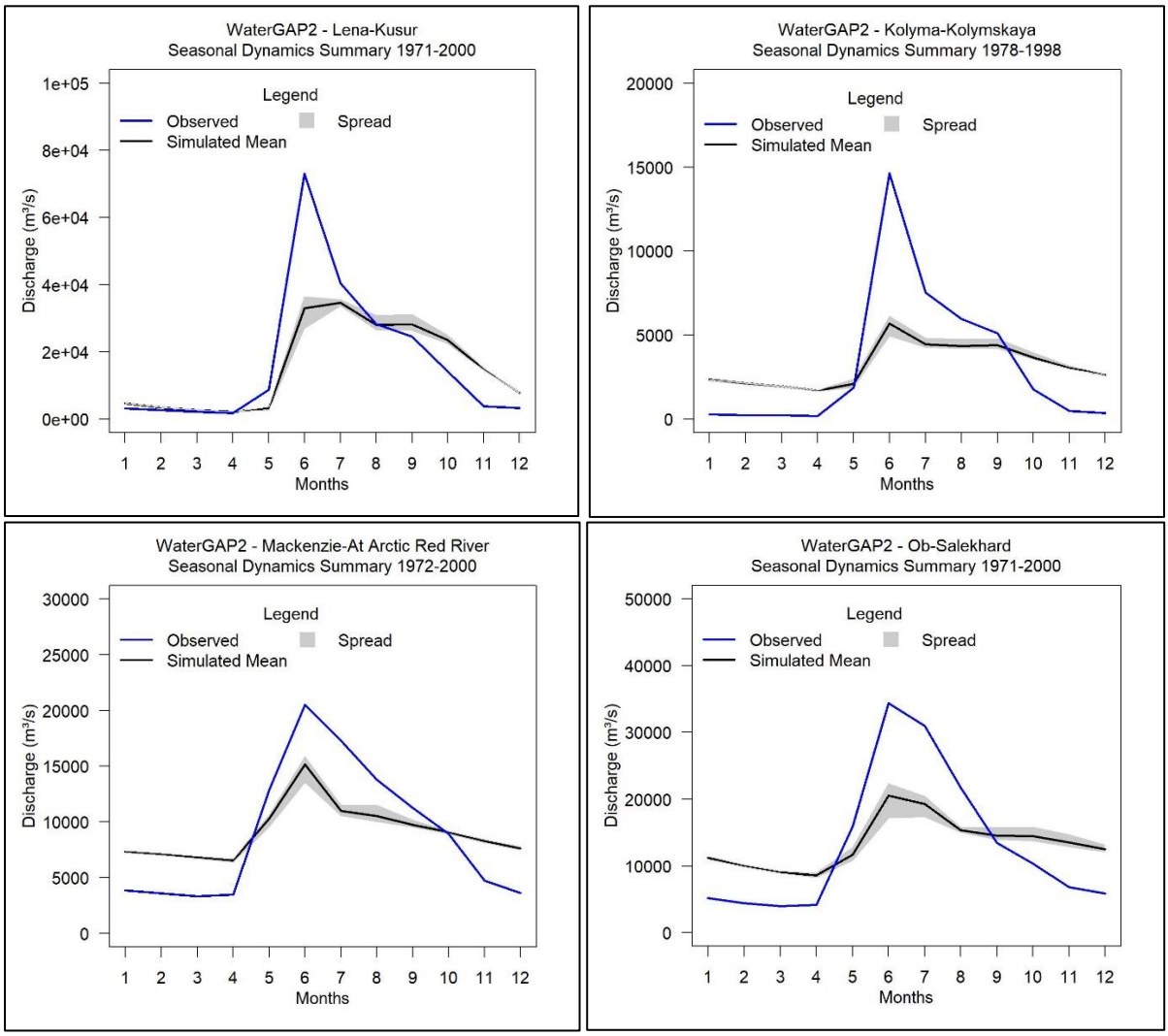


Figure 12: Comparison of observed and 4 different simulated long-term average monthly seasonal dynamics of discharges modelled by WaterGAP2 using 4 different climate forcing datasets at outlet gauging station of 6 river basins

The four colorful lines seen in the plots above indicate the simulated discharges from a model that was driven by four different climate forcing datasets and it could be confusing and at times difficult to assess when comparing against the observed discharge. Thus, in order to solve the problem, a mean simulated discharge was calculated from these four simulated results. Figure 13 highlights simulated mean discharge for each month along with the maximum and minimum discharges taken from the four simulated datasets. The maximum and minimum values are indicated as a spread (gray color) in the seasonal dynamics plots (Figure 13), which show the range of simulated values at a particular instance of time (each month). Similar plots, as seen in figure 13, were produced for the remaining models and gauging stations.



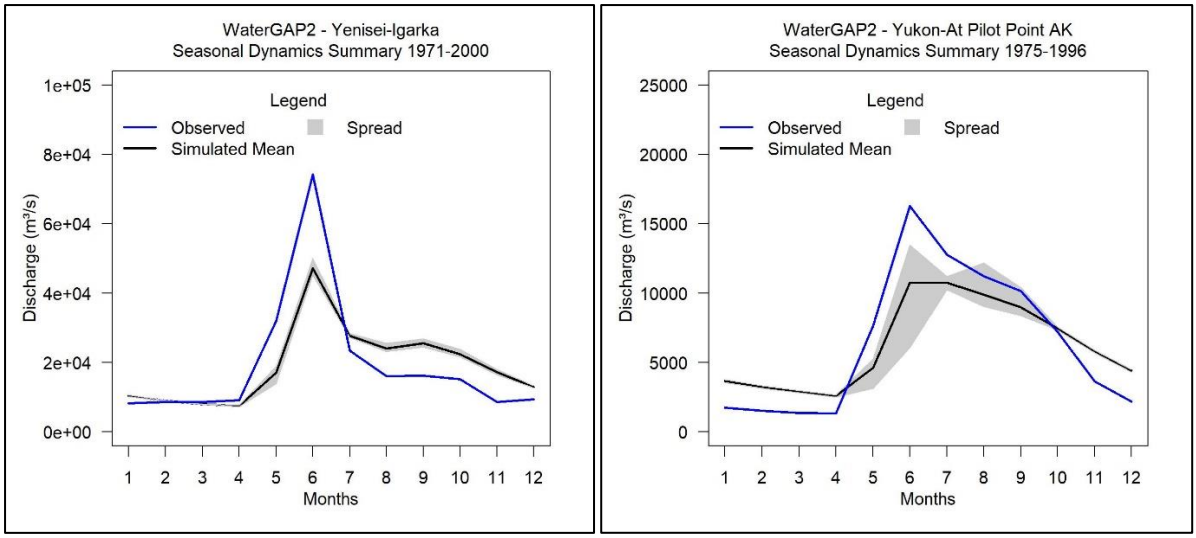


Figure 13: Comparison of observed and simulated mean with spread (maximum and minimum range) long-term average monthly seasonal dynamics of discharge modelled by WaterGAP2 at outlet gauging station of 6 river basins

5.7 Model Evaluation Methods

All models except WaterGAP2 used in this study were not calibrated. GHMs are usually applied with a coarse spatial resolution without calibration, and are often driven by biased climate input data, which could lead to significant biases in the simulation results [Hattermann et al. 2017]. Due to the higher uncertainties associated with the models in question, it was impossible to evaluate the models using the standard threshold values that were used in many research papers [Hattermann et al. 2017; Krysanova et al. 2018; Huang et al. 2017; Gudmundsson et al. 2012; Zaherpour et al. 2018; Moriasi et al. 2007]. Huang et al. [2017] chose the rating NSE ≥ 0.7, |PBIAS| ≤ 15 % and $|\Delta\sigma| \leq 15\%$ to denote a 'good' performance.

However, these threshold ratings did not work in this study as the models used coarser resolution for their simulations, and thus, the model output showed significant bias when compared with the observations. It seemed like almost all the models would perform poorly if the threshold ratings, as suggested by other researchers were used for the model evaluation in this study. Therefore, we (me and my institutional tutors) decided to adjust the threshold ratings of the three criteria. The threshold values for each criterion were released compared with the values used in the regional-scale studies.

The models were evaluated and rated on their performance using 3 efficiency criteria and their threshold values as listed below in Table 10, and then aggregated indices were estimated for every model and basin using rating scores of 1 (good performance), 0.5 (weak) and 0 (poor) for every criterion and gauging station.

Table 10: Thresholds values for Model Efficiency Analysis

NSE (Monthly)	NSE (Seasonal)	PBIAS and Bias in SD (Δσ)
≥ 0.5 = Good	≥ 0.7 = Good	(-25%, +25%) = Good
(0.3, 0.5) = Weak	(0.5, 0.7) = Weak	(-50%, -25%) or (+25%, +50%) = Weak
≤ 0.3 = Poor	≤ 0.5 = Poor	≤ -50% or ≥ +50% = Poor

The efficiency criteria values were calculated first for all models including the climate forcing datasets and using the threshold values from Table 10; they were rated with the scores ranging from 0 to 1 to indicate different levels of performance by the models. The rating score was then, applied separately in all climate forcing datasets for a given model at each gauging station. After scoring the individual criteria of each climate forcing datasets, they were added together.

The maximum score a model received for each criterion was 4. So, the total score from different efficiency criteria (Table 10) used in this study was 16 (4 from climate forcing datasets x 4 from efficiency criteria) in a given model. Therefore, to get the model performance index of each station, the total sum of rating score from these criteria was divided by the highest possible score that could be achieved, which was 16 in the given scenario. Then, the ratio was converted into a percentage to come up with the best and poor performing models for each gauging station.

Model Performance Aggregated Index

Once all the models were evaluated for all gauging stations using the rating scores based on threshold values adapted for efficiency criteria (Table 10), the final rating scores were aggregated together to get the total score for the basin. Then, the obtained aggregated score for a given model driven by 4 climate forcing datasets was divided by the highest possible score achievable by the basin and converted the ratio into a percentage to get a model performance aggregated index value. For example, to evaluate the model

performance for Lena basin, the rating score from all 4 gauging stations (Kusur, Verkhoyanski Perevoz, Tabaga and Hatyrik Homo) were added together and then that aggregated score was divided by the highest achievable score for Lena basin, which in this case was 64 (4 from climate forcing datasets x 4 from efficiency criteria x 4 gauging station). For other basins, the highest possible score depended on the number of gauging stations used to represent that basin and all other factors remained the same.

This method was adopted to make the model performance evaluation feasible. With the help of the performance aggregated index, the work of evaluating the participating models became quite convenient. It bolsters the confidence of the model used when dealing with climate change impact assessment in the future runs.

5.8 Linear Trend Analysis of Observed and Simulated Discharge Series

"Linear trend analysis of the time series data look for meaningful trends, and it is prevalent tasks carried out in the scientific disciplines. For classifying the usefulness of a trend, the statistical significance of a linear trend fit is often used to the time series "[Bryhn and Dimberg, 2011]. "The statistical significance of a trend depends on the number of data analysed. Even if the data are scattered far from the trend line, and their number is large, a trend may be statistically significant" [Bryhn and Dimberg, 2011].

A trend analysis is a statistical method used to determine the fundamental pattern of behaviour in a time series that otherwise would be partly or almost completely obscured by noise. Trend analysis could be described, in a broad sense, as a collection of methods to detect and forecast patterns and regularities in time series.

This study used a quality criterion for linear trend analysis called statistical significance. The observed and simulated monthly time series data were aggregated into annual mean values. After that, coefficients of linear line relationship such as y-intercept and slope, and p values were calculated using observed and simulated annual mean data. The statistical significance of the observed time series trend was checked for all 18 gauging stations first (Table 11).

Then, few gauging stations were chosen, which satisfied the conditions of having shown significant trends in their time series and which had the observations data for the analysis period of 30 years (1971-2000) (Tables 12 and 13). The stations with complete 30 years of observed data were chosen to make coherent and consistent comparison with the simulated data. The simulated data from ISIMIP2a runs contained complete time series values of 30 years without any missing data.

Here, the coefficients like y-intercept and slope were calculated to find out the nature of the linear trend lines. The slope helped determine if the trend of discharge was increasing or decreasing over the years in the gauging stations. If the slope is positive, then, the discharge trend is considered to be increasing, and if it is negative, then, the trend is decreasing. However, in this study, the focus was on finding the statistical significance of the time series trend, so, the p-value was used to fulfil the purpose.

"The p-value is used for describing the probability (from 0 to 1) in statistical significance tests in which a null hypothesis is rejected when the p-value is low. The 95% confidence level ($p \leq 0.05$) has traditionally been used for indicating statistical significance in such tests within a wide variety of academic research fields" [Bryhn and Dimberg, 2011].

Therefore, the same threshold limit of 5% (0.05) for p values were considered in this study for the trend to be regarded as statistically meaningful. The p-value must be very low to trust the calculated metric. The lower the p-value, stronger is the significance of the trend. Also, the p-value is not an indicator of the strength of the relationship, but just the statistical significance. The strength is actually measured by the coefficient of determination (r^2), which was not taken into consideration in this study.

Table 11: Linear trend analysis results of annual mean observed discharge time series for 18 gauging stations

Trend Analysis of Observed Discharge (Annual Mean)						
River Basins	Gauging Stations	Coefficients		p-value	Time Period	
		y-intercept	slope		Used	Available
Kolyma	Kolymskaya	3463.6	-0.0373	0.653	1978-1998	1978-2008
	Sredne-Kolymsk	3122.8	-0.13	0.0223	1971-2000	1927-2002
Lena	Kusur	16487.6	0.131	0.301	1971-2000	1935-2011
	Verkhoyanski Perevoz	5346.9	0.0212	0.679	1971-2000	1945-2002
	Tabaga	7248	-0.0162	0.897	1971-1992	1950-1992
	Hatyrik-Homo	1235.2	0.0802	0.0752	1971-1992	1967-1992
Mackenzie	At Peace Point Alberta	2156.5	-0.0024	0.923	1971-2000	1959-2014
	At Arctic Red River	8989.9	0.013	0.857	1972-2000	1972-2014
	At Fort Simpson	6545.64	0.0458	0.381	1971-2000	1938-2014
Ob	Salekhard	13721.276	-0.111	0.344	1971-2000	1954-2010
	Hanti-Mansisk	2046.5	0.118	0.084	1977-1997	1977-1997
	Kolpashevo	3786.2	-0.064	0.117	1971-1994	1936-1994
Yenisei	Igarka	25048.6	-0.688	0.0045	1971-2000	1955-2011
	Bol Porog	3576.1	-0.0078	0.909	1971-1990	1940-1990
	Pod Tunguska	10731.2	0.0031	0.966	1971-1993	1970-1993
Yukon	At Eagle AK	2525.1	-0.0172	0.347	1971-2000	1950-2016
	At Pilot Point AK	5684.03	0.123	0.0197	1975-1996	1975-2016
	At Nenana AK	665.2	0.0025	0.570	1971-2000	1962-2016

The linear trend analysis for the observed time series was carried out in the R environment. From Table 11, it was deduced that at least 4 out of 18 gauging stations showed statistically significant trends in their observed data by fulfilling the threshold criteria of p-value≤0.05. However, the two stations, namely, Hatyrik-Homo and Pilot Point AK, did not have the complete set of data for 30 years (1971-2000). The other two stations (Sredne-Kolymsk and Igarka), highlighted in red (Table 11), had the complete set of time series data available from 1971 to 2000 and were taken into consideration to check whether the simulated data from all nine models gave statistical significance trends or not. The significance test of simulated time series for all gauging stations was possible to perform regardless of the p-values, but due to time constraints, the significance of the linear trend for two gauging stations was considered in this study. The linear trend analysis for simulated time series was also carried out in the R environment.

Table 12 provides linear trend analysis results of annual mean simulated discharge time series for Sredne Kolymsk station of Kolyma basin. Here, the significance test was carried out for all participating models and based on the obtained p-value; it can be said that only DBH and LPJML models driven by all four climate forcing datasets were able to provide

significant trend from their simulated data. MPI-HM and MATSIRO models also showed significant trends in their simulated data-driven by only two climate forcing data. The rest of the models did not show a sign of a significant trend in their simulation results. The p-value of less than 5% obtained from the given analysis is highlighted in Table 12 in red colour. All models were able to estimate the decreasing trend (negative slope) in their simulated discharge over the years when compared with the observed discharge trend in Sredne-Kolymsk station.

Table 12: Linear trend analysis results of annual mean simulated discharge time series for Sredne Kolymsk, Kolyma

Trend Analysis of Simulated Discharge (Annual Mean)										
River Basin: Kolyma										
Gauging Station: Sredne Kolymsk (1971-2000)										
			WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
Coefficients	y-intercept	gswp3	2220.41	2497.62	2791.82	2412.30	3035.20	2328.11	1785.18	2911.46	1523.62
		princeton	2268.94	1743.53	2231.99	1761.12	2605.23	1281.45	654.18	2197.99	990.92
		watch	2223.06	2712.99	2623.54	2232.62	1837.52	2265.44	2164.62	2697.44	1339.26
		wfdei	2324.74	2702.53	2534.45	2223.38	1993.65	2265.18	2175.81	2712.04	1503.24
	slope	gswp3	-0.0071	-0.0922	-0.0666	-0.0783	-0.0671	-0.0661	-0.0441	-0.1042	-0.0387
		princeton	-0.0136	-0.0640	-0.0452	-0.0517	-0.0609	-0.0112	-0.0037	-0.0875	-0.0240
		watch	-0.0084	-0.0914	-0.0575	-0.0655	-0.0121	-0.0668	-0.0561	-0.0925	-0.0299
		wfdei	-0.0227	-0.1333	-0.0695	-0.0759	0.0440	-0.0870	-0.0636	-0.0966	-0.0269
p-value		gswp3	0.7519	0.0117	0.0566	0.0335	0.0527	0.0521	0.0858	0.0128	0.1527
		princeton	0.5920	0.0189	0.1131	0.0745	0.0554	0.6319	0.7896	0.0057	0.2096
		watch	0.6971	0.0212	0.0797	0.0600	0.6316	0.0357	0.2203	0.0149	0.2473
		wfdei	0.3335	0.0013	0.0276	0.0323	0.1356	0.0088	0.2474	0.0115	0.3795
p-value in %		gswp3	75.2	1.2	5.7	3.4	5.3	5.2	8.6	1.3	15.3
		princeton	59.2	1.9	11.3	7.4	5.5	63.2	79.0	0.6	21.0
		watch	69.7	2.1	8.0	6.0	63.2	3.6	22.0	1.5	24.7
		wfdei	33.4	0.1	2.8	3.2	13.6	0.9	24.7	1.1	38.0

Table 13 provides linear trend analysis results of annual mean simulated discharge time series for Igarka station of the Yenisei basin. Here, the significance test was also carried out for all participating models and based on the obtained p-value from the analysis; it can be said that all models driven by all 4 climate forcing datasets were able to provide significant trend from their simulated data except for LPJML. Also, PCR-GLOBWB model forced by the 'wfdei' climate forcing dataset failed to show a significant trend in its simulated data. The p-value of less than 5% obtained from the given analysis is highlighted in Table 13 in red colour. All models were able to estimate the decreasing trend (negative slope) in their simulated discharge over the years when compared with the observed discharge trend in Igarka station.

Table 13: Linear trend analysis results of annual mean simulated discharge time series for Igarka, Yenisei

Trend Analysis of Simulated Discharge (Annual Mean)										
River Basin: Yenisei										
Gauging Station:Igarka (1971-2000)										
			WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
Coefficients	y-intercept	gswp3	25502.07	30488.57	24445.27	18843.06	20680.62	21213.61	17046.43	17145.65	13897.84
		princeton	25225.81	28460.76	22366.28	18673.45	19834.17	16788.10	10653.78	16064.38	13857.79
		watch	24932.09	32513.90	24106.36	19148.99	13993.12	16780.95	13261.50	17589.21	14140.34
		wfdei	26739.59	32340.56	22901.98	18978.11	14664.27	16978.51	13106.38	17580.21	14665.39
	slope	gswp3	-0.7753	-1.5065	-0.6313	-0.7065	-0.6361	-0.8781	-0.6616	-0.2880	-0.5449
		princeton	-0.7231	-1.3282	-0.4606	-0.6310	-0.5604	-0.5300	-0.2645	-0.3510	-0.5393
		watch	-0.7002	-1.5924	-0.6345	-0.6755	-0.3204	-0.5626	-0.4713	-0.2852	-0.6186
		wfdei	-0.9614	-1.8147	-0.7194	-0.8749	0.0524	-0.7752	-0.6662	-0.3861	-0.7717
p-value		gswp3	0.00002	0.00002	0.00011	0.00011	0.00001	0.00003	0.00007	0.1489	0.00003
		princeton	0.00006	0.00002	0.00129	0.00009	0.00001	0.00006	0.00665	0.0555	0.00002
		watch	0.00008	0.00002	0.00003	0.00003	0.00105	0.00020	0.00024	0.1860	0.00002
		wfdei	0.00001	0.00001	0.00002	0.000004	0.3621	0.000003	0.00002	0.0724	0.000001
p-value in %		gswp3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.9	0.0
		princeton	0.0	0.0	0.1	0.0	0.0	0.0	0.7	5.6	0.0
		watch	0.0	0.0	0.0	0.0	0.1	0.0	0.0	18.6	0.0
		wfdei	0.0	0.0	0.0	0.0	36.2	0.0	0.0	7.2	0.0

The statistical significance test using p-value for linear trend analysis of observed and simulated time series provides confidence to the modellers and researchers. This gives solid proof that the predicted values from the models are trustable, and the models can correctly predict the trends when compared it with the observed trends. It also provides support when projecting the trends in the future runs or when dealing with climate change impact assessments. Due to its importance, the linear trend estimation and analysis is carried out in most of the scientific works and when dealing with mass data analysis.

5.9 Extreme Flows Analysis (High Flows and Low Flows)

Extreme flows take into consideration both high and low flow conditions. In this study, extreme flows were represented as a combination of both high and low flows. High-flows is emulated when there is a flood-like situation observed in the rivers whereas, in low-flows condition, there is a drought-like situation observed in the rivers. In Pan-Arctic region, during spring and summer periods high-flows occur in the Pan-Arctic watersheds, and during the winter period, the flow in the river is low because of snowfall and ice formation due to freezing temperatures. Thus, these high and low flows conditions were essential to analyze in this study as they frequently occur in this region and play a significant role in understanding the river discharge dynamics in both spatial and temporal extents.

For the extreme flow analysis, the daily discharges from both observed and simulated data were analysed. The daily time series discharges of outlet station of the basins were used for this study. The observations of only the basin outlet station were available to carry out the analysis of the extreme flows in the basins. The daily discharges were used to calculate various flow percentiles from the Flow Duration Curves (FDCs). "The characteristics of daily discharges were analysed using flow duration curves (FDCs), where every single discharge value is related to the percentage of time, which is equalled or exceeded" [Smakhtin 2000; Liersch et al. 2016]. FDCs summarize discharge variability of a time series and display the complete range from low flows to flood events. In order to analyse and visualize low and high flow characteristics, 10 percentile values ($Q_{0.01}$–$Q_{99.99}$) were used based on the entire daily discharge time series of the 30-year reference period. This method was applied to assess whether model performance is suitable to study high and low flow situations as well as their extremes [Liersch et al. 2016].

For the extreme-flows analysis, Liersch et al. [2016] approach was followed. In their research work, they used five different flow duration curve (FDC) percentiles each to represent the high-flows and low-flows conditions. The high flows (HF) were represented by flow percentiles such as Q_{10}, Q_5, Q_1, $Q_{0.1,}$ and $Q_{0.01}$ and the low flows (LF) were represented by Q_{90}, Q_{95}, Q_{99}, $Q_{99.9,}$ and $Q_{99.99}$. The analysis was carried out in the R environment, and the R script used for this analysis is provided in the Appendix. All 10 flow percentile values were calculated for both observed data and simulated data obtained from all nine models driven each by four climate forcing datasets using R. In order to compare the percentile values obtained from observed and simulated datasets, efficiency criteria, Percent Bias (PBIAS) was used. Here, only PBIAS was used as a criterion to evaluate the model performance in regards to extreme flows due to time constraints. Other efficiency criteria could be used to make the evaluation stronger and more convincing, but PBIAS should be enough to support the findings from the analysis.

Here, PBIAS was calculated by aggregating flow percentiles values for each flow condition. The 5 percentile values from observed time series and the difference between observed and simulated values were added together, to get the overall PBIAS value. This method of calculating PBIAS was applied to all model simulation results. This analysis process was carried out separately for high flows and low flows conditions. Then, the

obtained PBIAS values were rated based on threshold values provided in Table 10. The rating score of '1' was used for |PBIAS| ≤ 25% to indicate 'good' performance, the rating score of '0.5' was used for 25% ≤ |PBIAS| ≤ 50% to indicate 'weak' performance, and the rating score of '0' was used for |PBIAS| ≥ 50% to indicate 'poor' performance. This scoring was applied to all scenarios and in all basins to get the model performance index finally. The method of obtaining the model performance index by rating PBIAS in the analysis of the extreme flows was the same as the discharge time series analysis.

5.10 Snow Water Equivalent Time Series Analysis and Visualization

Reading and Manipulating Time Series Data

The method used for reading and manipulating the snow water equivalent data was the same as the one used for discharge variable. Here instead of choosing the 18 gauging stations for analysis as done for discharge data, 4 or 5 points in the basin were chosen according to the size of the basin. For Lena, Ob, and Yukon river basins 5 locations were chosen, and for the rest of the basins, 4 locations were chosen. The location of points in each basin was randomly chosen according to the directions like North, South, East, West, and Center. This way of analyzing the snow cover data was chosen because these points would help to get relevant information about the entire basin. As these points were distributed in all directions, they gave a good overview of the basins under study. For this analysis, 6 models (WaterGAP2, DBH, MPI-HM, MATSIRO, PCR-GLOBWB, and LPJmL) were considered as the snow cover data was not available for the other 3 models. That could be because those models might not have done simulations of the snowpack data. Analyzing the snow data might not have been in their scope of model simulation runs. The snow water equivalent simulated data was available in the monthly time series when extracted from the NetCDF files. Here, the data for 21 years (1980-2000) was analysed for the model validation runs and model performance evaluation. The mass data analysed for this variable was about the same as that used in the case of discharge. It was around 648 iterations (27 locations x 6 models x 4 climate forcing datasets). The run time in the R environment was quicker as the NetCDF files for this output variable only contained the monthly data (252 monthly series). It was quicker to analyse the snow data than the discharge data because of the lesser amount of time series data. The R script for the snow water equivalent time series analysis is available in the Appendix.

Aggregating Time Series Data into different Temporal Resolutions

Since the time series already had a monthly data available from the start, there was no need of aggregating the time series as it was done for discharge time series analysis. However, the long-term average monthly snow water equivalent time series were aggregated by calculation from the monthly time series in case of both simulated and observed data. The time series reduced from 252 monthly data (11 years) to just 12 values representing the overall mean for each month. The rest of the process carried out was the same as with discharge time series analysis.

Plotting Time Series

After completing the steps of reading, manipulating and aggregating the snow water equivalent time series data into the desired format, the R script (Appendix) was written to visualize the time series for both simulated and observed data. The R scripts helped to produce the plots of snow water equivalent time series in different temporal resolutions such as monthly, and seasonal. The long-term average monthly (seasonal) snow water equivalent series plot obtained from manipulating the monthly time series data are presented in the chapter "Snow Water Equivalent Data Analysis: Results and Discussion."

Location of Interests

In order to analyse and visualize the snow water equivalent data in the basins, the locations in the basins were chosen such that they cover the basin evenly and possibly in all directions. Therefore, the points of interest in different directions like North, South, East, West, and Center of the basins were considered for the analysis. These points of interest helped to represent and study the snowpack dynamics in the basins. To get geographic information such as longitude and latitude of these selected points, ArcGIS was used. Figure 14 presents the geographic information (latitude and longitude) of the selected points (green dots) in the basins under study. After receiving the geographic location data of the points from ArcGIS, grid cell values representing the longitude and latitude information of these points of interest were extracted from the NetCDF files.

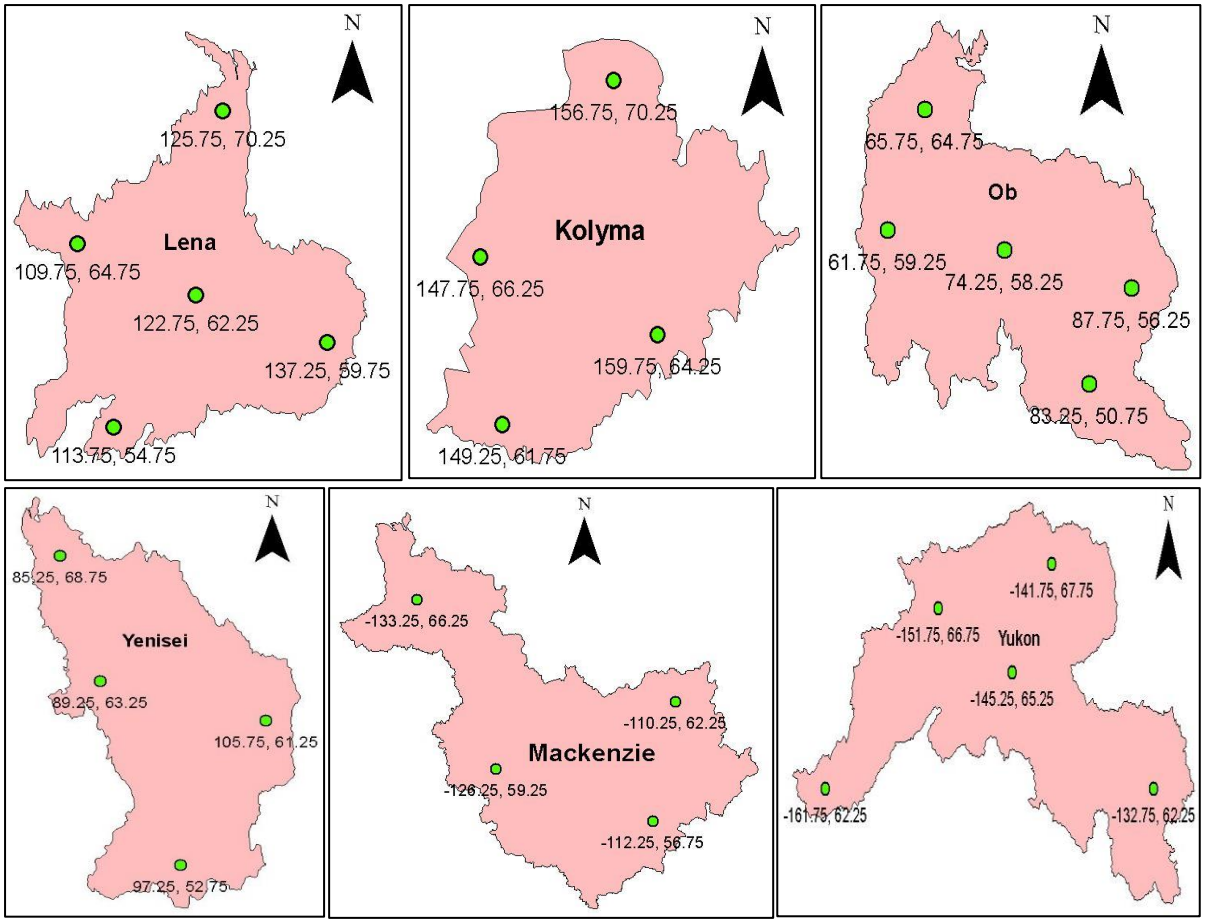


Figure 14: Location of Interests chosen in all river basins understudy to analyse snow water equivalent model output variable

6. DISCHARGE DATA ANALYSIS: RESULTS AND DISCUSSION

The desirable analysis of river discharge was obtained from a coherent impact model intercomparison by driving the models with climate forcing data from the same source and for the same periods. The processing of the simulated and observed discharge time series to get coherent and consistent datasets are discussed in the Data Processing chapter. Moreover, in the discharge time series analysis and visualization section, the methodology of how the work was initiated to analyse the discharge data is covered. In this section, the results obtained from the discharge data analysis are presented and discussed. The models were validated by comparing simulated and observed datasets (monthly hydrographs and seasonal dynamics) using both visual inspection and model efficiency criteria. Based on the validation, the best and poor performing models were selected for each gauging stations and the participating models' overall performance was then, evaluated both model and basin-wise using Model Performance Aggregated Index. The timeseries of the simulated and observed discharge were also used to analyse trends and extremes: percentiles of the high and low flows. The method of evaluating the extreme flows, including high and low flows, was similar to mean discharge evaluation. The details of all the hydrological indicators mentioned above are presented in the sections below.

6.1 Model Validation Runs

The validation runs of the hydrological models were carried out for each gauging station in the analysis period 1971-2000. Here, in this study, the models were validated using both subjective and objective assessment methods. The subjective assessment included a visual comparison to determine the closeness of fit between observed and simulated discharge. The objective assessment involved a mathematical analysis of the two datasets using the model efficiency criteria such as NSE, PBIAS, and $\Delta\sigma$.

6.1.1 Comparison of Simulated and Observed Data

In this study, the comparison between simulated and observed discharges was carried out using monthly hydrographs (monthly dynamics) and long-term average monthly discharge (seasonal dynamics) plots. These two hydrological indicators have unique characteristics, which are explained briefly in the Technical Concepts chapter. The two different time series lengths of the monthly dynamics and seasonal dynamics were used to analyse the discharge values. The seasonal dynamics indicator was only used for visual comparison (subjective assessment) between simulated and observed discharges whereas both monthly hydrographs and seasonal dynamics hydrological indicators were used for comparing simulated and observed discharges in case of the objective assessment using numerical efficiency criteria as explained in the Technical Concepts chapter.

Multiple plots of all nine participating models, which featured all the considered gauging stations of a basin together side by side, were produced to make the visualization of river discharge dynamics of the particular basin easier. These multiple plots reduced the number of plots that were being produced before and made the analysis work more convenient. The figures below show long-term average monthly (seasonal) discharge modelled using three randomly chosen models out of the nine for each basin and the plots from the other six remaining models are provided in the Appendix.

Figures 15, 16, 17, 18, 19, and 20 presented below show examples of the multiple plots featuring seasonal dynamics comparison plots between simulated mean and observed data at all gauging stations of the basin. A wide range of differences in the plots was visualized in the study areas. Each model simulated different long-term average monthly discharge values, and the differences can be easily noticed from the figures below. The significant differences in simulated values may be due to the various kinds of different uncertainties associated with each model. Results from the Global Hydrological Models in most cases showed varied uncertainty ranges (depending on each model) in terms of a considerable bias against the observations. Some models were able to simulate the river discharge more or less close to the observations, and few of them failed to simulate the river discharge as per the observations. Each river basin discharge analysis will be dealt with individually below.

Lena River Basin

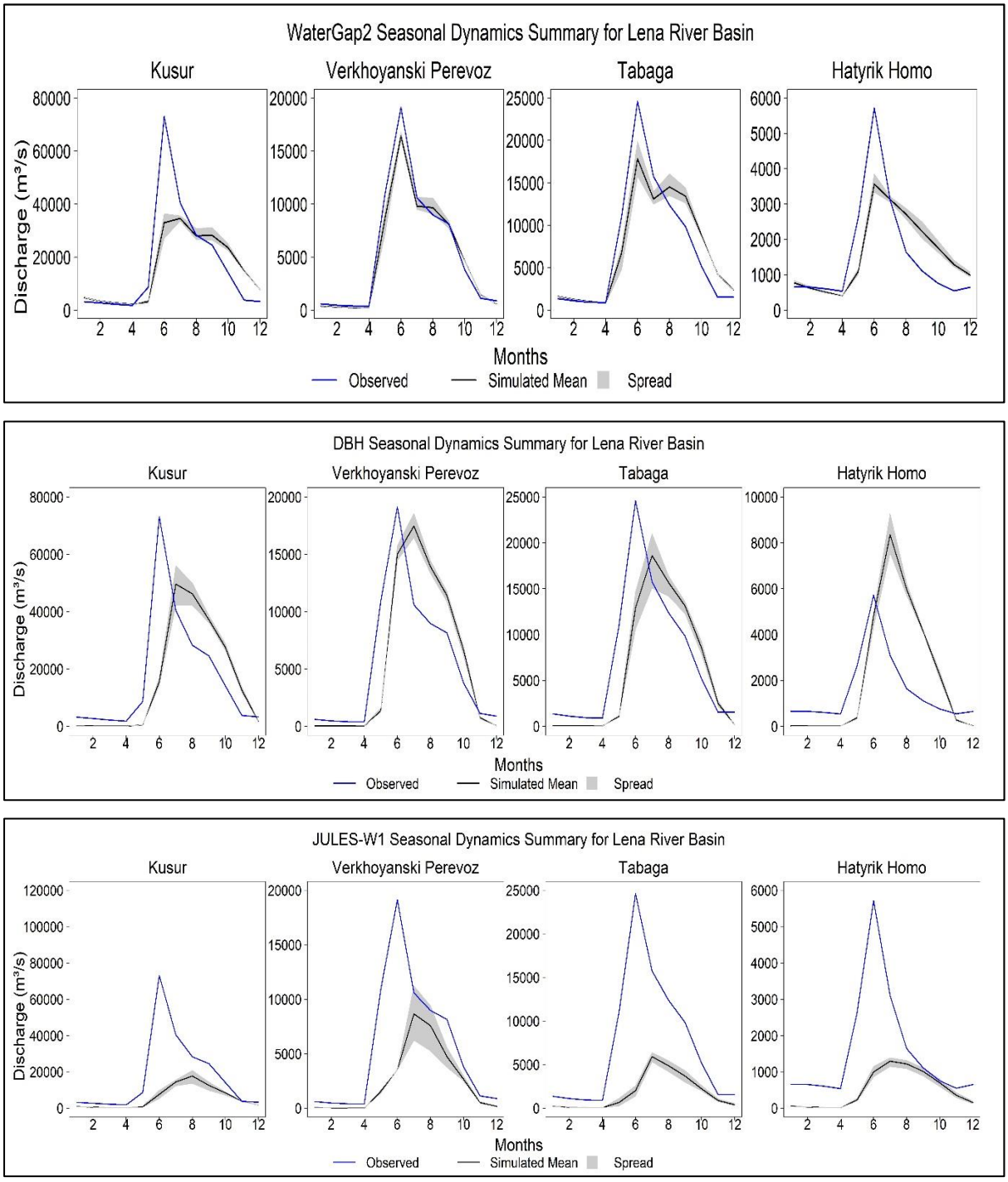
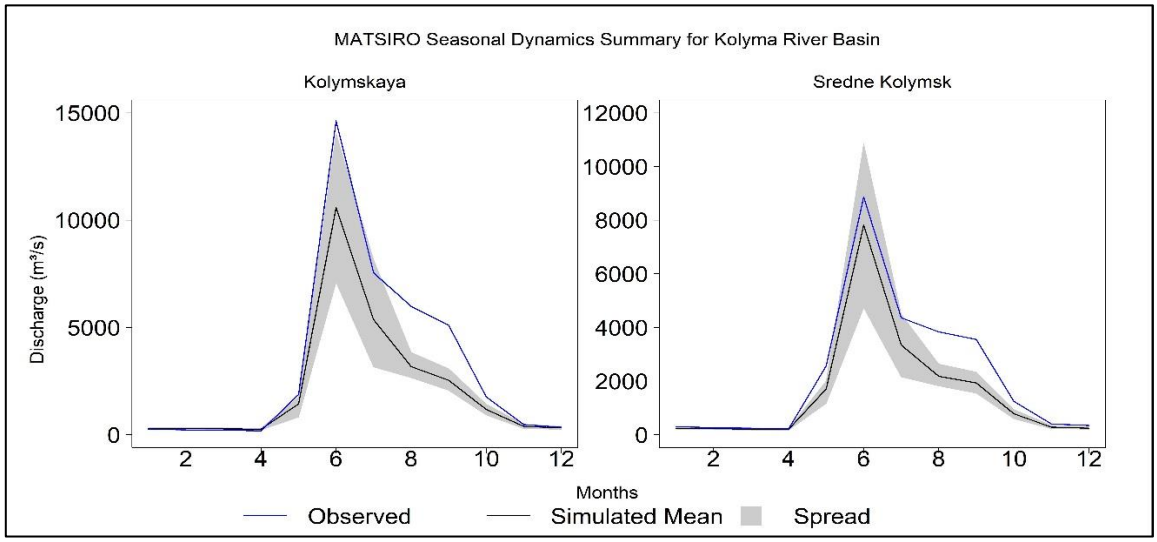
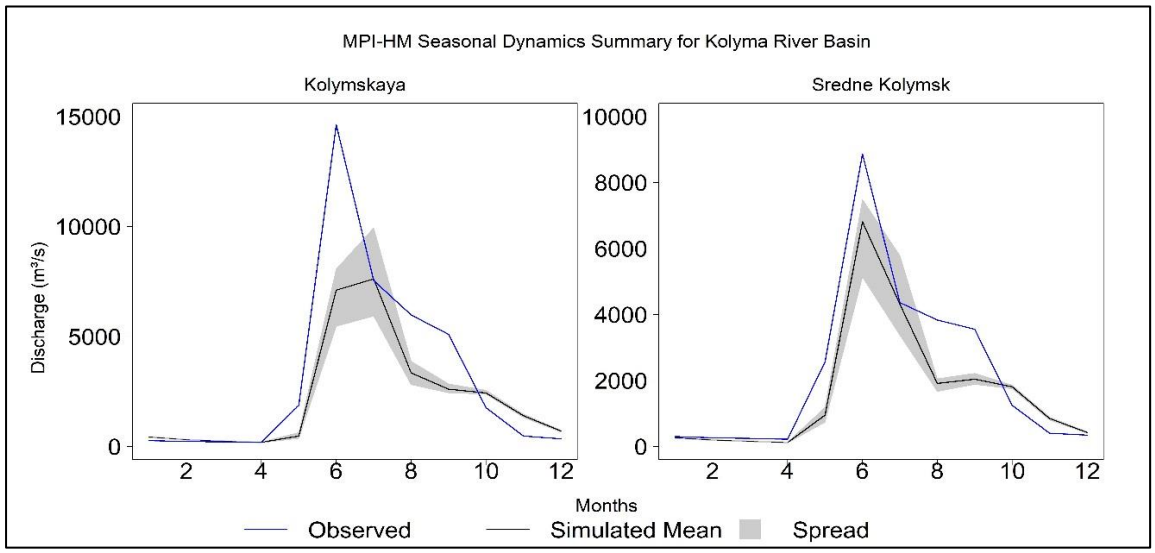


Figure 15: Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by WaterGAP2, DBH, and JULES-W1 at 4 gauging stations of Lena Basin

Kolyma River Basin





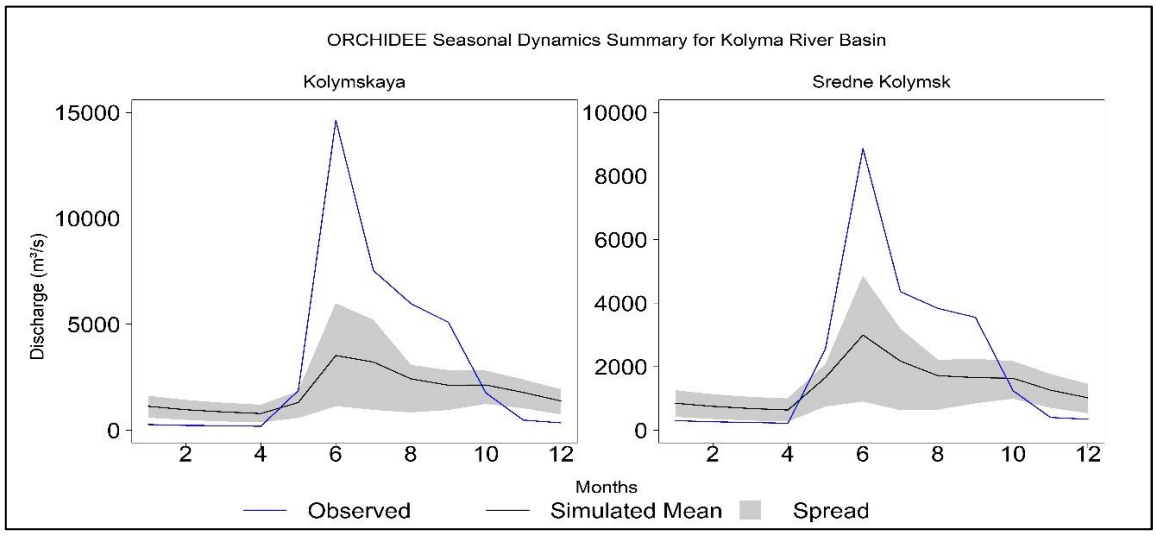


Figure 16: Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by MATSIRO, MPI-HM, and ORCHIDEE at 2 gauging stations of Kolyma Basin

Ob River Basin

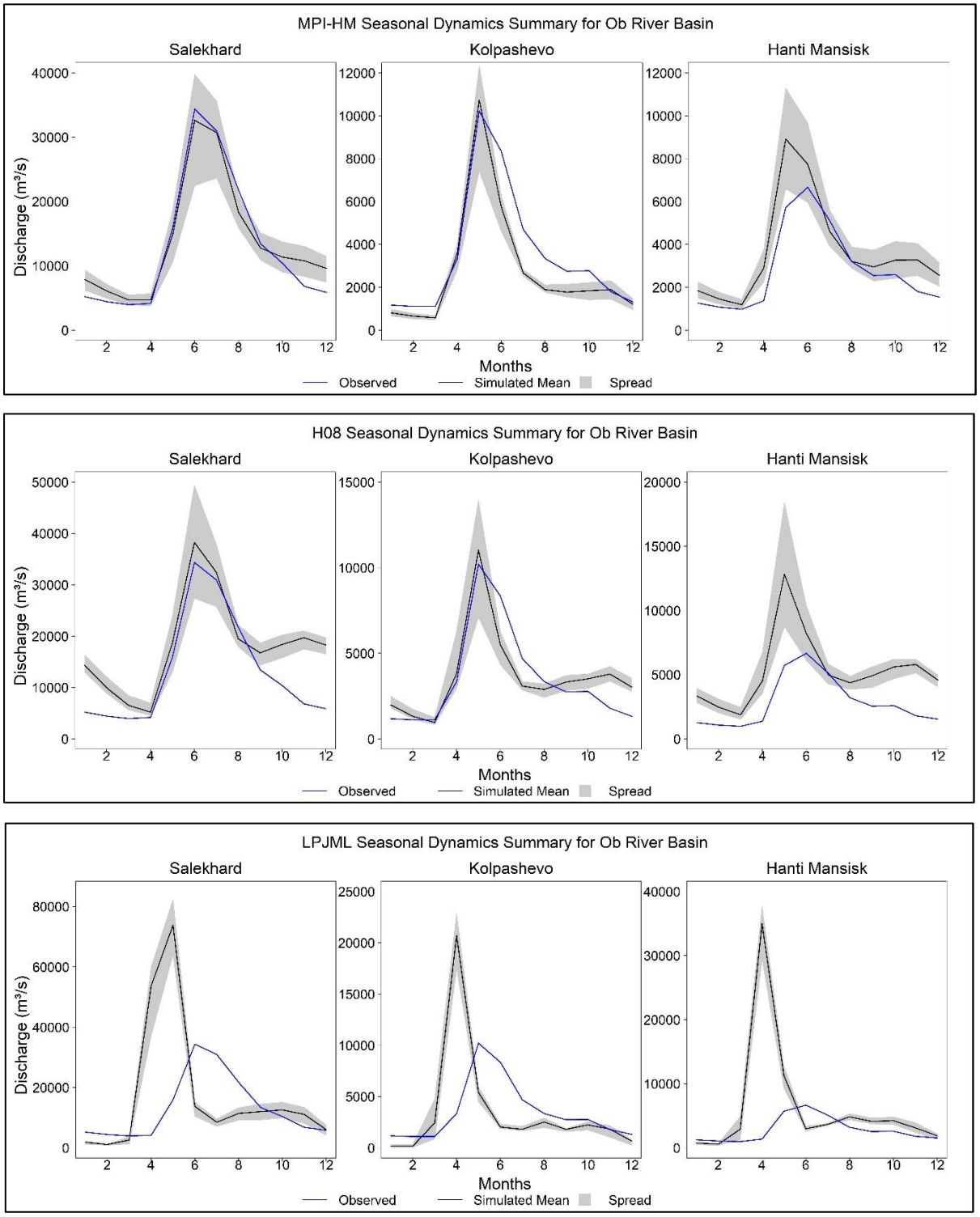


Figure 17: Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by MPI-HM, H08, and LPJML at 3 gauging stations of Ob Basin

Yenisei River Basin

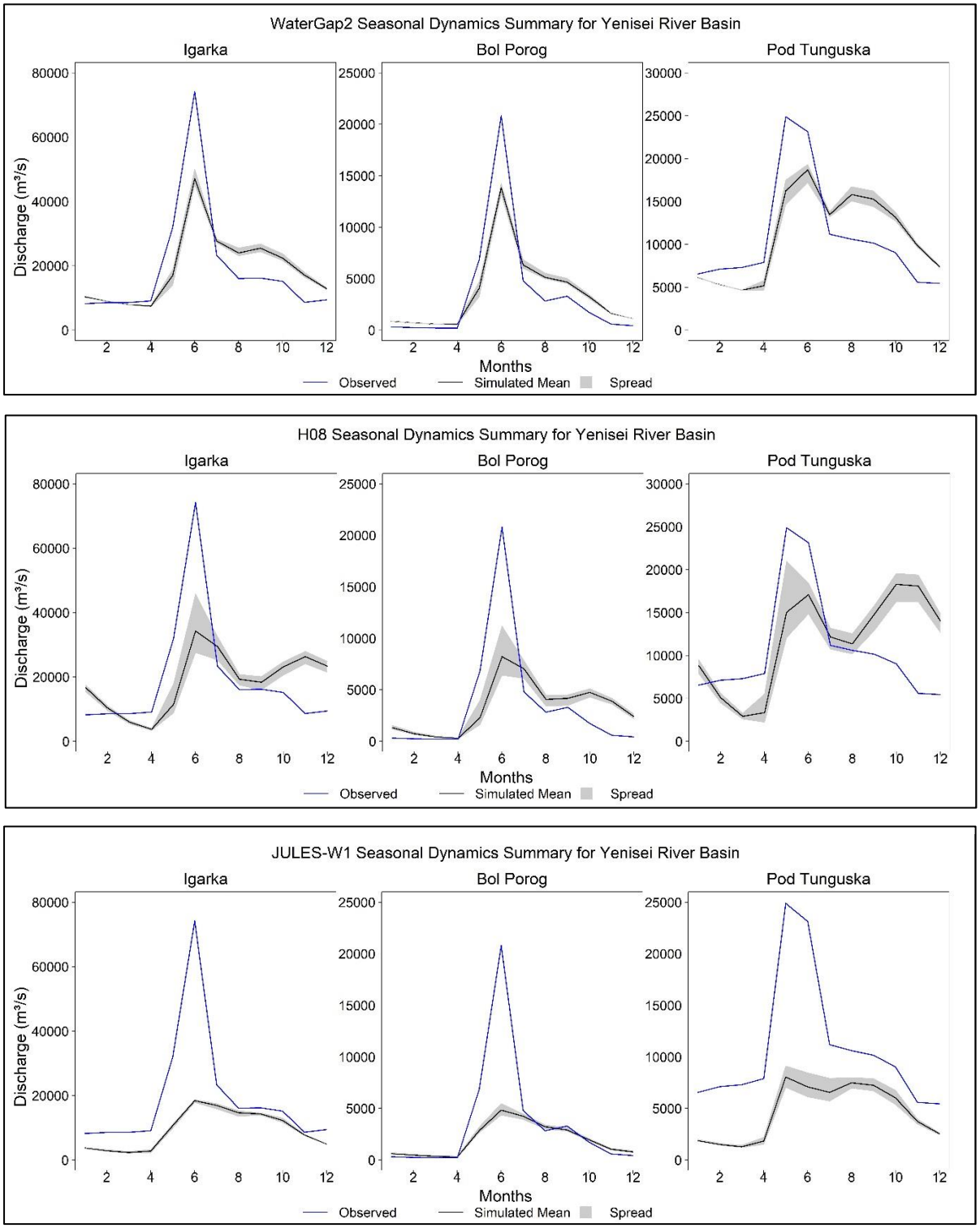


Figure 18: Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by WaterGAP2, H08, and JULES-W1 at 3 gauging stations of Yenisei Basin

Mackenzie River Basin

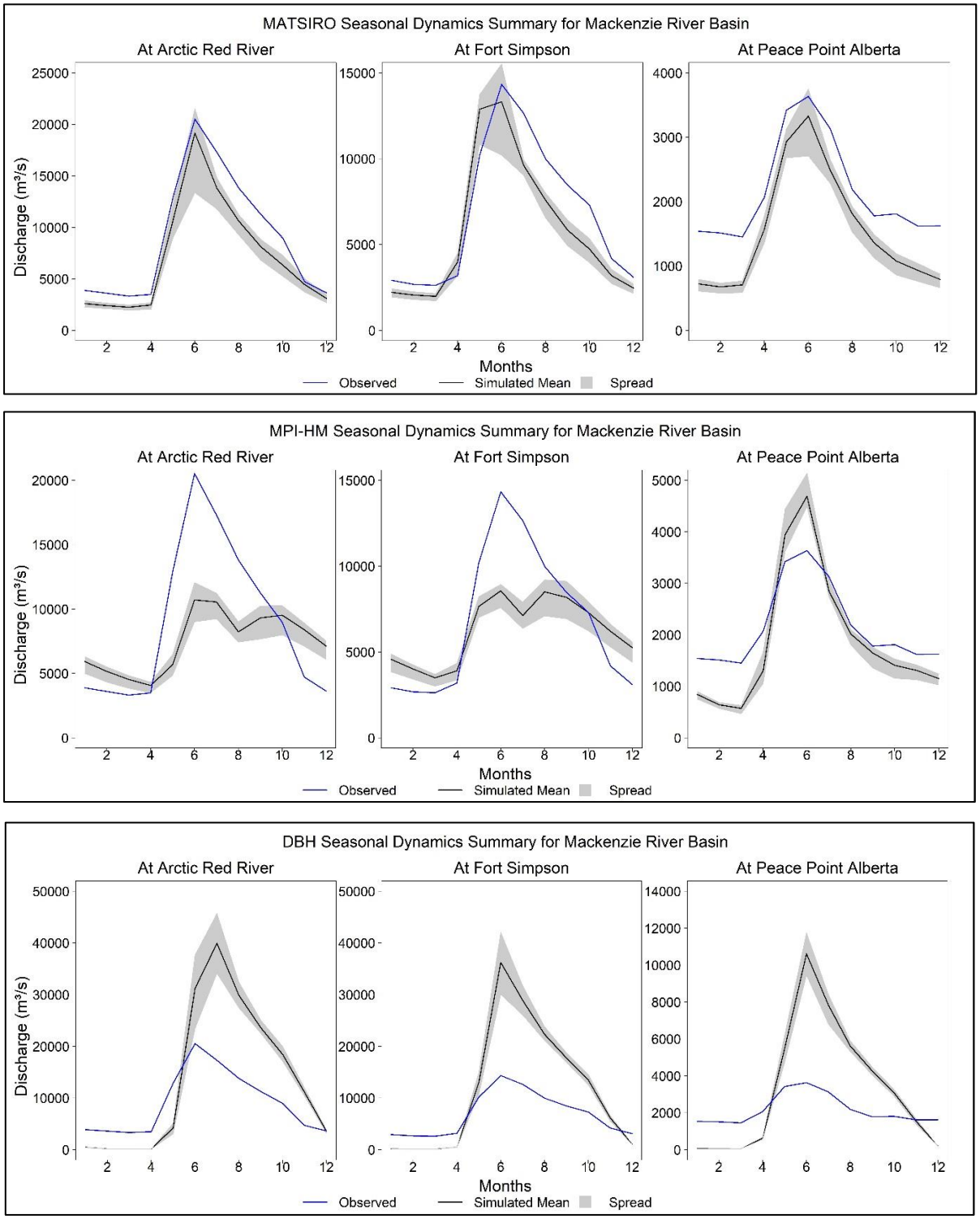


Figure 19: Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by MATSIRO, MPI-HM, and DBH at 3 gauging stations of Mackenzie Basin

Yukon River Basin

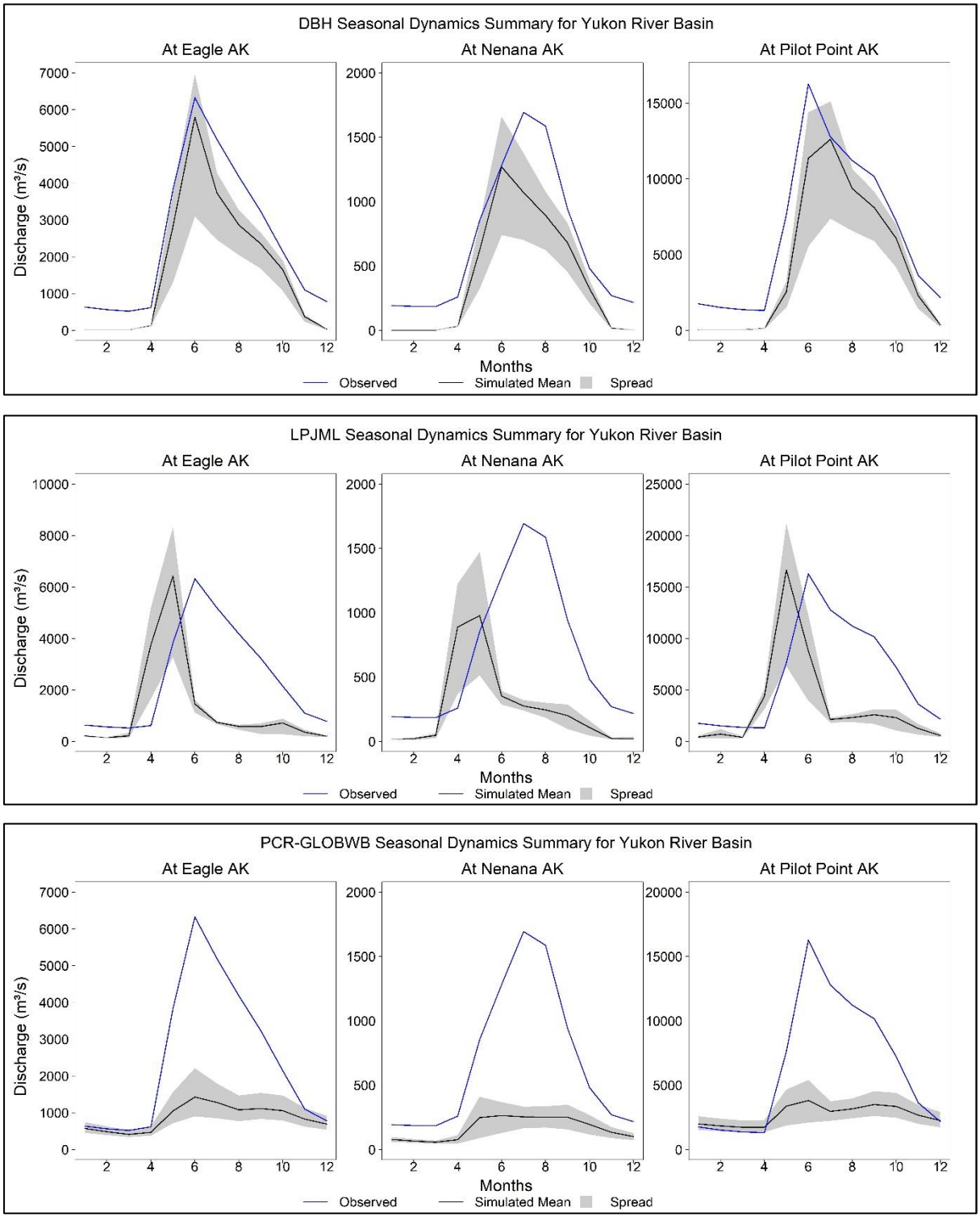


Figure 20: Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by DBH, LPJML, and PCR-GLOBWB at 3 gauging stations of Yukon Basin

6.2 Subjective Assessment Results

Lena River Basin

Most of the model simulations of river discharge were underestimated in the first half of the year (January to July) but overestimated during the other half, and they failed to capture the peak flow observed at the onset of the spring season. However, few models managed to capture the peak flow, but it was simulated either too early or too late than the observed period of peak flow. In many cases, the simulated discharge followed a similar seasonal runoff cycle, as seen from the observations. The discharge level increased during the spring and summer months, followed by a sharp decline during winter months. This phenomenon was captured by most of the model simulations. Few cases were discovered where the model simulations captured the observed discharge dynamics and seasonality behaviour accurately such as WaterGAP2 and MATSIRO in Verkhoyanski Perevoz station; MPI-HM and MATSIRO in Hatyrik Homo station but no one model simulated seasonal dynamics well at the outlet station Kusur. Several models (LPJmL, DBH, H08, JULES-W1, PCR-GLOBWB) showed lag while simulating the peak flow occurrence in the Lena basin.

Kolyma River Basin

The simulation results from MATSIRO (Figure 16) have managed to reproduce seasonal dynamics in one gauging station, Sredne Kolymsk quite accurately. The LPJML model quite accurately captured the peak flow at Kolymskaya station, but the summer flow was notably underestimated. Six out of nine models: ORCHIDEE, DBH, H08, JULES-W1, PCR-GLOBWB and WaterGAP2 significantly underestimated peaks and seasonal dynamics in both stations.

Ob River Basin

For MPI-HM model, the simulated mean dynamics was quite well when compared with observations at 2 stations out of 3 that were under consideration for this basin. In case of MATSIRO, the peak flow was well reproduced on average for Salekhard but spread resulting from climate input was large. In H08 model, the peak flow and dynamics in April - September were reproduced quite well in 2 stations out of 3 (Salekhard and Kolpashevo). Further, in ORCHIDEE, the peak flow and dynamics were reproduced quite well by the model mean in Kolpashevo but spread was large. Finally, LPJmL and DBH

showed significant overestimation whereas JULES-W1 and WaterGAP2 showed significant underestimation (except WaterGAP2 for Kolpashevo).

Yenisei River Basin

MATSIRO, MPI-HM, and ORCHIDEE models under performed in this basin as they failed to capture the peak flow. All three models showed significant underestimation of peak flows in all the gauging stations considered to represent this basin. Furthermore, all three models showed underestimation in reproducing the seasonal dynamics at the Pod Tunguska station. In LPJML model, the peak flow occurrence was not captured correctly in all stations (earlier peaks). In case of DBH, H08, and PCR-GLOBWB models, the simulated discharge was underestimated in the first half of the year, but later on, it changed into overestimation. Additionally, the peaks were significantly underestimated (expect DBH for Pod Tunguska: overestimated peak with a time lag). JULES-W1 did not perform well when assessing the plot visually. The simulated results were underestimated throughout the period, and they did not comply with the observed data.

Mackenzie River Basin

In MATSIRO, mean simulated results for 2 stations (Arctic Red River and Fort Simpson) out of 3 were quite good but showed underestimation at the Peace Point Alberta station. DBH showed significant overestimation of discharge. In case of H08, overestimations of peaks and dynamics were observed. LPJML showed significant overestimation of peaks occurring earlier than observed and too low discharge in the second half of the year. Finally, JULES and PCR-GLOBWB showed underestimation of discharge.

Yukon River Basin

WaterGAP2 performed quite well in 2 stations (Eagle AK and Nenana AK) out of 3 considering seasonal dynamics, though the peaks were underestimated. In DBH, the seasonal dynamics was captured but with some underestimation of peaks and dynamics. Further, in LPJML, the seasonal dynamics was not captured, peak discharge occurred earlier than in observations, and discharge in the second half of the year was too low. MATSIRO, MPI-HM, ORCHIDEE showed poor performance with significant underestimation of discharge and peaks of discharge. Finally, JULES-W1 and PCR-

GLOBWB showed very poor results with significant underestimation of discharge and peaks of discharge.

6.3 Objective Assessment Results

The subjective assessment does not guarantee what is observed visually from the plots may necessarily be the truth. The subjective assessment could be misleading and false at times, and the conclusions drawn solely from them might not be enough to convince the scientists, modelers, and users' community. Therefore, in addition to the visual inspection of data using plots, the efficiency criteria are also essential to validate the models and provide support to the visual observations made using the plots.

The objective assessment helps to reinforce the findings. Therefore, an objective assessment was carried out to compare the simulated and observed data in this study. For the assessment of the models, three different efficiency criteria such as NSE, PBIAS and Bias in SD were calculated in total from the monthly hydrographs and long-term average monthly (seasonal) discharges. NSE and PBIAS were calculated based on monthly hydrographs, whereas for the seasonal dynamics NSE and Bias in Standard Deviation (Bias in SD) were calculated. The details of these criteria and the reasons behind considering them in this study are provided in the Technical Concepts chapter.

As there were mass data to be assessed in this study, it was not possible to present all the calculated values in the main report and to discuss every single one of the values. Thus, Tables 14, 15, 16, 17, 18 and 19 presented below show the efficiency criteria values calculated using the simulated (monthly and seasonal dynamics) and observed data at the outlet station of the basins. The calculations were carried out individually using all climate forcing datasets in a given model, and this was done for all 9 participating models at 18 gauging stations. The tables for the rest of the stations apart from the outlet ones are provided in the Appendix.

Table 14: Efficiency criteria to assess simulated and observed discharge at Kusur, Lena

LENA - KUSUR									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.6	0.17	0.068	0.45	0.46	0.68	0.48	0.19	-0.061
princeton	0.5	0.21	-0.15	0.31	0.45	0.53	0.024	0.46	-0.042
watch	0.66	0.13	-0.11	0.43	0.32	0.53	0.27	0.23	-0.013
wfdei	0.67	0.17	-0.15	0.4	0.43	0.56	0.27	0.23	0.063

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-9.8	-11	-11.1	-31.6	1.9	-36.3	-48.1	-28.6	-71.5
princeton	-9.5	-14.6	-11.5	-31.4	0.7	-41.5	-64	-38.1	-64.8
watch	-9.6	0.5	-9.6	-29.3	-22.1	-40.3	-46.9	-24.2	-66.5
wfdei	-9.6	-3.4	-11.1	-33.1	-5.4	-40.1	-47.4	-25.9	-63.3

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.62	0.18	0.072	0.47	0.48	0.71	0.51	0.22	-0.088
princeton	0.51	0.21	-0.16	0.33	0.46	0.54	0.022	0.48	-0.063
watch	0.68	0.13	-0.12	0.46	0.33	0.55	0.29	0.28	-0.04
wfdei	0.69	0.17	-0.16	0.42	0.45	0.57	0.29	0.27	0.041

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-39.9	-14.5	-58.8	-48.8	-50.6	-36.7	-52.4	-15.9	-76.8
princeton	-40.1	-19.3	-57.8	-55.1	-48.9	-54.8	-80.7	-23.6	-66.8
watch	-39.3	-2.9	-55.8	-51.1	-62.2	-51.6	-69.3	-11.6	-70.9
wfdei	-38.8	-7	-59.8	-53.5	-54.6	-51.7	-69.7	-12.2	-69.2

Table 15: Efficiency criteria to assess simulated and observed discharge at Kolymskaya, Kolyma

KOLYMA - KOLYMSKAYA									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.46	0.65	0.57	0.56	0.67	0.83	0.48	0.34	0.41
princeton	0.4	0.44	0.32	0.48	0.51	0.56	-0.19	0.51	0.17
watch	0.45	0.68	0.42	0.67	0.34	0.84	0.15	0.37	0.32
wfdei	0.48	0.64	0.31	0.65	0.57	0.8	0.15	0.41	0.38

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-0.3	-22.4	-2.4	-22.7	10.5	-20.5	-38.8	-11.2	-48.4
princeton	-0.1	-45.1	-19	-41.6	-5.3	-51	-75.4	-34	-66.2
watch	-0.6	-12.7	-6.2	-26.9	-28.1	-26.4	-29.7	-16.5	-53.3
wfdei	-0.5	-26.9	-14.3	-30.5	2.3	-31.6	-31.8	-16.7	-45.4

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.46	0.69	0.62	0.66	0.69	0.9	0.54	0.41	0.43
princeton	0.37	0.44	0.33	0.5	0.52	0.55	-0.26	0.52	0.13
watch	0.44	0.72	0.45	0.73	0.33	0.87	0.23	0.42	0.32
wfdei	0.47	0.65	0.32	0.69	0.59	0.82	0.2	0.44	0.4

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-69.8	-23.7	-36.8	-29	-38.1	-3.7	-58	29	-55.4
princeton	-71.3	-48.4	-55	-54.3	-45.9	-54	-93.4	-10	-69.2
watch	-71.9	-15.4	-43.6	-38.7	-60.3	-29.1	-80.6	7	-59.4
wfdei	-70.1	-29.9	-54	-40	-44.6	-32.8	-82.4	10	-53.5

Table 16: Efficiency criteria to assess simulated and observed discharge at Salekhard, Ob

OB - SALEKHARD									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.52	-0.59	0.37	0.89	0.35	0.47	0.85	-5.1	0.29
princeton	0.36	0.39	0.5	0.73	0.41	0.48	0.4	-3.1	0.33
watch	0.56	-1.7	0.35	0.8	0.36	0.69	0.86	-5.9	0.34
wfdei	0.56	-1.2	0.4	0.86	0.36	0.68	0.69	-5.6	0.39

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	1.1	57	-4.3	-4	11.1	-4.4	10.9	33.2	-35.4
princeton	1.5	25.5	-0.54	-12.9	-14.3	-26.7	-31.1	3.5	-36.1
watch	1.6	77.8	-2.5	24.7	-23.5	-6	-4	49.6	-34.1
wfdei	4.4	68.2	-0.8	10.4	2.8	-10	-18.7	44.1	-30.9

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.53	-0.72	0.46	0.97	0.42	0.59	0.97	-4.7	0.28
princeton	0.36	0.38	0.54	0.77	0.46	0.5	0.44	-2.8	0.32
watch	0.57	-2	0.39	0.86	0.4	0.78	0.96	-5.6	0.33
wfdei	0.58	-1.4	0.43	0.95	0.48	0.74	0.79	-5.4	0.4

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-63.6	99	21	-4	-26	23	8	115	-45.6
princeton	-70.1	53	-34.5	-40.4	-48.7	-50.1	-62.3	74	-47.6
watch	-62	135	9	4	-53.7	-11.6	-16	132	-47.1
wfdei	-61.5	121	-31.4	-8.9	-33	-18.2	-37.5	126	-48

Table 17: Efficiency criteria to assess simulated and observed discharge at Igarka, Yenisei

YENISEI - IGARKA									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.59	0.13	0.56	0.58	0.34	0.72	0.51	-0.18	0.048
princeton	0.55	0.065	0.1	0.41	0.21	0.44	0.09	0.076	0.044
watch	0.65	0.072	0.27	0.58	-0.022	0.45	0.16	-0.36	0.03
wfdei	0.67	0.13	0.12	0.51	0.24	0.4	0.055	-0.3	-0.0072

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-0.1	-0.2	-0.3	-30.8	-19.1	-24.6	-38.2	-24.9	-49.5
princeton	0.1	-4.3	-4.8	-28.9	-20.9	-34.8	-56.5	-32.4	-49.5
watch	-0.6	6.9	-1.9	-28	-42	-36.2	-50.4	-22.5	-50.6
wfdei	0	-0.9	-10.4	-35.2	-26.5	-41.8	-57.4	-25.7	-52.9

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.64	0.14	0.59	0.63	0.39	0.79	0.56	-0.15	0.052
princeton	0.59	0.065	0.09	0.45	0.25	0.48	0.11	0.1	0.044
watch	0.71	0.082	0.28	0.64	-0.0061	0.5	0.2	-0.39	0.03
wfdei	0.73	0.16	0.13	0.56	0.28	0.44	0.084	-0.31	-0.0031

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-41.8	4	-38.6	-29.8	-57.3	-15.3	-32.8	6	-68.2
princeton	-40.6	-4.4	-52.5	-41.8	-59.8	-53.2	-66.2	-1.6	-67.1
watch	-39.1	12	-46	-33.3	-71.9	-47.1	-50.2	14	-68.3
wfdei	-36.1	2	-56.8	-35.9	-63.9	-53.1	-59.8	11	-69.5

Table 18: Efficiency criteria to assess simulated and observed discharge at Arctic Red River, Mackenzie

MACKENZIE - ARCTIC RED RIVER									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.62	-1.4	-0.073	0.29	0.67	0.72	0.23	-2	0.31
princeton	0.55	-1	-0.14	0.13	0.57	0.65	-0.52	-1.3	0.27
watch	0.57	-3.5	-0.5	0.39	0.34	0.82	0.11	-1.9	0.31
wfdei	0.57	-3	-0.36	0.34	0.62	0.82	0.04	-1.7	0.36

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	1.3	41.6	27.9	-26.7	-12.9	-24.3	-27.1	-20.6	-37.3
princeton	1.5	33.2	25.7	-18.7	-19	-29	-52.7	-31.8	-36
watch	1.2	71.1	38.7	-10.7	-29.8	-13.6	-25.5	-6.5	-33.3
wfdei	1.7	64.6	37.5	-13.5	-11.1	-13.1	-31.7	-8.4	-26.7

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.64	-1.4	-0.043	0.33	0.73	0.79	0.25	-1.8	0.31
princeton	0.57	-1.1	-0.14	0.17	0.63	0.69	-0.58	-1.1	0.27
watch	0.59	-3.7	-0.49	0.47	0.37	0.93	0.14	-1.6	0.32
wfdei	0.59	-3.1	-0.37	0.41	0.74	0.92	0.062	-1.4	0.37

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-56.3	118	11	-60.3	-42.6	0	-67.5	64	-52.9
princeton	-64.1	105	-1	-64.6	-47.8	-30.7	-93.6	26	-53.9
watch	-59.8	170	21	-57.8	-58.7	-3.4	-73.7	49	-53.3
wfdei	-58.7	158	-10.1	-61.1	-42.6	-4.9	-76.9	44	-53.2

Table 19: Efficiency criteria to assess simulated and observed discharge at Pilot Point AK, Yukon

YUKON - PILOT POINT AK									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.81	0.74	0.56	0.19	-0.0069	0.44	-0.072	-0.6	0.069
princeton	0.47	0.19	-0.079	-0.6	-0.57	-0.49	-1.1	-0.59	-0.44
watch	0.73	0.78	0.38	-0.0013	-0.53	0.25	-0.41	-0.6	-0.07
wfdei	0.81	0.79	0.22	0.16	-0.036	0.41	-0.35	-0.69	0.17

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-1.3	-26.7	-10	-48	-45.5	-40.4	-44.3	-35.2	-55.9
princeton	-5.8	-57.1	-40.1	-72.2	-68.6	-71.9	-83.5	-69.6	-74.4
watch	-3.5	-22.5	-18.4	-54.1	-68.4	-51.8	-55	-41.1	-60.4
wfdei	0.4	-18.6	-14.1	-48.5	-49.5	-44.8	-51.1	-32.9	-50.9

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.84	0.79	0.6	0.19	-0.044	0.47	-0.065	-0.55	0.05
princeton	0.5	0.18	-0.11	-0.69	-0.64	-0.56	-1.2	-0.62	-0.51
watch	0.76	0.84	0.4	-0.019	-0.61	0.23	-0.44	-0.54	-0.1
wfdei	0.85	0.86	0.23	0.16	-0.075	0.41	-0.38	-0.6	0.16

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-33.8	-1.1	-34.3	-58.3	-80.7	-14.3	-69.3	15	-60.1
princeton	-35.4	-43.7	-56.2	-84	-90.4	-78.8	-98.5	-60.5	-71.4
watch	-43.9	5	-41.1	-69.5	-90.2	-46.3	-85.2	-3.4	-62.9
wfdei	-33.8	11	-49	-62.4	-78.1	-35.8	-84.8	14	-57.3

The calculated values above are self-explanatory when compared with the threshold range of each criterion. For instance, the NSE value of 1 was regarded as the perfect fit between simulated and observed data. The NSE value ranging from 0.5 to 1 for the monthly dynamics was considered as good for global HMs. Furthermore, the NSE value between 0.3 and 0.5, was regarded as weak but within the acceptable levels of performance, so the model simulation results were accepted. However, for the NSE below 0.3, the model simulation results became unacceptable, and that model's performance was regarded as 'poor. '

Similarly, for PBIAS and Bias in SD, the threshold range remained the same as both were calculated in percentage. PBIAS and Bias in SD with value 0 indicated that there was no bias between simulated and observed data, which was then, considered as the optimal value. The optimal value (0) for PBIAS and Bias in SD indicated that the model's simulations correctly predicted the dynamics seen in the observations. The model was considered as the best one in this given scenario. Also, when the values of PBIAS and Bias in SD were close to zero in both negative and positive x-direction, then, the simulation results were accepted. Thus, the model simulations were regarded as accurate in this case. If the PBIAS and Bias in SD values were high, then the model simulation results became unacceptable, and they were regarded as 'poor' performing model.

Moreover, the negative PBIAS and Bias in SD values were considered as overestimation (simulated values are higher than the observed values) whereas, the positive PBIAS and Bias in SD values were considered as underestimation (simulated values are lower than the observed values). The use of these standard threshold values of the criteria made it very easy to evaluate the performance of the model, and they provided enough evidence to support the subjective assessment carried out above.

6.4 Discussions

As expected, most of the Global Hydrological Models showed varied uncertainty ranges in their simulated results and often a considerable bias when compared against the observations. Most of them failed to predict the peak flow values correctly at considered gauging stations, and there was either a lag or a lead in time of occurrence of the peak flow when compared against the observations. The occurrence of spring flood with high discharge levels in the basins is the most critical seasonal phenomenon observed in the Arctic region. In the spring period, high levels of runoff were observed in the rivers, and it was hard to simulate the floods in the model simulations because there was a lack in input data and the input data were biased to predict the higher flow levels. The uncertainties with input parameters in the models made it difficult to get the desired results at the end.

The Arctic hydrological processes such as snow and permafrost are prevalent in the Arctic basin. Snow plays a significant role, and it is directly linked to the river discharge dynamics in the region. Mostly the northern part and the mountainous regions of the basins are covered in ice throughout the year, which makes it challenging to simulate river discharges in the basin. Due to the complexities of the processes involved, the models did not simulate river discharge accurately and hence, could not predict the results as expected. The snow accumulation and melt in almost all the river basins considered influenced river discharge during the spring-summer period and also high flow peak.

Furthermore, the models in their simulation did not take into consideration the impacts due to human-made structures such as dams and reservoirs that are present in almost all river basins considered except for Lena and the main tributaries of Mackenzie basins, which ultimately led to more uncertainties in the model simulation results. The inclusion of dams and reservoirs when designing the models would help mitigate some of the uncertainties observed in the simulated discharges.

Some models underperformed in almost all the basins giving poor results. Their discharge dynamics and seasonality did not match closely with the observations. A possible reason for the poor result might be due to the biased climate data. GHMs also used coarser spatial resolution for their simulation, which made it hard to predict accurate results. Even a small overestimation of evaporation or underestimation of precipitation in the forcing may lead to a substantial underestimation of river discharge in the basin. The low runoff coefficient (i.e., the fraction of precipitation that reaches the basin outlet) may lead to significant overestimation of river discharge and vice versa.

The Global Hydrological Models under consideration were not calibrated except for WaterGAP2. Due to the absence of calibration process, some of the model parameters that were being used for simulations were not finetuned, which then resulted in a slight variation in the river discharge dynamics along with systematic and dynamics behaviours. Moreover, because of the fine-tuning done in WaterGAP2, the model performed better than most of the other models in all basins. Even though most of the models were not calibrated, it did not mean that the results of their simulations were all poor. Some of the uncalibrated models, such as MATSIRO and MPI-HM performed well while analysing river discharges and other models were not that far behind in producing quality simulation results. JULES-W1 and ORCHIDEE were the ones who did not perform well in most basins as they dealt with managing both energy and water balances.

6.5 Model Evaluation Results (Best and Poor Models)

The models were evaluated and rated on their performance using 3 efficiency criteria and their threshold values as listed below in Table 10, and then aggregated indices were estimated for every model and basin using rating scores of 1 (good performance), 0.5 (weak) and 0 (poor) for every criterion and gauging station. Figures 21, 22, 23, 24, 25, and 26 shown below, highlight the observed and simulated long-term average monthly discharge (seasonal dynamics) plots. Evaluation results for all gauging stations of river basins are shown. In each plot, the 2 best performing models and 2 poor performing models are displayed. The grey spread indicates the maximum and minimum range of simulated discharge from 4 climate forcing datasets.

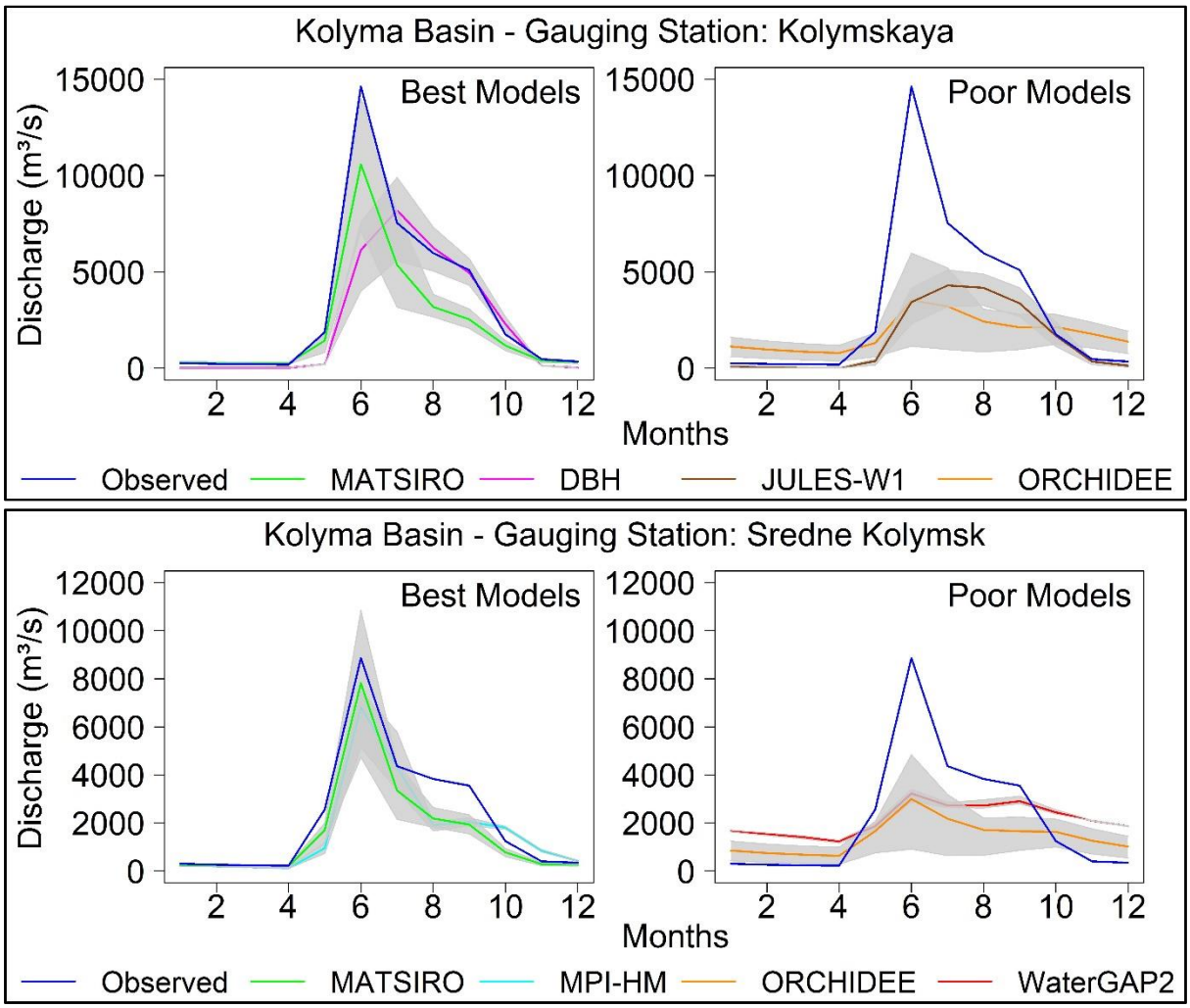


Figure 21: Best and Poor Performing Models in Kolymskaya and Sredne Kolymsk, Kolyma River Basin

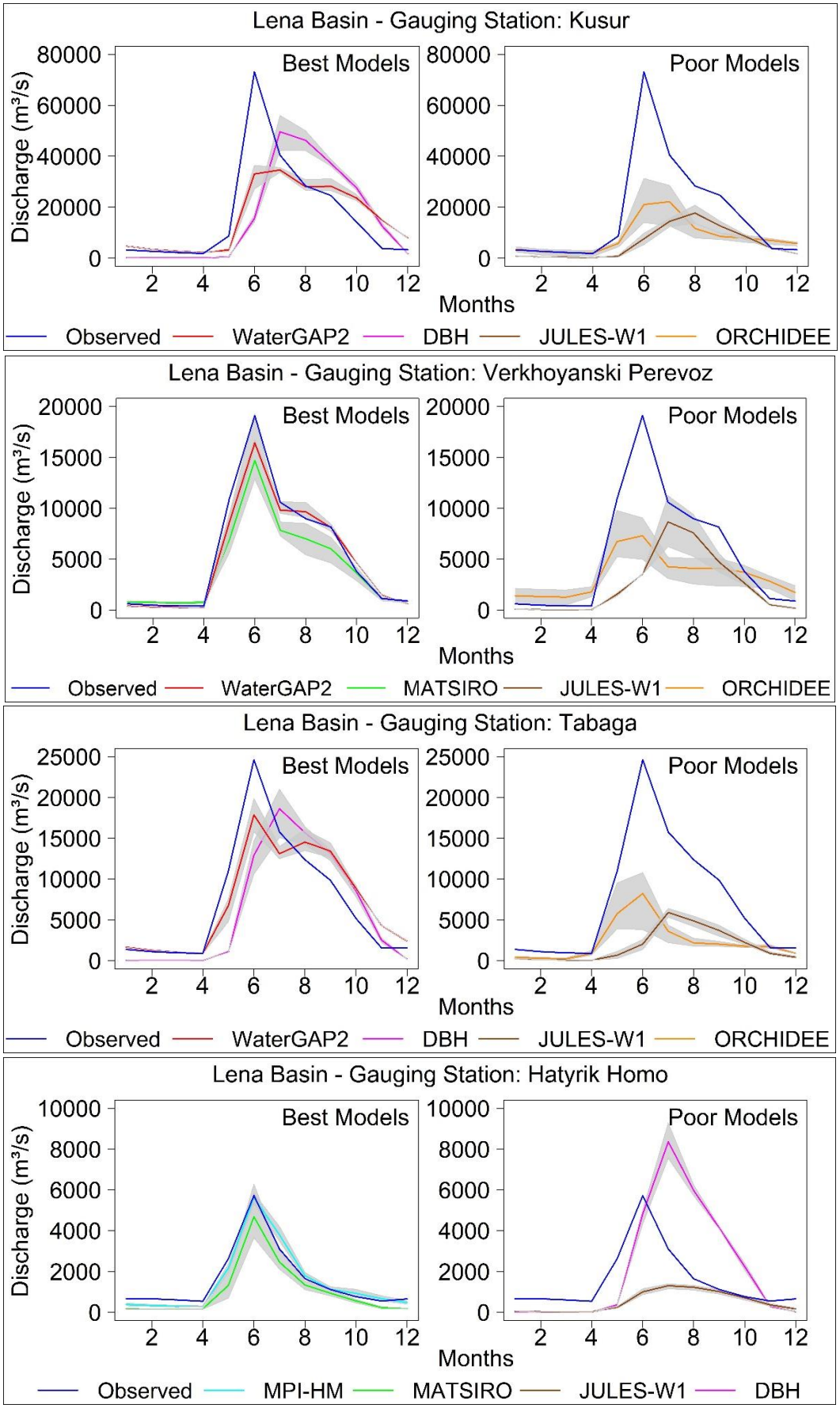


Figure 22: Best and Poor Performing Models in Kusur, Verkhoyanski Perevoz, Tabaga, and Hatyrik Homo, Lena River Basin

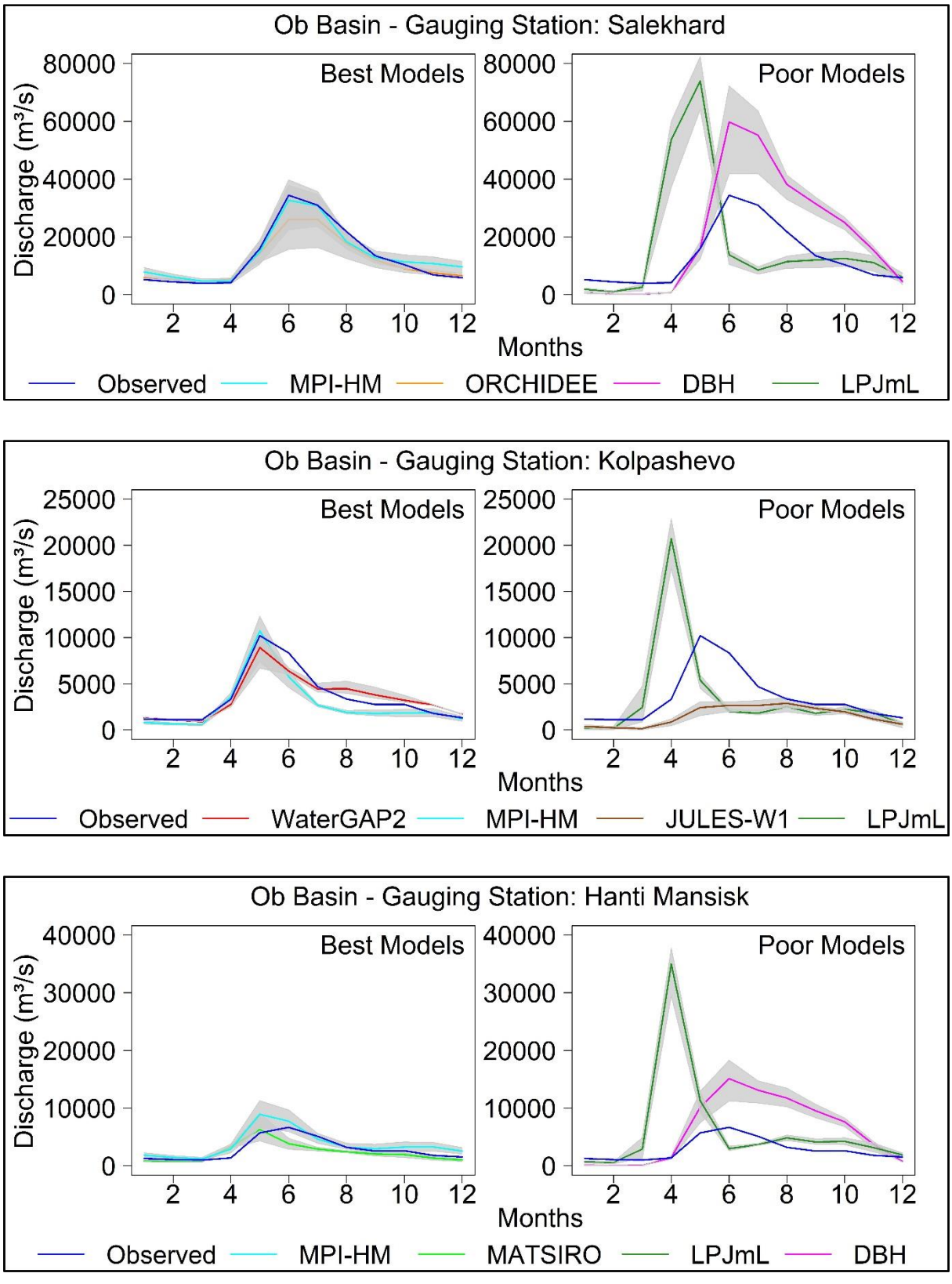
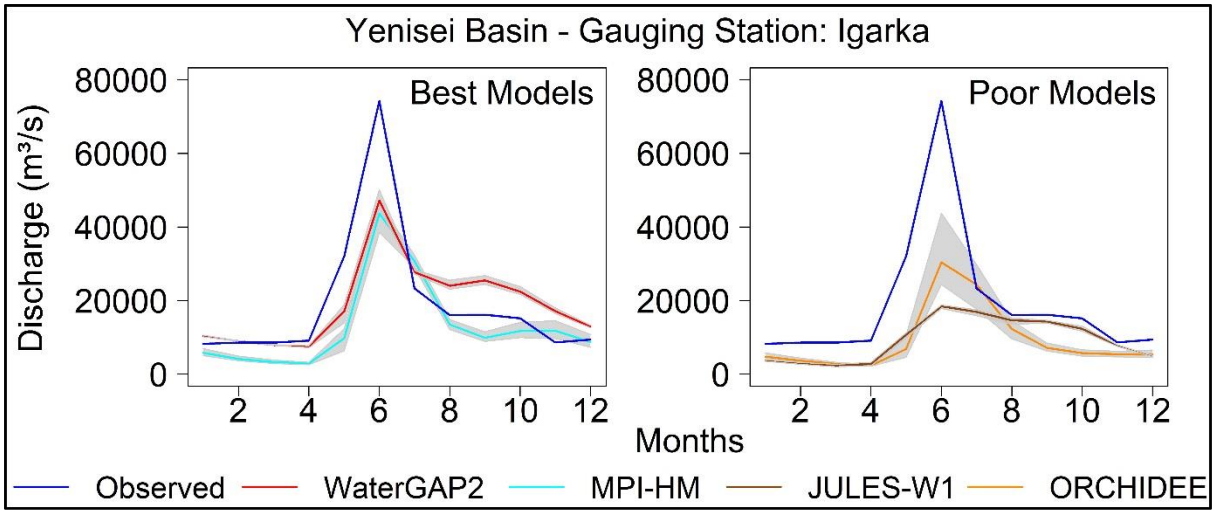
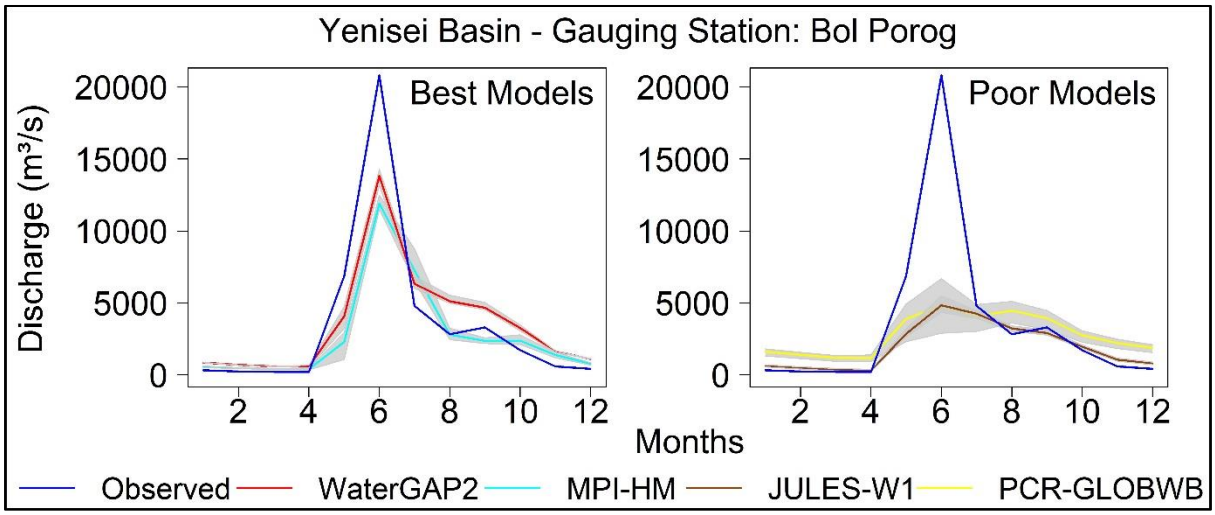


Figure 23: Best and Poor Performing Models in Salekhard, Kolpashevo and Hanti Mansisk, Ob River Basin





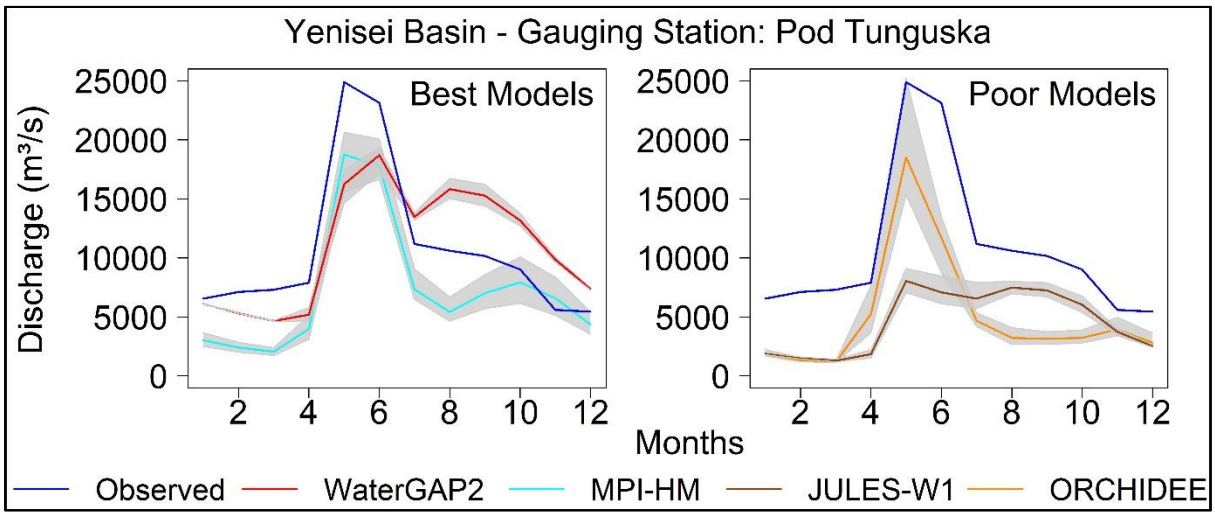
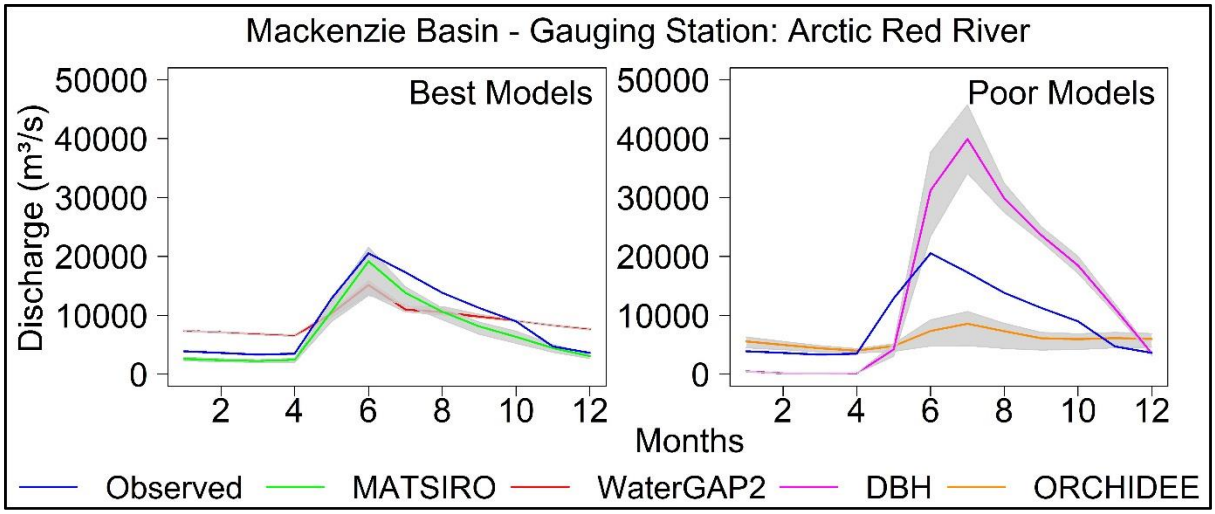
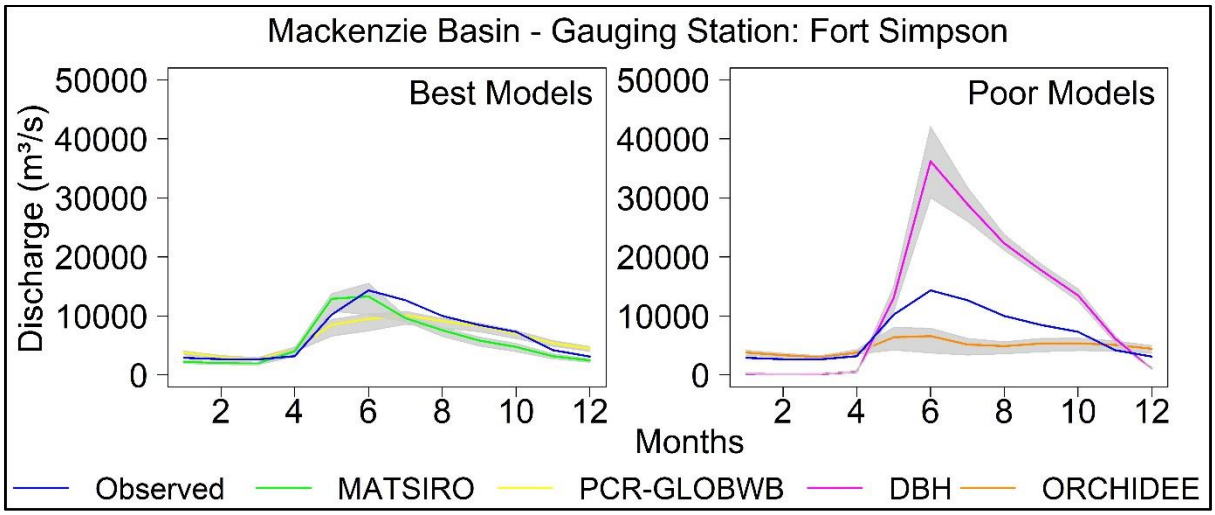


Figure 24: Best and Poor Performing Models in Igarka, Bol Porog and Pod Tunguska, Yenisei River Basin





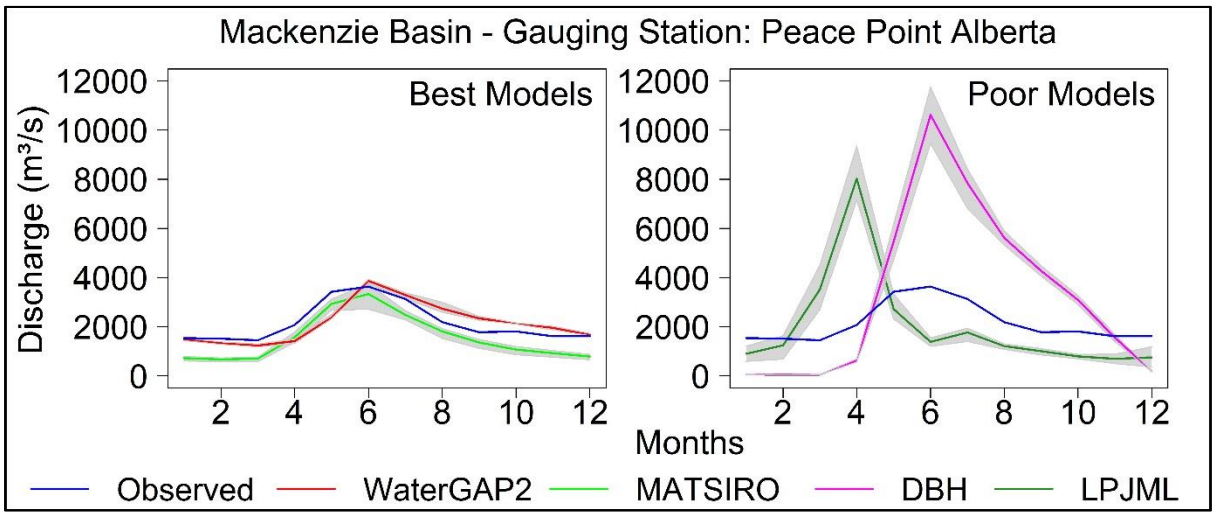
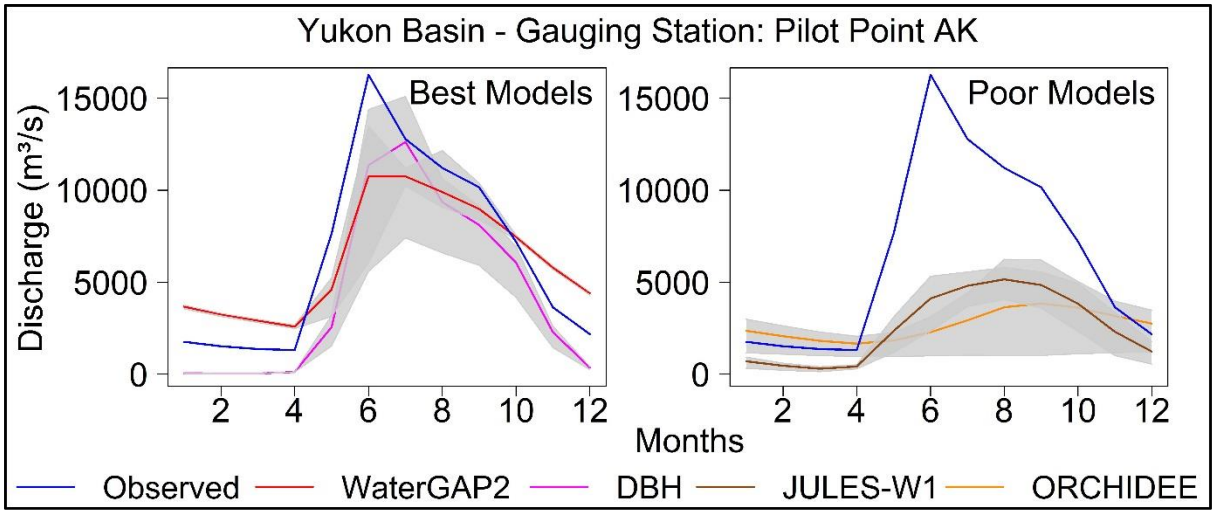
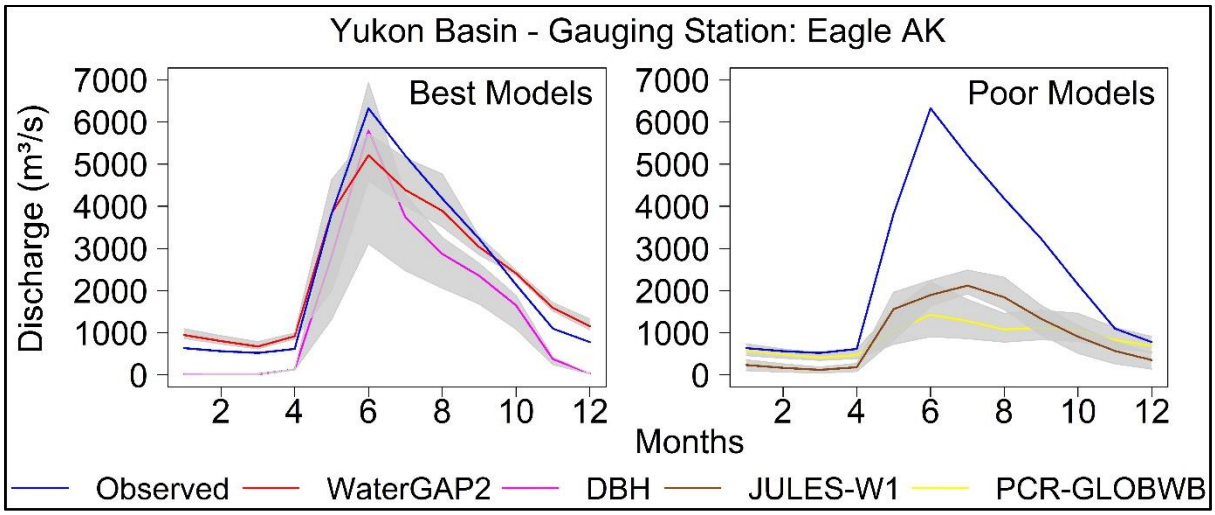


Figure 25: Best and Poor Performing Models in Arctic Red River, Fort Simpson and Peace Point Alberta, Mackenzie River Basin





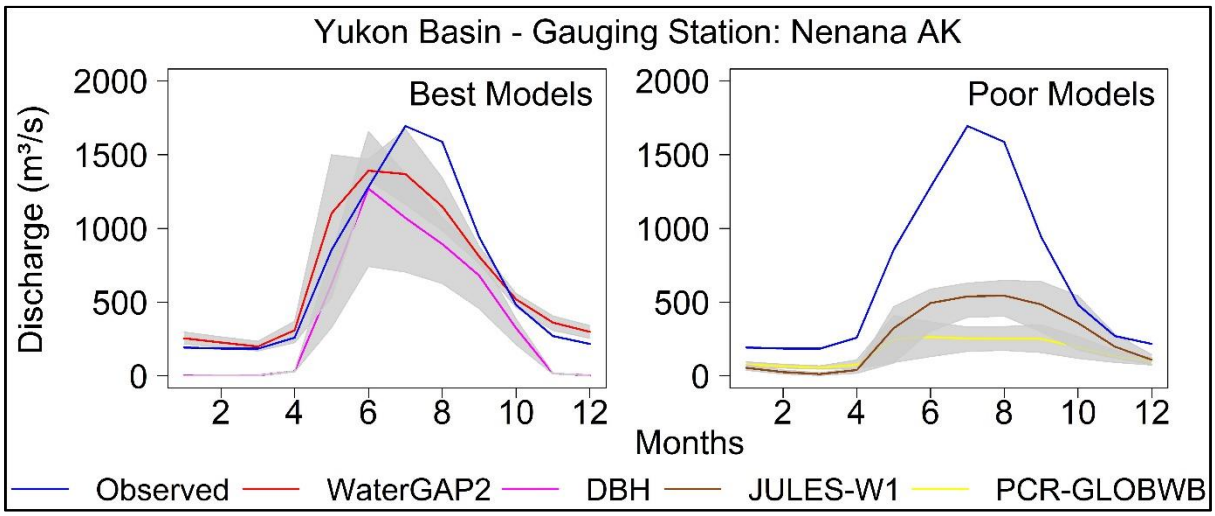


Figure 26: Best and Poor Performing Models in Pilot Point AK, Eagle AK and Nenana AK, Yukon River Basin

6.6 Model Performance Aggregated Index Results

The models were finally evaluated and rated using the model performance aggregated index. This performance aggregated index provided a clear picture of how the models behaved and performed to simulate discharges and capture the river dynamics of the basins accurately. The model performance results from this study are presented below in both tabulated format and bar graph plots. The model performance evaluation results were looked from the perspective of both model-wise and basin-wise. In the model-wise analysis results, each model was individually evaluated to get performance index for all basins under study (Table 20), and the tabulated results were also illustrated using the bar graphs (Figure 27). In the basin-wise analysis results, basins were evaluated individually based on the performance index from all models under study, and they were also illustrated using the bar graphs (Figure 28).

Table 20: Model Performance Aggregated Index Summary for each basin

Model Performance Aggregated Index									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
Kolyma	38	80	58	75	64	81	22	52	28
Lena	88	53	35	63	48	60	20	37	3
Ob	61	11	55	83	36	63	53	14	26
Yenisei	81	42	45	67	28	52	23	39	7
Mackenzie	68	4	32	35	63	77	22	28	31
Yukon	93	72	43	7	2	16	8	25	0

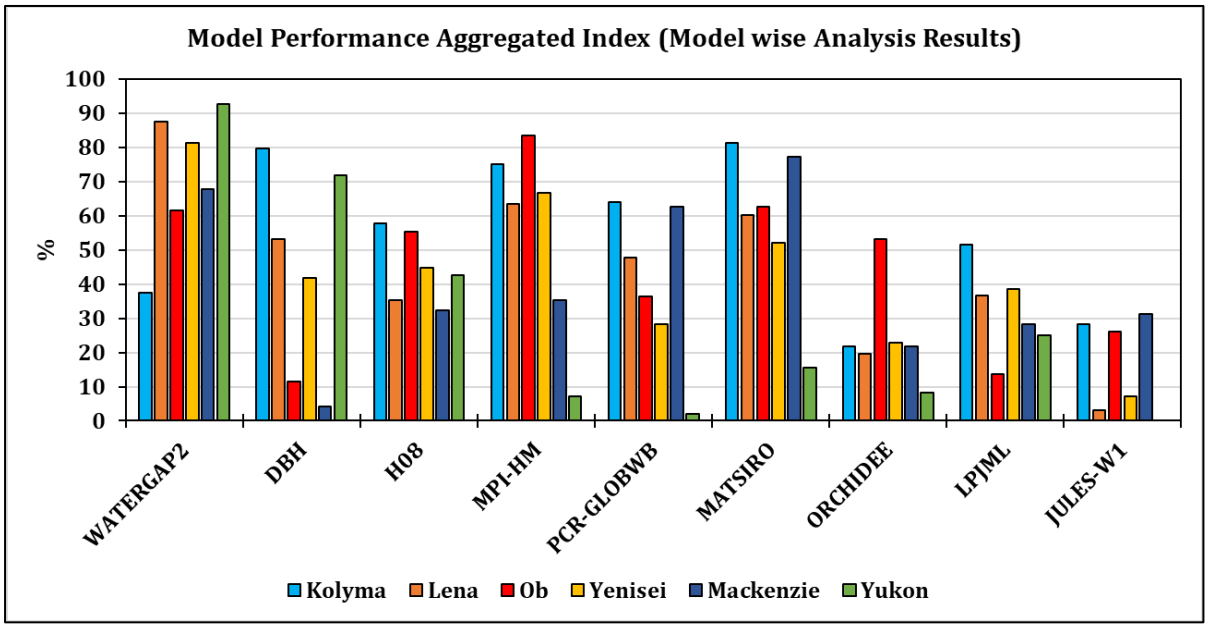


Figure 27: Model Performance Aggregated Index (Model-wise Analysis Results)

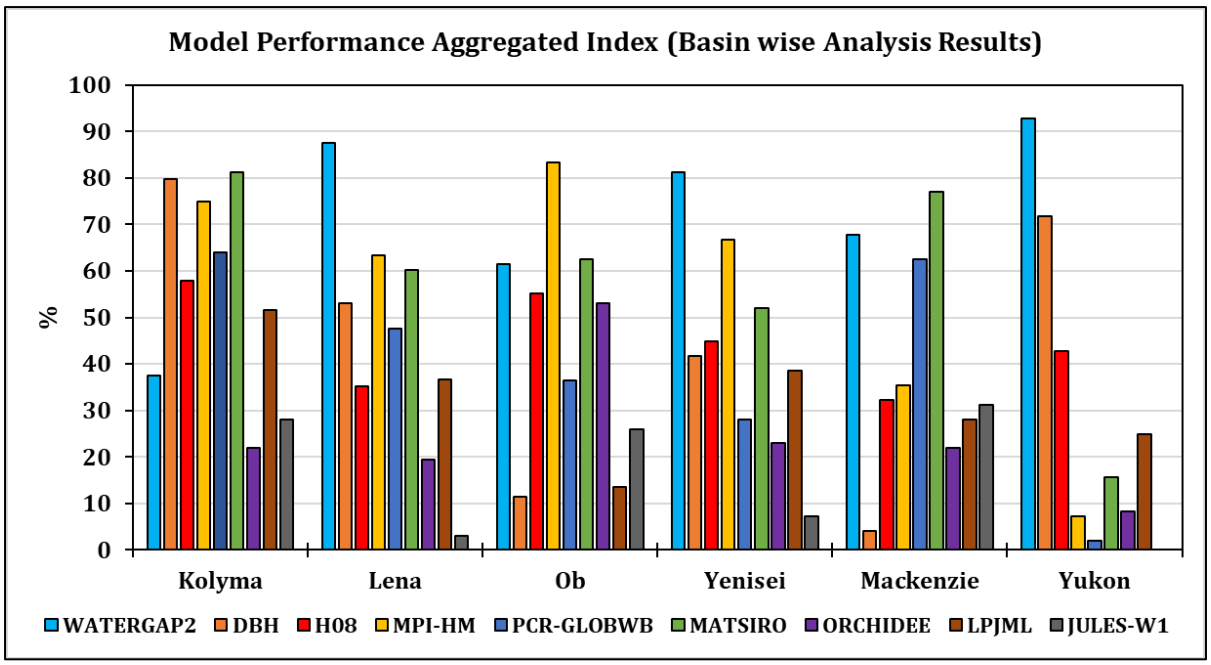


Figure 28: Model Performance Aggregated Index (Basin-wise Analysis Results)

From the bar graph plots presented above (Figures 27 and 28), it was deduced that WaterGAP2 model performed better in all basins except Kolyma basin. The performance index of WaterGAP2 in all these basins was above 60%, which was considered a very good result. This good performance was somehow expected from WaterGAP2 as it was the only model that was calibrated. Performance of MATSIRO and MPI-HM was quite good for seasonal dynamics as compared to other models, with the aggregated index exceeding 60% in 4 basins from 6.

In Mackenzie basin, MATSIRO outperformed WaterGAP2 in terms of model performance. Similarly, in Ob basin, MPI-HM was the best model when looking at its performance to simulate correct and near accurate river discharge in the basins. On the other hand, few models which were considered as Land Surface Models like ORCHIDEE and LPJML did not perform well in almost all the basins. Their performance was poor in most basins as they failed to simulate the river discharges accurately. The performance index was mostly below 40% for these two models in all basins except for a few exceptions.

For instance, if the threshold limit for model performance aggregated index was kept as 50% to be deemed as 'acceptable model', then from Figure 28, we can say that 6 models in Kolyma basin, 4 models in Lena basin, 5 models in Ob basin, 3 models in Yenisei basin,

3 models in Mackenzie basin and 2 models in Yukon basin were accepted as well-performing models. Many models tried to simulate river discharges as accurately as possible in Kolyma and Ob basins, whereas very few models simulated acceptable discharges in Yukon, Mackenzie, and Yenisei basins.

Similarly, it was deduced from Figure 27 that WaterGAP2 and MATSIRO gave acceptable performance in 5 basins out of 6, MPI-HM performed well in 4 basins, DBH performed well in 3 basins, H08 and PCR-GLOBWB gave acceptable performance in 2 basins, whereas ORCHIDEE and LPJML only gave acceptable discharges in 1 basin and finally, JULES-W1 did not give acceptable results in any of the basins. From the above observations and deductions, it was easy to rate and rank the participating models from best performing to poor performing in each river basin.

6.7 Extreme Flows Analysis Results

The model performance evaluation methods for analyzing high flows, low flows and extreme flows are provided in the Data Processing chapter. Here, the results of the analysis are presented and discussed. Table 21 shows the model performance index values in high flows condition for all river basins, and in Figure 29, the tabulated values from Table 21 are presented as bar-graph plots to visualize the results quickly. Most of the models were able to simulate the high flows satisfactorily when compared with the observations in the Ob basin (6 models out of 9 with the aggregated index exceeding 50%, and 4 models exceeding 60%). In Lena basin, most of the models failed to perform well except for LPJML. In Kolyma basin, only PCR-GLOBWB and MATSIRO performance index crossed 50% indicating their simulation results were in acceptable levels.

Similarly, in Yenisei basin, MPI-HM and LPJML models performed satisfactorily as their performance index crossed 50% mark. JULES-W1 and ORCHIDEE models performed poorly in all river basins as seen from the performance index obtained by these models in all basins. In Yukon, three models (WaterGAP2, DBH, and MATSIRO) performed better than the rest of the models. The model performance index as presented from the table and figure below was quite diverse, and it was challenging to come up with one model which was considered as the best one to simulate actual high flows condition in all basins.

Table 21: Model Performance Index in high flows condition for all river basins

Model Performance Index in High Flows Condition									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
Kolyma	0	38	13	38	75	50	0	38	0
Lena	0	38	0	0	0	13	0	88	0
Ob	50	0	50	88	88	63	88	0	38
Yenisei	25	38	13	50	0	13	0	100	0
Mackenzie	50	0	100	50	0	88	0	0	13
Yukon	75	75	25	0	0	63	0	38	0

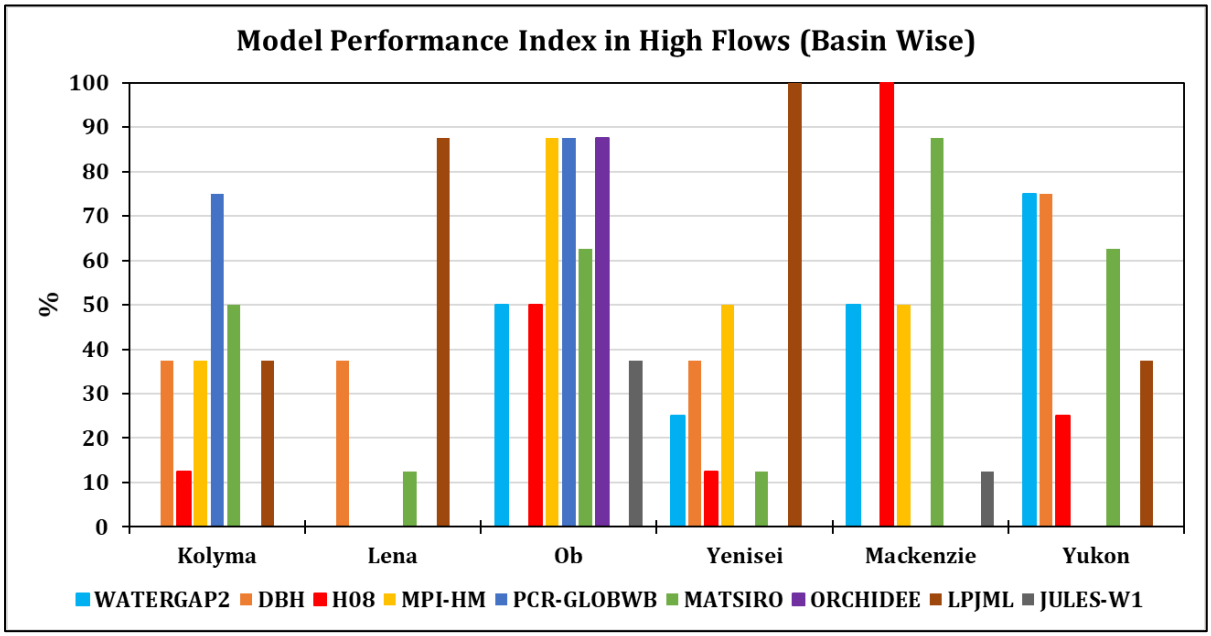


Figure 29: Basin-wise Model Performance Index in high flows

Table 22 shows the model performance index values in low flows condition for all river basins, and in Figure 30, the tabulated values from Table 22 are presented as bar-graph plots to visualize the results quickly. Five models (H08, MPI-HM, PCR-GLOBWB, MATSIRO, and ORCHIDEE) performed quite well in Ob basin as the model performance index crossed 70% for all the five models. Four models were able to simulate the low flows to some extent as the aggregated index exceeded 50% level when compared with the observations in Yenisei and Mackenzie river basins. In Yukon basin, most of the models failed to perform well except for ORCHIDEE, but the index value for ORCHIDEE was in the border of the threshold limit. In Kolyma basin, not even a single model was able to simulate low flows as all the models obtained 0% in their performance index.

In Lena basin, three models (MPI-HM, MATSIRO, and ORCHIDEE) performed satisfactory as their performance index crossed 50% mark. DBH and LPJML performed poorly in all river basins as seen from the performance index obtained by these models in all basins.

Their performance index was 0% in all basins. The model performance index as observed from the table and figure below was also quite diverse as observed in the high flow conditions and it was thus, difficult to come up with a model which was considered as the best one to simulate actual low flows condition in all basins.

Table 22: Model Performance Index in low flows condition for all river basins

Model Performance Index in Low Flows Condition									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
Kolyma	0	0	0	0	0	0	0	0	0
Lena	38	0	38	75	0	100	63	0	0
Ob	0	0	75	88	88	100	100	0	0
Yenisei	50	0	50	25	100	63	0	0	0
Mackenzie	0	0	50	63	0	88	13	0	63
Yukon	0	0	38	13	0	0	50	0	0

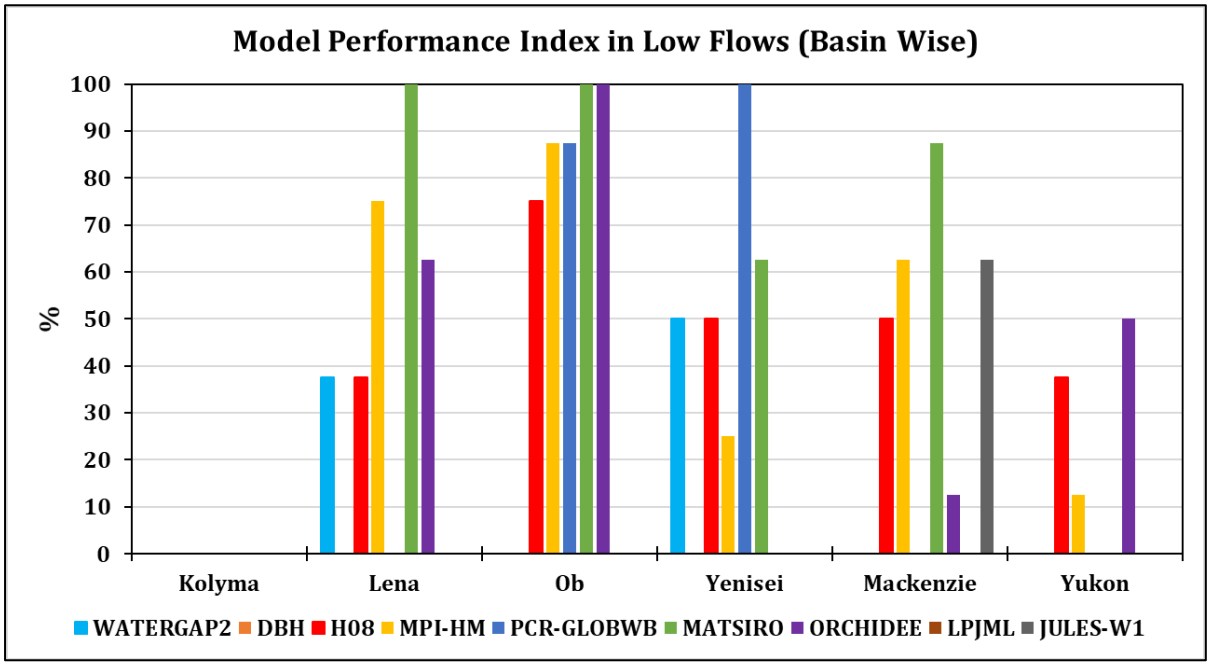


Figure 30: Basin-wise Model Performance Index in low flows

It was difficult to come up with the best model that could simulate the high and low flows accurately in each river basin. In this case, the rating scores of all models from the high flows and low flows were added together to get a combined rating score to represent the extreme flows. Then, the total rating score was converted into model performance index values in extreme flows condition for all river basins as presented in Table 23. The model performance index values for the extreme flows were then presented as the bar graphs in Figure 31. When combining the scores for the extreme flows, the models' performance in the basins did not show significant improvements as anticipated. Ob basin was the one which had 5 models performing well in this scenario with model performance index

crossing 60%. Mackenzie basin had only 3 models performing satisfactorily to predict the dynamics of the extreme flows. The rest of the basins did not have any good performing models whose performance index crossed 50%. All the models in the rest of the basins were performing below average except for a few cases like MATSIRO's performance in Lena basin, and PCR-GLOBWB and LPJML performance in Yenisei basin with exactly 50%.

Table 23: Model Performance Index in extreme flows condition for all river basins

Model Performance Index in Extreme Flows Condition									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
Kolyma	0	19	6	19	38	25	0	19	0
Lena	19	19	19	38	0	56	31	44	0
Ob	25	0	63	88	88	81	94	0	19
Yenisei	38	19	31	38	50	38	0	50	0
Mackenzie	25	0	75	56	0	88	6	0	38
Yukon	38	38	31	6	0	31	25	19	0

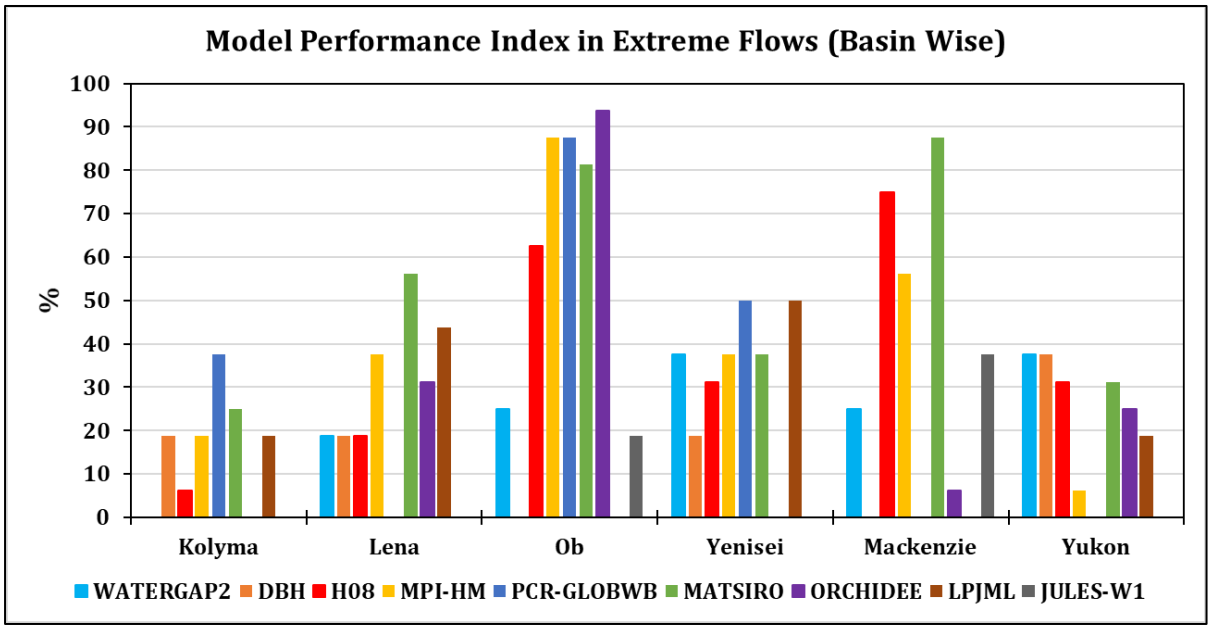


Figure 31: Basin-wise Model Performance Index in extreme flows

Initially, the model performance was evaluated separately for high flows, low flows, and extreme flows conditions basin-wise. In the next step, the models were evaluated for their overall performance by combining all the results of all river basins in respect to the extreme flows. In this case, the scores of a particular model were aggregated from all the basins and converted into the model performance aggregated index. The obtained total score from all 6 basins including the results from 4 climate forcing datasets simulations for a given model was divided by the highest possible score achievable by that model, which was 24 (6 basins x 4 climate forcing datasets x 1 efficiency criteria) and then, the

ratio was converted into the aggregated index as percentage. Figure 32 illustrate the model performance aggregated index obtained by applying the method as mentioned earlier in extreme flows conditions, which include both high and low flows. The performance aggregated index considered all river basins together as one single entity which is represented here as the Pan-Arctic region.

MATSIRO performed relatively better than the rest of the models, but its performance was only satisfactory with the index hovering around 50% in all flow conditions. MPI-HM performed slightly better than the other models, but its performance index was also below the satisfactory mark of 50%. Other models had a performance index below 40%, which did not indicate satisfactory performance in the overall sense for accurately simulating the extreme flows.

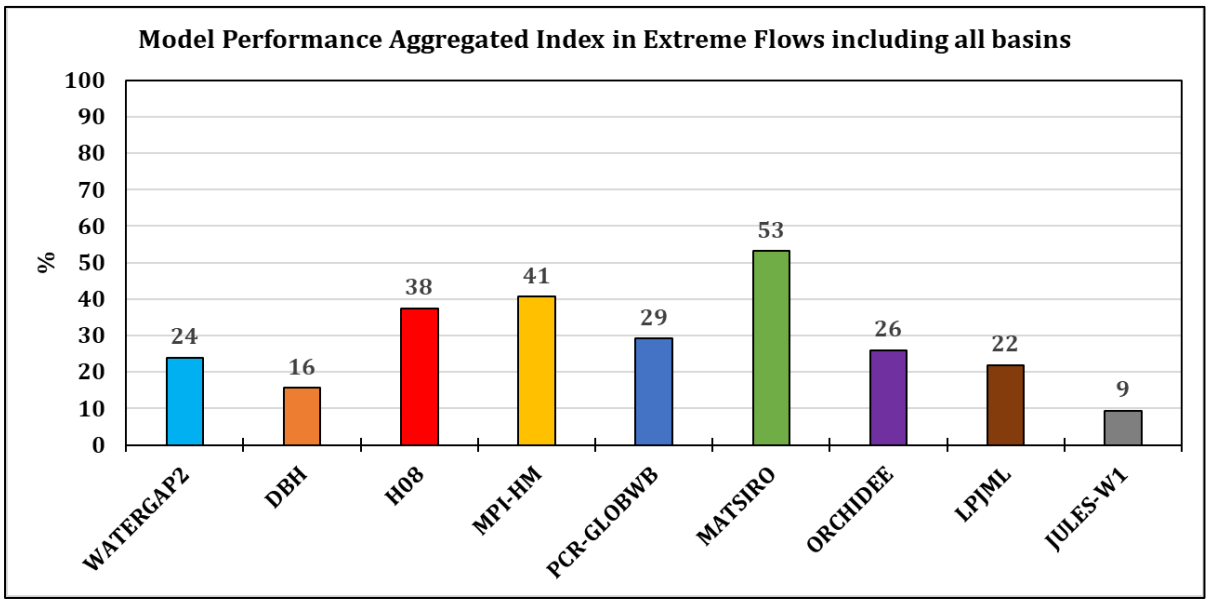


Figure 32: Model Performance Aggregated Index in extreme flows including all basins

The analysis and evaluation results showed a rather poor representation of high and low flows. The models should be improved to better simulate extreme conditions in the Arctic river basins.

7. SNOW WATER EQUIVALENT DATA ANALYSIS: RESULTS AND DISCUSSION

Snow is one of the important hydrological processes in the Pan-Arctic region. It plays an important role in the vertical water balance as it is directly linked with other various processes and can directly influence the river discharge dynamic. Snow Water Equivalent (SWE) is a standard snowpack measurement. It is the amount of water contained within the snowpack. It can be thought of as the depth of water that would theoretically result if the entire snowpack melted instantaneously. Snow water equivalent data were processed and analysed similarly as with discharge data. The methods used for processing the simulated snow water equivalent time series data from the NetCDF files were similar like what was done in discharge time series analysis. The details about processing the snow data have already been covered in the chapter "Data Processing" under "Snow Water Equivalent Time series Analysis and Visualization." Due to time constraints, this study only assessed the models subjectively using visual comparison between observed and simulated long-term average monthly snow water equivalent data. The source of both simulated and observed datasets has already been discussed above in the "Project Description" chapter.

7.1 Model Validation Runs

The main aim of this study was to evaluate the performance of different global hydrological models in simulating snow water equivalent and validate these simulation results against the observed data. The model validation was an essential step to assess the quality of the simulated data here. In this case, the snow water equivalent time series data were assessed subjectively and the performance of each model at each basin were evaluated. The models produced simulated time series data, which are then validated using the observation data collected using the remote sensing tools from GlobSnow Project. The R scripts for snow water equivalent time series analysis and visualization of both datasets are provided in the Appendix.

7.1.2 Comparison of Simulated and Observed Data

The simulated and observed long-term average monthly snow water equivalent values were compared after necessary data processing and manipulation of the monthly time series data. The process of converting monthly series into long-term average monthly series has already been discussed in the "Data Processing" chapter. For this analysis, a mean value was obtained in a given instant of time from four climate forcing datasets used in each model simulation. Six hydrological models (DBH, MPI-HM, LPJML, PCR-GLOBWB, MATSIRO, and WaterGAP2) were used. Then, the simulated mean of long-term average snow water equivalent for each model was compared with the observed data that could be seen from the plots presented in the figures below. The snow dynamics and seasonality behaviour of the values were discussed using visual inspection of the plots.

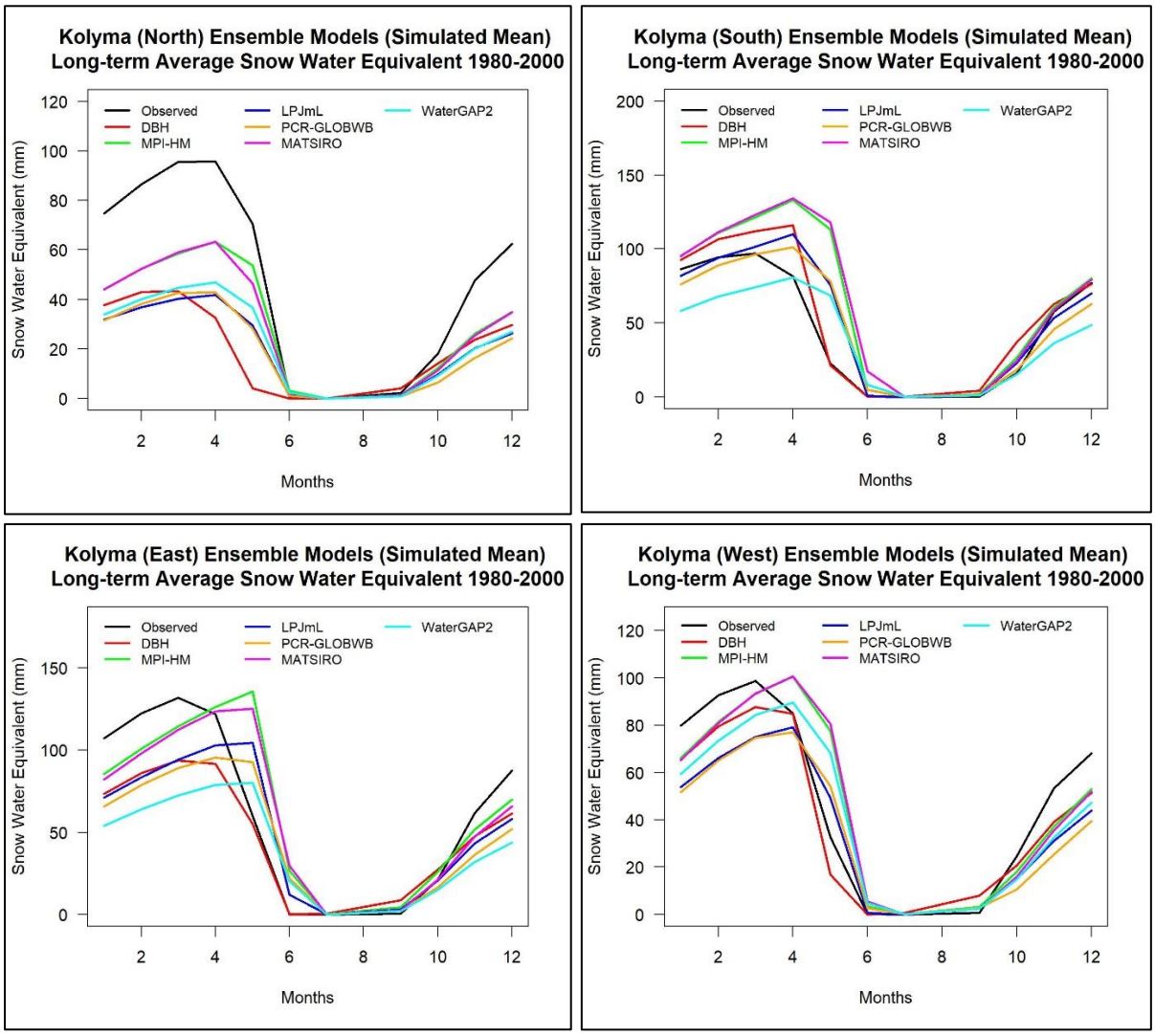


Figure 33: Comparison between simulated mean and observed long-term average monthly snow water equivalent at different locations in the Kolyma basin

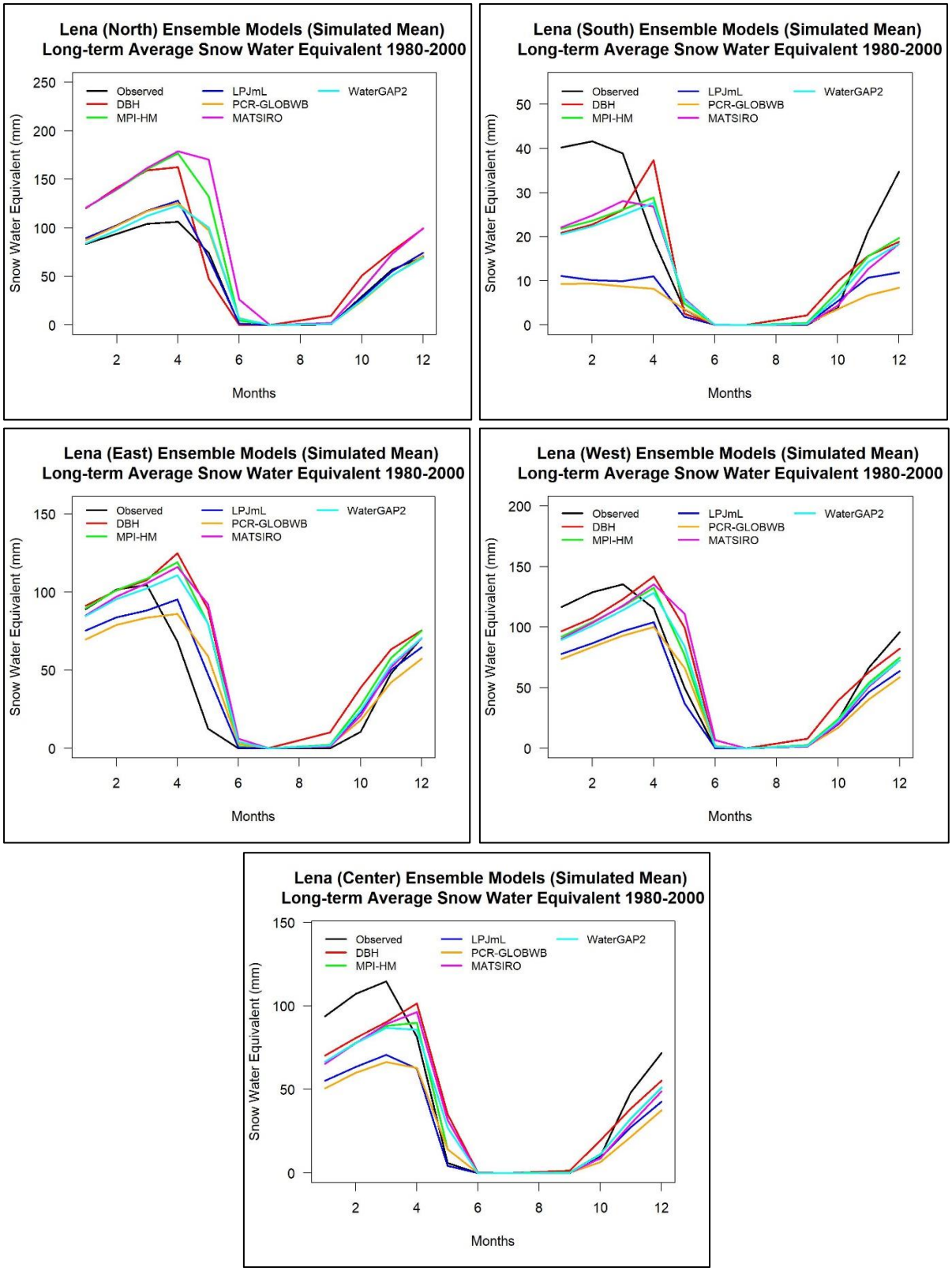


Figure 34: Comparison between simulated mean and observed long-term average monthly snow water equivalent at different locations in the Lena basin

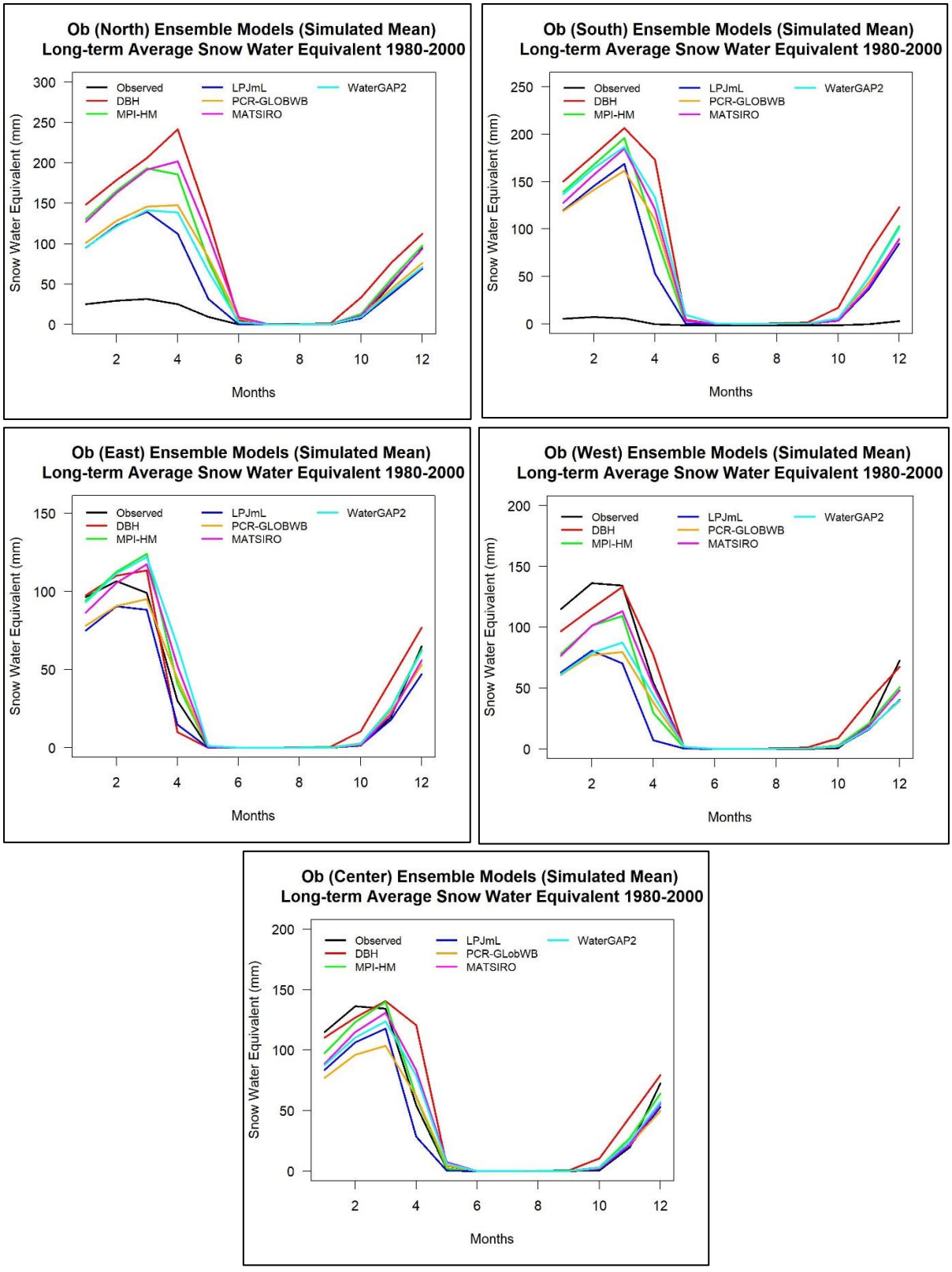


Figure 35: Comparison between simulated mean and observed long-term average monthly snow water equivalent at different locations in the Ob basin

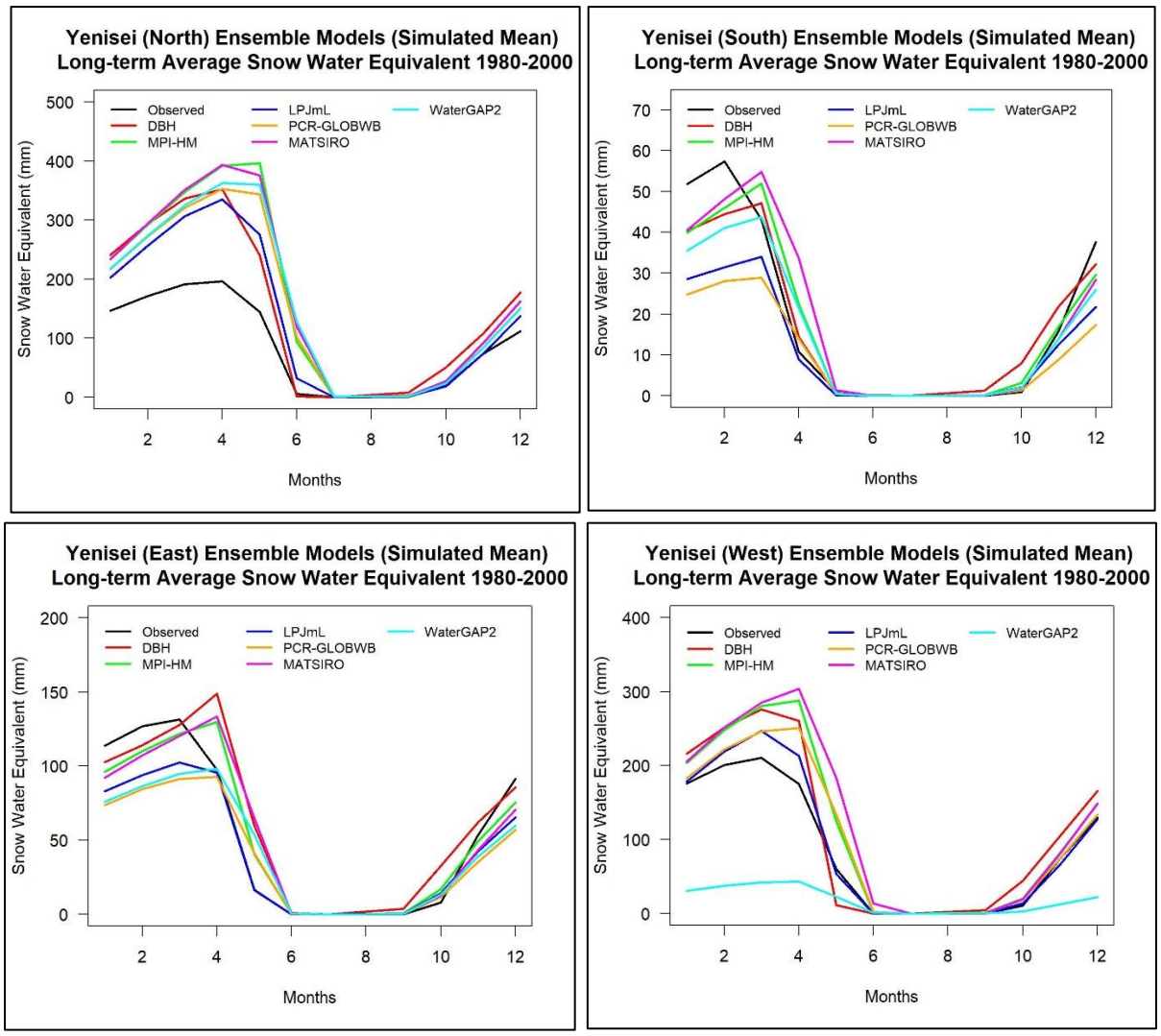
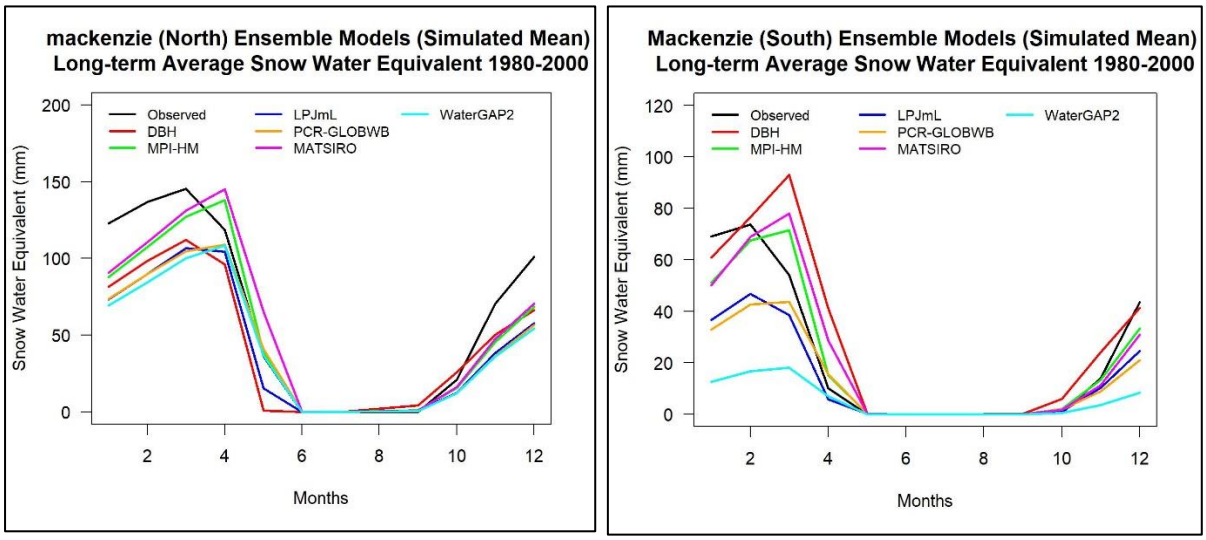


Figure 36: Comparison between simulated mean and observed long-term average monthly snow water equivalent at different locations in the Yenisei basin



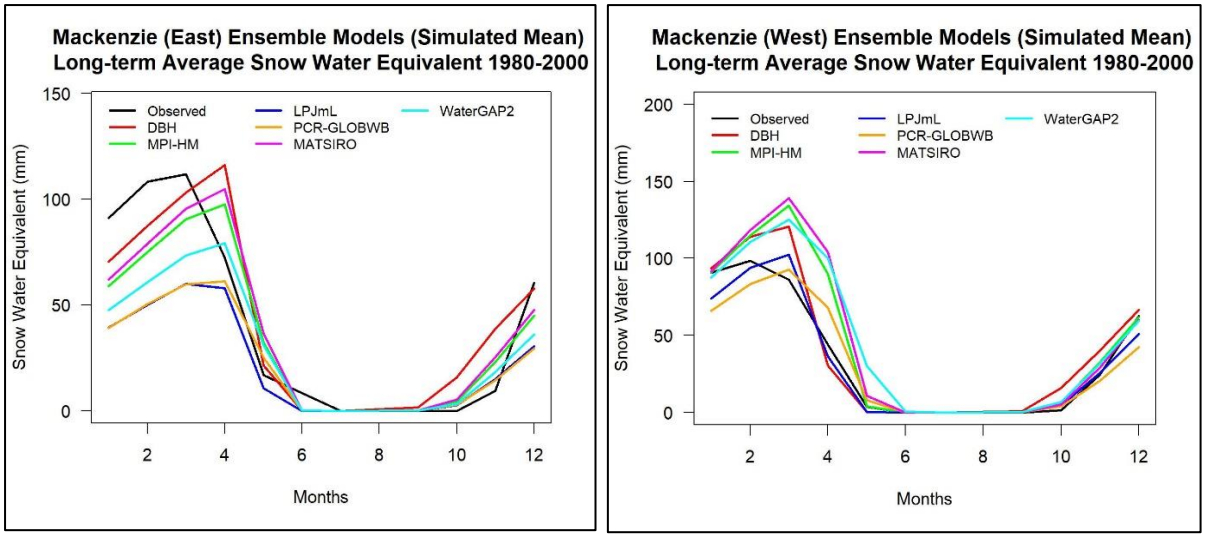
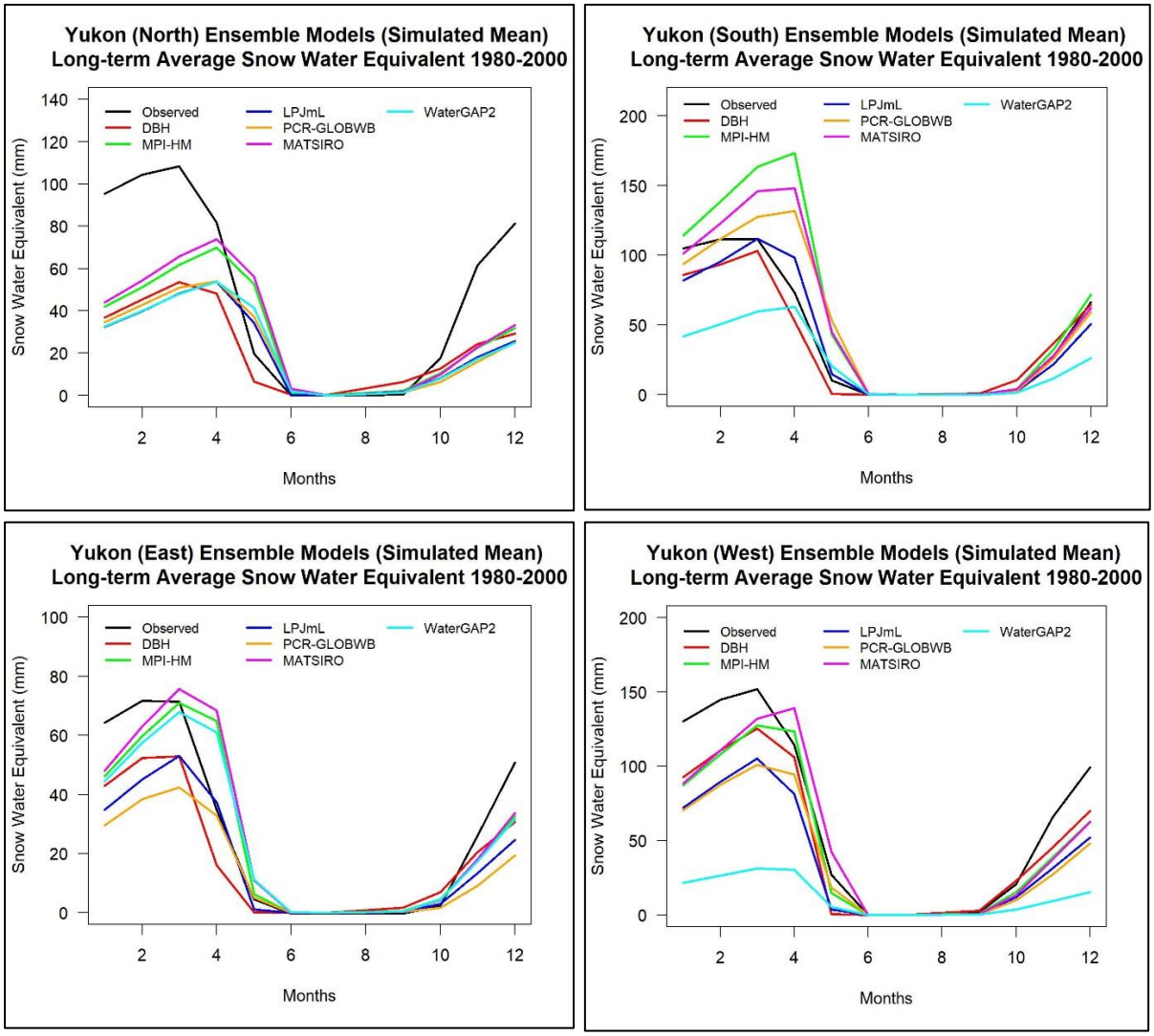


Figure 37: Comparison between simulated mean and observed long-term average monthly snow water equivalent at different locations in the Mackenzie basin



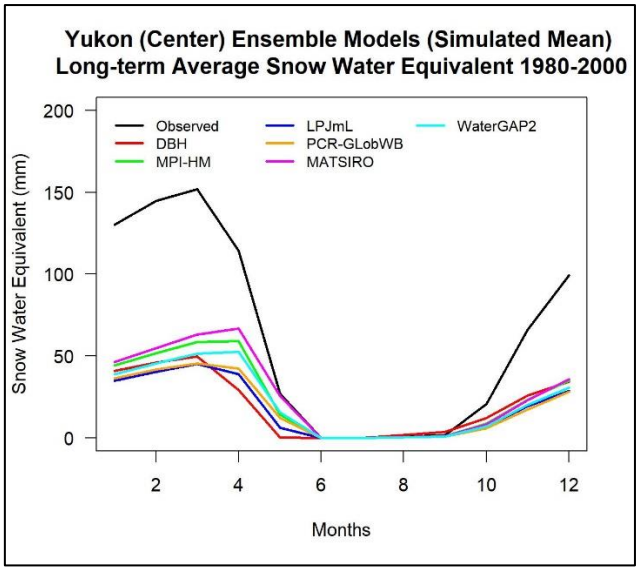


Figure 38: Comparison between simulated mean and observed long-term average monthly snow water equivalent at different locations in the Yukon basin

The SWE values of 0 mm denote snow-free areas, 0.001 mm denote areas with melting snow and, >0.001 mm denote areas with full snow cover [Luojus et al. 2010]. The areas that are identified as wet snow or have no SWE retrieval but are identified as snow-covered with the time-series melt detection approach are denoted by SWE value of 0.001 mm. The areas that are determined as snow-free or melted by the melt detection approach are denoted by SWE value of 0 mm. All the other areas show a retrieved SWE value (that is, in all cases greater than 0.001 mm) [Luojus et al. 2010].

All the above-presented plots (Figure 33, 34, 35, 36, 37 and 38) were easy to visualize with necessary details provided clearly for inspecting the dynamics and seasonality of both simulated and observed snow water equivalent values. There were underestimation or overestimation of the simulated results at different periods when compared against the observed values. It was observed from the figures that the snow water equivalent increased during winter months, decreased during spring months, and remained mostly at zero during the summer months. The snowpack is abundant during the cold and freezing winter period in the region leading to increase in snow water equivalent, but when the temperature rises during the spring season, the snow/ice starts to melt and flows directly into rivers and streams that lead to decrease in snow water equivalent levels. The hot summer months with higher temperature will completely melt the

snowpack in the region leading to 0 values all over the basins. From the subjective assessment, it was noted that some models in some points reproduced the long-term mean seasonal dynamics of snowpack quite well, e.g. WaterGAP2 in the Lena North, DBH in the Ob West, and LPJML in the Yukon South but in many cases they showed notable over- or underestimation of snowpack. It is advisable to assess the snowpack time series objectively using model efficiency criteria as discussed in the discharge time series analysis. These criteria will provide more evidence to evaluate the models and help to support the conclusions drawn from the subjective assessment.

In Kolyma basin, the periodic pattern was predicted accurately in all locations of interest. However, in the northern part, simulated data showed underestimation throughout the year when compared against the observed data. In the south and west locations, the simulation results showed overestimated during the summer period. In the east, the simulated results showed underestimation during the winter period. In Lena North, 3 models showed overestimation and the other 3 models performed well. In Lena South, none of the models were able to reproduce seasonal dynamics. In Lena East, all models did not reproduce dynamics from March until November. In Lena West, all models showed underestimation in winter and early spring but most of them showed overestimation in April-May. In Lena Center, all models showed underestimation in the winter period. In Ob basin, the difference between simulated and observed values in north and south locations was very big. In Ob East, all models showed acceptable performance. In Ob West, all models showed underestimation in the winter-spring period, and in Ob Center, most of the models showed underestimation in the winter months.

In Yenisei North, all the models showed overestimation in all months. In Yenisei South and East, almost all the models showed underestimation in the winter period and overestimation in the spring period. In Yenisei West, most of the models showed overestimation in all months. In Mackenzie North and East, all the models showed underestimation in the winter and in the early spring period. In Mackenzie South, most of the models showed underestimation in the winter months. In Mackenzie West, most of the models showed overestimation in February-March period. In Yukon basin, underestimation in simulated values was observed in all locations during the winter

period. The simulated values in the center of the basin had significantly deviated from the observed values. WaterGAP2 simulated poor results in the south and west part of the basin.

7.2 Discussions

In most of the locations in the basins, the simulated values were underestimated when compared with the observed data during the winter period, and the simulated values were overestimated during the spring period. The range of bias between simulated and observed data could be due to several reasons such as uncertainties in model parameterization, usage of remote sensing product to obtain data that lack accuracy and reliability of the recorded values, coarser resolution of input data used while modelling and others. The errors in the input of precipitation may directly affect the snowpack simulation results as they are closely linked. As stated in Luojus et al. [2010], the mountainous areas of Northern Hemisphere are masked out from the SWE product, which produced biased simulated values as most of the Pan-Arctic region is covered by mountainous areas. Since mountainous areas of the region are concealed while taking measurements of the snowpack in the region, the observed data could be compromised, and the accuracy of the data would be questioned. Luojus et al. [2010] also stated that SWE estimates for wet snow areas (where retrieval using the radiometer data is not feasible) are determined from the weather station data using kriging interpolation. Such an interpolation method may be another source of uncertainty.

Lastly, when long-term average monthly discharge analysis results were compared with long-term average monthly snow water equivalent analysis results, it was clear that these two variables were linked closely with one another. During the spring period, the simulated discharge values were underestimated, whereas snow water equivalent values were overestimated in the basins. The higher levels of snow water equivalent simulated by the models during the spring period prevented these models to accurately simulate the peak river discharge in the basins by predicting lower discharge levels. The opposite phenomenon took place during the winter period. The lower levels of snowpack simulated by the models resulted in the higher levels of river discharge simulated during that same period.

8. CONCLUSIONS AND RECOMMENDATIONS

This chapter includes the conclusions of the study. The areas that need further improvement are highlighted in the recommendation section, which also includes some future research works that could be continued from the findings of this study.

8.1 Conclusions

The main aim of this study was to evaluate the model performance using the output variables such as river discharge and snow water equivalent of the ISIMIP2a Global Hydrological Models (GHMs) for the Pan-Arctic river basins in the historical period 1971-2000. Therefore, this study focused on the model performance evaluation of 9 GHMs for 6 Pan-Arctic river basins, represented by 18 stream gauge stations using 3 efficiency criteria such as NSE, PBIAS and Percent Bias in Standard Deviation and visual comparison regarding river discharge. In case of snow water equivalent, the model performance evaluation of 6 GHMs for the same river basins was carried out using visual comparison only. The model performance evaluation for hydrological indicators such as monthly hydrographs, seasonal dynamics, high flows and low flows was performed using river discharge. In addition, the long-term average monthly snow water equivalent was used to evaluate model performance.

Global Hydrological Models, mostly uncalibrated, often showed bias in the long-term average seasonal river discharge when compared against observations, although they did in many cases reproduce the intra-annual variability well. Most of the models showed varied uncertainty ranges in the simulated results. The reasons behind the uncertainties could be due to biased climate input data along with coarser spatial resolution used in GHMs' simulation. Furthermore, model structure, its parametrization, and missing model calibration process could be responsible for large bias in GHMs' output. Biased climate input data such as precipitation, temperature, and other inputs may have caused erratic systematic (underestimation and overestimation) behaviour in discharge output of the river basins, and missing calibration of the 8 GHMs may have led to higher bias in the simulated discharge.

The models were evaluated and rated on their performance using efficiency criteria and their threshold values, and then model performance aggregated indices were estimated for every model and basin using rating scores of 1 (good performance), 0.5 (weak) and 0 (poor) for every criterion and gauging stations. As a result, WaterGAP2, MATSIRO, and MPI-HM performed better in most of the river basins than the other models. Other GHMs had considerable difficulties in representing the Pan-Arctic hydrological processes realistically.

Sredne Kolymsk (Kolyma) and Igarka (Yenisei), the gauging stations considered for linear trend analysis of time series data, showed a significant linear trend in observed and in simulated discharge time series from all the participating models. 4 models in Ob basin and 3 models in Yukon basin scored more than 60% in high flows condition. 5 models in Ob basin, 3 models in Mackenzie basin and 3 models in Lena basin scored more than 60% in low flows condition. The simulations showed a poor representations of high and low flows by other models in these basins, and by most of the models in the remaining river basins. MATSIRO performed relatively better than the other models when overall performance for the extreme flows condition was evaluated, but the performance aggregated index for MATSIRO was around 50%, which was considered as a satisfactory performance.

Lastly, regarding snow water equivalent analysis, some models in some points reproduced the long-term mean seasonal dynamics of snowpack quite well, e.g. WaterGAP2 in the Lena North, DBH in the Ob West, LPJML in the Yukon South, but in many cases, they showed notable over- or underestimation of snowpack. Most of the models in most of the locations underestimated snow water equivalent values during the winter period and overestimated values during the spring period. The reasons behind biased results could be similar to what was stated regarding discharge analysis previously. Looking at the seasonal dynamics plots of discharge and snow water equivalent; it could be said that these two variables were closely linked with each other, and these variables directly influenced the river dynamics in the Pan-Arctic region.

8.2 Recommendations

It is recommended to use good performing models based on the evaluation results from the study and use these models for climate impact assessment and projections in the future run. In addition, weighing coefficients could be applied to the model outputs depending on the models' performance for impact assessment. These models can also be used to assess climate change impact under different Representative Concentration Pathway (RCP), namely RCP2.6, RCP4.5, RCP6, and RCP8.5 scenarios.

In case of snow water equivalent analysis, objective assessment using numerical criteria is recommended for validating the models and evaluating their performance in detail. This assessment method will provide backup evidence to the findings from the visual comparison (subjective assessment) of the data. As snow is an essential hydrological process in the Arctic region, an in-depth study of snow water equivalent data will help better understand the water balance in the region.

It is recommended to put more efforts towards improving primarily the global hydrological models themselves and, secondarily the quality of climate input parameters and snowpack water measurements along with improving technical approaches to the model calibration process. Model simulation results could be improved through better quality input and evaluation datasets, the inclusion of missing physical processes, and better representation of existing processes in the models [Zaherpour et al. 2018]. More attention needs to be given by global-scale hydrological modelling community to improve the model's ability to predict the magnitude and timing of seasonal cycles.

9. REFERENCES

Abramowitz, G., Leuning, R., Clark, M. & Pitman, A. (2008). Evaluating the Performance of Land Surface Models. *J. Climate,* 21, 5468–5481. https://doi.org/10.1175/2008JCLI2378.1

Adhikari, R., & Agrawal, R.K. (2013). An Introductory Study on Time Series Modeling and Forecasting. *ArXiv, abs/1302*.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T. & Siebert S. (2003). Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrological Sciences*, 48(3), 317-337.

Arnell, N.W. & Gosling, S.N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 134(3), 387–401. doi:10.1007/s10584-014-1084-5

ASCE (1993). Criteria for evaluation of watershed models. *J. Irrigation Drainage Eng,* 119(3), 429-442.

Barnett, T.P., Adam, J.C. & Lettenmeier, D.P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438, 303–309.

Beck, H.E., et al. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52 (5), 3599–3622. doi:10.1002/2015WR018247

Bennett, K. E., & Prowse, T. D. (2010). Northern Hemisphere geography of ice-covered rivers. *Hydrol. Processes*, 24(2), 235– 240.

Best et al. (2011). The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4, 677-699.

Beven, K.J. & Smith, P.J. (2015). Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *ASCE Journal of Hydrologic Engineering*. doi:10.1061/[ASCE]HE.1943-5584.0000991

Bowling, L. C., Lettenmaier, D. P., & Matheussen, B. V. (2000). *Hydroclimatology of the Arctic drainage basin*, in The Freshwater Budget of the Arctic Ocean, edited by E. L. Lewis et al., pp. 57– 90, Springer, Dordrecht, Netherlands.

Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mård, J., Mernild, S. H., Prowse, T. D., Semenova, O., Stuefer, S. & Woo, M.K. (2016). Arctic terrestrial hydrology: A synthesis of processes, regional effects and research challenges. *J. Geophys. Res. Biogeosci.*, 121, 621–649.

Bring, A., Shiklomanov, A. & Lammers, R. B. (2017). Pan-Arctic river discharge: Prioritizing monitoring of future climate change hot spots. *Earth's Future*, 5, 72–92. doi:10.1002/2016EF000434

Bring, A. & Destouni, G. (2009). Hydrological and hydrochemical observation status in the pan-Arctic drainage basin. *Polar Research*, 28, 327-338. doi:10.1111/j.1751-8369.2009.00126.x

Brooks, R.N., Prowse, T.D., & O'Connell, I.J. (2013). Quantifying Northern Hemisphere freshwater ice, *Geophys. Res. Lett.*, 40, 1128– 1131. doi:10.1002/grl.50238

Bryhn, A. C., & Dimberg, P. H. (2011). An operational definition of a statistically meaningful trend. *PloS one*, 6(4), e19241. doi: 10.1371/journal.pone.0019241

Carmack, E. C., et al. (2016). Fresh water and its role in the Arctic Marine system: Sources, disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans, *J. Geophys. Res. Biogeosci.*, 121, 675–717. doi:10.1002/2015JG003140

Dankers, R., et al. (2014). First look at changes in flood hazard in the inter-sectoral impact model intercomparison project ensemble. *Proceedings of the National Academy of Sciences*, 111, 3257–3261. doi:10.1073/pnas.1302078110

Degtyarev V. (2016). *Lena River Basin (Russia)*. In: Finlayson C., Milton G., Prentice R., Davidson N. (eds) The Wetland Book. Springer, Dordrecht.

Dirmeyer, P.A., Gao, X., Zhao, M., Guo, Z., Oki, T. & Hanasaki, N. (2006). GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface. *Bull. Amer. Meteor. Soc.*, 87, 1381–1397.

Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth Syst. Sci. Data*, 10, 765-785. https://doiorg/10.5194/essd-10-765-2018.

Döll, P., Kaspar, F. & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation, *J. Hydrol*., 270, 105-134.

Encyclopaedia Britannica (2016). *Kolyma River* [online] Available at: https://www.britannica.com/place/Kolyma-River *[Accessed on June* 19, 2019].

ESRI (2001). ArcGIS 9: What is ArcGIS? http://downloads.esri.com/support/documentation/ao_/698What_is_ArcGis.pdf (Assessed on 16 July 2019)

Fox, J. & Andersen, R. (2005). *Using the R Statistical Computing Environment to Teach Social Statistics Courses.* Department of Sociology, McMaster University.

Frappart, F., Ramillien, G. & Famiglietti J.S. (2011). Water balance of the Arctic drainage system using GRACE gravimetry products. *International Journal of Remote Sensing*, 32(2), 431-453. doi: 10.1080/01431160903474954

Gerten, D., et al. (2004). Terrestrial vegetation and water balance: hydrological evaluation of a dynamic global vegetation model. *Journal of Hydrology*, 286, 249–270. doi: 10.1016/j.jhydrol.2003.09.029

Gosling, S. N. & Arnell, N. W. (2011). Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. *Hydrol. Process.*, 25, 1129-1145. doi:10.1002/hyp.7727

Gosling, S.N., Bretherton, D., Haines, K. & Arnell, N.W. (2010) Global hydrology modelling and uncertainty: running multiple ensembles with a campus grid. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368. http://doi.org/10.1098/rsta.2010.0164

Gosling, S.N., Zaherpour, J., Mount, N.J. et al. (2017) A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1°C, 2°C and 3 °C. *Climatic Change*, 141(3), 577-595. doi:10.1007/s10584-016-1773-3

Gudmundsson, L., et al. (2012a). Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. *Journal of Hydrometeorology*, 13(2), 604–620. doi:10.1175/JHM-D-11-083.1

Gudmundsson, L., et al. (2012b). Evaluation of nine largescale hydrological models with respect to the seasonal runoff climatology in Europe. *Water Resources Research*, 48(11). doi:10.1029/2011WR010911

Guimberteau, M., Ducharne, A., Ciais, P., Boisier, J., Peng, S., De Weirdt, M., Verbeeck, H., et al. (2014). Testing conceptual and physically based soil hydrology schemes against observations for the Amazon Basin. *Geoscientific Model Development*, 7,1115-1136.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1), 80-91.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrologic Eng.,* 4(2), 135-143.

Haddeland, I. *et al*. (2011). Multimodel estimate of the global terrestrial water balance: setup and first results. *J Hydrometeorol.* 12(5), 869–884.

Haine, T. W. N., et al. (2015), Arctic freshwater export: Status, mechanisms, and prospects. *Global Planet Change*, 125, 13–35. doi: 10.1016/j.gloplacha.2014.11.013

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., Tanaka, K. et al. (2008). An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing. *Hydrol. Earth Syst. Sci.*,12,1007-1025,2008

Hattermann, F.F., Krysanova, V., Gosling, S.N., Dankers, R., Daggupati, P., Donnelly, C., Flörke, M., Huang, S., Motovilov, Y., Buda, S., Yang, T., Müller, C., Leng, G., Tang, Q., Portmann, F.T., Hagemann, S., Gerten, D., Wada, Y., Masaki, Y., Alemayehu, T., Satoh, Y., & Samaniego, L. (2017). Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. *Climatic Change*, 141(3), 561-576. doi:10.1007/s10584-016-1829-4

Hinzman, L.D., et al. (2005). Evidence and Implications of recent climate change in Northern Alaska and other Arctic Regions, *Clim. Change*, 72(3), 251– 298. doi:10.1007/s10584-005-5352-2

Hipel, K.W. & McLeod, A.I. (1994). "Time Series Modelling of Water Resources and Environmental Systems", Amsterdam, Elsevier 1994.

Holmes, R.M.R., Coe, M.M.T., Fiske, G.G.J., Gurtovaya, T., McClelland, J.W., Shiklomanov, A.I., Spencer, R.G.M., Tank, S.E., & Zhulidov, A.V. (2013). *Climate change impacts on the hydrology and biogeochemistry of Arctic rivers,* in Climatic Change and Global Warming of Inland Waters: Impacts and Mitigation for Ecosystems and Societies, edited by C. R. Goldman, M. Kumagai, and R. D. Robarts. doi:10.1002/9781118470596.ch1

Houghton J.T., Ding Y., Griggs D.J., Noguer M., van der Linden P.J., Dai X., Maskell K. & Johnson C.A. (eds.) (2001). Climate change 2001: the scientific basis. Contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change. *Cambridge: Cambridge University Press*.

Huang, S., Kumar, R. & Flörke, M. (2017). Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. *Climatic Change*, 141, 399.

Instanes, A., Kokorev, V., Janowicz, R., Bruland, O., Sand, K., & Prowse, T.D. (2016). Changes to freshwater systems affecting Arctic infrastructure and natural resources. *J. Geophys. Res. Biogeosci.*,121. doi:10.1002/2015JG003125

ISIMIP (2018): *The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP): Mission & Implementation Document*, Authors: ISIMIP Coordination Team, Sectoral Coordinators & Scientific Advisory Board. https://www.isimip.org/about/ [Online Accessed on 5 April 2019]

ISIMIP2a (2018): *The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP): ISIMIP2a Simulation protocol*, Authors: ISIMIP Coordination Team, Sectoral Coordinators & Scientific Advisory Board. https://www.isimip.org/#isimip2a/ [Online Accessed on 5 April 2019]

Jin, X., Xu, C.Y., Zhang, Q. & Singh, V.P. (2010). Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *J Hydrol [Amst]*, 383(3–4), 147–155.

Kane, D.L., Hinzman, L., Gieck, R., McNamara, J., Youcha, E., and Oatley, J. (2008). Contrasting extreme runoff events in areas of continuous permafrost, Arctic Alaska. *Hydrol. Res.*, 39. doi:10.2166/nh.2008.005

Karlsson, J. M., Jaramillo, F., & Destouni, G. (2015). Hydro-climatic and lake change patterns in Arctic permafrost and non-permafrost areas. *J. Hydrol.*, 529(1), 134– 145. doi:10.1016/j.jhydrol.2015.07.005

Kauffeldt, A., et al. (2013). Disinformative data in largescale hydrological modelling. *Hydrology and Earth System Sciences*, 17(7]), 2845–2857. doi:10.5194/hess- 17-2845-2013

Kim, H. (2014) Global Soil Wetness Project Phase 3 [online] Available at: http://hydro.iis.u-tokyo.ac.jp/GSWP3/ [Accessed: June 26, 2019]

Koster, R.D., Suarez, M.J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, 105(D20), 24809– 24822. doi:10.1029/2000JD900327.

Krause, P.A., Boyle, D.P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment.

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F. & Kundzewicz, Z.W., (2018). How the performance of hydrological models relates to credibility of projections under climate change, *Hydrological Sciences Journal*, 63(5), 696-720. doi:10.1080/02626667.2018.1446214

Legates, D.R. & McCabe, G.J., (1999). Evaluating the use of ‘‘goodness-of-fit” measures in hydrologic and hydro climatic model evaluation. *Water Resources Research*, 35, 233–241.

Liersch, S., Tecklenburg, J., Rust, H., Dobler, A.H., Fischer, M., Kruschke, T., Koch, H., & Hattermann, F. (2016). Are we using the right fuel to drive hydrological models? A climate impact study in the Upper Blue Nile.

Liu, Y. & Gupta, H.V. (2007). Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resour Res*, 43(7), W07401.

Luojus, K., Pulliainen, J., Takala, M., Kangwa, M., Smolander, T., Wiesmann, A., Derksen, C., Metsämäki, S., Salminen, M., Solberg, R., Nagler, T., Bippus, G., Wunderle, S. & Hüsler, F. (2010). *GlobSnow-2 Product User Guide Version 1.0* European Space Agency Study Contract Report, Global Snow Monitoring for Climate Research

McClelland, J.W., Holmes, R.M., Peterson, B.J. & Stieglitz, M. (2004). Increasing river discharge in the Eurasian Arctic: Consideration of dams, permafrost thaw, and fires as potential agents of change, *J. Geophys. Res.*, 109, D18102. doi:10.1029/2004JD004583

Mernild, S.H., Liston, G.E., & Hiemstra, C.A. (2014). Northern Hemisphere glacier and ice cap surface mass balance and contribution to sea level rise. *J. Clim.*, 27(15), 6051– 6073.

Micklin, P.P., Gennadiyevich Tikhotskiy, C. and Owen, L. (2016). *Yenisey River* [online] Available at: https://www.britannica.com/place/Yenisey-River [Accessed on June 19, 2019]

Micklin, P.P., Owen, L. & Konstantinovna Malik, L. (2018). *Ob River* [online] Available at: https://www.britannica.com/place/Ob-River [Accessed on June 19, 2019]

Micklin, P.P., Owen, L. and Konstantinovna Malik, L. (2018) Ob River [online] Available at: https://www.britannica.com/place/Ob-River [Accessed: June 19, 2019]

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., & Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, 50(3), 885-900.

Müller Schmied, H., et al. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, 18, 3511–3538. doi:10.5194/hess-18-3511-2014

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. & Döll, P. (2016). Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, *Hydrol. Earth Syst. Sci.*, 20, 2877-2898. https://doi.org/10.5194/hess-20-2877-2016

Nash, J.E., Sutcliffe, J.V. (1970). River flow forecasting through. Part I. A conceptual models discussion of principles. *Journal of Hydrology*, 10, 282–290.

Newton, R., Pfirman, S., Schlosser, P., Tremblay, B., Murray, M. & Pomerance, R. (2016). White Arctic vs. Blue Arctic: A case study of diverging stakeholder responses to environmental change. *Earth's Future*, 4, 396–405, doi:10.1002/2016EF000356

Overgaard, J., Rosbjerg, D., & Butts, M.B. (2006). Land-surface modelling in hydrological perspective—A review. *Biogeosciences*, 3, 229–241.

Peterson, B.J., Holmes, R.M., McClelland, J.W., Vorosmarty, C.J., Lammers, R.B., Shiklomanov,A.I., Shikloman ov, I.A. & Rahmstorf, S. (2002). Increasing river discharge to the Arctic Ocean. *Science*, 298, 2171–2173.

Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P.J.F., Kim, H., Kanae, S. & Oki, T. (2012). Incorporating anthropogenic water regulation modules into a land surface model. *J. Hydrometeorol*., 13(1), 255-269.

Proshutinsky, A., Polyakov, I. & Johnson, M. (1999). Climate states and variability of Arctic ice and water dynamics during 1946–1997. *Polar Research*, 18, 135–142.

Prowse, T., Bring, A., Mård, J. & Carmack, E. (2015a). Arctic freshwater synthesis: Introduction. *J. Geophys. Res. Biogeosci.*, 120, 2121–2131. doi:10.1002/2015JG003127

Prowse, T., Bring, A., Mård, J., Carmack, E., Holland, M., Instanes, A., Vihma, T. & Wrona, F. J. (2015b). Arctic freshwater synthesis: Summary of key emerging issues, *J. Geophys. Res. Biogeosci.*, 120, 1887–1893. doi:10.1002/2015JG003128

Prowse, T., Alfredsen, K., Beltaos,S., Bonsal,B., Duguay,C., Korhola,A., McNamara,J., Vincent,W. F., Vuglinsky, V., & Weyhenmeyer, G. A. (2011). Arctic freshwater ice and its climatic role. *Ambio*, 40(1), 46–52.

Prudhomme, C., et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences*, 111, 3262–3267. doi:10.1073/pnas.1222473110

Prudhomme, C., et al. (2011). How well do large-scale models reproduce regional hydrological extremes in Europe? *Journal of Hydrometeorology*, 12(6), 1181–1204. doi:10.1175/2011JHM1387.1

Pulliainen, J. (2006). Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sensing of Environment*, 101, 257-269. doi: 10.1016/j.rse.2006.01.002

R Project (n.d.) What is R? Available from: http://www.r-project.org/about.html [Last cited on 2019 July 9].

Rafiei Emam, A., Kappas, M., Fassnacht, S. et al. (2018). Uncertainty analysis of hydrological modeling in a tropical area using different algorithms, *Frontiers of Earth Science*, 12, 661-671. https://doi.org/10.1007/s11707-018-0695-y

Robinson, J.L. (2019). *Mackenzie River* [online] Available at: https://www.britannica.com/place/Mackenzie-River [Accessed on June 19, 2019]

Robinson, J.L. (2016). *Yukon River* [online] Available at: https://www.britannica.com/place/Yukon-River [Accessed on June 19, 2019]

Schewe, J., et al. (2014). Multimodel assessment of water scarcity under climate change, 2014. *Proceedings of the National Academy of Sciences*, 111, 3245–3250. doi:10.1073/pnas.1222460110

Serreze, M.C., Barrett, A.P., Slater, A.G., Woodgate, R.A., Aagaard, K., Lammers, R.B., Steele, M., Moritz, R., Meredith, M., & Lee, C. M. (2006). The large-scale freshwater cycle of the Arctic. *J. Geophys. Res*., 111, C11010. doi:10.1029/2005JC003424

Setegn, S.G., Srinivasan, R., Melesse, A.M. & Dargahi, B. (2010). SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia. *Hydrol. Processes*, 24, 357–367.

Sevat, E. & Dezetter, A. (1991). Selection of calibration objective functions in the context of rainfall-runoff modeling in a Sudanese savannah area. *Hydrological Sci. J.,* 36(4), 307-330.

Sheffield, J., Goteti, G. & Wood, E.F. (2006). Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modelling, *J. Climate*, 19(13), 3088-3111.

Sitch, S., Smith, B., Prentice, I., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K. & Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9, 161-185.

Smakhtin, V. (2000). Estimating daily flow duration curves from monthly streamflow data, *Water SA*, 26, 13–18.

Stacke, T., Hagemann, S. et al. (2012). Development and evaluation of a global dynamical wetlands extent scheme*. Hydrology and Earth System Sciences,* 16, 2915-2933*.*

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P, Koskinen, J. & Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of spaceborne radiometer data and ground-based measurements. *Remote Sensing of Environment*, 115(12). doi: 10.1016/j.rse.2011.08.014

Takata, K., Emori, S. & Watanabe, T. (2003). Development of the minimal advanced treatments of surface interaction and runoff, *Global Planet. Change*, 38, 209–222.

Tang, Q., Oki, T. & Hu, H. (2006). A distributed biosphere hydrological model (DBHM) for large river basin. *Ann. J. Hydraul. Eng. JSCE*, 50, 37–42.

Traore, A., Ciais, P., Vuichard, N., Poulter, B., Viovy, N., Guimberteau, M., Jung, M., Myneni, R., Fisher, J. et al. (2014). Evaluation of the ORCHIDEE ecosystem model over Africa against 25 years of satellite-based water and carbon measurements. *Journal of Geophysical Research: Biogeosciences*, 119, 1554-1575.

Uddameri, V. (2017). *Using R to read NetCDF Files* WateR – Tutorials (Using R in Water Resources Engineering). Publication at: https://www.researchgate.net/publication/319463953

Van Beek, L.P.H., Wada, Y. & Bierkens, M.F.P. (2011). Global monthly water stress: 1. Water balance and water availability. *Water Resources Research*, 47(7), W07517. doi:10.1029/2010WR009791

Vörösmarty, C.J., Hinzman, L.D., Peterson B.J., Bromwich, D.H., Hamilton, L.C., Morison, J., Romanovsky V.E., Sturm, M. & Webb, R.S. (2001). The hydrologic cycle and its role in Arctic and global environmental change: a rationale and strategy for synthesis study. *Fairbanks: Arctic Research Consortium of the United States*.

Vörösmarty, C.J., Fekete, B.M. & Tucker, B.A. (1998). *Discarge compilation from The Global River Discharge (RivDIS) Project.* Distributed Active Archive Center, Oak Ridge National Laboratory, PANGAEA. https://doi.org/10.1594/PANGAEA.859439

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S. (2003). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.*, 39(8). https://doi.org/10.1029/2002WR001642

Wada, Y., Wisser, D., Bierkens, M. et al. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics*, 5, 15-40.

Waseem, M., Mani, N., Andiego, G. & Usman, M. (2017). A Review of Criteria of Fit for Hydrological Models. *International Research Journal of Engineering and Technology (IRJET)*, 4(11) 1765-1772.

Weedon, G.P., Gomes, S.S., Viterbo, P.P., Shuttleworth, W.J., Blyth, E.E., Österle, H.H., Adam, J.C., Bellouin, N.N., Boucher, O.O. & Best, M.M. (2011). Creation of the WATCH Forcing Data and Its Use to Assess Global and

Regional Reference Crop Evaporation over Land during the Twentieth Century. *J. Hydrometeor.*, 12, 823–848. doi:10.1175/2011JHM1369.1

Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J. & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res*., 50, 7505–7514. doi:10.1002/2014WR015638

Wickham H. (2015). R Packages Available from: http://blog.runsheng.xyz/attachment/r-packages.pdf [Last cited on 2019 July 9].

Wu, H. & Chen, B. (2015). Evaluating uncertainty estimates in distributed hydrological modeling for the Wenjing River watershed in China by GLUE, SUFI-2, and ParaSol methods. *Ecol Eng*, 76, 110–121.

Yang, D., Robinson, D., Zao, Y., Estilow, T. & Ye, B. (2003). Streamflow response to seasonal snow cover extent changes in large Siberian watersheds. *Journal of Geophysical Research*, 108:4578. doi:10.1029/2002JD003419

Yang, J., Reichert, P., Abbaspour, K.C. & Yang, H. (2007). Hydrological modelling of the Chaohe Basin in China: statistical model formulation and Bayesian inference. *J Hydrol [Amst]*, 340(3–4), 167–182.

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *J Hydrol [Amst]*, 358(1–2), 1–23.

Yang, D., Shi, X., & Marsh, P. (2014). Variability and extreme of Mackenzie River daily discharge during 1973–2011, *Quat. Int.*, 380-381, 159– 168. doi:10.1016/j.quaint.2014.09.023

Zaherpour, J., Gosling, S.N., Mount, N., Schmied, H.M., Veldkamp, T.I.E., Dankers, R., … Wada, Y. (2018). Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters*, 13(6), Article: 065015.

Zhang, K., Kimball, J. S., Mu, Q., Jones, L. A., Goetz, S. J., and Running, S. W. (2009). Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005, *J. Hydrol.*, 379(1), 92–110.

10. APPENDICES

Appendix-1 R Scripts

Appendix-2 Extra Figures and Tables

Appendix-3 Research Poster

Appendix-1 R Scripts

Reading NetCDF files and Discharge Time Series Analysis and Visualization (e.g. MATSIRO+gswp3 for Kusur, Lena)

```
# Load required libraries
library(ncdf4)
library(PCICt)
library(ncdf4.helpers)

#.....1971-1980 (1) .....#
# read netcdf file
nc_in_1 <- nc_open('matsiro_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1971_1980.nc')

# get lon, lat, t (time)
lon_1 <- ncvar_get(nc_in_1, "lon")
lat_1 <- ncvar_get(nc_in_1, "lat", verbose = F)
time_1 <- ncvar_get(nc_in_1, "time")

# create date vector
date_1 <- seq.Date(as.Date("1971-01-01"), as.Date("1980-12-31"), by = "days")

# convert time values into calendar dates
new_time_1 <- nc.get.time.series(nc_in_1, v = "dis", time.dim.name ="time")
new_time_1 <- as.character.PCICt(new_time_1)
new_time_1 <- as.Date(new_time_1)

# enter grid cell values representing longitude and latitude of Kusur
dis_daily.ts_1 <- ncvar_get(nc_in_1, "dis", start = c(614,40,1), count = c(1, 1, dim(time_1)))
dis.daily_1 <- data.frame(new_time_1, dis_daily.ts_1)

# aggregate daily to monthly time series
# Assisted by Anne Gaedeke
dis.monthly_1 <- aggregate(dis.daily_1[,2], by=list(date = cut(dis.daily_1[,1], breaks =
c("months"))), mean, na.rm = T)
dis.monthly_1$date <- as.Date(dis.monthly_1$date)

#subsetting the data/selecting few data from the whole dataset
dis.monthly_1 <- subset(dis.monthly_1, date >= '1971-01-01' & date <= '1980-12-01')

#data frame consisting of simulated data from 1971-1980
dis.monthly_1 <- data.frame(dis.monthly_1)
names(dis.monthly_1)[1] <- "Date"
names(dis.monthly_1)[2] <- "Simulated Data"

#.....1981-1990 (2) .....#
# read netcdf file
nc_in_2 <- nc_open('matsiro_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1981_1990.nc')

# get lon, lat, t (time)
lon_2 <- ncvar_get(nc_in_2, "lon")
lat_2 <- ncvar_get(nc_in_2, "lat", verbose = F)
time_2 <- ncvar_get(nc_in_2, "time")
```

```
# create date vector
date_2 <- seq.Date(as.Date("1981-01-01"), as.Date("1990-12-31"), by = "days")

# convert time values into calendar dates
new_time_2 <- nc.get.time.series(nc_in_2, v = "dis", time.dim.name ="time")
new_time_2 <- as.character.PCICt(new_time_2)
new_time_2 <- as.Date(new_time_2)

# enter grid cell values
dis_daily.ts_2 <- ncvar_get(nc_in_2, "dis", start = c(614,40,1), count = c(1, 1, dim(time_2)))
dis.daily_2 <- data.frame(new_time_2, dis_daily.ts_2)

# aggregate to monthly time series
dis.monthly_2 <- aggregate(dis.daily_2[,2], by=list(date = cut(dis.daily_2[,1], breaks =
c("months"))), mean, na.rm = T)
dis.monthly_2$date <- as.Date(dis.monthly_2$date)

#data frame consisting of both simulated data from 1981-1990
dis.monthly_2 <- data.frame(dis.monthly_2)
names(dis.monthly_2)[1] <- "Date"
names(dis.monthly_2)[2] <- "Simulated Data"

#.....1991-2000 (3) .....#
# read in netcdf file
nc_in_3 <- nc_open('matsiro_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1991_2000.nc')

# get lon, lat, t (time)
lon_3 <- ncvar_get(nc_in_3, "lon")
lat_3 <- ncvar_get(nc_in_3, "lat", verbose = F)
time_3 <- ncvar_get(nc_in_3, "time")

# create date vector
date_3 <- seq.Date(as.Date("1991-01-01"), as.Date("2000-12-31"), by = "days")

# convert time values into calendar dates
new_time_3 <- nc.get.time.series(nc_in_3, v = "dis", time.dim.name ="time")
new_time_3 <- as.character.PCICt(new_time_3)
new_time_3 <- as.Date(new_time_3)

# enter grid cell values
dis_daily.ts_3 <- ncvar_get(nc_in_3, "dis", start = c(614,40,1), count = c(1, 1, dim(time_3)))
dis.daily_3 <- data.frame(new_time_3, dis_daily.ts_3)

# aggregate to monthly time series
dis.monthly_3 <- aggregate(dis.daily_3[,2], by=list(date = cut(dis.daily_3[,1], breaks =
c("months"))), mean, na.rm = T)
dis.monthly_3$date <- as.Date(dis.monthly_3$date)

#subsetting the data/selecting few data from the whole dataset
dis.monthly_3 <- subset(dis.monthly_3, date >= '1991-01-01' & date <= '2000-12-01')

#data frame consisting of both simulated data from 1991-2000
dis.monthly_3 <- data.frame(dis.monthly_3)
names(dis.monthly_3)[1] <- "Date"
```

```
names(dis.monthly_3)[2] <- "Simulated Data"

###.....1971-2000 .....###

#data frame consisting of simulated data from 1971-2000
main_dis.monthly <- rbind(dis.monthly_1, dis.monthly_2, dis.monthly_3)
names(main_dis.monthly)[1] <- "Date"
names(main_dis.monthly)[2] <- "Simulated Data_1971-2000_gswp3"

#data frame consisting of observed data from 1971-2000
# Read in txt file
obs_main <- read.table("LENA_Kusur RU_0000014.txt", header = TRUE, sep = ",")

# select first two variables: date and mean discharge from the table
myvars_main <- c("date", "MEAN")
obs.subset_main <- obs_main[myvars_main]

#select observed monthly discharge data of certain period
obs.subset_main$date <- as.POSIXct(obs.subset_main$date, format = "%Y-%m-%d")
obs.subset_main$MEAN <- as.numeric(as.character(obs.subset_main$MEAN))

# subset data - 1971-2000
obs.monthly_main <- subset(obs.subset_main, date >= as.POSIXct('1971-01-31') &
                    date <= as.POSIXct('2000-12-31'))
main_obs.monthly <- data.frame(obs.monthly_main$MEAN)

#observed time series from 1971-2000
main_obs.monthly.1 <- data.frame(obs.monthly_main$date, obs.monthly_main$MEAN)
names(main_obs.monthly.1)[1] <- "Date"
names(main_obs.monthly.1)[2] <- "Observed Data_1971-2000"

#main data frame which contains Simulated and Observed Data from 1971-2000
main <- data.frame(main_dis.monthly, main_obs.monthly)
names(main)[1] <- "Date"
names(main)[2] <- "Simulated Data_1971-2000"
names(main)[3] <- "Observed Data_1971-2000"

# time series visualization
xdata <- c(main_obs.monthly.1$Date)
y1 <- c(main_obs.monthly.1$`Observed Data_1971-2000`)
y2 <- c(main_dis.monthly$`Simulated Data_1971-2000_gswp3`)

# plot the first curve by calling plot() function
# First curve is plotted
plot(xdata, y1, main="matsiro_gswp3_nobc_hist_nosoc_co2_dis_global_daily_1971_2000",
    type="o", col="blue", pch="o", lty=1, xlab="Year", ylab="Discharge" )

# Add second curve to the same plot by calling points() and lines()
points(xdata, y2, col="red", pch="+")
lines(xdata, y2, col="red",lty=1)

# Adding a legend inside box in graph coordinates.
legend("topleft", legend=c("Observed", "Simulated"), col=c("blue","red"),
     pch=c("o","+"),lty=c(1,2), cex=0.7,
```

```
    box.lty=1, ncol=2)

##time series for gswp3 climate model from 1971 to 2000
main.gswp3 <- main_dis.monthly

# set path and filename for simulated discharge data only
csvpath <- "D:/Data Analysis_matsiro/Lena/matsiro_Lena_all climate models_1971_2000_sim-obs ts/"
csvname <- "lena-kusur_matsiro_gswp3_sim_dis_1971-2000.csv"
csvfile <- paste(csvpath, csvname, sep="")
write.table(na.omit(main.gswp3),csvfile, row.names=FALSE, sep=",")

# set path and filename for observed discharge data only excluding the missing (NA) values
csvpath <- "D:/Data Analysis_matsiro/Lena/matsiro_Lena_all climate models_1971_2000_sim-obs ts/"
csvname <- "lena-kusur_obs_dis_1971-2000_without_NA.csv"
csvfile <- paste(csvpath, csvname, sep="")
write.table(na.omit(main_obs.monthly.1),csvfile, row.names=FALSE, sep=",")

# set path and filename for observed discharge data only including the missing (NA) values
csvpath <- "D:/Data Analysis_matsiro/Lena/matsiro_Lena_all climate models_1971_2000_sim-obs ts/"
csvname <- "lena-kusur_obs_dis_1971-2000_with_NA.csv"
csvfile <- paste(csvpath, csvname, sep="")
#write.table(na.omit(main_obs.monthly.1),csvfile, row.names=FALSE, sep="")
write.table(main_obs.monthly.1, csvfile, row.names=FALSE, sep=",")
```

Monthly Hydrographs/Dynamics Analysis (e.g. MATSIRO+gswp3 for Kusur, Lena)

```
# load r package libraries
library(PCICt)
library(hydroGOF)

# Read in csv files
gswp3_sim_dis <- read.csv("lena-Kusur_matsiro_gswp3_sim_dis_1971-2000.csv")
obs_dis <-read.csv("lena-Kusur_obs_dis_1971-2000_without_NA.csv")

#data frame consisting of simulated data from 1971-2000
sim_dis <- data.frame(gswp3_sim_dis)
names(sim_dis)[1] <- "Date"
names(sim_dis)[2] <- "Simulated Discharge_gswp3"

#delete selective date (deleting a row) from the simulated dataset to make it coherent with the observed dataset by managing missing data
sim_dis <- subset(sim_dis, Date!="1996-01-01")

#data frame consisting of observed data without NA from 1971-2000
obs_dis <- data.frame(obs_dis)
obs_dis$Date <- as.Date(obs_dis$Date)
names(obs_dis)[1] <- "Date"
names(obs_dis)[2] <- "Observed Discharge"

#combining simulated and observed data together
```

```
monthly_dynamics <- cbind(sim_dis, obs_dis$`Observed Discharge`)
names(monthly_dynamics)[3] <- "Observed Discharge"

#changing into calendar date and time
calendar_date <- as.Date(monthly_dynamics$Date)

# monthly dynamics visualization of both simulate and observed data in a single plot
#subjective assessment
xdata <- c(calendar_date)
y1 <- c(monthly_dynamics$`Observed Discharge`)
y2 <- c(monthly_dynamics$`Simulated Discharge_gswp3`)

# plot the first curve by calling plot() function
# First curve is plotted
plot(xdata, y1, main="Lena-Kusur_matsiro_gswp3_monthly_dynamics_1971-2000",
   type="l", col="black", pch=0, lty=1, lwd=1, xlab="Year", ylab="Discharge", ylim= c(0,
120000))

# Add second curve (gswp3) to the same plot
lines(xdata, y2, col="blue",lty=1, lwd = 1)

# Adding a legend inside box in graph coordinates.
# Note that the order of plots is maintained in the vectors of attributes.
legend(0, 119000, legend=c("Observed", "Simulated"), col=c("black","blue"), title="Legend",
     lty=1, cex=1, box.lty=0, ncol=2)
#Objective assessment
#calculate PBIAS, NSE, ratio of santard deviation (rSD) and Bias in SD
PBIAS <- pbias(monthly_dynamics$`Simulated Discharge_gswp3`,
monthly_dynamics$`Observed Discharge`)
PBIAS <- signif(PBIAS,3)
NSE <- NSE(monthly_dynamics$`Simulated Discharge_gswp3`, monthly_dynamics$`Observed
Discharge`)
NSE <- signif(NSE,2)
rSD <- rSD(monthly_dynamics$`Simulated Discharge_gswp3`, monthly_dynamics$`Observed
Discharge`)
rSD <- signif(rSD,3)
bias_in_SD <- (rSD - 1)*100
bias_in_SD <- signif(bias_in_SD,3)
r <- rPearson(monthly_dynamics$`Simulated Discharge_gswp3`, monthly_dynamics$`Observed
Discharge`)
r <- signif(r,2)
gof <- gof(monthly_dynamics$`Simulated Discharge_gswp3`, monthly_dynamics$`Observed
Discharge`)
R2 <- gof[17,1]
R2 <- signif(R2,2)
criteria_tests <- data.frame(PBIAS, NSE, rSD, bias_in_SD, r, R2, row.names = "ct_gswp3")
write.csv(criteria_tests, file = "gswp3_criteria tests.csv")
```

Seasonal Dynamics Analysis for Discharge (e.g. MATSIRO+gswp3 for Kusur, Lena)

```
# load R libraries
library(PCICt)
library(hydroGOF)
library(hydroTSM)
library(lubridate)

# Read in csv files
gswp3_sim_dis <- read.csv("lena-Kusur_matsiro_gswp3_sim_dis_1971-2000.csv")
obs_dis <-read.csv("lena-kusur_obs_dis_1971-2000_without_NA.csv")

#data frame consisting of simulated data from 1971-2000
sim_dis <- data.frame(gswp3_sim_dis)
names(sim_dis)[1] <- "Date"
names(sim_dis)[2] <- "Simulated Discharge"

#delete selective date (deleting a row) from the simulated dataset to make it coherent with the
observed dataset by managing missing data
sim_dis <- subset(sim_dis, Date!="1996-01-01")

#data frame consisting of observed data without NA from 1971-2000
obs_dis <- data.frame(obs_dis)
obs_dis$Date <- as.Date(obs_dis$Date)
names(obs_dis)[1] <- "Date"
names(obs_dis)[2] <- "Observed Discharge"
#creating a loop for selecting simulated data of 1 month from the time series and taking the avg
mon_sim <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12){
  month_select_sim <- sim_dis$`Simulated Discharge`[which(month(sim_dis$Date)==i)]
  mean_sim <- mean(month_select_sim)
  mon_sim <- append(mon_sim, mean_sim)
}

#creating a loop for selecting observed data of 1 month from the time series and taking the avg
mon_obs <- 0
for (i in 1:12){
  #data frames has two types Date and obs
  #dd1 <- obs_dis[!is.na(obs_dis$`Observed Discharge`),]
  month_select_obs <- obs_dis$`Observed Discharge`[which(month(obs_dis$Date)==i)]
  mean_obs <- mean(month_select_obs)
  mon_obs <- append(mon_obs, mean_obs)
}

#transform date to month names
months_name <- month(ymd(sim_dis$Date), label = TRUE, abbr = FALSE)
months_name <- months_name[1:12]
#rename if desired
names(months_name)[which(names(months_name)=="Date")]<-"Months"

#Seasonal Dynamics Dataframe
seasonal_dynamics <- data.frame(months_name, mon_sim[2:13], mon_obs[2:13])
names(seasonal_dynamics)[1] <- "Month"
names(seasonal_dynamics)[2] <- "Simulated Discharge_gswp3"
```

```
names(seasonal_dynamics)[3] <- "Observed Discharge"

# seasonal dynamics visualization of both simulate and observed data in a single plot
#subjective assessment
xdata <- c(seasonal_dynamics$Month)
y1 <- c(seasonal_dynamics$`Observed Discharge`)
y2 <- c(seasonal_dynamics$`Simulated Discharge_gswp3`)

# plot the first curve by calling plot() function
# First curve is plotted
plot(xdata, y1, main="Lena-Kusur_matsiro_gswp3_seasonal_dynamics_1971-2000",
   type="l", col="black", pch=0, lty=1, lwd=2, xlab="Months", ylab="Discharge (m^3/s)", ylim=
c(0, 100000), cex.lab=1,las=1,cex.axis=0.8)

# Add second curve (gswp3) to the same plot
lines(xdata, y2, col="blue",lty=1, lwd = 2)

# Adding a legend inside box in graph coordinates.
legend(1, 99000, legend=c("Observed", "Simulated"),
     col=c("black","blue"), title="Legend",
     lty=1, cex=0.8, box.lty=0, ncol=2, lwd = 2)

#Objective assessment
#calculate PBIAS, NSE, ratio of santard deviation (rSD) and Bias in SD
PBIAS <- pbias(seasonal_dynamics$`Simulated Discharge_gswp3`,
seasonal_dynamics$`Observed Discharge`)
PBIAS <- signif(PBIAS,3)
NSE <- NSE(seasonal_dynamics$`Simulated Discharge_gswp3`, seasonal_dynamics$`Observed
Discharge`)
NSE <- signif(NSE,2)
rSD <- rSD(seasonal_dynamics$`Simulated Discharge_gswp3`, seasonal_dynamics$`Observed
Discharge`)
rSD <- signif(rSD,3)
bias_in_SD <- (rSD - 1)*100
bias_in_SD <- signif(bias_in_SD,3)
r <- rPearson(seasonal_dynamics$`Simulated Discharge_gswp3`, seasonal_dynamics$`Observed
Discharge`)
r <- signif(r,2)
gof <- gof(seasonal_dynamics$`Simulated Discharge_gswp3`, seasonal_dynamics$`Observed
Discharge`)
R2 <- gof[17,1]
R2 <- signif(R2,2)
criteria_tests <- data.frame(PBIAS, NSE, rSD, bias_in_SD, r, R2, row.names = "ct_gswp3")
write.csv(criteria_tests, file = "gswp3_criteria tests.csv")
```

Seasonal Dynamics plot showing simulated mean discharge with maximum and minimum values as spread (e.g. MATSIRO+gswp3 for Kusur, Lena)

```
# load libraries
library(PCICt)
library(lubridate)

# Read in csv files
gswp3_sim_dis_k <- read.csv("lena-Kusur_matsiro_gswp3_sim_dis_1971-2000.csv")
princeton_sim_dis_k <- read.csv("lena-Kusur_matsiro_princeton_sim_dis_1971-2000.csv")
watch_sim_dis_k <- read.csv("lena-Kusur_matsiro_watch_sim_dis_1971-2000.csv")
wfdei_sim_dis_k <- read.csv("lena-Kusur_matsiro_wfdei_sim_dis_1971-2000.csv")
obs_dis_k <- read.csv("lena-kusur_obs_dis_1971-2000_without_NA.csv")

#data frame consisting of simuated data from 1971-2000
sim_dis_k <- data.frame(gswp3_sim_dis_k,
          princeton_sim_dis_k$Simulated.Data_1971.2000_princeton,
          watch_sim_dis_k$Simulated.Data_1971.2000_watch,
          wfdei_sim_dis_k$Simulated.Data_1971.2000_wfdei)
names(sim_dis_k)[1] <- "Date"
names(sim_dis_k)[2] <- "Simulated discharge_gswp3"
names(sim_dis_k)[3] <- "Simulated discharge_princeton"
names(sim_dis_k)[4] <- "Simulated discharge_watch"
names(sim_dis_k)[5] <- "Simulated discharge_wfdei"


#delete values of a selective date (deleting a row) from the simulated dataset created above
sim_dis_k <- subset(sim_dis_k, Date!="1996-01-01")

#creating a loop for selecting simulated data of 1 month from the time series and taking the avg
for gswp3
mon_gswp3_k <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12) {
     month_select_gswp3_k <-
          sim_dis_k$`Simulated discharge_gswp3`[which(month(sim_dis_k$Date) == i)]
     mean_gswp3_k <- mean(month_select_gswp3_k)
     mon_gswp3_k <- append(mon_gswp3_k, mean_gswp3_k)
}

#creating a loop for selecting simulated data of 1 month from the time series and taking the avg
for princeton
mon_princeton_k <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12) {
     month_select_princeton_k <-
          sim_dis_k$`Simulated discharge_princeton`[which(month(sim_dis_k$Date) == i)]
     mean_princeton_k <- mean(month_select_princeton_k)
     mon_princeton_k <- append(mon_princeton_k, mean_princeton_k)
}

#creating a loop for selecting simulated data of 1 month from the time series and taking the avg
for watch
mon_watch_k <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12) {
     month_select_watch_k <-
```

```
        sim_dis_k$`Simulated discharge_watch`[which(month(sim_dis_k$Date) == i)]
    mean_watch_k <- mean(month_select_watch_k)
    mon_watch_k <- append(mon_watch_k, mean_watch_k)
}

#creating a loop for selecting simulated data of 1 month from the time series and taking the avg
for wfdei
mon_wfdei_k <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12) {
    month_select_wfdei_k <-
        sim_dis_k$`Simulated discharge_wfdei`[which(month(sim_dis_k$Date) == i)]
    mean_wfdei_k <- mean(month_select_wfdei_k)
    mon_wfdei_k <- append(mon_wfdei_k, mean_wfdei_k)
}

#data frame consisting of simulated data from 1971-2000 without the date column
sim_dis_k_1 <-
    data.frame(mon_gswp3_k[2:13], mon_princeton_k[2:13], mon_watch_k[2:13],
mon_wfdei_k[2:13])
names(sim_dis_k_1)[1] <- "Simulated discharge_gswp3"
names(sim_dis_k_1)[2] <- "Simulated discharge_princeton"
names(sim_dis_k_1)[3] <- "Simulated discharge_watch"
names(sim_dis_k_1)[4] <- "Simulated discharge_wfdei"

# max, mean and min
max_k <- data.frame(apply(sim_dis_k_1, 1, FUN = max))
names(max_k)[1] <- "Maximum"

mean_k <- data.frame(apply(sim_dis_k_1, 1, FUN = mean))
names(mean_k)[1] <- "Mean"

min_k <- data.frame(apply(sim_dis_k_1, 1, FUN = min))
names(min_k)[1] <- "Minimum"

#creating a data frame to record summary of the simulated data
summary_k <- data.frame(max_k, mean_k, min_k)

#data frame consisting of observed data without NA from 1971-2000
obs_dis_k <- data.frame(obs_dis_k)
obs_dis_k$Date <- as.Date(obs_dis_k$Date)
names(obs_dis_k)[1] <- "Date"
names(obs_dis_k)[2] <- "Observed discharge"

#creating a loop for selecting observed data of 1 month from the time series and taking the avg
mon_obs_k <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12) {
    month_select_obs_k <-
        obs_dis_k$`Observed discharge`[which(month(obs_dis_k$Date) == i)]
    mean_obs_k <- mean(month_select_obs_k)
    mon_obs_k <- append(mon_obs_k, mean_obs_k)
}

#transform date to month names
months_name <- month(ymd(gswp3_sim_dis_k$Date), label = TRUE, abbr = FALSE)
```

```
months_name <- months_name[1:12]
#rename if desired
names(months_name)[which(names(months_name) == "Date")] <- "Months"

#combining simulated, observed data with the summary (max, mean and min)
seasonal_dynamics_k <-
      cbind(months_name, mon_obs_k[2:13], sim_dis_k_1, summary_k)
names(seasonal_dynamics_k)[1] <- "Months"
names(seasonal_dynamics_k)[2] <- "Observed discharge"

#time series plot of simulated (all climate forcing data) and observed data
xdata <- c(seasonal_dynamics_k$Months)
y1_k <- c(seasonal_dynamics_k$`Observed discharge`)
y2_k <- c(seasonal_dynamics_k$`Simulated discharge_gswp3`)
y3_k <- c(seasonal_dynamics_k$`Simulated discharge_princeton`)
y4_k <- c(seasonal_dynamics_k$`Simulated discharge_watch`)
y5_k <- c(seasonal_dynamics_k$`Simulated discharge_wfdei`)

jpeg("MATSIRO_Lena-Kusur_Seasonal Dynamics.jpeg", width = 1350, height = 1200, res = 240)
# plot the first curve by calling plot() function
# outline margin
par(mar=c(4,5,3,2) + 0.1)
plot(xdata, y1_k, type = "l", col = "black", lty = 1, lwd = 2, xlab = "", ylab = "", ylim = c(0, 110000),
      cex.lab = 1, xaxt="n", las = 1, cex.axis = 1)
mtext("MATSIRO - Lena-Kusur", side=3, line=1.5, cex=1, col="black")
mtext("Seasonal Dynamics 1971-2000", side=3, line=0.5, cex=1, col="black")
axis(1, at=1:12)
mtext("Discharge (m^3/s)", side = 2, line = 3.5)
mtext("Months", side = 1, line = 2)

# Add second curve (gswp3) to the same plot by calling matlines()
matlines(xdata, y2_k, col = "blue", lty = 1, lwd = 2)

# Add third curve (princeton) to the same plot
matlines(xdata, y3_k, col = "red", lty = 1, lwd = 2)

# Add fourth curve (watch) to the same plot
matlines(xdata, y4_k, col = "green", lty = 1, lwd = 2)

# Add fifth curve (wfdei) to the same plot
#points(xdata, y4, col="yellow", pch=4)
 matlines(xdata, y5_k, col = "yellow", lty = 1, lwd = 2)

# Adding a legend inside box in graph coordinates.
legend(0.75, 110000, legend = c("Observed","Simulated_gswp3","Simulated_princeton",
          "Simulated_watch","Simulated_wfdei), col = c("black", "blue", "red", "green", "yellow"),
          title = "Legend", lty = 1, lwd = c(2,2,2,2,2), cex = 0.95, box.lty = 0, ncol = 2)
dev.off()

#plot with obseved data and summary data (max, mean and min)
y6_k <- c(seasonal_dynamics_k$Maximum)
y7_k <- c(seasonal_dynamics_k$Mean)
y8_k <- c(seasonal_dynamics_k$Minimum)
```

```
df_k <- data.frame(x = seq(length(xdata)), Mean = y7_k, Min = y8_k, Max = y6_k, Observed =
y1_k)

jpeg("MATSIRO_Lena-Kusur_Seasonal Dynamics Summary.jpeg", width = 1350, height = 1200,
res = 240)

par(mar=c(4,5,3,2) + 0.1)
plot(Mean ~ x, data = df_k, ylim = c(0, 100000), type = "l", xlab = "", ylab = "", xaxt="n", cex.lab =
1,
     las = 1, cex.axis = 1)

mtext("MATSIRO - Lena-Kusur", side=3, line=1.5, cex=1, col="black")
mtext("Seasonal Dynamics Summary 1971-2000", side=3, line=0.5, cex=1, col="black")
axis(1, at=1:12)
mtext("Discharge (m^3/s)", side = 2, line = 3.5)
mtext("Months", side = 1, line = 2)
#make polygon where coordinates start with lower limit (min) and then upper limit (Max) in
reverse order
with(df_k, polygon(c(x, rev(x)), c(Min, rev(Max)), col = "grey80", border = FALSE))
matlines(df_k[, 1], df_k[, -1], lwd = c(2, 1, 1, 2), lty = 1, col = c("black", "grey80", "grey80",
"blue"))
legend(0.75,100000, legend = c("Observed", "Simulated Mean", "Spread"), pch= c(NA,NA,15),
        pt.cex = 2, col = c("blue", "black","grey80"), title = "Legend", lty = c(1,1,NA), cex = 1,
        lwd = c(2,2,2), box.lty = 0, ncol = 2)
dev.off()
```

Linear Trend Analysis of Observed Discharge at Igarka, Yenisei

```
# r librares
library(PCICt)
library(lubridate)

# Read in csv files
obs_original <- read.table("YENISEI_Igarka RU_0000050.txt", header = TRUE, sep = ",")

# select variables v1, v2
myvars <- c("date", "MEAN")
obs <- obs_original[myvars]

#select observed monthly dischage data of certain period
obs$date <- as.POSIXct(obs$date, format = "%Y-%m-%d")
obs$MEAN <- as.numeric(as.character(obs$MEAN))

# subset data - 1971-2000
obs <- subset(obs, date >= as.POSIXct('1971-01-31') & date <= as.POSIXct('2000-12-31'))

obs <- data.frame(obs$date, obs$MEAN)
names(obs)[1] <- "Date"
names(obs)[2] <- "Observed Discharge"

#omit the NA values from the dataset
obs <- na.omit(obs)

# aggregate to annual mean discharge
```

```
obs_annual <- aggregate(obs[,2], by=list(date = cut(obs[,1],breaks = c("year"))),
                        mean, na.rm = T)
obs_annual$date <- as.Date(obs_annual$date)
obs_annual <- data.frame(obs_annual$date, obs_annual$x)
names(obs_annual)[1] <- "Date"
names(obs_annual)[2] <- "Annual Observed Discharge"

#scatter plot of the observed discharge
x <- c(obs_annual$Date)
y <- c(obs_annual$`Annual Observed Discharge`)
plot(x, y, main="Trend Analysis: Yenisei-Igarka", xlab="Year", ylab="Observed Discharge")


#simple linear regression trend analysis for observed discharge
mod <- lm(y~x)
summary(mod)
attributes(mod)
coefficients <- mod$coefficients
abline(mod, col = 2, lwd = 2)
```

Extreme Flows Analysis (e.g. MATSIRO+gswp3 for Kusur, Lena)

```
# load libraries
library(PCICt)
library(hydrostats)
library(hydroGOF)

# Read in csv files
gswp3_sim_dis <- read.csv("Lena_Kusur_matsiro_gswp3_daily_dis_1971-2000.csv")
obs_dis <-read.csv("Lena_Kusur_obs_dis_1971-2000_without_NA.csv")

#data frame consisting of simulated daily time series data from 1971-2000
sim_dis <- data.frame(gswp3_sim_dis)
sim_dis$Date <- as.Date(sim_dis$Date)
names(sim_dis)[1] <- "Date"
names(sim_dis)[2] <- "Q"

#data frame consisting of daily observed data without NA from 1971-2000
obs_dis <- data.frame(obs_dis)
obs_dis$Date <- as.Date(obs_dis$Date)
names(obs_dis)[1] <- "Date"
names(obs_dis)[2] <- "Q"

#Calculate SIMULATED HIGH FLOW PERCENTILES
Q10_sim <- high.spells(sim_dis, quant=0.9)
Q10_sim <- Q10_sim$high.spell.threshold
Q10_sim <- signif(Q10_sim, 6)

Q5_sim <- high.spells(sim_dis, quant=0.95)
Q5_sim <- Q5_sim$high.spell.threshold
Q5_sim <- signif(Q5_sim, 6)

Q1_sim <- high.spells(sim_dis, quant=0.99)
Q1_sim <- Q1_sim$high.spell.threshold
```

```
Q1_sim <- signif(Q1_sim, 6)

Q0.1_sim <- high.spells(sim_dis, quant=0.999)
Q0.1_sim <- Q0.1_sim$high.spell.threshold
Q0.1_sim <- signif(Q0.1_sim, 6)

Q0.01_sim <- high.spells(sim_dis, quant=0.9999)
Q0.01_sim <- Q0.01_sim$high.spell.threshold
Q0.01_sim <- signif(Q0.01_sim, 6)
#Calculate SIMULATED LOW FLOW PERCENTILES
Q90_sim <- low.spells(sim_dis, quant=0.1)
Q90_sim <- Q90_sim$low.spell.threshold
Q90_sim <- signif(Q90_sim, 4)

Q95_sim <- low.spells(sim_dis, quant=0.05)
Q95_sim <- Q95_sim$low.spell.threshold
Q95_sim <- signif(Q95_sim, 4)

Q99_sim <- high.spells(sim_dis, quant=0.01)
Q99_sim <- Q99_sim$low.spell.threshold
Q99_sim <- signif(Q99_sim, 4)

Q99.9_sim <- high.spells(sim_dis, quant=0.001)
Q99.9_sim <- Q99.9_sim$low.spell.threshold
Q99.9_sim <- signif(Q99.9_sim, 4)

Q99.99_sim <- high.spells(sim_dis, quant=0.0001)
Q99.99_sim <- Q99.99_sim$low.spell.threshold
Q99.99_sim <- signif(Q99.99_sim, 4)

#calculate OBSERVED HIGH FLOW PERCENTILES
Q10_obs <- high.spells(obs_dis, quant=0.9)
Q10_obs <- Q10_obs$high.spell.threshold
Q10_obs <- signif(Q10_obs, 6)

Q5_obs <- high.spells(obs_dis, quant=0.95)
Q5_obs <- Q5_obs$high.spell.threshold
Q5_obs <- signif(Q5_obs, 6)

Q1_obs <- high.spells(obs_dis, quant=0.99)
Q1_obs <- Q1_obs$high.spell.threshold
Q1_obs <- signif(Q1_obs, 6)

Q0.1_obs <- high.spells(obs_dis, quant=0.999)
Q0.1_obs <- Q0.1_obs$high.spell.threshold
Q0.1_obs <- signif(Q0.1_obs, 6)

Q0.01_obs <- high.spells(obs_dis, quant=0.9999)
Q0.01_obs <- Q0.01_obs$high.spell.threshold
Q0.01_obs <- signif(Q0.01_obs, 6)

#Calculate OBSERVED LOW FLOW PERCENTILES
Q90_obs <- low.spells(obs_dis, quant=0.1)
Q90_obs <- Q90_obs$low.spell.threshold
```

```
Q90_obs <- signif(Q90_obs, 4)

Q95_obs <- low.spells(obs_dis, quant=0.05)
Q95_obs <- Q95_obs$low.spell.threshold
Q95_obs <- signif(Q95_obs, 4)

Q99_obs <- low.spells(obs_dis, quant=0.01)
Q99_obs <- Q99_obs$low.spell.threshold
Q99_obs <- signif(Q99_obs, 4)

Q99.9_obs <- low.spells(obs_dis, quant=0.001)
Q99.9_obs <- Q99.9_obs$low.spell.threshold
Q99.9_obs <- signif(Q99.9_obs, 4)

Q99.99_obs <- low.spells(obs_dis, quant=0.0001)
Q99.99_obs <- Q99.99_obs$low.spell.threshold
Q99.99_obs <- signif(Q99.99_obs, 4)

#Dataframe for HIGH FLOWS
Q10 <- data.frame(Q10_obs, Q10_sim, Q10_PBIAS)
row.names(Q10) <- "Q10"
names(Q10)[1] <- "Observed Discharge"
names(Q10)[2] <- "Simulated Discharge_gswp3"
names(Q10)[3] <- "PBIAS"

Q5 <- data.frame(Q5_obs, Q5_sim, Q5_PBIAS)
row.names(Q5) <- "Q5"
names(Q5)[1] <- "Observed Discharge"
names(Q5)[2] <- "Simulated Discharge_gswp3"
names(Q5)[3] <- "PBIAS"

Q1 <- data.frame(Q1_obs, Q1_sim, Q1_PBIAS)
row.names(Q1) <- "Q1"
names(Q1)[1] <- "Observed Discharge"
names(Q1)[2] <- "Simulated Discharge_gswp3"
names(Q1)[3] <- "PBIAS"

Q0.1 <- data.frame(Q0.1_obs, Q0.1_sim, Q0.1_PBIAS)
row.names(Q0.1) <- "Q0.1"
names(Q0.1)[1] <- "Observed Discharge"
names(Q0.1)[2] <- "Simulated Discharge_gswp3"
names(Q0.1)[3] <- "PBIAS"

Q0.01 <- data.frame(Q0.01_obs, Q0.01_sim, Q0.01_PBIAS)
row.names(Q0.01) <- "Q0.01"
names(Q0.01)[1] <- "Observed Discharge"
names(Q0.01)[2] <- "Simulated Discharge_gswp3"
names(Q0.01)[3] <- "PBIAS"

high_flows <- rbind(Q10, Q5, Q1, Q0.1, Q0.01, make.row.names = TRUE)

#Dataframe for LOW FLOWS
Q90 <- data.frame(Q90_obs, Q90_sim, Q90_PBIAS)
row.names(Q90) <- "Q90"
```

```
names(Q90)[1] <- "Observed Discharge"
names(Q90)[2] <- "Simulated Discharge_gswp3"
names(Q90)[3] <- "PBIAS"

Q95 <- data.frame(Q95_obs, Q95_sim, Q95_PBIAS)
row.names(Q95) <- "Q95"
names(Q95)[1] <- "Observed Discharge"
names(Q95)[2] <- "Simulated Discharge_gswp3"
names(Q95)[3] <- "PBIAS"

Q99 <- data.frame(Q99_obs, Q99_sim, Q99_PBIAS)
row.names(Q99) <- "Q99"
names(Q99)[1] <- "Observed Discharge"
names(Q99)[2] <- "Simulated Discharge_gswp3"
names(Q99)[3] <- "PBIAS"

Q99.9 <- data.frame(Q99.9_obs, Q99.9_sim, Q99.9_PBIAS)
row.names(Q99.9) <- "Q99.9"
names(Q99.9)[1] <- "Observed Discharge"
names(Q99.9)[2] <- "Simulated Discharge_gswp3"
names(Q99.9)[3] <- "PBIAS"

Q99.99 <- data.frame(Q99.99_obs, Q99.99_sim, Q99.99_PBIAS)
row.names(Q99.99) <- "Q99.99"
names(Q99.99)[1] <- "Observed Discharge"
names(Q99.99)[2] <- "Simulated Discharge_gswp3"
names(Q99.99)[3] <- "PBIAS"

low_flows <- rbind(Q90, Q95, Q99, Q99.9, Q99.99, make.row.names = TRUE)

flow_indicators <- rbind (high_flows, low_flows, make.row.names = TRUE)
write.table(flow_indicators, file= "lena_kusur_matsiro_gswp3_extreme_flows.csv", sep = ",")
```

Observed Snow Water Equivalent Time Series Analysis (Center of Lena Basin)

```
# load libraries
library(ncdf4)
library(PCICt)
library(ncdf4.helpers)

#.....1980-2000.....#
# read netcdf file
nc_in <- nc_open('GlobSnow_SWE_monthly_lena_0_5_197909-201412_v2.nc4')

# get lon, lat, t (time)
lon <- ncvar_get(nc_in, "lon")
lat <- ncvar_get(nc_in, "lat", verbose = F)
time <- ncvar_get(nc_in, "time")

# convert time -- split the time units string into fields
# Assisted by Anne Gaedeke
date <-seq.Date(from = as.Date("1979/09/01"), to = as.Date("2014/12/31"), by = "months")
# correcting date and time obtained from the NetCDF file
```

```
date.index <- 0:(length(date)-1)
date.new <- date[date.index  %in% time]

# choose the grid cell NORTH
swe_monthly.ts_n <- ncvar_get(nc_in, "SWE_avg", start = c(46,38,1), count = c(1, 1, dim(time)))
swe.monthly_n <- data.frame(date.new, swe_monthly.ts_n)

#subsetting the required set of time series data from 1980-2000
swe.monthly_n <- subset(swe.monthly_n, date.new >= '1980-01-01' & date.new <= '2000-12-
31')

#output of time series data in .csv format
write.table(na.omit(swe.monthly_n),file = "lena_north_swe_monthly_observed data_1980-
2000.csv", row.names=FALSE, sep=",")
```

Simulated Snow Water Equivalent Time Series Analysis (e.g. MATSIRO+gswp3 for Center of Lena Basin)

```
# r libraries
library(ncdf4)
library(PCICt)
library(ncdf4.helpers)

#.....1980-2000 .....#
# read netcdf file
nc_in <- nc_open('matsiro_gswp3_nobc_hist_nosoc_co2_swe_global_monthly_1971_2010.nc')

# get lon, lat, t (time)
lon <- ncvar_get(nc_in, "lon")
lat <- ncvar_get(nc_in, "lat", verbose = F)
time <- ncvar_get(nc_in, "time")

# convert into actual date
date <- seq.Date(from = as.Date("1971/01/01"), to = as.Date("2010/12/31"), by = "months")

# choose the grid cell CENTER
swe_monthly.ts_c <- ncvar_get(nc_in, "swe", start = c(606,56,1), count = c(1, 1, dim(time)))
swe.monthly_c <- data.frame(date, swe_monthly.ts_c)

#subsetting the required set of time series data from 1980-2000
swe.monthly_c <- subset(swe.monthly_c, date >= '1980-01-01' & date <= '2000-12-31')

#output of time series data in .csv format
write.table(na.omit(swe.monthly_c),file = "lena_center_matsiro_gswp3_swe_monthly simulated
data_1980-2000.csv", row.names=FALSE, sep=",")
```

Model Validation using long-term average monthly (seasonal) Snow Water Equivalent (e.g. MATSIRO for Center of Lena Basin)

```
# load libraries
library(PCICt)
library(zoo)
library(lubridate)

#..... center .....#
#Months Name
months <- read.csv("months name.csv")
#transform date to month names
months_name <- month(ymd(months$Date), label = TRUE, abbr = FALSE)
months_name <- months_name[1:12]
#rename if desired
names(months_name)[which(names(months_name)=="Date")]<-"Months"

#..... gswp3 .....#
# Read in csv files
gswp3_sim_dis_gs <- read.csv("lena_center_matsiro_gswp3_swe_monthly simulated data_1980-2000.csv")
obs_dis <-read.csv("lena_center_swe_monthly_observed data_1980-2000.csv")

#data frame consisting of observed data without NA from 1980-2000
obs_dis <- data.frame(obs_dis)
names(obs_dis)[1] <- "Date"
names(obs_dis)[2] <- "Observed Discharge_center"
obs_dis$Date <- as.Date(obs_dis$Date)

#delete certain values from observed dataset
# obs_dis <- subset(obs_dis, !(Date=="1996-10-31"))

#data frame consisting of simulated data from 1980-2000
sim_dis_gs <- data.frame(gswp3_sim_dis_gs)
names(sim_dis_gs)[1] <- "Date"
names(sim_dis_gs)[2] <- "Simulated Discharge_gswp3"
sim_dis_gs$Date <- as.Date(sim_dis_gs$Date)

#delete values of a selective date (deleting a row) from the simulated dataset created above
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1980" & format.Date(Date, "%m")=="05"))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1980-08-01' & Date <= '1980-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1981-07-01' & Date <= '1981-08-31'))
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1981" & format.Date(Date, "%m")=="12"))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1982-06-01' & Date <= '1982-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1983-07-01' & Date <= '1983-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1984-07-01' & Date <= '1984-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1985-08-01' & Date <= '1985-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1986-07-01' & Date <= '1986-08-31'))
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1987" & format.Date(Date, "%m")=="06"))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1987-08-01' & Date <= '1987-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1988-07-01' & Date <= '1988-08-31'))
```

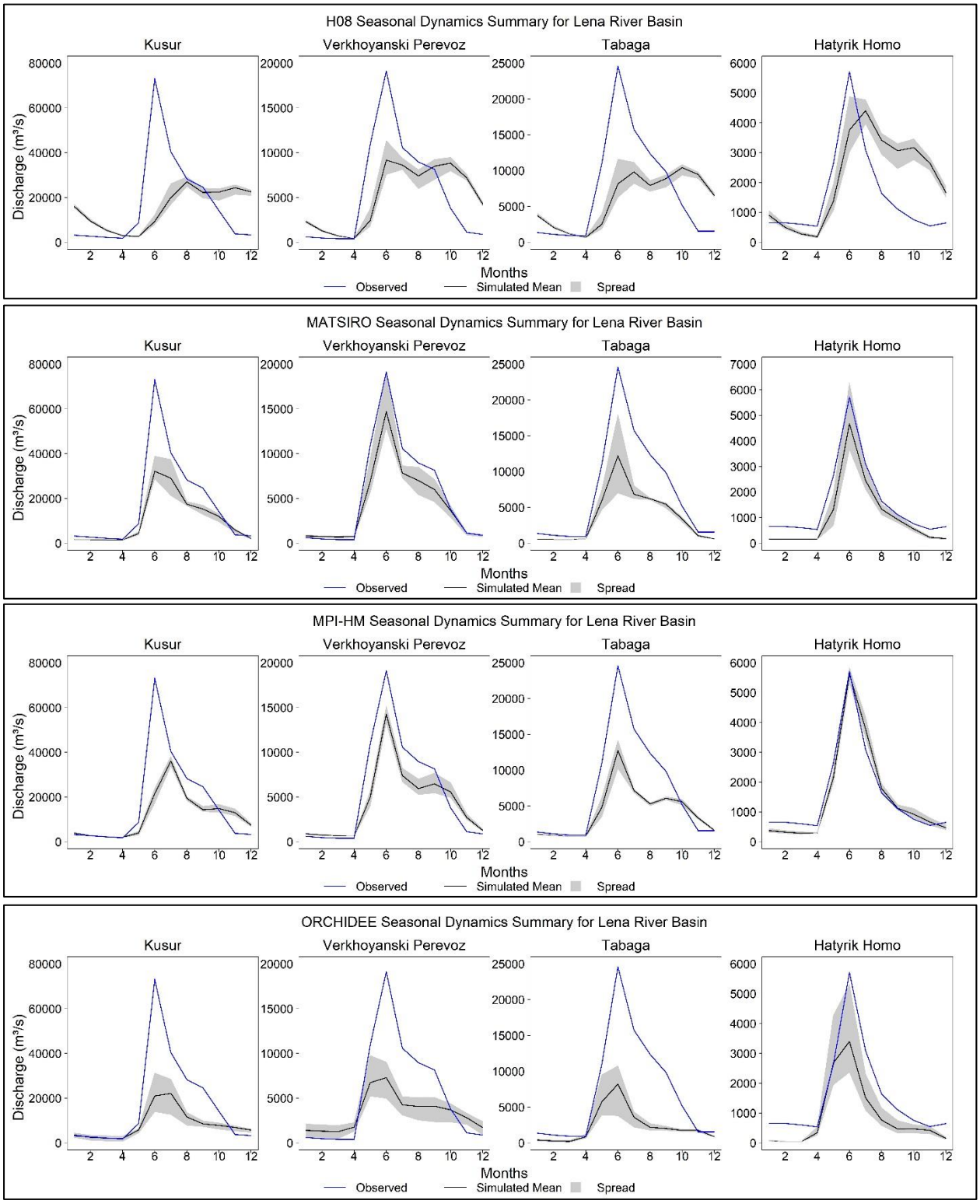
```
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1989-06-01' & Date <= '1989-08-31'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1990-08-01' & Date <= '1990-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1991-07-01' & Date <= '1991-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1992" & format.Date(Date,
"%m")=="08"))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1993-07-01' & Date <= '1993-08-31'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1994-07-01' & Date <= '1994-08-31'))
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1995" & format.Date(Date,
"%m")=="08"))
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1996" & format.Date(Date,
"%m")=="08"))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1997-07-01' & Date <= '1997-08-31'))
sim_dis_gs <- subset(sim_dis_gs, !(format.Date(Date, "%Y")=="1998" & format.Date(Date,
"%m")=="08"))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '1999-07-01' & Date <= '1999-09-30'))
sim_dis_gs <- subset(sim_dis_gs, !(Date >= '2000-07-01' & Date <= '2000-08-31'))

#creating a loop for selecting simulated data of 1 month from the time series and taking the avg
mon_sim_gs <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12){
  month_select_sim_gs <- sim_dis_gs$`Simulated
Discharge_gswp3`[which(month(sim_dis_gs$Date)==i)]
  mean_sim_gs <- mean(month_select_sim_gs)
  mon_sim_gs <- append(mon_sim_gs, mean_sim_gs)
}

#creating a loop for selecting observed data of 1 month from the time series and taking the avg
mon_obs_gs <- 0 #inilializing , your mean starts from here i.e 0
for (i in 1:12){
  month_select_obs_gs <- obs_dis$`Observed
Discharge_center`[which(month(obs_dis$Date)==i)]
  mean_obs_gs <- mean(month_select_obs_gs)
  mon_obs_gs <- append(mon_obs_gs, mean_obs_gs)
}

#Seasonal swe Dataframe
seasonal_swe_gs <- data.frame(months_name,  mon_obs_gs[2:13], mon_sim_gs[2:13])
names(seasonal_swe_gs)[1] <- "Months"
names(seasonal_swe_gs)[2] <- "Observed Discharge_center"
names(seasonal_swe_gs)[3] <- "Simulated Discharge_gswp3"

#repeat the above given codes by replacing gswp3 with princeton, watch and wfdei

#main data frame
main_center <- data.frame(seasonal_swe_gs, seasonal_swe_pr$`Simulated Discharge_princeton`,
seasonal_swe_wa$`Simulated Discharge_watch`, seasonal_swe_wf$`Simulated Discharge_wfdei`)
names(main_center)[1] <- "Date"
names(main_center)[2] <- "Observed Snow Water Equivalent"
names(main_center)[3] <- "Simulated Snow Water Equivalent_gswp3"
names(main_center)[4] <- "Simulated Snow Water Equivalent_princeton"
names(main_center)[5] <- "Simulated Snow Water Equivalent_watch"
names(main_center)[6] <- "Simulated Snow Water Equivalent_wfdei"
```

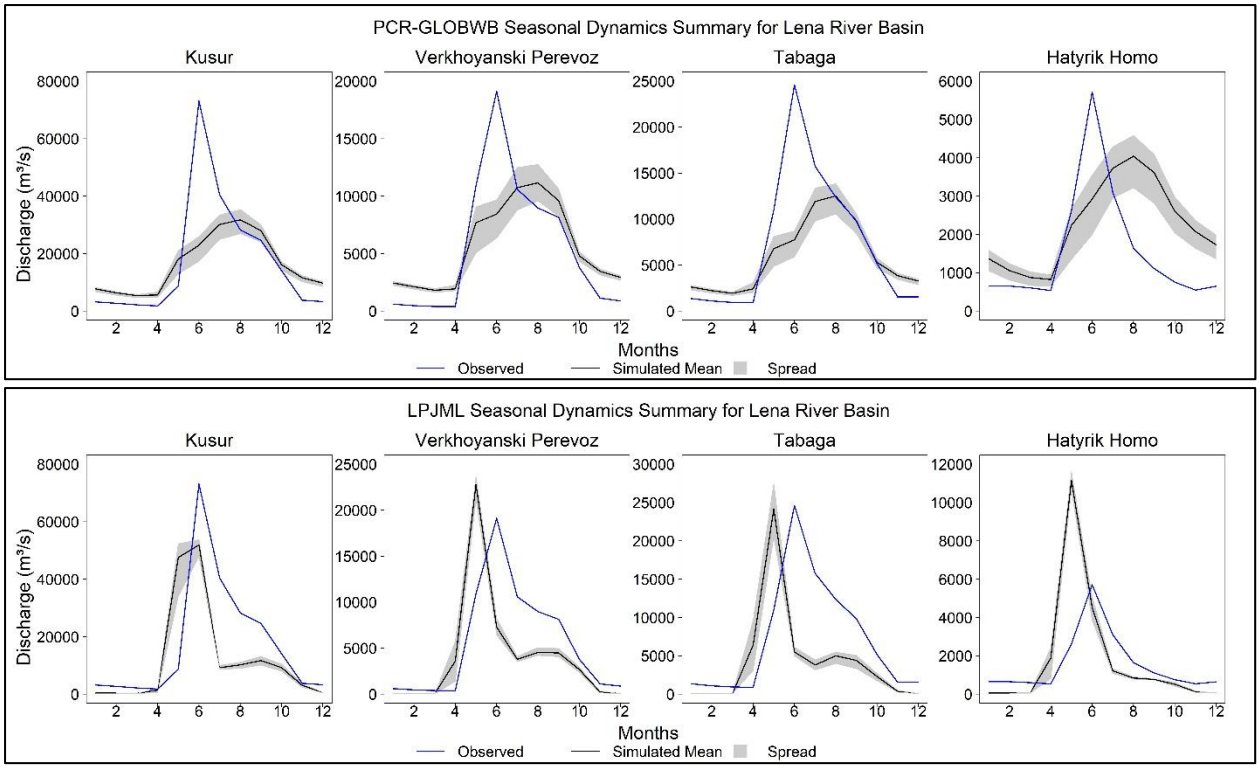
```
#data frame consisting of simulated data from 1971-1992 without the date column
sim_swe_c <-
  data.frame(mon_sim_gs[2:13], mon_sim_pr[2:13], mon_sim_wa[2:13], mon_sim_wf[2:13])
names(sim_swe_c)[1] <- "Simulated swe_gswp3"
names(sim_swe_c)[2] <- "Simulated swe_princeton"
names(sim_swe_c)[3] <- "Simulated swe_watch"
names(sim_swe_c)[4] <- "Simulated swe_wfdei"

# max, mean and min
max_c <- data.frame(apply(sim_swe_c, 1, FUN = max))
names(max_c)[1] <- "Maximum"

mean_c <- data.frame(apply(sim_swe_c, 1, FUN = mean))
names(mean_c)[1] <- "Mean"

min_c <- data.frame(apply(sim_swe_c, 1, FUN = min))
names(min_c)[1] <- "Minimum"

#creating a data frame to record summary of the simulated data
summary_c <- data.frame(max_c, mean_c, min_c)

#Simulated Mean Seasonal swe
sim_mean <- data.frame(months_name, mean_c)
names(sim_mean)[1] <- "Months"
names(sim_mean)[2] <- "Simulated Mean_matsiro"

write.table(na.omit(sim_mean),file = "lena_center_matsiro_swe_long term average simulated
mean data_1980-2000.csv", row.names=FALSE, sep=",")

#Obseverved Mean Seasonal swe
obs_mean <- data.frame(months_name, mon_obs_gs[2:13])
names(obs_mean)[1] <- "Months"
names(obs_mean)[2] <- "Observed swe"

write.table(na.omit(obs_mean),file = "lena_center_swe_long term average observed data_1980-
2000.csv", row.names=FALSE, sep=",")

#combining simulated, observed data with the summary (max, mean and min)
seasonal_swe_c <-
  cbind(main_center, summary_c)

# Snow water equivalent time series visualization
xdata <- c(months_name)
y1 <- c(seasonal_swe_gs$`Observed Discharge_center`)
y2 <- c(seasonal_swe_gs$`Simulated Discharge_gswp3`)
y3 <- c(seasonal_swe_pr$`Simulated Discharge_princeton`)
y4 <- c(seasonal_swe_wa$`Simulated Discharge_watch`)
y5 <- c(seasonal_swe_wf$`Simulated Discharge_wfdei`)

jpeg("Lena-center_matsiro_all climate forcing_Seasonal SWE.jpeg", width = 1350, height = 1200,
res = 240)
```

```
# plot the first curve by calling plot() function
# First curve is plotted
plot(xdata, y1, main= paste("Lena (Center) MATSIRO_all climate forcing datasets","\nLong-term
Average Snow Water Equivalent 1980-2000"),
    type="l", col="black", lty=1, lwd=2, xlab="Months", ylab="Snow Water Equivalent (mm)",
ylim= c(0, 155), cex.lab=1,las=1,cex.axis=1)

# Add second curve to the same plot by calling lines()
lines(xdata, y2, col="red",lty=1, lwd = 2)
lines(xdata, y3, col="green",lty=1, lwd = 2)
lines(xdata, y4, col="blue",lty=1, lwd = 2)
lines(xdata, y5, col="orange",lty=1, lwd = 2)

# Adding a legend inside box in graph coordinates.
# Note that the order of plots are maintained in the vectors of attributes.
legend(0.6, 158, legend=c("Observed", "gswp3", "princeton", "watch", "wfdei"),
      col=c("black","red","green","blue","orange"),
      lty=1, cex=0.97, box.lty=0, ncol=3, lwd = c(2,2,2,2,2))

dev.off()

jpeg("Lena-center_matsiro_Seasonal SWE_Simulated Mean.jpeg", width = 1350, height = 1200,
res = 240)

par(mar=c(4,5,3,2) + 0.1)
plot(xdata, y1, ylim = c(0, 120), type = "l", xlab = "", ylab = "", xaxt="n", cex.lab = 1, las = 1,
       cex.axis = 1, lwd = 2,col="blue")

mtext("MATSIRO - Lena-Center", side=3, line=1.5, cex=1, col="black")
mtext("Long term Average Snow Water Equivalent 1980-2000", side=3, line=0.5, cex=1,
col="black")
axis(1, at=1:12)
mtext("Snow Water Equivalent (mm)", side = 2, line = 2.5)
mtext("Months", side = 1, line = 2)

lines(xdata, y7, col="red",lty=1, lwd = 2)

legend(7,122,legend = c("Observed", "Simulated Mean"), pt.cex = 2, col = c( "blue", "red"),
          lty = c(1,1), cex = 1, lwd = c(2,2), box.lty = 0, ncol = 1)

dev.off()
```

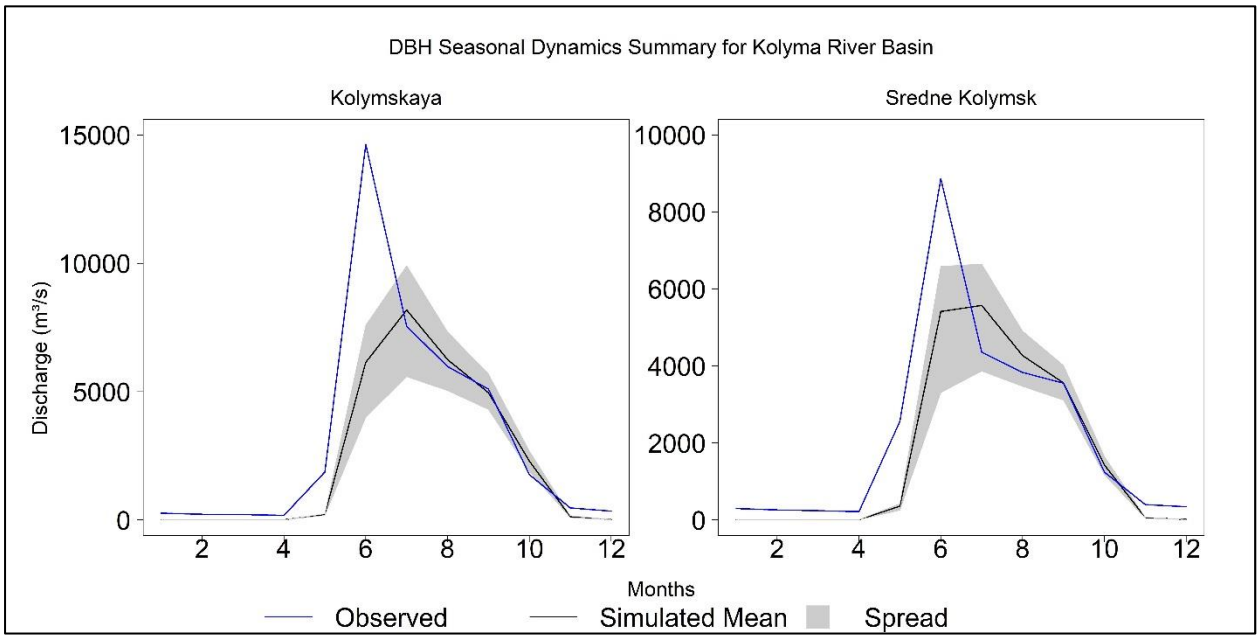
Appendix-2 Extra Figures and Tables

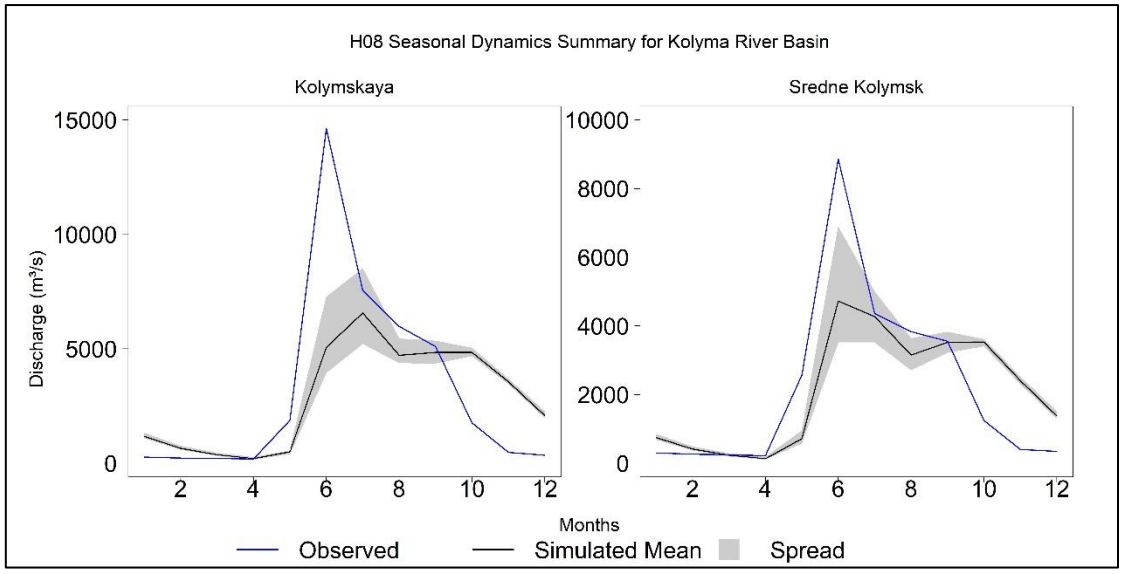
Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by remaining 6 models at 4 gauging stations of Lena Basin

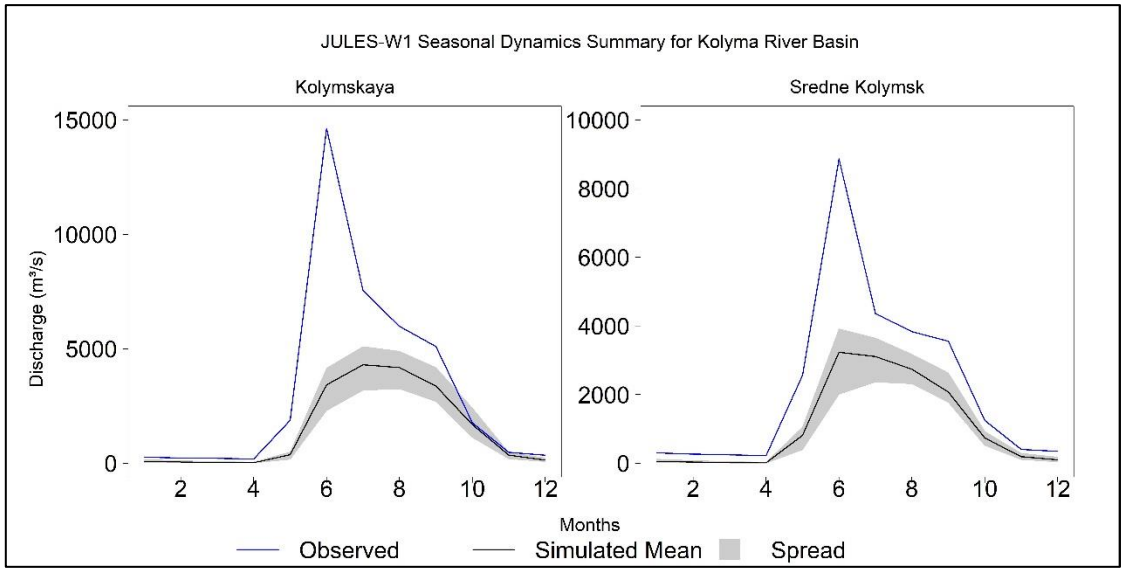


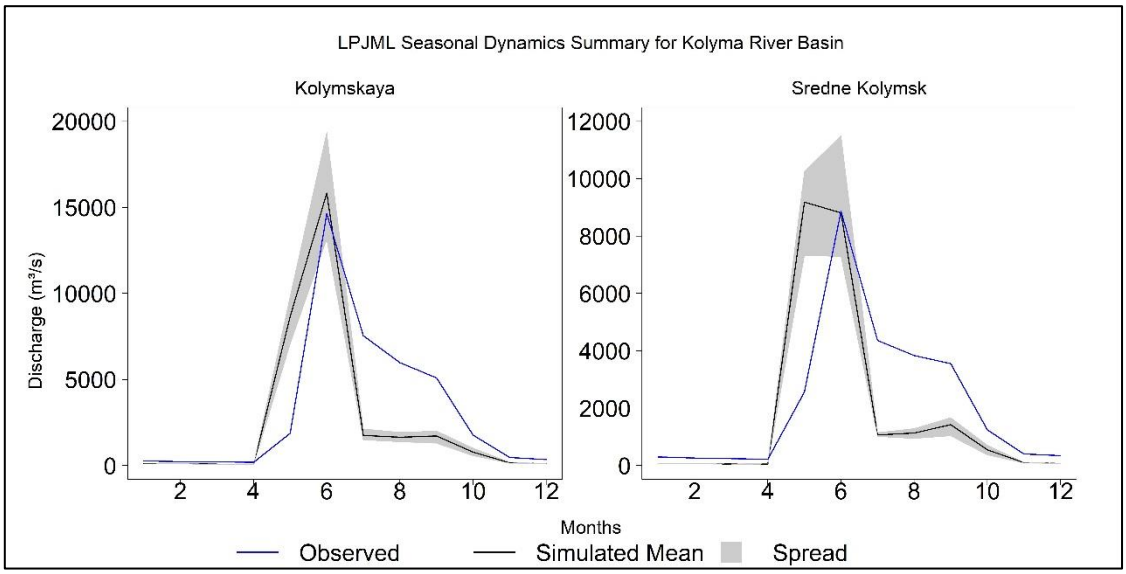


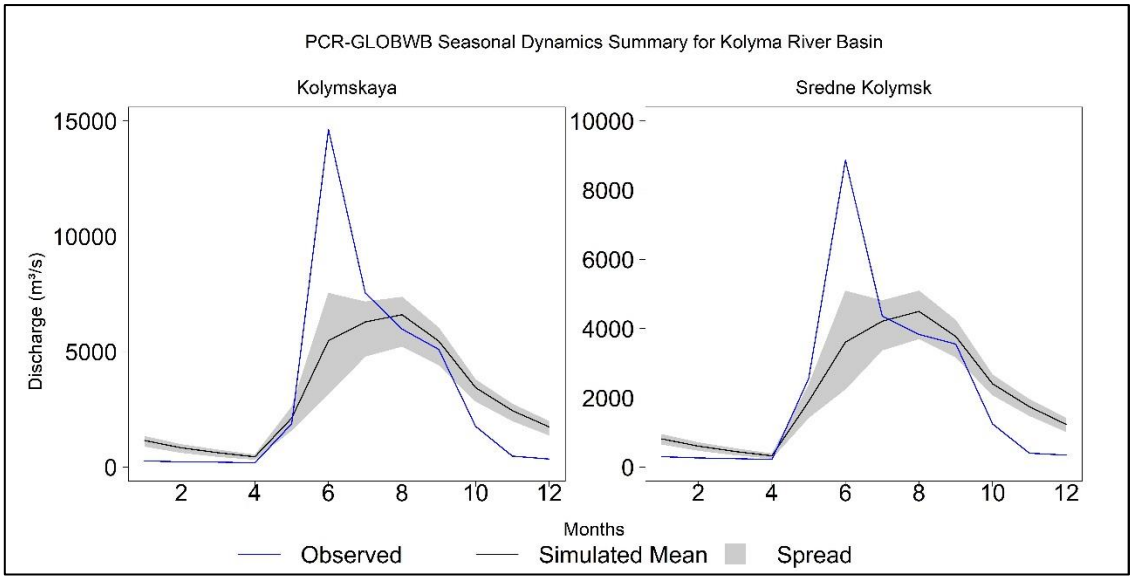
Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by remaining 6 models at 2 gauging stations of Kolyma Basin

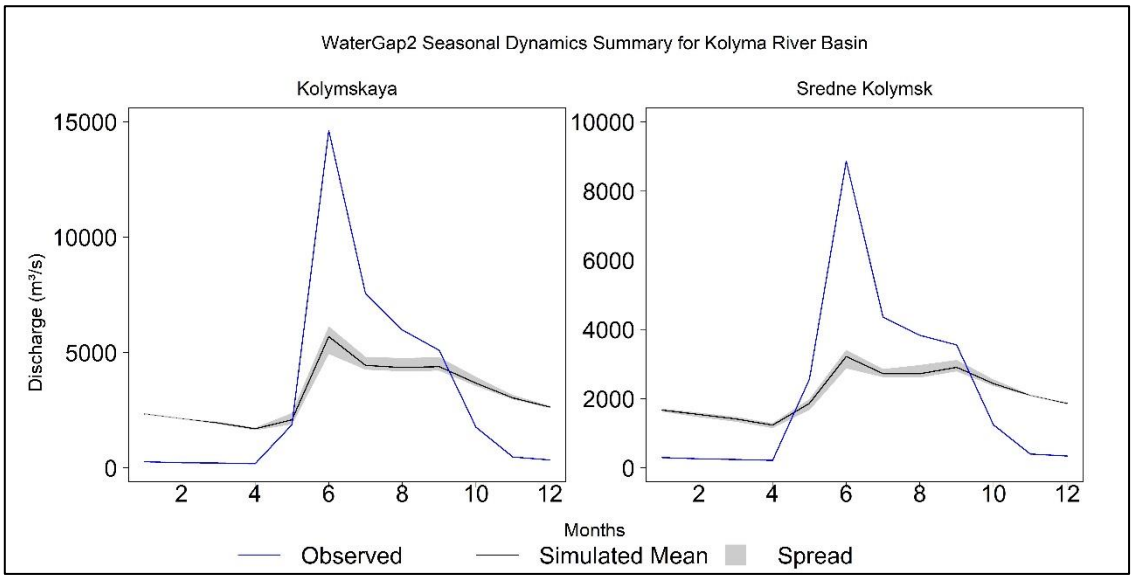



H08 Seasonal Dynamics Summary for Kolyma River Basin
Kolymskaya
Sredne Kolymsk
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

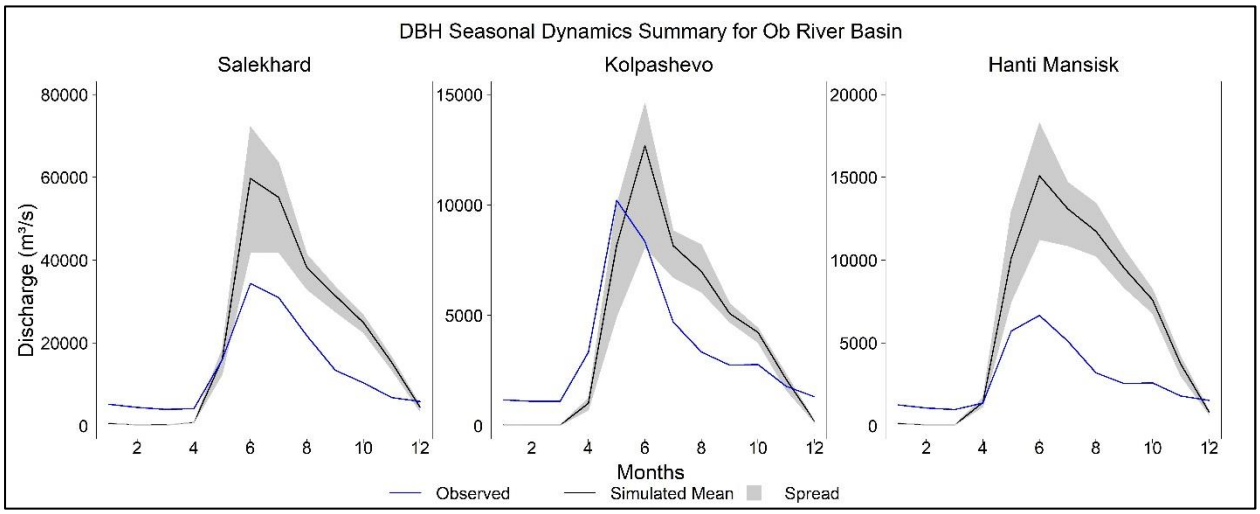

JULES-W1 Seasonal Dynamics Summary for Kolyma River Basin
Kolymskaya
Sredne Kolymsk
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

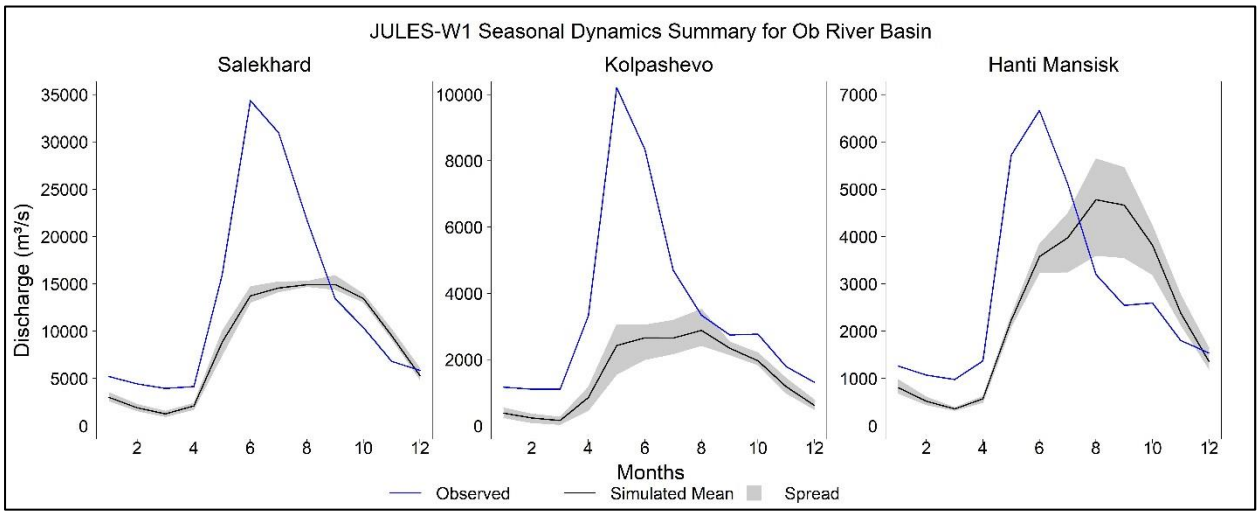

LPJML Seasonal Dynamics Summary for Kolyma River Basin
Kolymskaya
Sredne Kolymsk
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

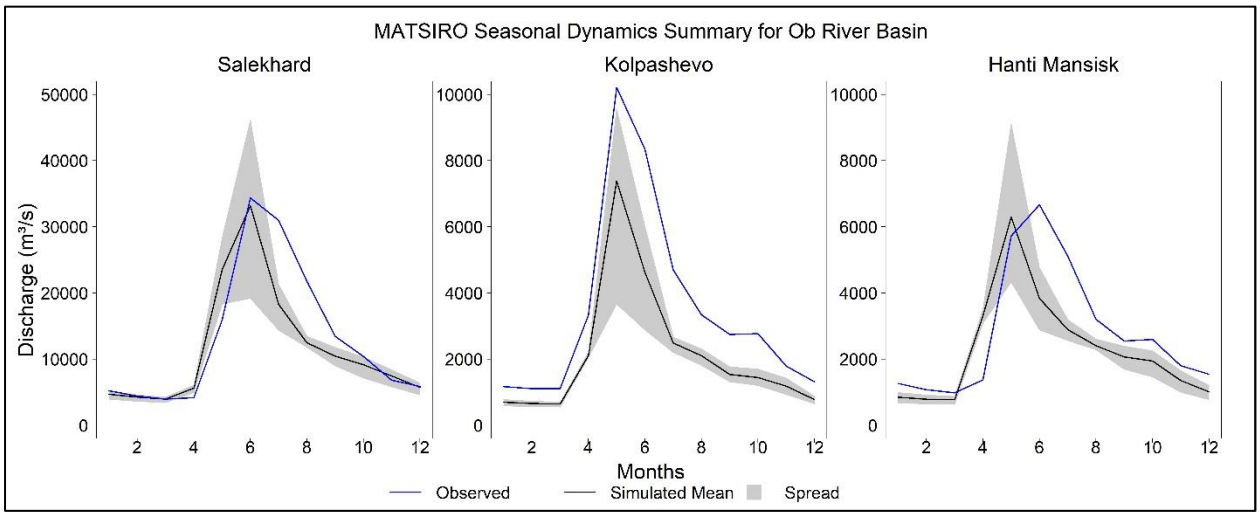

PCR-GLOBWB Seasonal Dynamics Summary for Kolyma River Basin
Kolymskaya
Sredne Kolymsk
Discharge (m³/s)
15000
10000
5000
0
10000
8000
6000
4000
2000
0
2
4
6
8
10
12
Months
Observed
Simulated Mean
Spread

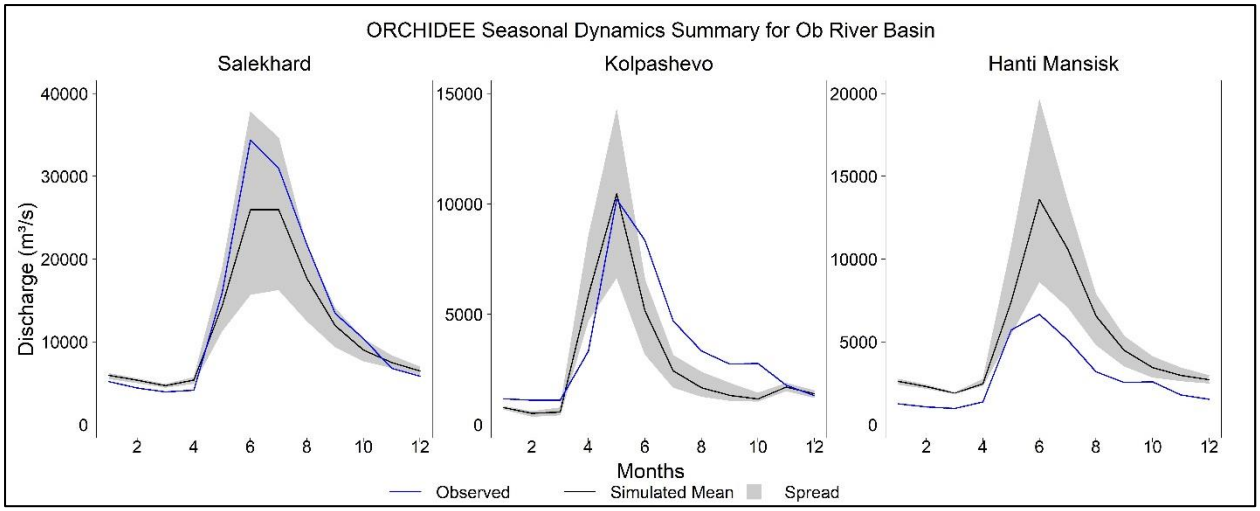

WaterGap2 Seasonal Dynamics Summary for Kolyma River Basin
Kolymskaya
Sredne Kolymsk
Discharge (m³/s)
15000
10000
5000
0
10000
8000
6000
4000
2000
0
2
4
6
8
10
12
Months
Observed
Simulated Mean
Spread

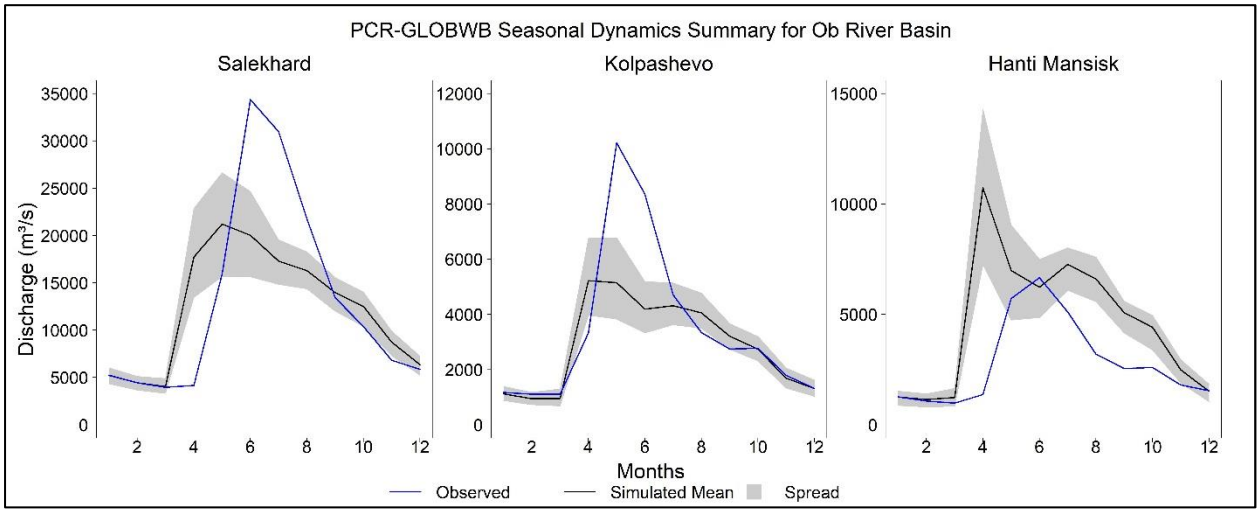
Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by remaining 6 models at 3 gauging stations of Ob Basin

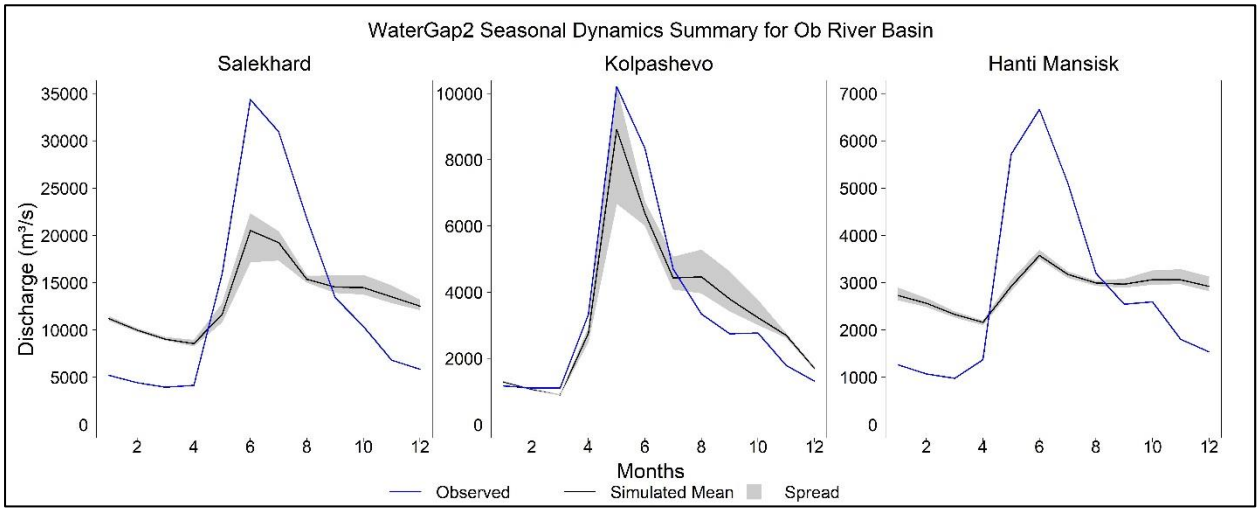




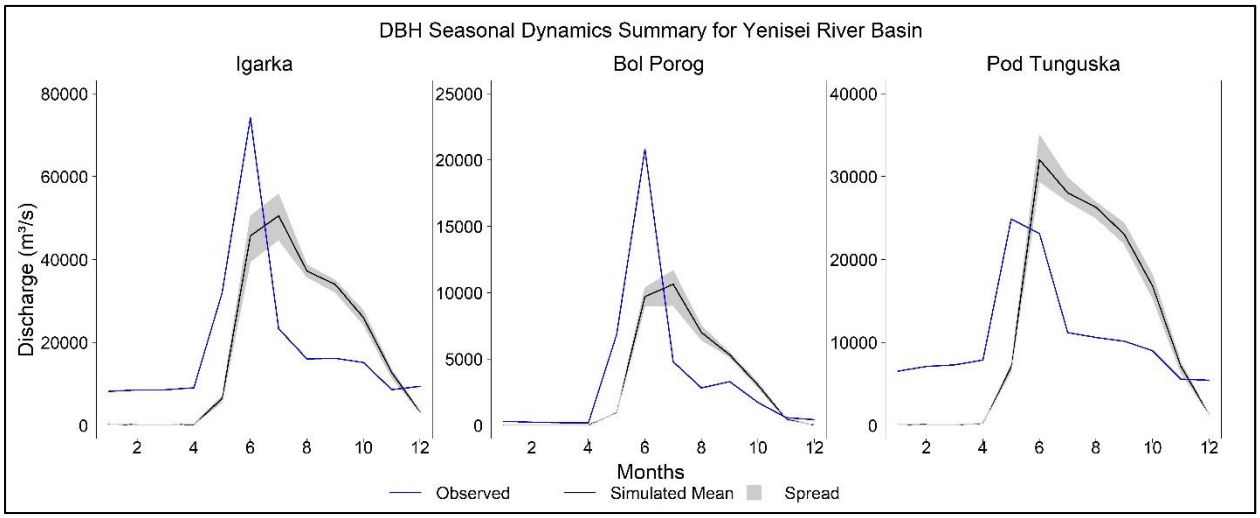


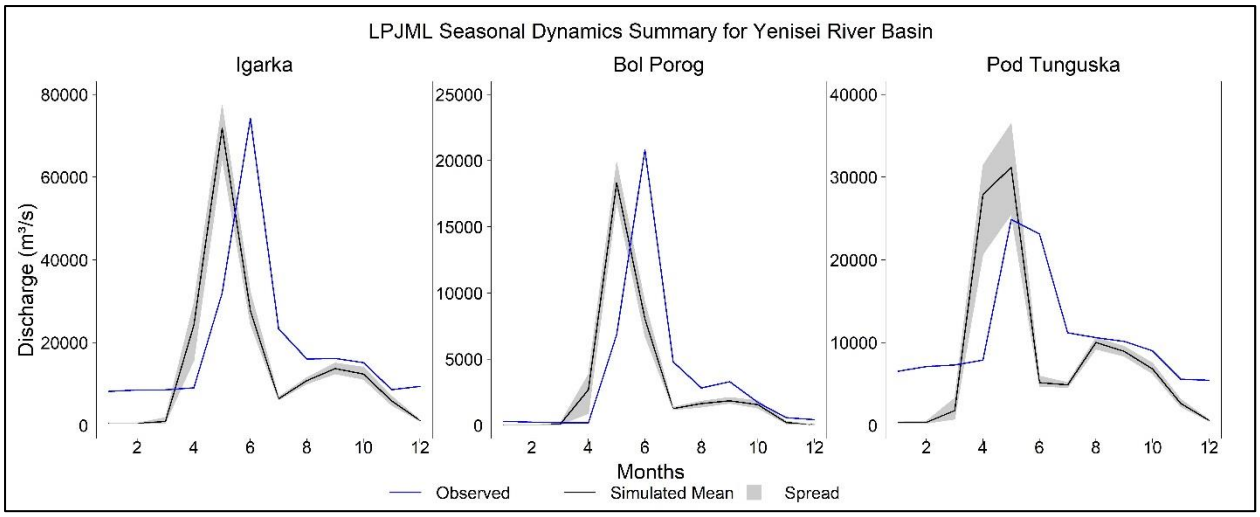

ORCHIDEE Seasonal Dynamics Summary for Ob River Basin
Salekhard
Kolpashevo
Hanti Mansisk
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

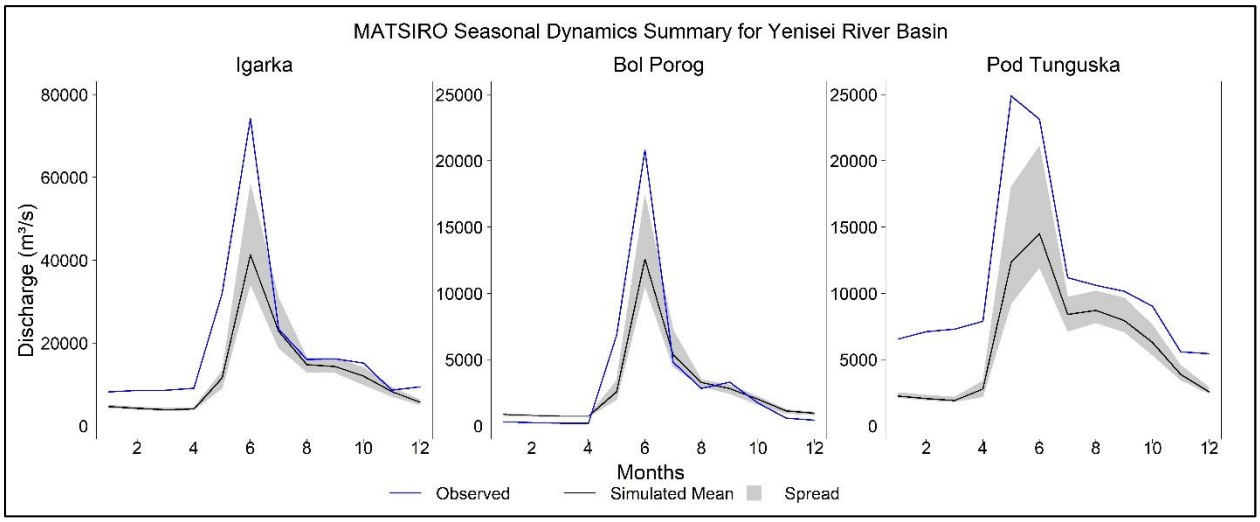

PCR-GLOBWB Seasonal Dynamics Summary for Ob River Basin
Salekhard
Kolpashevo
Hanti Mansisk
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

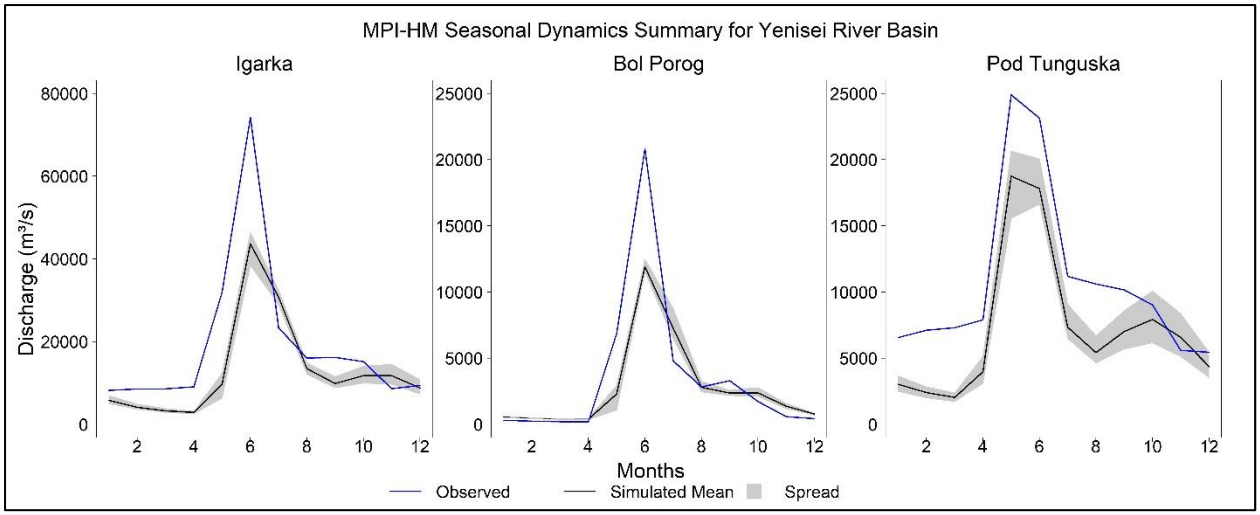

WaterGap2 Seasonal Dynamics Summary for Ob River Basin
Salekhard
Kolpashevo
Hanti Mansisk
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

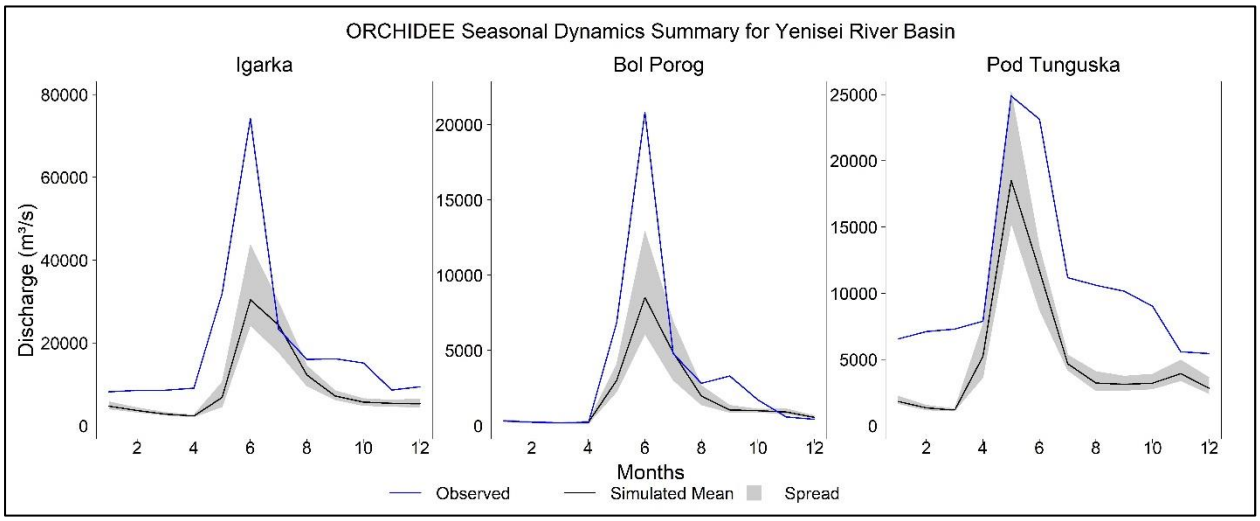
Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by remaining 6 models at 3 gauging stations of Yenisei Basin

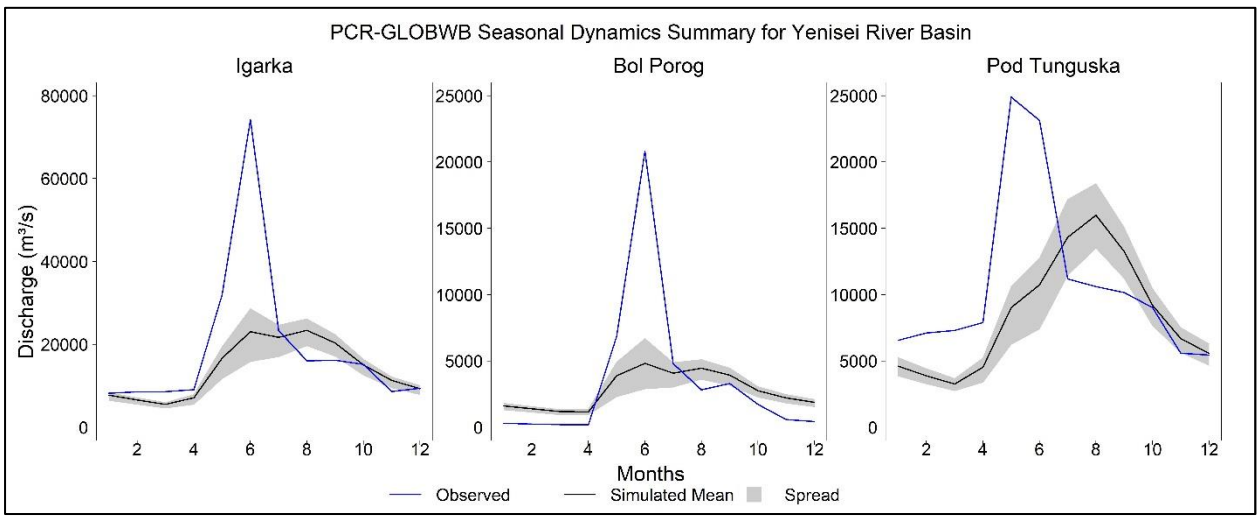




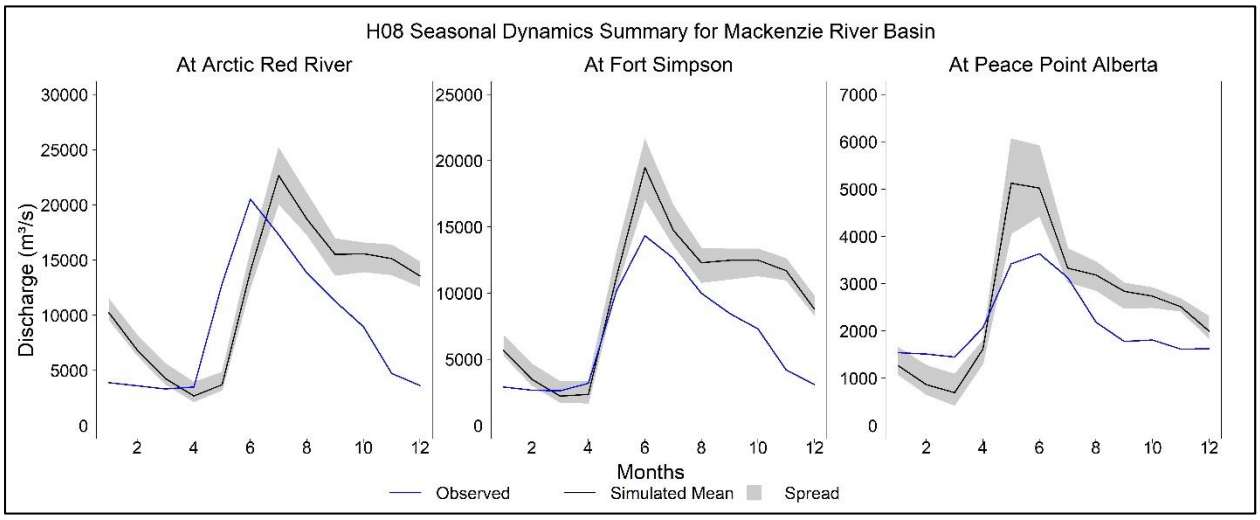


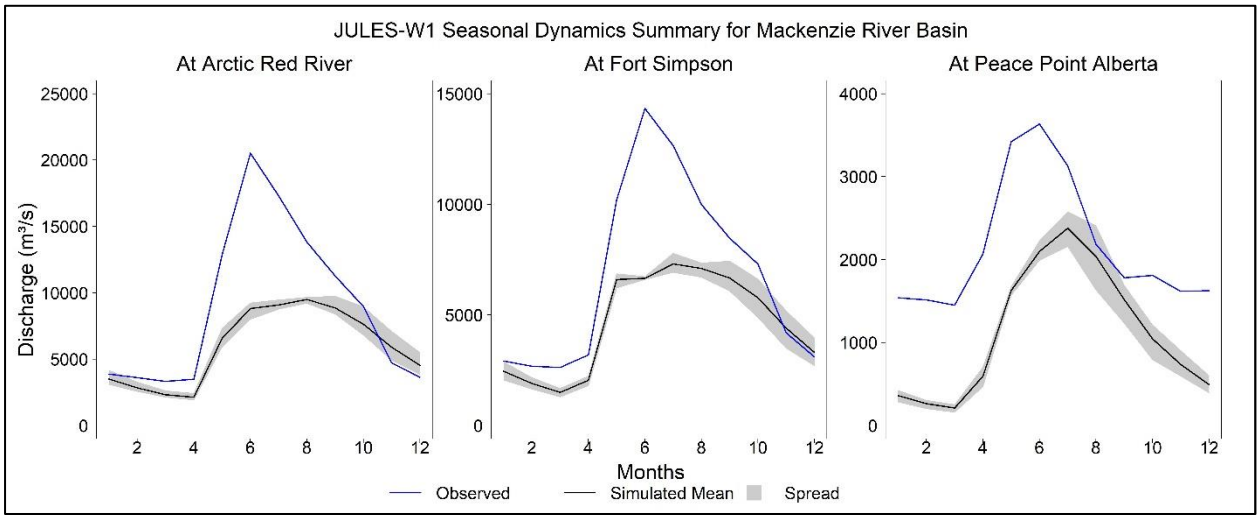

MPI-HM Seasonal Dynamics Summary for Yenisei River Basin
Igarka
Bol Porog
Pod Tunguska
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

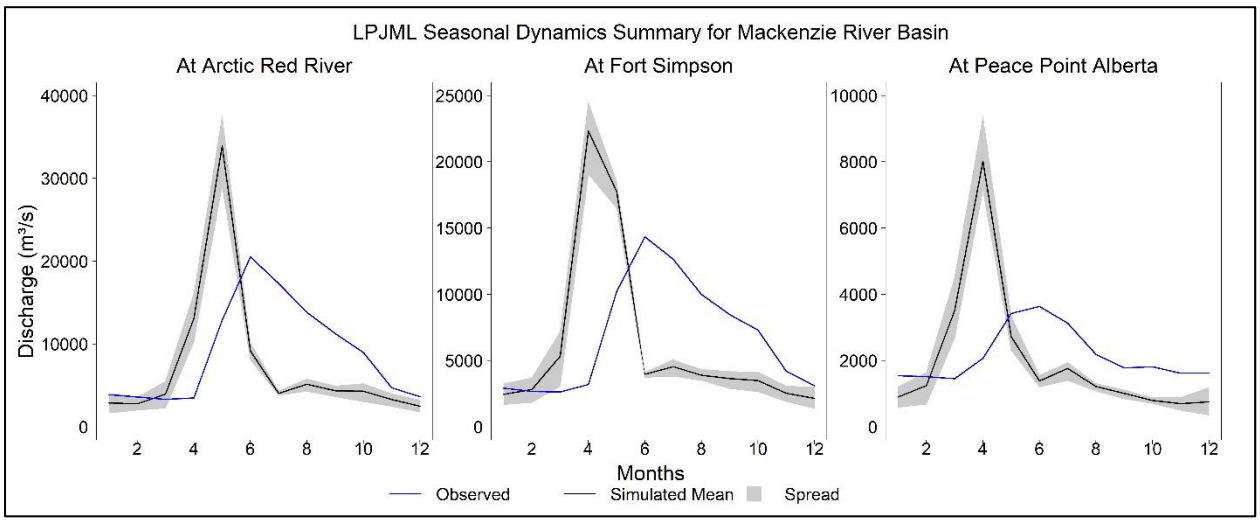

ORCHIDEE Seasonal Dynamics Summary for Yenisei River Basin
Igarka
Bol Porog
Pod Tunguska
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

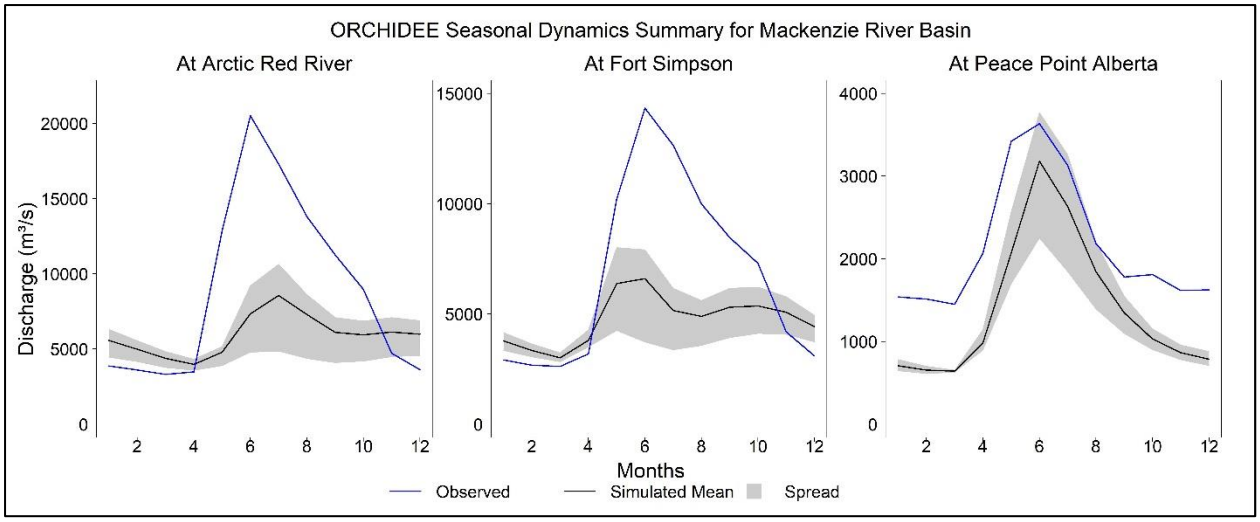

PCR-GLOBWB Seasonal Dynamics Summary for Yenisei River Basin
Igarka
Bol Porog
Pod Tunguska
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

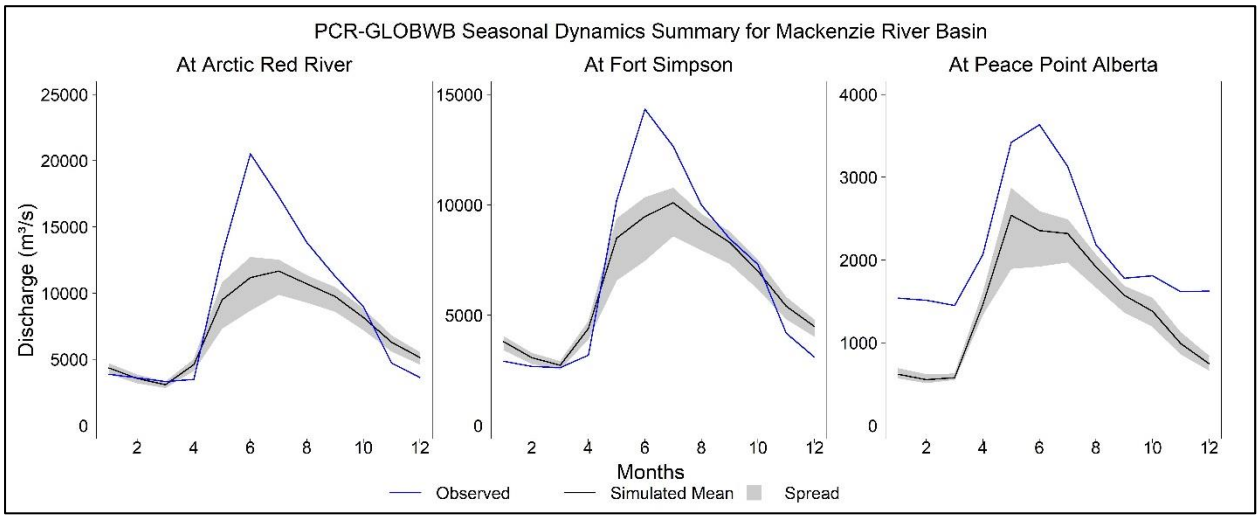
Multiple plots showing the comparison between observed and simulated mean seasonal discharge modelled by remaining 6 models at 3 gauging stations of Mackenzie Basin

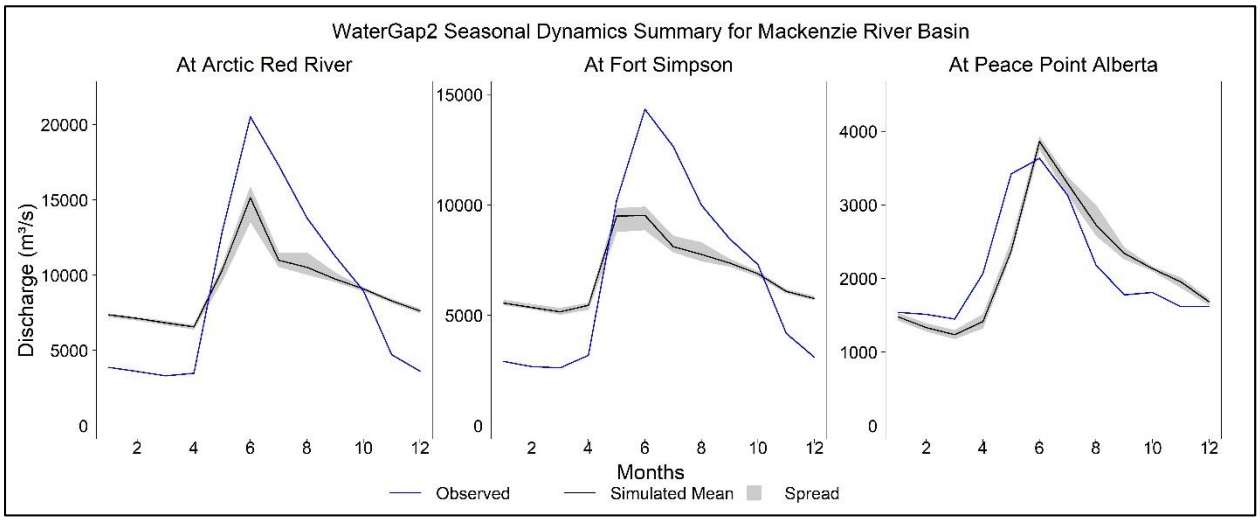




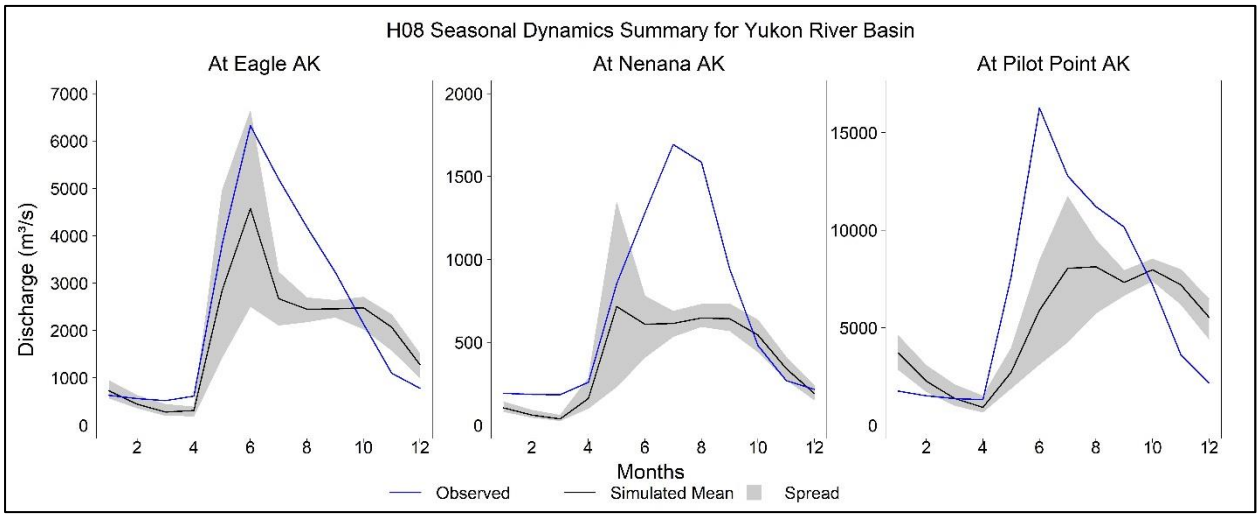


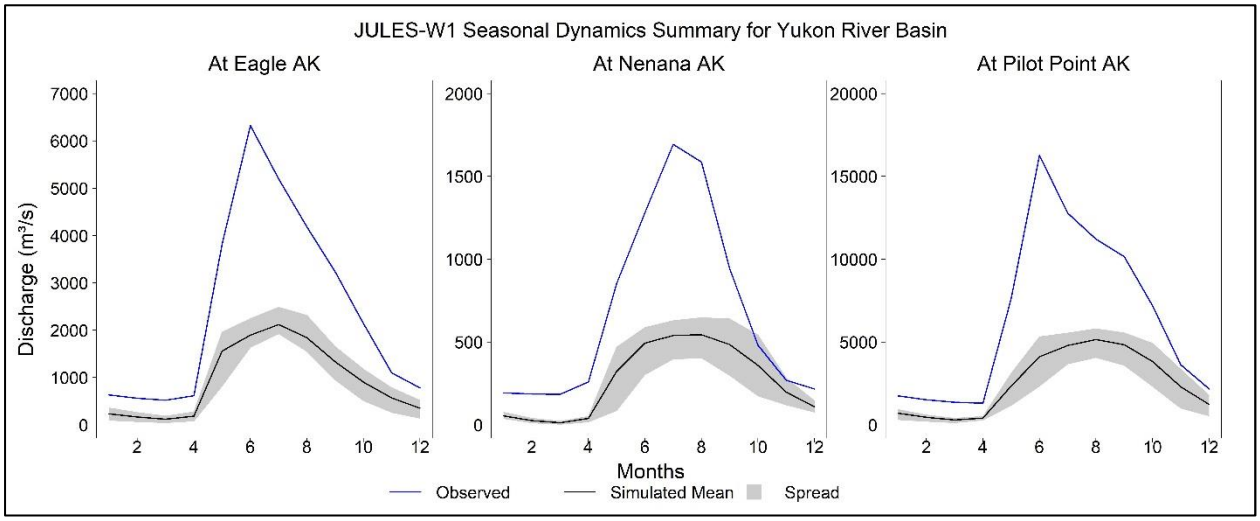

ORCHIDEE Seasonal Dynamics Summary for Mackenzie River Basin
At Arctic Red River
At Fort Simpson
At Peace Point Alberta
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

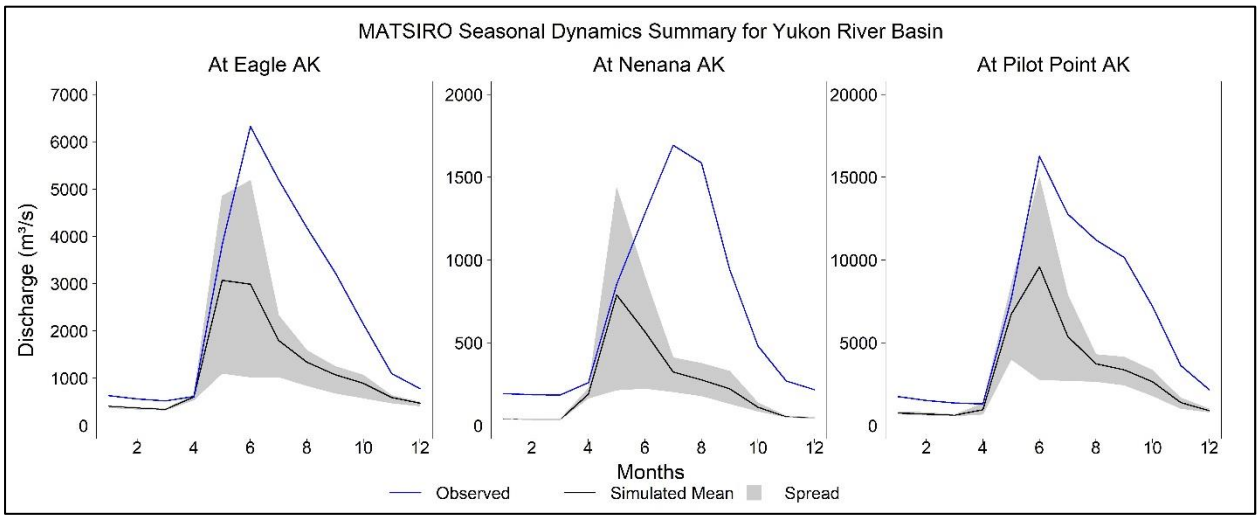

PCR-GLOBWB Seasonal Dynamics Summary for Mackenzie River Basin
At Arctic Red River
At Fort Simpson
At Peace Point Alberta
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

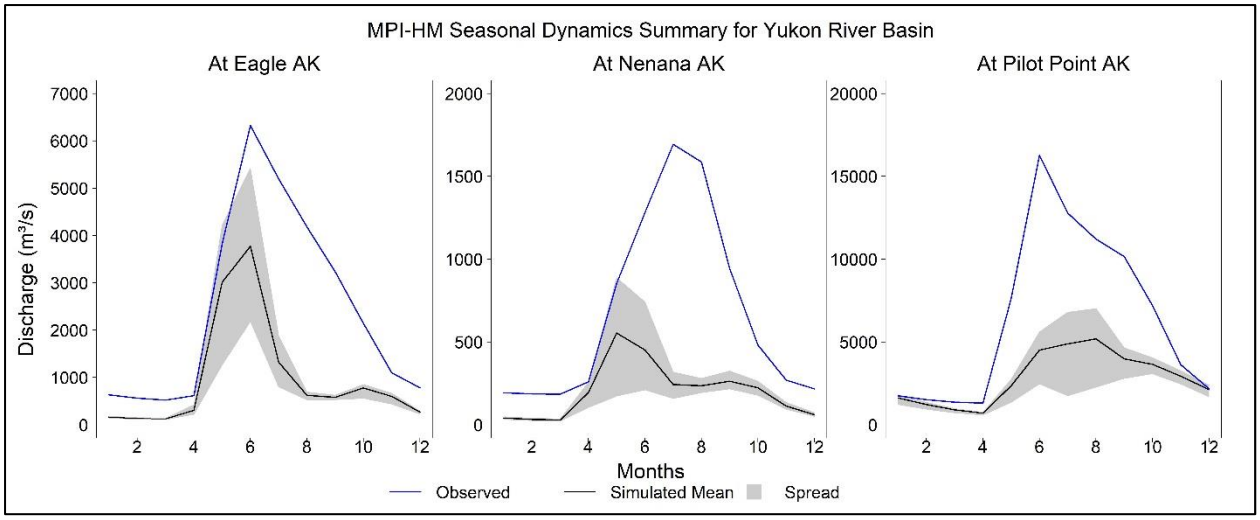

WaterGap2 Seasonal Dynamics Summary for Mackenzie River Basin
At Arctic Red River
At Fort Simpson
At Peace Point Alberta
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

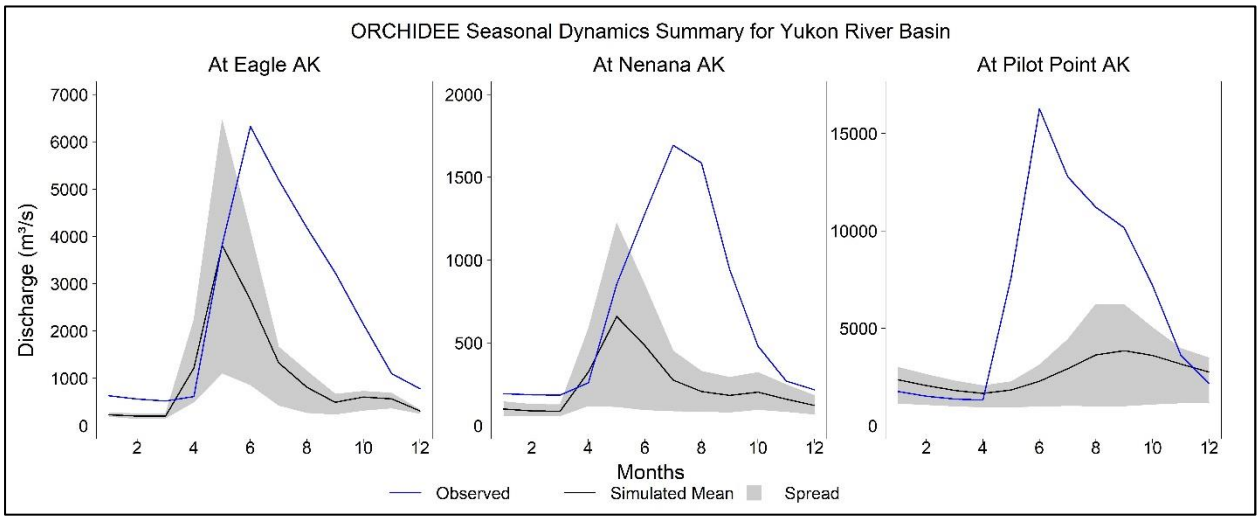
Multiple plots showing the comparison between observed and simulated mean seasonal dynamics of discharge modelled by remaining 6 models at 3 gauging stations of Yukon Basin

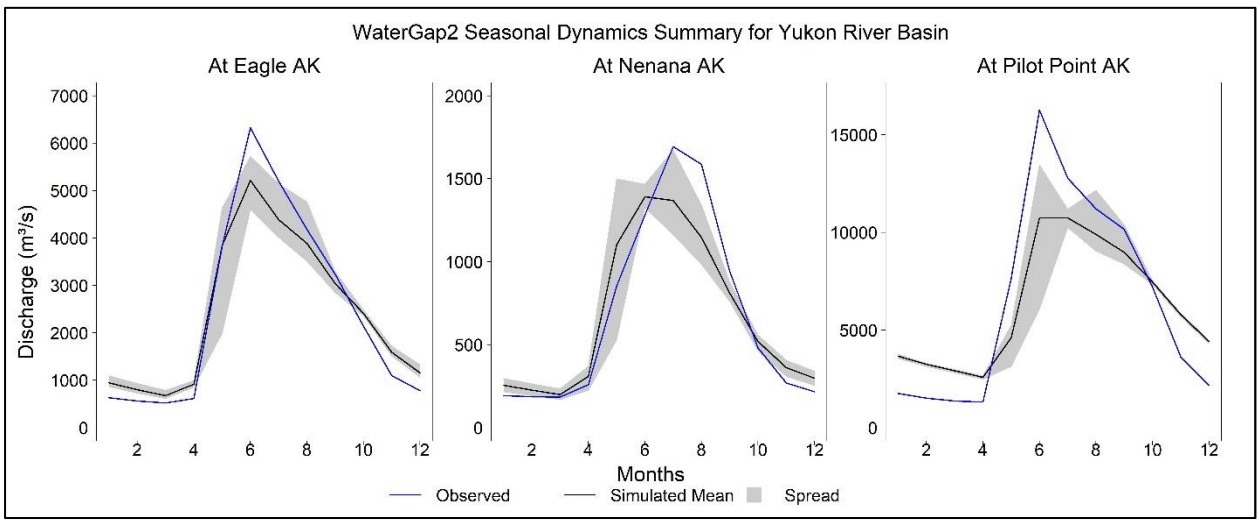







MPI-HM Seasonal Dynamics Summary for Yukon River Basin
At Eagle AK
At Nenana AK
At Pilot Point AK
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread


ORCHIDEE Seasonal Dynamics Summary for Yukon River Basin
At Eagle AK
At Nenana AK
At Pilot Point AK
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread


WaterGap2 Seasonal Dynamics Summary for Yukon River Basin
At Eagle AK
At Nenana AK
At Pilot Point AK
Discharge (m³/s)
Months
Observed
Simulated Mean
Spread

Efficiency criteria to assess simulated and observed discharge at Verkhoyanski Pervoz, Lena

LENA - VERKHOYANSKI PEREVOZ									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.91	0.49	0.46	0.73	0.62	0.82	0.38	0.13	0.1
princeton	0.9	0.49	0.3	0.72	0.61	0.84	0.16	0.15	0.17
watch	0.93	0.41	0.3	0.71	0.45	0.79	0.3	0.0046	0.15
wfdei	0.94	0.44	0.22	0.68	0.6	0.77	0.3	-0.016	0.18

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-7.4	-2.3	-8.4	-22.6	10.2	-23.1	-45.5	-23.2	-63.3
princeton	-6.9	1.2	-2.6	-13.8	14.4	-13.9	-49.3	-28.5	-46.9
watch	-7.9	6.8	-6.3	-22.3	-15	-24.4	-27.7	-21.8	-54.9
wfdei	-7.9	2.6	-9	-26.2	0.8	-26.4	-30.4	-23.3	-54.4

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.96	0.54	0.53	0.81	0.66	0.86	0.44	0.16	0.051
princeton	0.94	0.53	0.33	0.79	0.65	0.88	0.17	0.11	0.15
watch	0.97	0.45	0.34	0.77	0.46	0.82	0.44	0.0096	0.11
wfdei	0.97	0.49	0.25	0.74	0.64	0.8	0.43	-0.021	0.14

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-10.6	10	-42.7	-30.1	-35.3	-11.2	-47.8	-1.5	-63.3
princeton	-10	12	-38.9	-27.2	-29	-27.9	-72.9	6	-35.5
watch	-12.5	20	-38.6	-36.3	-51.6	-32.2	-69	7	-49
wfdei	-11.2	16	-44.4	-37.7	-41.4	-35.3	-68.9	8	-48.2

Efficiency criteria to assess simulated and observed discharge at Tabaga, Lena

LENA - TABAGA									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.76	0.5	0.37	0.49	0.54	0.67	0.23	-0.27	-0.23
princeton	0.71	0.5	0.057	0.35	0.52	0.28	-0.28	-0.073	-0.23
watch	0.84	0.55	0.17	0.52	0.33	0.48	0.081	-0.35	-0.18
wfdei	0.85	0.57	0.1	0.48	0.48	0.47	0.081	-0.31	-0.093

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.2	-16.8	-15.4	-40.6	-9	-40.7	-58.2	-38	-78.2
princeton	-0.7	-26.1	-21.4	-45.4	-14.2	-55.3	-77.4	-50.7	-77
watch	-0.9	-8.3	-15.3	-38.7	-32.1	-50.1	-67.8	-33.8	-75.5
wfdei	0.6	-10.9	-16.8	-42	-18	-50.2	-67	-34.9	-71.4

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.8	0.51	0.41	0.54	0.55	0.71	0.24	-0.25	-0.33
princeton	0.72	0.51	0.045	0.37	0.52	0.25	-0.34	-0.13	-0.33
watch	0.87	0.56	0.17	0.56	0.32	0.47	0.067	-0.41	-0.28
wfdei	0.88	0.59	0.091	0.52	0.5	0.46	0.068	-0.36	-0.18

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-20.4	-6.4	-51.3	-51.5	-43.8	-31.8	-53.4	-18.5	-76.7
princeton	-17.6	-18.9	-53.1	-60.1	-45.1	-62.7	-83.5	-25.4	-72.3
watch	-19.5	3	-47.5	-48.9	-58.7	-52.3	-66.5	-0.7	-73.3
wfdei	-16.9	-0.3	-51.8	-51.2	-50.4	-53.9	-66.6	-3.2	-70.9

Efficiency criteria to assess simulated and observed discharge at Hatyrik Homo, Lena

LENA - HATYRIK HOMO									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.53	-0.79	0.33	0.72	-0.011	0.71	0.58	-1.6	-0.14
princeton	0.54	-0.64	-0.018	0.78	-0.13	0.77	0.27	-1.2	-0.055
watch	0.53	-1.3	-0.075	0.74	0.16	0.54	0.29	-2	-0.06525
wfdei	0.55	-1.2	-0.09	0.73	0.087	0.68	0.37	-1.9	-0.00076

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	3.1	33.3	36.3	-3.6	62.4	-21.9	-14.2	16.4	-71.4
princeton	9.5	35.9	42.5	1.7	65.4	-28.8	-52.6	8.1	-66.7
watch	-2.9	51.1	31.2	-5.3	9.8	-46.6	-56.3	18.8	-67.1
wfdei	-0.7	48	34.1	-8.3	42.1	-36.6	-52.4	18.4	-63.3

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.58	-1.1	0.4	0.92	-0.18	0.89	0.81	-1.9	-0.31
princeton	0.58	-1	-0.086	0.93	-0.35	0.87	0.34	-1.6	-0.22
watch	0.59	-1.9	-0.12	0.98	0.099	0.61	0.42	-2.4	-0.22
wfdei	0.6	-1.7	-0.16	0.97	0.015	0.77	0.52	-2.2	-0.14

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-28.7	73	2	12	-19	16	12	106	-71.6
princeton	-27.4	75	-0.7	9	-17.8	-15.8	-47	101	-66.7
watch	-35	100	-4.6	4	-42.5	-32.3	-44.3	110	-67.8
wfdei	-33	96	-13.6	0	-29.2	-21.2	-37.4	107	-65.2

Efficiency criteria to assess simulated and observed discharge at Sredne Kolymsk, Kolyma

KOLYMA - SREDNE KOLYMSK									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.37	0.68	0.65	0.63	0.66	0.72	0.53	0.06	0.51
princeton	0.34	0.46	0.37	0.57	0.48	0.56	-0.2	0.38	0.22
watch	0.36	0.67	0.49	0.76	0.33	0.85	0.25	0.054	0.42
wfdei	0.39	0.68	0.39	0.75	0.51	0.83	0.18	0.1	0.52

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-5.6	-15.9	2.9	-16.6	13.2	-15.9	-34.5	-2.8	-44.1
princeton	-5.7	-41.4	-16	-37.9	-3.9	-47.7	-72.6	-28.9	-63.5
watch	-5.9	-6.6	-2	-21	-24.2	-19.3	-19.7	-8.7	-49.9
wfdei	-5.3	-18.2	-9.1	-24.2	-2.2	-24.7	-20.9	-9	-42.1

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.37	0.79	0.75	0.82	0.7	0.82	0.65	0.0064	0.58
princeton	0.32	0.51	0.44	0.66	0.53	0.59	-0.28	0.32	0.19
watch	0.35	0.78	0.58	0.87	0.36	0.94	0.39	-0.057	0.45
wfdei	0.38	0.77	0.45	0.85	0.57	0.9	0.34	0.001	0.59

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-75.1	-6	-20.3	-9.7	-30.2	18	-48.2	49	-44.8
princeton	-73.9	-37.7	-45.1	-41.6	-38.8	-48.1	-90.9	4	-62.1
watch	-76.7	3	-30.9	-19.2	-53.5	-13	-73.8	30	-49.9
wfdei	-75.2	-10.7	-42.3	-20.3	-42.3	-18.3	-76.3	31	-42.7

Efficiency criteria to assess simulated and observed discharge at Kolpashevo, Ob

OB - KOLPASHEVO									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.87	0.028	0.54	0.78	0.57	0.75	0.34	-2.6	0.031
princeton	0.68	0.42	0.59	0.65	0.34	0.12	0.32	-2	-0.34
watch	0.88	0.17	0.72	0.79	0.28	0.68	0.67	-3.5	-0.08
wfdei	0.9	0.21	0.69	0.8	0.47	0.7	0.67	-3.6	0.07

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-1	23.4	12.2	-21.5	2.5	-32.9	3.7	7.4	-52
princeton	-0.7	-12	-11.3	-36.3	-26.9	-54.7	-40.1	-20.8	-65.6
watch	-1.4	27.8	11.2	-9.8	-31.9	-34.2	-22.3	5.3	-56.4
wfdei	0	26.1	8.5	-14.9	-9.7	-34	-26.1	7.6	-51.6

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.91	0.018	0.61	0.82	0.61	0.79	0.43	-2.6	-0.025
princeton	0.71	0.42	0.62	0.65	0.35	0.073	0.34	-2.1	-0.43
watch	0.92	0.13	0.77	0.83	0.29	0.7	0.72	-3.7	-0.14
wfdei	0.94	0.17	0.72	0.83	0.58	0.72	0.72	-3.7	0.023

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-19.8	52	19	6	-29	-8	40	78	-57.9
princeton	-31.3	1	-45.6	-32.7	-48.3	-66.5	-33.1	63	-69.3
watch	-13.8	59	4	9	-56.8	-28.4	5	107	-63.9
wfdei	-10.9	59	-20.2	7	-39.5	-25.8	-6.3	110	-64.6

Efficiency criteria to assess simulated and observed discharge at Hanti Mansisk, Ob

OB - HANTI MANSISK									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.31	-7.1	-4.3	0.41	-3.9	0.09	-5.8	-27	0.077
princeton	0.23	-2.5	-0.54	0.73	-1.1	0.26	0.24	-16	0.18
watch	0.29	-8.7	-2.5	-0.38	-0.2	0.5	-2.2	-28	0.17
wfdei	0.31	-5.2	-0.8	0.5	-2.4	0.43	-0.75	-25	0.25

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.2	130	101	20.4	94.8	-12.9	129	130	-7.1
princeton	4	78.3	67.5	11.9	55.4	-29.3	39.5	81.5	-6.9
watch	0.1	150	103	60.5	22.3	-11.8	90.1	139	-13.9
wfdei	1.8	114	71.7	22.4	73.9	-23.1	65.7	123	-27.7

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.27	-7.9	-4.5	0.55	-4.3	0.34	-6.2	-30	0.082
princeton	0.16	-2.9	-0.61	0.82	-1.1	0.34	0.54	-19	0.16
watch	0.26	-10	-2.8	-0.4	-0.16	0.62	-2.3	-32	0.16
wfdei	0.29	-5.8	-0.82	0.66	-2.6	0.52	-0.64	-29	0.25

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-79.4	207	135	38	102	26	181	406	2
princeton	-80.4	134	5	-14.8	43	-42.1	9	308	-9.6
watch	-81.3	237	79	53	18	-16	110	420	-13.6
wfdei	-79.8	187	-15.3	11	75	-32.4	67	394	-33.3

Efficiency criteria to assess simulated and observed discharge at Bol Porog, Yenisei

YENISEI - BOL POROG									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.8	0.39	0.61	0.66	0.41	0.83	0.72	0.096	0.34
princeton	0.77	0.41	0.34	0.63	0.31	0.69	0.34	0.3	0.25
watch	0.78	0.38	0.37	0.66	0.081	0.65	0.45	0.034	0.29
wfdei	0.8	0.37	0.28	0.64	0.22	0.64	0.39	0.12	0.28

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	1.6	-8.6	0.4	-16.6	-6.1	-7.7	-26.3	-9.2	-40.7
princeton	1.2	-19.4	-7.5	-22.1	-12.7	-21.1	-55.9	-23.1	-45.1
watch	0.1	-6.1	-6.2	-22.3	-41	-22.9	-47.5	-14.2	-44.3
wfdei	1.4	-13.6	-12.3	-26.7	-24.9	-28.2	-53.8	-16	-47.2

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.83	0.43	0.65	0.73	0.4	0.9	0.79	0.13	0.34
princeton	0.79	0.44	0.35	0.68	0.3	0.71	0.36	0.29	0.23
watch	0.8	0.42	0.4	0.72	0.067	0.69	0.5	0.058	0.28
wfdei	0.83	0.4	0.29	0.7	0.22	0.66	0.42	0.12	0.26

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-32.4	-27	-44.9	-34.9	-69.1	-17.5	-34.4	-0.1	-70.9
princeton	-34.2	-38.3	-58.6	-40.1	-73.4	-50.1	-70.3	-13.2	-72.3
watch	-37.9	-25.7	-54.3	-43.8	-84.1	-48.3	-58	-11.7	-72.3
wfdei	-34.5	-32.6	-63.7	-43.9	-78.3	-52.4	-65.2	-10.2	-73.7

Efficiency criteria to assess simulated and observed discharge at Pod Tunguska, Yenisei

YENISEI - POD TUNGUSKA									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.43	-1.6	0.28	0.45	0.092	0.57	0.2	-1.6	-0.62
princeton	0.44	-1.7	-0.39	0.52	-0.028	0.23	-0.19	-1.2	-0.27
watch	0.61	-2	-0.16	0.57	-0.39	-0.017	0.0058	-1.7	-0.47
wfdei	0.61	-1.6	-0.32	0.45	-0.088	-0.06	-0.1	-1.6	-0.53

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	2.4	5.9	8.7	-37.9	-15.2	-34	-42.9	-22.6	-60.5
princeton	1.9	10.7	13	-25.7	-9.6	-38.3	-54.9	-25.5	-51.9
watch	0.9	15.3	10.3	-30.2	-37.1	-48.6	-55.4	-18.1	-57.2
wfdei	1.9	9.9	5.6	-37.3	-24.5	-50.3	-57.9	-21.1	-59.2

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.44	-1.8	0.35	0.5	0.076	0.59	0.24	-1.3	-0.75
princeton	0.45	-2	-0.45	0.56	-0.033	0.21	-0.23	-0.75	-0.34
watch	0.63	-2.3	-0.21	0.63	-0.45	-0.091	-0.032	-1.3	-0.58
wfdei	0.63	-1.9	-0.4	0.48	-0.068	-0.14	-0.16	-1.2	-0.65

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-23.8	87	-11	-11.2	-28.4	-0.7	6	55	-62.4
princeton	-20.8	90	-10.1	-17.1	-20.8	-37.1	-40.6	64	-52.3
watch	-21.1	104	-8.2	-9.7	-44.5	-44.4	-22.4	68	-58.2
wfdei	-19.2	94	-17.6	-16.5	-35.9	-47.5	-29	65	-60

Efficiency criteria to assess simulated and observed discharge at Fort Simpson, Mackenzie

MACKENZIE - FORT SIMPSON									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.57	-2.8	0.091	0.51	0.81	0.6	0.31	-3.2	0.35
princeton	0.56	-2.3	0.21	0.41	0.76	0.69	-0.47	-2.3	0.4
watch	0.53	-6.4	-0.28	0.53	0.55	0.72	0.31	-3.2	0.39
wfdei	0.55	-5.6	-0.11	0.49	0.72	0.71	0.16	-2.9	0.41

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.5	59.9	38.8	-18.4	-1.2	-17.9	-24.2	-12.9	-37.3
princeton	0.5	52.8	37.8	-6.9	-5.3	-20.6	-46.7	-21.6	-31.5
watch	0.4	89.2	48.8	-2.6	-19.5	-9.4	-20.9	1.7	-31.8
wfdei	2.9	83.5	48	-5.6	0.2	-9.6	-28.1	-0.5	-26.5

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.57	-3.1	0.16	0.54	0.86	0.66	0.32	-3	0.35
princeton	0.56	-2.4	0.28	0.45	0.83	0.74	-0.54	-1.9	0.42
watch	0.54	-6.8	-0.21	0.61	0.57	0.83	0.35	-2.7	0.4
wfdei	0.56	-5.9	-0.056	0.57	0.85	0.81	0.19	-2.4	0.43

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-60.9	170	35	-54.4	-31.2	11	-62.7	74	-45.5
princeton	-64.4	158	20	-53.8	-34	-19.2	-89.8	39	-45.8
watch	-63.1	231	46	-53	-49.9	2	-66	58	-46
wfdei	-61.9	219	7	-56.1	-32.3	-1.5	-71.8	50	-46.6

Efficiency criteria to assess simulated and observed discharge at Peace Point Alberta, Mackenzie

MACKENZIE - PEACE POINT ALBERTA									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.42	-5.7	-0.43	0.14	0.25	0.16	0.089	-6.4	-0.69
princeton	0.34	-5.1	-0.28	-0.0049	0.15	0.2	-0.44	-5.5	-0.52
watch	0.44	-9.2	-0.69	0.11	-0.2	0.4	-0.0099	-6.1	-0.52
wfdei	0.4	-8.7	0.0064	0.2	0.18	0.4	-0.014	-5	-0.34

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-0.1	42	18.2	-20.9	-31.7	-33.5	-33.6	-10.6	-55.8
princeton	0	37.7	17.1	-12.2	-33	-32.9	-47.7	-13.5	-45.4
watch	-0.5	63.4	23.4	-9.6	-43.6	-23.8	-27.4	0.3	-48.6
wfdei	0.5	61.5	24.3	-11.2	-27.9	-25.2	-32.5	-0.8	-44.5

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.59	-12	-0.86	0.12	0.029	-0.06	0.014	-9.3	-1.7
princeton	0.53	-10	-0.36	0.17	-0.012	0.088	-0.97	-6.5	-1.2
watch	0.73	-19	-1.3	0.37	-0.76	0.46	0.02	-6.3	-1.29
wfdei	0.64	-18	-0.062	0.47	0.23	0.43	-0.046	-5	-0.97

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	8	319	106	67	-3.9	35	19	214	-8.1
princeton	12	300	81	75	5	-0.5	-30.8	167	9
watch	-3.8	389	118	67	-28.2	24	32	153	-0.17
wfdei	-0.1	379	36	58	0	13	13	129	-1.4

Efficiency criteria to assess simulated and observed discharge at Eagle AK, Yukon

YUKON - EAGLE AK									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.89	0.84	0.61	0.27	0.083	0.46	-0.085	-1.5	0.25
princeton	0.78	0.27	0.27	-0.36	-0.45	-0.49	-0.8	-0.59	-0.2
watch	0.88	0.85	0.73	0.088	-0.57	0.27	0.066	-0.96	0.03
wfdei	0.88	0.85	0.58	0.096	-0.24	0.26	0.058	-1	0.048

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-0.4	-25	-8.3	-47	-50.1	-34.8	-39.2	-34.6	-51.3
princeton	-0.8	-58.3	-42.4	-75.5	-70.4	-73.3	-82.9	-69.5	-72.2
watch	-0.7	-19.9	-18.7	-58.4	-73.6	-49.7	-52.5	-42.8	-61.2
wfdei	-1.5	-24.8	-20.1	-58.5	-62	-50.2	-54.3	-41.6	-60.2

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.93	0.88	0.68	0.27	0.045	0.49	-0.081	-1.3	0.25
princeton	0.83	0.26	0.26	-0.45	-0.52	-0.58	-0.9	-0.62	-0.26
watch	0.93	0.9	0.76	0.048	-0.66	0.24	0.07	-0.83	-0.001
wfdei	0.92	0.88	0.6	0.054	-0.28	0.22	0.052	-0.9	0.006

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-15.2	3	-6.6	-15.8	-74	-17.3	-5.6	22	-56.6
princeton	-20.7	-46.1	-57.6	-70.8	-86.8	-86.1	-85.9	-54.8	-66.4
watch	-21.9	9	-28.2	-41.2	-90.1	-52.8	-42.1	-2.3	-62.2
wfdei	-19.4	3	-44.1	-39.9	-82	-52.4	-47.4	5	-63.5

Efficiency criteria to assess simulated and observed discharge at Nenana AK, Yukon

YUKON - NENANA AK									
NSE (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.66	0.69	0.21	-0.23	-0.31	-0.26	-0.17	-0.93	0.11
princeton	0.83	0.17	0.0099	-0.68	-0.76	-0.76	-1	-0.63	-0.42
watch	0.83	0.54	0.22	-0.68	-0.88	-0.53	-0.61	-0.76	-0.06
wfdei	0.57	0.62	0.36	-0.28	-0.39	-0.2	-0.26	-0.95	0.14

PBIAS (Monthly)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-0.2	-26.3	-30.6	-59.3	-67.1	-53.3	-40.1	-49.7	-56.6
princeton	-0.3	-61.7	-57.3	-80.8	-81.7	-82.4	-87.4	-76.3	-76.2
watch	0.1	-44.9	-48.2	-77.7	-85	-75.8	-70.6	-68	-60.9
wfdei	-7.3	-25.5	-34.3	-62.2	-69.3	-56.3	-59.8	-49.7	-50

NSE (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	0.74	0.76	0.23	-0.26	-0.36	-0.25	-0.1	-0.74	0.12
princeton	0.92	0.16	-0.013	-0.77	-0.83	-0.84	-1.1	-0.68	-0.46
watch	0.91	0.57	0.21	-0.76	-0.96	-0.59	-0.67	-0.74	-0.05
wfdei	0.73	0.75	0.4	-0.27	-0.43	-0.11	-0.26	-0.72	0.2

BIAS in SD (Seasonal)									
	WATERGAP2	DBH	H08	MPI-HM	PCR-GLOBWB	MATSIRO	ORCHIDEE	LPJML	JULES-W1
gswp3	-12	4	-35	-52.5	-78	-25.1	-42	-19.9	-57.4
princeton	-9	-47.8	-62.9	-85.4	-89	-86.8	-96.4	-69.5	-73.7
watch	-22.2	-26.8	-57	-81.6	-92	-76.8	-72.4	-54.5	-61.8
wfdei	-16.9	3	-48.7	-58.3	-78.1	-40.9	-64.9	-19.1	-54.2

Appendix-3 Research Poster of the study presented at 2019 cross-sectoral ISIMIP workshop in Paris

PERFORMANCE OF GLOBAL HYDROLOGICAL MODELS FOR CLIMATE CHANGE PROJECTIONS IN PAN-ARCTIC RIVER BASINS

Aashutosh Aryal, Anne Gädeke, Valentina Krysanova
Potsdam Institute for Climate Impact Research (PIK), Telegraphenberg A62, 14473 Potsdam, Germany
CONTACT: Anne Gädeke, Email: Anne.Gaedeke@pik-potsdam.de

Introduction

- Climate warming has strong implications for the (sub)Arctic since many biophysical states and processes are strongly influenced by the threshold and phase change of the freezing point
- Impact of climate change on the high latitude water cycle can be accessed via Global Hydrological Models (GHMs) which simulate cold region water cycle processes, including snow accumulation and melt and soil freeze and thaw processes
- A prerequisite for a model-based climate change impact assessment is a thorough model evaluation under historical climate conditions

Aim of Study

- Systematic assessment of the outputs of the ISIMIP GHMs for the Arctic river basins by doing model evaluation in the historical period 1971-2000 based on methodology suggested in Krysanova et al. (2018)
- Our methodology aims at increasing model confidence and decreasing model uncertainties in projecting climate change impacts for the future

Methods

- 9 Global Hydrological Models (GHMs) from ISIMIP2a: **DBH, H08, JULES-W1, MPI-HM, PCR-GLOBWB, VIC, WATERGAP2, LPJML and ORCHIDEE**
- 4 Climate Forcing Datasets: **gswp3, princeton, watch and wfdei**
- Historical analysis period: **1971-2000**
- Model Efficiency Criteria:
  - Mean monthly discharge: Nash-Sutcliffe Efficiency (**NSE**) and Percent Bias (**PBIAS**)
  - Long-term average monthly discharge (**Seasonal Dynamics**): Coefficient of Determination (r^2) and Percent Bias in Standard Deviation ($\Delta\sigma$)
- The models were evaluated and rated on their performance using 4 criteria and their threshold values listed below (Table 2), and then aggregated indices were estimated for every model and basin using rating scores of 1 (good performance), 0.5 (weak) and 0 (poor) for every criterion and gauge.

Table 2. Thresholds values for Model Efficiency Analysis

NSE	r^2	PBIAS and BIAS IN SD ($\Delta\sigma$)
≥ 0.5 = Good	≥ 0.7 = Good	(-25%, +25%) = Good
(0.3 , 0.5) = Weak	(0.5, 0.7) = Weak	(-50%, -25%) or (+25%, +50%) = Weak
≤ 0.3 = Poor	≤ 0.5 = Poor	≤ -50% or ≥ +50% = Poor

Study Areas

Table 1. Study Areas Details

River Basins	Countries	Basin Areas [M km²]	Gauging Stations
Kolyma	Russia	0.64	**Kolymskaya** Sredne-Kolymsk
Lena	Russia	2.42	**Kusur,** Verkhoyanski Perevoz, Tabaga, Hatyrik-Homo
Yenisei	Russia Mongolia	2.7	**Igarka** , Bol. Porog, Pod. Tunguska
Ob	Russia Kazakhstan China Mongolia	2.43	**Salekhard,** Hanti-Mansisk, Kolpashevo
Mackenzie	Canada	1.8	**Arctic Red River** , Peace Point Alberta, Fort Simpson
Yukon	Canada USA	0.83	**Pilot Point AK,** Eagle AK, Nenana AK

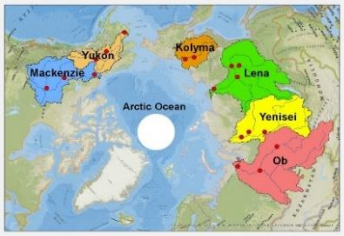


Figure 1. Pan-Arctic map showing the 6 river basins under study along with 18 gauging stations (red points)

Model Evaluation

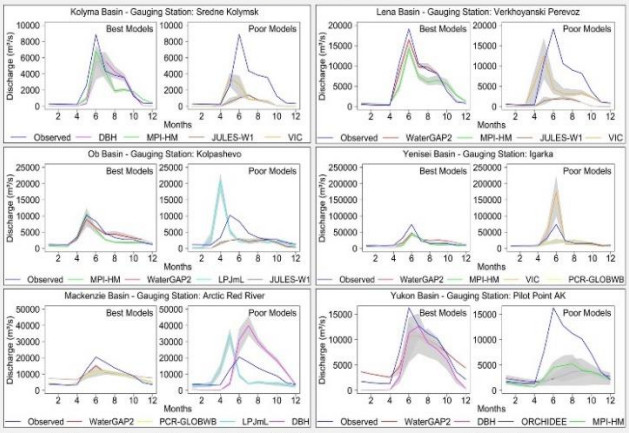


Figure 2. Measured and simulated (by GHMs) long-term average monthly discharge (seasonal dynamics). For each river basin one selected gauging station is shown. In each plot, the 2 best performing models and 2 poorest performing models are displayed. The grey spread indicates the maximum and minimum range of simulated discharge from 4 climate forcing datasets.

Preliminary Results

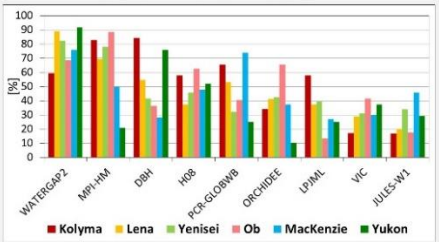


Figure 3. Aggregated indices of model performance of nine GHMs based on four Model Efficiency Criteria (see methods)

Conclusions

- WaterGap2 and MPI-HM models performed better in most of the river basins than the other models (Figure 3)
- Some GHMs have considerable difficulties in representing (sub)Arctic hydrological processes realistically (Figure 2)
- The model evaluation allows to assign weighting coefficients to the models based on their performance, and to apply them for climate impact assessment

Reference
Krysanova V., Donnelly C., Gelfan A., Gerten D., Arheimer B., Hattermann F. & Kundzewicz Z.W., (2018) *How the performance of hydrological models relates to credibility of projections under climate change*, Hydrological Sciences Journal, 63:5, 696-720, DOI: 10.1080/02626667.2018.1446214