Estimating Future Flood Risk in the Columbia River Basin Under Climate Change Using an Ensemble of Hydrologic Simulations

by

LAURA ELISABETH QUEEN

A THESIS

Presented to the Department of Computer and Information Science
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2019

An Abstract of the Thesis of

Laura Queen for the degree of Bachelor of Science
in the Department of Computer and Information Science to be taken June 2019

Title: Estimating Future Flood Risk in the Columbia River Basin Under Climate Change Using an Ensemble of Hydrologic Simulations

Approved: ______________________________________

Hank Childs

People have congregated along the Columbia River's banks throughout history, from the earliest settlements to contemporary metropolises, but this close proximity has posed a serious threat when extreme flooding occurs. Understanding how climate change will affect the future flood risk throughout the Columbia River Basin is imperative for risk mitigation and infrastructural planning. To address this question, we analyze an ensemble data set which provides daily streamflow values (1950-2100) for 172 different future projections for 396 locations in the Columbia River Basin. The ensemble members were created with a modeling decision chain which included two representative concentration pathways, ten global climate models, two meteorological downscaling methods, and four hydrological model setups. From the daily timestep streamflow data, we use extreme events from each water year to estimate flood flow values for floods with 10-, 20- and 30-year return periods. From this analysis, we find a substantive increase in flood risk for all simulated stream gauge locations in the Columbia River Basin. Our results emphasize how the hydrologic response to climate change at an streamflow location is intrinsically region and watershed dependent. Sites along the Columbia and Willamette Rivers are estimated to have a higher increase in flood-risk the further downstream the site is located. Sites along the Snake River, however, are estimated to have a lower increase in flood-risk the further downstream the site is located.

Acknowledgements

I would like to thank my committee members: Drs. Hank Childs, Phil Mote, and Barbara Mossberg. All three of you have had a substantive impact on my undergraduate and future career. Thank you, Dr. Mossberg, for encouraging me to apply for the NOAA Hollings Scholarship, the opportunity which lead me to discover climate and hydrologic modeling. Thank you, Dr. Phil Mote, for accepting me as a mentee upon my return to Oregon and advising me along each step of this entire project. Dr. Hank Childs, thank you for your continued support as a professor and advisor and, in particular, for your close involvement in the writing of this thesis. The contribution you have all made to this work and my education is invaluable to me.

Thank you as well to the team which has supported me every hour of every day for years – my family and friends. I am so lucky to have you all and simply would not be where I am without you – thank you.

This work benefitted from access to the University of Oregon high performance computer, Talapas.

Table of Contents

1. Introduction 1

2. Related Work 4

3. Methodology 7

    3.1. The Ensemble Data Set 11

    3.2. Flood Risk Metric 14

    3.3. Snow Dominance Metric 16

4. Results 17

    4.1. Evaluation of streamflow for key streamflow locations 17

    4.2. Evaluation of flood risk for the Columbia, Snake, and Willamette Rivers 22

    4.3. Evaluation of flood-risk across the Columbia River Basin 27

5. Conclusions 31

6. Bibliography 33

List of Figures

Figure 1: Annual Cycles for The Dalles, OR and Portland, OR. 10

Figure 2: Making the Ensemble Data Set: The Modeling Decision Chain 12

Figure 3: Simulated Stream Gauge Locations in the Columbia River Basin 14

Figure 4: Calculating a Flood Flow Value Example Using Simulated Streamflow 16

Figure 5: The Dalles, Brownlee and Portland Streamflow (1950-2100) 20

Figure 6: 10-yr flood-risk for the Willamette, Snake and Columbia given a 30-yr time window 23

Figure 7: 20-yr flood-risk for the Willamette, Snake and Columbia given a 30-yr time window 25

Figure 8: 30-yr flood-risk for the Willamette, Snake and Columbia given a 50-yr time window 26

Figure 9: Flood-risk across the Columbia River Basin 29

1. Introduction

Water cycles constantly through stages in the atmosphere, land and oceans as precipitation, runoff, evaporation and transpiration. A dynamic and naturally variable system, the hydrological (water) cycle sets the stage for all life to exist. The Earth's surface water, its rivers, lakes and oceans, have long provided resources as cultural, ecological, and economic agents in society. As a result, people have congregated along riverbanks and coastlines for millennia, from the earliest settlements to contemporary metropoles. The close proximity of civilization and rivers, however, poses a serious threat when extreme flooding occurs. Anthropogenic climate change is altering the hydrologic cycle in complex ways over varying time and geographic scales across the globe (Bates et al. 2008). The field of hydro-climatology studies the influence of climate upon the hydrologic cycle. It is critical we study hydroclimate extremes in order to understand the impacts of these changes and mitigate risks for nearby populations.

There are several approaches for studying the hydroclimate – theory, observation, and simulation. While each of these approaches have a role in moving our societal understanding forward, this thesis focuses on computer simulation in the form of hydrologic models. Hydrologic models are numerical models that represent the hydrologic system by using physical laws to simulate river flow over time. They provide a laboratory in which we can run experiments on the hydroclimate and predict potential impacts before they happen. Projections from calibrated, well-tested and validated models are more meaningful than extrapolations from observational data because they are designed to capture the essence of our natural world. Of course, these simulations are only useful if they accurately capture the complexities of the hydrologic system – a decidedly difficult thing to do.

An effective strategy for improving model confidence is running an ensemble of simulations. An ensemble is a set of projections produced by a sequence of model runs. This sequence is created by making a chain of decisions about the model's configuration and inputs. The model completes numerous runs given these different parameters and outputs an ensemble of future projections which simulate the hydroclimate until the end of the 21st century. The resulting analysis considers all members of the ensemble, finding not only an average future projection, but also a measure of confidence depending on how much spread there is between the ensemble members.

A key factor when using models to study the hydroclimate is ensuring the model's resolution is fine enough to capture the small-scale physics and diverse topography of river basins. This is of particular importance when studying highly mountainous river basins which vary in elevation and climatic regions in short distances. Some climate measures, like temperature and sea-level rise, are useful at the global level and can be analyzed using global climate models. For the practical planning of local issues such as water availability and flood defense, however, a global view is relatively meaningless. There is little use for a global average river flow because the study of the hydroclimate, particularly extremes like flooding, are intrinsically region and basin dependent.

This thesis considers the confluence of these topics – regional ensemble simulations, hydrologic extremes, and climate change – in the context of the Columbia River Basin (CRB). The Columbia River is significant in its size, resources, and populous basin. It is the fourth largest river by volume in North America, flowing through seven states and two countries, the United States and Canada. Hydroelectric dams on its main stem and tributaries produce nearly half of all U.S. hydroelectric power. The river meanders through a diverse array of landscapes and populations densities, from snow-capped peaks to inland deserts and large cities to rural expanses. Historic floods in the basin include the Heppner Flood of 1903, the second deadliest flashflood in U.S. history, and the Vanport Flood of 1948, which obliterated the now non-existent city of Vanport, OR. Rivers in the basin have reached flood levels as recently as April, 2019, when the Willamette River flowed over highways and neighborhoods in the Central Willamette Valley.

The central research question of this thesis asks how an ensemble of hydrologic simulations estimate the future flood risk in the CRB. This is not the first time that future flood risk in the CRB has been studied; this research is unique because we consider an ensemble of hydrologic model runs. The Pacific Northwest (PNW) presents a mosaic of regional hydroclimates. Using streamflow data for over 300 simulated stream gauge sites in the CRB, this thesis is a hyper-localized study of the future flood risk for each regional watershed. It differentiates between watersheds which are rain or snow dominant to ensure the results are of maximum relevance for local stakeholders. This analysis of future flood risk as projected by an ensemble of hydrologic simulations is expected to help inform the international renegotiation of the Columbia River Treaty between the U.S. and Canada.

2. Related Work

Flooding in the Pacific Northwest (PNW) is often attributed to extreme precipitation by the general public. In reality there is a complex array of processes which interact to lessen or amplify the effect that precipitation has on flooding. While there may be public doubt that changes in extreme precipitation and flooding are related to global climate change at all, direct effects of global warming have been shown to affect the hydrological response to precipitation regardless of whether precipitation trends change (Tohver et al. 2014). In other words, direct effects of climate change, like warming temperatures, elevating freezing levels, and retreating mountain snowpack, would change how the land distributes today's heavy precipitation, let alone the future's precipitation storms. As such, over the next few decades, climatic processes could change the frequency and magnitude of flooding in the PNW.

Extreme precipitation is projected to increase in the PNW (Duliére et al. 2013) and is associated with atmospheric river events. Atmospheric rivers are narrow corridors of water vapor which often release large amounts of precipitation as rain or snow. They are particularly prevalent along the West Coast and are a key feature in the hydrologic cycle in the CRB (Colle and Mass 1996; Garvert et al. 2007; Warner et al. 2012). Changes in atmospheric rivers could have significant effects on future hydrologic extremes in the PNW (Neiman et al. 2011). Global climate models have shown that extreme precipitation in the PNW is estimated to occur more frequently by the second half of the 21st century due to intensifying atmospheric rivers along the West Coast (Janssen et al. 2015, Wang and Kotamarthi 2015). Similarly, regional climate models have shown increases in extreme precipitation events across the PNW (Dominguez et al. 2012).

The interaction between the atmosphere and land in the hydrologic system highlights the importance of integrating precipitation and surface hydrology models to conduct hydroclimate analyses. The main method for using simulations to estimate flood risk in the PNW has been downscaling either global or regional climate model data and using this data as an input for regional hydrologic models. Tohver et al (2014) projected changes in flood risk in the PNW by using climate information from a global climate model as driving input for a hydrologic model (method described in Hamlet et al. 2013). Flood risk was assessed using the method of fitting generalized extreme value curves to projected streamflow for 197 sites in the PNW to estimate flood flow values (Hamlet and Lettenmaier 2007). The results found widespread increases in flooding across the PNW due to wetter winters and an increase in snow levels during storms associated with warmer weather. The largest increases in flooding were generally found in watersheds west of the Cascade ridge.

Tohver et al (2014) describes the future hydrologic extremes in the PNW and identifies important mechanisms contributing to the impacts of climate change. The results, however, are understated because the simulations were based on monthly climate model output. Daily changes in precipitation were assumed to scale to the changes in monthly precipitation. Thus, this analysis is missing valuable variability in precipitation which could have significant impact on the response of hydrologic extremes. Salathé et al (2013) responded to these limitations by using climate information from a regional climate model to drive the hydrologic model. Regional climate models have a finer resolution and can better represent features, such as atmospheric rivers, which determine local changes in precipitation and flooding. The results of the streamflow analysis from this modeling set-up also show increased flood risk for many sites across the PNW due to the combination of reduced snowpack and more intense precipitation events.

These prior studies suggest that future changes in extreme weather systems may have substantial effect on flood frequency. The hydrologic system in this region is highly complex because of the mesoscale processes which govern it. Mesoscale processes function on the scale between weather systems and microclimates. In this case, these processes include topographically forced air-flows such as downslope windstorms, the blocking and channeling of the winds by orography (mountain topography), and the prediction of precipitation over diverse terrain. This complexity is difficult to model and heightens the effects of methodological choices in simulating streamflow, underscoring the need for comprehensive multi-model ensemble analysis (Salathé et al. 2013). Our study follows Tohver et al (2014) and Salathé et al (2013) in being concerned with the future flood risk in the CRB under climate change. In addition, we follow similar techniques in using hydrologic simulation data for frequency analysis to characterize changing flood risk. However, our study contrasts and builds upon previous work because we use data from an ensemble of hydrologic simulations as opposed to a single hydrologic model with different climate scenarios. This study extends the previous work by considering a multi-model analysis for increased model confidence. In particular, we are seeking a better understanding of the spatial variability of flood risk in both the past and future climate.

3. Methodology

In order to accomplish our goal of estimating future flood risk in the CRB, we analyzed streamflow from an ensemble of hydrologic simulations. This project expands on previous work on future flood risk in the CRB by considering an ensemble of hydrologic simulations instead of a single projection. This data set was developed before the inception of this project (Chegwidden et al. 2018) to address the influence of methodological choices on streamflow projections. By understanding which modeling decisions had little to no effect on how streamflow is simulated, we were able to reduce the data set by a factor of four, from 160 different projections per streamflow location to 40 projections. This reduction loosened the computational constraints of this project; details of the reduction are in section 3.1 below. To better understand the changing flood risk in the CRB, we wrote computer programs to extract useful information from the daily streamflow data and present this information in meaningful visuals.

We defined two metrics which we calculate from the streamflow data: flood-risk and snow dominance. Both of these metrics are discussed broadly here and in detail in their respective sections, 3.2 and 3.3. Flood risk can be interpreted and quantified in many ways – by flood hazard, exposure, vulnerability or performance. This study provides an important insight on regional flood risk by focusing on the flood hazard component: the probability and magnitude of extreme streamflows. Streamflow in this data set has units of cubic meter per second (cms) and describes the volume of water which passes through a site each second.

We quantify flood risk by using common statistical procedures for extreme value analyses. Specifically, we considered the return period for extreme streamflow values from the data. A flood flow value is a large daily streamflow value which is associated with some return period. An example of a flood flow value, and a common term heard and often misconstrued, is the “100-year flood.” This term is intended to simplify the definition of a flood which has a 1% chance of occurring in a given year. It is often misinterpreted, however, as meaning a flood flow value which only occurs every 100 years. In reality, a 100-year flood flow value can happen two years in a row or even twice in the same year. It simply has a return period of 100 years, a probability estimated through a process called frequency analysis. By performing frequency analysis on our daily streamflow values, we can calculate flood flow values for various return periods during the beginning of our time window, the mid-20th century, and the end of the window, the late-21st century, and observe how these flood flow values have changed.

The snow-dominance metric is used to differentiate between rain- and snow-dominant watersheds to ensure our results are as locally relevant as possible. Climate change is likely to have different effects on streamflow locations in the CRB with annual cycles driven by different phenomena. An annual cycle describes the average streamflow for each month throughout a year or across years. A snow dominant watershed’s annual cycle peaks later in the water year (which begins in October) because the majority of the water contributing to the stream flow is coming from snowmelt in the spring months. A rain dominant watershed’s annual cycle peaks earlier in the water year, receiving the bulk of its flow from the rainy winter months. Figure 1 shows example annual cycles for two different streamflow locations: (1) The Columbia River at The Dalles, OR, representing a snow-dominant watershed and (2) The Willamette River at Portland, OR, representing a rain-dominant watershed. The figure shows the flow peaking in December at the rain-dominant site and in February-March for the snow-dominant site. To make our results valuable to local stakeholders, we introduce a metric of snow dominance to distinguish flood-risk trends between watersheds driven by rain or snow.

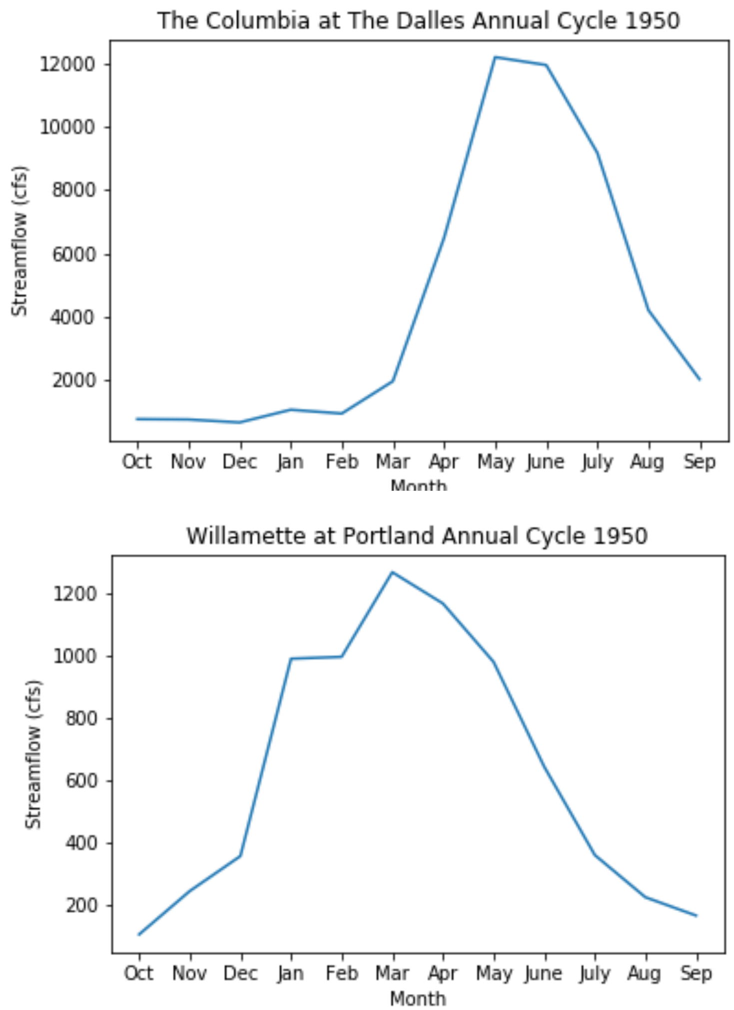


Figure 1: Annual Cycles for The Dalles, OR and Portland, OR.

The average streamflow for each month throughout 1950 is shown for (a) a snow dominant site: The Dalles, and (b) a rain-dominant site: Portland. The Dalles peaks later (May-June) because the majority of the water contributing to the streamflow is coming from snowmelt in the spring months. Portland peaks earlier (March) because the bulk of its flow comes from the rainy winter months.

The data set is encapsulated in a 25GB NetCDF file (Rew and Davis 1990). It is large enough to have required that we seek computational power beyond a single workstation. We accessed the necessary storage and computational power by using the University of Oregon's supercomputer, Talapas. Any program which reduced the dimensionality of the entire data set by extracting useful metrics, like flood-risk and snow dominance, ran on Talapas. The extracted data, much smaller in size than the original netcdf file, was then transferred to a workstation where we used the Python programming language (G. van Rossum 1995), particularly its xarray (Hoyer et al. 2017) and matplotlib (Hunter 2007) libraries, in the Jupyter Notebook software (Kluyver et al. 2016) to perform final computations and create plots of the data. The following sections provide details on the ensemble data set and the flood-risk and snow dominance metrics.

3.1. The Ensemble Data Set

While we analyze a state-of-the-art ensemble data set of simulations in this study, the development of this data set took considerable effort from an array of scholars before this project's inception. The data set was developed by scholars at the University of Washington and Oregon State University in the UW Hydro | Computation Hydrology Group and the Oregon Climate Change Research Institute respectively. The ensemble was created by running a chain of models with different permutations of modeling decisions. Each modeling decision determines how the model will ingest climate information and produce estimates of hydrologic impacts. The modeling decision chain used to create this data set, originally described and published in Chegwidden et al (2018), is shown in Figure 2.

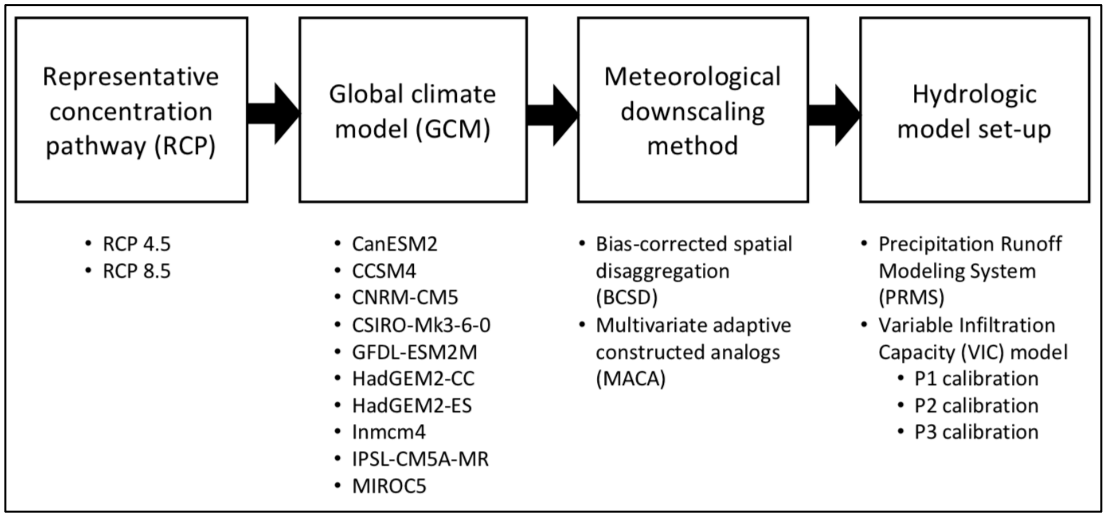


Figure 2: Making the Ensemble Data Set: The Modeling Decision Chain

This figure is from Chegwidden et al. (2018), the paper which describes the production of the ensemble dataset. The modeling decision chain consists of four decision points, represented by the four boxes. Every permutation of the bulleted list of modeling decision choices was considered to create the ensemble.

Representative concentration pathways (RCPs) describe different 21st century pathways of greenhouse gas (GHG) emissions and atmospheric concentrations, air pollutant emissions and land use (IPCC 2014). While the data set considers both an intermediate emission scenario (RCP 4.5) and a high emission scenario (RCP 8.5), our analysis only considers projections with RCP 8.5. In order to reduce the data set by a factor of two, thereby reducing necessary computations, we only consider the high emission scenario. Future work could consider a similar analysis on the RCP 4.5 projections to consider future flood risk under an intermediate emissions scenario.

Global Climate Models (GCMs) use RCPs to simulate the future of the Earth's climate. Research groups around the world have produced a large number of GCMs, which implement and simulate the Earth's system in different ways (Rupp et al. 2013). Output from GCMs is downscaled from their coarse native resolution (~150km) to the scale of the hydrologic model resolution (~6 km). After it has been downscaled, the climate information output from a GCM is used as the input for the hydrologic model. Using different downscaling methods does not have a significant effect on how streamflow is projected (Chegwidden et al. 2018). As such, we have chosen to only consider projections which used the Multivariate Adaptive Constructed Analogs (MACA) method. This again reduced our data and computations by a factor of two. We did, however, use all ten GCMs, as GCMs were shown to have significant impact on streamflow projections (Chegwidden et al. 2018).

Finally, four different hydrological models were used to develop this data set: three distinct implementations of the Variable Infiltration Capacity (VIC) model (Liang et al. 1994) and an implementation of the Precipitation Runoff Modeling System (PRMS) (Markstrom et al. 2015). The key difference between these models is what and how data were used to calibrate the model. Calibration in this case is adjusting the model such that its simulation for the historical period matches historical observational records as close as possible.

Overall, this data set contains 160 individual daily streamflow time series for 396 streamflow locations (shown in Figure 3 from Chegwidden et al 2018). In this study, we analyze 40 projections for each of the 396 streamflow locations. The reduction from 160 projections to 40 projections came from the choices to use only RCP 8.5 and MACA for the RCP and downscaling method decision points. For each site, we calculated the average flood risk and snow dominance values across the 40 projections. A description of the flood risk and snow dominance metrics follows below.

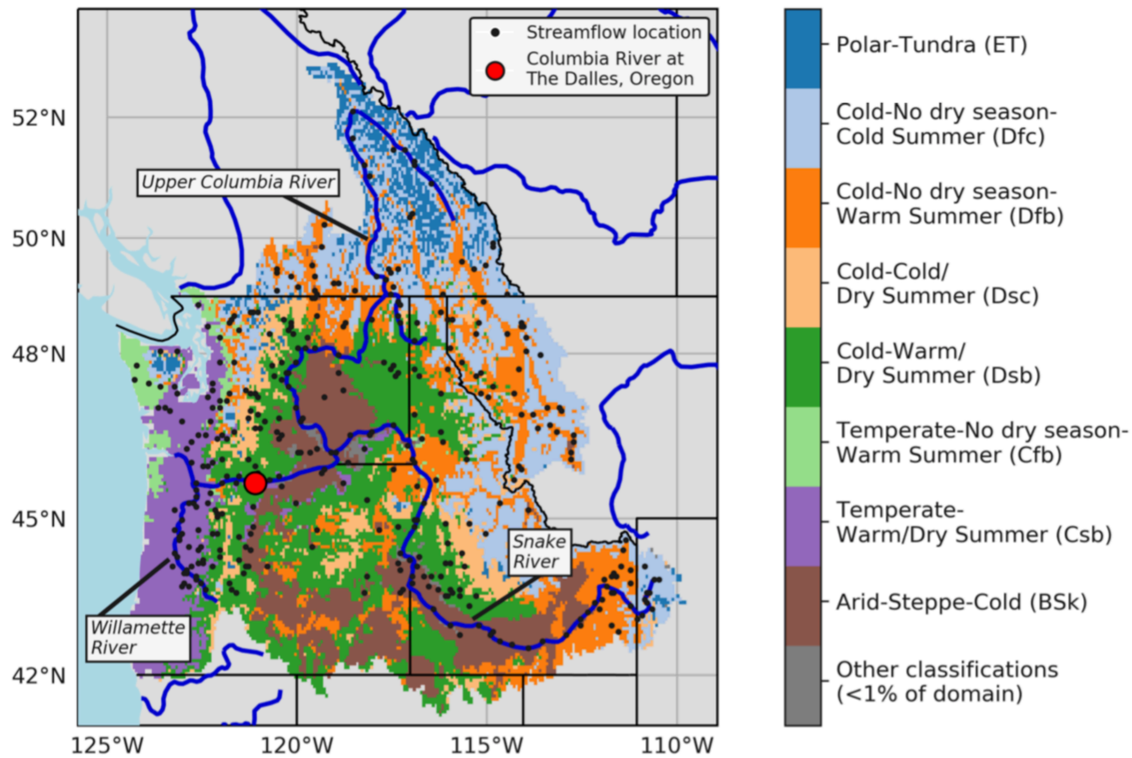


Figure 3: Simulated Stream Gauge Locations in the Columbia River Basin

This figure is from Chegwidden et al. (2018), the paper which describes the production of the ensemble dataset. The study domain is shown above using Köpen-Geiger climatic regions. The streamflow locations are shown in black.

3.2. Flood Risk Metric

The flood risk metric is quantified as the ratio of a future flood flow value (in volume per second) over the past flood flow value. For example, we can calculate a streamflow value associated with a 10-year flood for the future and the past. A 10-year flood is a flood which has a return period of 10 years, or a 0.1 probability of recurring in a year. If the future 10-year flood flow value is greater than the past 10-year flood, the ratio of future/past will be greater than one and will indicate an increase in flood-risk. If the ratio is less than one, this indicates a decrease in flood-risk.

Flood flow values for the future and past can be calculated from decades of streamflow data at either end of the total time range 1950-2100. Two options for snapshots of the future and past are 30- and 50-year time windows. The 30-year time window defines the past as 1950-1979 and the future as 2070-2099. The 50-year time window defines the past as 1950-1999 and the future as 2050-2099. Using the 30-year window, we consider floods with return periods of 10 and 20 years. Using the 50-year window, we consider floods with return periods of 30 years.

We use an empirical method to find flood flow values directly from the streamflow data in the past and future time windows. An example of finding the past (1950-1979) 10- and 20-year flood flow values for a single ensemble member at The Dalles, Or, is shown in Figure 4 below. Given a timeseries of daily streamflow, we find and sort maximum streamflow values for each year in descending order. We calculate the probability of exceedance (PE) (number of values above current value divided by the total number of values) for each max and select the max whose PE is closest to the value 1/(return period). This selected annual max is the flood flow value for the given return period. We calculate this value for the future and past and find the ratio of future/past to represent the change in flood risk. The flood-risk ratio is calculated for all 40 projections. The average of these 40 ratios is the final flood risk metric for the streamflow location.

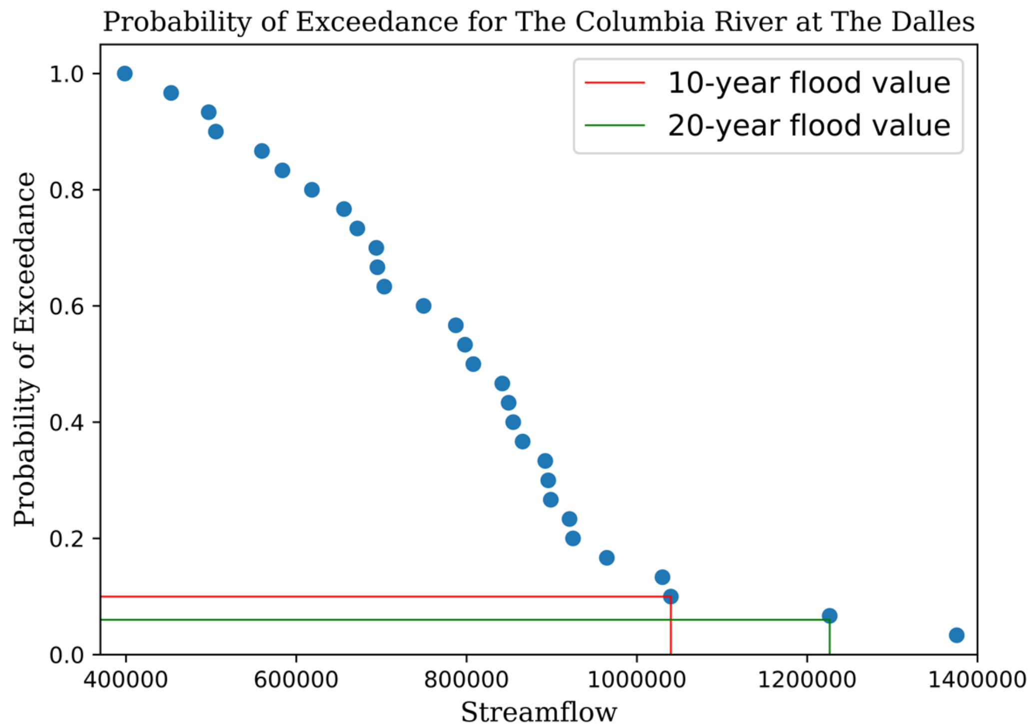


Figure 4: Calculating a Flood Flow Value Example Using Simulated Streamflow

The PE for annual maximums from 1950-1979 are shown for one projection (RCP 8.5, CanESM2, BCSD, VIC P1). The red line shows the maximum which corresponds to a 10-year flood flow value with a PE of 0.1. The green line shows the maximum which represents the 20-year flood flow value with a PE of 0.05.

3.3. Snow Dominance Metric

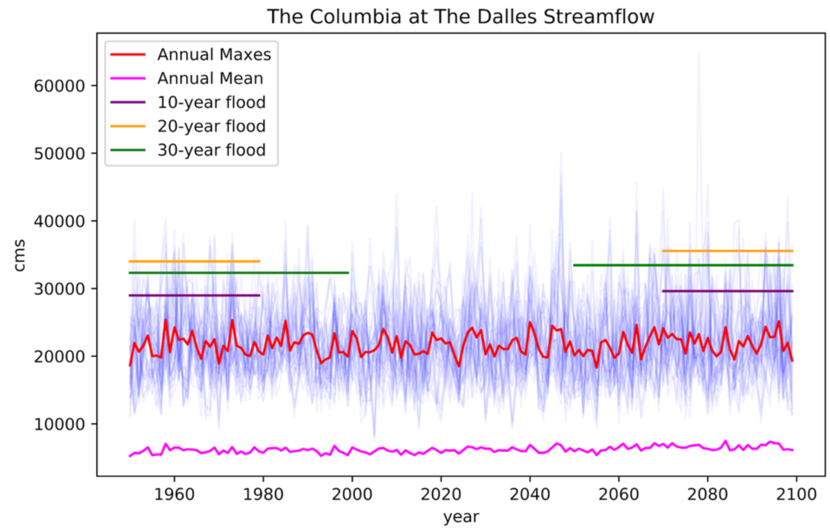
We use the centroid of timing as a proxy for snow dominance in this study. The centroid of timing is defined as the day of the water-year when half of the total annual flow has passed through a system. A snow-dominant watershed will have a later centroid because of the large contribution of snowmelt late in the water-year to the streamflow. A rain-driven watershed will have an earlier centroid due to heavy rainfall in the late fall and winter months. Refer to the annual cycles for The Dalles, OR and Portland, OR in Figure 1 for a visualization of differences in timing of streamflow between snow- and rain-dominant sites.

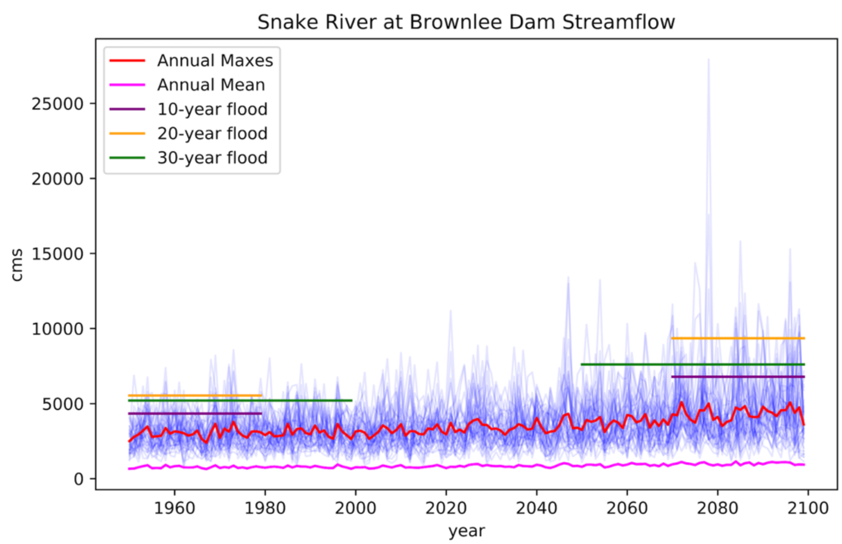
4. Results

The central research question of this thesis asks how an ensemble of hydrologic simulations estimate the future flood risk in the CRB. We break this question into three levels of focus: individual streamflow locations, main rivers, and across the entire basin. Analyses of key streamflow locations (described in 4.1) show how the simulations project daily streamflow over time. A further zoomed out analysis of the basin's three main rivers (described in 4.2) shows how trends between flood-risk and snow-dominance differ between the major watersheds in the basin. Finally, a view over all points in the basin (described in 4.3) gives a high-level perspective of the changing flood risk in the entire CRB.

4.1. Evaluation of streamflow for key streamflow locations

Looking closely at all 396 streamflow locations in the study area would be an unwieldly task. Therefore, we chose three significant rivers and sites along each to represent different watershed types: 1) a snowmelt dominant watershed: the Columbia River at The Dalles, OR; 2) a transient, mixed rain-snow watershed: the Snake River at Brownlee Dam, WA; and 3) a rain dominant watershed: the Willamette River at Portland, OR. Figure 5 plots a variety of projected streamflow data for these three locations. There are 40 timeseries displayed in blue; these show the annual maximum daily streamflow values for the 40 projections. It is these maximum streamflow values which are used in the frequency analysis to find the (10,20,30)-year flood flow values (described in 3.2). The 10- and 20-year flood flow values for the 30-year time windows (1950-1979 for past and 2070-2099 for future) are shown in purple and orange respectively. The 30-year flood flow value from the 50-year time window (1950-1999 for past and 2050-2099 for future) is shown in green. Figure 5 also shows the average annual daily maximum and mean timeseries over the 40 projections in red and magenta respectively. This plot shows differences between the streamflow locations in three key ways: (1) total streamflow magnitude, (2) the ratio between maximum and mean daily streamflow timeseries, and (3) decadal variability. The plots show how flood flow value calculations are influenced by the given time window and decadal variability. Each of these observations are described in its own paragraph in the remainder of this section.


The Columbia at The Dalles Streamflow
Annual Maxes
Annual Mean
10-year flood
20-year flood
30-year flood
cms
10000
20000
30000
40000
50000
60000
1960
1980
2000
2020
2040
2060
2080
2100
year


Snake River at Brownlee Dam Streamflow
Annual Maxes
Annual Mean
10-year flood
20-year flood
30-year flood
cms
0
5000
10000
15000
20000
25000
1960
1980
2000
2020
2040
2060
2080
2100
year

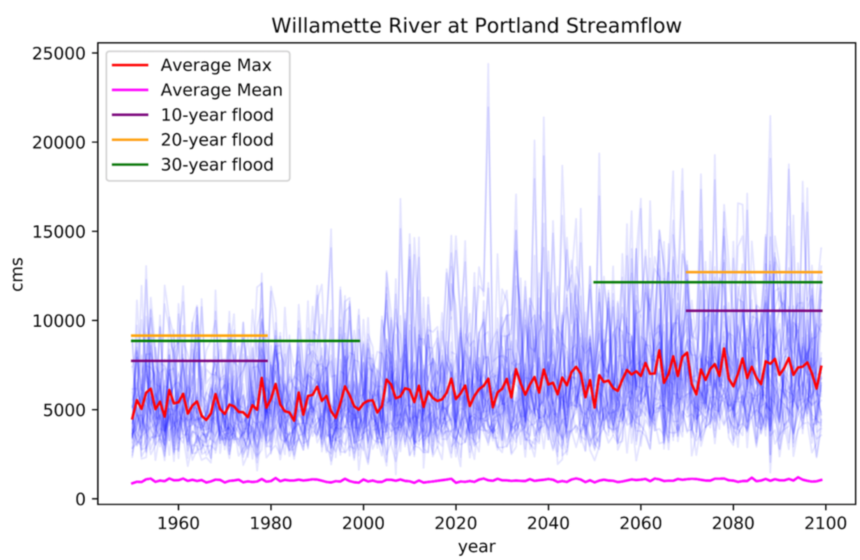

Willamette River at Portland Streamflow
Average Max
Average Mean
10-year flood
20-year flood
30-year flood
cms
0
5000
10000
15000
20000
25000
1960
1980
2000
2020
2040
2060
2080
2100
year

Figure 5: The Dalles, Brownlee and Portland Streamflow (1950-2100)

Timeseries for the annual maximum daily streamflow for each of the 40 projections are shown in blue, the average of these 40 projections in red, the average annual mean daily streamflow for the 40 projections in magenta, and the 10-, 20- and 30-year flood flow values calculated from the projections over 30 (10- and 20-year floods) or 50 (30-year flood) year windows in purple, orange and green respectively.

The difference in relative size of the watersheds is shown by the differing scales of the y-axes, which represent streamflow values in cms. The relative size of a basin is defined by the magnitude of its average mean streamflow (in magenta). This value can be seen in the figures in sections 4.2 and 4.3 as the size of the points. The Columbia River at The Dalles is one of the largest flows in the basin and is decidedly larger than the flows at both Brownlee and Portland with a mean flow about 6 times greater (~6,000 cms vs ~1,000 cms). Brownlee and Portland are similar sized flows, averaging about 1000 cms each for an annual mean daily streamflow value.

The change between the mean streamflow (in magenta) and max streamflow (in red) differentiates the significance of the consequences of flood increases between the sites. The Dalles' max streamflow jumps to levels about 3.5 times its mean flow, i.e. approximately 21,000 cms. While Brownlee and Portland have a similar average mean streamflow, Portland's average maximum streamflow is higher than Brownlee's with a scaling between mean and max of about 5 compared to Brownlee's 3. With maximum values starting at different orders of magnitude, the significance of an increase in flood flow value for The Dalles is much greater than an increase in Portland and Brownlee. An increase in flood-value in Portland, then, is slightly more significant than an increase at Brownlee. In other words, a flood ratio of 1.5, meaning the future flood flow value is 50% larger than the past value, would have different implications depending a basin's size and the relationship between the mean and maximum averages.

The flood flow values, plotted with the projections they are calculated from, show how running frequency analysis using data from different time ranges affects the resultant value. The 30-year flood flow values for each site, for example, are calculated from 50 years of streamflow data. The 30-year flood should theoretically be greater in magnitude than the 10- and 20-year floods because 30-year floods have a .03 chance of occurring in a year compared to .1 and .5 chances for 10- and 20-year floods respectively. Instead, the 30-year flood is shown to be less than the 20-year flood flow values. The 30-year flood is lower because the outlying peaks which have a great effect on the 30-year time window calculations have less weight when 50 years of data are considered. As such, there is a trade-off between choosing 50 years of data for better statistics but more compression of extremes and choosing 30 years of data for more representative variability but less robust statistics.

Figure 5 also shows differing decadal variability between the sites which suggests the effectiveness of using 30- or 50-year time windows as snapshots of the past and future depends on the streamflow location. The Dalles, for example, does not depict the average maximum streamflow changing significantly in the first or last 50 years, so using a 50-year time window to calculate flood flow value could be an appropriate approximation for past and future flood-risk. The smaller time window, with a weakened statistical analysis, is appropriate, however, for sites like Brownlee and Portland which do show decadal change. There is a visible increase in average maximum streamflow for both Brownlee and Portland. This increasing trend appears within the first and last 50 years, suggesting using 50-year time windows for these sites is not an appropriate snapshot of the past and future.

4.2. Evaluation of flood risk for the Columbia, Snake, and Willamette Rivers

Figures 6, 7 and 8 show results of an evaluation of 10-, 20-, and 30-year flood-risk for the Columbia, Snake and Willamette Rivers. For each streamflow location along these rivers, flood-risk ratio (described in 3.2) is plotted against snow-dominance (described in 3.3). The size of each point is relative to the size of the streamflow location, measured as its annual streamflow. The size of a site can also be interpreted as the location of the site along a river - how downstream it is. The size of streamflow location indicates how far downstream the site is along a river because the larger the site, the more of its flow is contributed by upstream tributaries.

These figures also indicate the spread between model projections by coloring each streamflow location according to the standard deviation of its projections. This coloring is an indication of the level of confidence we have in the model simulations. A high standard deviation would indicate a lot of spread between projection and thus a low level of confidence in the results. A low standard deviation indicates little spread amongst the projections and thus a high level of confidence in their results. We use a discrete colormap and make each point slightly transparent, so larger points do not entirely block smaller points. As such, points which are much darker in value are likely two or more points overlapping.

The relationship between flood-risk, snow-dominance and downstream location (size) for all three rivers are shown in the figures below. The [10,20]-year floods given a 30-year time window are depicted in figures 6 and 7 and 30-year floods given a 50-

year time window are shown in figure 8. A discussion of the trends within the rivers follows each figure. A discussion of ensemble spread and model confidence follows figure 8.

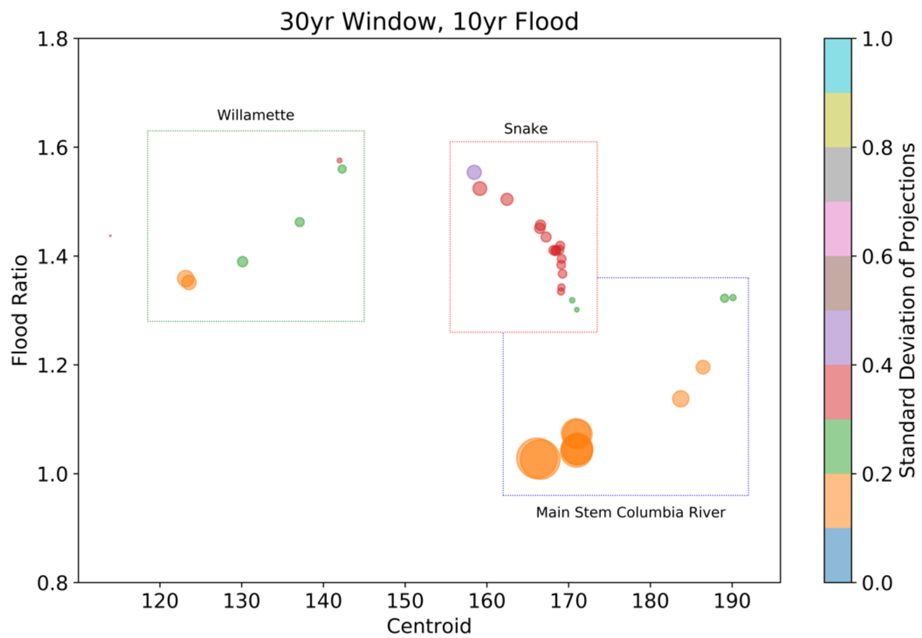


Figure 6: 10-yr flood-risk for the Willamette, Snake and Columbia given a 30-yr time window

The snow-dominance (centroid in number of days past October 1st) vs. flood-risk (flood ratio) for a 10-year flood from 30-year time windows is shown for each point along the Willamette, Snake and Columbia rivers. The point size represents the streamflow location's relative size (annual streamflow) and the color represents the spread (standard deviation) between the 40 projections.

Figure 6 shows a positive trend between 10-year flood-risk and snow-dominance for the Willamette and Columbia Rivers and a downward trend for the Snake River. The flood flow values were calculated given a 30-year time window (past: 1950-1979, future: 2070-2099). Points along the Willamette, shown on the left surrounded by a green box, have early centroid timing, classifying the Willamette as a rain-dominant watershed. For these streamflow locations, the later the centroid timing, the higher the flood-risk ratio. Sites increasing in centroid timing can be interpreted as sites rising in elevation, retreating into the Cascade Range, becoming more affected by snowmelt. A similar trend is shown for sites on the main stem of the Columbia River which are shown on the right surrounded by a blue box. Sites along the Columbia also increase in flood-risk as they increase in snow-dominance, despite the Columbia being a snow-dominant watershed. Sites along the Snake River, bounded by the center red box, show a contradictory trend. As sites become more snow-dominant their flood-risk decreases. The Willamette and Snake share similar flood-risk ratio values, ranging from 1.2 to 1.6. The Columbia's flood-risk ratios are lower, ranging from 1 to 1.3.

Figure 6 also reveals a relationship between a streamflow location's size and its snow-dominance. For all three rivers, catchment areas increase in size as their centroid timing decreases and they become more rain-dominant. Water flows down the Cascade mountains, as snowmelt or runoff, joining with rain in the lower valleys to create large rivers in rain-dominant watersheds. As such, it makes sense that moving down a river decreases the centroid (i.e. increases the rain-dominance) while increasing the catchment area. In terms of the figure, another way to say this is that sites on each river are ordered from downstream to upstream from left to right.

Because a site's downstream location is inversely proportional to its snow-dominance, the trend between flood-risk and downstream location is inversely proportional to the trend between flood-risk and snow-dominance. As such, the Willamette and Columbia Rivers share a negative trend between flood-risk and downstream location. This means that the further downstream a site is along the Willamette or Columbia Rivers, the lower its increase in flood-risk. The Snake river, in contrast, has a positive trend between flood-risk and location. The further downstream and site is along the Snake River, the higher its increase in flood-risk.

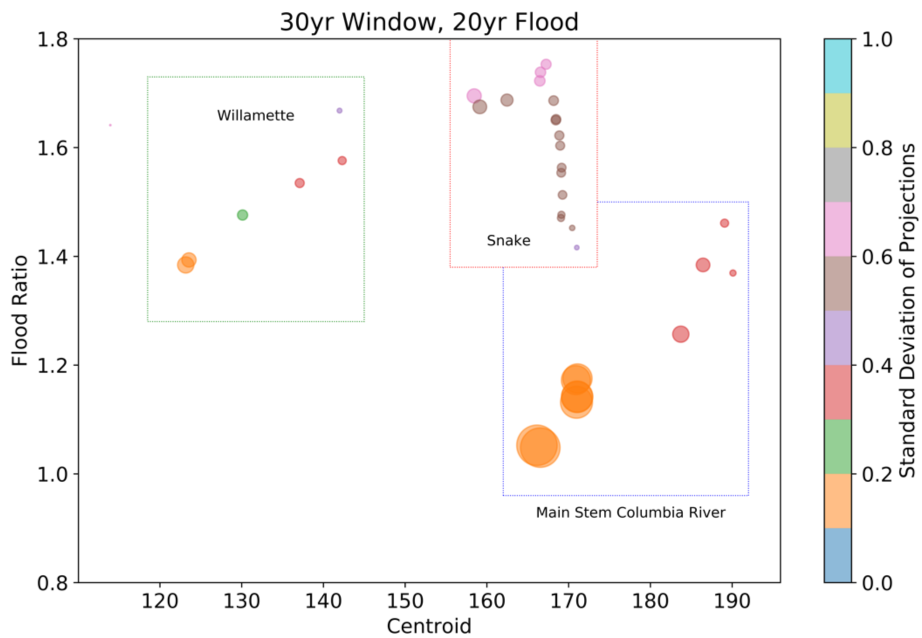


Figure 7: 20-yr flood-risk for the Willamette, Snake and Columbia given a 30-yr time window

The snow-dominance (centroid) vs. flood-risk (flood ratio) for a 20-year flood from 30-year time windows is shown for each point along the Willamette, Snake and Columbia rivers. The point size represents the streamflow location's relative size (annual streamflow) and the color represents the spread (standard deviation) between the 40 projections.

Figure 7 shows 20-year flood-risk given a 30-year time window. The figure shows similar (but messier) trends to those in figure 6. The main differences between flood-risk trends for 10- and 20-year floods are (1) the magnitudes of the flood ratios and (2) the rate at which flood-risk changes dependent on a site's snow-dominance or size. Again, the Willamette and Columbia have similar behavior. However, only the large downstream sites have similar flood-risk magnitudes as the 10-year flood. As the sites decrease in size and move upstream, the flood-risk magnitudes rocket to higher levels than seen in the 10-year flood figure. Along the Snake, all sites show a much higher flood-risk magnitude. Barring the few large sites, the downward trend is much steeper, showing flood-risk decreasing quickly as sites retreat upstream.

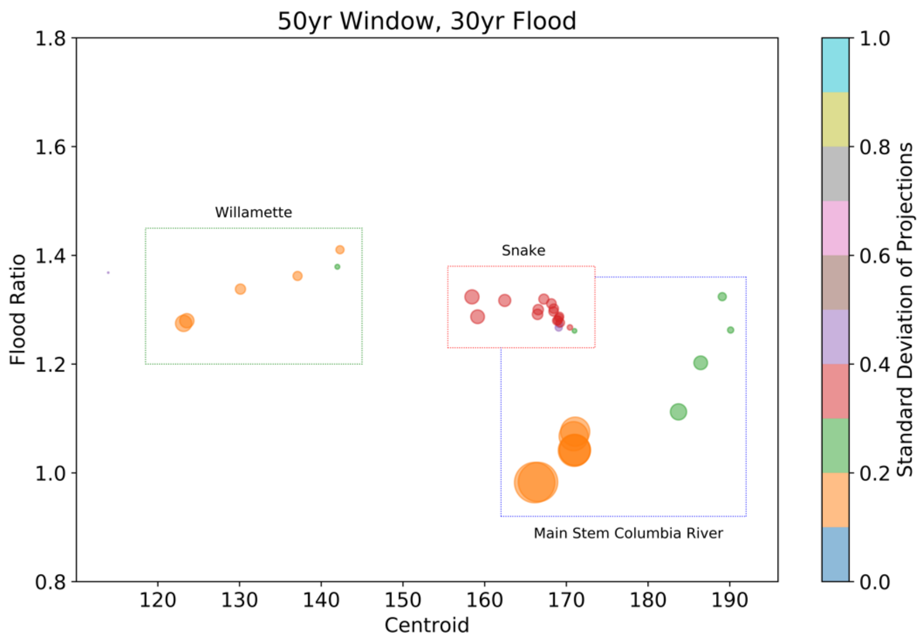


Figure 8: 30-yr flood-risk for the Willamette, Snake and Columbia given a 50-yr time window

The snow-dominance (centroid) vs. flood-risk (flood ratio) for a 30-year flood from 50-year time windows is shown for each point along the Willamette, Snake and Columbia rivers. The point size represents the streamflow location's relative size (annual streamflow) and the color represents the spread (standard deviation) between the 40 projections.

Figure 8 shows 30-year flood-risk calculated from a 50-year time window (past: 1950-1999, future: 2050-2099). The results show considerably lower flood-ratio magnitudes for each site and compressed trends between flood-risk, snow-dominance and streamflow location size. The results for the Columbia River show very similar increases in flood-risk and trends as the 10-year flood plots in figure 6. The Columbia River in all figures has shown the least response to changing return periods and time ranges. The Willamette and Snake, however, show a great decrease in flood-ratio magnitudes compared to Figures 6 and 7. Most sites along the Willamette which laid above a 1.4 ratio in both the 10- and 20-year flood figures now lay entirely between 1.2 and 1.4. The sites along the Snake River all have similar flood-risk magnitudes hovering around 1.3 and show little upward or downward trend between flood-risk and snow-dominance or streamflow location size.

The coloring in figures 6, 7 and 8 show varied spread between the model results for different streamflow locations. Model results for the Columbia and Willamette Rivers have little spread with most standard deviations between projections ranging from 0.1 to 0.4. The Snake River has more spread with most location standard deviations between projections ranging from 0.3 to 0.6. The downstream sites of the Columbia and Willamette are consistent in having very low spread and therefore high model confidence. The scenario with the most spread and least model confidence for each river is the 20-year flood given a 30-year time window.

4.3. Evaluation of flood-risk across the Columbia River Basin

Figure 9 shows results of an evaluation of flood-risk across the entire CRB. Figures for the three scenarios, [10, 20]-year floods given a 30-year time window and a 30-year flood given a 50-year time window, are shown in figure 9 below. Each point in the figures represents one of the streamflow locations shown in the study area (shown in figure 3). Adding the remaining streamflow locations outside the Columbia, Snake and Willamette Rivers reveals wide spread between projections (shown with cyan) for small outlying sites. These small streamflow locations are predominantly in mountainous, upstream locations with hydrologic behavior too small-scale to accurately simulate, even with the relatively high resolution in regional models. The streamflow locations with significant flows, however, whose flooding would pose the most risk, have relatively little spread amongst their projections.

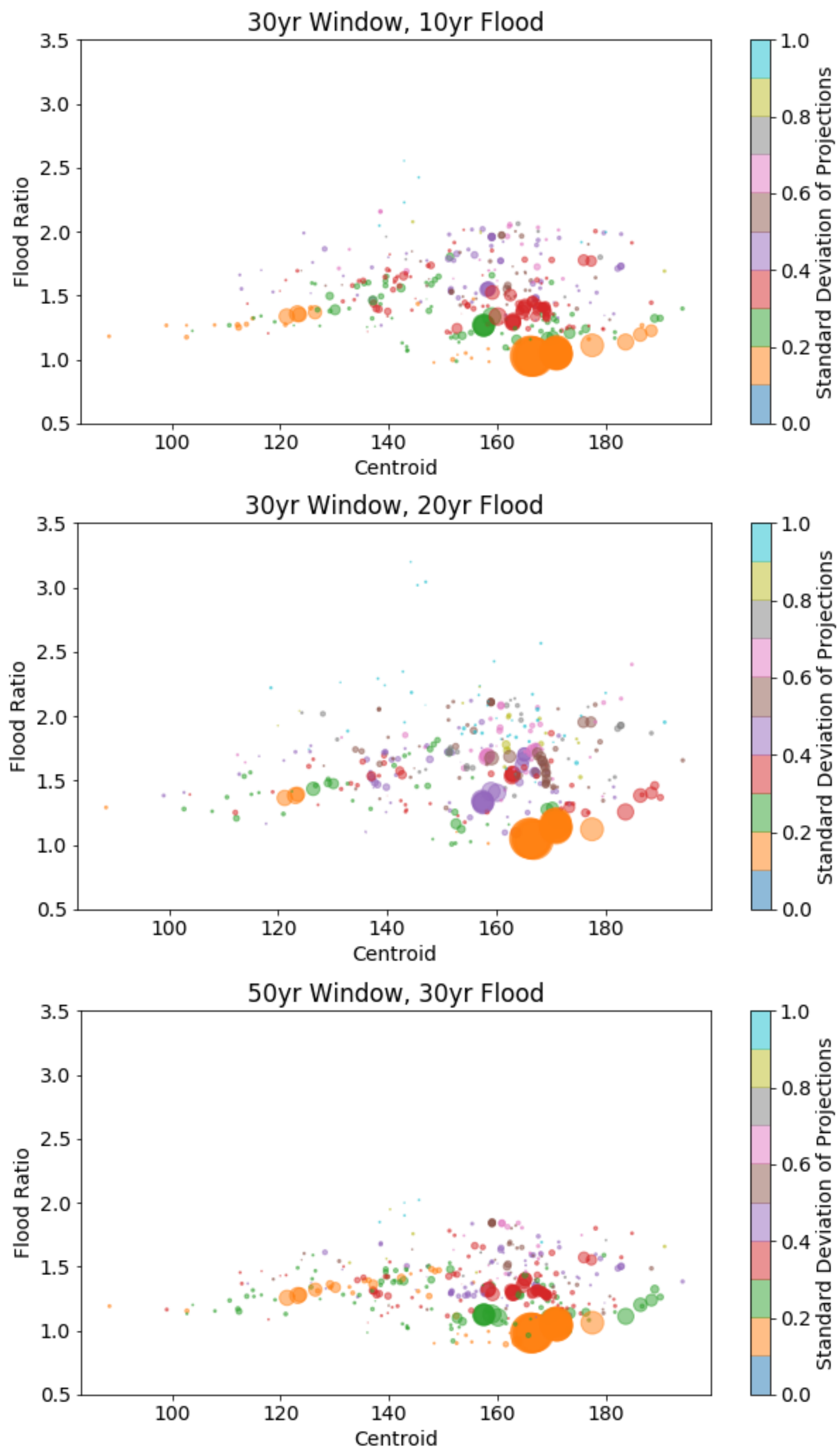


Figure 9: Flood-risk across the Columbia River Basin

10-year flood-risk (top), 20-year flood-risk (center), and 30-year flood-risk (bottom) are shown for all 300+ sites across the CRB. Point size represents streamflow location size and color represents the spread amongst the projections.

The figure shows increased flood-risk (a flood-risk ratio above one) for all sites across the basin. While there is no significant overall change in flood-risk between 10- and 20-year floods (figure 9 top and middle), the 50-year window (figure 8 bottom) is shown to compress the flood-risk values and trends. Performing a statistical frequency analysis over the 50-year window diminishes the influence of outlying maximums as shown in 4.1. This may imply there is too much decadal variability within the 50-year windows for most sites. If streamflow is already changing within 50 years, using a 50-year time window is skewing the flood-values for the past and future towards each other, making it a poor snapshot of the past and future.

5. Conclusions

This thesis sought to estimate future flood-risk in the CRB under climate change by using an ensemble of hydrologic simulations. An understanding of changes in flood-risk is imperative for risk mitigation and infrastructural planning, especially as extreme weather becomes more prevalent in a warming world. While flooding in the CRB has been studied using hydrologic models before, an ensemble of projections has never been used for flood-risk estimation in the basin. Methodological choices during model set-up influence how the model projects streamflow. This study uses an existing ensemble of projections produced by a sequence of model runs with permutations of modeling decisions to accomplish this analysis (Chegwidden et al. 2018). By taking an average of these many future projections, we lessen the uncertainly introduced by human-made decisions and focus on the features shared between the ensemble members.

Our analyses show increased flood-risk for all sites in the CRB. This result is in accordance with previous studies which analyzed single model projections (Tovher et al. 2014 & Salathé et al. 2013). Our results emphasize how hydroclimate variables, particularly extremes like flooding, are intrinsically region and watershed dependent. The location of a site, which river it is located along and how far downstream it is, determines the magnitude of the flood-risk increase. The figures detailing flood-risk for all sites show no clear trend between flood-risk, snow-dominance and streamflow location size. Isolating the Columbia, Snake and Willamette Rivers, however, shows clear trends. Sites along the Columbia and Willamette Rivers are estimated to have a higher increase in flood-risk the further downstream the site is located. Sites along the

Snake, however, are estimated to have a lower increase in flood-risk the further downstream the site is located.

The results of this study suggest that future changes in the Earth's climate will have an effect on flood-risk in the CRB. The results also emphasize the regionality of these hydrological effects. While a warming climate has been generally shown to bring an increase in precipitation, an increase in soil saturation, and a decrease in snowfall and resultant snowpack and melt, the summative effect that these changes have on streamflow location varies depending on the specific physics of the site's watershed. Further analyses on the timing of flooding across the basin and an introduction of statistical methods to extrapolate to larger flooding scenarios, like the 100-year flood, are needed to fully understand the risks attributed to future flooding under anthropogenic climate change.

6. Bibliography

Bates, B., Kundzewicz, Z. & Wu, S. *Climate Change and Water*. (Intergovernmental Panel on Climate Change Secretariat, 2008).

Chegwidden, O. S. *et al.* How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. *Earth's Future* **0**,

Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Colle, B. A. & Mass, C. F. An Observational and Modeling Study of the Interaction of Low-Level Southwesterly Flow with the Olympic Mountains during COAST IOP 4. *Mon. Wea. Rev.* **124**, 2152–2175 (1996).

Dominguez, F., Rivera, E., Lettenmaier, D. P. & Castro, C. L. Changes in winter precipitation extremes for the western United States under a warmer climate as simulated by regional climate models. *Geophysical Research Letters* **39**, (2012).

Dulière, V., Zhang, Y. & Salathé, E. P. Changes in Twentieth-Century Extreme Temperature and Precipitation over the Western United States Based on Observations and Regional Climate Model Simulations. *J. Climate* **26**, 8556–8575 (2013).

Garvert, M. F., Smull, B. & Mass, C. Multiscale Mountain Waves Influencing a Major Orographic Precipitation Event. *J. Atmos. Sci.* **64**, 711–737 (2007).

G. van Rossum, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

Hamlet, A. F. *et al.* An Overview of the Columbia Basin Climate Change Scenarios Project: Approach, Methods, and Summary of Key Results. *Atmosphere-Ocean* **51**, 392–415 (2013).

Hoyer, S. & Hamman, J., (2017). xarray: N-D labeled Arrays and Datasets in Python. Journal of Open Research Software. 5(1), p.10.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

Janssen E, Wuebbles DJ, Kunkel KE, Olsen SC, Goodman A. 2014. Observational- and model- based trends and projections of extreme precipitation over the contiguous United States. *Earth's Future* **2**(2): 2013EF000185.

Kluyver, T. *et al.* Jupyter Notebooks - a publishing format for reproducible computational workflows. in *ELPUB* (2016).

Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* **99(D7)**, 415–14 (1994).

Markstrom, S.L., Regan, R.S., Hay, L.E., Viger, R.J., Webb, R.M.T., Payn, R.A., and LaFontaine, J.H., 2015, PRMS-IV, the precipitation-runoff modeling system, version 4: U.S. Geological Survey Techniques and Methods, book 6, chap. B7, 158 p., https://dx.doi.org/10.3133/tm6B7.

Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M. & Wick, G. A. Flooding in Western Washington: The Connection to Atmospheric Rivers. *J. Hydrometeor.* **12**, 1337–1358 (2011).

Pierce, D. W. *et al.* Attribution of Declining Western U.S. Snowpack to Human Effects. *J. Climate* **21**, 6425–6444 (2008).

Rew, R. K. and G. P. Davis, "NetCDF: An Interface for Scientific Data Access," IEEE Computer Graphics and Applications, Vol. 10, No. 4, pp. 76-82, July 1990.

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C. & Mote, P. W. Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *Journal of Geophysical Research: Atmospheres* **118**, 10,884-10,906 (2013).

Salathé, E. P. *et al.* Estimates of Twenty-First-Century Flood Risk in the Pacific Northwest Based on Regional Climate Model Simulations. *J. Hydrometeor.* **15**, 1881–1899 (2014).

Tohver, I. M., Hamlet, A. F. & Lee, S.-Y. Impacts of 21st-Century Climate Change on Hydrologic Extremes in the Pacific Northwest Region of North America. *JAWRA Journal of the American Water Resources Association* **50**, 1461–1476 (2014).

Wang J, Kotamarthi VR. 2015. High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America. *Earth's Future* **3**(7): 2015EF000304.

Warner, M. D., Mass, C. F. & Salathé, E. P. Wintertime Extreme Precipitation Events along the Pacific Northwest Coast: Climatology and Synoptic Evolution. *Mon. Wea. Rev.* **140**, 2021–2043 (2012).