University
of Ljubljana
Faculty
for Civil and Geodetic Engineering

STEVEN BRAZDA

SNOWMELT FLOODS IN RELATION TO COMPOUND DRIVERS IN NORTH AMERICA AND EUROPE

MASTER'S THESIS

MASTER'S OF SCIENCE IN FLOOD RISK MANAGEMENT

Ljubljana, 2021

University
of Ljubljana
Faculty
for Civil and Geodetic Engineering

Candidate
STEVEN BRAZDA

SNOWMELT FLOODS IN RELATION TO COMPOUND DRIVERS IN NORTH AMERICA AND EUROPE

Master's thesis no.:

Mentor:
Assoc. Prof. Mojca Šraj, Ph.D.

Co-mentor:
Assist. Prof. Nejc Bezak, Ph.D.

Commission member:

Chairman of the Commission:
Prof. Dr. Matjaž Četina, Ph.D.

Ljubljana, 3. September 2021

"This page is intentionally blank"

ERRATA

Error page	Error bar	Instead	Let it be

ACKNOWLEDGEMENTS

I would like to acknowledge and greatly thank my two mentors for this thesis, Associate Professor Mojca Šraj and Assistant Professor Nejc Bezak of the University of Ljubljana. They offered advice and directed me down successful routes when I was struggling. I would also like to thank all the friends I have made along the way throughout this degree. We managed to help each other make it through this degree and pandemic with our constant support of each other, and constant positive attitude.

BIBLIOGRAPHIC-DOCUMENTALISTIC INFORMATION AND ABSTRACT

UDC:	**551.579:551.578.46: 504.4:556.166(043.3)**
Author:	**Steven Richard Brazda**
Supervisor:	**Assoc. Prof. Mojca Šraj, Ph.D.**
Co-supervisor:	**Assist. Prof. Nejc Bezak, Ph.D.**
Title:	**Snowmelt Floods in Relation to Compound Drivers in North America and Europe**
Document type:	**Master thesis**
Scope and tools:	**70 p., 7 tab., 49 fig., 10 eq., 2 ann.**
Keywords:	**Snowmelt driven floods, flood typology, soil moisture, compound events, climate zones, flood catchments**

Abstract

This study assessed the drivers of snowmelt floods in relation to compound events. Compound events are when multiple drivers/hazards occur in the same geographic region/time scale, thus amplifying their impacts. The climate drivers considered in this study are temperature, precipitation, snow thickness, snow liquid water equivalent, wind speed, vapour pressure and soil moisture content. 107 different catchments across North America and Europe were investigated, from the years 1979-2019. Each annual maximum flood was sorted into a different flood typology. These typologies were rain-on-snow floods, snowmelt floods, long precipitation floods, and short precipitation floods. These four flood typologies are all split into another two categories, with a wet initial condition and a dry initial condition.

The results indicate that the considered catchments have snowmelt floods being the dominant flood type. The high elevation catchments had the dominant typologies being short precipitation floods and long precipitation floods. This initially surprised us, as higher elevations have colder weather and thus are expected to be more influenced by snowfall. However, these mountainous regions often experience large summer rainstorms, thus creating the maximum annual flood. Lower elevation catchments had snowmelt driven floods as the dominant typology. The wet initial conditions were also much more prevalent than the dry initial conditions, proving the importance of the soil moisture condition. This was confirmed through an investigation of the relative influence of the climate factors on the determination of the river discharge. Here, soil moisture had the largest relative influence.

Results were determined for the seasonality of the floods as the average day of the year the floods occur on. Based on the results of this, there are geographic regions, such as northeastern North America and central Sweden and Norway, which have a strong seasonality and the floods occur at roughly the same time every year. This can lead to spatially compounding events, where many hazards occur in one geographic region at the same time, thus amplifying their impacts.

BIBLIOGRAFSKO-DOKUMENTACIJSKA STRAN IN IZVLEČEK

UDK:	**551.579:551.578.46: 504.4:556.166(043.3)**
Avtor:	**Steven Richard Brazda**
Mentor:	**izr. prof. dr. Mojca Šraj**
Somentor:	**doc. dr. Nejc Bezak**
Naslov:	**Poplave zaradi taljenja snega in sočasnega delovanja več dejavnikov v Severni Ameriki in Evropi**
Tip dokumenta:	**magistrsko delo**
Obseg in oprema:	**70 str., 7 pregl., 49 sl., 10 en., 2 pril.**
Ključne besede:	**Poplave zaradi taljenja snežne odeje, tipologija poplav, vlažnost tal, sestavljeni dogodki, podnebni tipi, porečja**

Izvleček

V okviru raziskave smo analizirali dejavnike, ki vplivajo na nastanek poplav zaradi taljenja snežne odeje in spadajo pod t.i. sestavljene dogodke. Do sestavljenih dogodkov pride ob sočasnem nastopu različnih naravnih pojavov ali nesreč, ko se le-ti pojavijo časovno ali prostorsko sočasno. Posledično je vpliv takih dogodkov večji kot zaradi običajnih dogodkov. V okviru naloge smo analizirali različne vplivne dejavnike kot so temperatura zraka, padavine, debelina snežne odeje, hitrost vetra, vlažnost tal, itd.. Izbrali smo 107 porečij v Evropi in Severni Ameriki, kjer so bili na voljo merjeni podatki o pretokih za obdobje 1979-2019. Največje letne poplave smo klasificirali glede na definirano tipologijo z upoštevanjem naslednjih tipov poplav: poplave zaradi kombinacije padavin in taljenja snežne odeje, poplave zaradi taljenja snežne odeje, poplave zaradi dolgotrajnih padavin in poplave zaradi kratkotrajnih padavin. Omenjene tipe poplav smo nadalje razdelili še glede na mokre in sušne začetne razmere na porečju.

Rezultati so pokazali, da se na izbranih porečjih najpogosteje pojavijo poplave zaradi taljenja snežne odeje. Na porečjih, ki so locirana na višjih nadmorskih višinah, se najpogosteje pojavljajo poplave zaradi kratkotrajnih ali dolgotrajnih padavinskih dogodkov. To je mogoče presenetljiv rezultat, saj so na višjih nadmorskih višinah temperature običajno nižje in bi pričakovali večji vpliv snega, vendar so bile poplave v primeru teh porečjih najpogosteje posledica večjih poletnih neviht. Na porečjih, ki so locirana na nižjih nadmorskih višinah, pa so poplave najpogosteje posledica taljenja snežne odeje. Nadalje smo ugotovili, da je predhodna vlažnost (mokri predhodni pogoji) eden izmed pomembnih dejavnikov nastanka poplav. Vlažnost tal je imela tudi največji relativni vpliv na merjene podatke o pretokih.

Izvedene so bile tudi analize sezonskosti. Zaznali smo območja, kjer je sezonskost zelo izrazita, kar pomeni, da se poplave skoraj vedno pojavijo v istem obdobju leta, kot je Severna Amerika ter osrednja Norveška in Švedska. To lahko pripelje do sočasnega pojava poplav na večjih območjih, kar še povečuje njihov vpliv in z njimi povezano gmotno škodo.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND SYMBOLS

α	Retention parameter in digital recursive baseflow separation method
AM	Annual Maximum
BFI	Baseflow Index Method
GRDC	Global Runoff Data Centre
IPCC	Intergovernmental Panel on Climate Change
LPF	Long Precipitation Flood
POF	Peak Over Threshold
ROS	Rain-on-snow flood
SMF	Snowmelt driven flood
SPF	Short precipitation flood

1 INTRODUCTION

Extreme weather events are a constant obstacle for societies, as they often occur with limited warning and can create devastating effects. Then, these extreme weather events can occur at the same time, and the resulting impacts on human society can be even greater. These events are known as compound events – when multiple hazards and drivers occur within the same spatial and/or temporal boundaries. Compound events were first introduced by the Intergovernmental Panel on Climate Change (IPCC) in 2012 (Vormoor et al., 2015). The IPCC presented the need for further research into the complexities, causes and consequences of compound events. Since 2012, there has been an increase in research related to compound events.

Definitions and typologies for compound events were conceived by Zscheischler et al. (2020). A hazard is the weather/climate event that causes impacts on human societies. Examples of hazards includes floods, heatwaves, droughts, wildfires, etc. These hazards are composed of various drivers; weather events which themselves may not be hazardous, but at certain scales and locations can create hazards. Examples of drivers include rainfall, snowfalls, cold fronts, high pressure systems, etc. Compound events can be categorized into four main typologies (Zscheischler et al., 2020):

Preconditioned events: In this scenario, a hazard is created or exacerbated by a pre-existing condition. The hazard is present only because of the existence of the preconditioned factor. Floods caused by snowmelts fall into this typology. Here, the snow cover is the pre-existing condition, snowmelt (or rainfall creating a rain-on-snow condition) is the driver, and flooding is the consequent hazard.

Multivariate events: In this scenario, multiple drivers and/or hazards occur in the same geographic region within a temporal boundary. This can occur when multiple drivers cause multiple hazards, or when single drivers create multiple hazards. An example of this is when a large storm with heavy rains and a storm surge hits a coastal region. This storm can cause coastal, fluvial, and pluvial flooding.

Temporally compounding events: In this scenario, a succession of hazards occur in a spatially bounded region. The succession of hazards can exacerbate the impacts from previous hazards or create new impacts. An example of this is when a series of large rainstorms (the drivers) cause flooding (the hazard) which consequently creates negative impacts.

Spatially compounding events: In this scenario, connected geographic regions experience multiple hazards within a given time period. This typology can have large, sweeping impacts across the region and the wider globe. An example of this can be seen in the 2010/2011 Australian floods, which occurred across much of the country (Zscheischler et al., 2020). This caused larger impacts as there were consequences to national insurers, federal recovery aid, and the mining and agricultural industries.

This thesis will focus on flooding compound events where snowmelt is one of the drivers. These events naturally occur in northern regions in the Northern Hemisphere, conversely in the Southern Hemisphere, and in mountainous regions. The hazard becomes possible when the snow depth increases over the winter season, then melts rapidly in the spring or with a sudden temperature rise. The rapid melting of snow saturates the soils and causes excessive amounts of runoff. Many rivers experience flooding into the floodplains each spring as vast snow-covered areas melt. When rain occurs simultaneously with the snowmelt, this can lead to further flooding.

One example of a snowmelt driven flood was seen in the 1997 Red River Flood in the United States and

Canada. The weather of the 1996-1997 winter in Southern Manitoba, North Dakota and Minnesota created the exact conditions for a devastating snowmelt flood (Burn, 1999). There were large amounts of precipitation in the late fall of 1996 which caused saturated soil to freeze over the winter, a cold and long winter led to snow accumulating for a long period of time, and heavy snowfall occurred creating a depth of snow of over 50 cm in much of the catchment. When this snow melted, the largest flood in the Red River in Manitoba occurred in over 100 years. The flood caused over $500 million Canadian (€332 million) in damages in the City of Winnipeg (Burn, 1999).

Snowmelt driven floods are vulnerable to climate change – as climate change alters expected weather patterns. One may expect when they hear the words "global warming" and think that snowmelt driven floods will be less frequent due to less snow from the warming climate. However, a more realistic scenario is that climate change will have a more regional impact; with some regions receiving more snowfall and some receiving less snowfall (Vormoor et al., 2016). Temperature patterns are also being changed, which largely impacts the snowmelts. Having sudden and intense temperature rises will cause the snow to melt at a faster rate, resulting in more hazardous flooding events. This thesis will contribute to the research of snowmelt driven floods as it will provide a greater understanding to the relationships between climate zones, catchment areas, catchment elevations and flooding typologies.

1.1 Literature review

Researching compound events is vital to better understand the connections between drivers, hazards and impacts. Raymond et al. (2020) described various methods appropriate to investigate compound events. Here, the investigations are grouped into three categories: statistical approaches, modelling approaches and socio-physical approaches. The statistical approaches include copulas, event coincidence analysis and complex networks. The modelling approaches include large climate model ensembles, integrated assessment models and hazard, catastrophe, and statistical dynamical models. The socio-physical approaches include adaptive pathways, storylines and scenario planning and stress testing (Raymond et al., 2020). Raymond et al. (2020) also suggests that climate change increases the need for investigating and determining the relationships between compound hazards, drivers, and impacts.

Sorting floods into various typologies has been completed in multiple studies. Sikorska et al. (2015) sorted floods occurring in Switzerland into six different categories: rain-on-snow floods, snowmelt floods, flash floods, short-rainfall floods, long-rainfall floods, and glacier-melt floods. The flood sorting was completed twice with two different methods of decision trees, one crisp decision tree and one fuzzy. Crisp decision trees have hard thresholds, and the floods are sorted into only one category. Fuzzy decision trees however allow a flood to be split into multiple different flood types if it shows characteristics of them. They study concluded that the dominant types of floods in the region are short-rainfall and long-rainfall floods, followed then by rain-on-snow floods. The drivers of floods were also investigated in a study from Merz and Blöschl (2003). The study investigated floods in Austria. They studied which catchments had floods driven by long-rain floods, short-rain floods, flash floods, rain-on-snow floods, and snowmelt floods. They found that there is a large regional variation throughout Austria in its various climate and terrain zones. The timing of floods was analyzed, showing which time of year the catchments are most likely to flood. Clustering of the catchment centroids also occurred. This was completed to analyze the spatial coherence of the flooding – determining which types of floods were most common in certain regions. The floods were then characterized into typologies of one of the above-mentioned drivers. The study found that the long-term rain events are the primary driver of floods in Austria.

Another study from Berghuijs et al. (2019b) investigated the drivers of the largest floods per year across Europe. The key drivers examined in this study were extreme precipitation, snowmelt, and a high antecedent soil moisture condition. The driver of extreme precipitation was determined to be the least dominant in generate large scale floods in Europe. However, in the mountain ranges of the Alps and the Carpathians, this driver was the most dominant. Snowmelt was the next most important driver in Europe – dominating areas in Eastern Europe and Scandinavia. However, it was found that the most important driver across Europe was a high antecedent moisture condition. This includes the Iberian Peninsula, British Isles and much of central Europe (Berghuijs et al., 2019b).

Vormoor et al. (2016) investigated the key drivers of floods in Norway – particularly rain-on-snow (ROS) events. They found that extreme precipitation and snowmelts are the key drivers in Norway. Looking at the trends of the snowmelt floods, it was found that due to climate change, the snowmelt and ROS floods are decreasing in frequency. However, these floods can still be extreme due to the snow depth varying from year to year. A similar study was conducted also for ROS events in Southern Germany. Sui and Koehler (2001) determined that the peak discharges in the winter are larger than in the summer months, despite precipitation being lower in the winter. This was due to snow falling and being stored on the ground. The snow would accumulate and then melt rapidly when the temperature rose. Thus, most of the precipitation would turn to runoff in a shortened period of time. 80% of the peak discharges in the study occurred in winter, and over 70% were from snowmelt or ROS events. The study confirmed that when ROS events occur, that the wettest Antecedent Moisture Condition (AMC III) in the SCS runoff generation method is present.

Harpold et al. (2014) studied the relationship between the timing of snowmelt and the peak soil moisture condition. They found that the timing of peak soil moisture was most strongly related to the days where the snowpack was completing its melting process. In nearly every case, the peak soil moisture condition occurred within 14 days of the disappearance of the snowpack due to melting. The peak soil moisture condition typically occurred a few days prior to the completion of the snowmelt, but at one of their sites in the Sierra Nevada Mountains, the peak soil moisture condition occurred after the snowpack had completely melted. They attribute this to the surface coverage. They also found that snow melts faster in open areas and prairies compared to areas which are forested, due to the solar radiation being absorbed by the trees rather than the snow.

Ho and Valeo (2004) studied the relationship between soil infiltration and soil temperature. They found that generally, frozen ground leads to a reduction in soil infiltration. This is however highly dependant on soil type. At test site in Calgary, Canada they found that the frozen soil created a nearly impervious cover. Whereas another test site in Ontario, Canada found that there was little difference in the infiltration between the frozen and unfrozen soil. The frozen ground also maintains a constant soil moisture throughout the winter. If there is a high soil moisture content in the fall when the ground freezes, there is a high soil moisture content in the spring when the snow begins to melt. When there is a low soil moisture content in the fall, the soil moisture content remains low throughout the winter. When the soil moisture condition is low, the soil tends to retain its infiltration capacities. The snowpack also has insulating capabilities, meaning that the soil remains frozen while there is snow coverage, and only melts once the snowpack melts.

Burn (1997) investigated the seasonality of floods in the Canadian Prairies, using directional statistics. The Julian Day of the flood was transformed into an angle where the first Julian day is straight up, along with day 365. The other days of the year form an angle along this circle. The average flood date can then be found using this method, along with a measure of the variability of the flooding season. This

study found that the majority of annual maximum floods in the Canadian Prairies occurred in the late Spring (Burn, 1997).

Balke and Nilsson (2019) investigated the synchronicity of floods in Europe, which are related to spatially compounding events. They found that there is an increasing synchronicity of floods in Central and Eastern Europe, which occur in the spring (Balke and Nilsson, 2019). Berghuijs et al. (2019a) also investigated the synchronicity of European floods. Synchronous areas are areas in which 50% of the rivers flood within one week of each other. The average European synchronous area is 148 km in radius, with the upper limits reaching over 250 km in northeastern Europe (Berghuijs et al., 2019a). Small synchronous areas in the Spain and France and the Carpathian Mountains are less than 100 km. They found that the synchronous areas have grown approximately 1.1% in the last 50 years, while varying across Europe.

1.2 Thesis objectives

The overarching objective of this thesis is to investigate the drivers of snowmelt affected floods. To accomplish this overall objective, a number of sub-objectives are to be met. These include:

1. Determine which climate factors are the most prevalent drivers of snowmelt affected floods.
2. Investigate the seasonality of yearly maximum floods in North America and Europe.
3. Sort the floods into various categories based on causes and determine which flood typologies are the most prevalent in different regions.
4. Determine the relationships between climate zone, catchment elevation and catchment area and snowmelt affected floods.

2 DATA

The study consists of 107 catchments located across Norther America and Europe. These catchments can be seen in Figure 1 and Figure 2 for North America and Europe, respectively. The 107 catchments were selected by hand in an attempt to gather catchments in varying climate zones, areas, elevations, and with limited missing values. This research required large amounts of data of different types. Required data includes the daily river discharge data, catchment area, catchment elevation, catchment climate zone, and various climatic factors. Any catchments with large amounts of missing discharge data were unselected from the data sample.

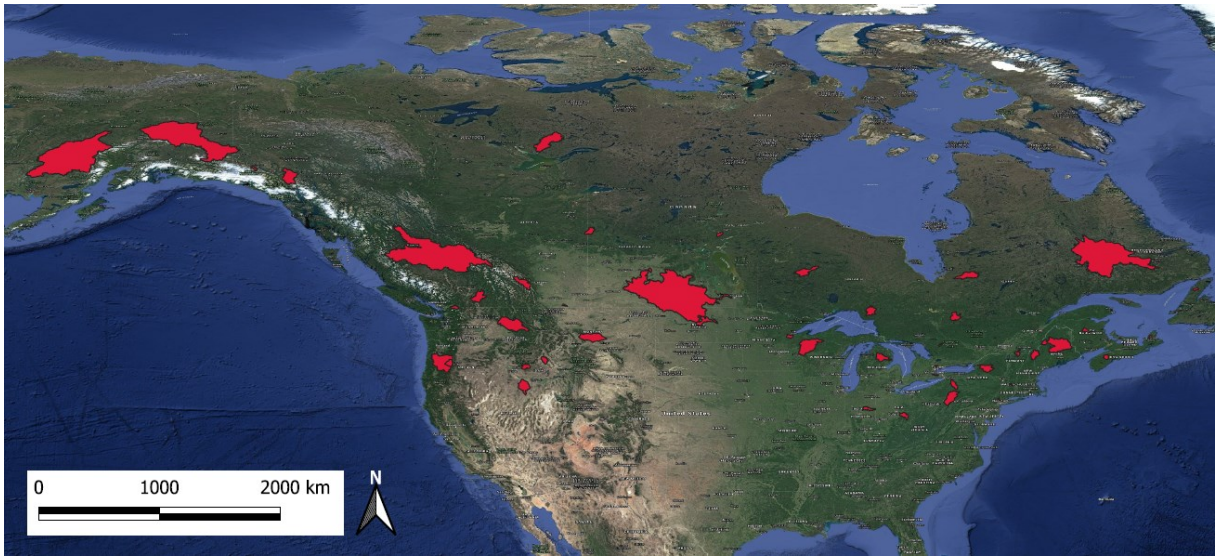

Figure 1: Considered catchments of North America

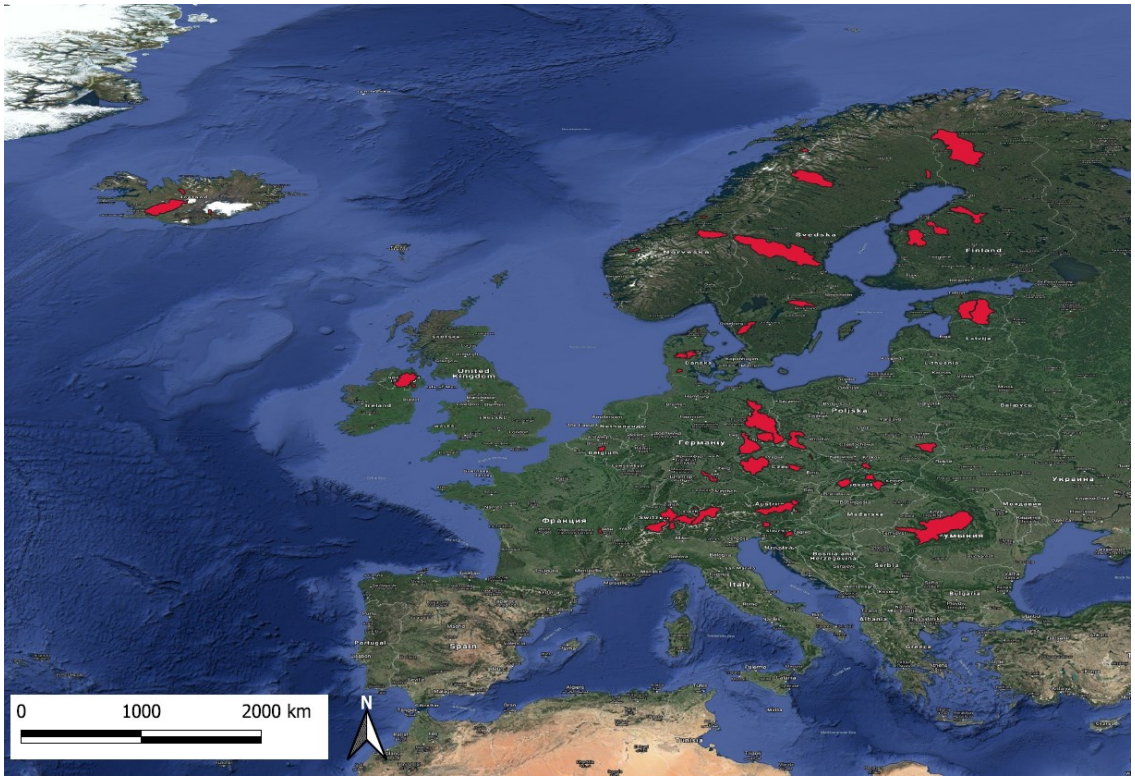

Figure 2: Considered catchments of Europe

2.1 River catchment data

2.1.1 Daily discharge data and catchment area shapefiles

The daily discharge data from the catchments was downloaded from the Global Runoff Data Centre (GRDC, 2020). The daily mean discharge was downloaded in text format (.txt), with data from 1979-2019. Additionally, the shapefiles, in GeoJSON format, were downloaded from the GRDC (GRDC, 2020). Not all catchments in the GRDC database have the associated area shapefile available for download, and thus this was another required criteria when selecting the catchments in the study. R was the primary software used to conduct the data analysis (R Core Team, 2020). When working in R, many packages are more efficient when working with shapefiles (.shp) rather than GeoJSON files, therefore the GeoJSON files were converted into .shp files (R Core Team, 2020). Table 1 displays the list of catchments chosen for the study. Figure 3 displays an example of the daily discharge time series data of the Penobscot River in the USA.

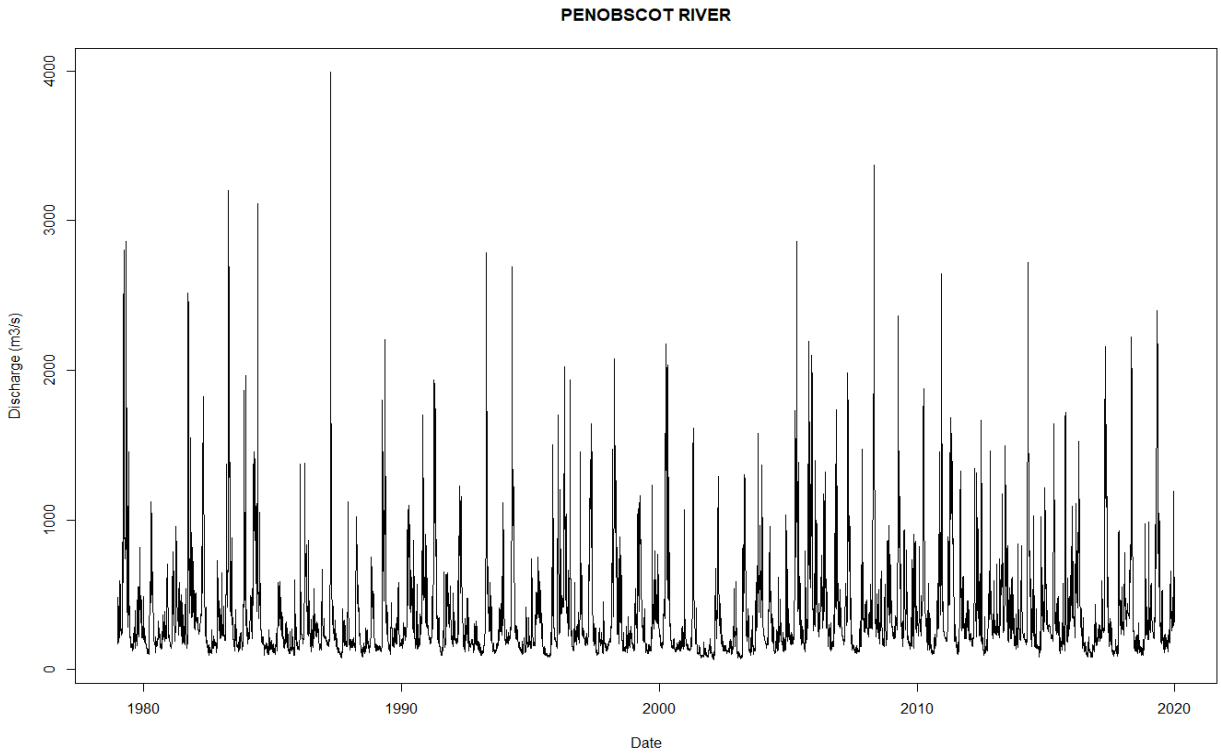

Figure 3: Example of river discharge time series

Table 1: List of chosen river catchments and their basic characteristics

Number	GRDC Number	Country	River Name	Climate Zone
1	6139360	France	AZERGUES	CFB
2	4244915	Canada	NORTHEAST MARGAREE RIVER	DFB
3	4243200	Canada	SAINT LOUIS (RUISSEAU)	DFB
4	4207910	Canada	CHILLIWACK RIVER	CFB
5	4102100	USA	KUSKOKWIM RIVER	DFC
6	4103600	USA	TANANA RIVER	DFC
7	4113360	USA	SHEYENNE RIVER	DFB
8	4115100	USA	WILLAMETTE RIVER	CSB
9	4115106	USA	MCKENZIE RIVER (TRIB. COLUMBIA)	CSB
10	4116350	USA	BRUNEAU RIVER	BSK
11	4115400	USA	SPOKANE RIVER	DSB
12	4116370	USA	PAHSIMEROI RIVER	DFB
13	4116461	USA	BOISE RIVER	DSB
14	4119441	USA	CHIPPEWA RIVER (TRIB. MISSISSIPPI)	DFB
15	4120321	USA	MUSSELSHELL RIVER	BSK
16	4123200	USA	WHITE RIVER (TRIB. OHIO)	DFA
17	4123280	USA	HOCKING RIVER	CFA
18	4143790	USA	MISSISQUOI RIVER	DFB
19	4132100	USA	NEMADJI RIVER	DFB
20	4134700	USA	AU SABLE RIVER	DFB

21	4136301	USA	GENESEE RIVER	DFB
22	4136500	USA	BLACK RIVER (TRIB. LAKE ONTARIO)	DFB
23	4147010	USA	PENOBSCOT RIVER	DFB
24	4147111	USA	ANDROSCOGGIN RIVER	DFB
25	4147470	USA	WESTFIELD RIVER	DFB
26	4147540	USA	SADDLE RIVER	CFA
27	4147726	USA	SUSQUEHANNA RIVER	DFB
28	4213465	Canada	IRONSPRING CREEK	DFB
29	4234550	Canada	NORTH MAGNETAWAN RIVER	DFB
30	4213650	Canada	ASSINIBOINE RIVER	DFB
31	4236300	Canada	EAST HUMBER RIVER	DFB
32	4203335	Canada	DUKE RIVER	DFC
33	4203910	Canada	TAKHINI RIVER	DSC
34	4207310	Canada	FRASER RIVER	DFC
35	4208370	Canada	YELLOWKNIFE RIVER	DFC
36	4213046	Canada	BOW RIVER	ET
37	4213700	Canada	TAYLOR RIVER	DFC
38	4214295	Canada	DILLON RIVER	DFC
39	4214420	Canada	PIPESTONE RIVER (TRIB. HUDSON BAY)	DFC
40	4214610	Canada	HARRICANAW RIVER	DFB
41	4214690	Canada	PONTAX RIVER	DFC
42	4215103	Canada	OKANAGAN RIVER	DFA
43	4220510	Canada	VERDIGRIS COULEE	CFB
44	4232750	Canada	WHITE RIVER (TRIB. LAKE SUPERIOR)	DFB
45	4236050	Canada	TWENTY MILE CREEK	DFA
46	4244500	Canada	CHURCHILL, FLEUVE (LABRADOR)	DFC
47	4244720	Canada	TORRENT RIVER	DFC
48	4244870	Canada	LITTLE SOUTHWEST MIRAMICHI RIVER	DFB
49	4244900	Canada	DUNK RIVER	DFB
50	4244985	Canada	LAHAVE RIVER	DFB
51	6140250	Czechia	BEROUNKA	CFB
52	6140700	Czechia	DIVOKA ORLICE	CFB
53	6142520	Slovakia	NITRA	CFB
54	6140450	Czechia	SAZAVA	CFB
55	6142680	Slovakia	VAH	DFB
56	6144100	Slovakia	SAJO	DFB
57	6233100	Sweden	VISKAN	CFB
58	6220500	Belgium	DYLE	CFB
59	6226700	Spain	VERO	CFA
60	6233220	Sweden	TAENNAN (LJUSNAN)	DFC
61	6233221	Sweden	LJUSNAN	DFC
62	6233440	Sweden	NYKOEPINGSAEN	CFB
63	6233680	Sweden	VINDELAELVEN (UMEAELVEN)	DFC

64	6233550	Sweden	KASSJOEAN (LJUNGAN)	DFC
65	6342520	Germany	ALTMUEHL	CFB
66	6233870	Sweden	SANGISAELVEN	DFC
67	6235100	Austria	BREGENZER ACH	ET
68	6242200	Austria	ILZBACH	CFB
69	6243030	Austria	INN	ET
70	6246611	Austria	MUR	CFB
71	6246700	Austria	MUERZ	CFB
72	6340600	Germany	VEREINIGTE MULDE	CFB
73	6340700	Germany	SCHWARZE ELSTER	CFB
74	6340510	Germany	HAVEL	CFB
75	6401090	Iceland	OELFUSA	ET
76	6401500	Iceland	DJUPA, FLJOTSHVERFI	ET
77	6401601	Iceland	SVARTA, SKAGAFIROI	ET
78	6457707	Poland	BOBR	CFB
79	6457880	Poland	LISWARTA	CFB
80	6458203	Poland	SKAWA	CFB
81	6458713	Poland	WIEPRZ	CFB
82	6545200	Slovenia	KRKA	CFB
83	6545190	Slovenia	SAVA	DFC
84	6604170	United Kingdom	NEVIS	CFC
85	6603300	United Kingdom	LOWER BANN	CFB
86	6605545	United Kingdom	SNAIZEHOLME BECK	CFB
87	6731175	Norway	EIDSELV	CFC
88	6731010	Norway	ENGESETELV	CFC
89	6731070	Norway	NORDELVA	DFC
90	6731501	Norway	GAULA	DFC
91	6731685	Norway	KOBBELV	DFC
92	6934800	Denmark	BREDE A	CFB
93	6744200	Romania	MAROS	DFB
94	6934250	Denmark	GUDENA	CFB
95	6854210	Finland	AHTAVANJOKI	DFC
96	6934300	Denmark	UGGERBY A	CFB
97	6854320	Finland	PYHAJOKI	DFC
98	6854710	Finland	OUNASJOKI	DFC
99	6854900	Finland	KYRONJOKI	DFC
100	6934100	Denmark	SKJERN A	CFB
101	6939500	Switzerland	RHONE	ET
102	6935145	Switzerland	RHINE RIVER	DFC
103	6935310	Switzerland	REUSS	DFC
104	6948120	Switzerland	MAGGIA (TRIB. PO)	DFC
105	6172050	Estonia	PAERNU JOGI	DFB
106	6172010	Estonia	LEVAJOKI	DFB
107	6172200	Estonia	EMAJOGI	DFB

2.1.2 Climate zone data

To determine the climate zones of each of the catchments, the Köppen-Geiger system was used (CCID, 2019). Figure 4 and Figure 5 show an example of the climate classification system in North America and Europe, respectively. The climate zone data was downloaded from the Climate Change and Infectious Diseases Group (CCID, 2019) as a .kmz file. Table 2 provides the list of climate zones in which the catchments are located in.

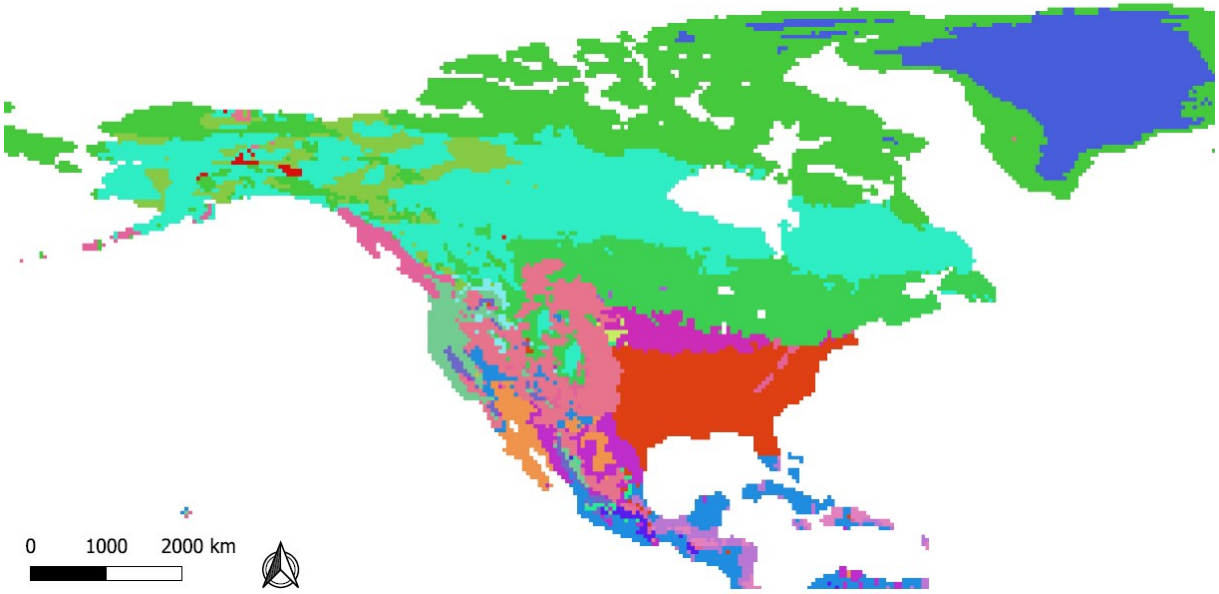

Figure 4: Climate classifications of North America

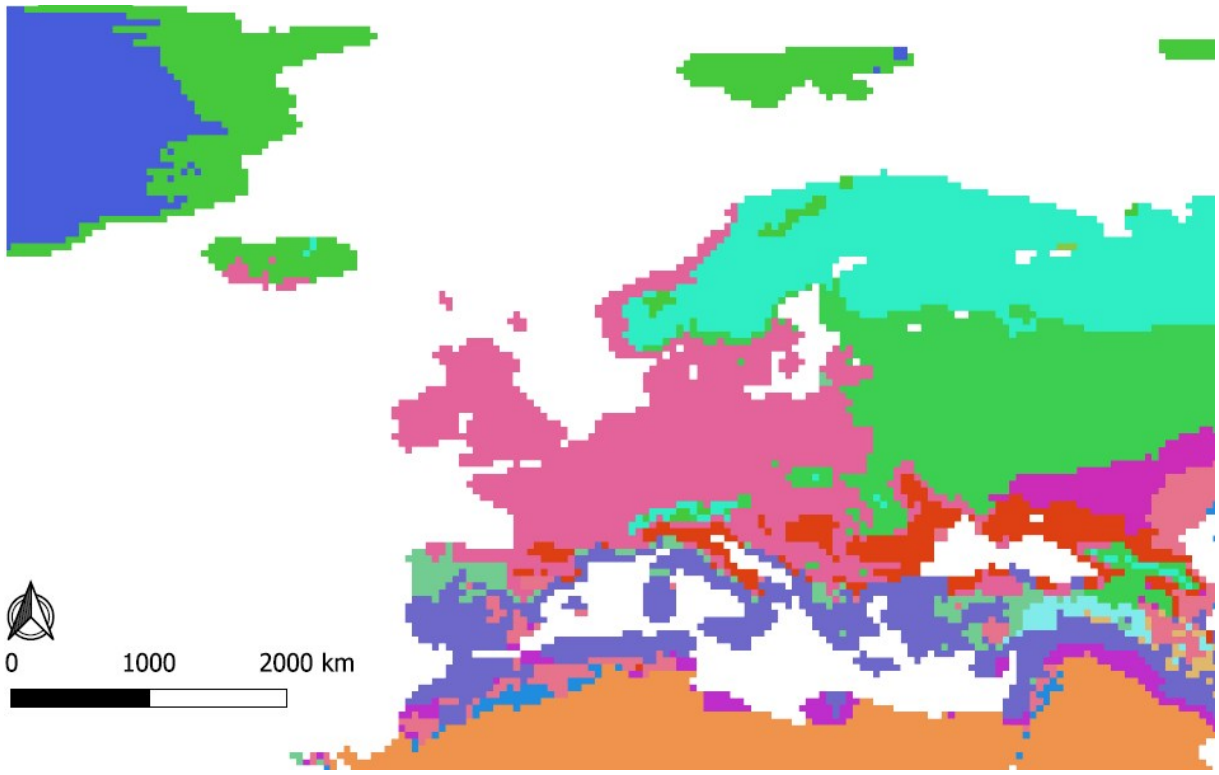

Figure 5: Climate classifications of Europe

Table 2: Selected climate zones

Short Name	Long Name
BSK	Cold Semi Arid
CFA	Humid Subtropical
CFB	Temperate Oceanic
CFC	Subpolar Oceanic
CSB	Warm Summer Mediterranean
DFA	Hot Summer Humid Continental
DFB	Warm Summer Humid Continental
DFC	Subarctic
DSB	Mediterranean Influenced Warm Summer Humid Continental
DSC	Mediterranean Influenced Subarctic
ET	Tundra

2.1.3 Catchment elevation data

Some catchment elevation data was provided from the downloaded shapefile from the GRDC (GRDC, 2020). However, approximately 50% of the catchments did not have this information available. Therefore, another data source was required. A .tiff file containing worldwide elevation data was downloaded from EarthEnv (Amatulli et al., 2018). The elevation data is a gridded dataset with a 1 km by 1 km grid. The data file was created by combining information from the Global Multi-resolution Terrain Data (GMTED) 2010 dataset, origially provided by the USGS and the 90 metre Shuttle Radar Topographic Missions (SRTM), originally provided by NASA (Amatulli et al., 2018). The GMTED and SRTM datasets are commonly used in elevation analyses, and the cross referencing of the two sources by Amatulli et al. (2018) allows for a reduction in error. Figure 6 visually displays the elevation data for the study, with the selected catchments outlined black.

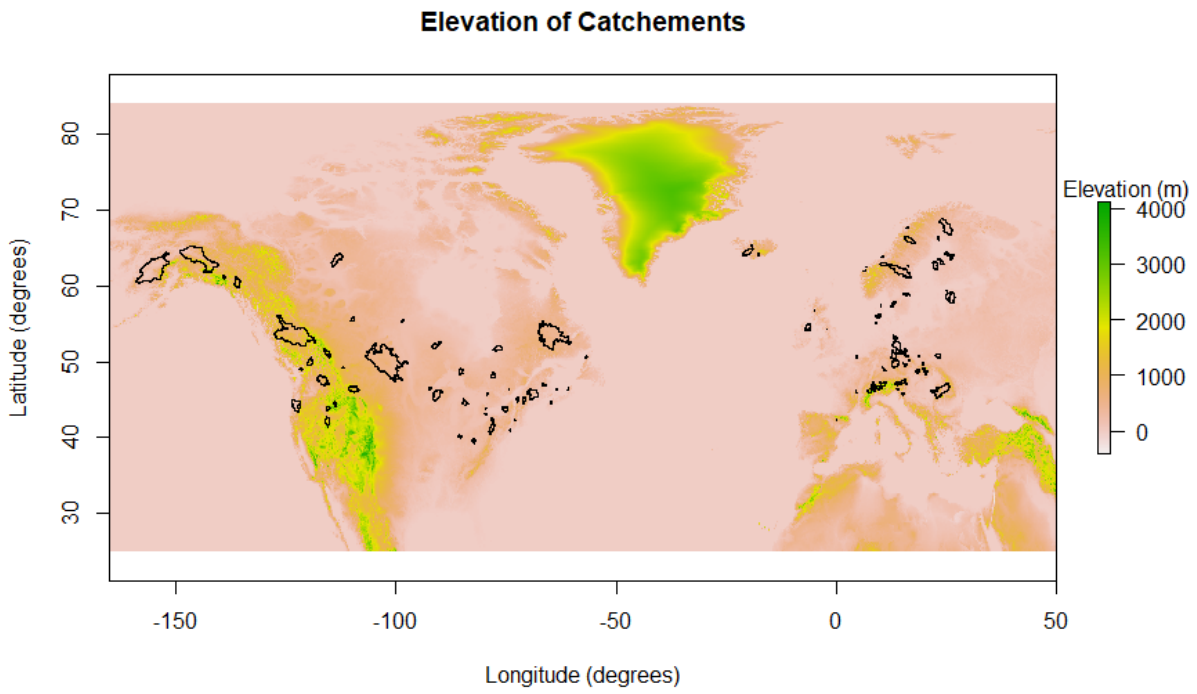

Figure 6: Elevation data with catchment outlines

2.2 Climate data

All climate data, except for the soil moisture data, was downloaded from the Agrometeorological Indicators Data Store from Copernicus (Copernicus, 2021a). The data was downloaded as NetCDF files (.nc) from the years 1979-2019. NetCDF files are gridded, layered raster files. The grid has dimensions of 0.1º by 0.1º, and each layer represented the mean daily data values. Each year was downloaded as an individual file, where each file contains 365 layers for each day of the year (366 layers for leap years). The downloaded data consisted of global data. There were options to download data from North America or Europe separately, however for continuity reasons it was important that data came from the same source to reduce errors. Table 3 displays all the climate information downloaded.

Table 3: Climate factors used in data analysis

Variable	Description	Unit
Temperature	Mean 24 hour air temperature at a 2 metre height	K
Precipitation	Total volume of water fallen over the 24 hour period, per unit area	mm/day
Snow Thickness	Mean depth of snow cover over the 24 hour period	cm
Snow Thickness Liquid Water Equivalent (LWE)	Mean depth of liquid over the 24 hour period assuming all snow melts and there is no runoff, soil penetration or evaporation	cm
Vapour Pressure	Mean water vapour pressure measured over the 24 hour period	hPa
Wind Speed	Mean wind speed at 10 m height	m/s
Soil Moisture	Volume of water in the top layer of soil (0-7 cm depth)	m^3/m^3

The soil moisture data was downloaded from the ERA5 hourly data on single levels from 1979 to present from the Copernicus Data Store (Copernicus, 2021b). This data was also downloaded as .nc files. However, there was no daily mean soil moisture data available. Therefore, the hourly data was downloaded, taken each day at noon. It was assumed that the soil moisture value at noon was the mean value for the entire day. All 24 hours of soil moisture data were not able to be downloaded and then processed into a mean value due to download times, download storage space limitations and data computational times. Figure 7 displays an example of the daily soil moisture data with the catchments overlaid.

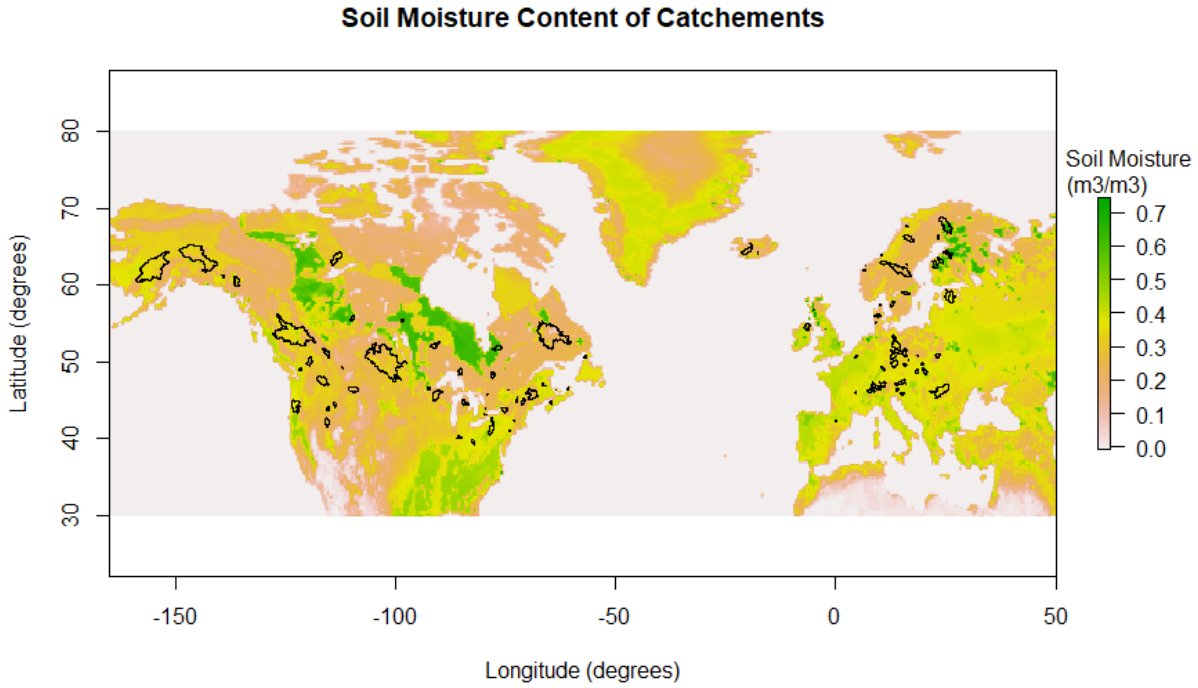

Figure 7: Soil moisture content (.nc file) of January 1, 1990 with overlaid catchment shapefiles

The climate data was developed from re-analysis methods. Reanalysis takes observed data points and inputs these into global climate models, based on physical principals. Air temperature and air humidity are often used as forced controls, to ensure that the model is accurate (Copernicus, 2021b).

3 METHODS

3.1 Data processing

Upon the completion of downloading the data, the data had to be processed, and trimmed down to manageable data sizes. The following sections describe the steps taken to accomplish the data processing.

3.1.1 Annual maxima flood selection

There are two primary methods often used for flood selection in hydrologic research. These are the Annual Maximum (AM) and Peak Over Threshold (POT) (Gottschalk and Krasovskaia, 2002). The AM method assumes that the largest discharge each year is a flooding event. The POT method sets a threshold in which any discharges occurring above this threshold is considered a flood event. Consequently, POT can result in multiple flooding events occurring in a year, and also years where no flooding events occur (Gottschalk and Krasovskaia, 2002).

The AM method was selected for the purposes of this thesis. Due to the nature of flooding affected by snowmelt, it is unlikely that multiple hazardous snowmelt driven floods will occur each year. This is due to snowfall having to be accumulated over time for significant water volumes to be stored in the snow. Additionally, due to the large variety of river catchment characteristics, with different areas, discharges, climate zones and terrain slopes, determining a consistent threshold may result in an excess of or a lack of flooding events. To determine the AM, the river discharge data was loaded into R, filtered by year, and then the function max() was used (R Core Team, 2020). This function returns the maximum flood quantity, then the data was searched and returned the Julian Day which corresponded to this flood volume.

3.1.2 Snow affected peak flood selection

The AM method provides the maximum flood each year, regardless of the time of year the flood occurs or the drivers of the flood. However, as this thesis is investigating snowmelt affected floods, for certain analyses the yearly maximum flood is of no interest. This occurs when there was a flood without any influence of snowmelt. Thus, the maximum flood that occurs each year with the presence of snowmelt was also found. To determine if the flood was impacted from snow, a threshold method was used based on the study completed by Sikorska et al. (2015). The threshold numbers were met if snow coverage in the catchment was greater than 5%, and if the snowmelt that occurs was greater than 1 mm. If the annual maximum flood meets these conditions the flood was considered to be impacted by snowmelt.

Each flood was checked to determine if the threshold conditions were met. If the conditions were not met, then the next highest daily discharge was found, and that new flooding event was checked once again for the threshold conditions. The process was repeated until a flood was found which met the threshold conditions and thus was a snowmelt affected flood. The function Rfast::nth in R was used to determine the subsequent peaks (Papadakis et al., 2021). The threshold conditions were checked for the first day of the hydrograph, when the baseflow was still equal to the total flow. The conditions could not be checked on the peak day of the hydrograph as there as some flood events where there was snow present at the start of the hydrograph but has subsequently melted completely during the rising limb of

the hydrograph. This results in no snow being present at the peak of the hydrograph. See Section 3.1.3 for how the start dates of the hydrograph were determined.

3.1.3 Hydrograph determination

The shape, duration, and start date of the hydrograph was required to conduct the analysis. The analysis requires the daily climate factors of the entire hydrograph, not just during the day of the peak. This is due to a lag time in many catchments, and the flood driving climate factors often occur on the days prior to the peak. Baseflow separation is a common technique for hydrograph determination. The baseflow in a river is considered to be the water that comes from groundwater sources, rather than from overland runoff like precipitation or snowmelt (Stoelzle et al., 2020). The overland runoff components influences the flow in the rivers on a scale of minutes to hours, whereas the baseflow components takes days to weeks to influence the river flows (Lyne and Hollick, 1979). The flood hydrograph begins when the overall flow exceeds the baseflow, and ends when the two are equal again, as there is no water remaining that is a part of overland runoff and the only water source for the rivers are the groundwater sources.

Two methods of baseflow separation were tested. These are the Baseflow Index Methods from the LF Stat Package in R (Koffler et al., 2016), and the Recursive Digital Filter Method from the EcoHydrology Package in R (Fuka et al., 2018). The BFI Method is presented in detail in the World Meteorological Organization's Manual on Low-flow Estimation and Prediction (World Meteorological Organization, 2008). The steps for calculating the BFI are as follows (World Meteorological Organization 2008) :

1. Split the discharge data into blocks of 5 days (from the entire time series from 1979-2019).
2. Select the minimum flow from the 5-day period and label it as Q_{min}.
3. In the series of Q_{min} values, identify the turning points. A $Q_{min(i)}$ is a turning point if it is less than 90% of both the $Q_{min(i-1)}$ and $Q_{min(i+1)}$.
4. Construct the baseflow hydrograph by drawing straight lines between each turning point.
5. Linearly interpolate the discharge values in between the turning points.

Following these 5 steps creates a daily baseflow value which can be compared against the daily flow discharge values to determine the hydrograph.

The Recursive Digital Method was completed by computing a formula to determine the overland flow component of the discharge. The formula to determine the overland runoff is (Lyne and Hollick, 1979):

$$f_k = \propto f_{k-1} + \frac{(1+\propto)}{2}\left(y_k - y_{k-1}\right) \quad (1)$$

Where f_k is the filtered overland flow component at the kth sampling instant, in this case it is the kth day of the discharge data series, y_k is the discharge, and α is the filter parameter. The filter parameter α is the degree of attenuation of the soil (Nathan and McMahon, 1990). Nathan and McMahon (1990) recommend values ranging from 0.9-0.95 for the degree of attenuation. The method was also run multiple times, where Nathan and McMahon (1990) recommend 3 runs, once forward, once backwards, and once forwards again. The higher the number of runs increases the amount of smoothing of the

curves. As f_k represents the overland flow component of the discharge, the baseflow component was taken as the overland component subtracted from the total flow.

When determining which method to proceed with for the baseflow separation the flood hydrographs must be inspected. An example can be seen in Figure 8. The total flow is visualized with the black line. The various baseflow methods are plotted below the total flow line. The BFI method is visualized with the orange line. Three different recursive digital method trials were completed with an α of 0.9, 0.925 and 0.95, seen with the yellow, pink, and green lines, respectively. The vertical red lines indicate the start and end of the hydrograph as determined by the BFI method. When looking at the second vertical line, indicating the end of the hydrograph, only the BFI method ends the hydrograph before the flow beings to increase again. The BFI method is the only baseflow method where the baseflow is equal to the total flow. Due to the duration of the hydrograph, the discharge in this example should be viewed as two separate events, of which only the BFI correctly identifies the end of the hydrograph. When investigating more hydrographs, this event was repeated numerous times. The digital recursive methods often create hydrographs lasting longer than one month. For the purposes of this thesis, the BFI method was used to determine all hydrographs.

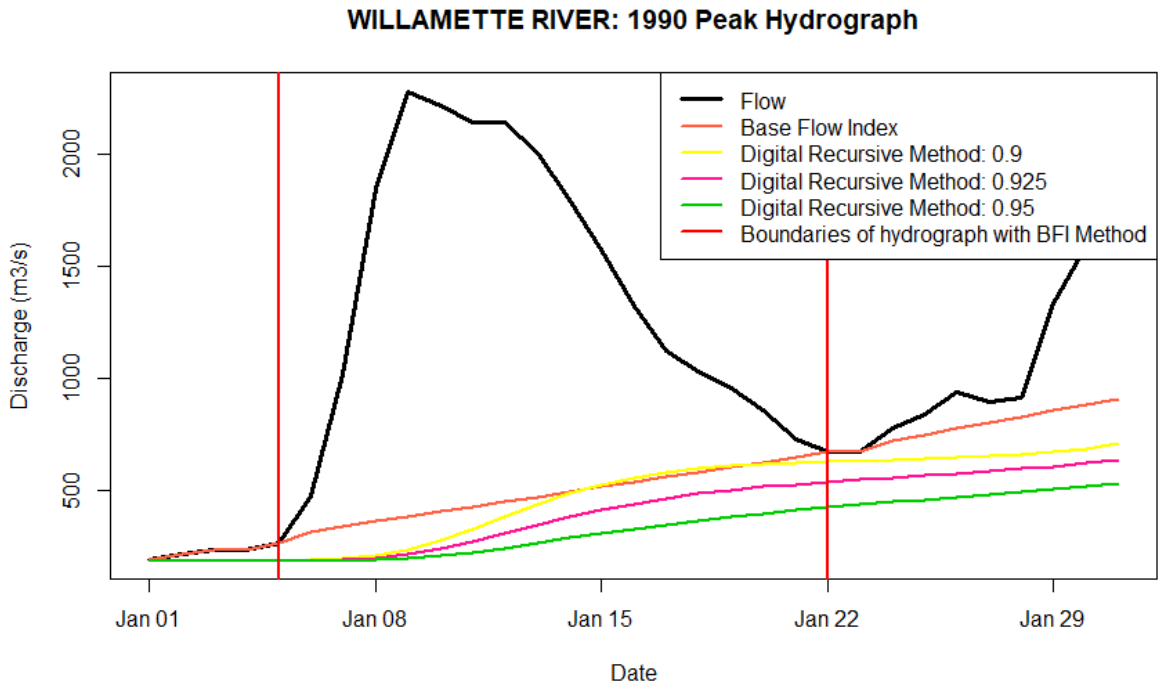

Figure 8: Example of baseflow separation for the Willamete River

As stated above, the start and end of the hydrograph are when the baseflow is equal to the total discharge. There are situations however when long hydrographs are created. This is often true when there is a long falling limb of the hydrograph, as seen in Figure 9. To reduce the overall duration of some hydrographs, allowing for more manageable data sizes for analyses, the start and end of the hydrograph were determined to be when the baseflow was within 10% of the original flow. In Figure 9 the hydrograph length was reduced by approximately one week due to this assumption, indicated by the vertical blue line.

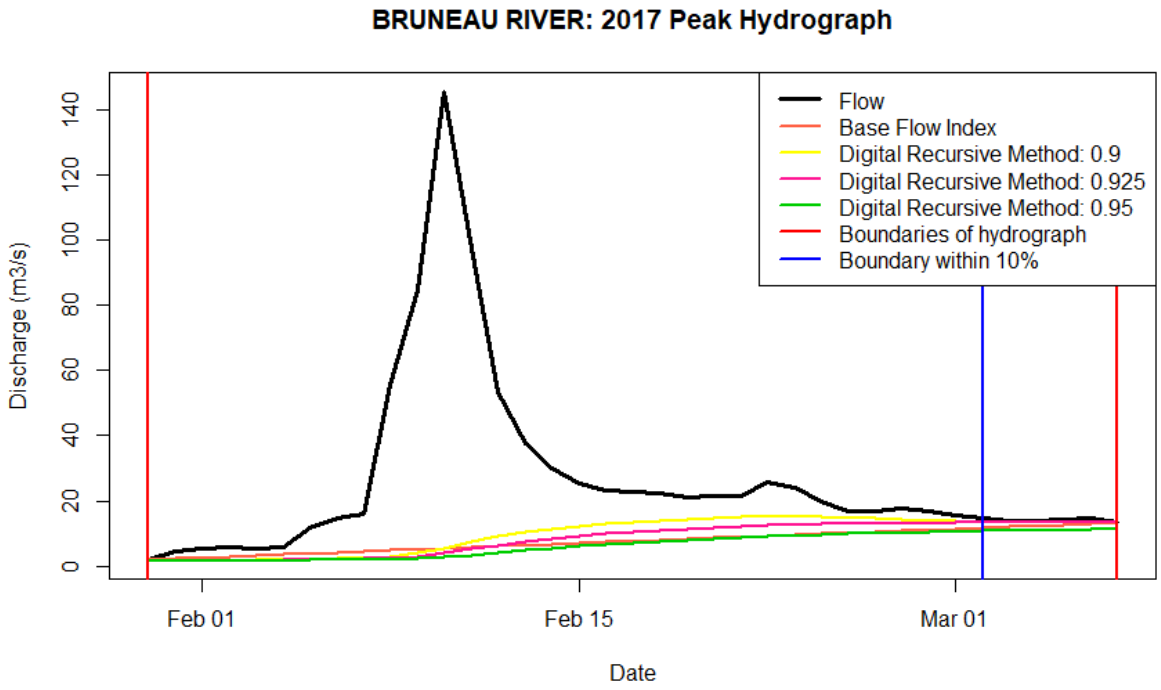

Figure 9: Bruneau River peak hydrograph

3.1.4 Climate data processing

The climate data had to be trimmed to the required areas, and the remaining data discarded for the reasons of data storage and processing times. The brick() function, from the Raster Package in R was used to upload the .nc data into R (Hijmans et al, 2021). Figure 7 displays an example of the plotted .nc data in R. Once uploaded, the data for each catchment could be extracted. Using the extract() function from the Raster Package (Hijmans et al., 2021), a list of values, cropped for a single catchment was created, one value for each grid cell in the catchment. The list of values was all values for the single day of the .nc file. Once this list was extracted, the maximum, minimum, mean and Gini values were found. The Gini was found by using the gini() function from the Reldist Package in R (Handcock, 2016). Outside of hydrology, the Gini value is commonly used as a way to express wealth disparities between the rich and poor populations in different countries (Hayes and Anderson, 2021). However, it can be used to measure any form of disparity. In this situation it represents how constant or inconsistent the climate value was across the catchment. For example, a high Gini value for precipitation suggests that there was heavy rainfall in one area of the catchment, but little in another part of the catchment.

This process was repeated for each of the 107 catchments in the study. Once this was completed, the data values were written into a matrix. This was then repeated for each day of the year. Each day of the year came as a unique layer in the .nc file. Once the year was completed, the table was completed, a new .nc file was uploaded and the process was repeated. This was then repeated for all years in the study, from 1979-2019. Once completed, the entire process was repeated for each of the downloaded climate variables. The final step was to convert the data tables into a more useable format, by combining columns into one table per catchment, with all the daily climate factors per year. The tables were exported as a .txt file, which allows for much quicker data processing times than the .nc files.

3.1.5 Climate data per hydrograph

The final step in the data processing was the creation of tables listing all the climate values for each hydrograph. As the hydrograph start date and end date are already known from the hydrograph determination, this process involved extracting the certain rows of climate data that related to each day of the flood. The days of the hydrograph were expressed as the Julian Day of each year, therefore the row of the table corresponding to the Julian Date was extracted. This resulted in a table being created that has all the climate data for each hydrograph in the study. Figure 10 displays an example of temperature data for the 2016 hydrograph of the Saddle River in the USA.

	Julian_Day	Year	Flow	Baseflow	Temperature_mean	Temperature_max	Temperature_min	Temperature_GINI
1	54	2016	2.356	2.311842	274.714	275.433	274.102	0.001
2	55	2016	9.090	2.310684	279.854	281.040	279.008	0.002
3	56	2016	28.317	2.309526	281.300	281.703	280.846	0.001
4	57	2016	6.145	2.308368	273.618	274.389	272.980	0.001
5	58	2016	4.191	2.307211	273.712	273.960	273.334	0.001
6	59	2016	3.794	2.306053	282.835	283.070	282.559	0.000
7	60	2016	3.568	2.304895	283.642	284.019	283.197	0.001
8	61	2016	3.228	2.303737	278.352	278.780	277.882	0.001
9	62	2016	3.851	2.302579	277.520	278.298	276.874	0.001
10	63	2016	3.058	2.301421	272.243	272.650	271.782	0.001
11	64	2016	2.888	2.300263	272.911	273.113	272.552	0.000
12	65	2016	2.792	2.299105	273.909	274.239	273.485	0.001
13	66	2016	2.631	2.297947	276.105	276.318	275.739	0.000
14	67	2016	2.560	2.296789	281.988	282.163	281.641	0.000
15	68	2016	2.517	2.295632	285.371	285.984	284.671	0.001
16	69	2016	2.472	2.294474	289.571	290.577	288.085	0.002

Figure 10: An example of hydrograph basic and temperature data for the Saddle River in the USA

3.1.6 Elevation data

The extraction of the elevation data was completed with a similar method to the extraction of the climate data. The elevation data was loaded into R using the raster() function from the Raster Package in R (Hijmans, 2021). Then the extract() function from the Raster Package (Hijmans, 2021) was used to create a list of data values inside each of the catchments. From this list of values, the mean, median and Gini values were determined for each catchment. The Gini value is a representation of how mountainous the catchment is, with a high Gini signifies a large range in elevation.

3.2 Analysis methodology

3.2.1 Flood type separation

In order to investigate the compound events, the first step was to sort the floods into different typologies, including pre-conditioned compound events and multivariate compound events. The method followed in this study was similar to a process completed by Sikorska et al. (2015). They split floods into 6 different typologies; flash floods, rain-on-snow floods, snowmelt floods, short precipitation floods, long precipitation floods, and glacier melt floods. We should note that due to the data collected, the determination of flash floods was not possible for this study. Precipitation data was downloaded only as daily means. However, flash floods are caused by short, very intense rainfalls. Sikorska et al. (2015)

classified flash floods as over 12 mm of rain falling in less than half a day, which this study did not have the data available to classify. Furthermore, glacial melt floods could also not be classified in this study due to a lack of glacial data.

As the purpose of this study was to investigate compound events, the selected main flood typologies (rain-on-snow floods, snowmelt floods, long precipitation floods, and short precipitation floods) were further split up into preconditioned events, where there was a wet antecedent moisture condition in the soil and snow. These resulted in 8 flood types presented in Table 4, which provides a summary of the rules created to sort the floods.

Table 4: Flood typology sorting requirements

Flood Type	Precipitation	Snow Cover	Snowmelt	Antecedent Moisture Condition	Other	Abbreviation
Rain-on-Snow Flood with Dry Conditions	>12 mm[(a)]	>5% [(a)]	>1 mm[(a)]	>20% Increase in Snow Density[(b)]		ROS-D
Rain-on-Snow Flood with Wet Conditions	>12 mm[(a)]	>5% [(a)]	>1 mm[(a)]	<20% Increase in Snow Density[(b)]		ROS-W
Snowmelt Flood with Dry Conditions	<12 mm[(a)]	>5% [(a)]	>1 mm[(a)]	>20% Increase in Snow Density[(b)]		SMF-D
Snowmelt Flood with Wet Conditions	<12 mm[(a)]	>5% [(a)]	>1 mm[(a)]	<20% Increase in Snow Density[(b)]		SMF-W
Long Precipitation Floods with Dry Conditions	>25 mm over 4 days[(d)]	<5% [(a)]	<1 mm[(a)]	<75% Soil saturation at start of hydrograph	Multiple Peaks[(c)]	LPF-D
Long Precipitation Floods with Wet Conditions	>25 mm over 4 days[(d)]	<5% [(a)]	<1 mm[(a)]	>75% Soil saturation at start of hydrograph	Multiple Peaks[(c)]	LPF-W
Short Precipitation Floods with	>12 mm in 1 day[(a)]	<5% [(a)]	<1 mm[(a)]	<75% Soil saturation at start of hydrograph		SPF-D

Dry Conditions						
Short Precipitation Floods with Wet Conditions	>12 mm in 1 day[(a)]	<5% [(a)]	<1 mm[(a)]	>75% Soil saturation at start of hydrograph		SPF-W

(a) (Sikorska et al., 2015)
(b) (Kuusisto, 1984)
(c) (Fischer et al., 2019)

The rules and thresholds for the flood type classification were determined from a literature review. Sikorska et al. (2015) concluded that if there was greater than 5% snow coverage in a catchment then it was assumed that the flood was influenced by snow, and the flood type was either a rain-on-snow (ROS) flood or a snowmelt (SMF) flood. If precipitation occurs falling onto the snowfall, then it was a ROS flood. The threshold of ROS floods was set to 12 mm, causing greater than 1 mm of snowmelt. The amount of snowmelt was found by taking the LWE and subtracting the previous day's LWE. The total solid snow thickness was not used in this calculation because snow thickness can reduce, thus changing the density of the snow, without creating snowmelt runoff (Kuusisto, 1984). An SMF flood occurs when there was greater than 1 mm of snowmelt, but less than 12 mm of precipitation falls.

When determining whether the snow cover was in a wet initial condition versus a dry initial condition the snow density was investigated. Kuusisto (1984) investigated the density of snow as it melts. The density of snow varies on many factors including temperature, snow thickness, precipitation, etc. Snow density also varies regionally, and thus directly comparing the snow densities between the catchments was unlikely to yield accurate results. However, the percent increase during the melting process can be compared. Kuusisto (1984) determined that during the final stages of the melting process that the snow density increased greater than 20%. This can be used as the threshold to determine whether the snow has an initial condition of being wet or dry. If, from the start of the hydrograph to the day of maximum snowmelt flood, the snow density increased by greater than 20% it was assumed that the snow was not close to melting and thus was in a dry initial condition. If the density increase was less than 20%, then it was assumed that the snow was already close to melting, and thus was a wet initial condition. The day of maximum snowmelt was used to determine the change in density, rather than the day of the peak of the hydrograph because there are situations whereby the day of the peak hydrograph, snow coverage and snow thickness was zero, indicating that all the snow has already melted. To determine the snow density, the LWE was compared to the total solid snow thickness.

If there was less than 5% snow coverage, then it was assumed that snowmelt was negligible in the flood volume and timing of the flood (Sikorska et al., 2015). In this scenario the cause of the flood was from precipitation. The precipitation floods were split into long precipitation floods (LPF) and short precipitation floods (SPF). An SPF occurs when all the rain falls in a single day, and that rainfall was greater than 12 mm. A long precipitation flood occurs when the rain falls from 2-4 days, and the total volume of rain cumulatively was more than 25 mm. These thresholds came from Sikorska et al. (2015) and Fischer et al. (2019). Additionally, if there are multiple peaks in the hydrograph, it was assumed to

be an LPF (Fischer et al., 2019). The findpeak() function from the "Pracma" package was used to determine the number of peaks in the hydrograph (Borchers, 2021).

To determine if the initial conditions for the precipitation floods were wet or dry, the soil moisture was used. The only soil moisture data available for download that fit the purposes of the thesis was volumetric soil water content data rather than saturation percent data. To determine the percent saturation all data for each catchment was searched through, and the largest single daily volumetric water content was found for each individual catchment. This maximum water content was assumed to be at 100% saturation. All other daily values were then taken as the numerator over this maximum saturation data, thus creating daily percent saturation content. The threshold for wet conditions was initial soil moisture contents of greater than 75%, and wet soil moisture conditions was less than 75%.

By applying these restrictions to each of the floods in the study, they can be sorted into these 8 categories of floods. If the flood does not apply to any one of the 8 pre-determined categories, it was placed in the category "Other". The process was completed once for the dataset containing the yearly maximum floods, and once containing the dataset containing the maximum floods occurring with snow present. Figure 11 visually displays the data sorting process.

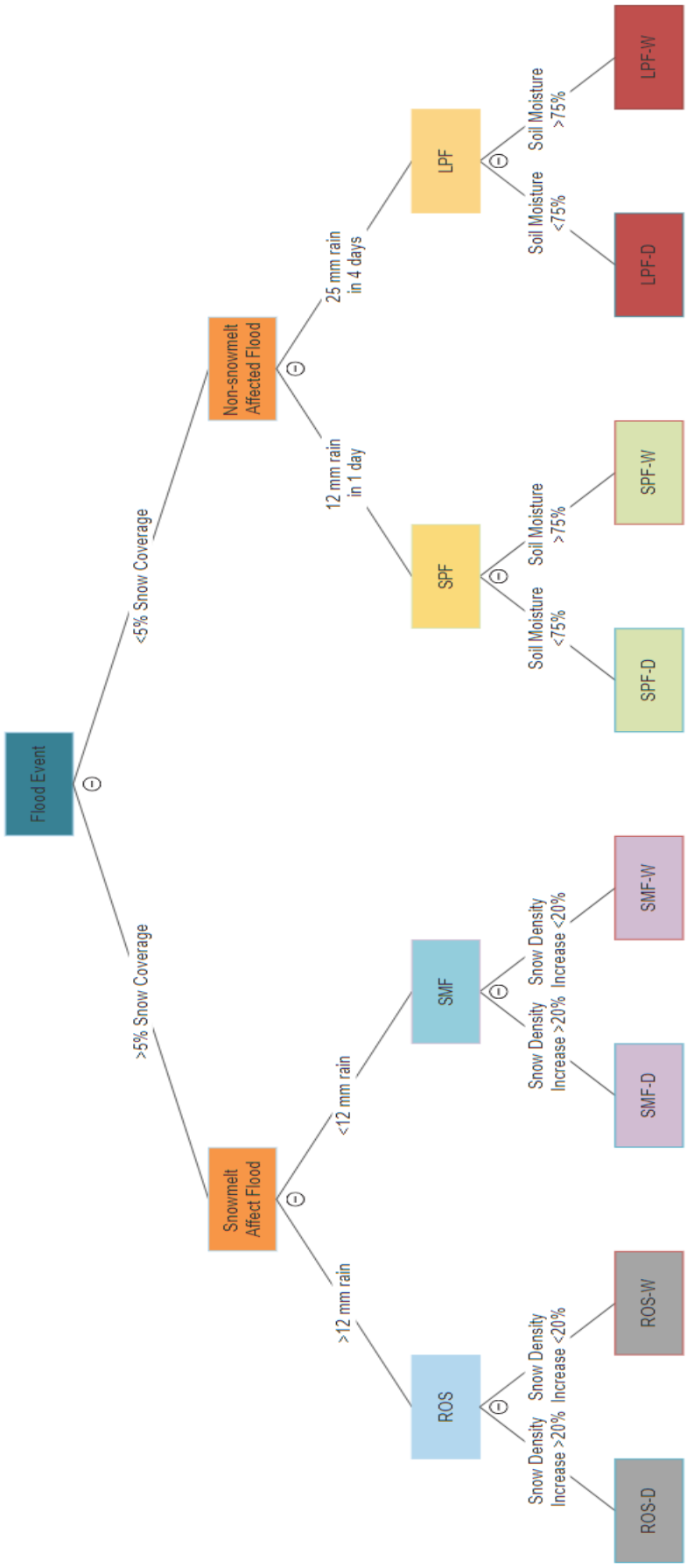

Figure 11: Decision tree for flood typologies

3.2.2 Seasonality

The seasonality of floods for each of the catchments was investigated following the method proposed by (Burn, 1997). This method determined the average date of the flood, therefore determining which season each catchment was most likely to flood. It also determined the variability of the seasonality. A large variability means that the floods are equally likely to occur in any season, whereas a low variability means that the largest flood is likely to occur in the same season.

The first step was to translate the peak flood dates into an angular value. This can be plotted onto a circular graph where January 1st is the value at 0 degrees, pointing straight up. With each subsequent day, the angular value increases until the final day of the year, December 31st is at 360 degrees, also pointing straight up. To do this, the maximum flood date must first be in Julian Date (JD) format, where January 1st has a value of 1, and December 31st has a value of 365 (or 366 on leap years). The JD can then be transformed into the angular value with the formula:

$$\theta_i = JD_i \frac{2\pi}{365} \tag{2}$$

θ_i is a value in radians. The θ value was then found for each of the floods per catchment. The θ value could then be broken into vector components, with an x and y component for each θ. The average x and y values could then be found using the following formulas:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \cos(\theta_i) \tag{3}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \sin(\theta_i) \tag{4}$$

Once these values were determined for the catchment, the mean direction of the flood could be determined using the formula:

$$\bar{\theta} = \tan^{-1}\frac{\bar{y}}{\bar{x}} \tag{5}$$

And then converted back into a JD with the formula:

$$JD_{mean} = \bar{\theta}\frac{365}{2\pi} \tag{6}$$

Finally, the variability could be calculated as the resultant between the x and y values with the formula:

$$\bar{r} = \sqrt{(\bar{x}^2 + \bar{y}^2)} \tag{7}$$

3.2.3 Generalized boosted regression trees

Generalized boosting regression trees are a type of machine learning often used in classification and regression analysis. Regression trees are a machine learning method in which an output is estimated,

based on the values of the inputs. A series of binary decisions splits the data into smaller and smaller sections until eventually the data is split into final categories (Elith et al., 2008). Regression trees are a common machine learning method as they are intuitive and easy for non-experts to understand. An example of a regression tree, estimating the flow in a hydrograph can be seen in Figure 12.

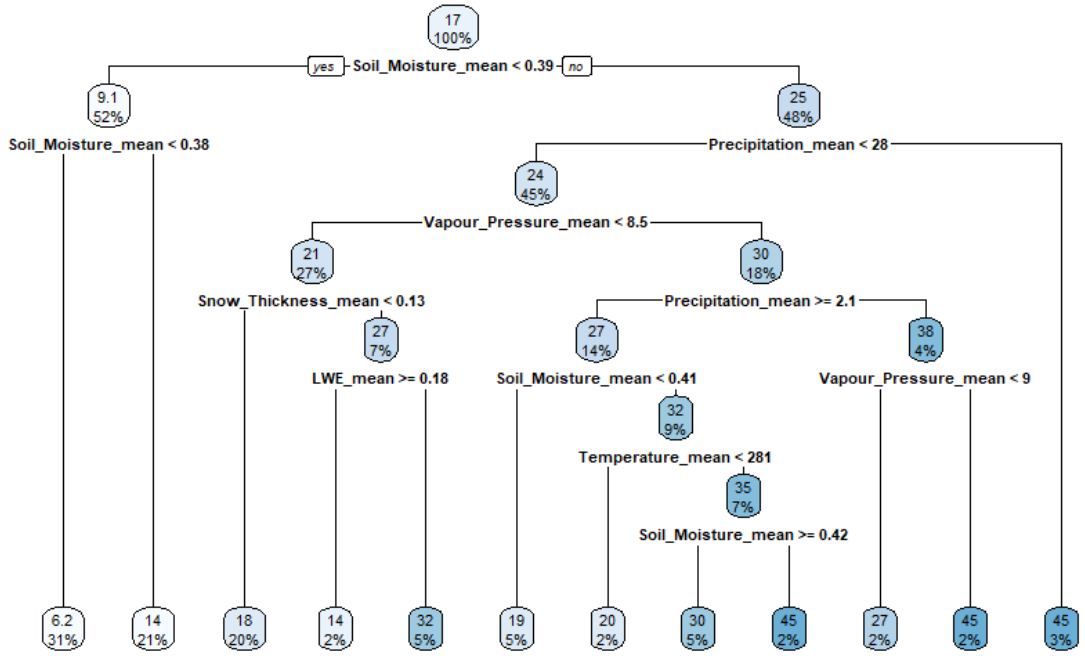

Figure 12: Regression tree example estimating river discharge

Boosting is a machine learning technique which reduces the overall error in the training stage. After the first model is built, boosting takes random data points and determines the error at each data point. The model is then re-created, and the error is determined for more random data points (Elith et al., 2008). The data points which had the error previously determined are weighted so that they are more likely to be chosen again. Boosting adds additional nodes to the trees until the output with the lowest overall error is found. The program keeps track of the various changes made to the models, and then is able to determine the most important variables in reducing the error in the prediction of the model (Elith et al., 2008).

The combination of boosting with regression trees combines both methods, resulting in more accurate results. Additionally, there is a measure of relative importance that each of the variables holds when reducing the error in the boosting of the regression trees. This relative importance can be used as a rank to determine which variables are the most influential in the prediction of the desired variable (Elith et al., 2008). The relative influence of a variable was found using the following formula from (Friedman, 2001):

$$I_J = \left\{ E_x \left[\frac{\partial \hat{F}_x}{\partial x_j} \right]^2 \times var_x(x_j) \right\}^{1/2} \quad (8)$$

Where I_j is the relative influence, F_x is the is the predicted outcome of the formula, and x_j are the individual inputs. The above formula is for a single regression tree however, and the program repeats the formula many times in order to reduce the overall error. Therefore, the final relative importance formula comes as:

$$I_j^2 = \frac{1}{M}\sum_{m=1}^{M} I_j^2\,(T_m) \tag{9}$$

Where T_m is a collection of decision trees with M members.

In this study case, the climate data was used to estimate the flow. The hydrograph data was loaded into a matrix, with each column representing a different climate variable, and each row representing a day, with the total rows equaling the hydrograph. The GBM Package in R was used to conduct the analysis (Greenwell et al., 2020). The function gbm() was used to predict the flow, based on the climate data. The gaussian distribution was used in the gbm package, and thus the squared error was the method in error reduction. The number of trees fit to determine the relative influences of the variables was 100 trees for each catchment.

3.2.4 Correlation matrices

A correlation matrix is a statistical tool used to determine the relationship between different variables. In this case, the relationship is between the river discharge and the climate factors. A strong correlation between the two means that the two are intricately linked, when one value changes, the other value is likely to change as well. It is a measure of the effect of change in one variable when the other changes (Benesty, et al., 2009). The values from the correlation analysis range from -1 to 0 to 1. Negative numbers indicate a negative correlation (as one variable increases, the other variable decreases), and positive numbers indicate a positive correlation (as one variable increases, the other variable also increases. Values closer to one and negative one indicate a stronger correlation, and values closer to 0 indicate a weaker correlation. The formula to determine the correlation coefficient is as follows (Benesty, et al., 2009):

$$r = \frac{n\sum xy - (\sum x \sum y)}{\sqrt{[n\sum x^2 - (\sum x^2)][n\sum y^2 - (\sum y^2)]}} \tag{10}$$

Where n is the number of pairs, x is variable one, and y is variable two. To determine the correlation of the different climate factors, the cor() function was used from the Stats package in R. The Stats package is a base package built into the R program with the authors and maintainers of the program being the R Core Team.

The correlation factors include the daily flow, mean temperature, mean precipitation, mean LWE, mean snow thickness, mean windspeed, mean vapour pressure, mean soil moisture, snow cover and snowmelt. A correlation matrix was made for each year, based on the daily climate factors on the days of the hydrographs, for every catchment. It was however important to summarize each of the catchments to determine the trends in the correlation values catchment wide. Therefore, the mean value was found for each of the absolute correlation values for each catchment.

3.2.5 Sorting catchments by characteristics

In order to analyze the data, the catchments were split by climate zone, area and elevation. This allowed for patterns and relationships to be determined.

3.2.5.1 Analysis based on climate zone

The downloading of the climate zone data was described in Section 2.1.2. The climate zone that covered the majority of the area was chosen as the dominant climate zone of the catchment. A summary of the climate zones can be seen in Figure 13. The graph shows that there are three dominant climate zones in the study; the temperate oceanic climate, warm summer humid continental climate and subarctic climate zones. The tundra climate zone also has 7 catchments, so it will also be included in the analyses. The other climate zones only have 1-3 catchments, which is not a large enough sample size for the analyses to be completed. The climate zones for the selected catchments on a map can be seen in Figure 14 and Figure 15.

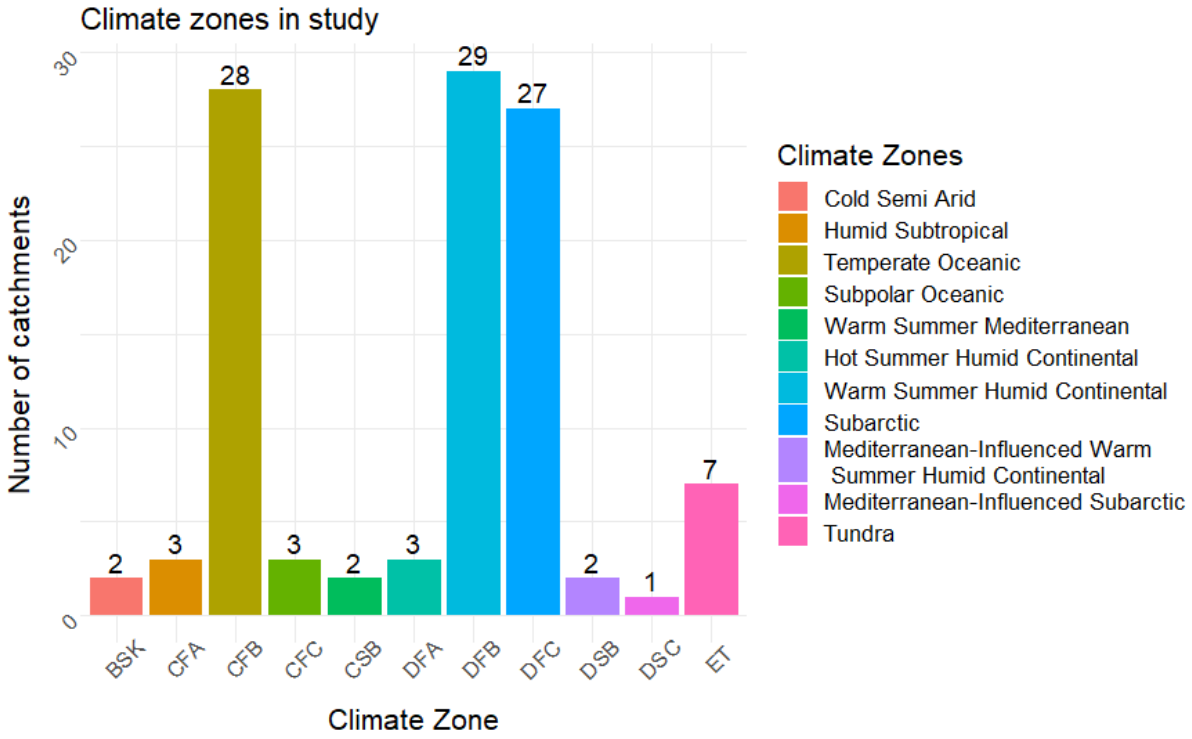

Figure 13: Distribution of climate zones of the considered catchments

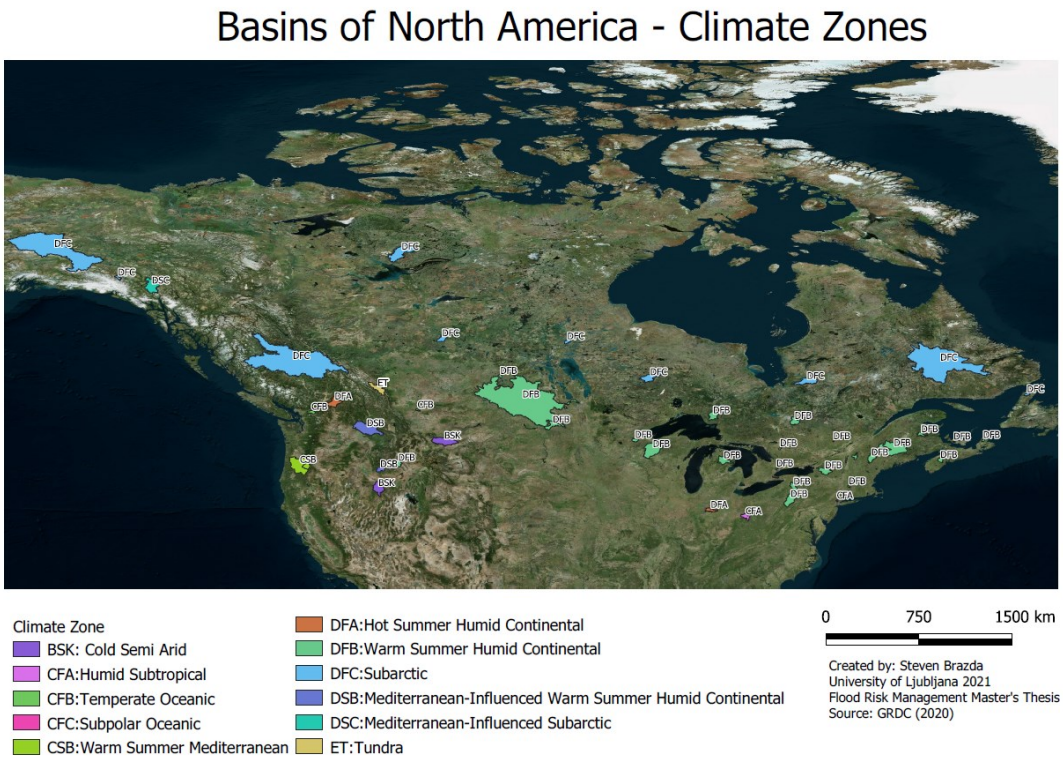

Figure 14: Climate zones for the catchments of North America

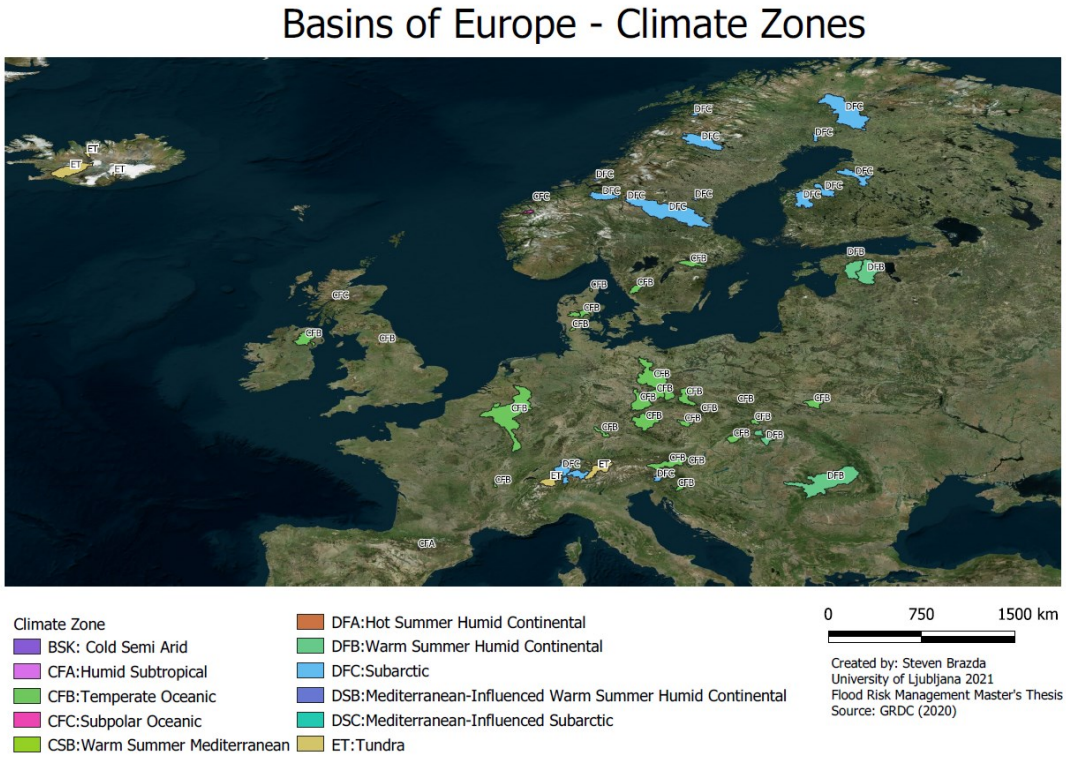

Figure 15: Climate zones for the catchments of Europe

3.2.5.2 Analysis based on area

Additional analysis were also conducted based on the areas of the catchments. The catchments were split into three categories, namely small, medium and large. Small catchments are less than 200 km^2,

large are greater than 10,000 km^2 and medium is in between 200 and 10,000 km^2. Table 5 provides the list of catchment sizes, and Figure 16 displays the distribution of the catchment sizes.

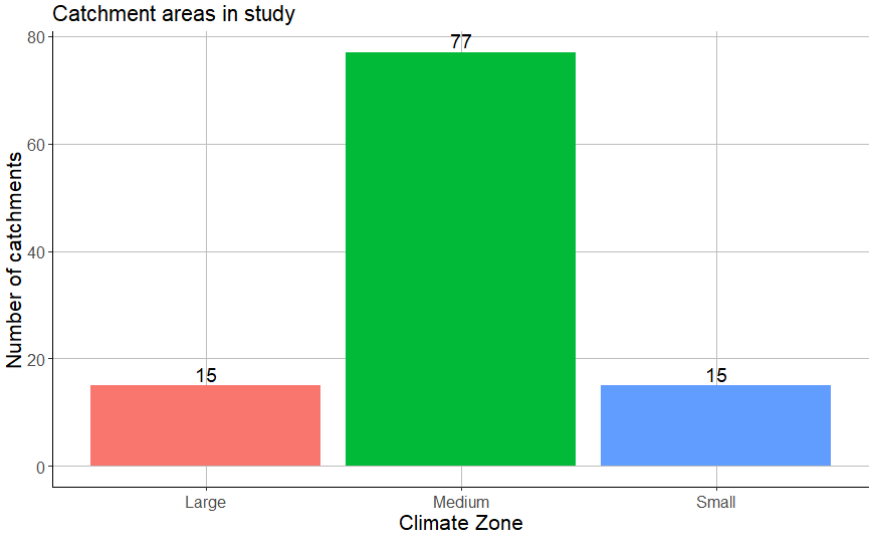

Figure 16: Catchment area distribution of the considered catchments

Table 5: Catchment size categories

Number	GRDC Number	Country	River Name	Area (km^2)	Size
1	6139360	France	AZERGUES	332	Medium
2	4244915	Canada	NORTHEAST MARGAREE RIVER	369	Medium
3	4243200	Canada	SAINT LOUIS (RUISSEAU)	40	Small
4	4207910	Canada	CHILLIWACK RIVER	1223	Medium
5	4102100	USA	KUSKOKWIM RIVER	80177	Large
6	4103600	USA	TANANA RIVER	64095	Large
7	4113360	USA	SHEYENNE RIVER	4959	Medium
8	4115100	USA	WILLAMETTE RIVER	18807	Large
9	4115106	USA	MCKENZIE RIVER (TRIB. COLUMBIA)	2399	Medium
10	4116350	USA	BRUNEAU RIVER	6969	Medium
11	4115400	USA	SPOKANE RIVER	15909	Large
12	4116370	USA	PAHSIMEROI RIVER	2133	Medium
13	4116461	USA	BOISE RIVER	2148	Medium
14	4119441	USA	CHIPPEWA RIVER (TRIB. MISSISSIPPI)	14478	Large
15	4120321	USA	MUSSELSHELL RIVER	10375	Large
16	4123200	USA	WHITE RIVER (TRIB. OHIO)	2078	Medium
17	4123280	USA	HOCKING RIVER	2430	Medium
18	4143790	USA	MISSISQUOI RIVER	1235	Medium
19	4132100	USA	NEMADJI RIVER	1125	Medium
20	4134700	USA	AU SABLE RIVER	4595	Medium
21	4136301	USA	GENESEE RIVER	2529	Medium
22	4136500	USA	BLACK RIVER (TRIB. LAKE ONTARIO)	4837	Medium

23	4147010	USA	PENOBSCOT RIVER	17317	Large
24	4147111	USA	ANDROSCOGGIN RIVER	3506	Medium
25	4147470	USA	WESTFIELD RIVER	249	Medium
26	4147540	USA	SADDLE RIVER	156	Small
27	4147726	USA	SUSQUEHANNA RIVER	7693	Medium
28	4213465	Canada	IRONSPRING CREEK	538	Medium
29	4234550	Canada	NORTH MAGNETAWAN RIVER	314	Medium
30	4213650	Canada	ASSINIBOINE RIVER	147113	Large
31	4236300	Canada	EAST HUMBER RIVER	196	Small
32	4203335	Canada	DUKE RIVER	615	Medium
33	4203910	Canada	TAKHINI RIVER	7247	Medium
34	4207310	Canada	FRASER RIVER	113759	Large
35	4208370	Canada	YELLOWKNIFE RIVER	10504	Large
36	4213046	Canada	BOW RIVER	5362	Medium
37	4213700	Canada	TAYLOR RIVER	891	Medium
38	4214295	Canada	DILLON RIVER	2312	Medium
39	4214420	Canada	PIPESTONE RIVER (TRIB. HUDSON BAY)	5620	Medium
40	4214610	Canada	HARRICANAW RIVER	3667	Medium
41	4214690	Canada	PONTAX RIVER	6075	Medium
42	4215103	Canada	OKANAGAN RIVER	5964	Medium
43	4220510	Canada	VERDIGRIS COULEE	343	Medium
44	4232750	Canada	WHITE RIVER (TRIB. LAKE SUPERIOR)	4121	Medium
45	4236050	Canada	TWENTY MILE CREEK	307	Medium
46	4244500	Canada	CHURCHILL, FLEUVE (LABRADOR)	91261	Large
47	4244720	Canada	TORRENT RIVER	644	Medium
48	4244870	Canada	LITTLE SOUTHWEST MIRAMICHI RIVER	1346	Medium
49	4244900	Canada	DUNK RIVER	137	Small
50	4244985	Canada	LAHAVE RIVER	1251	Medium
51	6140250	Czechia	BEROUNKA	8296	Medium
52	6140700	Czechia	DIVOKA ORLICE	183	Small
53	6142520	Slovakia	NITRA	2088	Medium
54	6140450	Czechia	SAZAVA	1492	Medium
55	6142680	Slovakia	VAH	1111	Medium
56	6144100	Slovakia	SAJO	1803	Medium
57	6233100	Sweden	VISKAN	2136	Medium
58	6220500	Belgium	DYLE	641	Medium
59	6226700	Spain	VERO	83	Small
60	6233220	Sweden	TAENNAN (LJUSNAN)	211	Medium
61	6233221	Sweden	LJUSNAN	20162	Large
62	6233440	Sweden	NYKOEPINGSAEN	2239	Medium
63	6233680	Sweden	VINDELAELVEN (UMEAELVEN)	6161	Medium
64	6233550	Sweden	KASSJOEAN (LJUNGAN)	168	Small

65	6342520	Germany	ALTMUEHL	1405	Medium
66	6233870	Sweden	SANGISAELVEN	519	Medium
67	6235100	Austria	BREGENZER ACH	228	Medium
68	6242200	Austria	ILZBACH	196	Small
69	6243030	Austria	INN	5755	Medium
70	6246611	Austria	MUR	6805	Medium
71	6246700	Austria	MUERZ	723	Medium
72	6340600	Germany	VEREINIGTE MULDE	6164	Medium
73	6340700	Germany	SCHWARZE ELSTER	3078	Medium
74	6340510	Germany	HAVEL	15472	Large
75	6401090	Iceland	OELFUSA	5769	Medium
76	6401500	Iceland	DJUPA, FLJOTSHVERFI	214	Medium
77	6401601	Iceland	SVARTA, SKAGAFIROI	351	Medium
78	6457707	Poland	BOBR	4359	Medium
79	6457880	Poland	LISWARTA	141	Small
80	6458203	Poland	SKAWA	834	Medium
81	6458713	Poland	WIEPRZ	2987	Medium
82	6545200	Slovenia	KRKA	997	Medium
83	6545190	Slovenia	SAVA	902	Medium
84	6604170	United Kingdom	NEVIS	78	Small
85	6603300	United Kingdom	LOWER BANN	5011	Medium
86	6605545	United Kingdom	SNAIZEHOLME BECK	11	Small
87	6731175	Norway	EIDSELV	391	Medium
88	6731010	Norway	ENGESETELV	39	Small
89	6731070	Norway	NORDELVA	194	Small
90	6731501	Norway	GAULA	3060	Medium
91	6731685	Norway	KOBBELV	306	Medium
92	6934800	Denmark	BREDE A	310	Medium
93	6744200	Romania	MAROS	27221	Large
94	6934250	Denmark	GUDENA	1296	Medium
95	6854210	Finland	AHTAVANJOKI	2393	Medium
96	6934300	Denmark	UGGERBY A	145	Small
97	6854320	Finland	PYHAJOKI	3646	Medium
98	6854710	Finland	OUNASJOKI	13198	Large
99	6854900	Finland	KYRONJOKI	4386	Medium
100	6934100	Denmark	SKJERN A	1087	Medium
101	6939500	Switzerland	RHONE	3763	Medium
102	6935145	Switzerland	RHINE RIVER	3221	Medium
103	6935310	Switzerland	REUSS	3398	Medium
104	6948120	Switzerland	MAGGIA (TRIB. PO)	929	Medium
105	6172050	Estonia	PAERNU JOGI	5158	Medium
106	6172010	Estonia	LEVAJOKI	94	Small
107	6172200	Estonia	EMAJOGI	8014	Medium

3.2.5.3 Analysis based on elevation

Additional analysis were also conducted based on the elevation of the catchments. The catchments were split into three categories, namely low, medium and high. Low catchments are less than 500 metres above sea level (masl), high catchments are greater than 1000 masl and medium is in between 500 and 1000 masl. Table 6 provides the list of catchcatchment elevtations, and Figure 17 displays the distribution of the catchment elevations.

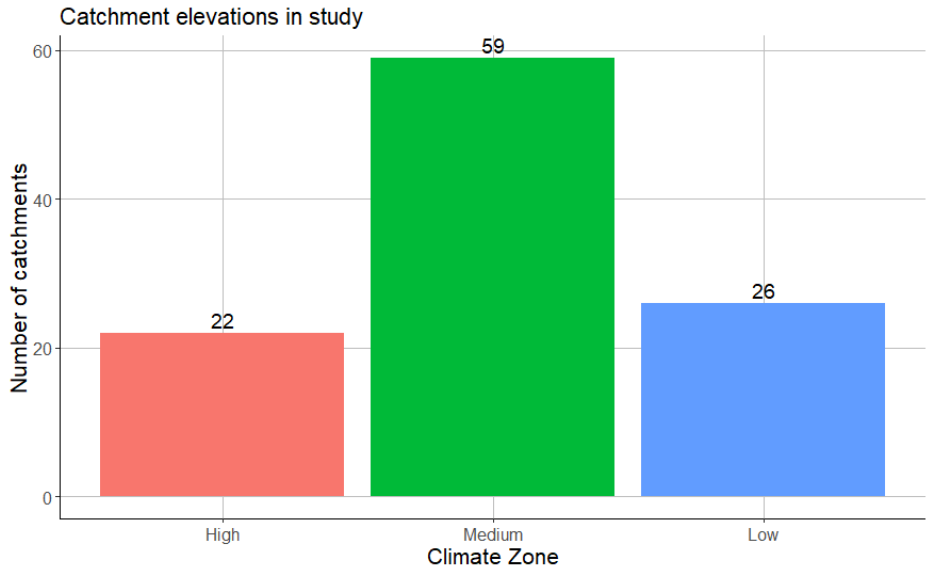

Figure 17: Catchment elevation distribution of the considered catchments

Table 6: Catchment elevation categories.

Number	GRDC Number	Country	River Name	Mean Elevation (m)	Elevation Class
1	6139360	France	AZERGUES	531	Medium
2	4244915	Canada	NORTHEAST MARGAREE RIVER	363	Low
3	4243200	Canada	SAINT LOUIS (RUISSEAU)	588	Medium
4	4207910	Canada	CHILLIWACK RIVER	1170	High
5	4102100	USA	KUSKOKWIM RIVER	436	Low
6	4103600	USA	TANANA RIVER	849	Medium
7	4113360	USA	SHEYENNE RIVER	495	Low
8	4115100	USA	WILLAMETTE RIVER	665	Medium
9	4115106	USA	MCKENZIE RIVER (TRIB. COLUMBIA)	1186	High
10	4116350	USA	BRUNEAU RIVER	1712	High
11	4115400	USA	SPOKANE RIVER	1000	Medium
12	4116370	USA	PAHSIMEROI RIVER	2245	High
13	4116461	USA	BOISE RIVER	1956	High
14	4119441	USA	CHIPPEWA RIVER (TRIB. MISSISSIPPI)	423	Low
15	4120321	USA	MUSSELSHELL RIVER	1442	High

16	4123200	USA	WHITE RIVER (TRIB. OHIO)	288	Low
17	4123280	USA	HOCKING RIVER	278	Low
18	4143790	USA	MISSISQUOI RIVER	350	Low
19	4132100	USA	NEMADJI RIVER	328	Low
20	4134700	USA	AU SABLE RIVER	359	Low
21	4136301	USA	GENESEE RIVER	569	Medium
22	4136500	USA	BLACK RIVER (TRIB. LAKE ONTARIO)	475	Low
23	4147010	USA	PENOBSCOT RIVER	268	Low
24	4147111	USA	ANDROSCOGGIN RIVER	577	Medium
25	4147470	USA	WESTFIELD RIVER	442	Low
26	4147540	USA	SADDLE RIVER	80	Low
27	4147726	USA	SUSQUEHANNA RIVER	533	Medium
28	4213465	Canada	IRONSPRING CREEK	561	Medium
29	4234550	Canada	NORTH MAGNETAWAN RIVER	404	Low
30	4213650	Canada	ASSINIBOINE RIVER	555	Medium
31	4236300	Canada	EAST HUMBER RIVER	271	Low
32	4203335	Canada	DUKE RIVER	1785	High
33	4203910	Canada	TAKHINI RIVER	1296	High
34	4207310	Canada	FRASER RIVER	1119	High
35	4208370	Canada	YELLOWKNIFE RIVER	358	Low
36	4213046	Canada	BOW RIVER	2083	High
37	4213700	Canada	TAYLOR RIVER	240	Low
38	4214295	Canada	DILLON RIVER	591	Medium
39	4214420	Canada	PIPESTONE RIVER (TRIB. HUDSON BAY)	377	Low
40	4214610	Canada	HARRICANAW RIVER	322	Low
41	4214690	Canada	PONTAX RIVER	228	Low
42	4215103	Canada	OKANAGAN RIVER	1084	High
43	4220510	Canada	VERDIGRIS COULEE	1004	High
44	4232750	Canada	WHITE RIVER (TRIB. LAKE SUPERIOR)	406	Low
45	4236050	Canada	TWENTY MILE CREEK	200	Low
46	4244500	Canada	CHURCHILL, FLEUVE (LABRADOR)	513	Medium
47	4244720	Canada	TORRENT RIVER	269	Low
48	4244870	Canada	LITTLE SOUTHWEST MIRAMICHI RIVER	374	Low
49	4244900	Canada	DUNK RIVER	60	Low
50	4244985	Canada	LAHAVE RIVER	174	Low
51	6140250	Czechia	BEROUNKA	496	Low
52	6140700	Czechia	DIVOKA ORLICE	701	Medium
53	6142520	Slovakia	NITRA	426	Low
54	6140450	Czechia	SAZAVA	527	Medium
55	6142680	Slovakia	VAH	1101	High
56	6144100	Slovakia	SAJO	506	Medium

57	6233100	Sweden	VISKAN	150	Low
58	6220500	Belgium	DYLE	113	Low
59	6226700	Spain	VERO	928	Medium
60	6233220	Sweden	TAENNAN (LJUSNAN)	931	Medium
61	6233221	Sweden	LJUSNAN	461	Low
62	6233440	Sweden	NYKOEPINGSAEN	74	Low
63	6233680	Sweden	VINDELAELVEN (UMEAELVEN)	713	Medium
64	6233550	Sweden	KASSJOEAN (LJUNGAN)	349	Low
65	6342520	Germany	ALTMUEHL	482	Low
66	6233870	Sweden	SANGISAELVEN	129	Low
67	6235100	Austria	BREGENZER ACH	1458	High
68	6242200	Austria	ILZBACH	404	Low
69	6243030	Austria	INN	2128	High
70	6246611	Austria	MUR	1263	High
71	6246700	Austria	MUERZ	1094	High
72	6340600	Germany	VEREINIGTE MULDE	436	Low
73	6340700	Germany	SCHWARZE ELSTER	156	Low
74	6340510	Germany	HAVEL	84	Low
75	6401090	Iceland	OELFUSA	498	Low
76	6401500	Iceland	DJUPA, FLJOTSHVERFI	790	Medium
77	6401601	Iceland	SVARTA, SKAGAFIROI	530	Medium
78	6457707	Poland	BOBR	321	Low
79	6457880	Poland	LISWARTA	260	Low
80	6458203	Poland	SKAWA	584	Medium
81	6458713	Poland	WIEPRZ	245	Low
82	6545200	Slovenia	KRKA	354	Low
83	6545190	Slovenia	SAVA	1180	High
84	6604170	United Kingdom	NEVIS	526	Medium
85	6603300	United Kingdom	LOWER BANN	99	Low
86	6605545	United Kingdom	SNAIZEHOLME BECK	448	Low
87	6731175	Norway	EIDSELV	526	Medium
88	6731010	Norway	ENGESETELV	193	Low
89	6731070	Norway	NORDELVA	343	Low
90	6731501	Norway	GAULA	725	Medium
91	6731685	Norway	KOBBELV	597	Medium
92	6934800	Denmark	BREDE A	28	Low
93	6744200	Romania	MAROS	623	Medium
94	6934250	Denmark	GUDENA	74	Low
95	6854210	Finland	AHTAVANJOKI	112	Low
96	6934300	Denmark	UGGERBY A	37	Low
97	6854320	Finland	PYHAJOKI	123	Low
98	6854710	Finland	OUNASJOKI	266	Low
99	6854900	Finland	KYRONJOKI	93	Low
100	6934100	Denmark	SKJERN A	59	Low

101	6939500	Switzerland	RHONE	2192	High
102	6935145	Switzerland	RHINE RIVER	2005	High
103	6935310	Switzerland	REUSS	1250	High
104	6948120	Switzerland	MAGGIA (TRIB. PO)	1519	High
105	6172050	Estonia	PAERNU JOGI	64	Low
106	6172010	Estonia	LEVAJOKI	51	Low
107	6172200	Estonia	EMAJOGI	76	Low

4 RESULTS AND DISCUSSION

4.1 General results and climate averages

Once the dates of peak flooding using the AM method were determined, various statistics were calculated for each catchment, which are summarized in Table 7. The average day of flood is the Julian Day of the year in which the flood occurs on the most. The variability (r) indicates how strong the seasonality is. A value closer to one indicates a strong seasonality, where the floods occur at the same time every year. A value close to zero indicates a weak seasonality, or the day of the peak flood is equally likely throughout the year. The average LWE is the LWE across the entire catchment taken during the first day of the hydrograph, as this can be considered the initial condition of the catchment. The average snow thickness, average soil moisture and average snow cover are also taken on the first day of the hydrograph.

Table 7: General results and climate factor averages for all catchments

	River Name	Average Flood Day	r (Variability)	Snowmelt Affected Floods (%)	Average LWE (cm)	Average Snow Thickness (cm)	Average Soil Moisture (%)	Average Snow Cover (%)
1	AZERGUES	20	0.55	28	0.1	0.3	0.87	26
2	NORTHEAST MARGAREE RIVER	82	0.54	0	2.6	11.4	0.88	90
3	SAINT LOUIS (RUISSEAU)	114	0.60	83	8.3	26.8	0.75	83
4	CHILLIWACK RIVER	355	0.26	88	8.2	33.9	0.87	79
5	KUSKOKWIM RIVER	170	0.56	90	13.6	49.1	0.92	61
6	TANANA RIVER	201	0.88	100	44.5	190.4	0.89	27
7	SHEYENNE RIVER	104	0.75	80	2.2	8.8	0.86	75
8	WILLAMETTE RIVER	16	0.88	100	5.1	22.0	0.93	75
9	MCKENZIE RIVER (TRIB. COLUMBIA)	24	0.69	98	11.6	48.6	0.86	91
10	BRUNEAU RIVER	125	0.80	63	1.4	6.1	0.48	37
11	SPOKANE RIVER	103	0.76	90	2.9	11.2	0.85	51
12	PAHSIMEROI RIVER	99	0.34	3	6.4	27.7	0.60	61
13	BOISE RIVER	134	0.94	100	16.9	46.9	0.89	91
14	CHIPPEWA RIVER (TRIB. MISSISSIPPI)	117	0.52	0	1.6	6.8	0.77	48
15	MUSSELSHELL RIVER	156	0.79	37	0.4	2.5	0.70	24
16	WHITE RIVER (TRIB. OHIO)	68	0.43	49	0.4	2.7	0.83	42
17	HOCKING RIVER	74	0.53	41	0.6	3.4	0.87	38
18	MISSISQUOI RIVER	87	0.47	76	3.4	15.1	0.63	76
19	NEMADJI RIVER	150	0.48	34	1.0	4.6	0.84	31
20	AU SABLE RIVER	102	0.72	74	0.9	4.7	0.65	48
21	GENESEE RIVER	44	0.56	80	1.4	8.2	0.90	74
22	BLACK RIVER (TRIB. LAKE ONTARIO)	88	0.71	85	4.7	20.0	0.77	85

23	PENOBSCOT RIVER	109	0.50	83	5.5	19.6	0.76	70
24	ANDROSCOGGIN RIVER	117	0.60	73	7.0	24.9	0.73	71
25	WESTFIELD RIVER	92	0.23	51	0.7	3.8	0.62	51
26	SADDLE RIVER	18	0.34	41	0.2	0.9	0.78	39
27	SUSQUEHANNA RIVER	49	0.44	66	1.7	8.9	0.88	63
28	IRONSPRING CREEK	106	0.92	83	1.3	5.7	0.84	81
29	NORTH MAGNETAWAN RIVER	96	0.88	98	5.3	19.9	0.74	90
30	ASSINIBOINE RIVER	122	0.74	71	0.5	2.1	0.82	56
31	EAST HUMBER RIVER	72	0.82	85	1.0	5.1	0.78	85
32	DUKE RIVER	194	0.96	100	147.4	811.4	0.72	58
33	TAKHINI RIVER	194	0.97	38	0.1	0.4	0.70	22
34	FRASER RIVER	160	0.96	97	2.8	7.1	0.82	31
35	YELLOWKNIFE RIVER	250	0.25	54	1.7	11.2	0.61	40
36	BOW RIVER	163	0.97	100	7.7	23.0	0.79	84
37	TAYLOR RIVER	165	0.73	34	0.3	1.5	0.86	32
38	DILLON RIVER	174	0.82	22	0.1	0.3	0.87	16
39	PIPESTONE RIVER (TRIB. HUDSON BAY)	157	0.72	60	0.6	2.6	0.73	54
40	HARRICANAW RIVER	125	0.87	54	0.8	2.8	0.69	47
41	PONTAX RIVER	126	0.85	93	7.8	25.2	0.92	92
42	OKANAGAN RIVER	174	0.72	41	1.9	5.4	0.68	19
43	VERDIGRIS COULEE	142	0.62	59	0.0	0.1	0.66	21
44	WHITE RIVER (TRIB. LAKE SUPERIOR)	132	0.87	78	0.7	2.3	0.72	70
45	TWENTY MILE CREEK	65	0.71	76	1.0	6.4	0.88	73
46	CHURCHILL, FLEUVE (LABRADOR)	141	0.93	100	10.8	33.3	0.83	92
47	TORRENT RIVER	134	0.84	83	6.7	21.2	0.94	81
48	LITTLE SOUTHWEST MIRAMICHI RIVER	112	0.71	93	7.1	24.2	0.88	90
49	DUNK RIVER	72	0.82	93	0.4	3.8	0.88	93
50	LAHAVE RIVER	62	0.54	78	1.7	8.0	0.90	76
51	BEROUNKA	80	0.42	60	1.2	5.9	0.87	54
52	DIVOKA ORLICE	40	0.54	76	2.2	10.5	0.85	74
53	NITRA	79	0.64	59	0.8	4.0	0.85	51
54	SAZAVA	66	0.43	68	2.3	10.1	0.87	62
55	VAH	150	0.51	36	1.6	5.7	0.85	32
56	SAJO	97	0.43	59	0.9	4.4	0.87	43
57	VISKAN	8	0.75	68	0.7	4.0	0.87	57

58	DYLE	347	0.37	19	0.1	0.5	0.82	16
59	VERO	355	0.49	0	0.0	0.2	0.74	11
60	TAENNAN (LJUSNAN)	142	0.99	100	10.3	29.6	0.92	97
61	LJUSNAN	166	0.57	60	2.1	6.5	0.78	36
62	NYKOEPINGSAEN	62	0.59	51	0.6	3.8	0.69	44
63	VINDELAELVEN (UMEAELVEN)	156	0.98	95	4.4	12.4	0.82	65
64	KASSJOEAN (LJUNGAN)	131	0.72	69	1.6	5.2	0.76	67
65	ALTMUEHL	47	0.64	71	0.6	3.0	0.95	64
66	SANGISAELVEN	136	0.76	76	4.0	12.3	0.70	54
67	BREGENZER ACH	196	0.58	49	2.2	10.1	0.87	49
68	ILZBACH	173	0.09	33	0.4	2.1	0.81	26
69	INN	176	0.92	79	3.1	10.3	0.86	71
70	MUR	190	0.67	43	1.8	6.1	0.85	21
71	MUERZ	160	0.39	41	1.5	5.8	0.89	40
72	VEREINIGTE MULDE	46	0.38	76	2.1	9.8	0.91	62
73	SCHWARZE ELSTER	42	0.53	71	0.7	3.7	0.91	61
74	HAVEL	57	0.71	54	0.2	1.0	0.90	44
75	OELFUSA	35	0.45	100	131.5	470.3	0.77	95
76	DJUPA, FLJOTSHVERFI	277	0.47	100	588.6	2497.8	0.72	100
77	SVARTA, SKAGAFIROI	126	0.77	95	18.3	56.5	0.72	93
78	BOBR	76	0.34	63	1.8	8.8	0.88	52
79	LISWARTA	106	0.48	41	1.1	4.8	0.70	41
80	SKAWA	158	0.46	32	0.8	4.5	0.87	28
81	WIEPRZ	90	0.51	56	0.8	3.7	0.92	52
82	KRKA	335	0.30	42	0.7	3.6	0.87	41
83	SAVA	299	0.73	29	0.9	4.6	0.83	28
84	NEVIS	8	0.59	66	0.7	4.2	0.86	69
85	LOWER BANN	6	0.64	44	0.0	0.1	0.88	21
86	SNAIZEHOLME BECK	11	0.72	59	0.4	2.3	0.92	59
87	EIDSELV	311	0.27	90	11.2	42.9	0.77	75
88	ENGESETELV	355	0.39	80	3.7	18.3	0.76	73
89	NORDELVA	22	0.40	90	2.8	13.7	0.88	88
90	GAULA	154	0.83	82	2.4	8.7	0.84	69
91	KOBBELV	249	0.21	78	8.9	32.5	0.84	83
92	BREDE A	8	0.72	45	0.3	1.9	0.78	45
93	MAROS	102	0.70	80	0.9	4.7	0.91	43
94	GUDENA	17	0.78	62	0.1	0.7	0.88	52
95	AHTAVANJOKI	101	0.44	68	1.9	7.2	0.90	64
96	UGGERBY A	13	0.62	46	0.4	2.3	0.83	46
97	PYHAJOKI	115	0.92	93	6.2	21.8	0.83	79
98	OUNASJOKI	135	0.99	100	13.7	37.5	0.94	100
99	KYRONJOKI	107	0.37	74	4.3	14.3	0.93	73
100	SKJERN A	12	0.72	41	0.6	3.0	0.80	41
101	RHONE	185	0.86	65	4.2	14.9	0.82	54

102	RHINE RIVER	202	0.60	78	3.1	12.0	0.83	56
103	REUSS	182	0.71	85	2.7	11.6	0.91	49
104	MAGGIA (TRIB. PO)	251	0.57	68	0.9	4.7	0.68	43
105	PAERNU JOGI	66	0.72	90	2.6	10.7	0.94	88
106	LEVAJOKI	64	0.54	88	2.2	9.1	0.93	88
107	EMAJOGI	88	0.70	78	1.0	3.9	0.91	72

Figure 18 to Figure 23 display the average LWE, snow thickness, and soil moisture values across all the catchments, separated by North American catchments and European catchments. A higher LWE, snow thickness, and soil moisture content are indicated by a larger circle on the plots. In North America, areas with a large amount of snow accumulation are in the northwest (in and near Alaska), in the Rocky Mountains of Western North America and northeastern North America (however, not directly on the Atlantic Coast). The largest snow accumulation, in both LWE and snow thickness, occur between the Canadian and USA border near Alaska. In Europe, the highest snow accumulation, in both snow thickness and LWE, occurs in Iceland, Scandinavia and to a lesser extent, in the Alps in Central Europe. The highest snow accumulation in Europe occurs in Iceland.

The areas of the highest initial soil moisture content at the start of the flood appear in the areas with the lower snow accumulation, although visually this is less distinct. In North America, the Central Prairies generally have a higher soil moisture condition than the mountainous areas and eastern areas. The Atlantic Coast also has a higher soil moisture condition, which happens to coincide with the lesser amount of snow accumulation in this region. The same phenomenon occurs in Europe. Here, Iceland, Scandinavia and the Alps have a lower soil moisture content than other areas such as the UK, Central Europe, and the Baltics.

This phenomenon could be due to the nature of the snowmelt, frozen soils, and soil moisture. Ho and Valeo (2015) studied the impacts of frozen ground on infiltration and runoff in Canada. They demonstrated that the soil moisture content throughout the winter stays constant. They reported that if there is a high moisture condition in the soil when the soil freezes, there will be a high soil moisture condition when the soil melts in the Spring. This is due to the inability of movement from the frozen water (Ho and Valeo, 2005). They found that when soil freezes, the infiltration capacity is decreased, and impermeability is increased. However, this is dependent on the soil type and type of surface coverage. In Alberta, Canada there were results that showed that in situations with a low initial soil moisture content, the frozen soil acted as a nearly impervious surface, despite the low soil moisture content. This test site had densely packed soil. Whereas another test in Ontario, Canada showed little difference in the amount of infiltration between the frozen and unfrozen soils (Ho and Valeo, 2005). The study also found that the ground often remains frozen during the snow coverage due to the insulating properties of snow.

Harpold et al. (2015) investigated the relationship between the timing of the snowpacks and the peak soil moisture condition. They found that the peak soil moisture condition that occurs is often in the final stages of the snowpack melt. At nearly all the sites in their study, located in the Northern United States, the peak soil moisture occurred within 5 days of the final melting of the snow (Harpold et al., 2015).

The changes in the amount of infiltration caused by frozen soils cannot yield any solid findings in this study due to the absence of soil temperature and soil type data. However, it can be assumed that because of the excess amount of snowmelt that occurs from the larger accumulation of snow, the regions with large amounts of snow require a lower initial soil moisture content. The lack of moisture in the soil can be made up for by the initial stages of the melting snow. Whereas in the catchments where there is little snow accumulation there must be a high antecedent moisture condition to create the yearly maximum flood. In the catchments where the ground freezes this antecedent moisture condition is likely determined in the fall (Ho and Valeo, 2005). Flood managing authorities can use this knowledge to prepare in advance. Soil moisture conditions in the fall, soil temperature, and snowpack accumulation data can be monitored and models made to predict the amount of runoff that will be created.

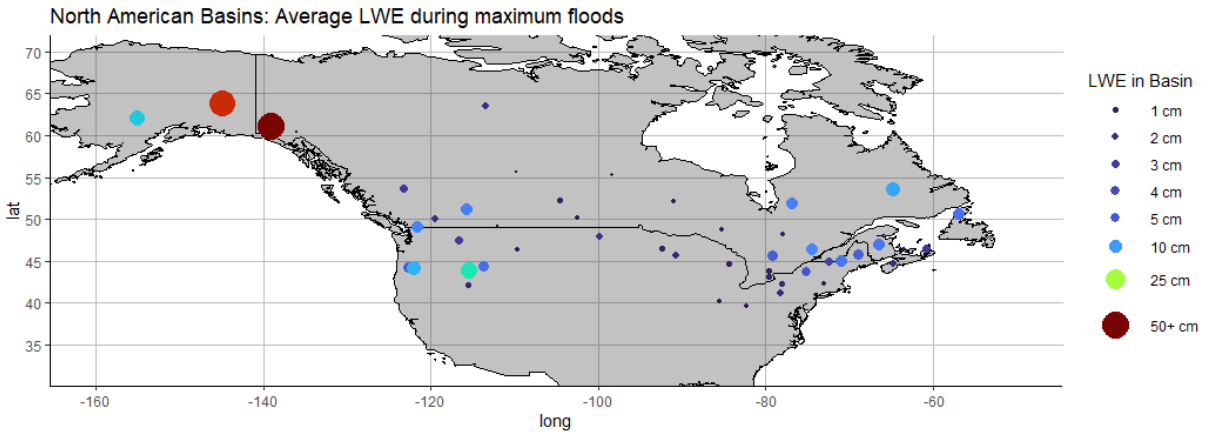

Figure 18: Average LWE during annual maximum floods in the period 1979-2019 in North American catchments

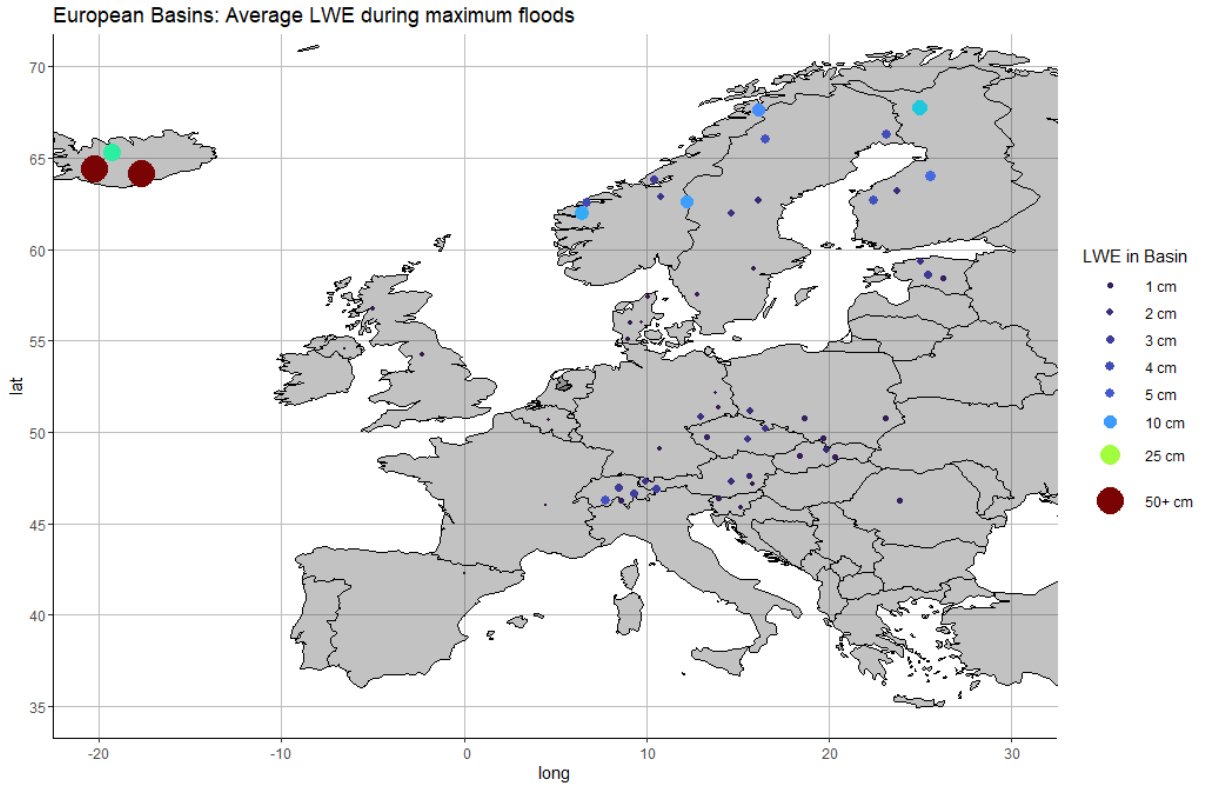

Figure 19: Average LWE during annual maximum floods in the period 1979-2019 in European catchments

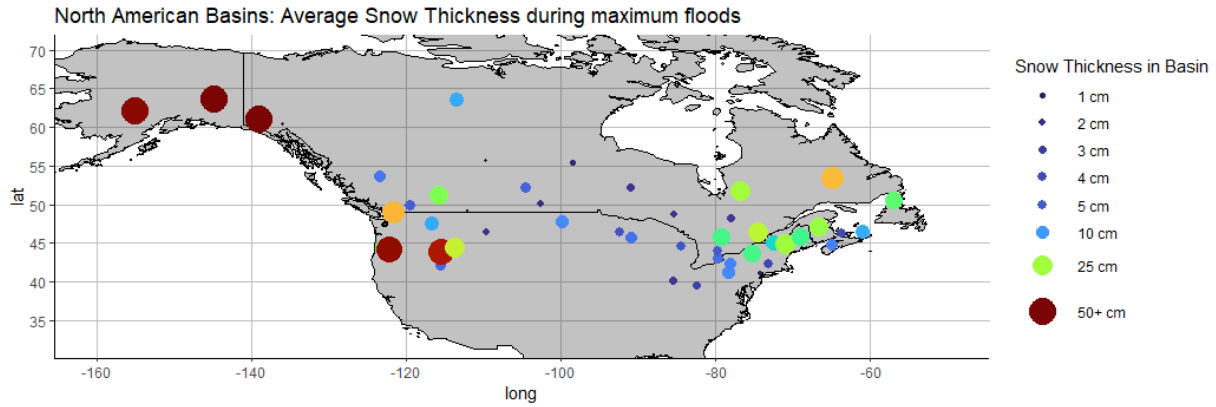

Figure 20: Average snow thickness during annual maximum floods in the period 1979-2019 in North American catchments

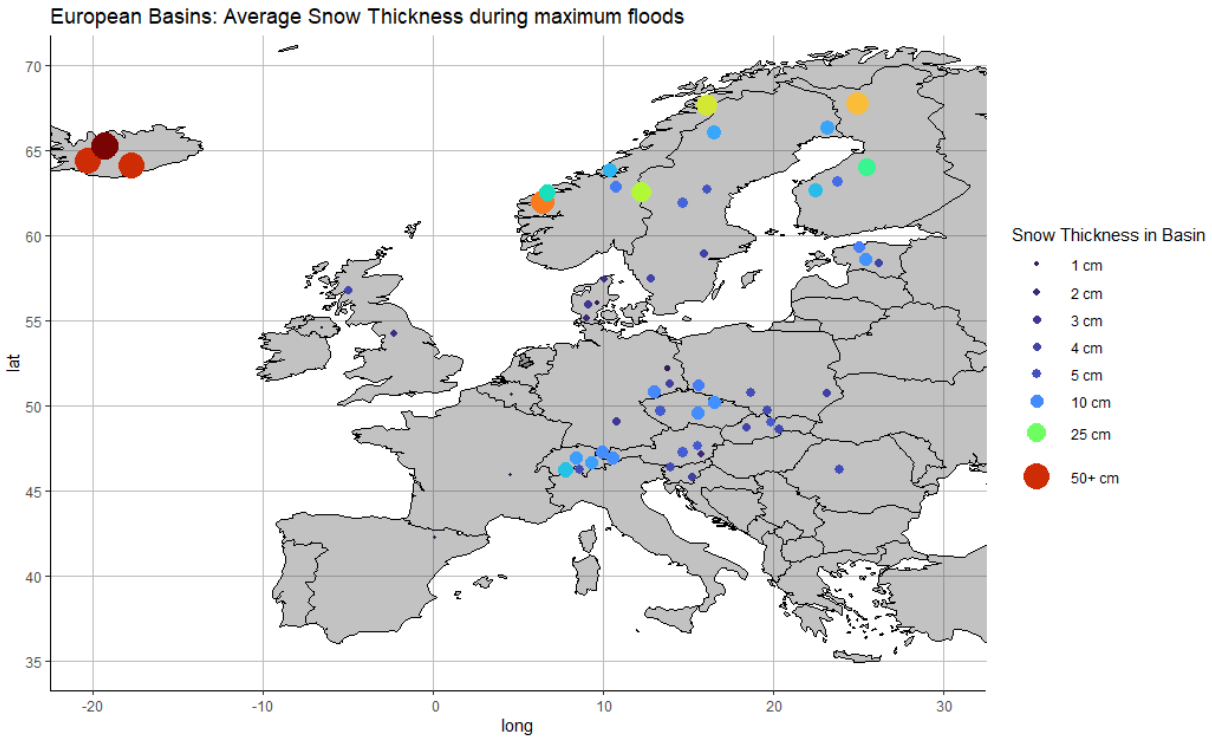

Figure 21: Average snow thickness during annual maximum floods in the period 1979-2019 in European catchments

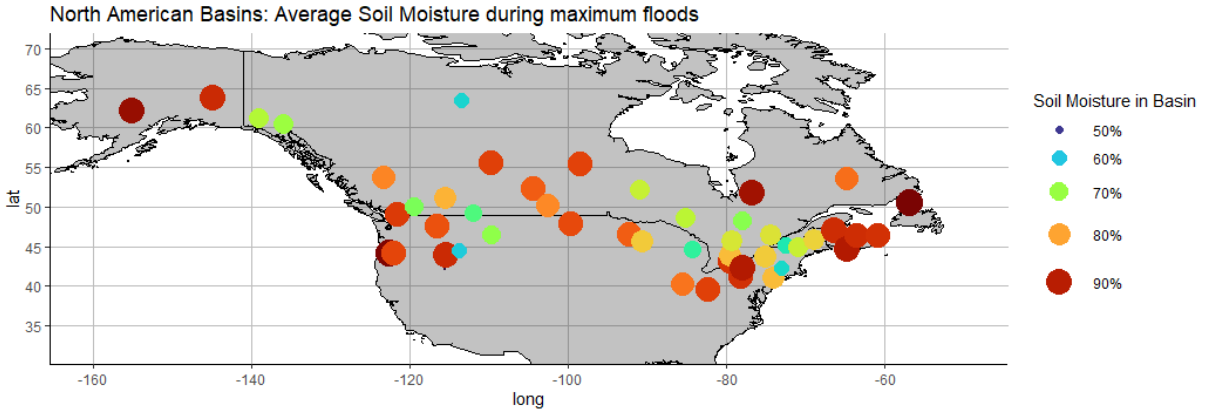

Figure 22: Average soil moisture during annual maximum floods in the period 1979-2019 in North American catchments

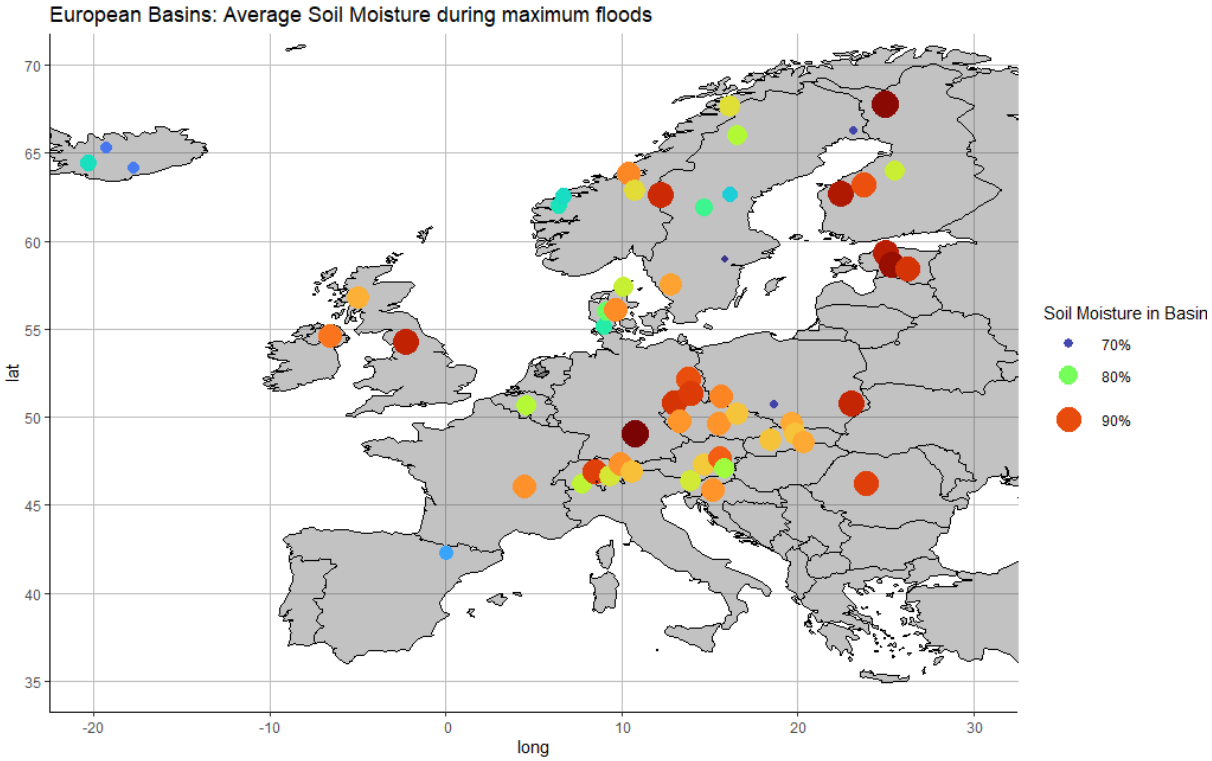

Figure 23: Average soil moisture during annual maximum floods in the period 1979-2019 in European catchments

4.2 Flood type separation

Every AM flood was categorized into the eight different flood typologies described in Section 3.2.1. If the floods did not belong to one of the eight categories it was assigned to an "Other" category. To present the distribution of the flood types, heat maps were created for the North American and European catchments in Figure 24 and Figure 25, respectively. A heatmap is a way to visually represent the magnitude of each flood typologies against each other. These heatmaps are presented as a percentage of the total floods in the catchment. The dark blue colour represents a low magnitude, whereas the red represents a high magnitude. The rows are sorted from the most southern catchment in the bottom row to the most northern catchment in the top row.

There are a few conclusions which can be drawn from the heatmaps. The first is that the majority of floods that occur in the selected catchments are snowmelt driven floods. The snowmelt driven floods (SMF-W and SMF-D) represent 39% of the total considered floods in North America and Europe. The wet condition is the dominant condition when looking at the snowmelt driven floods.

Another conclusion is the distinctions between the dry and wet initial conditions. In every main category, the rain-on-snow (ROS), snowmelt floods (SMF), short precipitation floods (SPF), and long precipitation floods (LPF), the wet conditions (-W) are always more prevalent than the dry conditions (-D). The dry initial conditions make up only 45%, 36%, 20%, and 19% of the ROS, SMF, SPF, and LPF, respectively. However, the floods which are affected by snow, the ROS and SMF floods have a higher percentage of dry initial conditions. The dry initial conditions for the snow affected floods is based on the snow density. Thus, it appears that the snow density is less important when determining the type of flood than the soil moisture. The ROS floods had 45% of the floods in the dry initial

condition, an almost equal amount between wet and dry. This is due to the fact that the rain quickly increases the snow density prior to melting (Harpold et al., 2015). Thus, the snow density for the ROS floods is a less important factor. When investigating the rain precipitation driven floods (LPF-W, LPF-D, SPF-W, SPF-D) the initial soil moisture condition is a key characteristic. There is less than 20% of the precipitation floods in total where the maximum annual flood started with less than 75% saturation. Based on this typology, it is reasonable to assume that the soil moisture content plays a key role in the scale of the flood – in addition to the flood typology.

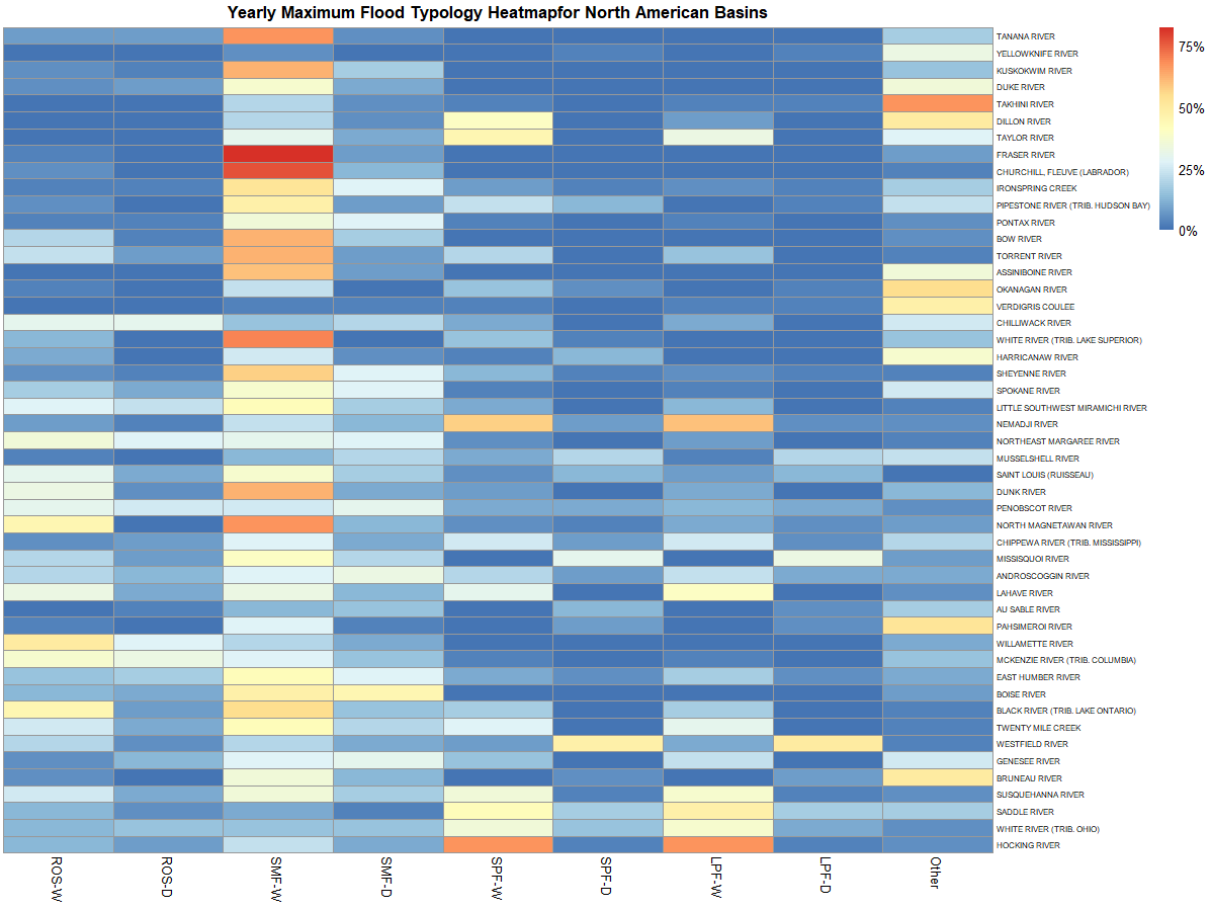

Figure 24: Heatmap for North American distrubution of flood typologies

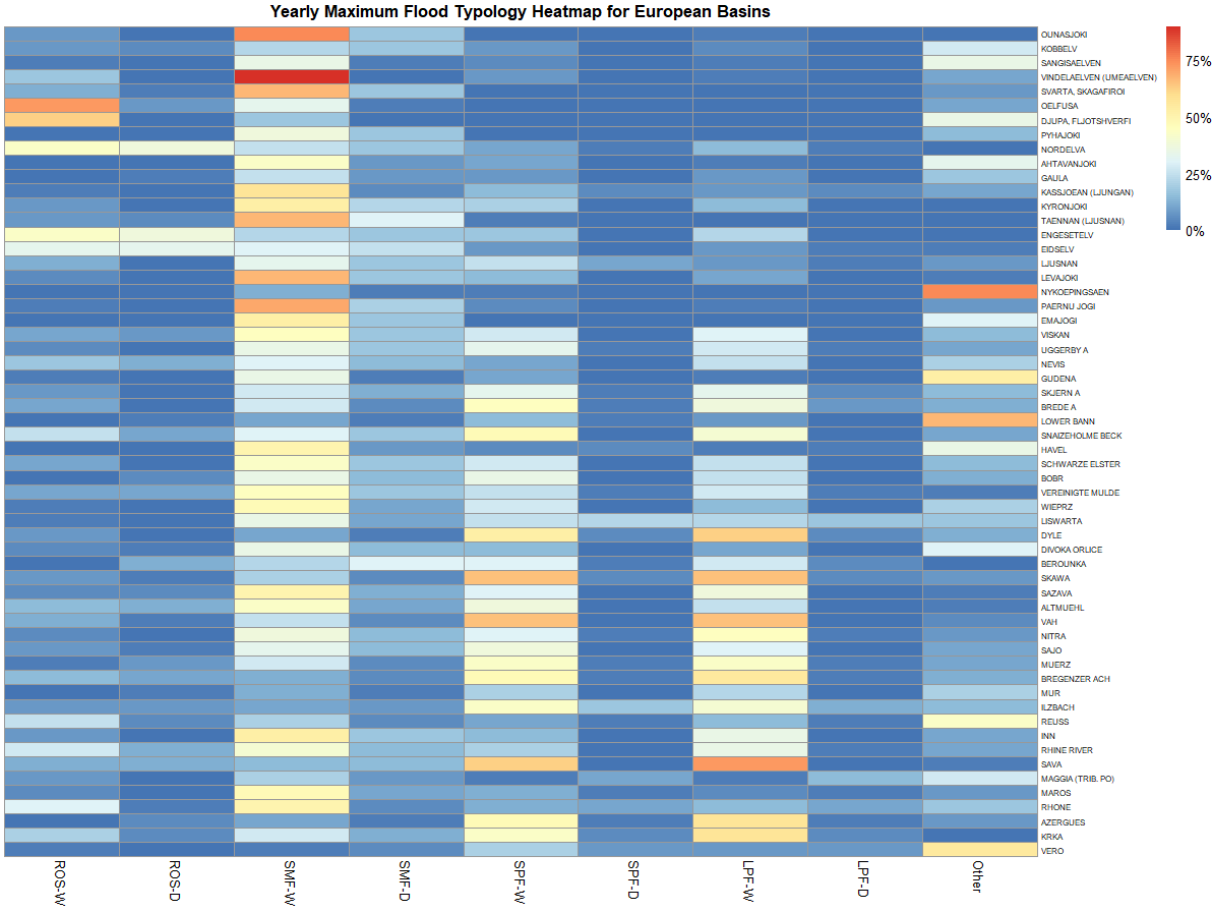

Figure 25: Heatmap for European distribution of flood typologies

The flood types were then summarized based on the climate zone of each catchment. The mean distribution of flood typologies was taken for each climate zone. It should be noted that only 4 of the climate zones, temperate oceanic, warm summer humid continental, subartic and tundra (CFB, DFA, DFC, and ET, respectively) have substanstial sample sizes. When inspecting Figure 26 the temperate oceanic and warm summer humid conintental climate zones (CFB and DFA) are more evenly distributed thoughout the different typologies, with the short and long precipitation floods being nearly equal to the snowmelt floods. The subarctic climate zone has the highest magnitude of floods in the SMF-W category. This is logical as this climate zone experiences lower winter temperatures, keeping the ground frozen and snow is able to accumulate. The tundra climate zone has the two dominant types being the ROS-W and SMF-W type floods. Tundra is the only climate zone of the four which has ROS-W as a dominant typology. This is likely due to the large amounts of rainfall found in many mountainous regions were the Tundra climate zones are located in (Sikorska et al., 2015).

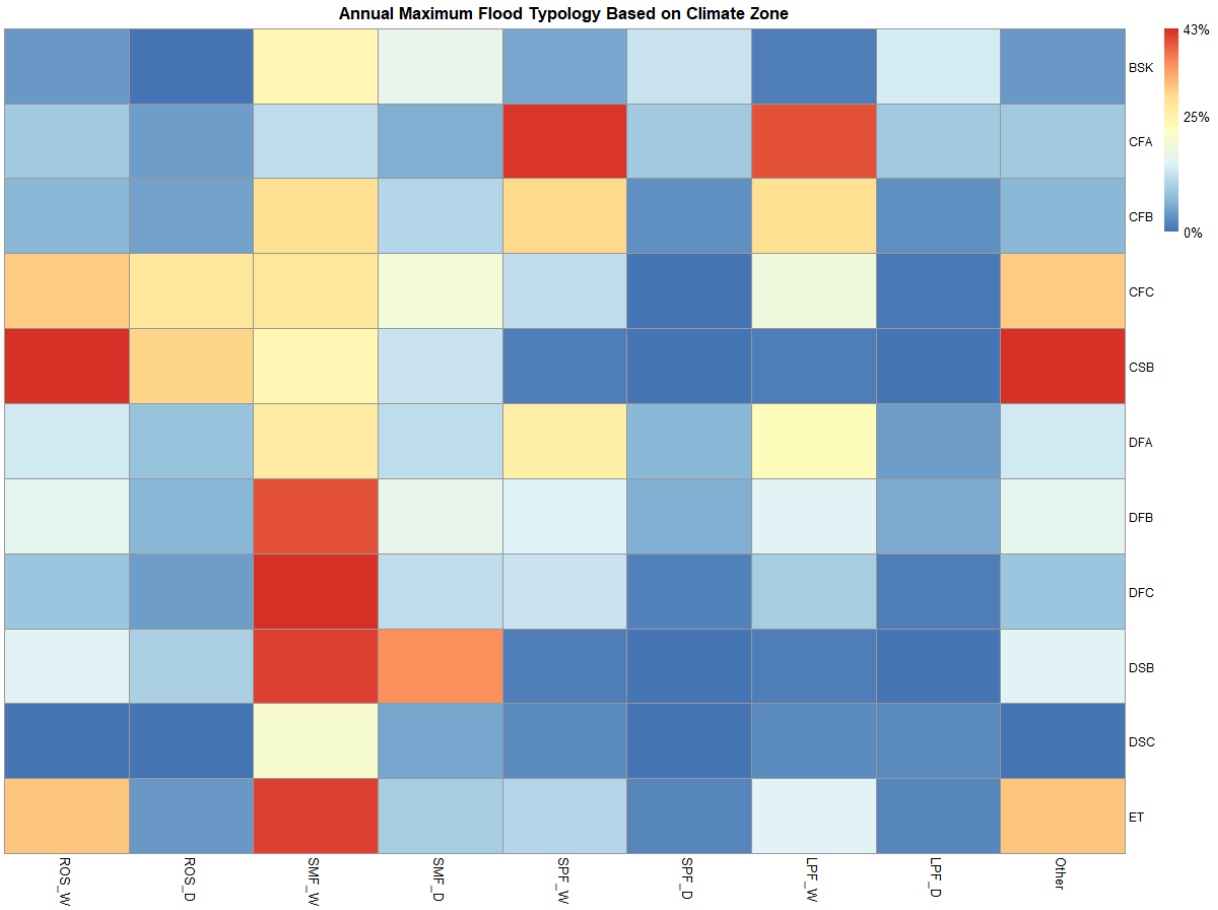

Figure 26: Heatmap for flood typologies by climate zone

Figure 27 displays the flood typologies based on the areas of the catchments. The large catchments are dominated by SMF floods, whereas the precipitation floods are more dominant in the medium and small catchments. This could be due to several reasons, one of which is the thresholds chosen for the sorting of typologies. The snow coverage threshold for the flood to be considered affected by snow is 5%, which was also used by Sikorska et al. (2015). In a large catchment, there is likely an area of the catchment which is covered by snow, thus it crosses the threshold as a snowmelt affected flood. Small catchments however may be completely without snow coverage, and thus do not cross the threshold. Another possible reason for this includes the soil-moisture concept discussed by Harpold et al. (2015). They state that peak soil moisture occurs within 5 days of the snowpack completely melting (Harpold et al., 2015). A high soil moisture content is required for floods to occur. For large catchments, a rainfall storm would have to be very widespread in order for the entire catchment to reach the soil moisture content required for an annual maxima flood to occur. However, if there is a melting snowpack across the entire catchment, this would allow for the entire catchment to reach the high soil moisture content required. However, one unexpected result is that long precipitation floods have a low count for large catchments. This is the opposite of the expected, as long precipitation storms have the capability to saturate the large catchments.

The medium and small floods are dominated by the short and long precipitation floods. This is likely because smaller sized storms are required for these catchments to become saturated. High intensity storms tend to be smaller in size, which is why short precipitation floods are more frequent in small and

medium catchments (Merz and Blöschl, 2003). Additionally, the time of concentration of small and medium catchments is less (Merz and Blöschl, 2003). Thus, long and short precipitation floods can saturate the entire small and medium catchments, resulting in the peak of the hydrograph being higher.

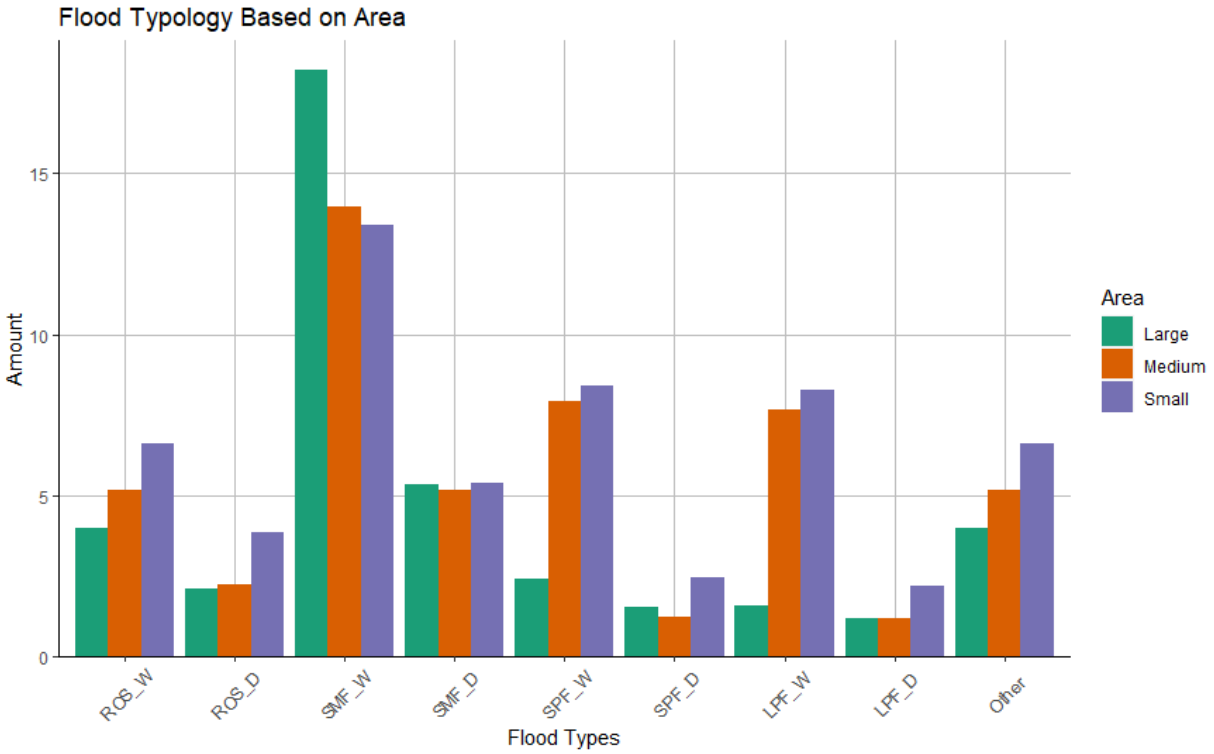

Figure 27: Flood typologies based on catchment area

Figure 28 displays the amount of floods in each typology based on the mean elevation of the catchments. One initially unexpected conclusion is that the high elevation catchments have a lower number of SMF and ROS floods than the medium and low elevation catchments. It was initially expected that the higher elevation catchments would have a higher number of SMF and ROS floods, as generally higher elevations receive more snowfall than lower elevation catchments. However, Sikorska et al. (2015) had similar findings in their study. They used two different methods to allocate the floods into different typologies. One was using a "crisp" method, which assigned one typology to each flood. The other was a "fuzzy" method, in which each flood could have aspects of multiple different typologies. When investigating high elevation catchments they found that the precipitation floods are the dominant typology when using the crisp method (Sikorska et al., 2015), similar to the findings in this study. However, when switching to the fuzzy method, they found that many of the floods, while the dominant typology remains precipitation driven, are partially SMF or ROS floods as well (Sikorska et al., 2015). However, the snow related aspects do not cross the thresholds for the flood to be classified as a snowmelt affected flood. They hypothesize that this is because of the large amounts of rainfall that falls in mountainous regions. Mountainous regions are also prone to flash floods (Sikorska et al., 2015), which are not a flood typology in this assessment. A study also completed by Berghuis et al. (2019b) investigated the drivers of European floods with three main categories: extreme precipitation, snowmelt and soil moisture excess. They found that the extreme precipitation was only the predominant driver in mountainous regions such as the alps (Berghuijs et al., 2019b). Despite there being large amounts of

snow in the mountainous regions, the rainfall that falls can create the dominant flood type being precipitation driven rather than snowmelt driven floods.

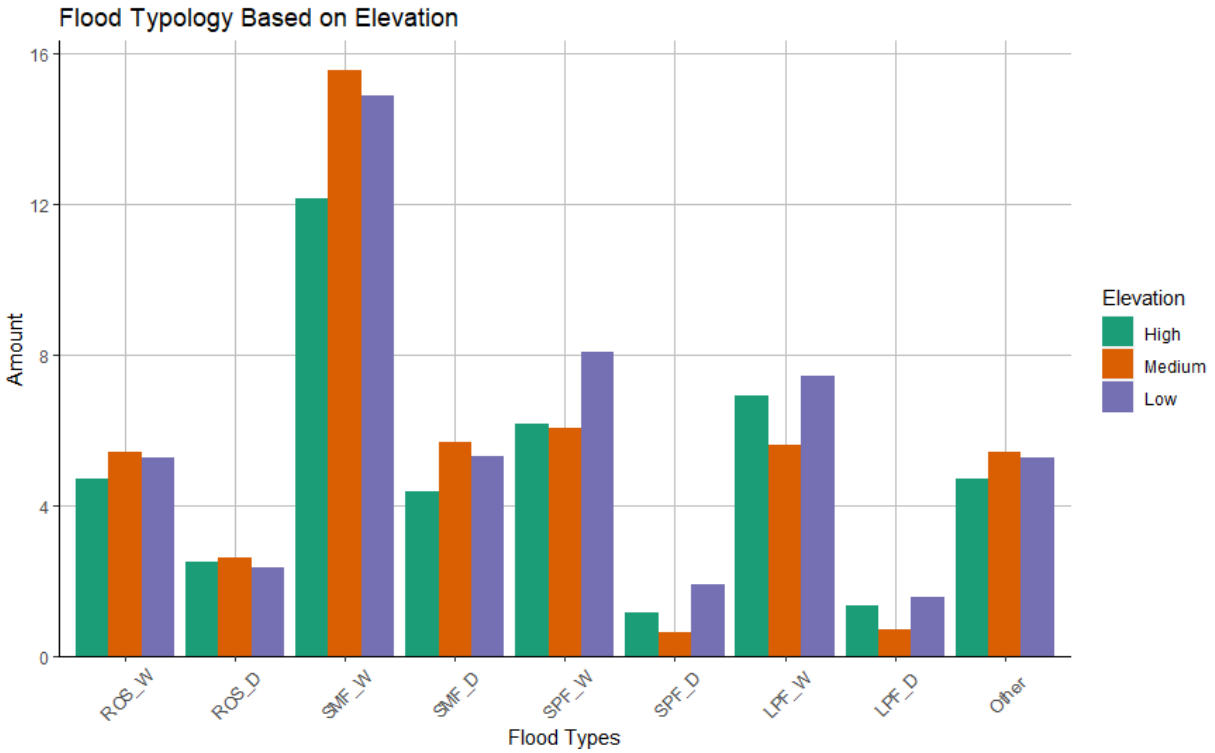

Figure 28: Flood typologies based on catchment elevation

Figure 29 and Figure 30 display the heatmaps for the maximum floods when snow was present in the catchment, not the annual maximum floods. The ROS and SMF floods increase greatly and the LPF and SPF floods decrease. There is a larger increase in the SMF floods, rather than the ROS floods. This can be seen in Figure 31. SMF floods increase by 29% whereas ROS floods increase by 17%. The total amount of SMF floods far exceeds the total amount of ROS floods as well when looking at the snowmelt affected floods. Therefore, it can be concluded that SMF floods are the dominant flood type in when snow is affecting the floods. SMF floods are typically caused by rapid increased in temperature when there is a large snowpack (Merz and Blöschl, 2003).

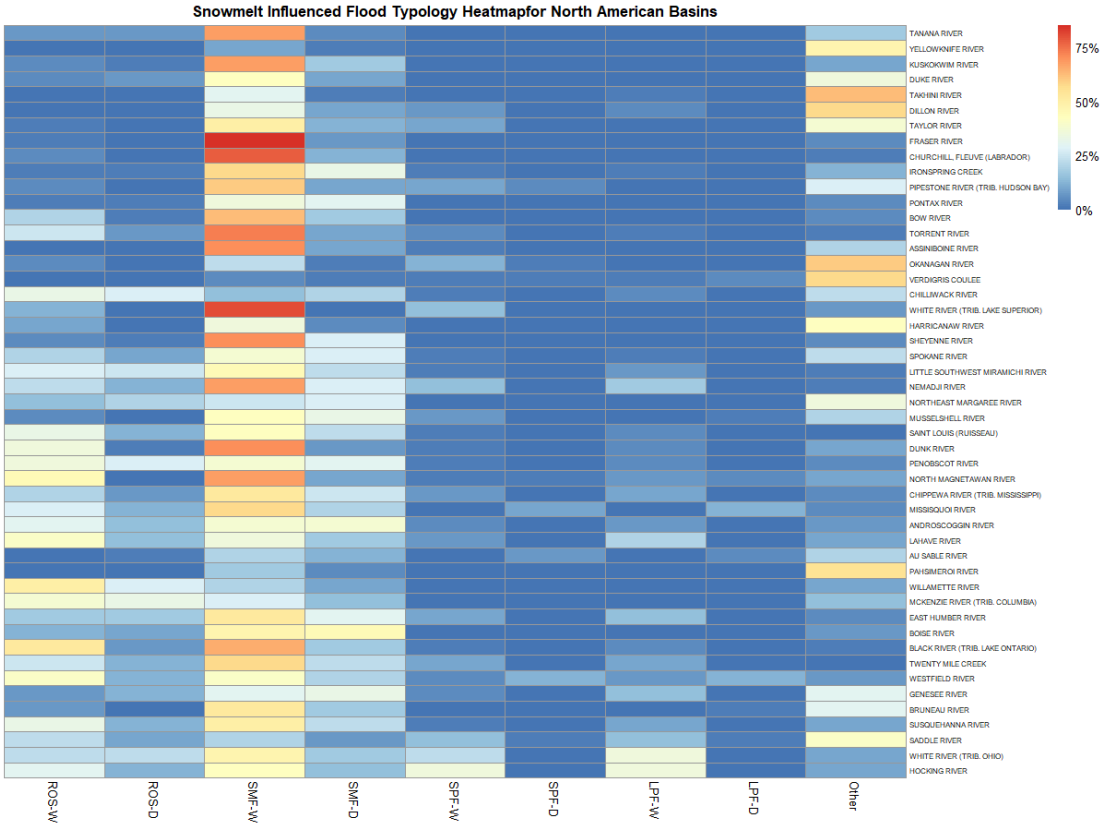

Figure 29: Heatmap for North American distribution of snowmelt affected flood typologies

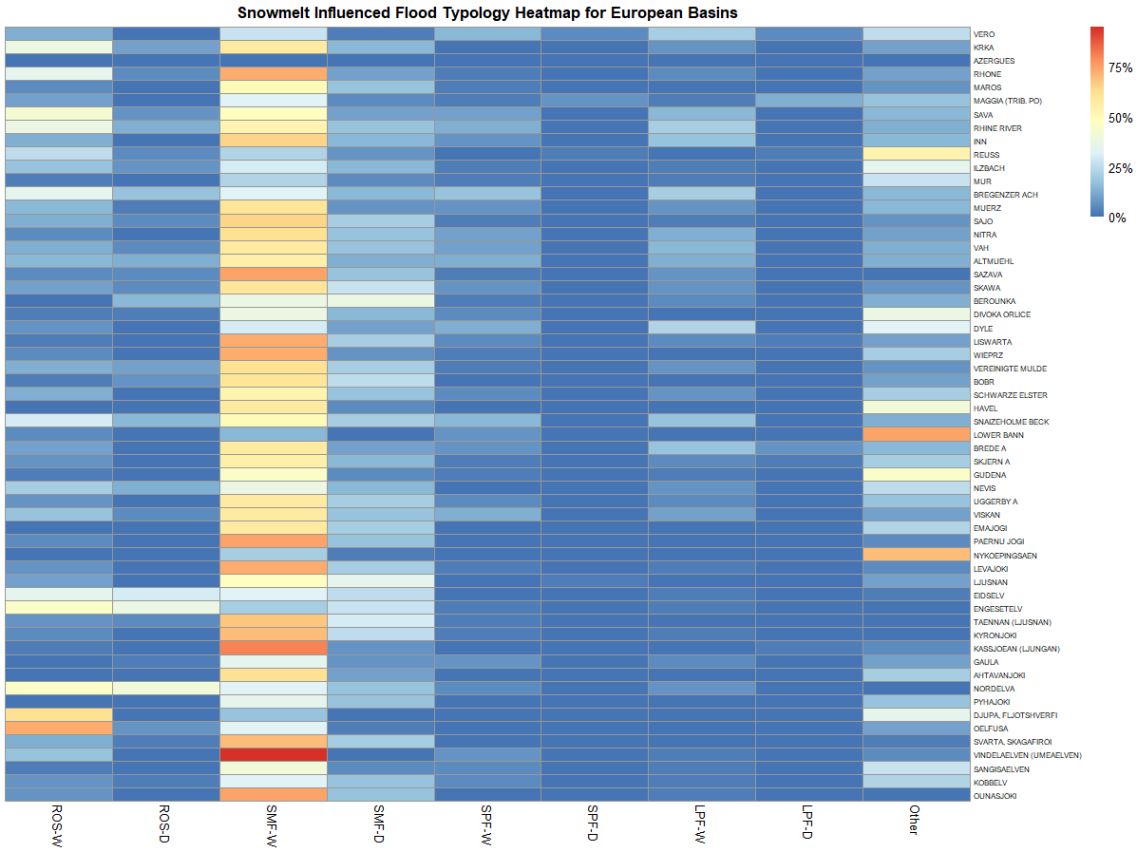

Figure 30: Heatmap for European distribution of snowmelt affected flood typologies

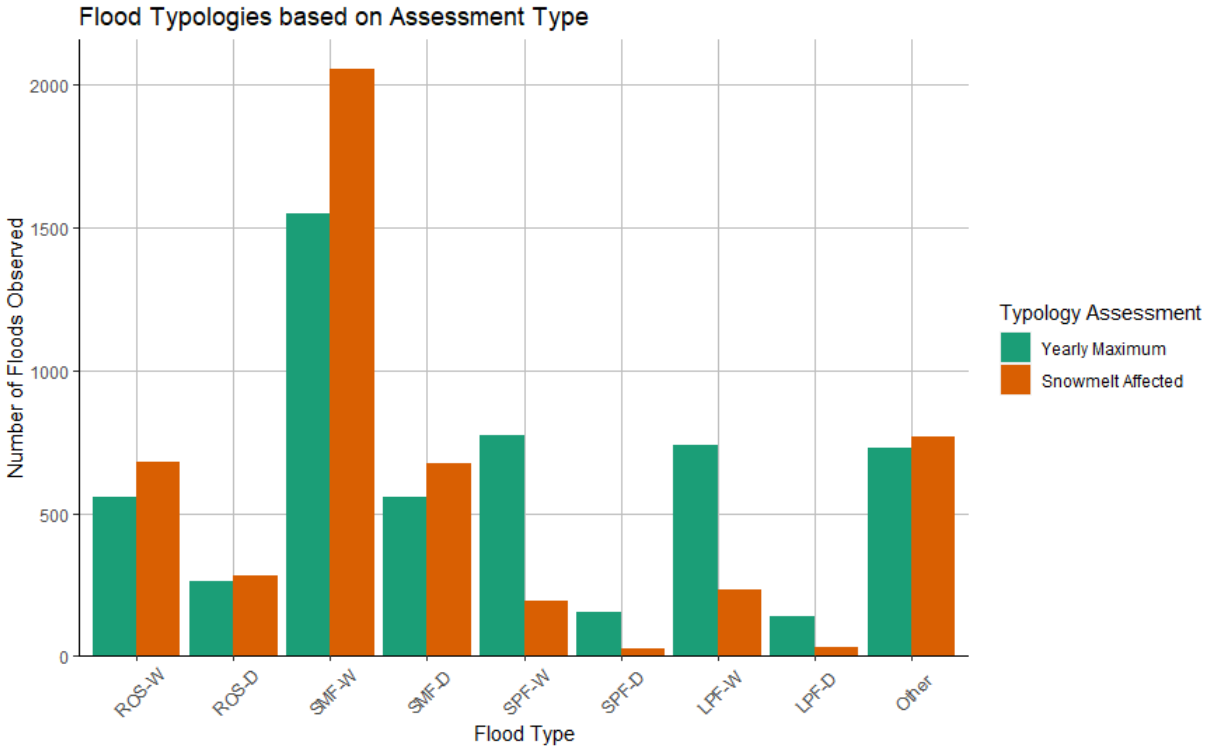

Figure 31: Flood typologies comparing the yearround assessment to the maximum snowmelt affected flood

4.3 Seasonality

4.3.1 Snowmelt affected floods

Figure 32 and Figure 33 display the percentage of floods which are impacted by snowmelt for the considered North American and European catchments, respectively. The percentage of floods that are affected from snow are geographically similar to the areas with large amounts of snow accumulation. Central North America has a lower percentage of snowmelt affected floods, which corresponds to this area having a lower snow accumulation. The areas with the highest accumulation, namely Alaska, the Rocky Mountains, and Eastern Canada all experience a high percentage of snowmelt affected floods. The same phenomenon occurs in Europe, where the largest snow accumulation coincides with the largest percentage of snowmelt affected floods. Scandinavia and Iceland all have large percentages of snowmelt floods. However, despite a large snow accumulation, the Alps do not have as high of a percentage of snowmelt affected floods as other areas. This coincides with the findings from the typologies sorting where the high elevation catchments have a higher amounts of precipitation driven floods. Despite large amounts of snow, the high elevations basins are not as affected by snowmelt influenced floods.

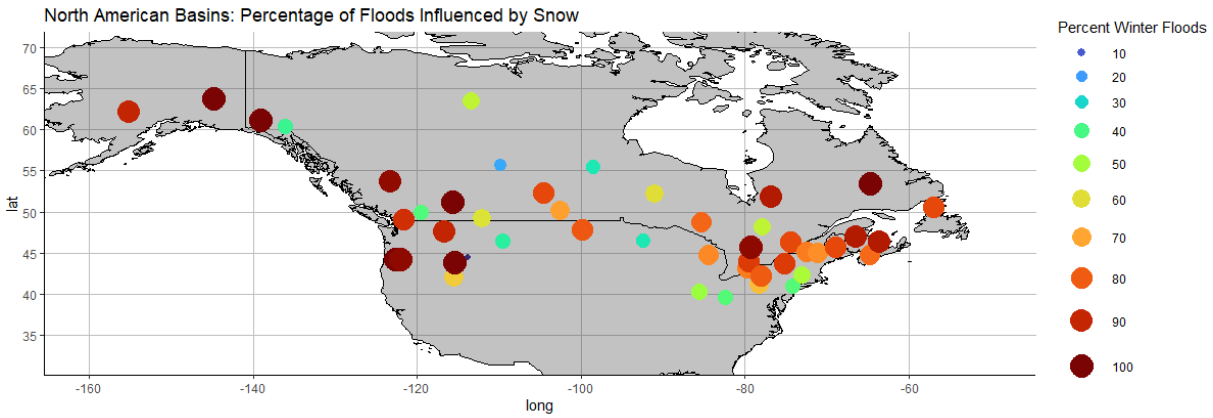

Figure 32: Percentage of floods in each catchment that was affected by snowmelt for North American catchments

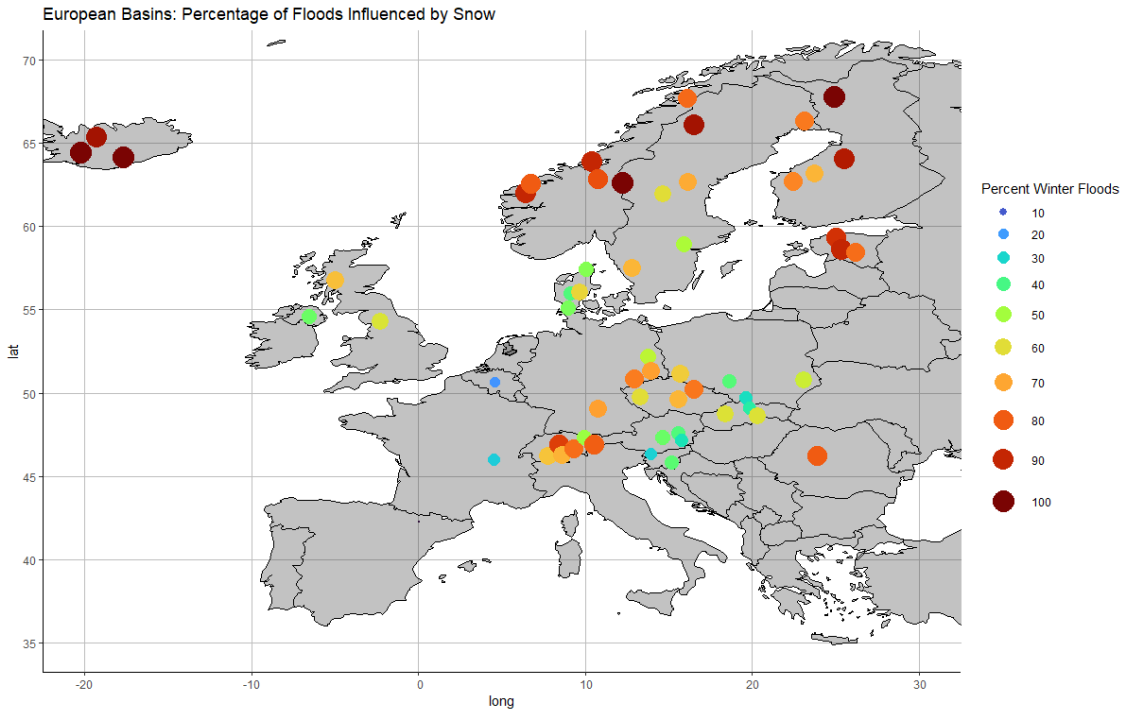

Figure 33: Percentage of floods in each catchment that was affected by snowmelt for European catchments

Figure 34 displays the distribution of snowmelt affected floods by climate zone in percentages. The results remain consistent with the earlier findings. The temperate oceanic climate, which has the lowest accumulation of snow, has the lowest percentage of snow related flood types. The highest concentration of percentages occurs around 50%, the lowest of all the four major climate zones. The warm summer humid continental climate zone has the largest range of percentages; however, the largest concentration and the mean are at around 80% snowmelt affected floods. The subarctic climate zone has a large dispersion of snowmelt affected floods, ranging from 25% to 100%, with the largest concentration at 90% and the mean at 80%. The tundra climate zone has the highest percentage of snowmelt affected floods, with the largest concentration and the mean above 90%. The tundra climate has the coldest temperatures, which allows for a large accumulation of snow.

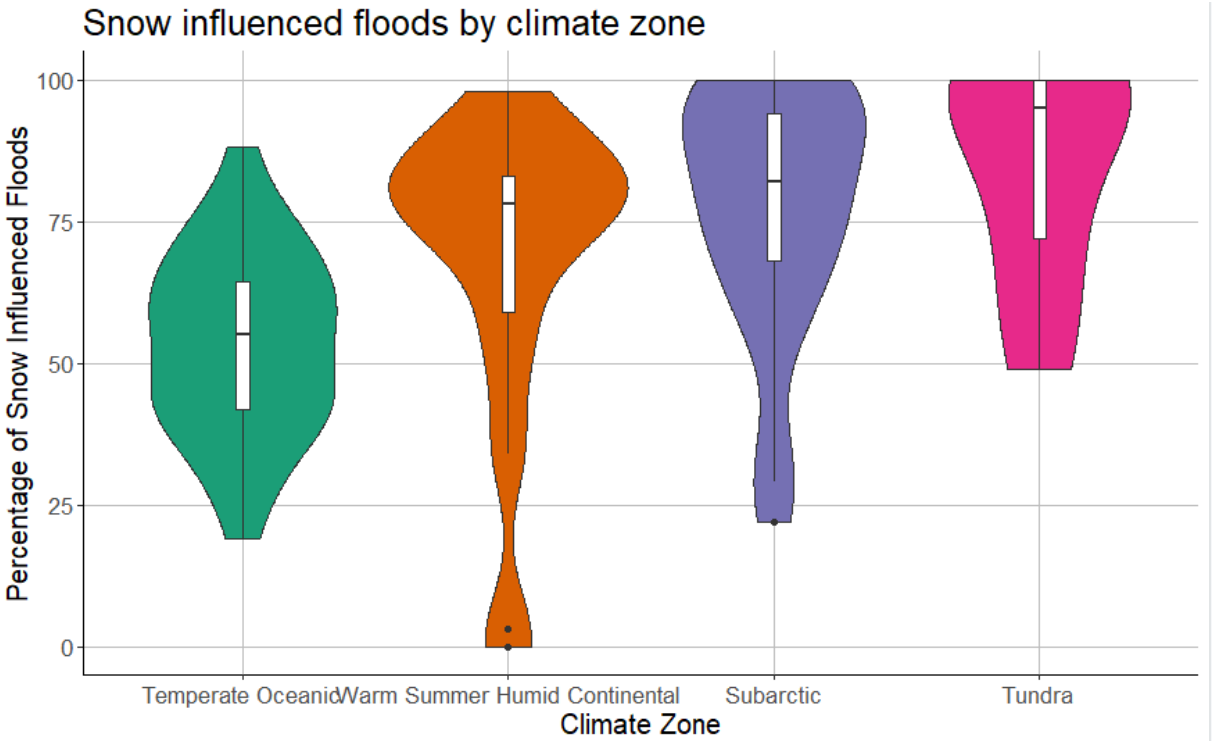

Figure 34: Percentage of snowmelt affected floods by climate zone

Figure 35 and Figure 36 display the percentages of snowmelt affected floods by area and elevation, respectively. The large catchments have the largest amount of snowmelt affected floods, with the mean at 80% and the highest concentration at 90%. The medium and small catchments have averages at around 70%. The catchments with high elevations have the lowest mean of less than 70%. The distribution of the high elevation catchments remains relatively constant from 30% all the way to 100%. It should also be noted that in all categories, there is little concentration at less than 30% snowmelt affected floods. Therefore, in the northern European and northern North American catchments, most catchments have at least 30% of their annual floods occurring where snow is present, proving the importance of snow in the water cycle and flood patterns of these catchments.

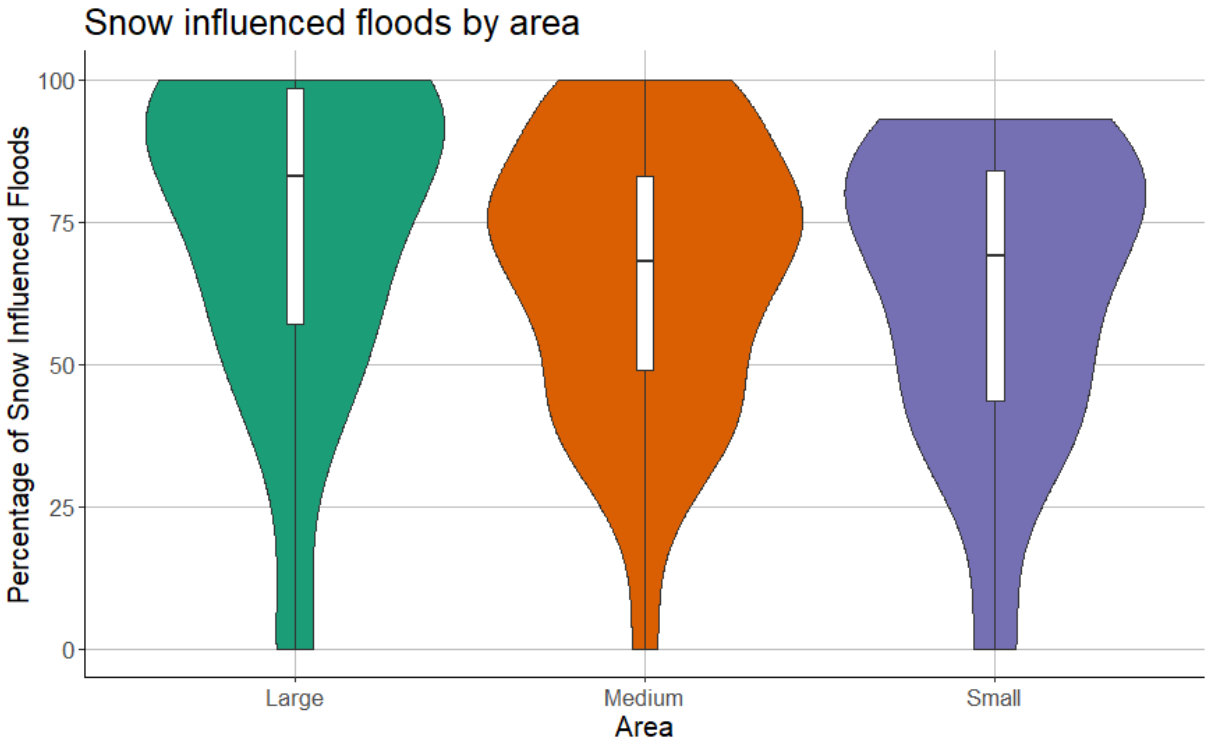

Figure 35: Percentage of snowmelt affected floods by area

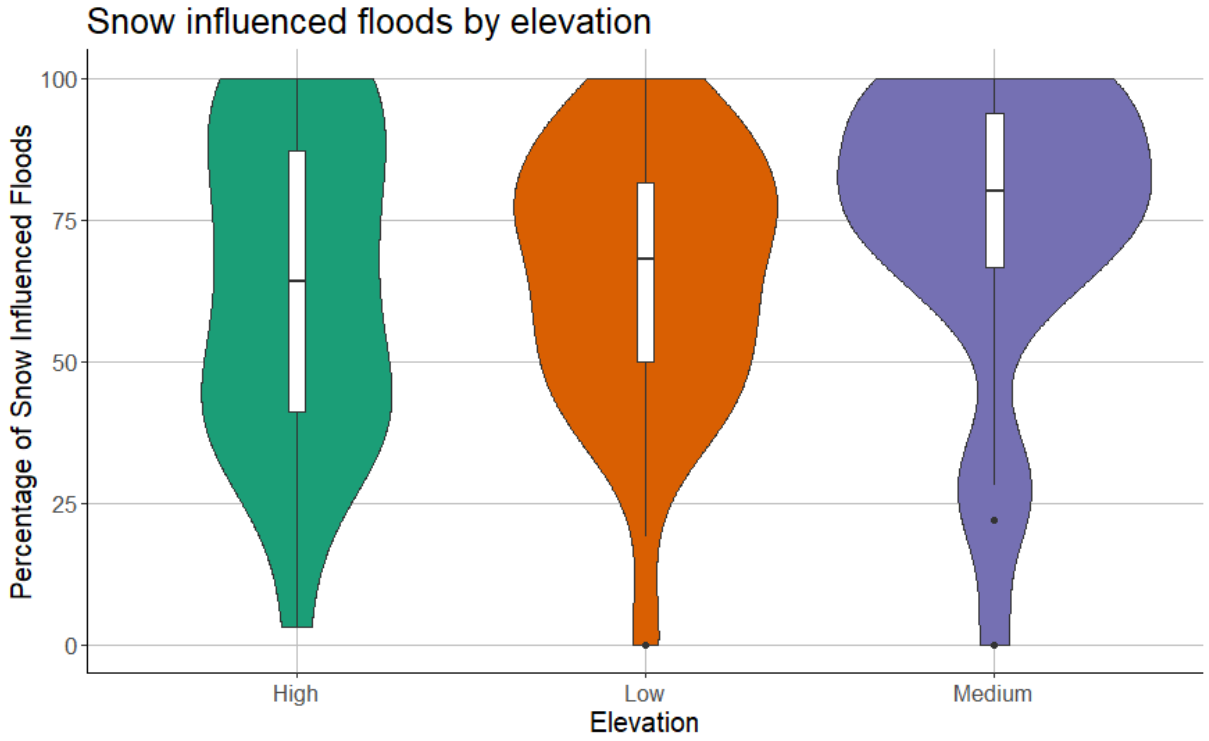

Figure 36: Percentage of snowmelt affected floods by elevation

4.3.2 Average day of occurrence of annual maximum flood

To determine the flood seasonality of each catchment, the method proposed by Burn (1997) was followed, as described in Section 3.2.2. A plot was created for every catchment in the study which

display its flooding seasonality. Figure 37 displays two different flood seasonality plots. In these plots, the red line indicates the average flood date. The height of each bar is the peak volume of discharge for each flood. The scale of the flooding is located on the left side of the plot. The plot for each of the 107 catchments can be found in Appendix A. Figure 37 displays two examples, namely the Eidselv River in Norway on the left, and the Willamette River in the United States on the right. The Eidselv River has a weak seasonality, with an r of 0.27, indicating that the floods are almost equally as likely in every season. The Willamette River has a strong seasonality, with an r of 0.87, indicating that the annual maximum flood is likely to occur at the same time every year. In this case, the Willamette River has winter flooding.

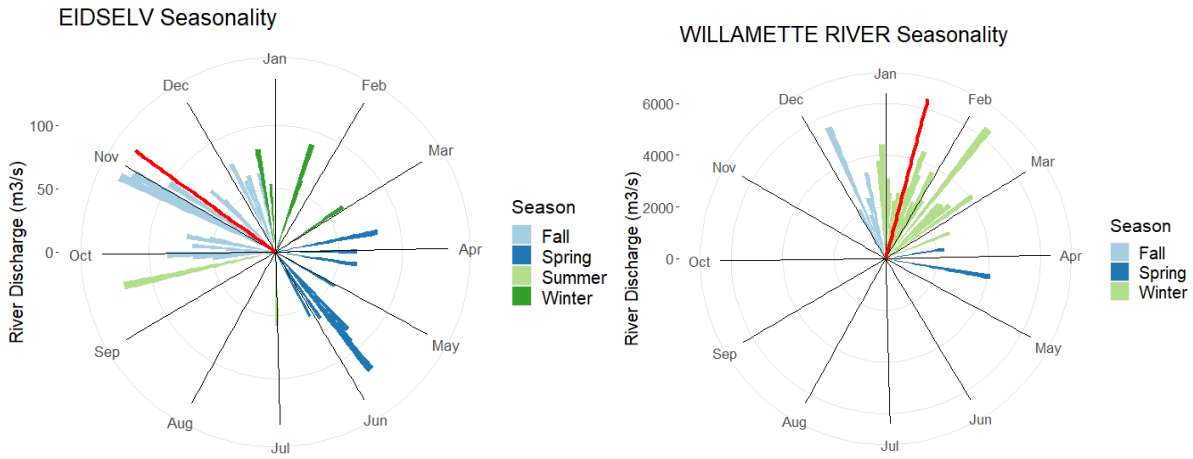

Figure 37: Seasonality of the Eidselv River and Willamette River

The average day of the annual maximum flood for each catchment was determined and are displayed on the maps in Figure 38 and Figure 39 for North America and Europe, respectively. The colour of the dot represents which month the maximum flood occurs in, and the size of the dot represents the variability in the day of the annual maximum flood. A large dot signifies a strong seasonality. A small dot signifies a weak seasonality.

In North America, the catchments in and near Alaska have a relatively strong seasonality, and the average flood days occur in the summer months. These catchments also have a very high percentage of snowmelt affected floods. This is likely due to the consistently large amounts of snow accumulated in this area. Snowpacks do not melt until later in the year due to low temperatures in this region, therefore the floods are occurring in the summer months. When the large amount of snow that has accumulated melts, this can often create the yearly maximum flood, hence the strong seasonality. In eastern North America many of the floods occur in February, March, and April. The consequences of having a vast area with the average flood days occurring at the same time can be very important – especially in relation to spatially compounding events. Spatially compounding events are when multiple events occur in the same geographic region during the same time frame, and can cause excess impacts (Zscheischler et al., 2020). Based on this map, the northeastern United States is susceptible to spatially compounding events, as the average flood dates are all in the spring. Central Canada has the average floods occurring in April, May, and June. In the Rocky Mountains, the flood variability is less consistent. Most of the catchments in the Rockies have a strong seasonality, with floods occurring in the summer months. Some catchments however have winter flooding, or a low seasonality, therefore no major patterns can be found.

In Europe, Central Europe has the potential for spatially compounding events, similar to northeastern North America. Here, the average day of flooding is in the spring. In Scandinavia, the average flood day occurs in September-December along the Norwegian coast, where there is a high percentage of snowmelt affected floods. This is likely due to the climate, and snow beginning to accumulate sooner than in other areas. In the interiors of Norway and Sweden there is a strong seasonality with the floods occurring in early summer. This can also indicate a potential area where spatially compounding events may occur. A study completed by Berghuijs et al. (2019a) investigated the synchronicity of floods across Europe. They found that Eastern Europe, the Netherlands and Northern Scandinavia all have large synchronicity areas, meaning that spatially compounding areas are prevalent in these regions (Berghuijs et al., 2019a). Iceland, which has a high percentage of snowmelt affected floods, seems to have no patterns in the seasonality, with the average flood days occurring in February, May, and October. In the Alps, the average flood occurs in the late summer months, with a relatively strong seasonality. This also coincides with the previous conclusions where the Alps experience many summer rainstorms, causing precipitation driven floods. In the UK, Ireland and Denmark, winter floods are most common, occurring in January. However, these regions have a low percentage of snowmelt affected floods, due to a lack of snow accumulation.

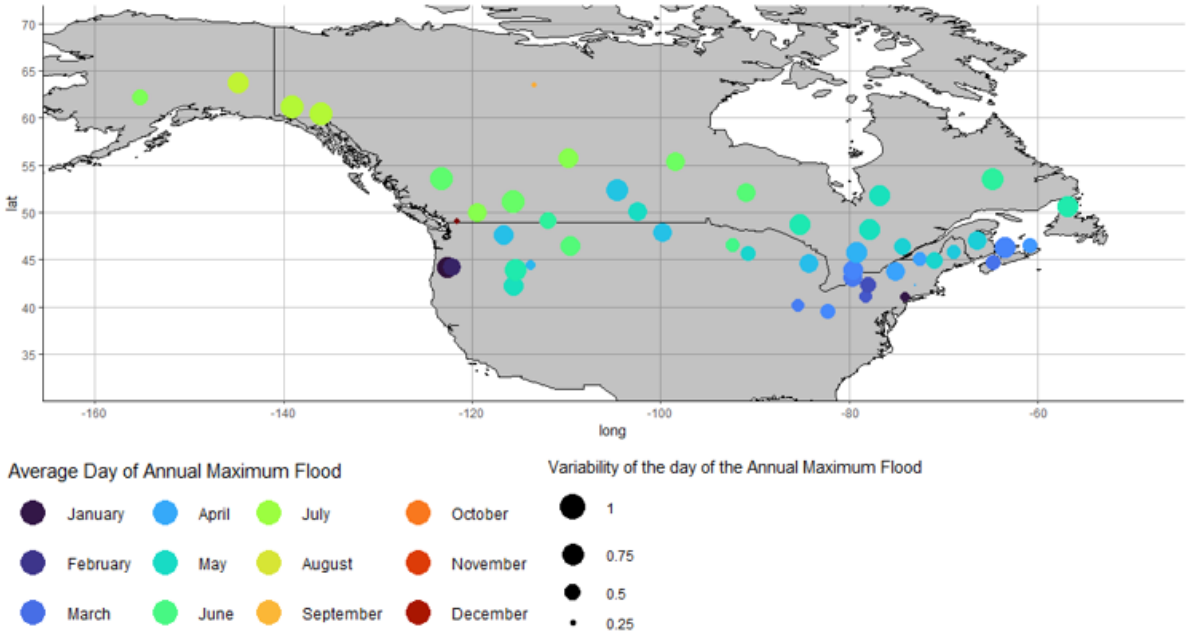

Figure 38: Average day of occurence of maximum annual flood and its variability for North American catchments

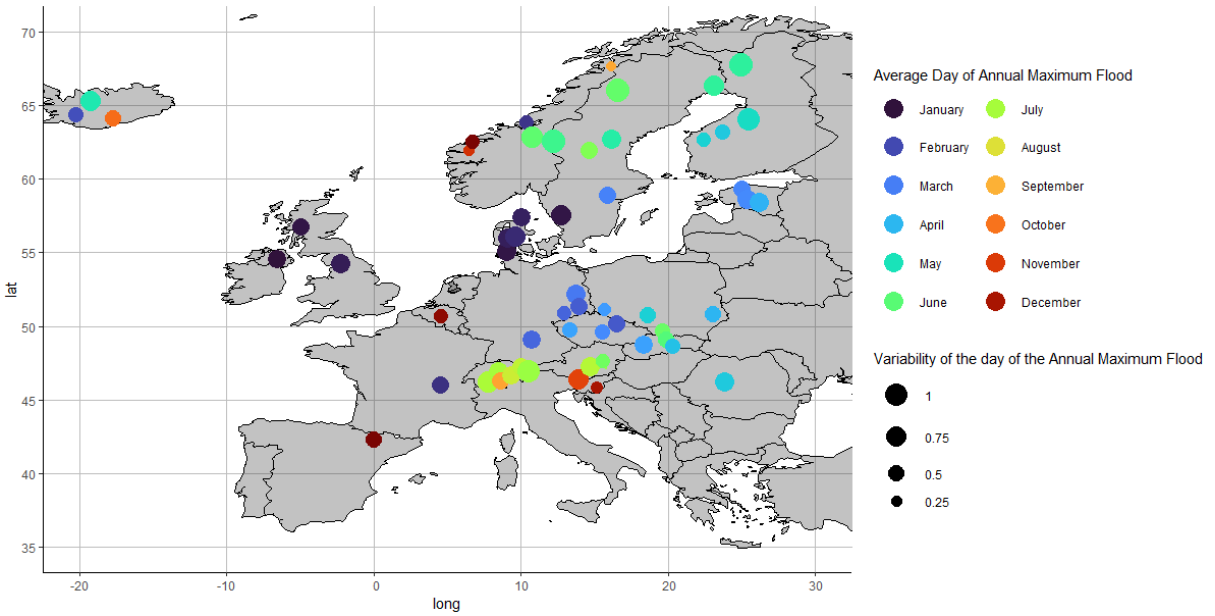

Figure 39: Average day of occurence of maximum annual flood and its variability for European catchments

When comparing the average day of the maximum flood by climate zone, it is clear that the warm summer humid continental climate zone has the strongest seasonality. There is a strong concentration of floods occuring from day 50 (February 19) to day 150 (May 30). The warm summer humid conintal climate zone makes up much of the northeastern United States, where there is the strong seasonality for spring flooding. It also makes up parts of Eastern Europe, including Estonia, Poland and Romania, where spring flooding also occurs. The temperate oceanic climate zone has the largest distribution of the average days. The subarctic and tundra climate zones have their largest concentrations later in the year, from days 150-200 (May 30-July 19). This is likely due to the colder weather weather and snow melting later in the year.

When investigating the relationship with the area, all three categories share a similar distribution, with the small catchments having the earliest mean date and the large catchments having the latest mean dates. Small and medium catchments have a distribution lasting all year long, whereas the large catchments it is concentrated in the Spring and Summer. When investigating the relationship in comparison to elevation, the results are similar to the expected outcomes. The high elevation category has floods occuring in the summer. This is likely due to the aforementioned summer storms and colder temperatures resulting in later snowmelts. The medium and low elevations have the highest concentration in the Spring, due to fact that these catchement are more affected by snowmelt.

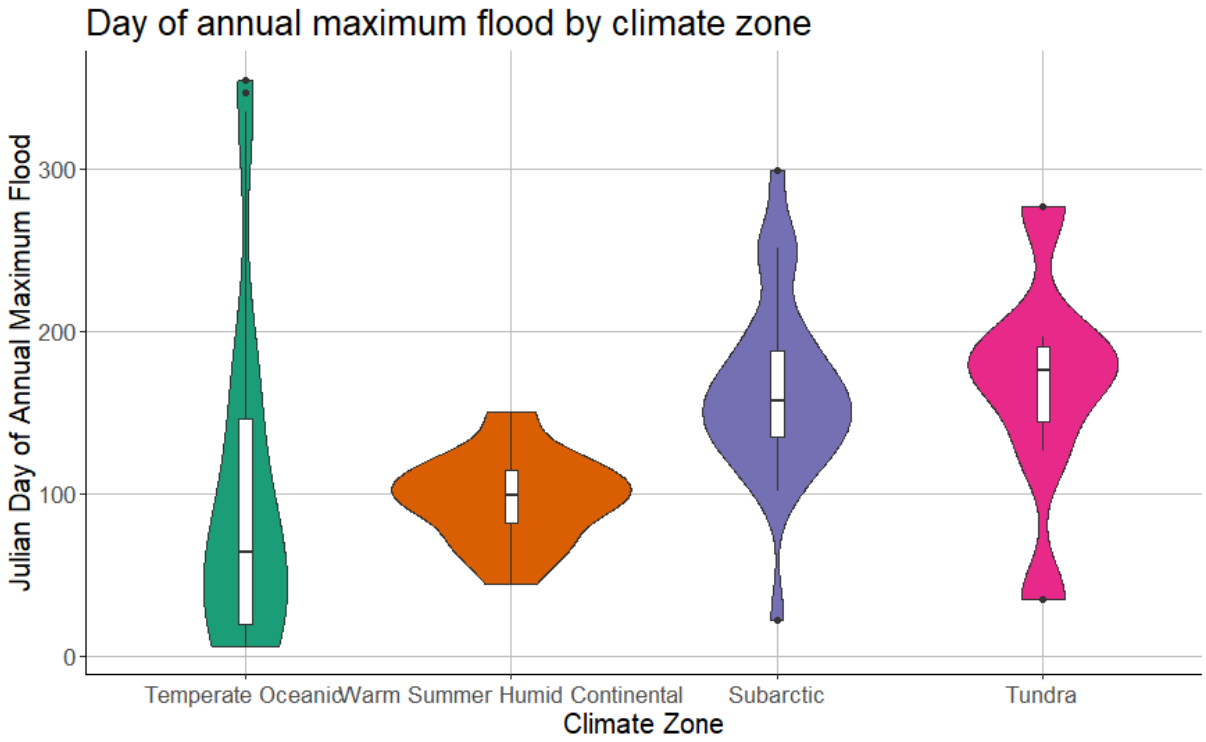

Figure 40: Average day of occurence of maximum annual flood and its variability by climate zone

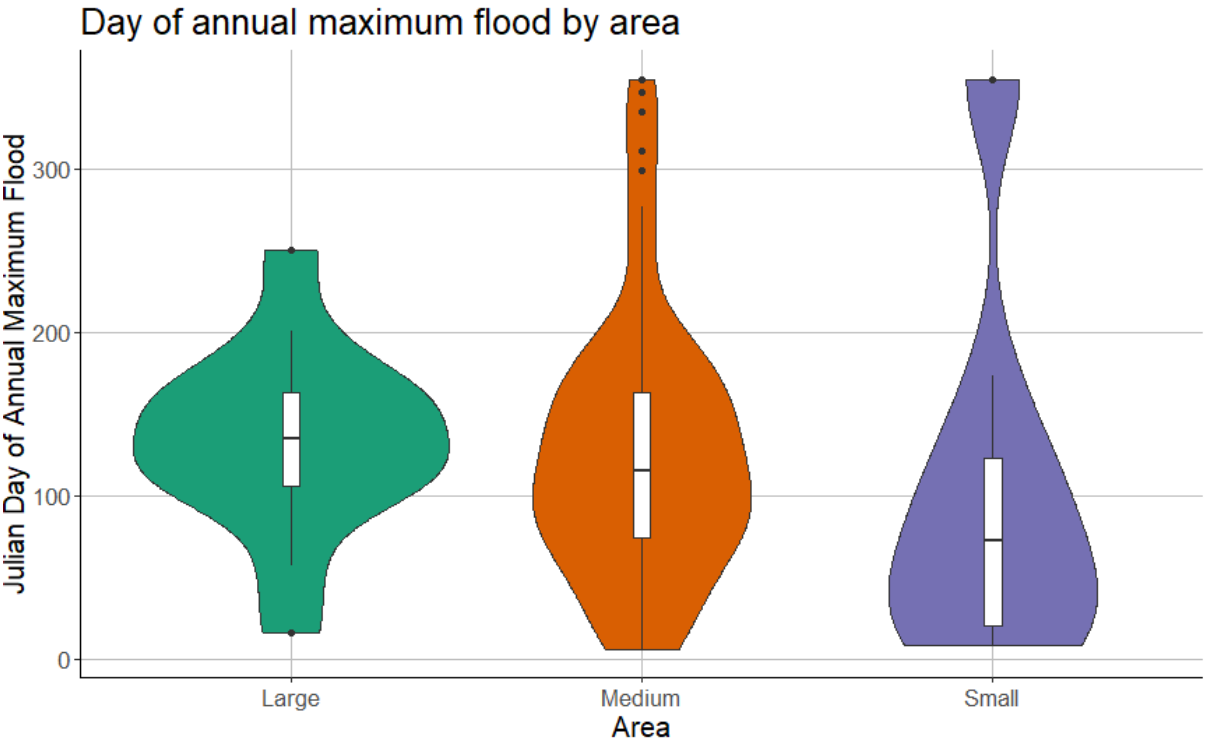

Figure 41: Average day of occurence of maximum annual flood and its variability by area

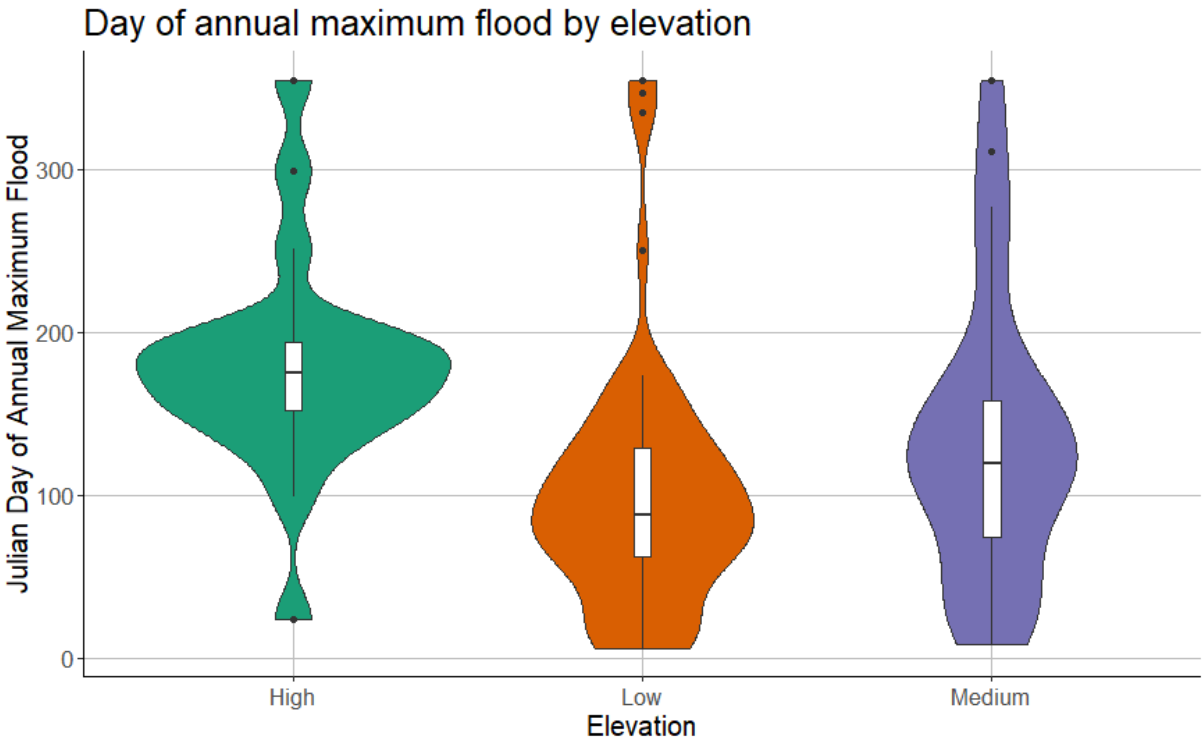

Figure 42: Average day of occurence of maximum annual flood and its variability by elevation

4.4 Relative influence of climate factors

To determine the relative influence of the climate factors, GBM regression trees were used, as described in Section 3.2.3. The value of the relative influence is normalized to be between 0 and 1, with 1 being the most important climate factor in determining the flow. The mean was taken of the relative influence of the climate factors per climate zone. This heat map can be seen in Figure 43. All climate zones were included in this heatmap, although the focus should be on the four climate zones of temperate oceanic, warm summer humid continental, subartic and tundra (CFB, DFA, DFC, and ET, respectively). However, the results do not vary drastically across the climate zone. The key takeaway is that soil-moisture is the dominant climate factor in determining the flow. In almost every climate zone the soil moisture content has a value 1, meaning it is the most important factor. This provides another example of the importance that soil moisture has on driving floods.

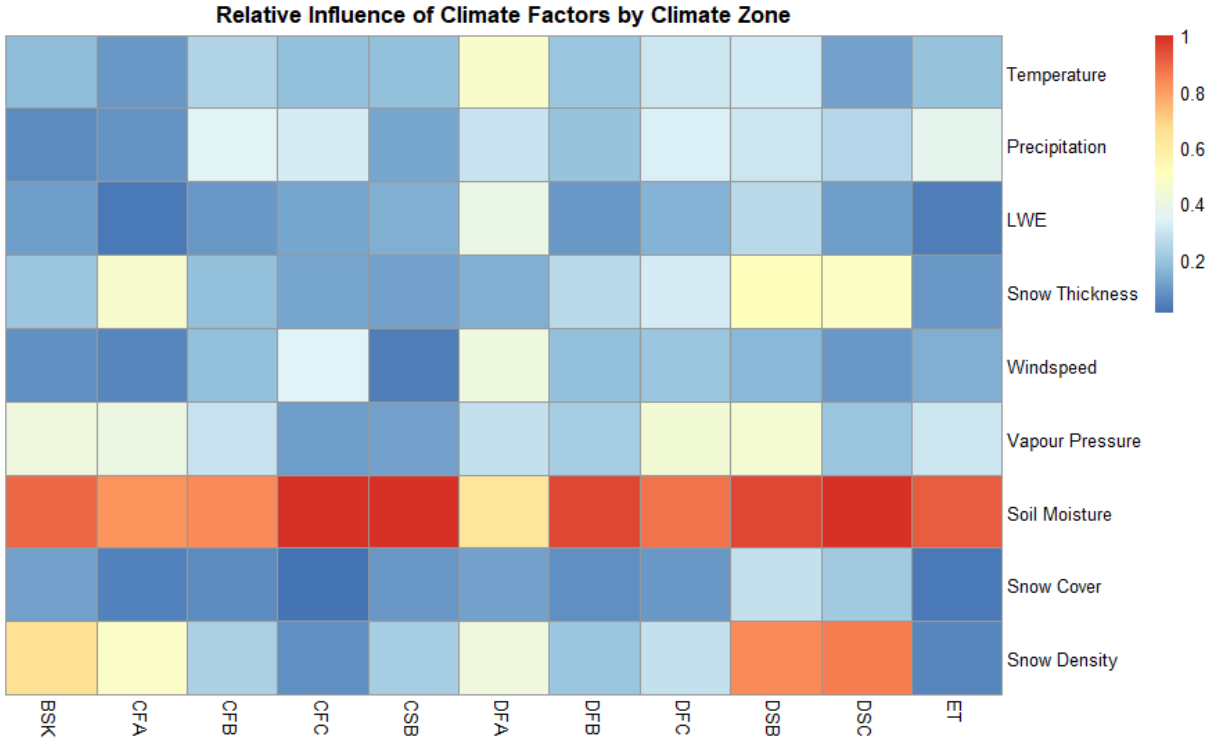

Figure 43: Mean relative influence of climate factors by climate zone

When investigating the relative influence of climate factors by area, the results coincide with the previous results. The large catchments, which have SMF floods as the predominant flood type, have an increase in the values of the snow related climate factors. There is an increase in the relative influence of the LWE, snow thickness, snow cover and snow density factors in the large area size category. This indicates that these values are more important when determining the flow in the river. In the large catchements, the snow density has a mean relative influence of greater than 0.5, the only climate factor other than soil moisture to cross this value. The snow related factors have the lowest relative influence in the small catchments.

The relative influence of the climate factors by elevation is shown in Figure 45. The high elevation catchments have the highest relative importance for the precipitation, and the lowest the for the snow related climate factors. This offers further support to the claim that the high elevation catchments have predominantly precipitation driven floods. The value of the soil moisture content is also highest in the high elevation catchments. In the medium and low catchments, the snow related climate factors have higher relative influences, which supports that these elevations are more dominated by winter floods.

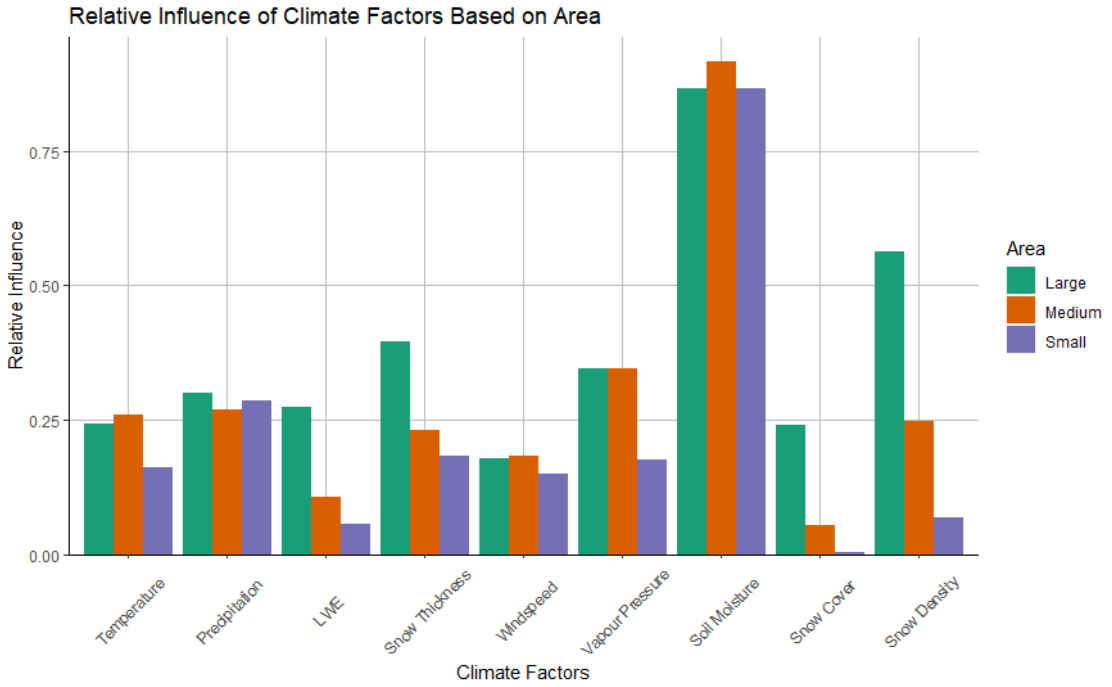

Figure 44: Mean relative influence of climate factors by area

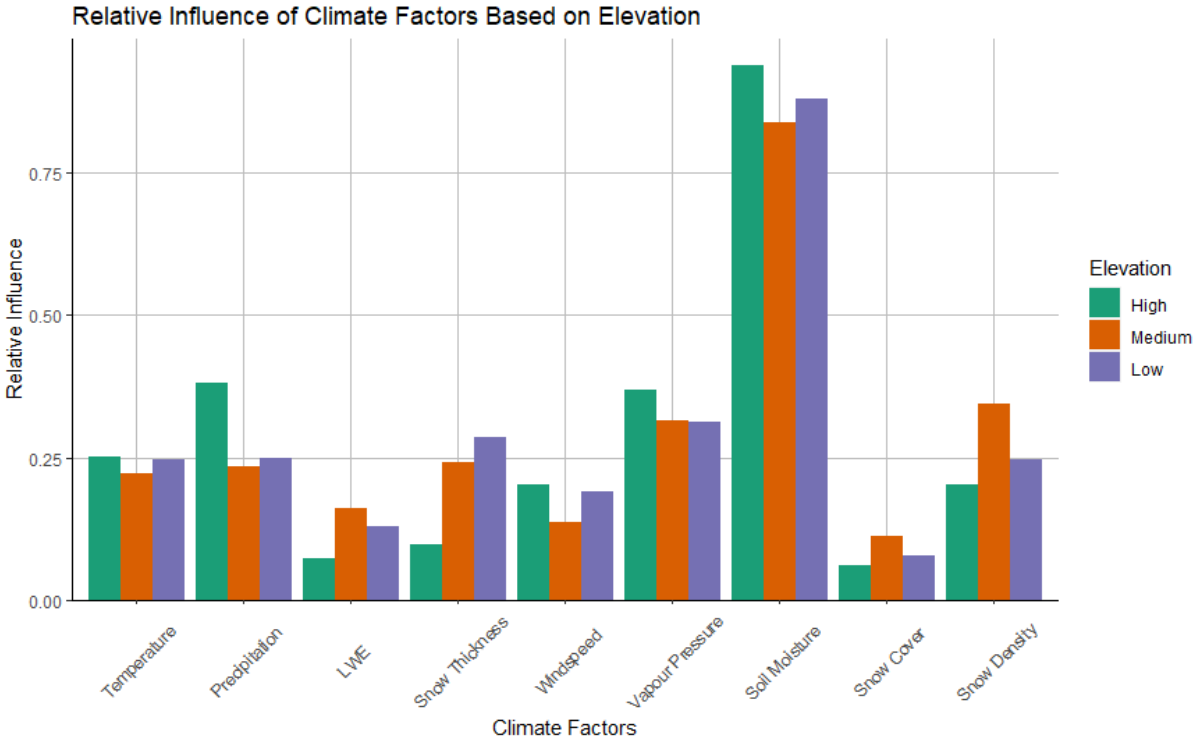

Figure 45: Mean relative influence of climate factors by elevation

4.5 Climate factors correlation

The climate factors were correlated against the flow for each flood per catchemnt. An example of two correllation matrices can be seen in Figure 46 for the Nemadji River in the USA. The left correlation matrix is for a flood that occured on March 16 in 2010. In this flood, there was initially 36 cm of snow cover, followed by a temperature rise from -1°C to 7°C. For this reason, the snow related climate factors had a much higher correlation value. The LWE, snow thickness, soil moisture, snow cover and snowmelt

are all highly correlated to the flow. Whereas in 1992 there was a flood that occured on July 5 when there was no snow present. Due to this, the snow related factors have no correlation to the flow. The precipitation and the windspeed are highly correlated to the flow. The Nemadji River has only 34% of its floods affected by snow, therefore more of the correlation matrices are similar to the 1992 correlation matrix. Correlation matrices for all catchments can be found in Appendix B.

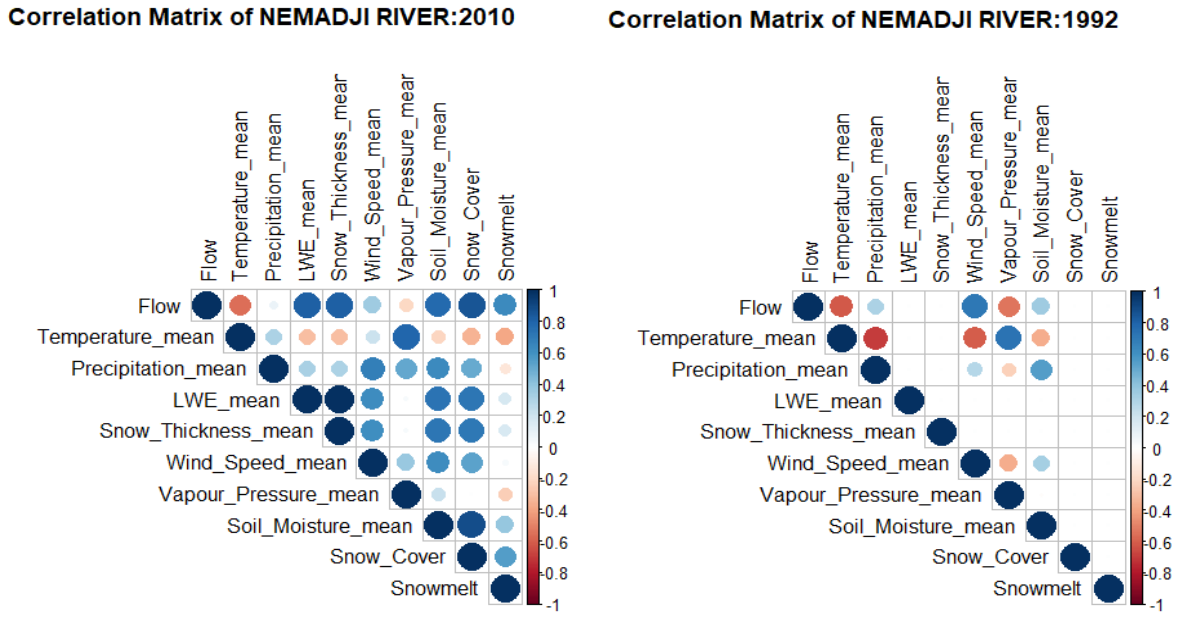

Figure 46: Correlation matrices of the Nemadji River

To compare all the catchments together, the correlation matrices were first converted to their absolute values, as both positive and negative correlations have a strong effect on the flow. Then, the mean value was taken for every catchment. Figure 47 displays the heatmaps comparing the correlation values to the climate zone. As with many of the findings in this study, the soil moisture remains the most highly correlated climate factor. Interestingly, in the subarctic climate zone, DFC, the flow was less correlated to the soil moisture and more correalted to the LWE and snow thickness. The subartcic climate zone does have a high percentage of maximum floods impacted by snowmelt, so this result is not unreasonable. The tundra climate zone, ET, has high correlation values for soil moisture, precipitation and temperature.

When investigating the climate factor correlation by area, shown in Figure 48, the soil moisture correlation value is surprisingly low for the large catchments. However, the snow related climate factors all remain the highest in their respective categories for the large catchments. This coincides with the large catchments having predominantly SMF floods as the snow factors all have a larger influence on the flow. For the precipitation, the small catchments have the highest values. Figure 49 displays the correlation values when sorted by elevation. Here, the results are also as expected and coincide with the previously found results. The precipitation climate factor has the highest value in the high elevation category. This supports the concept that large rainstorms are the dominant driver of floods in mountainous regions. The LWE and snow thickness have the highest values in the low category.

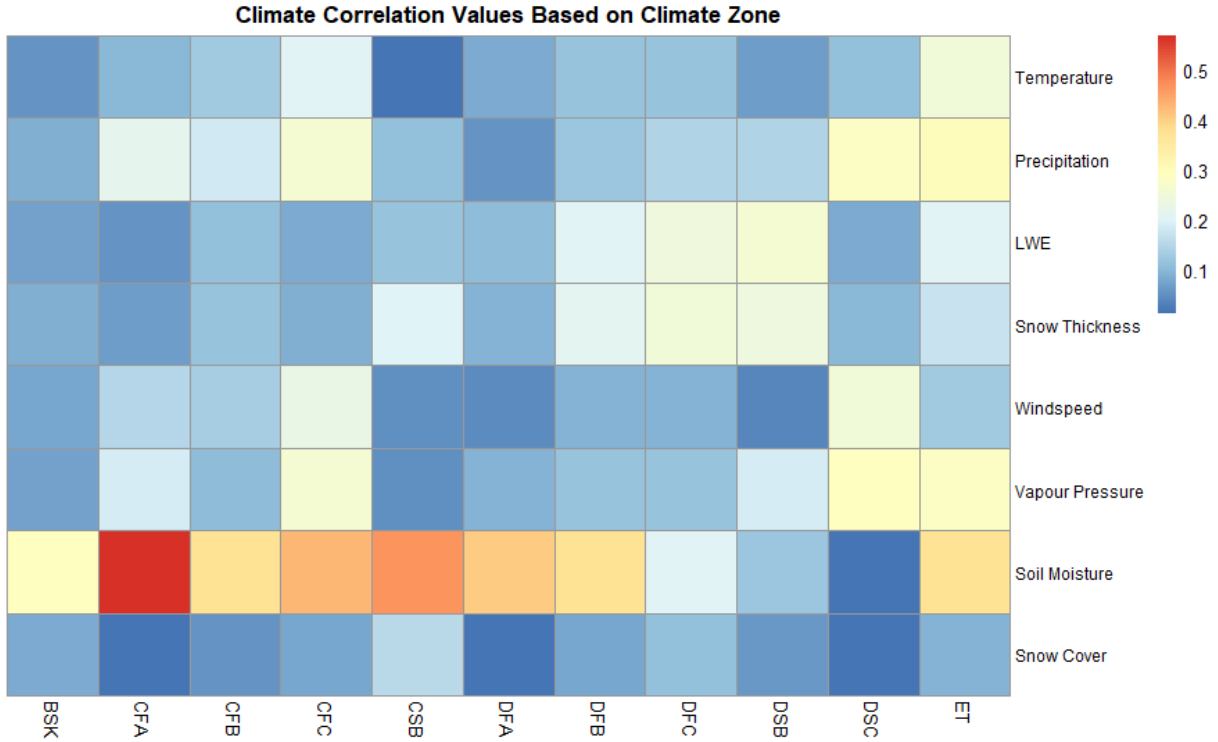

Figure 47: Mean correlation values of climate factors by climate zone

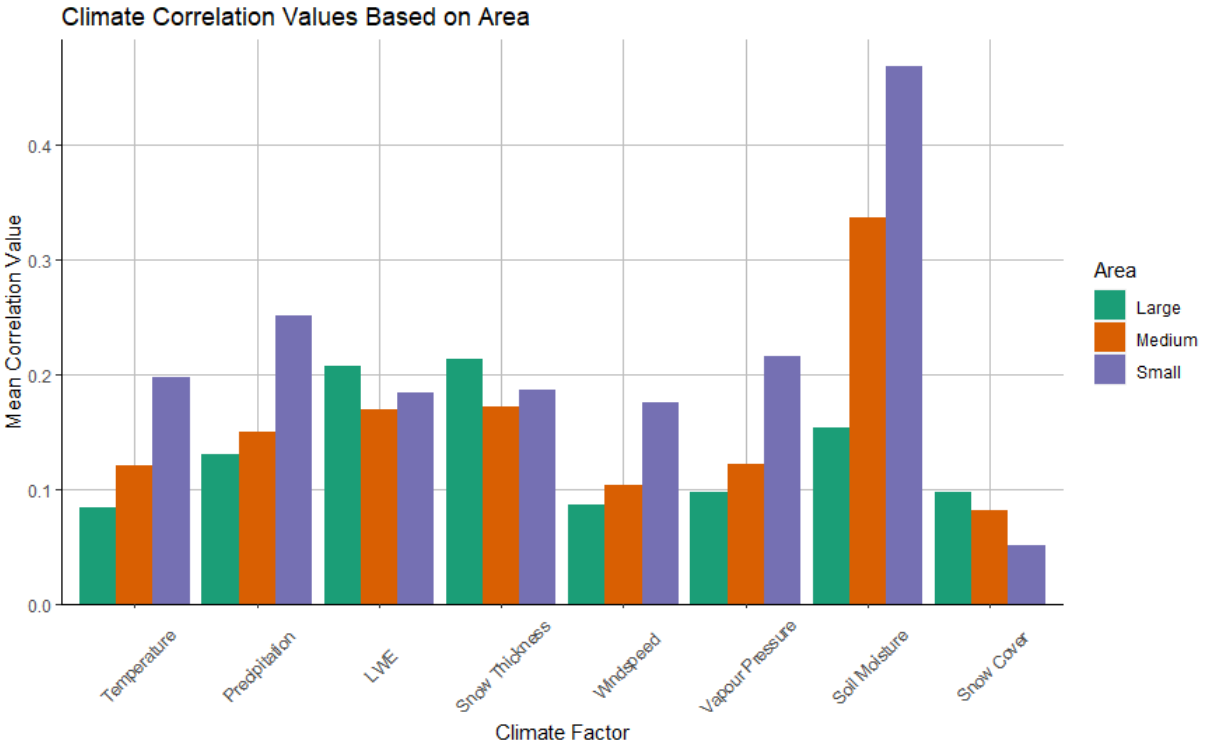

Figure 48: Mean correlation values of climate factors by area

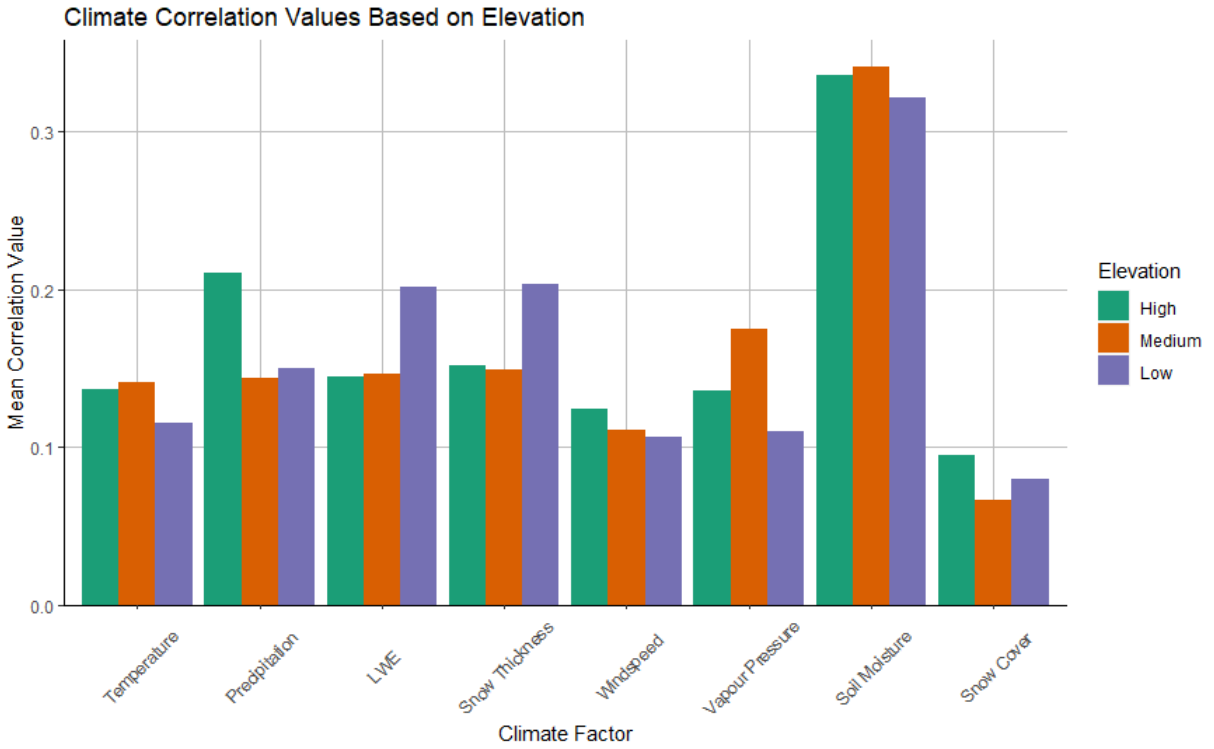

Figure 49: Mean correlation values of climate factors by elevation

4.6 Discussion related to compound events

This study focused on compound drivers in relation to snowmelt affected floods. Compound drivers are when multiple drivers/hazards occur in the same geographic region/time scale (Zscheischler et al., 2020). The classification of compound events was described in Section 1.0. To incorporate the concept of compound events into this study several actions were taken. When creating the typologies to categorize the floods into, all typologies are a compound event. To begin, an ROS flood is a typical example of what a compound event is. The snowpack on the ground is the pre-existing condition, and then the rain on top of the snowpack is an additional driver. When the two are combined, they can create the hazard. The other flood types, the SMF, SPF and LPF, are less typical compound events. However, with the application of the initial soil/snow moisture condition, these are also pre-conditioned compound events. Having a high antecedent soil moisture condition can be considered to be a pre-conditioned compound event (Zscheischler et al., 2020). Having a dry antecedent soil moisture condition does not qualify as a compound event, as a dry soil condition reduces the chances of flooding, rather than increasing it. A dry soil moisture condition can be a pre existing condition for other types of compound events however, such as drought or forest fires (Zscheischler et al., 2020). When inspecting the results of how many more floods occur in the wet initial condition rather than the dry initial condition, it becomes clear that the wet initial condition - compound events are naturally more dangerous than the dry initial conditions.

The study of the climate factors is an important part of the study of compound events. The key take-away from the study of the climate factors is how vital the soil moisture content is when determining if a flood will occur or not. While other factors like the snow thickness, snow cover and precipitation remain important, the importance of the soil moisture remains the largest in nearly every analysis. Soil moisture plays a key role in compound events, as it is often considered to be a pre-existing condition.

High levels of soil moisture can escalate other drivers like a light rainfall into flooding due to the lack of infiltration and excess runoff that is created (Zscheischler et al., 2020).

As described in Section 4.3, this study also has findings on spatially compounding events. Spatially compounding events are when multiple hazards occur in the same geographic region within a given time frame (Zscheischler et al., 2020). When certain regions all have their average flood days occurring in the same time frame, this can lead to increased impacts due to reduced access to resources and aid. There are certain regions in this study which are susceptible to spatially compounding events. Northeastern North America has a strong seasonality, with many of the floods occurring in the Spring. Central Sweden and Norway also have a strong seasonality, with many of the floods occurring in May and June. Central Europe, around Germany, Czech Republic and Poland, also have a tendency towards Spring floods, although the seasonality is less strong in this region.

An example of this spatially compounding event is occurring at the time of the writing of this thesis. Switzerland, Germany, Netherlands and Belgium all experienced catastrophic flooding (Gross, 2021). The floods have breached dykes, damaged bridges, submerged roadways, and at the time of writing, over 150 people lost their lives (Gross, 2021). Due to the large geographic scale of the flooding, rescue and aid efforts were delayed and have faced extra challenges. Many areas required assistance and there was an inability for transportation in and out of the most affected areas due to roads and bridges over a wide spread area being washed away (Gross, 2021). If the flooding only occurred in one region or river, then the rescue efforts would have been able to properly protect the residents in that area. However, due to the widespread damages and flooding, this is an example of a spatially compounding event.

4.7 Study limitations/future study recommendations

This study has certain limitations and possible next steps for studies in the future. The first limitation is the number of the river stations used and the geographic regions/climate zones where the catchments are located. The study was limited to the catchments with data from the GRDC, without missing data from 1979-2019 and had an assoicated shapefile. In the study, only four climate zones had a large enough sample size for a data analysis to be conducted on them. The introduction of more stations would allow for more climate zones to be incorporated into the analysis. Some geographic regions, such as Northern Canada and Eastern Europe do not have a many appropriate catchments from the GRDC. Therefore, catchements for this area are missing. Incorporating data from other sources would allow for a greater spatial analysis, and allow for the geographic gaps to be filled.

The next possible limitation of the research is the quality of the data. It is vital to have accurate data, especially for the LWE, snow thickness, and snow coverage data. The data in this study was re-analysis data from Copernicus. Xiao and Che (2020) conducted a study which compared the Copernicus ERA5 reanalysis snow coverage to other methods, including AMSR-E, AMSR2, and GlobSnow. The study showed that the ERA5 data is accurate in areas where there is a shallow snow depth, of less than 10 cm (Xiao and Che, 2020). The BIAS error is approximately zero for the majority of North America and Europe. In areas which recieve over 50 cm of snow, the reanalysis data also performs the best of all the methods. The reanlysis data has the best perfomance in areas of plains and forrest. There is however an over-approximately of the renalysis data in mountainous regions which experience less than 50 cm of snow, by up to 10 cm. In these regions, GlobSnow has the best performance. In future studies, having

area-specific data sources may produce more accurate results. However, the majority of stations in this study have the most accurate snowfall data.

The thresholds for sorting the catchments into various typologies were based of a literature review. However, a sensitivity analysis could be completed on the thresholds which determine the typology of the flood. This sensitivity analysis could allow for certain flood types to become predominant in certain areas. Additionally, a method similar to Sikorska et al. (2015) could be reproduced where the sorting of typologies has a fuzyy method appraoch. Here, one flood can be partially sorted into multiple typologies, giving a more accurate representation of what the drivers are for the flood. Finally, when decided on the flood typologies, this study focused on initial moisture content in the soil and snow as the main dividing factor. Additional studies can focus on precipitation intenstity, rate of the rise of temperature, etc.

5 CONCLUSIONS

This study assessed the drivers of snowmelt floods in relation to compound hazards. The climate drivers considered in this study are temperature, precipitation, snow thickness, snow LWE, wind speed, vapour pressure and soil moisture content. 107 different catchments across North America and Europe were investigated, from the years 1979-2019. Using the different climate factors, each annual maximum flood was sorted into a different flood typology. These typologies were rain-on-snow floods, snowmelt floods, long precipitation floods, and short precipitation floods. These four types of floods are all split into another two categories, with a wet initial condition and a dry initial condition.

The results indicate that the considered catchments have snowmelt floods being the dominant flood type. When comparing the relationships between the flood typologies and the catchment climate zones, catchment elevation and catchment areas further conclusions could be drawn. The most notable conclusions were that the primary flood type in the subarctic category was snowmelt floods with a wet initial condition, and in tundra climate zone that the snowmelt floods and rain-on snow floods were the primary flood types. When comparing the catchment by elevation, the high elevation catchments had the dominant typologies being the short precipitation floods and the long precipitation floods. This initially surprised us, as higher elevations have colder weather and thus are expected to be more influenced by snowfall. However, similar findings were also made by Siksorka et al. (2015) and supported by other later findings in this study. The precipitation driven floods at higher altitudes are often caused by summer rainstorms. Lower elevation catchments had snowmelt driven floods being the dominant typology. When comparing the typologies based on the catchment area it was found that the large catchments had a much higher portion of snowmelt driven floods, while the medium and small catchments had a more precipitation floods.

The other key takeaway is that the wet soil condition floods have a much higher frequency than the dry soil condition floods. This is result is expected and known, having already been researched in great depths. However, the sorting of typologies was able to quantify this. The high frequency of the wet soil initial conditions indicates that this can be a type of compound event – where a high antecedent moisture condition would qualify as a pre-existing event. This was already suggested by Zscheischler et al. (2020), however the findings from this study support their claims.

The seasonality of the floods was also determined, both in the percentage of the floods that occur with the influence of snow, and the average day of the year the floods occur on. Based on the results of this analysis, there are geographic regions, like northeastern North America and central Sweden and Norway, which have a strong seasonality and the floods occur at roughly the same time every year. This can lead to spatially compounding events, where many hazards occur in one geographic region at the same time, thus amplifying their impacts (Zscheischler et al., 2020).

The relative influence of the climate factors and the correlation values of the climate factors to the flow were also investigated. These findings supported the earlier findings, where the soil moisture content plays a key role in the determination of floods. The soil moisture content had the strongest relative influence of all the climate factors, as well as having the strongest correlation values. This, once again, confirms how important the pre-conditioned compound events of a high antecedent soil condition are.

Other results also supported earlier conclusions drawn. In high elevation catchments, precipitation had a stronger relative influence and correlation while the snow related climate factors had reduced importance. Also, the large catchments had stronger influence of the snow related climate factors, while the small catchments had less influence of these snow related climate factors.

REFERENCES

Amatulli, G., Domisch, S., Tuanmu, M., Parmentier, B., 2018. Data Descriptor : A suite of global , cross-scale topographic variables for environmental and biodiversity modeling. Scientific Data 5: 1–15.

Balke, T., Nilsson, C., 2019. Increasing Synchrony of Annual River-Flood Peaks and Growing Season in Europe. Geophysical Research Letters 46: 10446–10453. https://doi.org/10.1029/2019GL084612

Benesty, J., Chen, J. Huang, Y., Cohen, I., 2009. Pearson Correlation Coefficient, in: Springer Topics in Signal Processing. p: 1–4. https://doi.org/10.1007/978-3-642-00296-0

Berghuijs, W.R., Allen, S.T., Harrigan, S., Kirchner, J.W., 2019a. Growing Spatial Scales of Synchronous River Flooding in Europe. Geophysical Research Letters 46: 1423–1428. https://doi.org/10.1029/2018GL081883

Berghuijs, W.R., Harrigan, S., Molnar, P., Slater, L.J., Kirchner, J.W., 2019b. The Relative Importance of Different Flood-Generating Mechanisms Across Europe. Water Resoures Research 55: 4582–4593. https://doi.org/10.1029/2019WR024841

Borchers, H., 2021. Package "pracma." URL https://cran.r-project.org/web/packages/pracma/pracma.pdf (accessed 5.1.21).

Burn, D.H., 1999. Perceptions of flood risk: A case study of the Red River flood of 1997. Water Resoures Research 35: 3451–3458.

Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology 202: 212–230.

CCID, 2019. The climate of the European Alps: shift of very high resolution Köppen-Geiger climate zones 1800-2100 [WWW Document]. Climate Change and Infectious Diseases Group URL http://koeppen-geiger.vu-wien.ac.at/alps.htm (accessed 3.1.21).

Copernicus, 2021a. Agrometeorological indicators from 1979 to present derived from reanalysis [WWW Document]. URL https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.6c68c9bb?tab=form (accessed 3.15.21).

Copernicus, 2021b. ERA5 hourly data on single levels from 1979 to present [WWW Document]. URL https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview (accessed 3.20.21).

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecolology 77(4): 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Fischer, S., Schumann, A., Bühler, P., Schumann, A., 2019. Timescale-based flood typing to estimate temporal changes in flood frequencies. Hydrological Sciences Journal 64: 1867–1892. https://doi.org/10.1080/02626667.2019.1679376

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29: 1189–1232.

Fuka, DR, Walter MT, Archibald JA, Steenhuis TS, and E.Z., 2018. Package 'EcoHydRology.' URL https://cran.r-project.org/web/packages/EcoHydRology/EcoHydRology.pdf (accessed

5.1.21).

Gottschalk, L., Krasovskaia, I., 2002. L-moment estimation using annual maximum (AM) and peak over threshold (POT) series in regional analysis of flood frequencies. Norsk Geografisk Tidsskrift 56: 179–187. https://doi.org/10.1080/002919502760056512

GRDC, 2020. Global Runoff Data Centre (GRDC) [WWW Document]. Federal Institute of Hydrology (BfG), Koblenz, Ger. URL https://www.bafg.de/GRDC/EN/Home/homepage_node.html (accessed 3.1.21).

Greenwell, B., Boehmke, B., Cunningham, J., 2020. Generalized Boosted Regression Models. URL https://www.rdocumentation.org/packages/gbm/versions/2.1.8 (accessed 6.15.21).

Gross, J., 2021. Europe Flooding Deaths Pass 125, and Scientists See Fingerprints of Climate Change. NY Times.

Handcock, M., 2016. Package "reldist." URL https://cran.r-project.org/web/packages/reldist/reldist.pdf (accessed 6.15.21).

Harpold, A.A., Molotch, N.P., Musselman, K.N., Bales, R.C., Kirchner, P.B., Litvak, M., Brooks, P.D., 2015. Soil moisture response to snowmelt timing in mixed-conifer subalpine forests. Hydrological Processes 29: 2782–2798. https://doi.org/10.1002/hyp.10400

Hayes, A., Anderson, S., 2021. Gini Index [WWW Document]. Investopedia. URL https://www.investopedia.com/terms/g/gini-index.asp (accessed 7.1.21).

Hijmans, R., 2021. Package "raster." URL https://cran.r-project.org/web/packages/raster/raster.pdf (accessed 3.1.21).

Ho, C.L.I., Valeo, C., 2005. Observations of urban snow properties in Calgary , Canada. Hydrological Processes 19: 459–473. https://doi.org/10.1002/hyp.5544

Koffler, D. Gauster, T., Laaha, G., 2016. Calculation of Low Flow Statistics for Daily Stream Flow Data. Packag. "lfstat" 0.9.4. URL https://cran.r-project.org/web/packages/lfstat/lfstat.pdf (accessed 4.1.21).

Kuusisto, E., 1984. Snow accumulation and snowmelt in finland.

Lyne, V., Hollick, M., 1979. Stochastic Time-Variable Rainfall-Runoff Modelling. Australian National Conference 79: 89–93.

Merz, R., Blöschl, G., 2003. A process typology of regional floods. Water Resoures Research 39: 1–20. https://doi.org/10.1029/2002WR001952

Nathan, R. J., & McMahon, T.A., 1990. Evaluation of Automated Techniques for Base Flow and Recession Analyse. Water Resoures Research 26: 1465–1473.

Papadakis, M., Tsagris, M., Dimitriadis, M., 2021. Package "Rfast." URL https://cran.r-project.org/web/packages/Rfast/Rfast.pdf (accessed 6.1.21).

R Core Team, 2020. A Language and Environment for Statistical Computing. URL https://www.r-project.org/ (accessed 3.1.21).

Raymond, C., Horton, R.M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., Bowen, S.G., Camargo, S.J., Hess, J., Kornhuber, K., Oppenheimer, M., Ruane, A.C., Wahl, T., White, K., 2020. Understanding and managing connected extreme events. Nature Climate Change 10: 611–621. https://doi.org/10.1038/s41558-020-0790-4

Sikorska, A.E., Viviroli, D., Seibert, J., 2015. Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. Water Resoures Research 51: 7959–7976. https://doi.org/10.1002/2015WR017326

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., and Tallaksen, L.M., 2020. Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions. Hydrology and Earth System Sciences 24.

Sui, J., Koehler, G., 2001. Rain-on-snow induced flood events in southern Germany. Journal of Hydrology 252: 205–220. https://doi.org/10.1016/S0022-1694(01)00460-7

Vormoor, K., Lawrence, D., Heistermann, M., Bronstert, A., 2015. Climate change impacts on the seasonality and generation processes of floods – Projections and uncertainties for catchments with mixed snowmelt/rainfall regimes. Hydrology and Earth System Sciences 19: 913–931. https://doi.org/10.5194/hess-19-913-2015

Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D., Wong, W.K., 2016. Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway. Journal of Hydrology 538: 33–48. https://doi.org/10.1016/j.jhydrol.2016.03.066

World Meterorological Organization, 2008. Manual on Low-flow Estimation and Prediction.

Xiao, L., Che, T., 2020. Evaluation of Remote Sensing and Reanalysis Snow Depth Datasets over the Northern Hemisphere during 1980 – 2016. Remote Sensing 12: 1–31.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R.M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M.D., Maraun, D., Ramos, A.M., Ridder, N.N., Thiery, W., Vignotto, E., 2020. A typology of compound weather and climate events. Nature Reviews Earth and Environment 1: 333–347. https://doi.org/10.1038/s43017-020-0060-z

LIST OF APPENDICES

APPENDIX A: SEASONALITY PLOTS

These plots can be found in the attached folder labelled Appendix A: Seasonality Plots.

APPENDIX B: CORRELATION MATRICES

These plots can be found in the attached folder labelled Appendix B: Correlation Matrices.