M.Eng Project Report

DROUGHTS AND DEEP WATER

A DATA SCIENCE EXPLORATION OF DYNAMIC WATER RIGHTS AND VULNERABILITIES IN THE UPPER COLORADO RIVER BASIN

Ananya Gangadhar (ag757)
Systems Engineering, M.Eng '21

In collaboration with

Dr. Antonia Hadjimichael
Prof. Patrick Reed

TABLE OF CONTENTS

1. BACKGROUND

1.1. COLORADO RIVER

The Colorado River is one of the largest rivers in North America, covering a distance of 1,400 miles from southwestern United States into Mexico. The Colorado River basin is divided into two parts — the Upper and Lower. The Upper Colorado River Basin spans Colorado, Utah, Wyoming, New Mexico and Arizona, and supplies around 90% of the water in the river basin (OWDI Drought website). This project will focus on a subbasin of the Upper Colorado — the Upper Colorado River Basin within the state of Colorado (UCRB) occupying over 9,900 square miles, shown in Figure 1. This subbasin is also known as Water Division 5 in the state of Colorado.

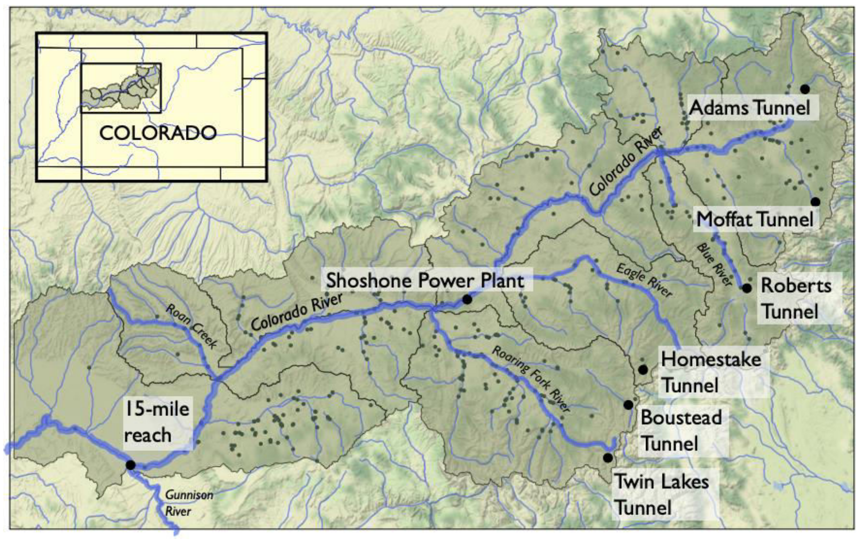

Figure 1. Map of the UCRB in the state of Colorado showing the water rights as blue dots, important transbasin diversion tunnels, and the Shoshone Power Plant (Hadjimichael at al., 2020b).

1.2. PRIOR APPROPRIATION DOCTRINE

Water use rights vary drastically across the United States. States in the northeast typically have Riparian water rights which is a water allocation system for landowners with properties along the riverbanks. However, states in the west follow the prior appropriation ("first in time, first in right") doctrine where the first person to take a quantity of water from the river for "beneficial use" has the right to that quantity of water. Additional people can gain new water rights, as long as their water rights do not infringe upon the older water rights. The older water rights therefore get preference in water allocation in such river systems. This is also the case in the UCRB where the seniority of water rights plays a major role in defining stakeholder relationships (Hadjimichael et al., 2020a).

1.3. WATER USERS IN THE BASIN

One of the main uses of water in the UCRB is irrigation, and there are several thousand irrigation ditches along the river and its tributaries to irrigate an area of approximately 390 square miles. Eighty percent of Colorado residents live east of the Continental Divide, whereas most of the precipitation falls west of the

Divide, which channels the water down the Colorado river. As a result, there are several transbasin diversions from the UCRB exporting around 460,000 acre feet of water for agricultural, municipal and industrial use. There are also several power plants along the UCRB that use its water for power generation, and the Shoshone Power Plant (shown in Figure 1) is one of the most important rights in the basin. In totality, the UCRB supports agricultural, industrial, municipal and recreational activities worth approximately $300 billion in the state of Colorado alone. However, this economy is at risk with the increasing hydrological stresses and demands within the basin (Hadjimichael et al., 2020b). The Colorado river is experiencing a severe and historic 20-year drought since 2000. The shortages in streamflow caused by the drought, combined with institutionally complex water allocations means that the state agencies must consider the resource competition among thousands of stakeholders while also accounting for potential vulnerabilities to climate change, demand growth due to increasing population within the basin, as well as infrastructure changes (Hadjimichael et al., 2020b).

2. Project Scope

Previous research on the UCRB includes bottom-up vulnerability assessments of water users in the basin, and evaluations of their vulnerability to water shortages as well as their robustness to uncertain future conditions (Hadjimichael et al., 2020a; Hadjimichael et al., 2020b; Quinn et al., 2020). In a bottom-up approach, several states of the world are generated that take into account hydrologic, social and institutional changes. The performance of a given water right is then evaluated in all these states of the world and the most important user vulnerabilities and uncertainties are identified. This approach allows the exploration of vulnerabilities while taking into account the "deep uncertainties" inherent in such complex systems. Through their analysis, Hadjimichael et al. (2020a, 2020b) show that vulnerability and robustness vary significantly across water users, but also across different definitions of acceptable performance. Further, these studies show that the uncertain factors driving water shortage also differ from user to user, suggesting the presence of complex relationships between the hydrologic, societal and institutional components of the system. These highly complex interactions between the different water users and the hydrology of the system, as well as between individual users, ought to be explored in order for the Colorado agencies to determine how these users will be impacted by future stressors like climate change, demand growth and infrastructure changes.

This project studies how past droughts and other hydroclimatic events have propagated through the UCRB water allocation network to affect users of varying seniority, location and demand. To do this, the project uses publicly available datasets of all historical water withdrawal requests from users in the UCRB, which can then be analyzed over time to dynamically map the interactions between the multitude of users in the basin. The project collects and uses large hydrologic and climate datasets available for the UCRB in order to analyze the impact of hydrological conditions at different times on the allocation network. Insights from this project can help illuminate how the presence of human institutions interacts with hydroclimatic stressors to shape the shortages experienced by individual users. Even though previous studies have illustrated the complexity of these relationships, how users affect each other and how those effects change under more stressful conditions remains unclear.

The goal of this project is to build a set of network visualizations of user interactions in the UCRB study the complex relationships between users and hydrological conditions. This information can be used by local agencies to better understand how the hierarchy of water rights affects individual users during times of water stress, as well as how to leverage specific critical water rights to improve water supply for the basin as a whole.

3. Project Logistics

3.1. Timeline

This project spans two semesters — Fall 2020 and Spring 2021 and follows a typical data science project workflow of data collection followed by data processing, exploration, analysis and visualization. In Fall 2020 there are three main project tasks. The first task is project set up and involves the creation of a syllabus, defining the problem and reviewing existing literature to understand the system context. The second task is training in all the skills necessary for this project which include but are not limited to bash scripting, Git, running parallel code, working with remote servers, and coding in Python and R. The third task is to collect the UCRB water rights and hydrological data and perform some initial visual exploration of this data. In Spring 2021 the project focus is on analyzing user networks, building network visualizations, and framing the results into a cohesive narrative. This timeline is formalized as a syllabus with periodic milestones and scheduled testing in order to keep the project on track for completion by May 2021.

3.2. Organization

All documents for this project are stored either on GitHub or on Box. All of the code (Python, R, or otherwise) is stored on GitHub under the *Shortages_network* project repository which is updated weekly. Project progress can be tracked in the weekly slide decks made for each meeting which are stored on Box along with all previous drafts of this report.

4. Data Collection

4.1. CDSS Database

The data for this project is primarily collected from the Colorado Decision Support System (CDSS) which is a collection of databases, models, data viewing and management tools jointly developed by the Colorado Water Conservation Board and the Division of Water Resources to support water resource planning in the state of Colorado. The CDSS website provides Representational State Transfer (REST) services, which allows users to access the databases, submit query requests and download data.

4.2. Types of Data Collected

There are two broad categories of data used in this project. The first is data on water rights and includes — 1) information about each basin structure and its associated water rights, and 2) time series call analysis showing the percentage of each day that a specified structure was out of priority and the downstream call in priority.

The second category is hydrological data of the UCRB and includes — 1) streamflow, and 2) climate conditions (precipitation, minimum and maximum daily temperatures, etc.) measured as a time series across gages in the basin.

4.3. Automating Data Collection

Since the CDSS database is RESTful, a data request can be set up by constructing a Uniform Resource Locator (URL) which is a web address that contains the query and format of data needed. This project involves the collection of data from thousands of stations and users, necessitating the need of a programming language like Python to automatically generate the URLs and submit requests to CDSS.

In order to control website traffic, the CDSS REST services places a daily limit on the number of rows of data that can be requested. For anonymous users this limit is 1,000 calls per day and 600,000 rows of data, while registered users may download up to 1,000,000 rows of data daily. For this project, a registered account was set up to maximize how much data could be collected per day.

The water rights and hydrological data all have different parameters and require separate Python scripts to set up queries. Each Python script iterates through a list of all gages (or structures), checks the list for data which has already been collected, selects a new subset of gages with under 1 million rows of data, generates and submits a formatted query using the open-source Requests library in Python, and downloads the data. Each output file is named after the gage (or structure) identifier. This process is computationally expensive and is executed on a remote cluster called THECUBE managed by the Center for Advanced Computing at Cornell University. The output can then be downloaded to a local machine using any secure file transfer method. In order to maximize computational efficiency on the cluster and prevent overloading any one node, each Python data collection script is modified to automatically distribute tasks evenly among five processing cores in the cluster and run them in parallel.

5. Data Processing

The CDSS datasets are typically of high-quality and are largely free of any missing values or unfamiliar formatting that may throw errors, necessitating very little cleaning. However, there are some caveats to keep in mind with the datasets used in this project.

The first pertains to the climate data. A climate gage can collect more than one kind of measurement, which means that climate gages that measure several kinds of metrics will have extremely large file sizes that may exceed the CDSS REST limit of 50,000 rows of data per page of request. To overcome this issue, separate requests are submitted for type of climate data and the outputs are stored under separate directories.

The second caveat is regarding the difference between a structure and a water right. Each structure has an associated unique identifier called its Water Diversion ID (WDID), and all the water rights data collected is organized by WDID. However, each WDID can hold multiple water rights of varying degrees of seniority since water rights are tradeable. Additionally, user interaction networks in this project have been built on annual time scales, so the structure call analysis data is reformatted in Python to be organized by year for the entire basin rather than by individual WDID across its lifespan.

6. Data Exploration

6.1. Basin structures

There are several thousand water rights in the UCRB. This project considers one diversion structure in the basin as equivalent to one user, and each diversion structure has a unique identifier (WDID). As mentioned in the above section, each user/diversion structure location may have more than water right associated with it. There are over 24,000 structures in the UCRB or Division 5. However, most of them divert extremely small amounts of water and not all are active in a given year.

6.2. Structure Call Analysis

The CDSS database Analysis tools can perform a call analysis that returns a time series showing the percentage of each day that a structure was out of priority by a downstream call. This data is collected for each structure in the basin, identified by its WDID. Each row of data contains the structure put out priority (Analysis WDID), the percentage of the day it was put out of priority, the most downstream structure controlling and diverting the entire available natural stream flow (Location WDID) as well as the structure whose water right priority is the most junior diverter within the call reach (Priority WDID).

6.3. Streamflow

The CDSS database has daily streamflow measurements from 393 streamflow stations or gages in the UCRB. Each gage has a unique Station ID. Not all 393 gages are currently active, and these gages have a wide range of operational lives, shown as a histogram in Figure 2. Future analysis should only include currently operational gages in order to make meaningful observations about the entire system. The period of focus in this project will be 2000-2019 since this also coincides with the most recent and ongoing 20-year drought in the UCRB.

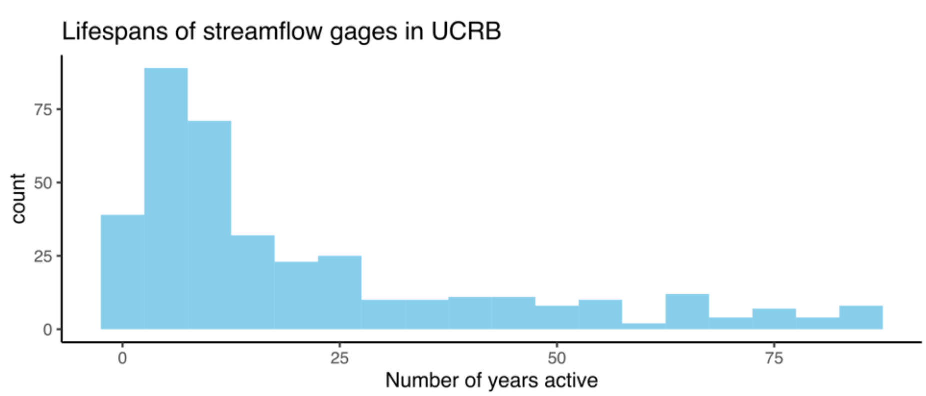

Figure 2. Histogram showing the number of operational years of 393 streamflow gages in the UCRB.

6.4. CLIMATE

There are 1180 climate gages in the UCRB, each measuring one of ten types of climate data — 1) Maximum temperature, 2) Minimum temperature, 3) Precipitation, 4) Snow, 5) Snow depth, 6) Evaporation, 7) Snow-water equivalent (estimated depth of liquid water contained within snowpack), 8) Mean temperature, 9) Solar (radiation energy per area), 10) Vapor pressure. Each gage has a unique Station ID. Table 1 below shows the number of gages measuring each type of data in Division 5. However, not all of them are currently operational.

Table 1. The number of gages in Division 5 corresponding to each data type.

Measurement Type	Number of gages
Maximum Temperature	110
Minimum Temperature	110
Precipitation	340
Snow	271
Snow Depth	229
Evaporation	6
Snow-Water Equivalent	59
Mean Temperature	13
Solar	14
Vapor Pressure	14

A map showing the locations of each diversion structure, streamflow gage and climate gage is shown in Figure 3. This visual map is created in Python using the cartopy library. Figure 3 shows the distribution of all water rights and stations in Division 5; however, it does not provide any information on which water rights and/or gages were active at any given point of time. For visualizing the spatial as well as temporal distribution of structures and gages, a GIF of Figure 3 animated through time can be found in the GitHub project repository.

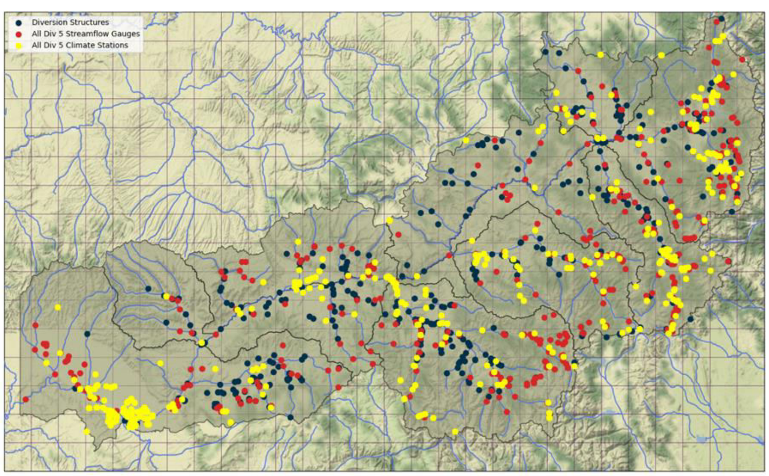

Figure 3. Map of Division 5, or the UCRB, showing the water rights diversion structures (blue), streamflow gages (red) and climate stations (yellow).

7. Geospatial Network Analysis

7.1. Structure Call Analysis Networks

Directed network graphs are constructed for every annual structure call analysis using the NetworkX library in Python. In directed graphs, the edges all have a direction from one node to another and this is visually represented with an arrow on the edge. Each node represents a user in the UCRB while each edge represents a call put out by one user on the other. The source node is the user putting out a call to divert streamflow while the destination node is the user being put out of priority by the call. Each node has a set of associated attributes that include 1) latitude, 2) longitude, 3) stream mile, 4) net absolute volume of water allocated to that user, and 5) structure name. Each edge has an associated weight that represents the number of days in that year that at least one of the water rights owned by the destination node user was put out of priority by at least one of the water rights of the source node user.

There are two possible ways to construct the structure call analysis graphs based on the columns of data obtained from CDSS. The first is to create a network using Analysis WDID (destination) and Location WDID (source) as edges, while the second is to use Analysis WDID (destination) and Priority WDID (source) as edges. Both Location WDID and Priority WDID are putting Analysis WDID out of priority in a given call. The difference between the two is that Location WDID is the most downstream diversion structure that is controlling and diverting the entire available natural stream flow while Priority WDID is the structure whose water right priority is the most junior diverter within the call reach. For this project, Priority WDID will be used as the source node. This is because graphs with Location WDID as the source are much less-connected and offer little new insight since it is well known and expected that senior water rights will put the more junior water rights out of priority. Using Priority WDID will allow for more highly connected graphs that reveal relatively junior water rights that end up diverting water away from users in the UCRB.

Figure 4 shows the network graphs mapped geospatially using the latitude and longitude coordinates embedded in each node. The three years for which networks are shown in Figure 4 are chosen based on mean annual streamflow in the UCRB during the chosen period of study (2000-2019). 2011 has the maximum streamflow, 2007 the median and 2002 has the lowest streamflow. The hydrograph for the mean annual streamflow is shown in Figure 5.

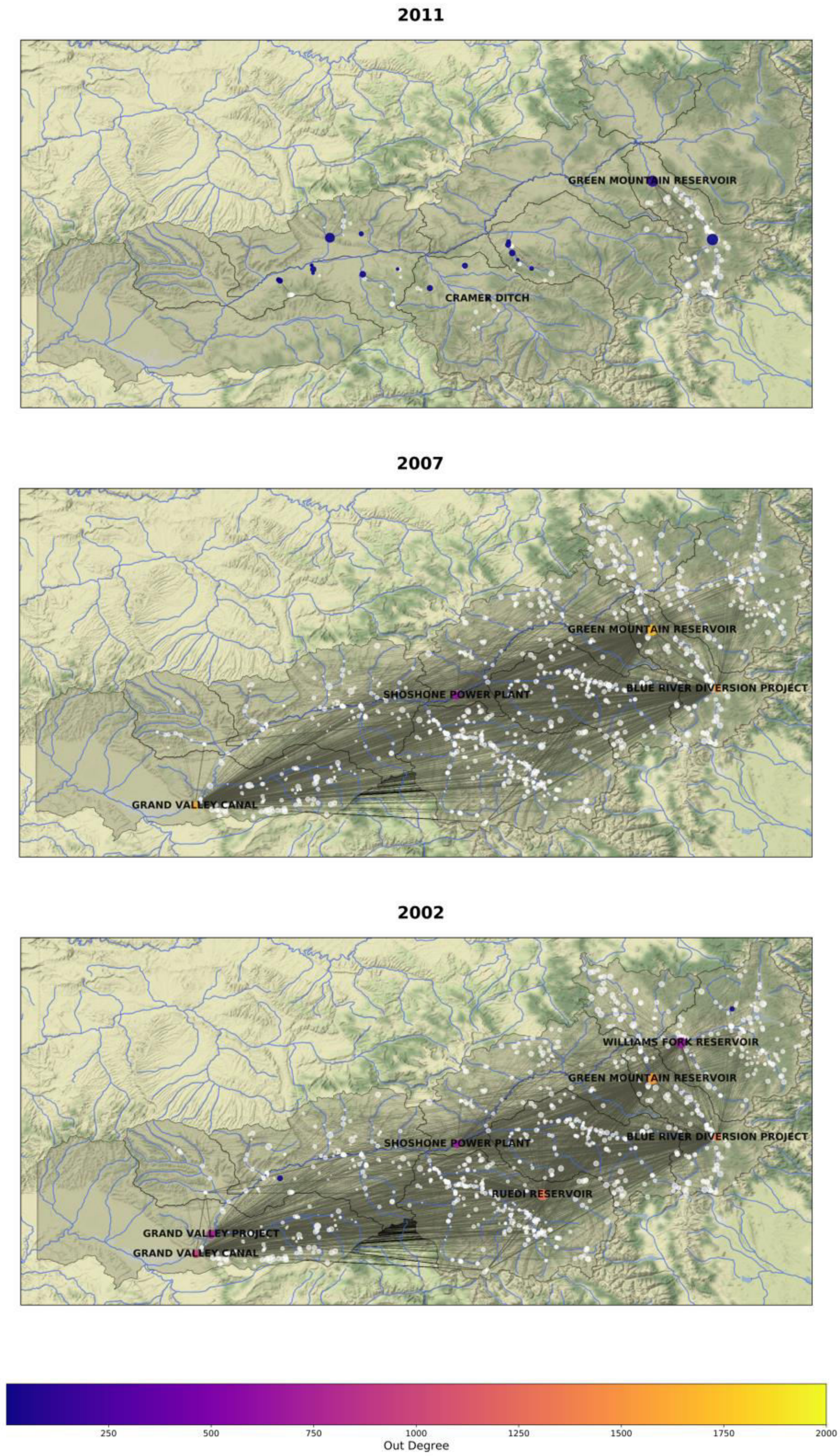

Figure 4. UCRB annual user interactions mapped geospatially as networks for the years with (from top to bottom) minimum, median and maximum mean annual basin streamflow. Each node represents a WDID structure while each edge represents an interaction where one user (source) puts the other (destination) out of priority in the given year. The nodes are colored by out-degree, with the ones in white having out-

degree of zero. The nodes are sized by their log-scaled net absolute allocation (in cfs). All nodes with out-degree greater than 20 are labelled.

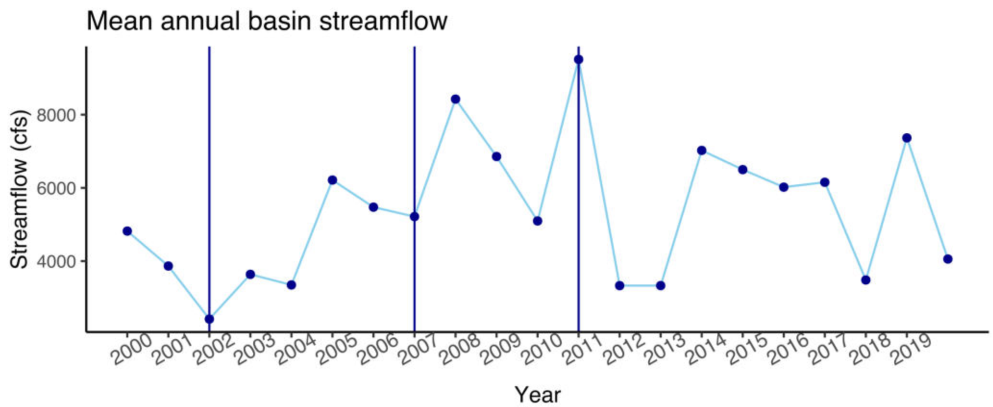

Figure 5. Hydrograph showing the mean annual streamflow in the UCRB from 2000-2019. Highlighted with blue vertical lines are the three years chosen — corresponding to the maximum (2011), median (2007) and minimum (2002) mean annual basin streamflow.

From Figure 4 it is evident that during the wet year (2011) there is less competition for water and as a result, fewer users put calls on the river to divert flow away from other junior users. In the dry year (2002), there are many more calls put out in the UCRB by several senior users like Shoshone Power Plant and Grand Valley Canal which results in a greater number of users being put out of priority. The networks are not well-connected and only a handful of nodes have out-degrees greater than zero in each case. Additionally, in the wet year (2011) the network consists of several smaller independent components.

7.2. Number of Network Components

The layout of Figure 4 makes it difficult to discern the components of larger networks like the median year (2007) or the dry year (2002). Figure 6 shows the number of components in each annual network from 2000-2019. The trend shown in Figure 6 does not correlate with streamflow, however there is a drastic increase in the number of network components after 2007. The cause for this increase is not clear, but may potentially be institutionally driven, and is worth exploring in the future.

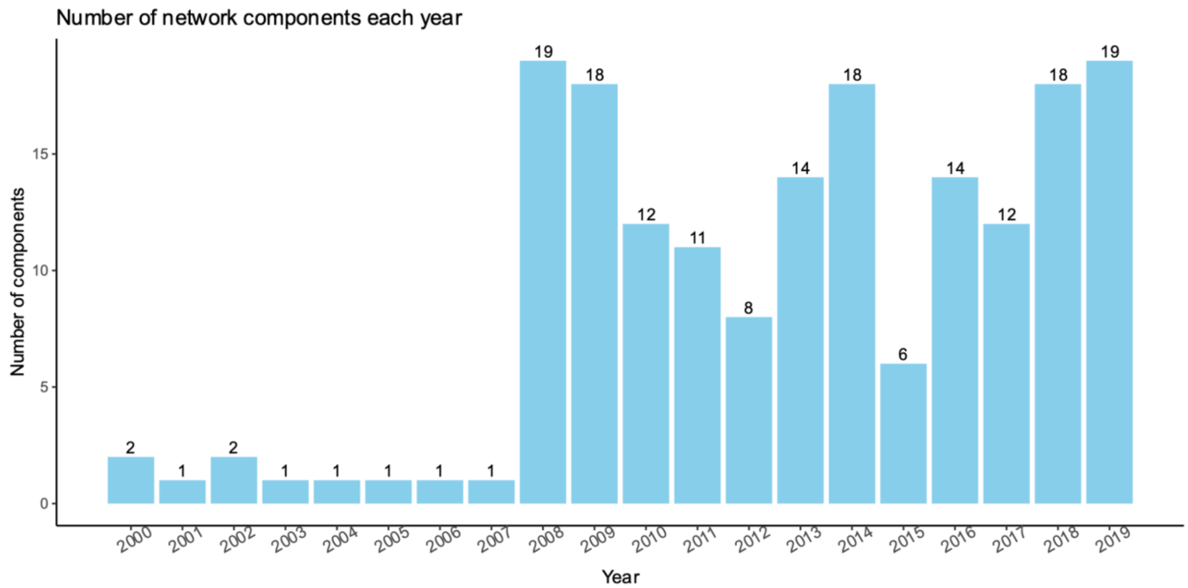

Figure 6. Bar graph showing the number of components in each user network from 2000-2019.

8. Alternative Network Visualizations

There are several ways to visualize the UCRB user networks besides the geospatial mapping. The igraph and ggraph libraries in R are used in this project to create alternate visualizations from which new insights may be derived. In order to visualize the user interactions with greater clarity, subgraphs of the annual user networks that include only the interactions between the 200 senior-most WDIDs in the basin are used.

8.1. Circular Layout

The ggraph library in R allows visualizing networks with all the nodes placed along the circumference of a circle. Figure 7 shows subnetworks with the 200 senior-most users in the UCRB for the three chosen years. In the wet year (2011), none of the senior users put calls on each other. However, in both the median (2007) and dry year (2002), senior users like Shoshone and Grand Valley Canal are responsible for putting several other users in the subnetwork out of priority. Sylvan Lake, McMahon Ditch and McMahon Reservoir No. 2 lose the most quantity of water in the median (2007) and dry (2002) years.

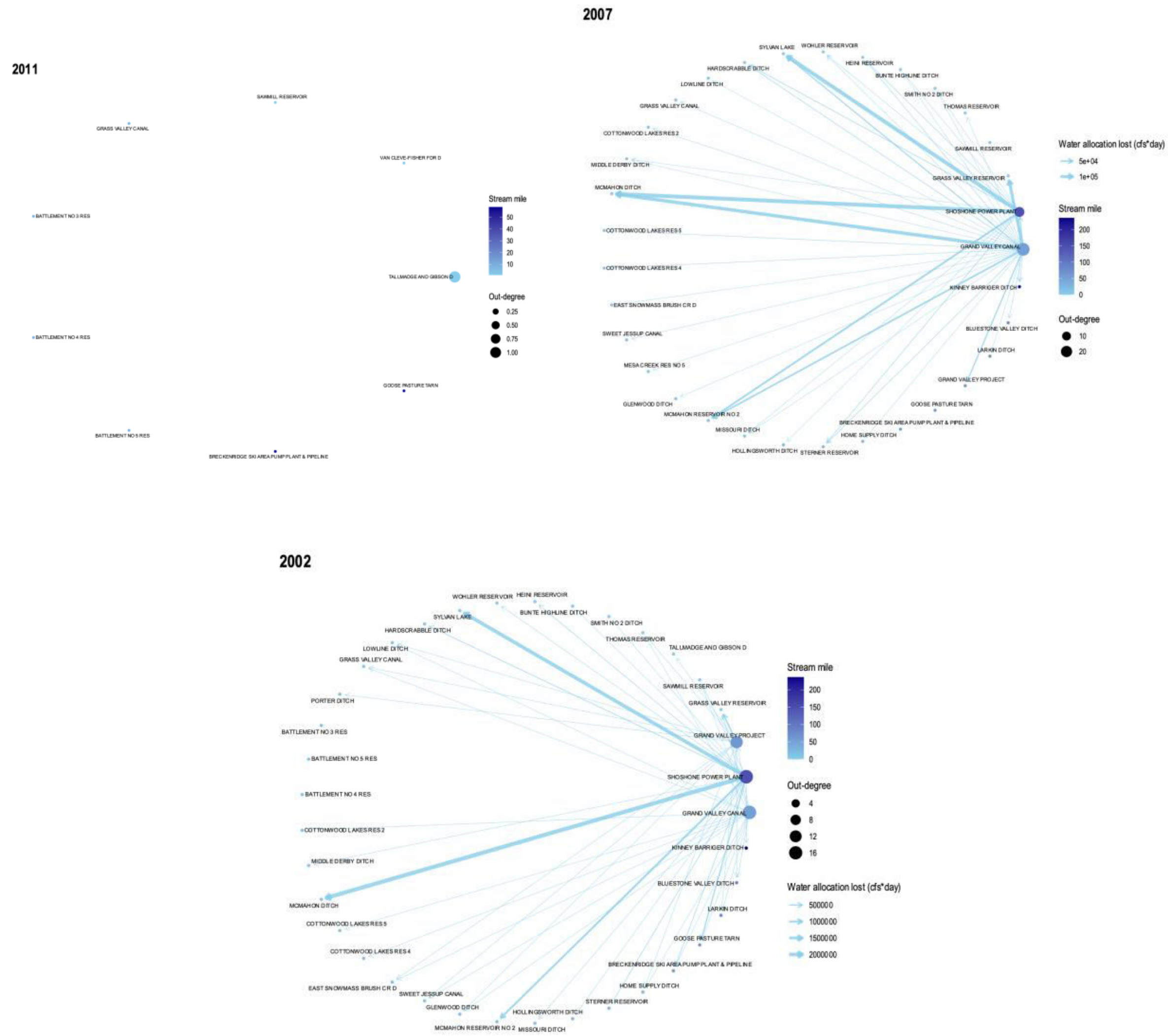

Figure 7. Circular layout graphs showing subnetworks of the UCRB that include only the 200 senior-most water rights in the basin. The nodes are sized by out-degree and colored by stream mile. The edge weights represent the volume of water (in cfs*day) lost by the user put out priority.

8.2. Number of Senior Users

From Figure 7, it appears that the number of senior users putting each other out of priority increases from wet years to dry. Figure 8 shows the number of senior users in each annual UCRB network from 2000-2019. This value is equivalent to the number of nodes in each circular layout subgraph. While the trend in Figure 8 does not correlate to UCRB streamflow, the number of senior users putting each other out of priority is increasing over time. This is likely due to the effects of the 20-year drought which carry over each year as reservoir levels continue to decrease.

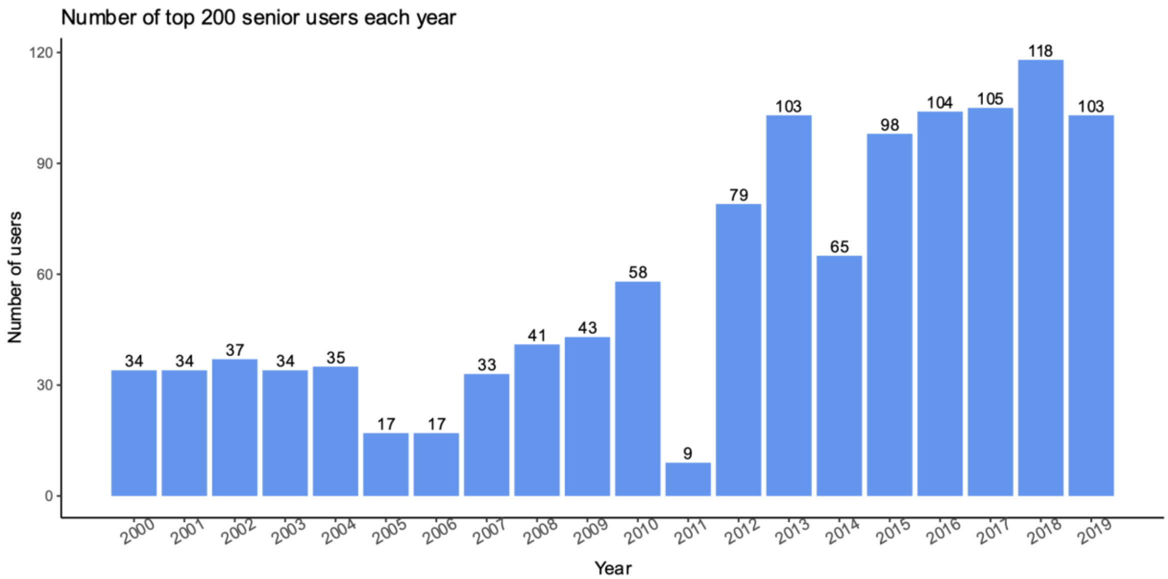

Figure 8. Bar graph showing the number of top 200 senior users in the UCRB that appear in each annual network from 2000-2019.

9. Conclusion

In this project we looked at structure call analysis data of user water rights in the UCRB to study the evolution of user interactions and water shortages. While the UCRB user interactions are primarily institutionally driven with preference to senior users, the situation is worsening due to hydrological stresses in the form of an ongoing bidecadal drought. One of the most senior and influential users in the basin is the 100-year-old Shoshone Power Plant, which regularly halts operations for repairs, and could soon be shut down permanently which, as this project reveals, would fundamentally reshape the structure of user networks in the UCRB.

This project focuses on user interactions and hydrological activity between 2000-2019 but hydrological trends often occur over much larger time scales. Future work should look at UCRB system behavior in the twentieth century as well, and check if outliers in 2000-2019 remain outliers over a century as well. Additionally, this project looks at user interaction dynamics on an annual time scale but a finer resolution, say monthly, could reveal more interesting trends in calls put out by senior users.

Furthermore, this project demonstrates that the use of network visualizations to study dynamics in hydrologically stressed river basins can reveal trends and insights that cannot be captured by an oversimplified basin model. User networks allow the study of individual interactions between two given users and can be a valuable tool for educating stakeholders in the UCRB about the uncertainty and impacts of future stressors like climate change, long-term hydrological stresses, and massive infrastructure changes.

10. Acknowledgements

I would like to first and foremost thank Prof. Patrick Reed and Dr. Antonia Hadjimichael of the Reed Research Group at Cornell University for introducing me to this wonderful project, and for being the most supportive and brilliant mentors I could have asked for. Further, I would also like to thank Prof. Scott Steinschneider for his valuable insights and advice over the course of this project. Lastly, I would like to thank the Systems Engineering Department at Cornell University for offering me the flexibility to tailor the M.Eng project to meet my future career goals.

References

Hadjimichael, A., Quinn, J., Reed, P., 2020a. Advancing Diagnostic Model Evaluation to Better Understand Water Shortage Mechanisms in Institutionally Complex River Basins. Water Resources Research 56, e2020WR028079. https://doi.org/10.1029/2020WR028079

Hadjimichael, A., Quinn, J., Wilson, E., Reed, P., Basdekas, L., Yates, D., Garrison, M., 2020b. Defining Robustness, Vulnerabilities, and Consequential Scenarios for Diverse Stakeholder Interests in Institutionally Complex River Basins. Earth's Future 8, e2020EF001503. https://doi.org/10.1029/2020EF001503

OWDI Drought. (n.d.). Retrieved December 07, 2020, from https://www.doi.gov/water/owdi.cr.drought/en/

Quinn, J.D., Hadjimichael, A., Reed, P.M., Steinschneider, S., 2020. Can exploratory modeling of water scarcity vulnerabilities and robustness be scenario neutral? Earth's Future e2020EF001650. https://doi.org/10.1029/2020EF001650