Examining the Hydrologic Properties of the Missouri River Basin

https://github.com/cwatson1013/Hydrologic_Data_Analysis_Final_Proj

Rachel Bash, Keqi He, Caroline Watson, and Haoyu Zhang

<Arrow brackets are used for annotating the RMarkdown files. Text within these brackets should not appear in the final version of the PDF document>

<**General Guidelines**> <1. Write in scientific style> <2. Global options for R chunks should be set so that only relevant output is displayed> <3. Make sure your final knitted PDF looks professional. Format tables appropriately, size figures appropriately, make sure bulleted and numbered lists appear as such, avoid awkwardly placed page breaks, etc.>

1 Rationale and Research Questions

The Missouri River is the largest river in North America (2,540 miles) and has the second largest watershed (529,000 mi^2/339 acres, U.S.-Canada). Its watershed covers portions of ten states, which account for approximately one-sixth of the continental United States, as well as a small part of Canada. The headwater is located in the Bitterroot Mountains River of northwestern Wyoming and southwestern Montana. Demands for managing the river for the benefit of human livelihood has resulted in drastic modification in the river and the floodplains. Numerous reservoirs and dams have been constructed, of which six major dams were built on the mainstream. Now, the river is used intensively in multiple ways, including municipal, agricultural, hydropower, recreation, flood control etc.

Within the 328 million acres of the basin's total area in the United States, 95% is related to agricultural uses, while the rest dedicated for recreation, fish and wildlife, and urban. More than half of the total is pasture and range grassland primarily for grazing, and cropland consists of almost 104 million acres, which is 32% of the whole basin. Irrigated land comprises 7.4 million acres, and 6.9 million acres are intensively cropped. Water bodies, on the other hand, cover 3.9 million acres. In spite of the low proportion of water areas (1.2%), they are the pivotal foundation for agricultural or other usages, and thus critical to the whole region's economy.

Along with the agricultural, urban, and industrial development in the region is nutrient loading and enrichment in water bodies, especially for nitrogen (N) and phosphorus (P). Agricultural input through fertilizer is the predominant anthropogenic source for nutrient in water bodies in the whole basin. Regardless of the major anthropogenic source, nutrient enrichment is considered nationally as one of the leading factors for water quality impairment. According to USEPA 303(d) lists, more than 160 stream reaches, lakes, or reservoirs were reported by USEPA to suffer nutrient-related impairment in 2006.

In addition to change in nutrient concentration, discharge appears to be highly variable in the basin, and both severe drought and flooding events occurred in the basin in the past. For example, in the spring and summer of 2011, an unprecedented flooding event caused over $2 billion damage FEMA disaster declaration was made in all states along the Missouri River. Subsequently, in 2012, a drought even struck the Central Great Plain, including the basin, and inflicted at least $12 billion of loss before July, 2012. Recently, another flooding event occurred in the spring of 2019.

Given all the background information above, we would like to know the current state of Missouri River and its tributaries, with a focus on the changing patterns in discharge and nutrient levels. Since regions along the downstream are more likely to be impaired by nutrient loading accumulated from upstream, in this project study sites were concentrated in the southeast of the whole basin (Figure 1). We are interested in how the dramatic change in discharge (i.e. water quantity) could potentially interact with nutrient enrichment (i.e. water quality). Also, we examined a few specific flooding events, during which changes in both water quality and quantity were well recorded, so that we could make concrete inference on the interplay between quantity and quality. Finally, based on the pattern in the past and the best model we could fit, we attempted to predict the likely future conditions and trends in the Missouri River Basin. Our research questions and acccompanying hypotheses are below:

1. How have changes in discharge (i.e. water quantity) interacted with nutrient enrichment (i.e. water quality) in the Missouri River Basin?
    a) Nutrient levels have increased over time
    b) Discharge has become more variable over time
    c) Nutrient levels increase with discharge
2. What effects do specific flood and drought events have on the water quality and quantity of rivers in the Missouri River Basin areas of interest?
    a) Rivers will exhibit a flushing behavior due to the land use and type of flow during storms
    b) Discharge will decrease during drought, causing nutrient levels to also decrease due to decreased overland flow.
3. What factors contribute to the variability of total nitrogen in the rivers?
    a) Land use, year, discharge, phosphorus, and HUC region will contribute to the variability of total nitrogen across sites
4. Given past and current data, what can we predict about the future state of water in the Missouri River Basin?
    a) Total flow in the Missouri River Basin is decreasing (non-stationary) over time
    b) The future situation of the river basin will see the continuation of current trends of decreasing overall volume of flow.

2 Dataset Information

The data we are analyzing comes from the United States Geological Survey (USGS) database called the National Water Information System interface, or NWIS. We pulled data from the interface using the R package `dataRetrieval`. Because we are interested in the lower Missouri River basin, we pulled sites from each HUC4 subbasin from 1020 to 1030. We chose these subbasins because they had a variety of tributaries that all flowed into the Missouri River, representing a variety of river sizes and lengths. We filtered these subbasin queries to only show us sites that contained discharge, nitrogen, and phosphorus data. Once we found the sites with all of this data, we chose 2 sites from each HUC sub basin for a total of 22. We chose the two sites from each HUC sub basin by comparing the periods of records for each of our chosen variables and finding the sites with the longest periods of records. We chose to look at two sites per HUC region (for a total of 22) in order to maintain a digestable scope. We retrieved data on water quantity, water quality (N, P concentrations), pH, and coliform concentrations (Table 1).

Only seven sites within our HUC subbasin boundary contained any high frequency nitrogen data. Therefore, we also looked at these 7 sites in order to do analyses and answer our research question about flooding.

We have three main datasets:

- The daily values dataset containing discharge, nitrogen, and phosphorus data for 22 sites.
- The water quality dataset containing nitrogen, phosphorus, and pH for 22 sites and total coliform for 6 sites.
- The high freqency dataset containing high frequency data for nitrogen and discharge for 7 sites.

<Add a table that summarizes your data structure (variables, units, ranges and/or central tendencies, data source if multiple are used, etc.). This table can be made in markdown text or inserted as a `kable` function in an R chunk. If the latter, do not include the code used to generate your table.>

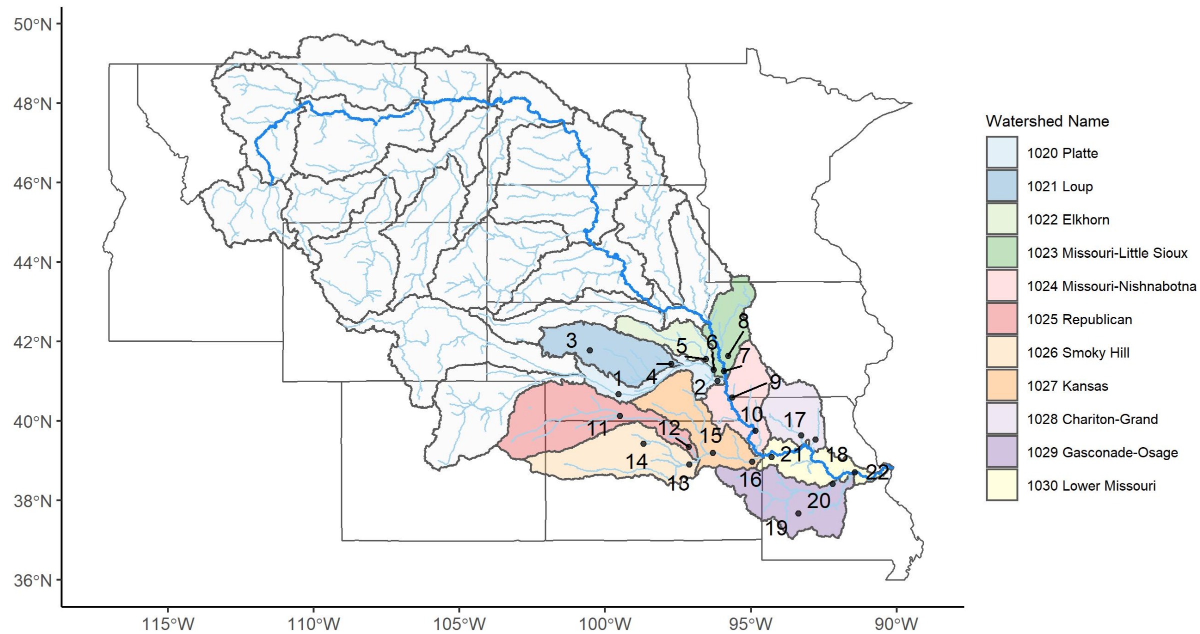

Figure 1: Map of USGS sites used for long term analysis.

3 Exploratory Analysis

<Insert exploratory visualizations of your dataset. This may include, but is not limited to, graphs illustrating the distributions of variables of interest and/or maps of the spatial context of your dataset. Format your R chunks so that graphs are displayed but code is not displayed. Accompany these graphs with text sections that describe the visualizations and provide context for further analyses.>

<Each figure should be accompanied by a caption, and each figure should be referenced within the text>

4 Analysis

<Insert visualizations and text describing your main analyses. Format your R chunks so that graphs are displayed but code and other output is not displayed. Instead, describe the results of any statistical tests in the main text (e.g., "Variable x was significantly different among y groups (ANOVA; df = 300, F = 5.55, $p < 0.0001$)"). Each paragraph, accompanied by one or more visualizations, should describe the major findings and how they relate to the question and hypotheses. Divide this section into subsections, one for each research question.>

<Each figure should be accompanied by a caption, and each figure should be referenced within the text>

4.1 Question 1: How have changes in discharge interacted with nutrient enrichment in the Missouri River Basin?

4.2 Question 2: What effects do specific flood and drought events have on the water quality and quantity of rivers in the Missouri River Basin areas of interest?

4.3 Question 3: What factors contribute to the variability of total nitrogen in the rivers?

4.4 Question 4: Given past and current data, what can we predict about the future state of water in the Missouri River Basin?

5 Summary and Conclusions

<Summarize your major findings from your analyses in a few paragraphs. What conclusions do you draw from your findings? Relate your findings back to the original research questions and rationale.>

6 References