

Proposal and Analysis Plan

Gaby Garcia, Walker Grimshaw, Tristen Townsend, Yixin Wen

October 25, 2019

What questions will your team address?

1. To what degree does climate change affect discharge in Alaskan streams and rivers?

What hypotheses will your team address?

- Hypotheses should relate directly to your questions. Each numbered hypothesis should match up with the corresponding numbered question.
- There may be multiple working hypotheses for a single question. If this is the case, note each hypothesis as 1a, 1b. etc.

1a. Streams at varying latitudes will have different responses to changing climate with monotonic trends.

1b. If streams are in areas that experience greater mean air temperature increases during spring seasons, then they will have greater discharge.

1c. The frequency of high discharge events will have increased through time due to climate change.

What dataset(s) will your team analyze?

- USGS Alaska discharge data
- Alaska site locations (spatial)
- NOAA Alaska mean air temperature data

How will you set up and manage your project repository?

We will have one master repository and ensure proper communication to avoid merge conflicts.

Create a table of variables you will analyze.

- Column 1: Variable
- Column 2: Units (if known)
- Column 3: Dependent (response) or independent (predictor) variable?
- Column 4: To which hypothesis(es) does this variable pertain?

Note: You may not know all of the individual variables you plan to analyze at this point. It is sufficient to describe what type of variable you anticipate using (e.g., land cover) and decide on specifics later

Variable	Units (if known)	Type of Variable	Hypothesis
Discharge	Cubic feet per second	Response	1a, 1b, 1c
Site Number	Latitude/Longitude	Predictor	1a
Date Time	Year/Month/Day	Predictor	1c
Date of First Snowmelt	Year/Month/Day	Predictor	1a, 1b, 1c
Air Temperature	Celsius	Predictor	1b
Precipitation	Millimeters	Predictor	1a, 1c

Variable	Units (if known)	Type of Variable	Hypothesis
HUC 8 Watershed Size	Square Meters	Predictor	1a, 1b
Permafrost Melt	Qualitative	Predictor	1b
Glacial Coverage/Melting	Qualitative	Predictor	1b

What tasks will your team conduct?

Data acquisition

Discharge data for the streams will be obtained from the USGS using the dataRetrieval package. The site locations for these same streams will also be obtained from the USGS using the dataRetrieval package. The historic air temperature and precipitation data will be retrieved from NOAA Climate Data Online (CDO) Database at <https://www.ncdc.noaa.gov/cdo-web/>. HUC8 watershed size data will be retrieved from <ftp://rockyftp.cr.usgs.gov/vdelivery/Datasets/Staged/Hydrography/WBD/National/GDB/>. Time permitting, we will retrieve qualitative data related to permafrost melting and glacial coverage/melting from scientific literature.

Data exploration

Goals for data exploration:

- Determine which Alaskan sites to choose
- Use the function ‘whatNWISdata’ to access metadata about our chosen sites, including what variables have been monitored at that site, start and end dates, and the number of samples for a given variable
- Determine which 10 USGS sites feature the most data points
- Looking at data distributions
- Any transformations required for non-normally distributed data?
- Do we have consistent data intervals for time series analysis?
- Create correlation plots

Data wrangling

- Remove NAs from quantitative datasets
- Join site data with latitude/longitude
- Fill in any missing data to perform time series
- Join all data sets so measurements are present for all sites

Data analysis and visualization

- Use high frequency data of discharge to do monotonic trend analysis and also use ARMA model to predict the discharge in the future
- Plot the decomposed dataset and the prediction of discharge
- Plot ggplot of discharge vs mean air temperature
- Time series analysis of temperature, precipitation, date of snowmelt and how they relate with discharge
- Map of ten sites