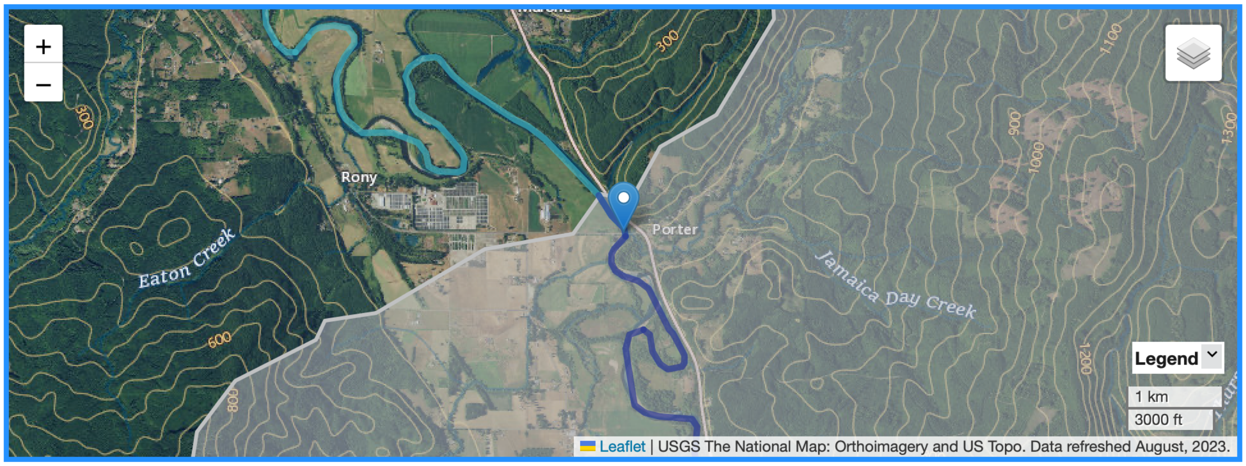
Project

Using this link, we collected the suspended sediment concentration and discharge observations at USGS station 12031000 CHEHALIS RIVER AT PORTER, WA. The following figure shows the location of the USGS station.



Let's see what these variables are!

Suspended sediment concentration (SSC) refers to the amount of solid particles that are suspended in water. These particles can include silt, clay, sand, and other materials. The concentration is typically expressed in terms of mass or volume of sediment per unit volume of water. Common units for SSC include milligrams per liter (mg/L) or parts per million (ppm). River discharge refers to the volume of water flowing through a river channel over a specific period of time. It is typically measured in cubic meters per second (m^3/s) or cubic feet per second (cfs). River discharge is a fundamental and important hydrological parameter as it reflects the movement of water within a river system.

See the dataset in the attached excel file. This data is collected at daily time scale from October 01, 1961, to September 29, 1971. Here is description of the datasets.

Description
Suspended Sediment Concentration (SSC), milligrams per liter (mg/L)
Discharge, cubic feet per second (cfs)

Müller and Förstner (1968) reported that the discharge–sediment relationship of a basin can be expressed by the empirical power function $\mathrm{SSC} = \alpha \times \mathrm{Q}^{\beta}$ (SSC, kg/m^3; Q, m^3/s), and α and β are parameters. (1 mg/L = 1×10^{-3} kg/m³ and 1cfs ≈ 0.0283168m³/s)

As we know this equation can be re-written as:

$$\mathrm{SSC}(\mathrm{Q_i} \mid \alpha, \beta) = \alpha \times \mathrm{Q_i}^{\beta}$$

If we assume there is Gaussian error associated with the observed SSC values, the probability for any data point under this model is expressed as:

$$P(Q_i\mathrm{SSC}_i \mid \alpha, \beta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-[\mathrm{SSC}_i - \mathrm{SSC}(Q_i \mid \alpha, \beta)]^2}{2\sigma^2}\right]$$

Here σ represents an unknown measurement error, which we'll treat as an uncertain parameter.

Multiplying these for all data points i gives the likelihood:

$$P(\{Q_i\}\{\mathrm{SSC}_i\} \mid \alpha, \beta, \sigma) \propto (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i-1}^{N}[\mathrm{SSC}_i - \mathrm{SSC}(Q_i \mid \alpha, \beta)]^2\right]$$

a) Use the flat prior assumption for β and visualize the ensemble of curves.
b) Let's assume the following is the best choice of prior for the model parameters α and β.

$$P(\alpha, \beta) \propto (1 + \beta^2)^{-3/2}$$

And let's assume that σ follows the Jeffreys prior. Use the MCMC approach to fit a curve of best-fit along with the 2-σ uncertainty region.
c) Use KDE to visualize the resulting traces marginalized over the uncertain parameter σ.
d) Use Least Squares method to find the model parameters α and β. Visualize the result similar to Figure 2.4 in lecture 2.
e) Compare the results from Bayesian approach (part b) with frequentist approach (part d).

Please ensure that you include your Python script, along with all figures and results, in your report, otherwise your project will not be graded. (***10 points***)

References

Müller, G., Förstner, U., 1968. General relationship between suspended sediment concentration and water discharge in the alpenrhein and some other rivers. Nature. https://doi.org/10.1038/217244a0