**Impacts of Elevation and River on partial CO2 concentrations**

1. Introduction

Streams play an important role in the emissions of greenhouse gases (GHGs) such as CO2. To better characterize this role, data were collected on three rivers on the Tibetan Plateau: the Yangtze (YZ), the Yarlung-Tsangpo (YT), and the Yellow (YL). Sampling was performed in 2014-2015, collecting partial CO2 pressure (pCO2) at each river site in μatm (Qu et al., 2017) as well as site elevation (in meters) and river name. Since both intra- and inter-river elevations variations may influence atmospheric pressure on dissolved gases, and river characteristics can impact CO2 emissions, we investigated how elevation impacts pCO2 and if those impacts vary across rivers.

1. Statistical Procedures Used

All analyses were conducted using R (R Core Team, 2024). The pCO2 observations for the three rivers were visualized with enhanced strip charts from the catstats2 package (Greenwood, 2024) in Figure 1. Summary statistics for pCO2 and elevation by river are provided in Table 1 based on the modelsummary package (Arel-Bundock, 2022). The estimated means pCO2 of YZ and YL were similar at 1054.25 and 1083.364 μatm, respectively, while the mean of YZ was noticeably lower at 595.4 μatm. However, pCO2 observations, particularly in the YT and YL rivers, had a wide range of pCO2 values, from 300 to more than 1700 μatm. The number of sites varied by river with nYZ = 4, nYT = 15, and nYL = 11, respectively, providing unbalanced design with respect to the rivers and a total sample size of n = 30. A scatterplot was utilized to examine the relationship between pCO2 and elevation by river with both linear and nonparametric smoothing lines (Figure 2, ggplot2 package, Wickham, 2016). From Figure 2, a negative relationship between pCO2 pressure and elevation was evident but the slope of the relationship for YT was more negative than for YZ or YL. There may be some curvature in the relationship for YT and increasing variance in the YT observations as elevation increases or what might be a possible outlier.

Linear models were used to model pCO2 using the lm function. After fitting a preliminary model of μ{pCO2|River&Elevation} ~ River\*Elevation, a suite of diagnostic plots were generated using the ggResidpanel package (Goode et al., 2024, Figure 3). In the Residuals vs. Fitted plot (Fig. 3, upper left), the spread of the residuals increases as fitted values increase, which raises concerns about the assumption of constant variance and a slight curve may suggest a violation of the linearity assumption. From the QQ-plot (Fig. 3, upper right), there is evidence of a violation of the normality assumption with a clear right-skewed distribution of the residuals. To address the potential violations of linearity, constant variance, and normality assumptions, the response variable pCO2 was log transformed (natural log) and the model was refit and new diagnostic plots were produced (Figure 4). After log transformation of the response variable, there was little or no evidence against the assumptions of constant variance (Residuals vs. Fitted does not show changing spread in the residuals as a function of fitted values), linearity is less clearly violated (limited curvature in Residuals vs. Fitted), and normality of residuals (QQ-plot of residuals does not show clear deviations from normality of the residuals). The partial residuals in the effects plots (Fox and Weisberg, 2018, Figure 5) for this model suggest that there is a weak negative relationship between the YZ river and elevation, there is a strong negative relationship between the YT river and elevation, and a moderate negative relationship between the YL river and elevation. The YL river residuals also display potentially missed curvature that could be due to an influential point with leverage for the negative relationship.

To address the question of interest, it was necessary to determine if a River by Elevation interaction term was needed in the model. To assess this, a Type II F-test (car package, Fox and Weisberg, 2019) was used on the μ{logpCO2 } ~ River \* Elevation model. Weak evidence was found against the null hypothesis of no interaction between River and Elevation on the log-pCO2 was found (F(2,24) = 1.0759, p-value = 0.357), so the interaction term was dropped from the model.

Diagnostic plots of the additive model with River and Elevation did not indicate further issues with the normality or constant variance assumptions (Figure 6). However, one observation had a Cook’s distance value of 0.65 in the Residuals vs Leverage Plot (Figure 6, lower right), qualifying as potentially influential observation. Examining that YL observation more closely, this point had both the highest elevation (4091 m) and pCO2 (1771 μatm) measurement in the data set. This was contrary to the generally observed trend of decreasing pCO2 values with increasing elevation (see Figure 2). However, without information about this observation that explicitly showed it to be in error, we did not exclude it from the data analysis.

Due to the sampling of sites with multiple observations in each river and some closer or further apart geographically, there might be an issue with a violation of the independence assumption as some sites might be more similar than others even after accounting for river and elevation information. If river is not included in the model, the repeated measures on a river would create a clear violation of the independence assumption. Because samples were taken sequentially in time in the study years, there could be an additional violation of independence by some observations being taken closer in time and others taken later. The results may be biased because the sampling locations in the rivers were not randomly selected and easy-to-access sites may have been selected and they might have systematically higher or lower pCO2 on average than the population of sites.

1. Summary of Statistical Findings

The final estimated model was: , where River is a three-level categorical variable represented by the indicator variables (which takes on a value of 1 if the River is YT, and 0 if not), and (which takes on a value of 1 if the River is YL, and 0 if not). This means that the third level of River, YZ, is treated as the reference.

A Type II F-test was generated to assess including River in the model. Accounting for elevation, there is very strong evidence against the null hypothesis of no difference in the true mean log.pCO2 for all three rivers (F(2, 26)= 6.75, p-value = 0.0044), so we would conclude that there is some difference in mean log.pCO2 across the rivers. Accounting for the river, there is moderate evidence against the null hypothesis that elevation is not linearly related to log.pCO2 pressure (2-sided t-test, t(26)= -2.19, p-value = 0.0381), so we would also conclude that there is a linear relationship between elevation and pressure after accounting for rivers. The model has an R-squared of 0.5567, which suggests that a model with river and elevation explains 56% of variation for logpCO2 which suggests that this model is a good fit.

For two otherwise similar locations that differ in elevation by 1 m in elevation, the median pCO2 pressure of the higher elevation location is 0.99977 times as much as the lower elevation, controlling for river (95 % CI: 0.99956 to 0.99998). Controlling for elevation, the median pCO2 pressure in river YT is 0.54 times as much as YZ (95 % CI: 0.368 to 0.795) and the median pCO2 pressure in river YL is 0.82 times as much as YZ river (95% CI: 0.53 to 1.27). To visualize the impact of both River and Elevation on the response log.pCO2, effects plots for the additive model with partial residuals are displayed in Figure 7.

1. Scope of Inference

There is no indication given in the study design that the rivers, river sites, or sampling times on the rivers were randomly selected, so it is not possible to infer the results of this study beyond the river observation points and times included in the study. These results only apply to the years of data collection of 2014 and 2015. Likewise, it was not possible to randomly assign Elevation or River to study sites, therefore we could not conclude that Elevation or River caused changes in pCO2 pressure. Keeping these limitations of the study in mind, with this analysis we were able to show that both River and Elevation impact the log-partial CO2 pressure in those rivers of the Tibetan Plateau considered in this study. Based on Figure 7, both river identity and elevation have similar impacts on the log.pCO2 in the model together, with estimated changes in the mean of -0.00023 across the Elevation and … for … across the …[Q4]

**References:**

[Q5 – complete remainder of references section, remember – alphabetical order and consistent citation style.]

Greenwood, M. (2024) catstats2: Upper Level Statistics for Montana State University Bobcats. R package version 0.2.

R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

**Figures:**

[Q6 – make figure 1 and insert here]

Figure 1. Enhanced stripchart of pCO2 by river.

[Q7 – make figure 2 and insert here]

Figure 2. Scatterplot of …

[Q8 – make figure 3 and insert here]

Figure 3

[Q9 - make figure 4 and insert here]

Figure 4

[Q10- make figure 5 and insert here]

Figure 5

[Q11 - make figure 6 and insert here]

Figure 6

[Q12 - make figure 7 and insert here]

Figure 7

**Tables:**

[Q13 – make table 1 and insert here]

Table 1. Table of …

[Q14] Replace this page with knitted Rmd work here by combining PDF files with this document after PDF’ing it.