

Cache Slough Complex Turbidity Analysis

Steven Cognac

2021-12-02

Research Question

In the Cache Shag Slough Complex (CSC) in the San Francisco Estuary, does turbidity (as measured in Formazin Nephelometric Units) vary with wind speed (miles per hour) and tidal stage (using gage height as a proxy).

Some turbid water in Cache Shag Slough Complex.

Introduction

Overview: In the lower San Francisco Estuary there is evidence of a decadal decrease in wind speed (Bever et al, 2017) from 1995 through 2015. This decrease in wind speed has been shown to be have corresponding effects on habitat, and specifically turbidity. I'm interested in understanding to what extent turbidity varies with wind speed and tidal stage in the upper San Francisco Estuary.

Background: Turbidity is a metric of the total suspended solids in the water column and is a measure of water clarity. In freshwater systems, low turbidity is typically good because it indicates high water clarity and a healthy ecosystem. On the reverse side, high turbidity is generally bad as it blocks sunlight from reaching aquatic plants and can smother aquatic organisms leading to plant die-offs. High turbidity can also be a vector of contamination. For example, particles, pathogens, and other contaminants (lead, mercury) can adsorb to suspended solids and increase the distribution of contaminated sediments.

Motivation: However, not all turbid waters are equal. In estuaries, a healthy amount of turbidity is an important habitat condition. Without listing the suite of ecosystem services estuaries provide, spawning and rearing habitat for forage-fish species are major components. In the San Francisco Estuary, turbidity changes can have major management implications. For example, the SF Estuary is home to the endangered Delta Smelt, which is a fish species that thrives in relatively high-turbidity water. Another example is the management of development and restoration projects along the SF Estuary. Understanding background conditions is important for setting water quality thresholds for work in and along the estuary. The goal of this project is to understand if the relationship between turbidity, wind, and tidal stage hold true in the CSC Complex.

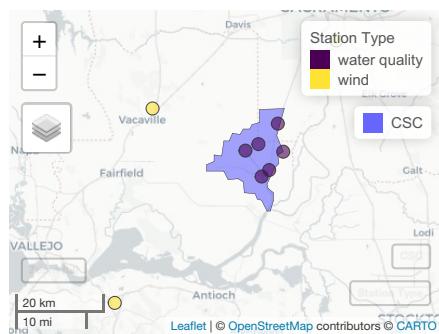


Figure 1: Map of wind and water station locations in relation to CSC

Hypothesis

My null hypotheses (H_0) is that, in the CSC Complex, there is no relationship between hourly or daily turbidity compared to wind speed and gage height.

My alternative hypothesis (H_a) is that, in the CSC Complex, there is a relationship between hourly or daily turbidity compared to wind speed and gage height.

Here is my regression equation:

$$\text{Turbidity}_i = B_0 + B_1 \text{WindSpeed}_i + B_2 \text{GageHeight} + u_i$$

Data Sources and Cleaning

The CSC includes portions of the Yolo Bypass, a network of major and minor sloughs, agricultural land, and the Sacramento River/Deepwater Channel. Five water quality stations and three wind stations were used for the analysis. This analysis expands upon the Bever et al. analysis by looking at the 2015 - 2021 time period. All stations were chosen based on location and data availability.

Turbidity & Gage Height Data Turbidity and gage height data downloaded from the United States Geological Survey (USGS) via the “dataRetrieval” R library. The library pulls hydrologic data from the National Water System Information System (NWIS) via the National Water Dashboard. Data was generally available in 15-minute time intervals. However, turbidity data is typically messy due to sensor fouling and hence contains lots of data gaps. A fair amount of tidying was required.

Wind Data US Local Climatological Data (LCD) downloaded from NOAA NCEI online portal map tool. Data available as hourly summaries from airport and other prominent weather stations. Data follows ISO 19115-2 Metadata standards for describing geographic information.

Data Cleaning Data for the three wind stations and six water quality stations were averaged together into hourly and daily averages, respectively. Those tables were merged together.

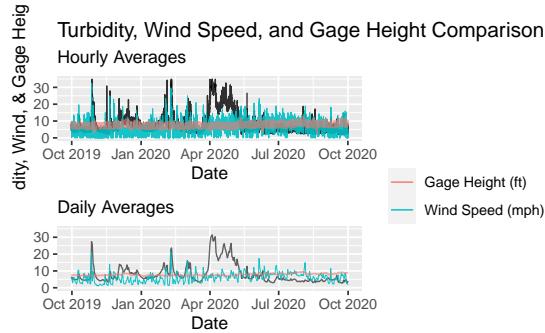


Figure 2: Correlation plots of turbidity compared to wind speed and gage height

Statistical Analysis Plan

To assess whether turbidity varies with wind speed or gage height, I ran two linear regressions. One with hourly data and another with daily data. Analysis was completed in RStudio.

The general steps of this analysis plan include the following: 1. Identify question 2. Select independent and dependent variables(based on Bever et al, 2017) 3. Data download, clean, and merge 4. Visualize relationships 5. Conduct regression 6. Test OLS assumptions 7. Interpret Results 8. Conclusions and Future Research

Visualize Relationships

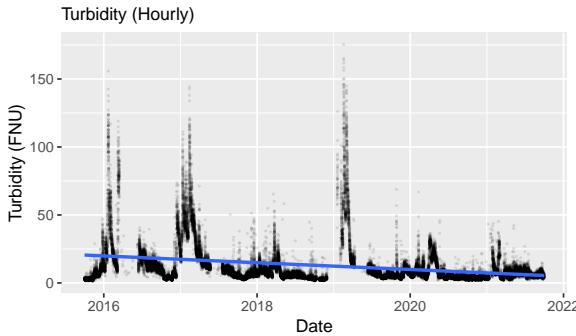


Figure 3: Plot of turbidity over time

Relationship plots of turbidity and the two independent variables (wind speed and gage height) are included below. From 2015 to 2021, there is a decrease, on average, in hourly and daily turbidity levels. We can see gage height has a strong positive correlation with turbidity for both hourly and daily data. We also can see clustering of gage height data, especially in the hourly data. We also observe a weak positive correlation between wind speed and turbidity.

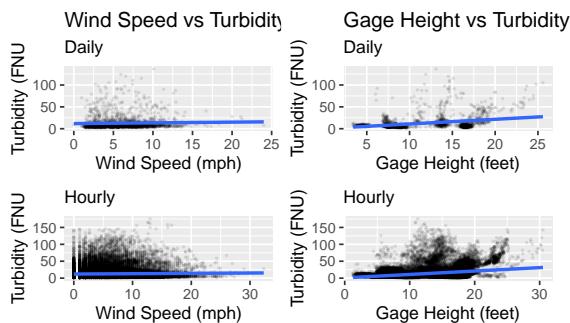
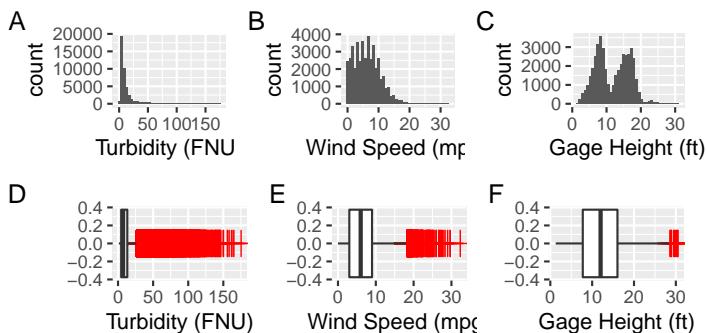


Figure 4: Correlation plots of turbidity compared to wind speed and gage height

Histogram and Boxplots



Source: NOAA NCEI and USGS

Figure 5: Histogram and box plots of independent and dependent variables

Regression Analysis & Interpretation

Hourly Data

Daily Data

Observations	43815			
Dependent variable	turbidity_fnu			
Type	OLS linear regression			
F(2,43812)	2206.85292			
R ²	0.09152			
Adj. R ²	0.09148			
	Est.	S.E.	t val.	p
(Intercept)	-0.20424	0.22951	-0.88990	0.37352
HourlyWindSpeed	0.15303	0.01790	8.54918	0.00000
gage_height	1.01404	0.01532	66.19358	0.00000
Standard errors: OLS				
Observations	1935			
Dependent variable	turbidity_fnu			
Type	OLS linear regression			
F(2,1932)	108.70531			
R ²	0.10115			
Adj. R ²	0.10022			
	Est.	S.E.	t val.	p
(Intercept)	-1.45005	1.18921	-1.21934	0.22286
DailyWindSpeed	0.23576	0.11070	2.12971	0.03332
gage_height	1.07881	0.07354	14.66993	0.00000
Standard errors: OLS				

Intercept (B_0): Tells us the predicted turbidity is -1.45 FNU when wind speed and gage height is zero. Logically this does not make sense as the turbidity scale typically ranges from 0 to ~500.

Coefficient on wind speed (B_1): There's a predicted increase of 0.15-0.23 turbidity units (FNU) for every mile-per-hour increase in wind speed, holding the gage height fixed.

Coefficient on gage height (B_2): There's a predicted increase of 1.01-1.07 turbidity units (FNU) for every foot increase in gage height, holding the wind speed fixed.

Adjusted R^2 : About 9-10% of the variation in turbidity can be explained by wind speed and gage height. Though turbidity data was filtered to remove values over 300 FNU, there are still many extreme turbidity events that approach this outlier threshold. Adjusted R^2 is very sensitive to outliers so it makes sense we have a low value.

P-Value: For both the hourly and daily regressions we reject the null hypothesis that wind speed and gage height has an effect of 0 on turbidity. With hourly data, $p-value < 0.000$. For the daily data, we get $p-value < 0.000$ for gage height and $p-value < 0.03$ for wind speed. We can say there is a **statistically significant correlation** for both variables.

Testing OLS Assumptions

We can assume assumption 1 (linear relationship) and 2 (X variables are exogenous $E[u | X] = 0$) of OLS hold up. However, the linear relationship is weak. Assumption 3 requires our X variables to have variation. Based on

the relationship plots above, we can see this holds true.

Assumption 4 of OLS is more difficult to prove. It requires population disturbances u_i to be independently and identically distributed as normal random variables with mean zero and variance σ^2 . To test assumption 4 I will generate residuals from the daily regression model and plot them.

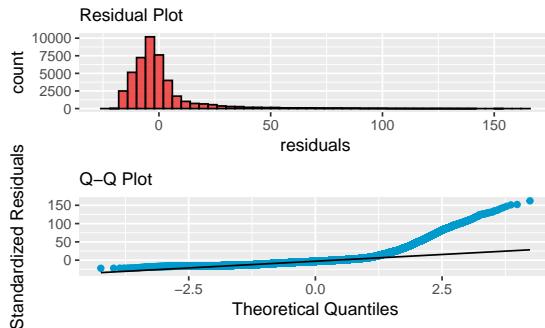


Figure 6: Histogram and box plots of independent and dependent variables

The residuals are not normally distributed as there is a long right tail with high outliers. This indicates that the regression analysis is under-predicting average hourly turbidity levels. The Q-Q Plot indicates the distribution is generally normally distributed for values up to the positive 1 theoretical quantile. This suggests that the model is not normally distributed for extreme high values.

Conclusions and Future Analysis

This analysis indicates that wind speed and tidal stage (gage height) correlate with higher turbidity levels at both the hourly and daily intervals. However, based on the adjusted R^2 value, omitted variable bias is likely an issue. Two critical omitted variables we could add to the model are upstream sediment loading, wind-wave resuspension, and an interaction between some of the variables.

Based on the residual plot, assumption 4 of OLS may not be satisfied and OLS may not be the estimator with the lowest variance. Also, because OLS is sensitive to outliers, I removed all values over 300 FNU in the turbidity dataset. This helped with the analysis, but still limits the use of OLS with the dataset. An alternative to removing values is to winsorize the dataset such that all extreme values are capped at a specified percentile range. Another option could be to choose a different regression technique. Partial least squares (PLS) is helpful when comparing multiple continuous dependent variables or when the independent variables are highly correlated. If wind-wave resuspension was added in a future analysis, PLS may prove to be a better regression method.

Break time variable up into categorical years + month columns so you could compare correlation on turbidity between months of the year. Wind speed in the Delta is relatively low during the late fall and winter months and then increases during the spring and summer (Schoellhamer et al. 2016).

References

Bever, Aaron J., Michael L. MacWilliams, and David K. Fullerton. 2018. “Influence of an Observed Decadal Decline in Wind Speed on Turbidity in the San Francisco Estuary.” *Estuaries and Coasts* 41 (7): 1943–67. <https://doi.org/10.1007/s12237-018-0403-x>.

Schoellhamer, D.H., G.G. Shellenbarger, M.A. Downing-Kunz, and A.J. Manning. 2016. Review of Suspended sediment in lower south Bay relevant to light attenuation and phytoplankton blooms. In Lower South Bay Nutrient Synthesis. San Francisco Estuary Institute & Aquatic Science Center, pp. 23–56.