Reviewer 1: Thank you for recognizing the importance of this study and for taking the time to review it. Your input is invaluable, and your comments will help improve our manuscript. Below, we provide our responses (in blue font) following each reviewer comment (in black font).

General comments:

This paper looks at how major precipitation events influence dissolved oxygen changes—an integrative measure of estuarine resistance to disturbance—in the NERRS estuaries across the US. There are not many comparative estuarine studies that leverage the network of NERRS data, so this paper fills an important research gap in that respect. The synthesis of the input dataset is a feat in and of itself. Additional strengths of the work are the graphical synthesis of complex results (Figure 5) and a detailed discussion.

Thank you for the encouraging remarks.

However, the paper also has major weaknesses that deserve further attention in a revision. The biggest weakness is the attempt to infer drivers of resistance across a very small number of estuaries through bivariate correlations. As the authors acknowledge in lines 395-397, resistance is associated with myriad specific factors, and these factors interact. Potential multicollinearity within the dataset between percent urbanization and other factors such as watershed contributing area, flushing time, tidal influence, etc. makes it challenging to infer how urbanization influences resistance on the basis of a bivariate correlation. Ideally, one would address collinearities in potential predictors and construct a generalized linear model to account for covariates and isolate contributions of factors like urbanization to resistance in order to draw inferences. However, with this small dataset, this is probably not an option.

We agree that multicollinearity between variables can be problematic for statistical analysis, particularly in the case of some multivariate analyses. Because of that, we previously examined correlations between all variables and provided a correlation matrix in the supplemental material (Fig. S8). Most variables were largely uncorrelated. To clarify this, in the revised manuscript we will add a statement to the methods that multicollinearity was assessed with reference to the supplemental figure.

Also, we note that some multicollinearity is typical in environmental datasets and that collinear variables do not pose an issue for univariate analyses. However, they complicate the construction and interpretation of some multivariate models. For example, temperature and salinity are often highly correlated, and yet they both have important independent impacts on estuarine dynamics (e.g., through rate kinetics/growth vs. stress tolerance mechanisms). In the multivariate approach proposed by this reviewer, one variable should be omitted from the model because including multiple collinear predictors reduces model performance. In cases where two collinear variables show a similar relationship, we cannot infer if one or both of the variables might be a mechanistic driver.

In univariate models, both can be kept, allowing readers to view results regarding the predictor most relevant for their system. For example, relationships between salinity and resistance may be more interesting to managers of more exposed coastal estuaries with large tidal influences, while relationships with temperature may be more interesting to those working in estuaries cut by deep rivers. We therefore think it is valuable to present independent analyses for all possible relevant parameters even if two variables are correlated (which occurred infrequently in this dataset).

Additionally, as described in lines 139-146 and 239-242, National Estuarine Research Reserve (NERR) water quality data were collected at different intervals (i.e., 15-min time series versus monthly), which means that we would not be able to include all predictive variables into the same model without collapsing them to the same resolution/time interval. This would result in a loss of substantial amounts of information for sensor-based (15-min interval) measurements. In other words, statistical approaches like the Generalized Linear Model or the Principal Component Analysis for this particular dataset cannot investigate physicochemical variables like salinity, water column depth, turbidity, temperature, nutrients, chlorophyll-*a*, and LULC collectively without losing information.

Also, the reviewer's example of collinearity between urbanization and the watershed contributing area is not clear to us. For correlations, we reduced the watershed area by using a defined zone (10-km radius) around monitoring stations. This allowed us to keep realistic the area of the watershed that could contribute to the water quality at a monitoring site within short time scales, and made the analysis more comparable area-wise. To clarify the exact breakdown of LULC within each watershed at each estuary, we will include an additional supplemental table in the revised manuscript.

I instead recommend a few options to address this issue:

1. Quantitative approach: Fully characterize the set of potential drivers of resistance hypothesized to be important and how they vary across estuaries (and potentially across sites, being careful to avoid pseudoreplication—inclusion of sites as multiple datapoints when they all correspond to one independent variable like percent urbanization). This should likely include variables related to estuarine circulation, like flushing time and stratification indices, and potentially watershed area.

We agree with the reviewer that estuarine circulation expressed through metrics like flushing time and stratification indices could impact estuarine processes. However, flushing time and stratification indices are not standard NERR metrics, and are not widely available in NERR metadata files or literature. To recognize the importance of estuarine circulation, in the revised manuscript we will add discussion on flushing time and stratification that were not considered in the analysis but may influence estuarine resistance to precipitation. We will also add estimated ranges for residence time, accompanied by citations, for each site to Table S1.

While we were unable to locate comparable data for circulation across all five estuaries, we note that the NERR dataset provides a fairly comprehensive assessment of key estuarine variables, as determined through a series of workshops and careful consideration among staff members. Data includes water quality variables (water temperature, depth, salinity, sp. cond., DO, pH, turbidity, chl-*a*), nutrient variables (N-species, PO_4^{-3}, other optional species), and weather variables (air temperature, precipitation, humidity, pressure, PAR, wind speed and direction). Also, NERR sample collection and processing is standardized, including the placement of sondes within the water column. The value of such a data product is that it allows direct comparisons across different estuaries. Additional site and estuary notes including average channel width and depth, generalized land-use description surrounding each site, equipment placement, equipment maintenance history, notable historical developments (i.e., site history), site-altering events, etc. are available in metadata files that can be requested at https://cdmo.baruch.sc.edu/get/landing.cfm.

Then conduct a principle components analysis followed by a Varimax rotation to identify how the predictor variables group together and how the estuaries vary with respect to each other and the predictor variables. The Varimax rotation makes each factor more interpretable.

Thank you for this suggestion. We agree that multivariate approaches, like PCA, can be powerful tools. Unfortunately, for this study, data were collected at different intervals which restricts the PCA analysis for the full range of predictive variables used in this study (see discussion above).

Yet, to demonstrate the robustness of our results, we performed the PCA analysis with and without Varimax rotation using the simultaneously-collected sensor-based measurements only (salinity, turbidity, water column depth, and temperature). This analysis is consistent with the data subset used in the linear regressions presented in Figure 3 in the manuscript. We found that the results between PCA and linear regression analysis were in agreement (please see *Figure R1* below). Particularly, with PCA when applied for all estuaries regardless of the site or salinity level (i.e., no grouping into low- and high- salinity), we found that salinity had the least influence on either PC1 or PC2, unlike temperature, depth, and turbidity loadings. In parallel, our continental-scale (i.e., all estuaries combined and regardless of salinity) linear regression analysis between resistance and the four variables showed the existence of significant relationships (dotted green lines) between resistance with turbidity, temperature, and depth, while the relationship between resistance with salinity was not significant.

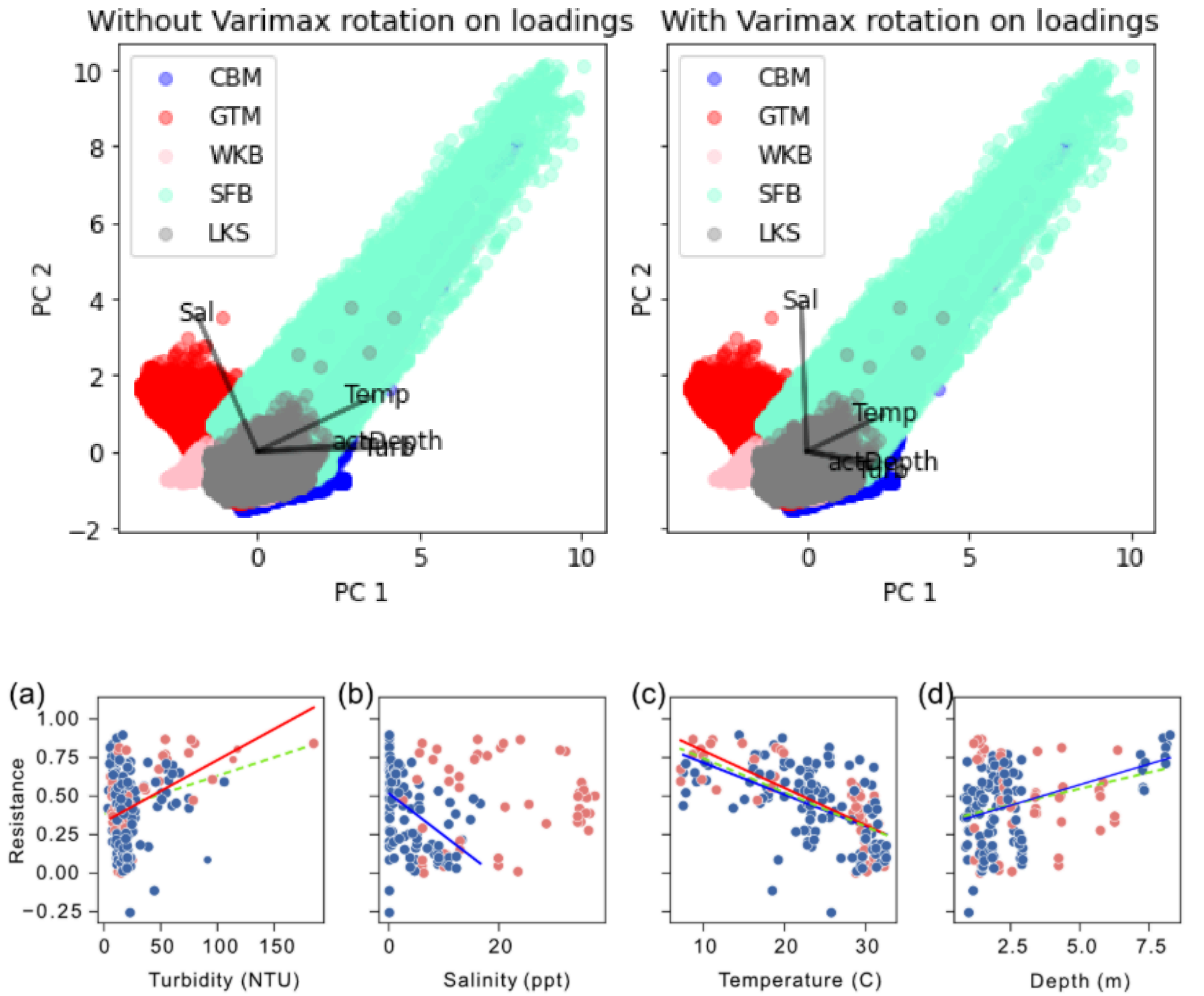


Figure R1. (top panel) Principal component analysis without and with Varimax rotation. *(bottom panel, from Figure 3 of the manuscript)* Relationships of continental-scale and salinity-based resistance with physicochemical factors. Continental-scale regressions considered all monitoring locations across all estuaries. Significant relationships ($p < 0.05$) are shown in dotted-green, solid red, and solid blue lines for continental-scale, high-salinity and low-salinity estuaries, respectively.

After considering the issues raised by different time intervals and by collinearity (please see the comment above), and by examining the results obtained with PCA versus linear regression analysis shown in *Figure R1* above, we feel that our overall statistical approach is robust and informative. We will add the PCA plots to the supplemental material.

Then, plot the factors from the Varimax rotation against resistance, or consider how resistance differs among different groups that emerge from the factor analysis.

Please see the comment above.

2. Qualitative approach: Alternatively, eliminate the cross-estuarine correlations from the analysis and take a case-study approach to elucidate how storm events impact resistance in different types of estuaries. For this approach to be successful, the case-study discussions should be tied to hypothesized detailed mechanisms of how storm events would impact resistance in different types of estuaries.

Understanding what factors are most likely involved in estuarine resistance to storms regardless of the type of estuary is important for large-scale models and for informing management decisions regarding estuarine water quality. Because the aim of the current study is to identify generalizable physicochemical and land-use factors that relate to estuarine resistance to storm events, we intend to keep cross-estuarine correlations.

Nonetheless, while our study is focused on generalizable physicochemical and land use factors that relate to estuarine resistance to storms, we recognize that many estuarine differences could lead to variation in responses to storms (please see lines: 29-31; 341-344; 373-374; 406-411; 437-447; 531-525). To further highlight how storm events impact resistance in different types of estuaries, in the revised manuscript we plan to enhance the discussion and add a conceptual-model figure as suggested by the reviewer in the comment below.

Regardless of which option is chosen, I recommend moving some of the excellent literature review in the Discussion to the Introduction and synthesizing it to present a coherent conceptual model of the different ways in which physicochemical variables and storm events are expected to interactively influence resistance. A conceptual figure would be nice as well. Without this theoretical underpinning to orient readers at the outset, the detailed and understandably messy dataset and results are challenging for readers to wade through.

Thank you for your suggestions. To aid the readership, in the revised manuscript we will present an additional figure– a conceptual model– describing the interaction between physicochemical variables, storm events, and resistance.

Another big-picture factor to consider further is that some of the estuaries may be less impacted by precipitation events than by flow events controlled by upstream reservoirs. One factor that should probably be discussed is whether some estuaries may appear more resistant to rain events because upstream reservoirs act as flood-detention buffers.

Thank you for drawing our attention to this detail. In the revised manuscript, we will expand the discussion on factors other than precipitation events that could control DO dynamics including upstream reservoirs and residence time.

Briefly, upon reviewing NERRS metadata files (https://cdmo.baruch.sc.edu/get/landing.cfm), three estuaries noted that there were upstream dams and/or other structures but that they were unlikely to influence site hydrology (see the information from NERRS metadata sheets below).

(1) At Lake Superior NERR (LSNERR), the Oliver Bridge monitoring site is reported to be located 12 miles downstream of the Fond de Luc dam. Although the site is hydrologically influenced by Lake Superior seiche and St. Louis River discharge, the potential influence of discharge-regulating dams on dissolved oxygen at this site has not been characterized.

(2) At Guana Tolomato Matanzas (GTMNERR), the hydrology of the GTM system is reported to be 'somewhat' altered by a dam across a portion of the Guana River, dikes, drainage ditches, and inland wells. However, the potential influences of upstream structure on resistance at GTM is not evident from our results (see Figure 2). Compared to, for example, SFB– a highly urbanized estuary– the resistance at GTM is lower.

(3) At San Francisco Bay (SFBNERR), the Suisun Bay which incorporates First and Second Mallard (FM and SM, respectively) monitoring sites, is equipped with a marsh salinity control gate. However, there is no mention in the metadata files that the gate was operated preceding, during, or after the rain events selected for this study (i.e., we assumed that the gate had no impact on the measurements used for this manuscript).

In general, a table and/or text providing a more detailed characterization of the estuaries (beyond land-use) is a needed addition.

To address the reviewer's suggestion, we will use any available information (i.e., NERR metadata files, literature, reports) to enhance site characterization in the revised manuscript. Specifically, we will:

1) Add columns to Table S1 which will contain:
    a) Approximate mean tidal fluctuation
    b) Tidal regime (e.g., minimally tidal, microtidal, moderately tidal)
    c) Residual current type (e.g., well-mixed, partially mixed)
    d) Tidal flow type (e.g., diurnal tidal flow; mixed, semi-diurnal tidal flow)
    e) Site characteristics (e.g., presence of upstream/downstream structures, WWTPs, submerged vegetation, dominant sediment type) and estimated residence time.
2) Add a supplemental figure showing temporal DO at each site during wet and dry years.

We also would like to draw the reviewer's attention to the supplemental material, which contains a wealth of additional information on the estuaries we investigated. In particular, Figures S1-S3, S6-S8, as well as Figure 1 and Table 1 that provide detailed information on long- and short-term precipitation patterns across broader area and within each estuary, temperature dynamics at each monitoring location, nutrient concentrations (i.e., DIN, PO_4^{-3}) and N:P at each monitoring location.

If a PCA is attempted as suggested in #1 above, it would be ideal to include variables that characterize hydrologic differences among the estuaries or across storms, such as a normalized increase in inflow (if available) and statistics about storm durations, which could be another driver of cross-estuary differences in resistance.

For the use of PCA and availability of estuarine circulation metrics (i.e., flushing time and stratification indices), please see the comments above.

Also, we note that we regressed precipitation amount (reported in Table S2) against resistance values at each estuary. The only significant relationship between precipitation and resistance was observed for Guana Tolomato Matanzas (GTM) estuary (reported in Figure S11).

Specific comments:

The abstract should more specifically state the big-picture implications of these results. What matters about them and why?

In the revised manuscript, we will expand the abstract to include more information on why it is important to understand the factors that associate with estuarine resistance to precipitation and their generalizable patterns.

Line 133-136: Is the 10-km lengthscale, then, the spatial scale expected to drain to the point of interest over the sensitive timescale of days? How the two parts of this sentence link is not entirely clear. If runoff generation length scales are smaller, a larger proximity zone would presumably add noise to statistical relationships.

We agree that the ideas presented in the sentence were poorly connected. We will rewrite this sentence in the revised manuscript.

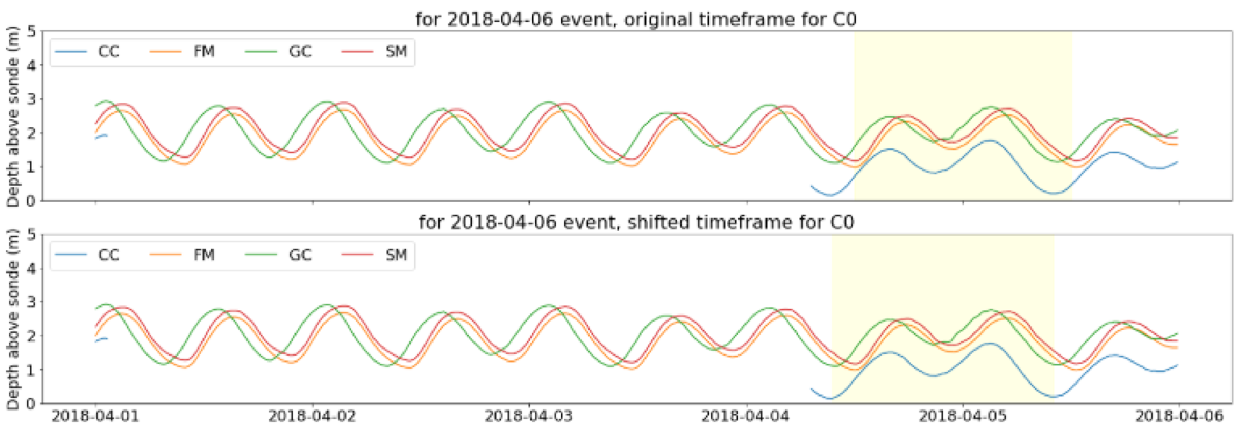
Line 176: Should mention atmospheric rivers for CA

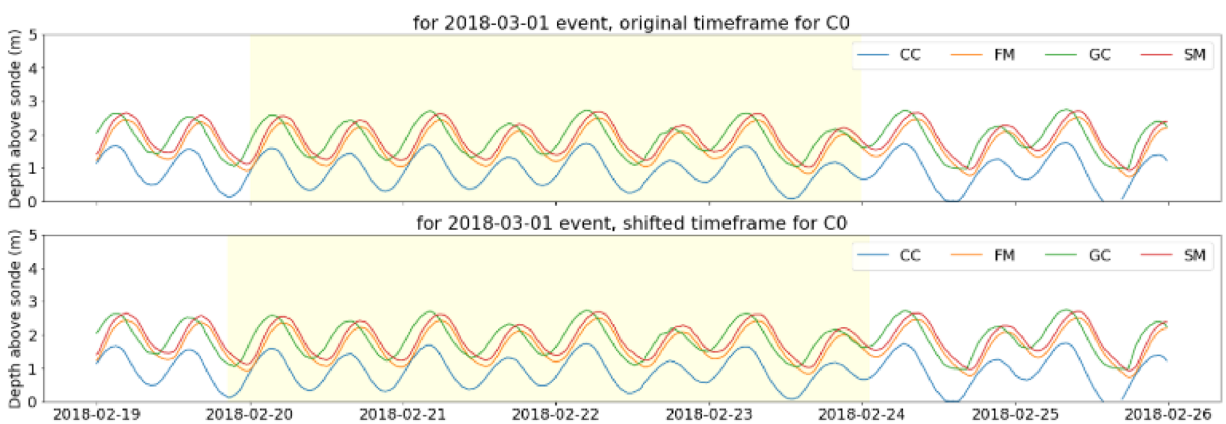
We will add the atmospheric rivers of California that result in precipitation events during the cool season to the revised manuscript.

Line 220-221: Was the tidal cycle accounted for in selecting the window of time for averaging? Averaging over partial tidal cycles may skew results in a way that could produce spurious differences among sites in the same estuary or across estuaries.

When selecting the timeframe for calculating average C_0 values we were unable to consider the exact tidal cycle at each monitoring location due to non-uniform timing of the tide across stations. Please see the temporal change in depth at each monitoring station and the lag between the arrival of the tides at SFB in *Figure R2* below.

Overall, the timeframes selected for the C_0 calculations incorporate most of the tidal cycles. To assess this reviewer's concern, we re-evaluated C_0 for three events for monitoring locations at the San Francisco Bay (SFB) estuary by shifting the original timeframes to coincide with the tidal cycle. We selected the SFB estuary because it experiences largest tidal fluctuations from the set of estuaries used in this study. The resulting C_0 with shifted timeframes were virtually unchanged from the original values (please see *Table R1* below). We will keep our approach for calculating C_0 values.





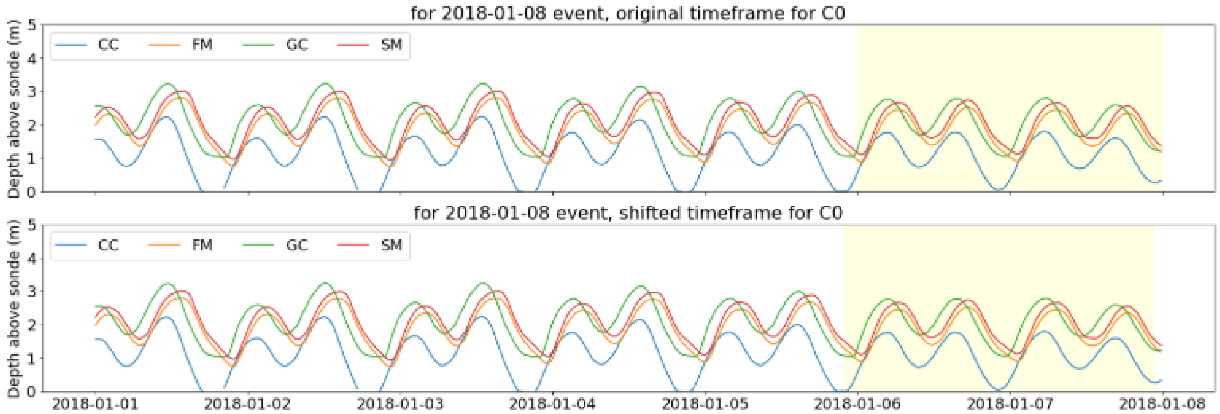


Figure R2: Examples of time delay between tides at monitoring locations at the San Francisco Bay estuary and the original and shifted timeframes used for calculation of average C_0 values.

Table R1: Average C_0 with original and shifted timeframes for three precipitation events at monitoring locations at San Francisco Bay (SFB) NERR.

Monitoring location at SFBNERR	Average C0 for event: 2018-04-06		Average C0 for event: 2018-03-01		Average C0 for event: 2018-01-08	
	Not accounted for the extent of the tidal cycle	Accounted for the extent of the tidal cycle	Not accounted for the extent of the tidal cycle	Accounted for the extent of the tidal cycle	Not accounted for the extent of the tidal cycle	Accounted for the extent of the tidal cycle
China Camp	9.06	9.04	9.30	9.32	8.234	8.231
Gallinas Creek	8.58	8.61	9.71	9.73	7.6041	7.6046
First Mallard	6.49	6.46	8.067	8.064	7.45	7.44
Second Mallard	7.37	7.35	NA	NA	8.82	8.79

Line 231-232: Provide a table of predictor variables

Thank you for the suggestion. While the predictor variables are reported in Figure 3 and Table S4, to aid the readership, we will add the list of predictor variables to the methods in the revised manuscript.

Line 479: N loading and low DO are not always associated. San Francisco Bay is one place where they are not, so this estuary is not the best one to highlight earlier in the paragraph.

The reviewer is right regarding the lack of association between N loading (as DIN) and low DO in the San Francisco Bay estuary, although studies suggest that this might not always have been the case. The reduction in primary production, which affects DO, was previously shown to relate to changes in N species entering the SFB estuary. Particularly, the reduction in primary production in SFB was related to increase in ammonium, which inhibits the uptake of nitrate. The authors attribute the shift in N species loading to changes in wastewater treatment practices in San Francisco (Dugdale et al., 2007, Parker et al., 2012).

Nonetheless, we acknowledge that the message in the manuscript is misleading. We will revise this paragraph.

- Dugdale, R. C., Wilkerson, F. P., Hogue, V. E., and Marchi, A.: The role of ammonium and nitrate in spring bloom development in San Francisco Bay, Estuar. Coast. Shelf Sci., 73, 17–29, https://doi.org/10.1016/j.ecss.2006.12.008, 2007.
- Parker, A. E., Hogue, V. E., Wilkerson, F. P., and Dugdale, R. C.: The effect of inorganic nitrogen speciation on primary production in the San Francisco Estuary, Estuar. Coast. Shelf Sci., 104–105, 91–101, https://doi.org/10.1016/j.ecss.2012.04.001, 2012.

Line 532-534: Please rewrite this paragraph with a more coherent narrative. The topical sentence, focused on variation within single estuaries, is not supported by the following sentences, focused on a comparison between two estuaries (CBM and GTM). The topical sentence also focuses on relationships between physicochemical factors and resistance, whereas the supporting sentences focus on the response of physicochemical factors to precipitation.

Thank you for pointing this out and apologies for the confusion. We will rewrite this section in the revised manuscript.

Technical comments:

Line 87: Reverse the order of the clauses in the previous sentence, since this sentence references the later study, which is in the former part of the sentence (could be confusing!).

We will revise this sentence.

Line 105: Add “of” after “understanding.”

We will add “of” after “understanding” in the revised manuscript.

Line 205: Replace “infer” with “be related to”

We will replace “infer” with “ be related to” in the revised manuscript.

Line 241: Alway --> always

Thank you for catching this. We will fix it in the revised manuscript.