Overarching thoughts:

In general, the work is an important contribution to a developing field of understanding biogeochemical variability in coastal zones and in particular, estuaries. The authors have a robust, long-term data set that merits publishing. I think much of the information could be trimmed down to focus primarily on the carbonate chemistry and air-sea fluxes, and a lot of information in tables could be moved to supplementary material to streamline this and keep the more novel aspects of the work at the forefront. This would also aid with the overall length. Additionally, see cautionary comments on the air-sea flux estimates. Authors could also benefit from some re-organization within the methods discussion to clearly define all data streams being used, and a separate section for applied statistical methods. Much of this information is buried in the discussion.

Intro:

49-50: Could specify that *an increase* or *addition of* CO2 acidifies seawater.

56-57: Vague, I would remove or specify. As it reads now, you say that 'the speciation of inorganic carbon in seawater is important […] because the ocean contributes substantially to the global carbon budget, which is important to understand due to climate change". What do you mean? I see the motivation behind understanding C chem variability but not so sure about the carbon budget point here.

70: Maybe I am missing something here: 'mean diel ranges exceed 0.1 but single day ranges can exceed 1' – isn't the 0.1 value stated here somewhat meaningless if the time frame over which you are averaging isn't stated? (Can you add "mean *seasonal* *weekly* or *monthly* ranges can exceed 0.1 and single day ranges can exceed 1 pH unit" or something along these lines).

93: What are "relatively high CO2 fluxes" (do you mean *influx, efflux*, or just generally larger magnitude fluxes regardless of direction?).

Materials and Methods:

Figure 1: Can you add a scale bar. From the yellow arrow it looks like the study actually did not take place in the estuary, but rather just outside. It is interesting because the paper to this point feels framed around estuarine C chem, but the location of the sensors based on Fig. 1 is an in-between spot, that seems more in the GOM than the estuary itself. This is fine, but the discussion to follow will need to carefully discuss GOM versus estuarine influences as this clearly will get a lot of both.

Section 2.2: Could provide a few more details on calibration procedures in this section – see Rivest et al. (2016) or Bresnahan et al. (2014). Was the -0.05 correction applied across the entire dataset as a calibration, or was this a slow drift? If the latter, how was this correction performed?

Section 2.4
The air-sea flux methods make a lot of assumptions. Recognizing that these can be difficult to measure, I think the overarching methods applied are acceptable, but should be some careful discussion or sensitivity analysis performed. For example, alterations in the air-sea gradient can drastically alter the estimated fluxes. I am not sure that using global average atmospheric xCO2 values is appropriate, but I have not run the sensitivity analyses to check. How often were these global values applied when local data were not available? If infrequently, it would be good to state and authors should discuss the continuity between these values. For instance, if the mean diel range in seawater pCO2 was 58 uatm, and the variability in the global estimates from the air-side concentrations eclipse this seawater variability, this would not be appropriate. Similarly, reporting the distance between the sampling site and the stations that wind and CO2 data were pulled from can help readers ID how large these assumptions are. Can authors clearly state over what frame data are averaged as well? Are you using hourly avg wind and CO2 data in calculations? Daily? Etc.

Standard parametric statistical test applied here, no time series analysis or more advanced techniques – may be worth noting that all data met assumptions and if not were there any transformations etc.

A very brief background on the T/B methods might be helpful as many of this paper's discussion rests on this

~145: What is the sampling duration and frequency? As in (e.g.) is each 1 hr measurement had a single collection point, once per minute then averaged over the entire hour?

~162: Were discrete/bottle samples for pH and pCO2 also collected from the cooler and taken back to the lab for these analyses or for comparison with the sensors' continuous sampling? If so, where are these methods? (you go on in the next section to describe discrete sampling, but it is unclear if these discrete sampling data were used in any sensor calibration procedures? Or just to corroborate uncorrected sensor data?) It is unclear if these discrete samples are their own analysis or not.

258: Takahashi et al. 2002 is not in the works cited.

Results:

There is a lot of information in tables, some of which might be better placed into a supplementary table to reduce length.

Figure 2: Figure text states that panels E-H covers the same dates as panels A-D, but the x axes do not match this (typo in caption).

Table 1: Why do the winter CO2 flux values have one value in parentheses only? I am confused by the caption "CO2 fluxes were calculated using the Jiang et al. (2008) wind speed

parameterization for gas transfer velocity, and ranges of CO2 flux that are given in brackets represent means calculated using parameterizations from Ho et al. (2006) and Raymond and Cole (2001), respectively." What do you mean by respectively? Do you mean that Ho et al. are in brackets and Raymond and Cole are in parentheses? Or why are some values in both parentheses and brackets? Not getting the coding here.

Table 2: You note significant difference by the p-value that summer temperature are higher at night than during the day. While true according to your statistical test, the error bar will overlap very considerably, I would check you test and if correct, make statements accordingly – the difference in reality seems marginal.

Table 3: Figure caption describes a lot of the statistical methods. I would remove from the figure captions and place into the statistical analyses section above (I might suggest making statistical analyses its own section entirely, rather than mixing this information in the same section as the air-sea fluxes etc). For example, just define alpha in this section, you already stated you will use a Westfall adjustment, and if described in the methods, you do not need to describe this or why individual one-way ANOVAs were conducted in the caption.

Table 5: Are these mean +/- SE or SD? Can authors provide the sample size in addition to the p-value for the t-tests run? Hard to judge based on p values alone. Additional test statics and pertinent info could be placed into a supplementary table.

447: Format error (one extra tab)

Discussion:

508-509: Does this really remove the influence of tides? It seems more like you can just analyze controls during these tidal periods rather than you removing their influences? Rephrasing might fix this.

538-551: This is the first mention of chlorophyll and these other env. parameters being used for further analysis. This should be in the methods, in addition to the statistical and data processing techniques used in analysis (ie Pearson's correlation in Table 6).

560-561: Why is water temperature warmer at night? Isn't this unusual? Can you explain why?

579: Spelling (asterisks)

605-611: To my earlier point about air-sea flux methods, I wonder if these results would change with more precise estimates of atmospheric pCO2 levels.

632-635: Given the analysis, it seems like rather than running isolated linear regression between two factors, the authors could consider a 'full' model where available data across all the aforementioned parameters are included and those that are significant are selected. This would be a more typical approach to explain variation.

651: Typo ('both pCO2 and pH')

657: Belongs in methods

684-685: What do you mean by tidal control? As in, in the winter, the nighttime typically corresponded with tidal phases that carry lower pCO2 water into the sampling site at night? Some discussion of mechanism might be worthwhile here.

700: As mentioned above, if DO is part of this analysis, needs to be mentioned and described before this point. Line 706 feels like it belongs in the results.

719-727: I feel this is beyond the scope of this paper and detracts from the main point. If authors insist on including, these methods also need to be described earlier. It is possible this could be move to supplementary information, but I do not think this is necessary. Also – do you think that if it is an urban water way, this boat traffic would further elevate atmospheric CO2 levels? Would this be captured by the station you pull data from?

728-738: This feels like it belongs on conclusions or could be removed altogether. At present, it seems misplaced.

743-747: Methods/statistical analyses. (also – I see you got to the 'full' model approach I suggested above eventually. If the methods are more clearly described, then readers will knows this up front.).

796: Can the authors not find any references other than pers comms to compare their variability to? It seems like this statement could be backed up with references as well.

821-824: Extraneous. This has already been described.

862: When you say *in situ* monitoring, do you mean continuous, sensor-based sampling versus periodic sampling?

870: Yes – I agree with authors on this point. If other methods such as eddy covariance were employed to estimate air-sea fluxes, you may find diurnal differences (I have, also in an estuary). But of course, maybe not! Not sure of this system. Generally, I appreciate the discussion of the sources of error on these measurements that follows.

937-958: This seems a bit in the weeds and not very relevant to this MS. Can this be removed or streamlined.

1003: Font typo

1008: Missing period after contract No. 1605.