General Comments:

This contribution from Granville and Berg covers a highly relevant and timely topic of air-sea gas exchange in seagrass ecosystems. The authors combined underwater EC with other in-situ data sources, and external data products to investigate the topic. While an impressive dataset is presented, and a number of relevant questions are discussed, I do not see a clear argument presented in the manuscript, either related to a new concept or the questioning of existing ones. Instead, the paper touches briefly on several topics, including k parameterization, seagrass ecosystem co2 fluxes and their drivers, as well as different wind speed products and their impacts. The cost of this broad spread is that there is limited quantitative effort placed on any of these subjects, as well as a lack of uncertainty assessments. Again, I believe the topic is relevant and the dataset of high quality, but would suggest the authors rethink their analytical framework and potentially focus on a subset of the topics. More detailed comments follow:

Response: Thank you for your review. We appreciate the time and effort you put into your comments, and find that they will improve the manuscript. Specific responses to the comments are given below.

The overall goal of this study was to derive air-water CO2 fluxes at a fine temporal frequency over a temperate seagrass meadow, and to challenge existing assumptions regarding deriving such fluxes by showing how much they can vary depending on how they are calculated, even when using generally accepted k parameterizations. This extreme variation or uncertainty is too often overlooked in published studies presenting gas exchange estimates for shallow near-shore environments.

With that purpose, the study describes the different components that went into deriving air-water CO2 fluxes and to illustrate the challenges and importance of doing this work on a local scale. This includes:

1. **Using an *in situ* wind sensor instead of regional data, which is a common practice. As we demonstrated, which regional wind station is selected can impact the wind speed values and therefore, the k values and CO2 fluxes.**
2. **Selecting a best-fitting existing k600 model for our site. Published and widely used k600 parameterizations result high variation of calculated k values and gas fluxes. This is of particular concern in shallow coastal bays, which can have site-specific characteristics of both the open ocean and rivers. Using some in-situ data where k was derived from air-water O2 fluxes measured on-site by upside down aquatic eddy covariance helped us make an informed decision regarding existing k600 model was best for deriving the CO2 fluxes.**

3. **Use the appropriate regional wind data (when *in situ* data were not available), the k600 model that was most representative of our site, and sensors that measured pCO2 and environmental variables at a fine temporal frequency to derive the multi-day and multi-year CO2 flux dataset.**

Technical comments:

I also have two related thoughts about the application of O2 fluxes to derive a k for CO2. The first, and to me more pressing of the two, is related to chemical enhancement, which contributes to increased CO2 transfer especially at high pH, but importantly does not exist for O2. The problem is that the calibration window used to parameterize Eq 3 was in the nighttime, when chemical enhancement was relatively low. As the daytime pCO2 lows (close to 200 µatm) were likely also accompanied with a high pH, this would mean a relatively high kCO2/kO2 that is not captured by Eq 3. Chemical enhancement of kCO2 can be on the order of a few cm/hr, which is especially relevant during periods of low wind speed which were quite common for all of the sites for the first half of the study.

I also don't believe that the Sc scaling captures well the difference in gas transfer between co2 and o2 during breaking wave conditions (which I assume you had, at least during the storm). Basically, my concern is based on the understanding that bubble-driven transfer contributes proportionally more for O2 than for CO2 (Woolf 1993; Asher and Wanninkhof 1998). This presents a problem, as the period used for calibrating the O2 -> CO2 conversion was calm, and I assume without breaking waves. My impression is that the conversion from O2 to CO2 as formulated in Eq 3 would only really be valid when sea state conditions are similar to the calibration window.

Taking these two points together into consideration, the approach used here to convert from kO2 to kCO2 would likely underestimate kCO2 at low wind speed, and overestimate at high wind. A few ideas on how to improve this: one approach would be simply to only apply Eq 3 for wind speed within the range of the calibration window, and not to extrapolate beyond that. Or, the authors could apply a range in plausible n exponents in Eq3, and use these to give a uncertainty bound for calculated kCO2. Another possibility would be to resolve this with a discussion in the text.

Response: Two main points are raised here:

With respect to point 2, we will state that UAEC works well for calm conditions without breaking surface waves/ripples and bubble entrainments. It should be noted that under such conditions where traditional k parameterizations are most uncertain like low wind conditions, UAEC has its strengths.

With respect to point 1 we will, if invited to revise our manuscript, evaluate in details how much chemical enhancement of CO_2 exchange can affect our results. An initial search pointed us to several papers, including:

- **Smith 1985 (doi:10.1111/j.1365-3040.1985.tb01674.x)**
- **Wanninkhof and Knox 1996 (doi:10.4319/lo.1996.41.4.0689)**
- **Bade and Cole 2006 (doi:10.1029/2004JC002684)**

Our understanding from those papers is that chemical enhancement of CO_2 fluxes is a mainly potential problem at high pH. Measurements at our South Bay site revealed relative low pH values with a mean of 8.1 for June (Berg et al 2019; doi:10.1002/lno.11236). Also, the detailed evaluation by Wanninkhof and Knox 1996 concluded that: "chemical enhancement has a minor effect on air-sea gas exchange of CO_2 under average oceanic turbulence conditions. However, for the equatorial CO_2 source regions, with high temperatures and low winds, the calculated CO_2 enhancement amounts to 4–8% of the total exchange."

Line by line comments:

30: Please consider rephrasing the first sentence. CVEs don't transport terrestrial C to the ocean, but rather intercept it on their way. More broadly, since this study focuses on inorganic carbon, so I would suggest to support this opening paragraph with some statements linking OC storage and air-water CO2 flux in CVEs.

Response: Good point – We will rephrase the first sentence as suggested and add supporting statements linking OC storage and air-water CO2 flux in CVEs.

35: The uncertainty in GHG fluxes does not directly affect Corg stock variability. Please rephrase.

Response: We will rephrase to clarify that contributions of GHG fluxes to the Corg stock are uncertain.

47: I don't think Ollivier et al 2022 measured air sea flux directly. They just used CRDS to measure the air-water CO2 gradient. There's also the gradient flux approach, although I can't recall a paper using it in seagrass meadows.

Response: Thank you for the correction. We will revise this sentence.

Figure 1: I have trouble placing where exactly the Delmarva peninsula is in broader regional context. Maybe it would be easier to see if the inlet map was zoomed into the eastern seaboard, rather than zoomed out to all of north America.

Response: We will revise the map to implement this change.

109-114: Not sure you need such detailed description of the Corg stocks here. If you need to trim length, this would be a good start.

Response: We will take this into consideration.

144: It might be nice to see where exactly this ERA5 grid cell is on the map in figure 1. Does it overlap with land at all, and are the field sites within its bounds?

Response: We will add this information to the map.

167: Would it be useful to say here why it only measured quality fluxes <4.5 m/s?

Response: This study was the first time a UAEC system was deployed in a tidal coastal system and we selected the 4 hour deployment period for this study as an example because it illustrates the impact tidal current flow can have on gas exchange rates under calm wind conditions; an effect that is usually not accounted for in traditional k parameterizations. In general, UAEC can work for higher velocities than 4.5 m/s. Breaking waves and the resulting bubble entrainment become considerations at higher wind speeds.

Equation 3: See comments above about conversion of kO2 to kCO2. Basically, I'm concerned about the application of k calculated for a non-reactive gas (O2) being used to derive k for a reactive gas (CO2).

Response: Please see our response above.

222: this is only the manufacturer stated precision/accuracy of the sensor, right? Still, this means an uncertainty of ca 40 ppm, which is ca 10% of the mean pCO2, and 20% of the typical daily low around 200. I feel it might be useful to quantify the impact of this analytical uncertainty on derived CO2 fluxes, and compare to the uncertainty coming from k parameterization.

Response: Correct, this is the manufacturer stated precision/accuracy of the sensor off-the-shelf. However, what is currently written in the text does not provide the full explanation of uncertainty related to the sensor. We sent the sensor back for annual calibration by the manufacturer and received a specific calibration curve that we used for data calibration and quality control. Although we listed the information about the precision and accuracy, we used

our individualized calibration curve to calibrate the sensor to the actual concentration. We will add in these details and reconsider how this information is presented. We will also add information about the analytical uncertainty this adds to derived fluxes.

224: How were DIC measurements used to validate (or was it to calibrate) pCO2 from the sensor?

Response: We took bottle samples at our site during our deployments and ran them for DIC and total alkalinity. pCO2 was derived from the bottle samples and salinity, temperature, and pH collected at the same time. We compared results from the pCO2 sensor at the same time point as the bottle sample was taken for validation. This was on top of the calibration mentioned in the previous answer. We will include more details about this process in the text.

225: Does the NOAA carbon tracker provide locally-downscaled CO2 values, or is this a regional or national average? Please state.

Response: The tracker provides regional averages of the CO2 mole fractions. Atmospheric pCO2 values were then determined using the regional mole fractions and locally measured variables. We will state this.

263: I don't think a "relative SD" would be the best comparison here. The absolute magnitude differences across k models is by definition larger at higher wind speed, but the relative difference is larger at low wind speed. The way the comparison is reported here is misleading, because it is generally the difference in k models at high wind that has the largest effect on flux parameterization.

Response: The intention of depicturing the relative standard deviation is to illustrate the uncertainty (or variability) of flux estimates which is the largest at low wind velocity. For that reason, we prefer to keep the figure as it is, but will add the reviewers point specifically in the discussion.

Figure 4b: I'm confused by the reported grey SD bars in the U10 vs k600 figure. Why to they span across zero? In this context, I don't see what a negative k600 would imply.

Response: Mean k600 is above zero. The gray relative SD bars show that the coefficient of variation in k600 is much wider when wind speeds are lower. We did not intend to imply a negative k600; the bars are just statistical. We will add text to the legend to clarify this.

Section 3.2.3: Is this entire section only using the 4-hr period? If so, I think a quick explanation of why only this time was used, and why it was chosen for UAEC kO2 work. Wouldn't more data have helped with choosing a parameterization for section 3.3?

Response: We can add an explanation of why only this time was used, and why it was chosen for UAEC kO2 work. As mentioned in a previous response, selecting this low-wind period for

this study illustrates the impact tidal current flow can have on gas exchange rates under calm wind conditions; an effect that is usually not accounted for in traditional k parameterizations. There were some breaking waves present in other parts of the deployment as well.

299-304: I don't think it really means anything, that the standard deviations of the modeled k600 values overlap and are "not statistically different" because they were modelled from a range in wind speeds….the larger the range in wind speed, the larger the standard deviations in k600. If, on the other hand, your calibration window had consistent wind speed, then the SDs would be much tighter and not overlap.

Response: To clarify, the range in wind speed during this calibration period ranged between 1.5-4.5 m/s, not the full range of wind speeds shown in Figure 5. In this part of the study, our goal was to determine which of the 11 models was the best fit for our site. Assessing standard deviations was one way to do that, but we also assessed the relative root mean square error and produced k vs. U10 curves, which showed the models that were not statistically different also had the tightest fit to the observed data. We will consider adding in one of those additional figures to better illustrate this and rely less on standard deviation.

Section 3.3.1: Here, I think it is missing, a consideration of the uncertainty in k parameterization on modelled CO2 fluxes. 11 k models were used earlier, so I do not see why only one is chosen for CO2 flux derivations.

Response: One of the goals of this study was to constrain CO2 fluxes at a fine temporal frequency over multiple days over the seagrass meadow. As part of this, we wanted to identify one model that could be used to do that at our site. We will clarify that in the discussion.

Section 3.3.2 & 3.3.4: Please clarify which k model was used for the storm fluxes. Was it only N00? If so, I again here think it would be helpful to at least consider a range in k models.

Response: The same k model, N00, was used for the full dataset. Overall, N00 was selected as most representative at our site given the field conditions we encountered.

370: Where in figures 1 or 3 is this underestimation shown, and what is being compared with what? It certainly makes sense that the land-based stations would more often underestimate offshore wind speed, but I don't see where this is quantified in the paper.

Response: Figure 1 is cited in case the reader wants to check the location of the stations. In Figure 3, you can see that the lines representing the land-based stations (Hog, Kiptopeke, Wachapreague), often underestimate the *in situ* wind sensor located at our site. All of the regional wind stations are being compared to our *in situ* wind sensor, because it is very common for people to not use *in* situ sensors and to make estimates from available regional

stations. Figure 3 provides a visual representation (one deployment) of a larger dataset where this is a consistent pattern. To note, our site is not offshore, but it was best represented by the "offshore" wind sensor in the Chesapeake Bay (Rappahannock). The "land based" stations are all located closer to our site than Rappahannock, and unlike Rappahannock, they are a part of the Virginia Coast Reserve like our site – Hog is on a barrier island, Kiptopeke is at the land-water interface, and Wachapreague is over a channel. The methods related to this are in section 2.2 and the results are in section 3.1. We will reconsider the term "land based" and "offshore" and add further explanation in the text about the importance of the regional study to make this clearer.

375: Would it be reasonable here for you to suggest a better way to handle the modeling of

fluxes when only land-based wind speed is available? Maybe some kind of roughness length scaling?

Response: Our recommendation would be to include the boundary layer correction referenced in line 377 (Hsu, 1986; Schwing and Blanton, 1984), or to use the ERA5 reanalysis if available. We will include that as an explicit recommendation in the text.