In this submission by Gutiérrez-Loza et al (GL), a lengthy dataset of eddy covariance data is presented, and used to probe the physical processes driving gas transfer in shallow coastal waters. In particular, results are separated into low and high wind-speed regimes, in which measured gas transfer velocity (k) is correlated with various environmental drivers. GL finds that water-side convection dominates k at lower wind speeds, but combined wind, wave, sea spray, and other forcing take when wind exceeds 8 m/s. My general impression of GL is that the methods are appropriate and largely well described (could use a few clarifications as discussed below), and the discussion and major conclusions are largely supported by the analysis conducted. I also find the scope of the work to be appropriate for publication in BG. Provided that the authors can address the primary concern raised below, and make some improvements to the methods (as also described below), I feel that this manuscript can make a nice contribution to the gas transfer literature.

I have two primary concerns about the work presented by GL:

- First, I am not certain that residual k (k_r) formulated in this manner (k_measured – k_W14) indeed has the effect of "removing the wind-speed dependency from k_660" (line 177, interpretations thereafter). This is because the W14 parameterization was developed for the ocean basin scale, using a (revised) estimate of the bomb 14C inventory and a global wind product. As described in the *comments and recommendations* section of W14, this formulation is intended to be used for "regional-to-global flux estimates of CO2". The W14 formulation is also intended for longer (multiple hours) time scales by squaring wind speed and averaging over 6+ hours, so I am not sure this is appropriate to compare with the 30-minute averaging intervals used in GL. Put simply, W14 describes the global relationship between wind and k, but does not necessarily *isolate* the effect of wind speed on k (except for the y-intercept which is forced through 0). Therefore, while k_r as formulated should correct for *some* of the wind-speed dependency from k_660, I do not feel that it can do so comprehensively at the time-scale applied here.
  With this being said, I am open to the use of k_r in this way, provided that the authors 1) disagree with my explanation above and can provide a reasonable rebuttal explaining so, or 2) if they decide to revise the manuscript text to verbally describe this issue, or 3) if they can calculate a new k_r in a way that better incorporates the uncertainties in the wind-based parameterization (W14 here).

- Secondly, it is not clear what benefit was gained by working with such a long (nearly a decade) EC dataset, as the correlation-based analysis of GL is similar to those applied to shorter-term datasets from the same measurement platform. I understand that a detailed time-series analysis was beyond the scope of this work, but a short discussion may be useful. So, maybe the authors can offer some advice as to the time required to capture the full range in gas transfer variability? i.e., for readers planning a similar coastal EC deployment, is it enough to measure for a year, or do we need many years to capture the variation in physical forcing described in GL?

Line-by-line comments:
92: I see that instruments were located 9m above the ground surface, but how high is this above the sea surface?
98: Was z/L uniform across wind directions within the southeast window?
110: Are there ancillary measurements of T (e.g. from a shaded thermometer, or maybe a closed-path IRGA) to show whether or not solar heating of the sonic anemometer affected the Ts record?

Data processing: I would like to see more detailed statistics showing how many 30-min datapoints were rejected according to individual screening criteria. Comparing, for example, figure 5 with the full time-series in figure 4, it appears that a large majority of data failed the screening criteria. If so, this needs to be fully explained in the methods.

174: The generation of excessive negative k values is a frequent criticism, and major caveat, of EC-based gas transfer studies. So, what is the justification for removing -k values when they otherwise meet the screening criteria applied to the rest of the dataset? Doesn't removing negative values artificially decrease the variability in calculated k?

177: As per the discussion above, I do not agree that k_r "remov[es] the wind-speed dependency from k_660". Given that the majority of the analysis in GL revolves around k_r, I think some additional description of the W14 parameterization and it's applicability to the current study site is warranted.

245: This enhanced wind dependence of k under unstable conditions is consistent with prior work in the Baltic (https://doi.org/10.1007/s10546-018-0408-9) and elsewhere (https://doi.org/10.1007/s10546-018-0408-9; https://doi.org/10.1002/lno.11620). Since these conditions are associated with the largest deviation of measured k_660 from k_W14, can the authors offer any further ideas as to the major driving cause?

259: Couldn't the lack of relationship between the wave field and k_r be in part explained by the fact that (as explained by the authors), the waves here are not swell but rather locally-generated by wind? I.e., one would expect a strong correlation between wind and wave height here?

299: Maybe I missed it, but I do not see where the formulation of McGillis 2001 is compared with the k values calculated by GL.