

We thank the anonymous Reviewers for their considerations/comments. We hope to have improved the manuscript based on the suggestions and comments provided by the referees.

General comment: We revised our methodology to calculate epsilon in response to the reviewer comments. We also carried out additional minor changes to the methodology as outlined below. The revisions lead to only slight changes in epsilon estimates with no major consequence for the results or conclusions of this manuscript.

A more detailed answer/change to each specific comment follows.

Note that in the “change” section we referred to the lines in the manuscript with tracked changes.

Reviewer 1

Comment:

Dissipation rates of turbulent kinetic energy were estimated using a bulk approach (from channel slope and flow depth), as well as from local measurements of turbulent velocity fluctuations. I suggest adding and to discuss a comparison of both dissipation rate estimates, as the bulk approach can be more easily applied to field conditions.

Answer:

Thank you. We will add a brief discussion about this issue.

Change:

We have expanded the 3.5 Section with a discussion about this issue (see L. 369-374).

Comment:

I suggest mentioning the range and variability of measured CO₂ fluxes in addition to dissolved concentration and gas exchange velocity.

Answer:

We will insert the range of CO₂ fluxes in the text.

Change:

We added the range of measured fluxes and associated variability in the first paragraph of the result and discussion section (see L. 278-282).

Comment:

Fig. 2: why not using a scaled x-axis (instead of a categorical axis), where significant regressions could be added to the graph.

Answer:

We initially used a scaled x-axis for velocity and discharge but then we decided it was better to use a categorical one to have a clear visual comparison of the mean values obtained with the different designs and deployments. Linear regressions for slope and discharge were computed and discussed in the text but not shown in the plot (see L. 266-282).

Change:

We have not made any change here.

Comment:

Comparison of the scaling coefficient (alpha) in the equation relating k₆₀₀ to dissipation rates to other studies: energy dissipation rates depend on the depth at which measurements were taken (see e.g. Esters et al. 2017, doi:

10.1002/2016JC012088 for wind-driven systems). In streams, turbulence is driven by bed friction, leading to a different depth-dependence of dissipation rates (and values of the "constant" α). Dissipation rates from bulk scaling (see above), in contrast, assume uniform distributions. This issue could be discussed further when comparing the results to other studies.

Answer:

We agree, we will add a more complete discussion about this issue in the revised text.

Change:

We have added the sentences: " According to Esters et al. (2017), the coefficient α follows the strong depth-dependency in ϵ . While Esters et al. (2017) refer to standing waters, this can also be expected to be true in running waters where epsilon varies with depth as a result of bottom friction (Sukhodolov et al., 1998). Therefore, an in-depth comparison between our results and previous studies proves difficult because of the different measuring depths across the studies. (see L. 356-360)."

For instance, Zappa et al. (2007) measured ϵ from few cm to 3 m below the surface, instead Vachon et al. (2010) used ϵ at a depth of 10 cm.

Comment:

line 452: I assume that the ADV velocities were rotated into a vertically oriented coordinate system before all subsequent analyses?

Answer:

Yes, we agree. We will add a sentence to explain the rotation, e.g. "We rotated the velocities to an earth coordinate system with an along-flow (u), cross-flow (v) and two duplicate up-down (w1, w2) components following standard transformations provided by Nortek (2020)."

Change:

To clarify the rotation further, we added "vertically-oriented" to "earth coordinate system" (see L. 493).

Comment:

line 483: lower bound of the wave number for spectral fitting: the lower bound should not be defined by water depth, but by the distance of the ADV sampling volume from the water surface (as this defined the largest isotropic eddy).

Answer:

Thank you very much for pointing this out. We agree on this point and we will correct the Ms accordingly.

Change:

We recalculated epsilon based on the distance of the ADV sampling volume from the water surface used as the lower bound of the wavenumber for spectral fitting. We changed the methods description (L. 518) and online R code accordingly.

Comment:

line 501: estimates of energy dissipation rates are typically log-normally distributed (see e.g., Baker et al. 1987, [https://doi.org/10.1175/1520-0485\(1987\)017<1817:STITSO>2.0.CO;2](https://doi.org/10.1175/1520-0485(1987)017<1817:STITSO>2.0.CO;2)). Arithmetic averaging may therefore be not appropriate.

Answer:

Thank you very much for pointing this out. We have verified that our data was log-normally distributed.

Change:

We now use the geometric mean to characterize the central tendency in epsilon values across the flume, following Baker and Gibson (1987). We clarify this in the methods description (L. 537-540). We also rephrased the text slightly to further improve clarity.

Reviewer 2

Comment:

L26 greenhouse gas emissions

Answer:

Thank you. We will correct it.

Changes:

We have changed Green House Gas with greenhouse gas (see L. 26).

Comment:

L33 might be helpful to show Eq. 1 here

Answer:

Ok, we will fix it.

Changes:

We have moved the Equation 1 and its description in the introduction section. We have slightly adapt the text to fit this rearrangement (see L. 33-37 and L. 155-159).

Comment:

L37 Yes, k can vary in space and time, which is a very important characteristic. I suggest expanding the aspect of spatiotemporal heterogeneity. Maybe the authors can add some examples or numbers to give us a better understanding how much k can vary in space and time in rivers? This could be then also used in the discussion of spatial k_{600} of drifting k_{600} .

Answer:

Thank you for this suggestion. We will expand the issue of spatiotemporal heterogeneity of K in the revised text.

Change:

We have added the following two sentences that provide two examples of the variability of k_{600} in space and time (see L. 43-46).

“For instance, Jeffrey et al. (2018) observed a 230-fold variation in the gas exchange velocity between two nearby stations along an estuary in the mid-coast of New South Wales, Australia. Likewise, Natchimuthu et al. (2017) estimated that flood events could produce a 7-fold increase of k in a hemiboreal catchment in Southwest Sweden.”

Comment:

L40 Are there k models for rivers other than from Raymond et al. 2013? If yes, do they also use wind, current and slope? I'm surprised to see only one reference here.

Answer:

In L40 we quoted the paper from Raymond et al. 2012 (not Raymond et al. 2013). In this paper the authors summarized many of the available equations relating gas transfer velocity to the hydraulic geometry. We will replace “Raymond et al. 2012” with “Raymond et al. 2012 and references therein”

Change:

We have changed the references (see L. 47).

Comment:

L55 a floating "flying" chamber design with flexible chamber walls has also been successfully used by Rosentreter et al. 2017 and Jeffrey et al. 2018.

Answer:

We will add the suggested references. Thank you.

Change:

We have inserted the suggested references (see L. 64).

Comment:

L58 Yes, local CO₂ sources such as groundwater inputs change surface water CO₂ concentration, but how would they interfere with local k?

Answer:

Local groundwater inputs rich in CO₂ do not affect k. We argued that the chamber method allows direct point measurements of gas fluxes and these latter could be helpful to observe the spatial heterogeneity of gas fluxes in a stream sourced by pointy CO₂ groundwater inputs. We will better clarify this point in the revised text.

Change:

We have rephrased the sentence (see L. 68-69) with "Furthermore, chambers allow direct point measurements of gas fluxes, that are especially useful in headwater streams typically characterised by spatially heterogeneous conditions and complex CO₂ patterns (Ploum et al., 2018; Rawitch et al., 2019)."

Comment:

L69-70 This may be exaggerated. For example, the study by Rosentreter et al. 2018 compared k of CO₂ in mangrove surrounded creeks, lakes, main river channel, and a bay and in direct comparison to a dual tracer experiment and found good agreement between the two methods (5% discrepancy). Lorke 2015 compared drifting vs anchored chamber measurements. Jeffrey et al. 2018 compared chamber measurements in different sections of an estuary. etc... so this has been discussed before and also quantified.

Was there a fan attached inside the two chambers? Did you test for evenly distributed air circulation inside the chamber?

Did you test for temperature artefacts inside the chamber? Was the temperature constant during chamber incubations?

Answer:

All the studies mentioned in this comment (Rosentreter et al. 2018, Lorke et al. 2015, Jeffrey et al. 2018) discuss or quantify the uncertainty associated to the estimate of k via the chamber method but using the water CO₂ concentrations obtained via other techniques. In our manuscript we discuss and quantify the uncertainty associated to the estimates of k derived only from CO₂ observations inside the chamber. This will be further stressed in the revision of the paper. A fan was not used inside the chamber and the temperature was always almost constant during the chamber incubation. This information will be added in the Method section.

Change:

We have stressed our point by changing the disputed sentence (see L. 79-80) with:

"Furthermore, the uncertainty associated with k estimates relying exclusively on CO₂ concentrations gathered using chamber-based CO₂ measurements has not been discussed nor quantified by previous studies."

We have specified the chambers have not a fan (see L. 95-96) and we measured almost constant temperature during the deployments (more than 80% of the runs experienced a variation in temperature less than 2.5 °C) (see L. 148-149).

Comment:

L126 what CO2 sensors? Please add brand, model, and accuracy of CO2 sensor and CO2 analysis.

Answer:

The CO2 sensor specifics are indicated in L. 101-104.

Changes:

We have not made any change here.

Comment:

L127 Roughly, how long did you conduct chamber incubations (runs) for? minutes, half an hour? an hour?

Answer:

Chamber incubation time for steady deployments ranges between 12 mins to 45 mins. This will be specified in the revised methods.

Changes:

We have inserted the range for incubation time of steady deployments (see L. 140-141) and we have explicitly written to refer to Table 2 (see L. 140) for the incubation time of drifting deployments.

Comment:

L134-135 Is this a problem? Even if chamber concentrations inside were not atmospheric, you can still use the change of concentration for estimating k, no? If you measured CO2 every 30sec over the duration of the chamber incubation, then you have a start and end concentration over time (F) that you applied in Eq.2 and Eq.3. Meaning only the difference between start and end concentration is important (slope) and not the concentration itself. I'm curious to hear if the authors agree or disagree.

Answer:

The equilibration condition at the moment of incubation does not affect per se the measure of k. Thus, we agree on the possibility of estimating k also in the case the chamber is not perfectly equilibrated to the atmospheric value at the moment of incubation. In general, ensuing a perfect equilibration before incubation is important to have a realistic representation of the CO2 fluxes occurring at the water-atmosphere interface. Moreover, if the chamber is not equilibrated to the atmosphere you might have a non-homogeneous CO2 concentration inside the chamber at the beginning of the incubation and this could generate a potential bias in the typical exponential curve during the incubation (i.e. Eq. 2 and 3 could be no longer valid). Also, ensuing equilibration to the atmospheric value before incubation guarantees the maximization of the concentration gradient and this leads to a clearer CO2 signal.

Changes:

We have not made any change here.

Comment:

L139 Were the atmospheric concentrations outside close to 400 ppm?

Answer:

Yes, we will explicitly indicate it around L. 139.

Changes:

We have specified that the sensor showed atmospheric concentration close to 400 ppm (see L. 153).

Comment:

L193 do you mean increasing "linear regression"? If yes, what was your threshold r^2 ?

Answer:

The first data quality check we used was to consider only the curves that showed a monotonous increase (or decrease) in CO₂ concentration inside the chamber volume during the incubation process. This was just a visual data quality check. Then we used the NSE coefficient to discriminate the ones with the highest performance. This will be clarified in the revised Methods.

Changes:

We have specified that the first was just a visual test (see L. 206).

Comment:

L250-251 this sentence could be deleted as this is also mentioned in the Table 3 caption.

Answer:

We will delete it. Thank you.

Changes:

We have removed the sentence (see L. 263-265).

Comment:

Figure 4b shouldn't this be k600, not k?

Answer:

No, the figure is referred to a single deployment (there is no need to standardize the k to k600 in this case).

Changes:

We have not made any change here.

Comment:

While this study greatly contributes to our understanding of appropriate chamber design and conditions (drifted vs anchored) of the chamber method in general, I wonder how good this chamber method is in predicting the CO₂ flux in comparison to other k methods and empirical k models? For example, were CO₂ fluxes measured in the flumes better predicted by k600 derived from the chambers measurements in this study than predicted from k600 models (e.g. Raymond et al. 2013, Ulseth et al. 2020)? Or more practically, would the authors recommend to use FF chamber anchored mode over the k600 model by Ulseth et al. (2020) based on energy dissipation for estimating CO₂ fluxes in rivers? Do the empirical models under or overestimate the flux?

Answer:

We thank you for your comment. In this paper, we do not have a direct comparison between the chamber method and other possible k methods. Hence, we are not able to state that the chamber method might be better than other k methods. In general, we observed k estimates from FF chamber in line with empirical models from Raymond et al. 2012 and Ulseth et al. 2019. We are not able to argue that our chamber estimates under or overestimates the flux with respect to these models. Your questions are all valuable, and they provide exciting hints for further research. We are conscious that to have measure of k from other methods might improve the paper and give a further support to the validity of our chamber method. We hope to have the possibility to investigate this further.

Change:

We have no made changes here.

Additional changes:

We corrected a typo (see L. 481): The ADV sampling depth was in fact 5-6 cm, not 3.5 cm as originally indicated.

We corrected a typo in L. 518 (m³ should read m).

We added details on the quality of our ADV data to estimate ε . Specifically, we now state in L. 540-541 that 82% of all ADV measurements passed all quality requirements (e.g. SNR > 70, assumption of isotropic turbulence, positive R² in spectral fitting, etc.) and were used for ε estimation.

We noticed a mistake in the description of the velocity time series despiking procedure. To identify spikes in the velocity time series, we now use the three-dimensional phase-space method by Goring and Nikora (2002) with modifications by Mori et al. (2007). We did not use the approach originally described that identifies spikes as observations outside a multiple of the standard deviation of neighboring data. We preferred the phase space method because it is not sensitive to subjective threshold levels and window sizes, and because it is well established for despiking ADV data. We interpolated spikes using cubic polynomials, not running medians, following procedures by Mori (2020). We adjusted the methods description (see L. 488-495) and online R code accordingly and provide further references that describe the method in detail.

We changed the order of despiking and coordinate transformation, following recommendations by Doroudian et al. (2010, *Limnology and Oceanography: Methods* 8). We now first despike the data and then transform the coordinate system. We adjusted the methods description and online R code accordingly and moved the sentence in L. 482-483 to L. 493-495.