

We are grateful to Reviewer 2 for these constructive comments and concerns. Below we address the comments raised and provide corresponding amendment done to the original manuscript.

In their Technical Note entitled, “*in-situ* measurement of flux and isotopic composition of CO<sub>2</sub> released during oxidative weathering of sedimentary rocks”, Soulet et al. report the results of a “proof-of-concept” study aimed at determining the release rate of carbon dioxide from outcrops of weathered shales and partitioning this carbon dioxide between inorganic and organic sources using C isotopes. The authors clearly describe the design, implementation, and data analysis for their rock weathering chambers in such a way that I am confident that I, or any other researcher, could implement this technique elsewhere. While I ultimately think that this paper should be published in *Biogeosciences*, I have a few comments that I would like to see the authors address (described below).

Firstly, I am confused by the distinction between a “direct” and “indirect” measure of a chemical weathering reaction. On Page 1 line 12, the authors imply that tracking reaction products (e.g., dissolved sulfate in rivers) is an indirect method. However, as carbon dioxide is also a reaction product, I do not see how their method is any more direct than measuring sulfate concentrations. Moreover, the relationship between the amount of product consumed (carbonate or organic carbon) and the amount of carbon dioxide release can be strongly modulated by the buffering capacity of natural waters. As a result, tracking carbon dioxide release may lead to a different assessment of the extent of reaction relative to a product that doesn’t partition into both the fluid and gas phase (e.g., sulfate ion). That being said, I do agree that their method provides a different perspective on weathering reactions than measuring the dissolved or solid phase chemistry of rivers. In particular, I think the Soulet et al. method averages over very different temporal and spatial scales (see below) that make it a nice complement to river based approaches. Perhaps some more clarity as to what the authors mean by direct versus indirect would be helpful.

RE: The words “direct” and “indirect” were used to refer to the way the flux of CO<sub>2</sub> emitted to the atmosphere during oxidative weathering of rocks has been measured/estimated in the literature. In other words, whether CO<sub>2</sub> was being tracked directly, or by another product of the reaction (e.g. Re, or SO<sub>4</sub>) which is what we meant by ‘indirectly’. However, we acknowledge that in two occurrences it was not clearly stated and agree with some of the reviewer’s comments above. We have thus modified the manuscript accordingly (see P1 L11 and P2 L19 in the revised manuscript).

The different stoichiometries for carbonate weathering by sulfuric acid (CWSA) presented as equations 2 and 3 have appeared elsewhere in the literature. However, I am not convinced that, in the context of this paper, there is a real distinction that can be made. The dissolution of one mole of calcium carbonate releases one mole of calcium ion (Ca<sup>2+</sup>) and one mole of dissolved inorganic carbon (DIC; 2H<sup>+</sup> + CaCO<sub>3</sub> → Ca<sup>2+</sup> + H<sub>2</sub>CO<sub>3</sub>), which is equivalent to 2 units of alkalinity per unit of DIC. The generation of sulfuric acid from pyrite oxidation can titrate these 2 units of alkalinity leading to a net reaction for CWSA that results in 0 units of alkalinity generation per unit of DIC generation (CaCO<sub>3</sub> + H<sub>2</sub>SO<sub>4</sub> → Ca<sup>2+</sup> + SO<sub>4</sub><sup>2-</sup> + H<sub>2</sub>CO<sub>3</sub>; equivalent to Equation 2 of Soulet et al.). Equation 3 of Soulet et al. predicts 1 unit of alkalinity generation per unit of DIC generation. In this way, it can be viewed as a 50/50 mixture of carbonate weathering by carbonic and sulfuric acids instead of a distinct pathway for CWSA.

RE: We agree with Reviewer 2. However, we make the distinction between these two pathways in our manuscript in order to link our work to a wider problem that includes the impact of oxidative weathering of rocks (including CWSA) on the CO<sub>2</sub> concentration of the atmosphere over different timescales. We feel it can be better understood for a wider community using these two (and too) simple equations: Equation 2 implying the “immediate” release of CO<sub>2</sub> to the atmosphere, and Equation 3 implying the release of CO<sub>2</sub> over the timescale of 10,000 to 1,000,000 years. Depending on the fluxes involved, these pathways could thus impact climate over different timescales.

Furthermore, the idea that Equation 2 reflects an “immediate” release of carbon dioxide to the atmosphere misses the fact that the aqueous chemistry of weathering fluids will strongly modulate this flux. If there is sufficient generation of alkalinity from silicate weathering, the carbon dioxide produced from CWSA will partition more into the dissolved phase despite generally following the stoichiometry of Equation 2. Similarly, springs developed in carbonate terrains that lack abundant pyrite still degass carbon dioxide into the atmosphere despite the fact that the reaction for carbonate weathering by carbonic acid is often written as generating

bicarbonate ion. In other words, without more constraints on the fluid composition, it difficult to directly relate the extent of an individual weathering reaction to changes in carbon dioxide concentrations (e.g., see Soetart et al. 2007 Maine Chemistry).

RE: In the context of our study, "immediately" has to be compared to the timescales of 10,000 to 1,000,000 years. We acknowledge that in details aqueous chemistry of weathering fluids may modulate the CO<sub>2</sub> flux, but from a geological point of view (10<sup>4</sup> to 10<sup>6</sup> year), this flux of CO<sub>2</sub> is an "immediate" response to oxidative weathering of rocks.

In general, this study lacks replication. While I do not think that this is a critical issue, it'd be worth acknowledging some of the limitations and/or adding more analysis where possible. For example, two chambers are shown in Figure 1C. Is there not two chamber's worth of data to show?

RE: We do acknowledge that our methodology lacks replication, that's why we attempt comparing our results to other methods on other catchments despite issues of scales (see below). However, we are working in natural settings and we expect changes in the CO<sub>2</sub> flux and isotopes in response to seasonal physical-meteorological changes in the catchment area. So we do not expect to find the exact same results for a single chamber over time, and for different chambers at the same time. Please also see our reply to Reviewer #1.

Based on both reviewers' comments we added a section regarding the limitation of our methodology (section 3.3 in the revised manuscript P12 L15-27).

Similarly, I am not sure if I found at what depth below the land surface the chamber was placed. Presumably this depth will have a large effect on the results. What depth was selected and why?

RE: We drilled the chambers on bare rock outcrops, and in places where we could not see roots (see Fig. 4). These outcrops make up 68% of the surface area of the catchment (Mathys et al., 2003; Cras et al., 2007) and are key parts of the landscape contributing to weathering, solute production (Cras et al., 2007) and sediment production (Mathys et al., 2003; Graz et al., 2012). These pieces of information were added in section 2.1 of the revised manuscript (P3 L25-32 and P4 L1-9).

Then, chambers are in the unsaturated zone, and the depth at which they were drilled depended on the accessibility in the field. Chamber H6 was drilled at ~2 meters above the Laval stream (P8 L3 in the revised manuscript).

Page 2 Line 24 - There are many other papers that have used S (and O) isotope ratios to partition the sulfate budget including some that precede the Calmels et al. 2007 paper. For example:

- \* Cameron, Eion M., et al. "Isotopic and elemental hydrogeochemistry of a major riversystem: Fraser River, British Columbia, Canada." *Chemical geology* 122.1-4 (1995): 149-169.
- \* Spence, Jody, and Kevin Telmer. "The role of sulfur in chemical weathering and atmospheric CO<sub>2</sub> fluxes: evidence from major ions,  $\delta^{13}\text{CDIC}$ , and  $\delta^{34}\text{SSO}_4$  in rivers of the Canadian Cordillera." *Geochimica et Cosmochimica Acta* 69.23 (2005): 5441-5458.
- \* Das, Anirban, Chuan-Hsiung Chung, and Chen-Feng You. "Disproportionately high rates of sulfide oxidation from mountainous river basins of Taiwan orogeny: Sulfur isotope evidence." *Geophysical Research Letters* 39.12 (2012).
- \* Turchyn, Alexandra V., et al. "Isotope evidence for secondary sulfide precipitation along the Marsyandi River, Nepal, Himalayas." *Earth and Planetary Science Letters* 374 (2013): 36-46.
- \* Hindshaw, Ruth S., et al. "Influence of glaciation on mechanisms of mineral weathering in two high Arctic catchments." *Chemical Geology* 420 (2016): 37-50.
- \* Torres, Mark A., et al. "The acid and alkalinity budgets of weathering in the Andes–Amazon system: Insights into the erosional control of global biogeochemical cycles." *Earth and Planetary Science Letters* 450 (2016): 381-391.

RE: Thanks. We added Spence and Telmer (2005) and Hindshaw et al. (2016) in the revised version of our manuscript.

Page 6 Line 15 - I'd appreciate a few sentences that explain directly how  $V_{ch}$  and  $S$  were determined. I assume that the dimensions of the drill hole and the assumption that it was shaped as a perfect cylinder were used. However, this ignores the fact the chamber walls are rough and not perfectly impervious. As a result, you are likely to get carbon dioxide from pores and cracks that intersect the chamber walls as alluded to on Page 9 Line 6. I'd appreciate some additional discussion on how this effects area-normalized estimates of carbon dioxide production rates.

RE:  $V_{ch}$  and  $S$  were indeed determined assuming that the drilled hole is a perfect cylinder. We do agree that CO<sub>2</sub> from a certain thickness around the drilled hole contributes overwhelmingly (compared to the CO<sub>2</sub> flux produced at the rock-chamber interface) to the flux we measure.

However, it has to be noticed that we want to provide the community with a flux of CO<sub>2</sub> emitted from the rock natural surface to the atmosphere. This flux includes the CO<sub>2</sub> produced at the rock-atmosphere interface and the CO<sub>2</sub> produced over a certain thickness from the weathered rock. In Draix, the thickness of the regolith is up to 10 to 20cm thick (Oostwoud Wijdenes & Ergenzinger, 1998; Mathys and Klotz, 2008). This means that the CO<sub>2</sub> flux from the rock to the atmosphere is produced over a thickness of 10 to 20cm. Thus, when we drill a 40cm-long hole, rather than creating a new weathering surface at the rock-chamber interface, we instead create a headspace that makes us able to measure a realistic flux of CO<sub>2</sub> from the rock to the atmosphere when we lower the pCO<sub>2</sub> to ~400ppm (atmospheric pCO<sub>2</sub>).

We do agree that it was not clearly stated in the paper. We added some lines to state this point in the section 3.1 P9 L23-27 in the revised manuscript.

Furthermore, are their constraints from porosity, permeability, grain-size, and/or fracture density measurements that can inform the "effective" volume that the chamber samples? Or, could the mass of CO<sub>2</sub> removed during the first few flushes inform this volume? Being able to determine an "effective" volume (as controlled by porosity, permeability, fracture density, grain size, etc.) would help others trying to replicate the methodology in determining if a site would be appropriate based on rock properties.

RE: There are some estimates of the rock properties for the Laval Catchment (Mathys et al., 2003; Oostwoud Wijdenes & Ergenzinger, 1998; Traveletti et al., 2002). These suggest i) the upper ~3 cm are loose material composed of mm-to-cm fragments of marls, ii) from ~3 to ~10cm is the regolith of marl more or less fragmented, iii) from ~10 to 20 cm is the compact lower regolith keeping the marl structure but not its cohesion, and iv) the bedrock (unweathered marl). The porosity has been determined to be 0.17-0.23 (Traveletti et al., 2012). We have added these details to the study area part of the manuscript to help others seeking to replicate the methodology (Section 2.1 P4 L1-9).

However, what matters for the "effective" volume is the connected porosity and gas permeability which is, as the reviewer states, is probably linked closely to the fracture density. There are no measurements of this parameter at the field site and so we cannot use the mass of CO<sub>2</sub> removed during the first flushes to inform us of this "effective" volume.

The purpose of our Technical Note is to show that one can measure reliable CO<sub>2</sub> flux to the atmosphere using a cylindrical chamber and trap enough CO<sub>2</sub> in the field to partition its source through its isotope composition (notably using <sup>14</sup>C which requires larger volumes of CO<sub>2</sub> to be collected). The controls on this flux (of which rock properties and connected porosity are likely to be one) cannot be assessed without more measurements at a range of chambers, and at a range of field sites.

Page 8 Line 27 - How realistic is it that the chamber has such a high pO<sub>2</sub>? My understanding of evidence from the oxygen isotopic composition of sulfate (e.g., Calmels et al. 2007), pyrite reactions fronts (Brantley et al. 2013 ESPL), and gas chemistry in wells (Kim et al. 2014 GCA 2017 GCA) is that oxidative weathering takes place under relatively low pO<sub>2</sub> conditions for many systems. Does this mean that your method provides a maximum estimate of reaction rates?

RE: Weathering occurs not only at the atmosphere-rock interface but over at least a certain thickness into the rock (Petsch et al., 2000; Bolton et al., 2006). Nevertheless, at the atmosphere-rock interface pO<sub>2</sub> is that of the atmosphere. From the chamber point of view, a pO<sub>2</sub> of that of the atmosphere replicates what occurs when the rock is exposed to the atmosphere while pO<sub>2</sub> probably decreases in depth in the rock.

It has to be noticed that our field site is not comparable to those described in the citations provided by Reviewer 2. In the suggested studies, erosion is much lower, leading to a very thick weathering front of 20 m or more (e.g., Brantley et al., 2013). In Draix – our field study – erosion removes 1cm of rock in average per year but it can be more. So the weathering front is probably far less thick, and pO<sub>2</sub> higher.

Page 9 Line 9 - For the analysis of CO<sub>2</sub> fluxes, it is stated that 3-4 flushes are necessary to get the "true" flux determination. What statistical criteria was this determination based on? Similarly, what is the basis for designating 6 minutes as the amount of time to fit the carbon dioxide accumulation curve (Page 6 Line 11)? How do the calculated averages and standard deviations of CO<sub>2</sub> flux vary with measurement / integration time?

RE: When the flux measured vs. number of repeats is examined (Figure R1), one observes a decrease in the flux that reaches steady values after 3 to 4 repeats, while statistically the last four are indistinguishable within 2 $\sigma$ . We are deliberately vague in our manuscript as this feature can change depending on the chamber and flux. In practice, the number of repeats on which flux is averaged has to be adapted based on the results observed.

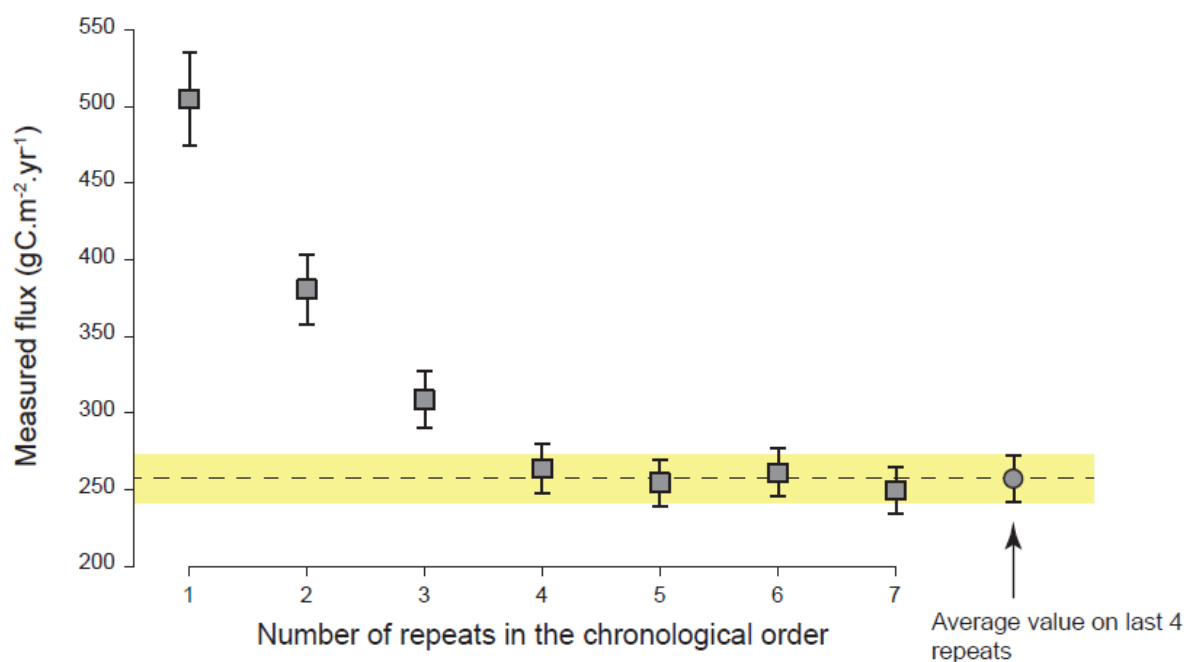


Figure R1: Evolution of the measured flux with the number of repeats (grey filled squares, error bars are 2 $\sigma$ ). Dashed line is the averaged flux over the last 4 repeats (257 $\pm$ 8 gC/m<sup>2</sup>/yr) and yellow bar represents the 2 $\sigma$ -domain of the averaged flux.

Regarding the window for the flux measurements. If we pick 1 to 8 minutes of fitting, the results all agree within 2 sigma (Figure R-2).

The fitting window has to be specified and in our case, 6 mins were chosen as a trade-off. In our manuscript, we present a series of active trapping for which we left the chamber replenish with CO<sub>2</sub> for more than 6 mins. However, when on field, we are not necessarily trapping CO<sub>2</sub>. Instead we are sometimes only measuring fluxes. In these cases, for logistical reasons mainly related to the time we can spend in the field daily, we are monitoring over shorter periods of 7 minutes. Thus fitting over 6 mins was sensible. This parameter can be modified as soon as it is specified. For example, Pirk et al (2016) chose 3-minutes fitting windows. The starting point of the fitting window may also impact slightly the results, although providing similar results within 2 sigma, if it is set to pCO<sub>2</sub> close to atmospheric values.

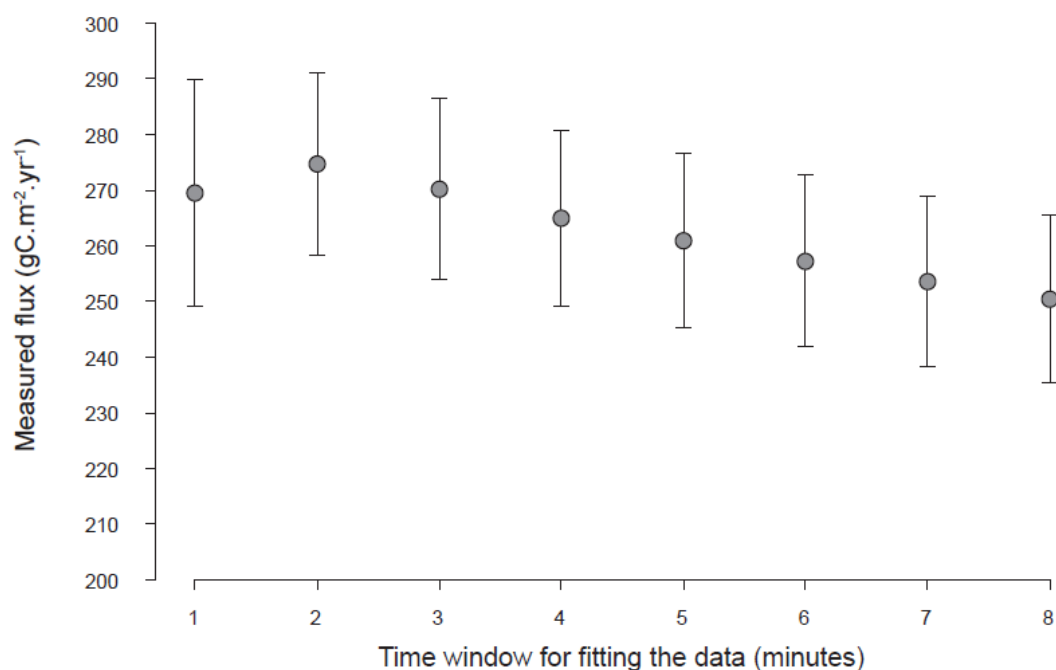


Figure R2: Evolution of the average flux (over the last 4 repeats) with changing fitting time windows from 1 to 8 minutes (grey filled circles, error bars are  $2\sigma$ ). All data agree within  $2\sigma$ . Please, note that scale of y-axis is different from Figure R1.

Page 9 Line 20 - I am not convinced that the difference between the 2 carbon isotopic samples reflects process and not fractionation. The analysis of carbon dioxide fluxes explicitly assumes that there are leaks in the system, which may induce fractionation. Similarly, two different methods were used for these samples. Finally, if the balance between oxidative reactions can vary daily, then why is the entire difference in the isotopic composition of CO<sub>2</sub> derived from the passive trapping method assumed to result from fractionation. In general, a better discussion of which isotopic signals are attributed to environmental process vs. sampling-induced fractionation and why would be helpful.

RE: Previous studies when developing the passive method quantified an isotopic fractionation (Garnett et al., 2009; Garnett and Hardie, 2009; Garnett and Hartley, 2010). In contrast, the pump/active method doesn't fractionate, as shown by e.g. Hardie et al 2005.

We agree that our discussion about fractionation was a bit short. We expanded this discussion and highlighted potential limits but also benefits of using the passive and active sampling methods.

Please note that we changed the  $3.5 \pm 0.45$  ‰ fractionation values (based on merging values provided in Garnett et al., 2009; Garnett and Hardie, 2009) by the now accepted value of  $4.2 \pm 0.3$  ‰ (Garnett and Hartley, 2010) based on a laboratory assessment. This value is indistinguishable from the value obtained in Garnett and Hardie (2009) of  $4.0 \pm 0.2$  ‰. The new applied value minimally changes the source partitioning results and does not changes our interpretations.

Page 10 Line 15 - **This is very interesting!** In other words, the release ratio of inorganic to organic carbon determined by carbon isotope ratios is different than the relative abundances of inorganic and organic carbon present in the underlying rock. Specifically, the isotopic method "sees" more organic carbon than would be expected if one "unit" of rock was congruently weathered. Does this make sense with what is known about carbon and sulfur reaction fronts in weathering profiles?

RE: Once again, in Draix the weathering profile is probably thin (several decimeters) compared to other weathering profiles published (several meters; e.g., Brantley et al., 2013). At this stage, we reiterate the explanation we provided in the original version of our manuscript (P10 L18-20). The dissolution of carbonate depending on the oxidation of sulphides, it is therefore likely that it only occurs locally where sulphides are

concentrated. In comparison the oxidation of organic carbon appears to occur homogeneously in the rock mass. We agree it is interesting and worthy of future study.

Page 10 Line 21 - While I understand the motivation behind including section 3.3, I think that the different area normalization schemes between the chambers and river-based measurements precludes direct comparison. The area normalization in river systems refers to the catchment area. However, weathering takes place at depth within porous media such that the true surface area of reactive material that rivers source solutes from is likely poorly approximated by the catchment surface area. In the chamber experiments, the area normalization refers to the surface area of the chamber walls, which likely more closely approximates the true "reactive" surface area (see above). At the very least, this discrepancy between area normalization schemes should be discussed before generating comparisons between the different datasets. Depending on how reactive surface area scales with catchment area, the fact that the chamber-based estimates are close in magnitude to the river-based estimates may actually mean that there is a large discrepancy in the rates that they predict.

RE: We somewhat agree and have added discussion to the revised version which relates to the referee's comment. This is a common issue when referring to element fluxes per unit of surface area. With more CO<sub>2</sub> flux measurements, alongside solute-based weathering estimates, we will be in a better position to probe these differences in more detail.