Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-304-AC1, 2020

Interactive comment on "Resistance and resilience of stream metabolism to high flow disturbances" *by* Brynn O'Donnell and Erin R. Hotchkiss

Brynn O'Donnell and Erin R. Hotchkiss

ehotchkiss@vt.edu

Received and published: 17 December 2020

With the exception of the "GENERAL RESPONSE", all of our response text sections begin with "RESPONSE" immediately following reviewer comments.

GENERAL RESPONSE:

This text is included in both responses to reviewers, with specific responses to reviewers below.

Here we respond more generally to questions about why/how we selected isolated flow events and the resulting number of events suitable for our analyses (n = 15 events over

5 years). We emphasize that we focused on quality over quantity when selecting for and analyzing stream metabolism before, during, and after high flow events. Our methods were chosen to address a lingering knowledge gap in our understanding of ecosystem processes: how biological processes (gross primary production and ecosystem respiration, GPP and ER) respond to and recover from discrete higher flow disturbances during storms, how those two processes compare to one another, and which environmental drivers may best explain these dynamics. Potential metabolic responses include subsidy (increasing rates due to higher substrate concentrations), stress (decreasing rates due to physical or chemical disturbances), or no change, which allows for our work to build on concepts fundamental to biogeochemistry and ecology. An additional knowledge gap is how different processes (i.e., GPP, ER) may respond differently to high flows. How we chose to quantify changes in metabolism during higher flow acknowledges the "pulsing steady state" of ecosystems in a novel way. In our revised manuscript, we will better introduce and identify how and where these different concepts apply to, inform, and are answered by our research.

The goal of this work was to assess how metabolism responded to and recovered from higher flow events that were also isolated flow events. Indeed, this decreased the number of suitable events for analysis. But our choice of methods allowed us to focus on response/recovery to discrete disturbances and avoid biased comparisons of pre/during/post multiple high flow (but not isolated) events that encompass time periods that are long enough (e.g., weeks) where pre/post comparisons are less meaningful. Perhaps we could have selected a more pristine stream with less flashy hydrology at the start of this project, but another motivation of our work is to better understand processes in less pristine ecosystems (historically understudied because they are more challenging sites to obtain high-quality metabolism estimates from, another factor that decreased the number of events with appropriate data for our analysis). Despite having "only 15 events", most past analyses included a similar or fewer number of events (e.g, n=10 in Reisinger et al. 2017) over a shorter time period. Our work fills in substantial knowledge gaps: we analyzed across seasons (not only summer months or a short sensor deployment period) and high flow magnitudes (not only base flow or the highest flow disturbances).

After all appropriate QA/QC measures, we had 1375 days of metabolism estimates over 5 years (which were reported in full in O'Donnell & Hotchkiss 2019 Water Resources Research). To calculate resistance and recovery, we needed consecutive days of high-quality metabolism estimates, which further limited the number of high flow events appropriate for our analyses. For example, in 2016 there were 52 (out of 352) days with quality-checked sensor data that had a 50% flow change relative to the day prior. After looking at these 52 storms and selecting those that had 3 days before and 3 days after without any other flow events, we had 12 that were isolated. After quality-checking our metabolism estimates for all of those days, we had 4 high flow events from 2016 that passed all quality-checking steps required for this analysis.

/ end of GENERAL RESPONSE

REVIEWER 1:

General comments:

The authors analyse the effects of 15 isolated storm events on stream metabolism, focusing on subsidy-stress hypotheses and drivers of response and recovery. The authors make use of a five-year high-temporal resolution dataset to address their four hypotheses. While the general ideas and approach are interesting and worthy of study, the current manuscript requires major revisions before publication. Some of the suggested revisions are substantial changes to analysis and fall more into the "reject and resubmit" category, but the authors of course can provide adequate justification for not conducting these changes.

RESPONSE: We appreciate this reviewer's feedback acknowledging the novelty of our work, encouraging us to clarify our research objectives, and highlighting areas where we can better justify the methods used to address current knowledge gaps related to how ecosystem metabolism responds to and recovers from high flow events. It appears that placing our work in the context of subsidy-stress, pulsing baselines, and ecosystem resistance/resilience “muddied” our communication of research objectives, which we will clarify and focus in a revised manuscript. We highlight how we will step through the relevance of multiple ecological concepts to this research in our general response above. We respond to specific suggestions in more detail below.

Specific comments:

Major:

1. After reviewing the approach and the data, and because it underlies all results in the paper, I think the authors need to present a stronger rationale (which currently does not exist)âĂŤor change the approachâĂŤ for how they arrived at their use of arbitrary discharge thresholds (e.g., 50%, 10%) in discriminating the “isolated events”. An option could be a simple sensitivity analysis. Additional rationale should support the use of cumulative daily discharge, as opposed to other commonly used metrics in hydrology for event detection. I further think using cumulative discharge may be obscuring some results and indeed missing many events. I first started down this path of inquiry because only 15 isolated events over five years seemed to be a small sample size. Would the same thresholds result in different events if applied to depth or even comparing maximum daily discharge as opposed to cumulative daily discharge? This, I realize, may be a bit of a task because it requires an entirely new analysis, but I think that the authors need to consider this route and defend their assertions more fully. If more events could be used based on a simple adjustment like threshold choice, there could be a much more robust sample size to draw inference from, and would make this a much stronger paper.

RESPONSE: We responded to concerns about “only 15 isolated events” in our general response at the start of this document, but add a bit more detail here. Again, what we sacrificed in quantity we gained in quality. A full dataset for a single day required

Printer-friendly version

Discussion paper

quality-checked sensor data as well as metabolism estimates that passed all QA/QC steps, which means some isolated events were excluded from our analysis. We excluded values of physicochemical parameters that were below the 1% or above the 99% quantile and removed physicochemical measurements we knew were inaccurate due to sensors being out of the water during low flow or not working properly (e.g., turbidity at zero). This is covered in this text in the results and in greater detail in O'Donnell and Hotchkiss 2019. In our revised manuscript we will clarify how and why we selected isolated storms. The 50% change in flow for our high flow events ensured those events were indeed outside of a pulsing baseline flow. We defined a flow event as >10% change in Q when comparing the high flow changes to prior metabolic rates, as smaller changes in Q may still influence metabolism. In prior analyses leading up to those presented in our manuscript, we did test different thresholds of flow change or different discharge metrics, and settled on our current method because after exploring the trade-off between different delta Q thresholds, number of quality-checked events, and differences between ambient stream flow and higher flow events.

2. Similarly, I think this paper could be much stronger by including as many events as possible, regardless of whether they are "isolated". As discussed in the Introduction, the pulsing fluxes, whereas the presented approach discounts them. I understand that perhaps the authors were particularly interested in capturing "resilience" metrics, which may require a period of calm after the storm, so to speak. But, one can imagine a much richer analysis if, for example, the authors calculated some kind of "resistance" metric for as many events as possible, but parsing which ones were preceded by large events. And, for "resilience", the authors could still calculate the time to return to pre-(initial)event conditions, but just parse which of these "initial" events had subsequent events. Without much effort, the authors could even estimate the subsequent rate of events and its influence on "resilience". One can imagine a figure of, for example, ΔGPP vs. ΔQ where points are colored by their recovery time and sized by their subsequent rate of events. I mention these suggestions because the current methods seemed disingenuous in taking an arbitrarily "neat" approach to this potentially very

fruitful test of the pulsing paradigm. Another important point in this regard is how to take into account when a rain event occurs during the day. For example, if a rain event occurs at 23h, it seems like your approach considers its effect for the previous day, when it is probably more appropriate to consider its effect starting for the following day. (This is understood by the authors in their approach in section 2.5.1, but their simple correlation approach does not effectively get at this idea). The current approach also likely discounts many possible events for this reason. I am aware that this may be a big ask of the authors, and, if different routes are taken, I still suggest that they provide stronger rationale for the (apparently) arbitrary decision for identifying events.

RESPONSE: We like the idea of the reviewer's suggested plot of "ΔGPP vs. ΔQ where points are colored by their recovery time and sized by their subsequent rate of events", and will see if such a plot might help better communicate our results. We disagree that our decisions were "arbitrary" and we do not think it is appropriate to change our research objectives for this work because we could analyze the data for a different purpose or with a different set of QA/QC standards to gain more events. As the reviewer acknowledged, our focus was to understand metabolic resistance to and recovery from isolated high flow events. Furthermore, we were interested in assessing these dynamics in a hydrologically flashy stream draining a heavily modified landscape. We agree these other topics are very interesting and look forward to seeing more research addressing this type of proposed work, but they are not the aim of our paper. We will ensure our objectives are focused, well-described, and clearly justified in our revised manuscript. Again, a higher-level response to this and #1 above is in our overall summary at the top of this document.

3. In the same vein, the authors should provide some kind of justification or sensitivity analysis for both of their critical choices in calculation of their resistance/resilience metrics. The first is the choice of "...three days prior to define a range of antecedent metabolism for each isolated flow event." (Lines 140–141).

RESPONSE: This is a good question and one that we will further describe in our revised manuscript. Our choice of three days was the result of balancing best practices from published papers on similar topics (e.g,. 4 days prior stable baseflow, Reisinger et al. 2017) while still analyzing as many events with appropriately QA/QC'ed data as possible.

The second is the choice of defining "X prior [as] the maximum or minimum value of GPP or ER from the antecedent range..." (Lines 147–148). Why not the median or mean, which would represent more of the "equilibrium" of the previous period?

RESPONSE: We wanted to capture the pulse of days prior (Figure 1). The pulse isn't captured in a mean or median, and analyzing the metabolism data in this way can result in estimating a departure from a mean or median that is erroneously considered to be different from baseflow tendencies when in reality it is within the ambient pulse of the system. We will revisit where we highlighted this choice in our methods (we adopted this more conservative method for measuring magnitude of departure and recovery that is more appropriate for variable ecosystems) to ensure this decision is well-described in our revised manuscript.

And why not use the most similar previous day in terms of driving forces–in particular, light availability (this seems especially relevant for the recovery interval!). I'm sure the authors considered such options in their initial work, but they need to do more work to convince the audience of their presented approachâĂŤor take a different one if the evidence from sensitivity analyses suggests that they should. Both of these choices are major factors in the subsequent analyses because they define the metrics used, and because these choices do not appear to have literature support/precedent, they need to have clear rationale.

RESPONSE: While an interesting idea, this approach requires too many assumptions about our ability to predict GPP and ER based on light or temperature or other environmental data alone. While our knowledge of metabolism in streams and rivers is growing rapidly, we are not at a point where we can justify how we would select a "most

Printer-friendly version

Discussion paper

similar previous day" for comparisons between the range of baseline metabolic rates and responses/recovery from high flow events with the environmental data we have for this site.

4. Lines 154–156: "To quantify the resilience of GPP and ER, we estimated recovery intervals (RI) by counting the number of days until metabolic rates returned to within the range of pre-event values, signifying a return to antecedent dynamic equilibrium (Figure 3)." This is a good illustration of a potential issue/untapped possibility with the current approach. If you look at the data for the event shown in Figure 3, depth increases in that event by approximately 0.12 m, which decreases light availability by approximately 13% (according to exponential attenuation). This is nearly exactly the difference between GPP on 7 February and 9 February, both of which had nearly identical incoming light signals (making them comparable).

RESPONSE: This is an interesting observation that we will explore and include in a revised manuscript if deemed an appropriate way to discuss results within the context of our research objectives.

5. Lines 157–160: "To ensure additional flow events did not obscure the recovery interval of GPP or ER, we stopped counting RI the day before the next event (i.e., if another flow event happened four days later, we stopped counting RI at 3 days), and have noted this in our results as days+ and used different symbols in data figures." Why? As far as I can tell, the authors throw these data points out in their analysis, and only reference them in Table 3 (which already uses asterisks to note the issue). Is this to note that the system was on its way to "recovery"? Maybe it would be better to just show a recovery rate, instead of a time, which could result in more data points being included. So, instead of the time it takes to get back to some baseline (which I argue above is a bit arbitrary), you can calculate the rate of increase in GPP over a period (which could be equal to the baseline period that you settle on). Let's say an event occurs and on that day GPP was 5 g O2 m-2 d-1 ; the subsequent days maybe it's 8, then 10 g O2 m d . The rate of increase could then be 2.5 g O2 m-2 d-1 ((10-5) / 2).

Then, even if a subsequent event occurs, you can still compare the rate of increase before that event. A rate also seems like it could be more comparable/scalable across systems in contrast with a number of days. I don't presume to have the best idea here, but I think an approach like the one outlined above could increase inclusion of useful data points, and thereby lead to more useful inferences.

RESPONSE: This is a great point, and an idea we explored early on but did not include in our final analysis. However, seeing that there is interest in recovery rates in addition to recovery days, we will update our revised manuscript to include this metric. We will still include the number of days to recovery, when measurable (yes, the days+ is to note it was on track to recovery, but did not return to the range of baseline metabolic rates before the next high flow event). We will keep RI estimates to allow us to compare our results to other studies (e.g., Figure 8) for broader discussion.

6. Lines 165–166: "We assessed three categories of potential predictors of metabolic resistance and resilience: antecedent conditions, characteristics of the isolated flow event, and characteristics of the most recent prior flow event." Antecedent conditions and characteristics of the recent prior flow event (especially the latter) are unrelated to any stated hypothesis and appear to come out of nowhere. There needs to be clear rationale in the Introduction that leads the reader to understand why you are doing this.

RESPONSE: We will update the objectives and hypotheses outlined in the introduction to clarify our interest in estimating metabolic responses to flow change and potential drivers of those responses.

7. Generally speaking, I had difficulty with the entire Results section, which I think needs a complete rewrite. Some specific details are presented below, but I glossed over several in the interest of time. This section needs to link to stated hypotheses (in the order that they are stated in the Introduction) and test them directly without including spurious tests and weak assertions.

RESPONSE: Agreed. Thank you for the detailed comments below and the reminder to ensure the introduction and discussion are better aligned. We will revisit and revise with these comments in mind.

8. Figure 5 as presented is not informative. What do the authors want the reader to understand from this figure?

RESPONSE: We respectfully disagree that showing these results graphically is not informative, which is why it's cited numerous times in the manuscript. We want the readers to understand the relationship between recovery and magnitude of departure for GPP and ER, which this figure visualizes.

Is the R based on a linear regression for all of the points or just the black circles? What is the slope of the regression and the p-value? How does the slope compare to the 1:1 lines?

RESPONSE: R is based on all the points. We will clarify this in our analysis section and the figure legend. P-values are included in the text (lines 194 and 206). We will add them to the figure legend as well.

The second panel (right, ER vs GPP recovery interval) is not related to any stated hypothesis.

RESPONSE: While we do not agree that every plot or table is only justified if related to a specific hypothesis, we will clarify that in trying to understand how metabolism changes with higher flow (and why), we are interested in the differences between GPP and ER as well. This is included in Figure 1, but can certainly be highlighted better in the text as well in a way that justifies the inclusion of the second panel.

The text discussing this Figure does not support the points on the figure, particularly for the high stated value of ER stimulation = 0.22 (Lines 189–190: "...The magnitude of departure for ER (M ER) ranged from -0.59 to 0.22, with a median of 0.").

RESPONSE: We will check this mismatch between the plot and our text - it appears part of the y axis may have been cut. The revised manuscript will correct accordingly.

Looking at this figure also raises red flags about how the authors defined stimulation/repression. How do the extremely small changes in magnitude shown here compare to the uncertainty in GPP/ER, which are never discussed or propagated through any of these analyses? For example, is a 1% increase (i.e., M = 0.01) detectable if uncertainty is considered?

RESPONSE: We note that our metrics were indeed detectible relative to metabolism estimates (with low uncertainty, as shown in Fig 3, A4-A18, and supplementary data files). Consequently, we disagree that this approach “raises red flags”.

The authors should improve this figure substantially, or remove it/leave it as a table. One possible idea is to color or size points based on the event size. Moreover, based on this figure, I am not sure I believe the results on Line 193–195 (italics mine): “Although GPP exhibited stronger responses across isolated flow events than ER, M GPP and M ER were positively correlated (R 2 = 0.39, p = 0.007, Figure 5) and not significantly different (p = 0.06, α = 0.05).” Just an eyeball test makes this seem unreasonable. ER magnitudes on average are about 0.

RESPONSE: We will consider these additional comments and questions about this figure as we decide whether to convert to a table or update it for the revised manuscript.

9. Figure 6 is not easily understood and appears to simply repeat the information on Table 4 in a cluttered way. What key piece of information is the reader supposed to understand from this? The results of the controls on process response in this section 3.3 is quite difficult to connect with any prior hypotheses and leaves the reader uninformed. There are two figures and a table with only six sentences to describe them in this section. One of the stated hypotheses (H2) is never even formally tested here, and only the resistance metric is tested for H3 (somewhat, in Figure 7), which included both resistance and resilience metrics.

RESPONSE: We will move either Table 4 or Figure 6 to the appendix in our revised manuscript.

10. As far as the Discussion and Conclusion, I have many comments, but the issues all stem from previous issues relating to hypotheses, methods, and results. If the authors apply any of my suggested revisions to their approach, they will inevitably have to rewrite these sections. So, I have not provided many specific comments out of the interest of time, but a few key ones are here. I again suggest to organize the Discussion (and entire document) in order of the hypotheses as they are presented and as makes logical sense. As written, the Discussion jumps around in its assertions and ideas. Finally, much of the content in these sections is hypothetical and rhetorical, with little critical analysis of the results actually presented in the manuscript and how they relate to the broader literature.

RESPONSE: As stated above in response to similar comments, we will keep this in mind when we revise the manuscript.

Minor:

1. Ideas of pulsing steady state could be clarified a bit with regard to the study design and terminology throughout. In the Introduction, the authors note "Frequent disturbances generate oscillations that form a pulsing steady state (sensu Odum et al. 1995) that includes ambient variability in processes (Resh et al., 1988; Stanley et al., 2010)." (Lines 21–23). So, flow disturbance regime defines the pulsing steady state of lotic systems. But, the authors then use–incorrectly I think–the periods outside of flow disturbance to define a "pulsing steady state" (or at times, "pulsing equilibrium", like in Figure 1, and "dynamic equilibrium", like in Figure 3, and "antecedent equilibrium" on Line 187), to which they then compare to periods with flow disturbance. The approach is clear, but there is some circular reasoning with respect to the definition of pulsing steady state. I recommend perhaps using different terminology for these two concepts. One idea could be to use something like "ambient equilibrium" for metabolism under baseflow conditions, and "pulsing equilibrium" to refer to the larger scale, (inter)annual behaviour of as originally conceived by Odum. I think these small changes would improve the clarity of the study design and arguments within.

RESPONSE: We agree these small changes would improve clarify and will revise accordingly when we update our manuscript.

2. Similarly, I do not think that "resilience" is appropriately used throughout the manuscript, first defined by the authors on Lines 59–60: "We can also quantify post-disturbance ecosystem responses by estimating resilience: the time it takes for a process returns to equilibrium following a disturbance (Carpenter et al., 1992)." We have first of all the issue of "return[ing] to equilibrium", which is not so clear based on the previous definition of a pulsed equilibrium that includes disturbance. In a system organized by regular disturbance regimes, the idea of resilience to that same disturbance regime is a bit convoluted. In contrast, the idea of a "recovery interval" to previous ambient conditions is clear and appropriate. Resilience in this context might make more sense if there were alternative metabolic equilibria that the stream could occupy, where each of these equilibria were tolerant to different levels disturbance. Ultimately, this is a choice of language and does not affect the analyses presented and if the authors opt to keep their current choice, I suggest spending some more time to expand these ideas/defend their use out in the Introduction and Discussion.

RESPONSE: We agree, and will keep these comments about word choice in mind as we revise our manuscript.

3. Line 73: "(H0) some flow events will not push GPP and ER outside of their pulsing equilibrium." Should this by "(H4)"? Or is this some kind of null hypothesis? Consider renumbering, or placing this at the beginning of the sentence–seems strange to go from 1–3 then back to 0.

RESPONSE: We will re-order or change to H4 in our revised manuscript.

4. Lines 70–71: "...(H2) there will be a stimulation of GPP and ER at intermediate flow disturbances due to an influx of limiting carbon and nutrients...". Is this stream known to be limited by carbon and nutrients? What is the timeframe for stimulation? It seems like the influx of carbon and nutrients would pass through the system quite quickly in

this small stream, and would not be easily acquired/processed by organisms. In larger systems with long recession curves, I think this perspective can make sense, but this hypothesis does not seem well supported in the Introduction as currently written.

RESPONSE: We will update the introduction and site information to provide better context for this work. While the speed at which nutrients and carbon travels during storms will increase, the ability of microbes to respond to increases in carbon and nutrients is not limited to microbes in larger rivers (e.g., Demars 2019). And yes, the stream is carbon limited. We have unpublished data on carbon and nutrient limitation at our study site that we can mention as part of the site information in our revised manuscript.

5. Lines 71–72: "...(H3) metabolic resistance and resilience will change with the size of the event, with larger flow disturbances inducing more stress due to enhanced scour..." The point about scour here seems important. Scour is a function of shear stress, which itself is a linear function of depth. The authors focus on discharge as their subsidy/stress driver, but I wonder if water depth would be more appropriate? Because depth only increases to the square-root of discharge (for a large range of depth-discharge in their Supplemental data), a quadrupling of discharge only results in a doubling of benthic shear stress. I don't expect for the authors to redo any analyses with this perspective, but I do think this kind of information would be useful to include especially in the Discussion so that future works would consider this as well. It also could be used as a future framework to further test the idea of subsidy/stress balance. Depth is a first-order control on both light availability and shear stress at the benthos, making it a more appropriate indicator of stress than discharge.

RESPONSE: Thanks for this comment. We agree that is another useful way to consider these processes, and will include it in our revised discussion.

6. The light data (first referred to on lines 92–93) appear to be in units of μA according to the supplementary material ("ODonnellHotchkiss_SuppData_ReadMe.pdf", under

point “1”). I am not familiar with this unit (is it micro-amperes?) for sunlight, and I think this needs some clarification. The light data in the data file itself appear to range between 0 and 1, but the streammetabolizer model take data in PAR (units μmol m s), which can be upwards of 1000 by noontime. I'm sure this is not a major issue, but I do not think the results will be replicable as currently presentedâĂŤthose units, if fed into streammetabolizer, will lead to very strange outputs I think. The sensor used (according to O'Donnell and Hotchkiss 2019) is a Campbell CS300, which should output data in typical units like W m .

RESPONSE: The units for light do not matter because how we use light in the metabolism model (Eq 1) is a ratio of light at time i divided over the sum of light over the entire day: PARi / sumPAR. We just realized we need to correct this in our revised manuscript, as it's noted as PARt, not PARi, in equation 1. That said, we will check units and make sure they are the same throughout data files and the revised manuscript.

7. Line 140: “To acknowledge the pulsing, day-to-day variability...” I don't think “pulsing” is appropriate or needed here.

RESPONSE: We will remove pulsing and state “To acknowledge the day-to-day variability. . ..”

8. Line 152: “...suppression...” please check for the consistent use of suppression and repression (and others) throughout.

RESPONSE: We will check for consistent use of suppression and repression throughout.

9. In section 3.2 “Metabolic resistance and resilience”, it would be very helpful to explicitly organize/label these paragraphs according to your numbered hypotheses from the Introduction. For example, Lines 187–192: There is no directly stated connection between any of the statements presented here and the actual hypotheses.

RESPONSE: We will update the hypothesis section of the introduction to include discussion that is not just focused on S-S hypotheses.

10. Lines 194–196 bring up another issue with the idea of "magnitude" (italics mine): "M GPP was less than M ER for nearly all flow events, except for one in which M GPP and M ER were both zero and two where M GPP and M ER were both small (Figure 5, Figure 195 A19)." The general idea of magnitude is that is not directional. I would argue that the magnitude of GPP response was greater than that of ER, and that they both had similar directional change (decrease in process magnitude). Consider different language throughout.

RESPONSE: Thanks for this perspective. We will use appropriate terms related to magnitude and direction in our revised manuscript.

11. Lines 198–199: "Similarly, the only other event that stimulated GPP (M GPP = 0.03) had no ER response, suggesting many flow disturbances may decouple GPP and ER." This seems like an unsupported assertion (which should be in the Discussion, if anywhere) based on one event with an extremely small signal.

RESPONSE: We will refine the language in the results to better reflect the magnitude of responses/signals. We will move the "suggesting many flow disturbances may decouple GPP and ER" to the discussion with additional support for this statement from other results.

12. Table 3: n/a is not clearly defined.

RESPONSE: Thank you for catching this. We will define n/a in the updated manuscript. It is when ER or GPP did not deviate from the antecedent range, and therefore had no recovery interval.

13. Lines 208–209: "Although GPP and ER are linked processes, the variables that were moderate or strong predictors of resistance or resilience ($r > 0.5$)." Why is 0.5 the threshold for being a strong predictor? That's only 25% of the variance explained.

RESPONSE: We will revisit all descriptive wording used and make sure it matches the thresholds we set in our statistical analyses. We will cite references supporting our decision for these thresholds in the revised manuscript.

14. Lines 210–211: “Because the median RI ER was zero, bivariate correlations could not be used to determine potential predictors of ER resilience.” Another reason to consider rate instead of day count.

RESPONSE: As described above, our revised manuscript will include both.

15. Lines 214–215: “Overall, there were multiple environmental controls on metabolic resistance or resilience that were strongly correlated with either GPP or ER, but no significant drivers of both GPP and ER resistance and resilience.” This is not supported by the figure or the table.

RESPONSE: We will update to refer to Table 4, which supports this statement.

16. Line 219: “Notably, ER was more resistant than GPP (Figure 1).” Figure 1 is a conceptual figure and does not support this statement.

RESPONSE: We will remove the citation to Figure 1 and replace with a citation referring readers to Figures 5 and 7.

17. Line 239–240: “In assessing metabolic responses and recovery from smaller flow events relative to the dynamic equilibrium of metabolism at baseflow, we found some of the shortest metabolic recovery intervals recorded in the literature (Figure 8; Table A1).” Do these other studies use the exact same methodology as you? How are they comparable? Are they similar sized streams? You should compare and contrast more here.

RESPONSE: We will revise to include types of sites and landscapes in our discussion. The methods we used to generate this graph (from reported metabolism rates in papers) were standardized. We selected calculations/metrics that could make use of the most publications that included metabolism during high and low flows. Metabolism modeling methods were similar among projects.

18. Line 259–260: “Contrary to our predictions, the size of the most recent antecedent flow disturbance had a positive relationship with M GPP and M ER (Figure A19).” Where is this prediction?

RESPONSE: We will articulate this prediction more clearly in the introduction when we revise to highlight that the aim of this paper is not only to address subsidy stress hypotheses and differences between GPP and ER, but also to understand which environmental variables are related to how metabolism responds to and recovers from higher flow events.

Technical Corrections:

1. Equation 1 (Line 110) seems boiler-plate and unnecessary.

RESPONSE: While it is indeed a commonly used (but often slightly modified) equation, we respectfully disagree with the reviewer’s opinion that is unnecessary. We strive to ensure that our work is understandable and repeatable without requiring readers to visit many other papers to understand our methods.

2. There are extra parentheses in Figure 2c description for “((m d))”

RESPONSE: Thank you for catching this. We will remove the extra parentheses.

3. Figure 3 should describe what the error bars are on the GPP estimates.

RESPONSE: We will update our figure legend to define error bars for GPP.

4. Lines 163–164: “Quantifying how different antecedent conditions induce variable responses from GPP and ER is critical to furthering our understanding of stream ecosystem responses to flow disturbances.” This belongs in the Introduction, not the Methods.

RESPONSE: We will move this method justification to the end of the introduction where we talk about overall project objectives that are not limited to subsidy-stress analyses.

5. Lines 167–168: “Antecedent medians for turbidity were estimated from seven days

prior due to missing sensor data." This is not clear, please explain what this means. There was always missing data for turbidity within the three days prior to an event? I can't imagine turbidity changes very much at baseflow.

RESPONSE: We had to remove poor-quality data from the turbidity dataset and chose to set methods that would accommodate the most storms possible for our analysis. We compared the outcome of changing the days prior for events with turbidity data available for both 3- and 7-day analyses and found no difference in the results. We will include that information in our revised manuscript to clarify this difference.

6. Lines 190–191: "Three of 15 flow events stimulated ER, 5 repressed ER, and ER did not deviate from the antecedent equilibrium for 7 events (i.e., M ER was 0)." It's more common to use numerals for numbers greater than 10, and to spell the numbers out for numbers less than 10.

RESPONSE: We will write out lower numbers per journal guidelines.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-304, 2020.