Response to Anonymous Reviewer 2

We thank the reviewer for a comprehensive review. Below we provide our point-by-point replies (black color) to the reviewer comments (blue color). New or modified text in the revised manuscript is presented in italics.

1. The paper aims to provide explanation for deviations from long-term C-Q behaviour for different types of hydrological conditions. The authors claim that they are first in doing so, but the only novel thing in this study is a large number of catchments that are investigated. The discussion and implications are pretty much the same as in other studies by the research team, highlighting the incremental character of this study. Thus, to grant the publication of this paper, the authors need to convince the readers about novelty of their work, in light of recent publications in this field.

We apologize that the novelty of our study was not clearly highlighted in the manuscript. The main novelty of our study is in combining the hydrological event classification framework with long-term low-frequency data. To the best of our knowledge runoff event characteristics to explore nitrate dynamics were previously mostly considered in high-frequency studies across individual or few catchments (e.g. Bauwe et al., 2015; Knapp et al., 2020; Heathwaite and Bieroza, 2021), while only Minaudo et al. (2019) and Pohle et al., (2021) considered hydrological conditions using a large sample of catchments and low-frequency data. Combining the information about the hydrological events at the time of sampling with low-frequency data enabled to find systematic deviations in the long-term C-Q relationships induced by different hydrological conditions in a large sample of catchments. Moreover, our large dataset of catchments with contrasting characteristics allows for deducing mechanisms behind the spatial variability of nitrate C-Q deviations across German catchments. Finally, the abundance of low-frequency data worldwide and the transferable nature of the applied event classification framework paves the way to further applications in contrasting environments to better understand scatter in long-term C-Q relationships.

To clarify and highlight the novelty of the study we propose to add/modify the following lines of the manuscript.

Low-frequency data combined with runoff event classification in a large number of contrasting catchments

Abstract L18: “This study combines a hydrological runoff event classification framework with low-frequency nitrate samples in 184 catchments to explore the role of different runoff events in shaping long-term C-Q relationships and their variability across contrasting catchments. ”

Introduction L48: “The scatter of C-Q relationships might also be related to hydrologic conditions at the time of sampling (Knapp et al., 2020, Musolff et al., 2021), which are investigated for a large number of catchments only by a few recent studies (Minaudo et al., 2019; Pohle et al., 2021).”

Spatial variability of C-Q deviations across contrasting catchments

Abstract L25: “Using long-term, low-frequency nitrate data we demonstrate for the first time for a large set of catchments that runoff event types shape observed scatter in long-term C-Q relationships according to the level of hydrologic connectivity characteristic of each runoff event type. In addition, we hypothesize that the level of biogeochemical attenuation taking place in catchments can partially explain the spatial variability of the scatter during different event types.”

Conclusions L452: “Moreover, we inferred using catchment descriptors physical mechanisms that possibly explain the spatial variability of this scatter.”
Systematic deviations of C-Q relationships linked to the runoff event types

Introduction L96: “Our study aims for the first time to investigate the presence of systematic deviations in long-term C-Q relationships produced by different runoff event types in a large set of catchments.”

Results L271: “We found systematic differences in the direction and magnitude of deviations of nitrate concentrations (∆res50) from the long-term C-Q relationships during different types of runoff events despite the large variety of study catchments (Fig. 5).”

Transferability of our methods

Discussion L440: The abundance of low-frequency data worldwide and the transferable nature of the applied event classification framework provide the means for further applications in contrasting environments to better understand the origins of scatter in long-term nitrate C-Q relationships.

2. I understand that the authors want to show off the contributions from their own team, but there are plenty other papers, not published by your group, that you could refer to in your discussion. We apologize if we have overlooked relevant references in our manuscript. To show contributions from a larger number of research groups, we will modify the cited papers by adding or replacing references. We propose the following changes:

L33: Moreover, due to long-lasting legacy effects a delay in reducing riverine nitrate concentration was reported in many catchments (Tesoriero et al., 2013; Meter and Basu, 2017; Bieroza et al., 2018; Chang et al., 2021).

L38: The shape of C-Q relationships encodes export patterns and reflects the temporally varying quantities of critical substances such as nutrients delivered to streams (Godsey et al., 2009; Meybeck and Moatar, 2012; Rose et al., 2018).

L43: Differences in long-term C-Q-relationships among catchments can be associated with differences in the availability and spatial distribution of solute sources (Musolff et al., 2017; Dupas et al., 2019; Zhi et al., 2019; Casquin et al., 2021), their hydrologic connectivity (Seibert et al., 2009; Dupas et al., 2016; Covino, 2017) and biogeochemical processes within the soil and stream that can retain or permanently remove nitrate from streamwater (Mulholland et al., 2008; Dupas et al., 2016; Moatar et al., 2017; Benettin et al., 2020).

L74: At seasonal scale nutrient transport to streams can be increased with higher hydrologic connectivity in catchments with abundant sources (Martin et al., 2004; Veith et al., 2020; Guillemtot et al., 2021).

L285: We argue that during snow-impacted events hydrologic connectivity of sources is high due to elevated wetness conditions (Stieglitz et al., 2003) which is consistent with previously reported high nitrate concentration during the winter period (Martin et al., 2004; Ocampo et al., 2006; Yang et al., 2018).

L299: Rain.dry.uniform and Rain.dry.patchy events occur more often during the dry season when nitrate concentrations are reported to be lower in several studies (House et al., 2001; Guillemtot et al., 2021).

L384: Deep sedimentary aquifers have a high potential for denitrification due to a great availability of electron donors, longer transit times, and more anoxic conditions due to sufficient reduction capacity (Kunkel et al., 2004; Wendland et al., 2008; Knoll et al., 2020) producing a lower nitrate supply in deeper soils compared to shallow soil (Dupas et al., 2016).

L400: Many studies have highlighted the importance of agricultural sources for nitrate export patterns in several catchments (e.g., Moatar et al., 2017; Minaudo et al., 2019; Casquin et al., 2020; Weber et al., 2020).

L423: A reduction in the frequency of snow-impacted events was already shown in Germany over the last decades (Fontrodona Bach et al., 2018; Chan et al., 2020; Taszarek et al., 2020).

Specific comments

3. Line 16 grammar

Thank you for this suggestion. We will modify the text.
Although previous studies investigated the origins of this scatter in individual or in a few catchments, the role of different runoff event types across a large set of catchments is not yet fully understood.

There are in fact considerable differences between Winter et al. (2021) and this manuscript. The work of Winter et al. focuses on the variability between runoff events during a limited 4-years period considering only samples taken during runoff events and using high-frequency data in only 6 neighboring catchments in Central Germany. The study finds that variability of hysteresis patterns decreases from runoff events induced by rainfall with dry antecedent conditions to snow-impacted events. In contrast, this manuscript uses low-frequency across 184 catchments data and investigates the effect of runoff event types in long-term C-Q relationships. The increase of nitrate concentration during snow-impacted and the decrease decrease during rainfall events with dry antecedent conditions of Winter et al. (2021) is also confirmed in our work. However, a much larger number of catchments with contrasting characteristics used in this study allow us to investigate systematic nitrate deviations from long-term C-Q relationships across catchments and attribute spatial patterns of deviations to potential physical mechanisms using catchment characteristics. We will modify L16 mentioned by the referee to clarify the differences between the two studies.

We apologize if we have overlooked relevant papers on the scatter in C-Q relationships due to the different storm responses. We have added now additional references to the Introduction.

The cause of this scatter can also be traced to a variety of responses observed at the event-scale in several studies with high-frequency data in single or a few catchments (e.g., Bowes et al., 2015; Lloyd et al., 2016; Koenig et al., 2017; Gorski and Zimmer, 2021).

Disparate patterns of the event C-Q relationships in a catchment over time are mainly attributed to varying dominant flow sources (e.g., groundwater, shallow subsurface flow), antecedent wetness conditions (Inamdar et al., 2006; Vaughan et al., 2017; Knapp et al., 2020), time of fertilizer application (Bowes et al., 2015; Dupas et al., 2016; Outram et al., 2016), biogeochemical cycling (Heathwaite and Bieroza, 2021) and runoff event characteristics or types (Butturini et al., 2006; Bauwe et al., 2015; Chen et al., 2020; Knapp et al., 2020; Heathwaite and Bieroza, 2021).

In contrast, negative deviations occur mostly for rainfall-induced events with dry antecedent conditions, indicating the occurrence of lower nitrate concentrations in river flows than their long-term pattern values during this type of events.

It is not clear if you analyse high-frequency or low-frequency C-Q data, this should be clarified at the very beginning of the paper. Without this information it is difficult to judge the quality of your hypotheses.

Thank you for pointing this out. We will modify the text to clarify this issue:

“Our study relies on low-frequency nitrate data, which is often used to build long-term C-Q relationships (e.g. Cartwright et al., 2020, Diamond and Cohen 2018). However, studies with high-frequency data...”
found large variability in the C-Q patterns during events (e.g., Knapp et al., 2020; Dupas et al., 2016; Vaughan et al., 2017) that might add scatter to the long-term C-Q relationship.

6. Figure 1 should be part of methods or results but not introduction

Thank you for this suggestion. We will move the figure to the Methods section.

7. Hypothesis 1 is not clear. Do you mean individual C-Q points?

Thank you for pointing this out. To clarify this we will modify it as follows:

“I Do samples collected during different event types deviate differently from the long-term C-Q relationships observed at the catchment outlets?”

8. Not clear how daily discharge data can provide information about short storm events with duration of hours?

Thank you for your question. We use daily streamflow to identify events. This implies that the shortest event that can be captured has a duration of at least 1 day. Any event shorter than 1 day cannot be captured with the available data. We will modify the following sentence in Line #135 for clarification:

“The method includes baseflow separation, precipitation attribution and an iterative procedure to adjust site-specific thresholds for the refinement of multi-peak events. We use daily streamflow data to identify events. This implies that only events longer than 1 day are captured.”

9. In this sense, using a term ‘event classification’ is misleading. I would rather use classification of ‘hydrological conditions’.

We prefer to keep the term “event classification” instead of “hydrological conditions” as the former more accurately represents the information combined in the event types and is a standard in the hydrological literature (e.g., Bauwe et al., 2015; Chen et al., 2020; Ross et al., 2019; Xie et al., 2019). Apart from information on hydrological conditions often used (i.e., wetness conditions) it also includes information on the nature of precipitation events and spatial distribution of soil moisture.

10. Since you have low-frequency samples they are sampled randomly over the hydrograph. So samples that belong to the same hydrological condition can have been sampled on a rising, falling limb of the hydrograph or baseflow conditions. Thus, some of your scatter in each hydrological condition group can be attributed to when on the hydrograph your samples were taken. Please clarify.

I have just noticed that Reviewer 1 expressed similar concerns regarding the role of C-Q hysteresis. This is a key weakness of your approach.

We appreciate the reviewer’s comment on the possible effect of hysteresis at the event scale. We agree that this effect requires additional attention in the manuscript. The first reviewer have raised a similar concern, therefore below we repeat our response to reviewer 1 (Comment 1).

We quantified the proportion of samples taken in the rising limb, falling limb and near to the discharge peak (near-to-peak) of the event hydrographs. The rising limb starts at the beginning of the runoff event and finishes one day before the day of the peak discharge. The falling limb starts one day after the day of the peak discharge and finishes at the end of the runoff event. The beginning and the end of the runoff events are obtained from the runoff event detection method explained in detail in the original manuscript (Lines 132-135). We defined near-to-peak as samples collected from one day before to one day after the day of the peak discharge. We allowed some overlap between near-to-peak and other two groups to use a larger number of samples than considering samples collected on the day of the peak of discharge only. Of the total samples taken during runoff event types 34% correspond to the rising limb, 55% to the falling limb and 30% to near-to-peak (11% of the samples were collected during the day of the peak discharge). This information will be shown in Figure S6a in the revised manuscript. In addition, we quantified the
deviations of the long-term C-Q relationship ($\Delta_{res50}$) for samples taken during the rising limb, falling limb and near-to-peak. We computed the deviations for these three groups of samples following the same bootstrapping procedure shown in the Method section (Lines 165-172) of the original manuscript.

The new results provided in Figure S6b show that the deviations from the long-term C-Q relationships for different event types are very similar for all three cases (samples taken during falling limb, rising limb or near the event peak) and resemble the deviations that we have previously observed for all collected samples (Figure 5 in the main manuscript). This suggests that the relative time of sampling during an event does not affect deviations from the long-term C-Q relationships that we detected for different event types.

![Figure S6](image)

**Figure S6.** a) Number of samples per catchment per event type corresponding to the samples taken during the rising limb, falling limb, or near to the peak (i.e., samples taken from one day before to one day after the peak of the hydrograph). b) Median deviations of nitrate concentrations from the long-term C-Q relationships ($\Delta_{res50}$) for samples taken during the rising limb, falling limb, and near to the peak. Deviations are computed analogously as for Fig. 5 in the main manuscript. The three first columns of the heatmap correspond to one of the long-term export patterns (i.e., dilution (slope b<0), neutral (slope b~0), and enrichment (slope b>0)), and the fourth column corresponds to all study catchments. Bold font and * indicate significant differences (Kruskal-Wallis test, p<0.05) between median deviations across catchments for each event type and median deviation across catchments of all nitrate samples. At least 5 catchments with sufficient data (more than 10 samples per event type) are required to evaluate the significance of the deviations. Gray squares indicate cases where this requirement is not met.

We will insert the following description in the revised manuscript in the Method section.

L181: “Low frequency datasets such as the one used in our study might contain samples collected during different phases of the event hydrograph (e.g., falling or rising limb). This might hamper the interpretability of the results due to possible bias in observed nitrate concentration linked to the time of sampling and the hysteresis effect revealed in high-frequency observations (e.g., Lloyd et al., 2016; Vaughan et al., 2017). In fact, Pohle et al. (2021) showed systematic differences in nitrate concentration between samples collected during rising and falling limbs for numerous catchments in Scotland. To understand the potential effect of the hysteresis on the deviations from long-term C-Q ($\Delta_{res50}$) we repeat the bootstrapping procedure described above considering samples collected during the rising limb, falling limb and near the event peak (near-to-peak). The rising limb of a runoff event starts at the beginning of
the event and finishes one day before the day of the peak discharge. The falling limb starts one day after the day of the peak discharge and finishes at the end of the runoff event. Moreover, we defined near-to-peak as samples collected from one day before to one day after the day of the peak discharge. Of the total samples taken during runoff event types 34% correspond to the rising limb, 55% to the falling limb and 30% to near-to-peak. Notice that definition of near-to-peak samples allows some overlap with the other two groups of samples to use a more balanced number of samples than considering samples collected on the day of the peak of discharge only (only 11% of the samples were collected during the day of the peak discharge). “

We will add the following lines in the Result section.

L236: The time of sampling in runoff events did not interfere with our main results (Fig. S6b). Although some data limitations for some group of samples (gray tiles in Fig. S6b), we could reproduce the analysis for most of the cases. We found that similarly to our results using all the samples (Fig. 5b) values of Δres50 for samples taken during the rising limb, near to the peak and falling limb, are positive for Rain.on.snow and Mix events, negative for Rain.dry.patchy and Rain.dry.uniform, and intermediate for Rain.wet events.

We will add the following text discussing the results of the additional experiment to the Discussion section of the original manuscript.

L443: Although the presence of the hysteresis effect might considerably affect nitrate concentration during rising and falling limbs of the event hydrograph in some catchments (Pohle et al, 2021) we found a similar direction of deviations from the long-term C-Q relationships when we considered samples taken during rising limb, falling limb and near to the peak (Fig S6b). Hence, our results suggest that the variability potentially added by the presence of hysteresis patterns is lower than the deviations observed for different event types from the long-term C-Q relationship. Increasing availability of high-frequency datasets coupled with new statistical modeling approaches might be used in the future to evaluate hysteresis-related effects in the existing long-term C-Q datasets to further disentangle inter- and intra-event variability of nitrate dynamics at larger scales.”

Referencenes:


Vaughan MCH, Bowden WB, Shanley JB, Vermilyea A, Sleeper R, Gold AJ, Pradhanang SM,


