Response to the interactive comment of Anonymous Referee #1 on

"Concentration-discharge relationships vary among hydrological events, reflecting differences in event characteristics" by Julia L. A. Knapp et al.

General comments:
This study examines the concentration-discharge (c-Q) relationships of several solutes at the scale of individual storm events and across the entire two-year study period in the Swiss Erlenbach catchment. The authors use the similarities and differences among solute c-Q slopes and intercepts to make inferences about the timing and hydrologic sources of streamwater in the catchment. The study also correlates the c-Q slope and intercept values with a wide variety of environmental controls to identify the most important regulators of solute transport within the Erlenbach catchment. I commend the authors for undertaking a sampling regime intense enough in both its frequency and duration to permit the development of such a unique dataset. High-frequency precipitation and stream sampling can be difficult during a single event, not to mention across multiple events. It is unfortunate that the nature of the Erlenbach catchment's hydrology was such that the study's sampling regime could not fully capture the c-Q responses across the complete hydrograph of the sampled storms. However, the authors made a convincing case for constraining their analysis to only the recession limb data, and they were mostly careful to not extent their inferences beyond where their limited dataset would allow them to go. I include the caveat "mostly" because I do question the description of a two-year, growing season only dataset as "long-term". This is perhaps a minor quibble, but I think it would be more accurate to describe the full dataset as "interannual" rather than "long-term". Overall, the authors present an interesting study of c-Q relationships at the catchment scale, highlighting important differences in biogeochemical responses observed across a range of temporal scales. This manuscript may be acceptable for publication in HESS, provided the authors address the specific comments outlined below.

We thank the reviewer for her/his positive assessment of our study. We understand the concerns regarding the term "long-term" for a two-year data set. The terminology was intended to more clearly separate observations based on the longer dataset from those of individual hydrologic events. Nevertheless, we understand that it may be misleading. In the revised version of the manuscript, we intend to use the term "multi-year" instead.

Specific comments:
L99: How long did the precipitation funnel sit out in the open prior to the onset of rain-fall? I am curious whether some of the solute concentrations in the early precipitation samples during events might have been biased by the addition of dry deposition on the funnel surface.

The collection and analysis of precipitation was automated, like the streamwater sampling. Therefore, the funnel was not cleaned and rinsed before every sample. However, rainfall events at Erlenbach occur roughly every second day, and the effect of dry deposition accumulating on the funnel was likely small. To account for the effect of dry deposition, we used (rainfall-) volume-weighted precipitation concentrations in our calculations (Tables 1 and 2). Small-volume samples at the start of a rain event with high solute concentrations due to dry deposition therefore had little weight in the calculation. Furthermore, our analysis is mostly based on the streamwater samples. The precipitation samplings procedure is described for completeness, and because the solute concentrations in precipitation are investigated as potential environmental drivers to explain the observed cQ slopes and intercepts (sections 2.6 and 3.4).

L121-122: Related to my previous comment, I'm curious how often it was the case that the first precipitation sample collected during an event was classified as an outlier. It might not affect your overall results much, but depending on how important a source dry deposition is in the study catchment, it might be worth considering. Also, I think it's important to provide additional details about the identification of outliers in the dataset. Was there a threshold that you set, or did you truly just "eyeball" the dataset? For example, how did you identify outliers in the case of solutes that had more variable (less tight) spreads? What proportion of the entire dataset was identified as outliers?

We excluded very few precipitation samples from our analysis, usually only if there was a known problem with the precipitation collector. Precipitation solute concentrations were generally very variable, even within an individual rain event. Generally, we observed a gradual decrease in solute concentrations as the rain event progressed. A possible explanation of this gradual decrease in precipitation solute concentrations could be due to the wash-off of dry deposition from leaves and the rain-out of aerosols in the atmosphere, and therefore likely reflect the normal processes occurring in the landscape.
Eliminating outliers in streamwater samples was mainly based on visual inspection. During the hydrograph recession, solute concentrations at Erlenbach generally change gradually from sample to sample (given our very high sampling frequency). Sudden jumps and outliers due to anomalous instrument behavior or sample handling were therefore often easy to detect. We furthermore compared among solute groups and instruments. Suspected anomalies in cation and anion data were compared across instruments, and if these anomalies were inconsistent across instruments, the respective data points were removed. We were generally very conservative in the removal of outliers.

L125 and L134: The word "aggregated" is kind of vague. Does this mean you averaged the data? Calculated the median?

Because streamwater sampling is instantaneous (not composite) in our field laboratory, we picked the closest 10-min data point for any quantity related to streamflow (discharge, EC in streamwater), air temperature and groundwater level. For the precipitation samples, the 10-min precipitation rates were cumulated from the start of the event or the last sampling time until the time of sampling. We will clarify this in the revised manuscript: "For the purpose of this study, only every second EC measurement was used to match the 10-min measurement frequency of river discharge" (lines 125-126), and "From the 10-minute discharge, air temperature, and groundwater level time series, we extracted those data that were closest to the sampling times of each streamwater samples. For the precipitation samples, associated precipitation amounts were calculated as cumulative sums from the 10-minute tipping bucket recordings." (lines 134-135)

L174-175: Did you consider how this flux index might be affected by accumulated dry deposition on your precipitation collector? This would be more important for some ions (e.g., nitrate, chloride, maybe sulfate) compared to others.

Acknowledged. It is indeed possible that dry deposition affected the flux budgets. However, dry deposition of soluble compounds *should* be part of the flux budgets, since they are real input fluxes that need to be taken into account if one wants an accurate picture of whether the catchment is a net source or sink of the solutes in question. In any case the effect should be small; given that it rains frequently at Erlenbach, the amount of dry deposition was probably relatively negligible (see explanation above).

L205-210: How many of the 30 events that you sampled fell into the category of being "not well constrained"? I can understand why you would want to limit your analysis to only those events for which the c-Q relationships were relatively straightforward, but this approach also kind of seems a little like "cherry picking" to me... From a practical standpoint, I completely understand the need to make such decisions about whether and when to exclude data (assuming they constitute a small percentage of the overall dataset) but I wonder if by limiting your dataset in this way, it also means that you're excluding some potentially important information about biogeochemical processes at event timescales. Those high RSEs are caused by something, and if they are attributable (even in part) to environmental and/or hydrologic drivers, there could be some very useful insights to be mined out of that variability.

Acknowledged. We excluded cQ relationships from events and solutes with high relative standard error (RSE) because they did not allow us to fit a power-law relationship, and thus would not allow for a consistent interpretation of the recession behavior. Nevertheless, we understand the concern that we may be excluding potentially important information. We thus redid the analysis with all calculated slope and intercept values (weighted by the inverse of the respective RSE) and we obtained a very similar result of the correlation analysis shown in Figure 5 of the manuscript.

The number of events excluded due to high RSEs was small for most solutes (usually none, and a maximum of 7 out of the 30 events), except manganese and copper (for which 13 and 17 events out of the 30 events had to be excluded because of high RSE, respectively). The generally low concentrations of these two solutes had high fractional variability, resulting in no clear power-law relationship for many events. We will include a note on this in the revised version of the manuscript.

L678-680 (Figure 4): Somewhat related to the previous comment: do you seen smaller uncertainties within events relative to the variability between events because you have removed from your analysis the storms with elevated RSEs?

This is a valid concern. We will update the figure (see new version below) to include the uncertainties in slope and intercept of all events, even those excluded from further analysis. As you can see, this mostly does not invalidate our statement that "the uncertainty in the estimated slopes and intercepts is mostly smaller than the variability between events, …". An exception may be manganese.

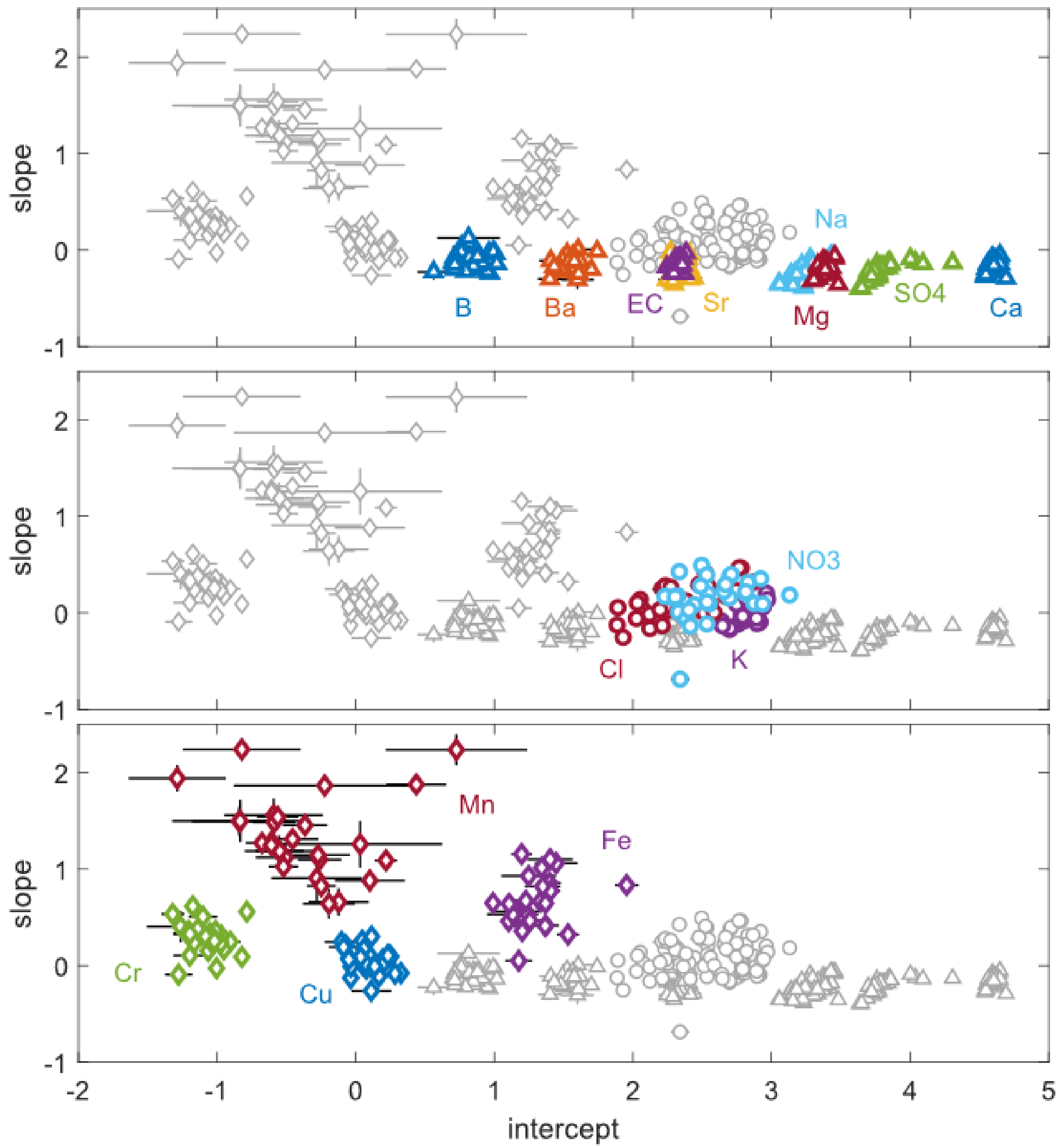

Figure 4 updated. Scatter plots of event cQ slopes and intercepts of the 14 different solutes and EC (error bars indicate one standard error). Solutes from different dominant sources cluster and exhibit similar ranges of variability. (a) Groundwater-sourced solutes cluster closely around similar slopes and show little inter-event variability in both slopes and intercepts. (b) Intercepts and slopes of solutes with significant atmospheric input (i.e., chloride, nitrate, and potassium) vary substantially among events. (c) The slopes of trace metals are generally higher than those of the other solutes (indicating predominantly mobilization behavior), and are also highly variable among events. The uncertainty in the estimated slopes and intercepts is mostly smaller than the variability between events, indicating that the observed inter-event variability in slopes and intercepts reflects real-world behavior rather than sampling and measurement noise.