

## **Response to the interactive comment of Reviewer #2 on**

**“A lab in the field: high-frequency analysis of water quality and stable isotopes in streamwater and precipitation” by Jana von Freyberg, Bjørn Studer and J.W. Kirchner**

*Comments of the reviewer are shown in italics.*

Responses from the authors are presented in regular font below each comment. Citations from the manuscript are in Times New Roman, changes of the text in the underlined.

General comment: In their work ‘A lab in the field: High frequency analysis of water quality and stable isotopes in streamwater and precipitation’ von Freyberg et al. present the set up and proof-of-concept of a new automatic sampling system for high- resolution measurements of stable water isotopes and stream solutes. The analytical components mainly consist of a laser spectroscopy system for stable water isotope analysis and an ion chromatograph for the measurement of anions and cations. Both instruments are joined by a sophisticated peripheral set up. Particular emphasis has been put into the reduction of carry over effects during operation when switching between the two water sources sampled.

The paper is clearly structured and easy to read. The system’s set up is impressive and the authors made great effort with a proof-of-concept. The precision achieved for both systems is excellent. I recommend to accept the paper for publication. However, I see some aspects of their set up that should be acknowledged in a revised version:

1. *The idea of a lab in the field is nice, but in fact the system requires a proper housing and full power connection. The system is far from being mobile and field deployable. To my understanding it is still a lab in a house (which is located in vicinity to a catchment).*

With the title of the manuscript we emphasize that a small-scale laboratory was set up at a field site next to a stream to allow for real-time, high-precision analysis of water samples. We do not claim in the manuscript that the analysis system is mobile or field deployable. All instruments were housed in a wooden hut, which was not temperature regulated, and required an area of only around 3m<sup>2</sup>. Thus, the analysis system could potentially be set up at any other location with road access and sufficient power supply. In the revised manuscript, we will include a photograph of the analysis system in the field to better illustrate the compact setup in the hut.

2. *Automatic wet chemistry instrument set ups are routinely operated by water provisioning services (e.g. in sewage treatment works) and by larger state environmental agencies or research facilities for surface water monitoring. Such systems are expensive (and therefore hardly used in basic research projects), but in place. See reviews from Gray et al. 2006, Environmental Chemistry, <http://dx.doi.org/10.1071/EN05059> or Bende-Michl and Hairsine 2010, Journal of Environmental Monitoring, DOI: 10.1039/B910156J.*

We thank the reviewer for these additional references, which we can implement into the revised manuscript.

3. *Even though the authors achieved a reasonable high sampling frequency of 30 min this set up does not allow to capture the potentially very high temporal change of stable water isotopes in*

*precipitation. This short term variation (that is missed by aggregating to 30 min composite samples or by sampling only the last precipitation sequence of a 30 min period) might be responsible for some of the differences found for the chemical- and isotope-based hydrograph separations. For high temporal resolution application see results in Pangle et al. 2013, already cited in the paper, or Moerman et al. 2013, Earth and Planetary Science Letters, <http://dx.doi.org/10.1016/j.epsl.2013.03.014>. I do not fully understand why the two sampling systems for isotope and chemical compositions are not operated independently, but parallel. In a synchronized, parallel set up, the systems could sample stable water isotopes independently in the order of minutes utilizing the full power of the continuous water sampler while (composite or grab) samples for the ion chromatograph are sampled in 30 min resolution.*

We agree that a higher sampling frequency would be desirable to resolve the isotopic variability in precipitation, however, our system was developed to measure many precipitation events in order to identify changes short-term catchment hydrological processes over time periods of weeks and months. In the study mentioned by the reviewer, Pangle et al. (2013) only sampled one event at 34min intervals. Thus, our study provides a more detailed insight into the variability of precipitation isotopes simply because more events were captured. Moerman et al. (2013) also sampled only one event at 1-4min intervals. This study also acknowledges that more events would have to be sampled at high frequencies to confirm the representativeness of the monitored event.

4. *The current set up is limited to two water sources. Given the set up and limitation of 30 min sampling frequency the system is only partly extendable with regard to sampling additional sources such as groundwater, tributaries or nearby standing water bodies. The sampling of additional sources, however, is needed to partially resolve the differences in the various hydrograph separations outlined in Chapter 5.*

We understand this point, and of course, given the sufficient resources, one could always install another isotope-IC system to measure other water sources. We clearly acknowledge this situation at the end of section 5.2, by pointing out that the age distribution of the pre-event water cannot be fully quantified with our data set, and thus, additional end-members would have to be determined: “Thus, the history of both events suggests that pre-event storage in this catchment was isotopically heterogeneous. This observation is unsurprising, given the pervasive heterogeneity of typical catchments, but a more detailed explanation is not possible with our spatially limited data set. Spatially distributed measurements, such as from groundwater and soil water storages, would help in constraining the individual end-members that contribute to streamflow (e.g., Hangen et al., 2001). Additional high-frequency time series of the groundwater table and soil moisture profiles would allow for documenting the effects of antecedent wetness conditions on the response times and on the activation of different storages at the site.”

#### **Specific comments**

5. *L16: Missing space between value and unit, check throughout the manuscript*

We will change that.

6. *L47 ‘much longer much smaller’ – I do not get what you intend to say.*

This was a typesetting mistake. We will correct that. “As pointed out by Kirchner et al. (2004), sampling at intervals much longer than the hydrological response times of a catchment may result in significant losses of information. “

7. *L48-52 There are two new papers out by Aubert and Breuer 2016, PLoS ONE, doi:10.1371/journal.pone.0153138 and Aubert et al. 2016, Scientific Reports, DOI: 10.1038/srep31536, that explicitly show the value of high resolution measurements of nitrate, might be worth considering here (or at least to have a look at).*

We thank the reviewer for this suggestion and will consider implementation of the references into the revised manuscript.

8. *L101 You sampled an internal standard every 3 h (see Line 280). So why did you not use these standards to correct absolute isotope values?*

The results presented in the original manuscript used only the factory calibration of the isotope analyzer. We have now used the internal standard to calibrate to reference standards, and will present the calibrated values in the revised manuscript, along with a more detailed description of our calibration and drift correction procedures.

9. *L246 An installation in a hut with power supply is not an installation in the field as the title of the paper indicates. At least not to my understanding of a field deployable system. I suggest to revise the title.*

We believe that the title concisely and accurately describes the contents of the manuscript. Our paper describes what is literally a laboratory in the field (with, as in most laboratories, power and communications and protection from weather). The point of the paper is to examine the relative merits of putting the laboratory out in the field, rather than transferring samples from the field back to a central laboratory far removed from the field site. Nowhere do we claim that the instrumentation package that we describe is a "lab in a box" or a "field deployable system" in the sense suggested by the reviewer. Instead, the manuscript quite explicitly describes the level of infrastructure that is required, and the tradeoffs involved vis-a-vis more conventional field sampling approaches.

10. *L257 I do not agree that the correlation of daily precipitation is good and therefore allows to derive subdaily precipitation patterns. We know how variable precipitation can be in space. As almost 20% of the variance of daily sums cannot be explained by the correlation, I can only speculate on the potential differences in hourly or even higher frequency precipitation rates. For future application of the system, I strongly suggest to add an independent met station to the field lab, which is negligible in costs compared to all the other instruments and peripherals used to set up this lab.*

In contrast to weather conditions in summer, when local convective storms cause highly variable, intermittent precipitation rates over time and space, the weather in northern Switzerland during winter is mainly characterized by large-scale frontal events. This results in longer, less variable and strongly autocorrelated precipitation events (Molnar and Burlando, 2008), which are likely to be captured by several meteorological stations in close vicinity of our field site. Unfortunately, the only rain gauge at our field site was not heated, and thus did not capture precipitation rates during snowfall periods. For this reason, we rely on data from the MeteoSwiss network, which used heated rain gauges. In order to account for the uncertainty pointed out by the reviewer, in the future version of the manuscript we will utilize hourly precipitation rates from the MeteoSwiss CombiPrecip dataset. This dataset is based on a geostatistical combination of radar estimates (1km<sup>2</sup>-resolution) and rain-gauge measurements, to reduce uncertainties due to spatial variability of precipitation patterns.

Reference: Molnar P, Burlando P. Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland. *Water Resour Res.* 2008; 44(10).

- 11.** *L264 Is there any reason to use a large bucket of 10 L rather than a substantially smaller container to sample from? Or utilize an even larger container to produce composite rather than grab samples.*

We used a long and narrow bucket with a siphon at the outlet. Fresh streamwater continuously flowed in from the top of the bucket and left the bucket through the siphon outlet at the bottom. Thus, sediment and organic material settled down and a less turbid streamwater sample could be retrieved from the upper part of the bucket. This configuration helped to slow down the clogging of the filters substantially.

- 12.** *Chapter 5 As the paper is mainly a description of the technical set up, I wonder a bit on the extensive evaluation of the event/pre-event water contribution in this chapter. While I see this a very interesting aspect of the utilization of the system in a fully operational application, I think this section can be reduced for this more technical paper. In fact, the discussed potentially missing end member to better understand the observations (groundwater, soil water, spatially variable precipitation) reflects the limitation of the current set up where only two water sources can be sampled in high temporal resolution (see also general comment, bullet point (4)).*

We agree with the reviewer and try to will shorten Chapter 5 in the revised manuscript.

- 13.** *L423-483 The difference of the isotope and chemical tracer derived hydrograph separation are enormous. The dissimilarity of results is so large, that the overall applicability of the approach seems to be questionable. As the authors stress, it goes beyond the scope of this paper to resolve the issue and further end members need to be identified to explain the observations. Thus, the discussion remains at least partly speculative. The current set up of the system does not allow to sample more sources in high resolution. To this end, the system described in the paper is limited to investigate more simple mixing processes of two end members, rather than complex mixing processes typical for catchments. Part of the observation might be due to the non-conservativeness of the chemical tracers. At least for NO<sub>3</sub>- I doubt its feasibility to be used as a real tracer, particularly in highly biological reactive top soils. Another potential explanation is variable source areas and their connectivity to the stream, with source areas being switched on/off during events.*

The point of this section of the paper is precisely to demonstrate that reactive chemical tracers (including electrical conductivity) are unreliable for separating event and pre-event water in the hydrograph. It is widely understood that this can be problematic, but our data clearly show how problematic this indeed can be. We make this point because we are frequently asked to review manuscripts that use electrical conductivity for hydrograph separation, apparently unaware of the problems that this approach poses. Our point is explicitly not that we have identified all the end-members that one would need to sample to more fully characterize the different sources of "old water" within the catchment.

- 14.** *L444 revise expression: moisture cannot be mineralized*

We will change that.

- 15.** *Chapter 5.4 This section misses a real discussion. There are at least a few papers that look into the effect of sampling frequency on hydrograph separation or mean transit time estimation that*

could be discussed here. Stockinger et al. 2014, <http://dx.doi.org/10.1016/j.jhydrol.2016.08.007>; Timbe et al. 2015 doi:10.5194/hess-19-1153-2015; Birkel et al. 2012 DOI: 10.1002/hyp.8210; Inamdar et al. 2013 DOI: 10.1002/wrcr.20158.

We thank the reviewer for the suggested references, which we will include into the discussion section of chapter 5.4:

“Thus, sampling at longer time intervals increases the risk of missing this critical peak response; if the sample is taken before or after the maximum hydrochemical response, the event-water signal in streamwater ( $C_s$ ) may be too weak, which will inevitably underestimate event-water fractions, or even lead to unrealistic negative values. Capturing the short-term responses of environmental tracers also helps in better quantifying transit time distributions (e.g., Birkel et al., 2012; Stockinger et al., 2016; Timbe et al., 2015) and in constraining concentration-discharge models (e.g., Stelzer and Likens, 2006; Jones et al., 2012).”

16. L514 I do not agree that you can conclude from Fig 3 that a 3 h sampling frequency would be sufficient. Maybe you can conclude this from Fig 11. However, in L516 you note yourself that even intervals of 3 h can result in a significant loss of information.

We will clarify the statement: “Figure 10 shows that 3h sampling intervals would still be sufficient to capture the major isotopic responses in streamwater, including during low-intensity precipitation events. However, there are also several storm periods (e.g., Events #7 and #8) during which the short-term variability cannot be resolved at this lower sampling frequency.”

17. L566 Despite the technical achievement and the effect of the high frequency sampling, the authors could stress even more the highly uncertain results of their hydrograph separation (even though this should not be the major focus of the paper). Combining the results of Fig 9 and Fig 11 I conclude that whatever model you have, it could easily be verified by an ‘appropriate’ selection of solutes and their sampling frequency, as the uncertainty to derive a ‘true’ hydrograph separation is very large.

A ‘true’ separation of the hydrograph into event and pre-event water was carried out by using high-frequency measurements of stable water isotopes, which are considered to be ideal conservative tracers. As pointed out in the manuscript, large uncertainties can be explained by the variability of the isotopic signal in precipitation (i.e., isotope values in precipitation vary by much more than the analytical uncertainty of the instrument).

Hydrograph separation based on solute data might have smaller uncertainties (due to smaller variability of solute concentrations in precipitation), however, event-water fractions can be largely over- or underestimated. As we show for Events #1 and #2, for instance, the results are not consistent, and thus the selection of ‘appropriate solutes and their sampling frequencies’ is not at all straight forward. The key point is that one should not be free to separate the hydrograph onto event and pre-event water based on reactive tracers, since they give results that are inconsistent with the isotopes, which are nearly ideal tracers.

18. Chapter 6 (or elsewhere in the discussion). After I read the conclusions, I wonder why the authors have decided to include an ion chromatography system that limits really high resolution measurements and therefore also limits sampling further water sources. At least for some of the solutes measured, ion selective probes or UV hyperspectral photometry sensors (reviewed by Rode et al. and already included in the paper) could be used as an alternative analytical system with substantially lower costs as well.

We agree that ion selective probes or UV hyperspectral photometry sensors are cheaper but measurement stability and accuracy are also substantial issues with such instruments. Of course, there are many alternatives of how an on-line system such as ours could be set up, however, each one has advantages and disadvantages.