Response to the interactive comment of Reviewer #1 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.

1. *General comment: This paper describes the development and performance of an advanced high-frequency analyser of both water isotopes and major ions in stream flow and precipitation. The authors provide a thorough account of the instrument design and operation and an assessment of analytical performance. The integration of the many components represents a significant engineering effort. The instrumentations analytical data quality is impressive, in particular the precision of isotope analysis. The instrumentation is described as a 'lab in the field' and high quality data can certainly be produced in real time during extended deployments. However, I question whether it can be described as a true field instrument considering its limited portability, multicomponent complexity (Fig. 1) and relatively high requirements for power and shelter. The field deployment described was in an outbuilding of the research institute, presumably with mains power but it is unclear what environmental conditions the instrument was exposed to (e.g. range of temperature fluctuations).*

   General response: we thank Dr. Niels Munksgaard for his thoughtful comments on our paper, and for developing his diffusion-based water isotope sampling system, which was the inspiration for the continuous water sampler that was used in our work.

   The terminology "lab in the field" indicates the use of analytical instruments that are usually employed in a laboratory environment (i.e, isotope analyzer Picarro L2130-i, ion chromatograph). Thus, the title makes clear that a small-scale laboratory was set-up at a field site to allow for real-time water sample analysis at high-precision. Nowhere did we claim that this was a "lab in a box" or a "field instrument". We point out in the manuscript that substantial infrastructure (accessibility, power access) is required to run the lab continuously. A completely remote system powered by solar panels would be impractical because of the high power requirements.

   All instruments were located in a small, wooden hut without additional insulation. The outside air temperatures, which are also shown in Figure 5a, were on average 2°C and ranged between -4 and 17°C. In the hut, temperatures were generally around 14°C and ranged between 7 and 23°C, because of the heat produced by the instruments. Humidity in the hut was around 40%.

2. *A photograph of the actual setup used would be a useful addition.*

   A photo of the setup will be added to Fig.1.

3. *The last 1/3 of the paper (Section 5 - comparison of event water fractions) is concerned with the interpretation of the month-long field trial. This section seems somewhat too long given that the main aim of the paper (as per the introduction) is the description of the development and field*

trial of the instrument (these are adequately described through sections 1-4). Moreover, the interpretation provided in section 5 is somewhat hampered by lack of data on important potential contributions to stream flow (soil and groundwater) as only two endmembers (base flow and precipitation) are considered. This limitation is clearly acknowledged by the authors. A substantial shortening of section 5 should be considered, although a shortened comparison of event-water fractions based on isotopes and ions should be retained as it provides a very good demonstration of the value of high frequency isotope measurement compared to the more traditional use of solute tracers in discrete samples taken at longer intervals. The manuscript is clearly written and the conclusions are sound and well supported by the data presented.

Sections 5.1 and 5.2 comprise only around ¼ of the whole manuscript (1905 of 7724 words), which includes the theory of HS, a detailed description of two contrasting (and thus interesting) events, and the conceptual description of runoff generation mechanisms at the site. We believe that these topics help the reader to understand how high-frequency time series of water isotopes and major ions can be used to study catchment hydrological processes. We will look for ways to streamline this section, while retaining its value for the reader (which is also recognized by the reviewer).

Specific comments:

4. *2. Methodology: For the laboratory based tests the analysis system was not calibrated as only relative isotope values were required - however, it is not clear if full calibration to the VSMOW scale or only drift correction was performed in the field tests – please expand on this (P 8 L280).*

The results reported in the manuscript were based only on drift correction using secondary isotope standards (Fiji and Evian bottled water) rather than calibration to the VSMOW scale. However, we have now calibrated these secondary standards to primary IAEA standards (SLAP, VSMOW, GISP), and field results in the final version of the paper will be both drift-corrected and calibrated to the VSMOW scale.

5. *Figures 5 and 6 display actual field data, e.g. in Fig. 5 data is shown relative to GMWL and LMWL so this comparison would require that full calibration was performed.*

As explained in our response to Comment 4., the final version of the paper will report values that are both drift-corrected and calibrated to the VSMOW scale, and thus will be comparable to the GMWL. Figures 5 to 8 will be updated accordingly.

6. *One of the limitations / uncertainties in the calculation of event-water fractions is (as stated by the authors) the precise definition of end member compositions. As the isotope composition (O and H) is often regarded as the most reliable tracer of event water it could be argued that the highest possibly frequency of isotope measurement of both stream and precipitation water should be prioritised. In this regard, it seems illogical that the measurement of isotope compositions was limited to 30 min intervals in order to synchronise data with the IC measurements which required this amount of time. This is particularly the case when a continuous water isotope instrument was used, wouldn't it be more beneficial to fully utilise its capability to perform truly continuous analysis, especially for precipitation which can vary significantly over much shorter time intervals than 30 min? At 30 min intervals, a 'conventional' CRDS instrument with a sequential injection/evaporation cycle could equally well have been used (apart from possible maintenance requirements). A similar auto sampling system was used for the IC and presumably this required regular maintenance.*

A "truly continuous" analysis is not possible with this instrument (or any other isotope analyzer that we know of) due to the memory effects within the instrument itself. As Fig. 3 clearly shows, significant memory effects persist within the sampler and analyzer for at least 10 minutes after the previous sample injection. Thus, sampling at 10-minute intervals (for example) would produce measurements that are strongly affected by the previous samples, as the isotopic signals overprint each other.

These memory effects might not be so problematic if we were only analyzing streamflow which changes relatively gradually. But instead, the sampling system must switch between rainwater and stream water during precipitation events, and thus sample carryover effects could lead to substantial distortions in subsequent calculations (such as hydrograph separations). We have chosen the 30-minute sampling interval in the interests of minimizing carryover effects, even if (say) 20-minute sampling would be potentially achievable.

7. *3.1 Optimisation of sample injection. . . The explanation provided for the delta dependence on hydraulic head difference (P 5 L 152 and Fig. 2) may not be the full explanation. The Picarro CWS uses a high air flow rate (possibly ≈500 mL/min as I recall) and this has the effect that the vapour generated is not in isotopic equilibrium with the liquid water flowing inside the ePTFE tube. The vapour is significantly depleted in d18O and d2H compared to equilibrium values due to the faster diffusion through the membrane of the light isotopologues compared to the heavy. This effect does not matter much as long as air and water flows and temperatures are kept constant for both sampling and calibration standards. However, the large fractionation effect probably tends to make the system prone to artefacts such as an altered pump rate. The use of a relatively cheap peristaltic water pump as opposed to the CWS supplied diaphragm pump would provide improved flow and lift characteristics (as does the expensive Dosimo pumps used in this study).*

Dr. Munksgaard's comment provides additional background to the explanation that we already give in the manuscript, starting on line 153: "Because the water is much colder than the surrounding air as it enters the membrane chamber, it is continuously warming as it travels through the membrane tube. At greater head gradients (and thus smaller flow rates), the sample will travel more slowly through the membrane chamber and will warm up more. As a consequence of higher water temperatures, water can be expected to diffuse more rapidly through the membrane and the resulting vapor can be expected to be less fractionated relative to the liquid phase (Kendall and McDonnell, 1998), as observed in Fig. 2." We can of course modify this explanation to include the additional point that evaporation through the membrane is highly fractionating (which we thought was sufficiently obvious that it did not need to be said).

A cheap peristaltic water pump might also be an improvement over the diaphragm pump that is supplied with the CWS, but would present its own maintenance issues (aging and wear of pump tubing, for example). We used the Dosino pumps, despite their higher cost, because they provide direct control over fluid flow rates.

8. *P 6 L 213. Drift is attributed to biofilm growth, was this growth assumed or actually observed?*

During prior tests at this field site the membrane was removed and a biofilm could indeed be observed. We can modify the text to make this clear.

9. *Possibly temperature drift (inlet air and water) was also a factor in the field deployment? Was the instrumentation exposed to outdoor temperature fluctuations or was temperature regulated indoors?*

The hut was not temperature controlled and thus the isotope analyzer was exposed to some temperature variations. The instruments inside the hut produced heat and thus, average temperature in the hut was about 12°C higher than outside. The variations of outside air temperature were clearly reflected inside the hut and in the membrane temperature of the CWS (streamwater temperature mirrors air temperature as well). Nevertheless, there was no long-term temperature trend that correlates with the drift observed in vapor concentrations (towards lower values). Except for diurnal patterns, air temperature and water temperature were rather stable - except for a warm period between 20 and 24 February 2016 when outside air temperature reached 17°C (Fig.5a). Inside the hut up to 23°C were reached during that period. This warm period did not manifest itself either in the vapor concentrations or in the isotope data of the routinely measured drift control, because the inlet air and water temperatures are both regulated using Peltier thermoelectric controllers. Once the air and water enter the membrane chamber, however, the flow rate of the water determines how much it is heated by the (much warmer) air. Thus, the temperature of the water at the membrane itself depends on the water flow rate, which is why we regulate the flow rate using the Dosino dosing pumps.

10. *P 6 L 217. How constant was the memory effect? If relatively constant, a data correction could be applied. Presumably it would be a function of analysis time (washout effect).*

To assess the stability of the memory effect during the 48-hour experiment, we calculated the percent carry-over for each Nanopure water sample injected immediately after Fiji water (when the isotopic difference between the two samples is the largest). We obtained rather stable values of percent carry-over that were (average±standard deviation) 1.25±0.35% and 0.89±0.44% for δ^2H and δ^{18}O, respectively. No statistically significant trend could be observed in the percent carry-over during the 48-hour experiment.

As mentioned already in P9 L323 in the first version of the manuscript, we applied a correction for memory effects by using

$$C_{true,i} = \frac{C_i - (1-X) \cdot C_{i-1}}{X}$$

where $C_{true,i}$ is the true value, C_i is the measured value, C_{i-1} is the value of the immediately previous injection and X is the memory coefficient. We will keep this correction procedure in the revised version of the manuscript.

4. Application in the field: P 9 L 323. It is unclear what ‘opposite behaviour’ means, a number of interpretations are possible. . .please clarify

We have clarified this statement: “Figure 5 shows that, for instance, precipitation samples became isotopically heavier during Events #2 and #8 when air temperature increased, while the precipitation samples became isotopically lighter during Events #1, #3 and #5, when air temperature decreased.”

11. *5. Comparison of event-water. . . P 11 L 407: consider using ‘precipitation’ instead of ‘moisture’*

As we also consider initial discharge as a proxy for antecedent moisture conditions rather than antecedent precipitation per se, we would prefer not to use the term “antecedent precipitation” conditions here.

12. *P12 L 435, 442: I agree this is likely - as has been shown by Tweed et al. 2016 (Hydrol. Process. 30, 648–660 2016). It would also be relevant to cite this publication in the Introduction as it* *appears to be the first study using continuous real- time isotope monitoring to trace contributions to storm flow.*

We will include this reference in the revised manuscript.

13. *P 13 L 463 and Figure 9: Event #7 results (»100% event water) needs commenting on. . .*

For Event #7, the large calculated event-water fractions (and standard errors) can be explained with the first δ^{18}O measurement of this event, which was isotopically very similar to the pre-event water signature (C_E=-11.5‰, C_P=-10.9‰). The incremental, volume-weighted mean of the event-water end member was thus isotopically heavier than the streamwater end member, resulting in a smaller difference to the pre-event water end member signature (Eq. 3). Precipitation samples after this first, less-δ^{18}O-depleted sample had an average δ^{18}O value of - 16.4±0.69‰ (±1 standard deviation, n=6). For δ^2H, such a strong effect did not occur and we could obtain reasonable isotope-based hydrograph separation results similar to the chemical hydrograph separation.

14. *P 13 L 466 onwards: Seems unlikely there was spatial variation in rainfall in such a small catchment (≈ 0.5 km2). Possibly O and H isotope compositions of other contributing (but not measured) water sources (groundwater, soil water, over land flow) had variable d18O / d2H relations (d-excess values). Since these were not accounted for, the simple two-component fraction calculations based on O and H could differ.*

We have moved this possible explanation to the end of this section and included the comment of the reviewer accordingly:

"Another way of viewing this problem is that the precipitation event may have mobilized a third pre-event water storage with unknown isotopic composition (e.g., Tetzlaff et al., 2014). It is further possible that the δ^{18}O- δ^2H relations (d-excess) of contributing water sources (groundwater, soil water, overland flow) were variable over time, resulting in different event-water fractions based on both isotopes."

15. *P 14 L 523: Somewhat ironically this section concludes that 3-hourly sampling would have been sufficient to capture the stream water events and would result in similar calculated event-water fractions. Consequently, the stream data could have been monitored using a conventional auto sampler at relatively low cost and with much simpler operation. While this may not be the case in all storm water events it may be the case that it is precipitation monitoring that will benefits the most by continuous isotope instrumentation due to the very rapid (minutes) changes that can occur in precipitation isotope values.*

We thank the reviewer for this comment, which we will implement into the revised manuscript: "Additionally, the rapid changes observed in precipitation isotopic composition (Fig. 6 and 5) suggests that high-frequency measurements are crucial for adequately represent the signature of the event-water end member."

As for the question of whether 3-hourly sampling could be done by conventional autosamplers: of course it is possible to use autosamplers at any sampling frequency, but higher sampling frequencies will necessarily entail more frequent field visits and greater numbers of bottles to be handled in the lab (with the associated quality control issues).

16. *Concluding remarks: The limitations regarding field deployability of the system (my general comments) and possible options for improvements in this regard could be expanded upon.*

The manuscript is quite explicit about the limitations of the analysis system because of the complexity of the instrumentation and its space requirements. The power requirement for the whole analysis system can only be estimated from the specifications given by the manufacturers. Based on these information, the number presented in the manuscript (around 1.7kW) considers the maximum power requirement of all instruments, for instance during warm-up. During steady-state operation we expect this number to be much smaller, however, we did not measure the power consumption directly, and thus do not know the exact number. Instead of presenting the maximum power requirement of all instruments, in the future version of the manuscript we will emphasize that line power would be optimal to allow for continuous instrument operation. Other alternatives, such a free-standing solar power system or a generator, would be possible, but these would be expensive and have their own reliability and maintenance issues.

Regarding possible options for improvements of the system regarding its field deployment we want to point out that the isotope analyzer with CWS was already optimized for “field applications” by the manufacturer. In contrast, the IC is a typical laboratory experiment that was not used in such an environment before. We are not instrument design engineers and thus will refrain from making specific recommendations for improvements beyond those we have tested ourselves.

17. *Table 2: stream stage unit must be m (not cm)*

We will change that.

18. *Figure 3: Note that the recorded water vapour concentrations (≈ 18,200 ppm) corresponds to a T of ≈16.2 °C supporting the explanation given by the authors that water warmed up beyond the 15°C setting of the diffusion cell of the CWS.*

We thank the reviewer for pointing this out.

19. Figure 6: GMWL and LMWL require references (especially the latter)

We included following reference for GMWL: Gat J, Mook WG, H.A.J. M. Environmental Isotopes in the Hydrological Cycle: Principles and Applications: International Atomic Energy Agency; 2001. In Fig. 6, we have changed the term “LMWL” to “Linear fit”.

20. *Figure 9: Error bars and their large variation between events need mention in legend and main text*

The error bars of the IHS are larger than those of the CHS because of the larger temporal variability of the isotope values in precipitation; that is, isotope values in precipitation vary by much more that the analytical uncertainty of the instrument.

We will add this information into the main text and into the caption of Fig. 9: “The larger uncertainties of the IHS results compared to CHS can be explained with the large temporal variability of the isotope values in precipitation, which substantially exceeds analytical uncertainty during most events.“

21. *Figure 10, 11: Legends should clarify that the 3, 6, 12, 24 hour ‘sampling intervals’ were derived by re-sampling of the 30min data.*

We will include this information in the figure captions.