**Authors' response to referee comment 1**

**General comments:**

**Referee Comment**: Carry-over effects: The manuscript describes how the sample (12ml) remains in the sampling tube until it is injected into the vial (P3 L27-31). Due to the under-pressure in the tube, a new sample fills the tube when the previous sample leaves it. I'm wondering about the carry-over effects due to the temporary sample storage in the tube, which might be significant, e.g. for instance for streamwater sampling when precipitation events cause drastically changing solute concentrations compared to baseflow conditions. Can you elaborate on potential carry-over effects in the tubing and what could be done about it (e.g., flushing with air or sample water)?

**Authors' response**: Carry-over effects might occur with the device setup as presented in our paper, in particular, as referee #1 points out, when the chemical composition varies strongly between consecutive samples. Carry-over effects could be effectively prevented by thoroughly flushing the tubing with sample water, either prior to sample pre-collection, or prior to sample injection. If such a flushing step is implemented, sample pre-collection becomes obsolete. During development of the presented device we regarded minimizing both the power-consumption and the technical complexity as a higher-priority requirement than preventing carry-over effects through flushing. However, a flushing step could still be implemented without the need of any fundamental changes to the current system. It is important to bear in mind that, in some sampling scenarios, flushing is not a viable option, especially in scenarios where the sample water is not provided in sufficient quantity or continuity, for example during rainwater or cave dripwater sampling. In sampling scenarios focussing on water isotopes, carry-over effects are likely to be minor as the water molecules to be analysed for oxygen isotope composition do not strongly bond to the tubing's wall, but are readily flushed out of the tubing during sample injection. Furthermore, the isotopic composition of natural waters is unlikely to change drastically between consecutive samples.

**Referee Comment**: If the sampling aims at analysing organic constituents, biofilm growth inside the tube might alter the sample, especially when the sample interval is long, e.g. several days? What could be done to prevent biofilm growth?

**Authors' response**: As the tubing is contained within a sealed case protecting the tubing and sample vials from sunlight, the probability of biofilm growth is already diminished compared to a system exposed to light. As some microorganisms are capable of forming biofilms in the absence of light, to further prevent the formation of biofilms, antimicrobial coatings could be applied to the inner walls of the tubing, such as antibiotics, biocides or colloidal silver coatings that are commonly used on medical devices to prevent infection (e.g. Ramasamy & Lee, 2016). The most practical solution to the potential problem of biofilm growth is probably the use of silver plated metal tubing instead of the FKM tubing presented in the paper.

**Referee Comment**: Fractionation effects during sample storage: During the third experiment you conclude that no alteration of the sample occurred because of the constant $\delta^{18}O$ values (Fig. 6). Do you get the same results when using $d^2H$? Since your samples were analysed with a LGR, both isotopes should be measured simultaneously.

**Authors' response**: Yes, the δD results (see Fig. 6b) also confirm the long-term stability of the samples: Again, if the vials were not airtight, evaporation would have led to a preferential removal of isotopically light water molecules from the water samples due to their higher vapour pressure (e.g.

Hoefs, 2015) and, consequently, to an increase of the δD value of the remaining water sample over time. Such a positive trend is not present in the δD data and the results from the repeated measurements agree well with the initial ones. The difference in δD values between initial and repeated measurements ranges from -0.30 ‰ (lt20 and lt23) to 0.70 ‰ (lt02-05), but averages out at 0.0 ‰ over all measurements (median also 0.0 ‰) indicating that there is no systemic discrepancy between initial and repeated analyses (Fig. 6b).

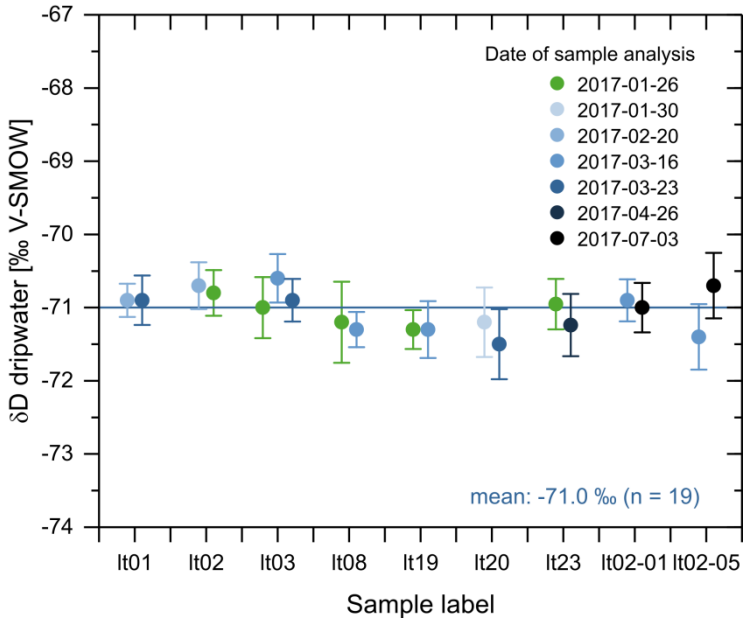


**Fig. 6b: Results of repeated δD measurements (circles in tones of blue) measured in the automatically collected samples together with the original δD data from Fig. 5 (green circles) plotted against their respective label ("lt" stands for Laichinger Tiefenhöhle). The darker the tones of blue, the later the respective measurement was repeated.**

**Changes to the manuscript**: Insert Fig. 6b in the Supplements and include a reference to Fig. 6b in the text.

**Referee Comment**: Check-standard during long-term sampling: In the case study, the GUARD system was operating over a period of 5-days and $\delta^{13}C$ was measured in the 22 drip water samples. How can you be sure that the $\delta^{13}C$ values you have measured were not affected by the sampling process or the storage? In order to quantify drift effects or alterations due to sample processing, it would have been ideal to regularly sample a check-standard with known $\delta^{13}C$ in addition to the drip water samples. I would recommend to at least address this issue in the interpretation section of the results.

**Authors' response**: The purpose of the case study performed in the cave "Kleine Teufelshöhle" was to monitor the changes in dripwater $\delta^{13}C_{DIC}$ values with varying cave $pCO_2$ after the dripwater had equilibrated with the cave atmosphere via $CO_2$ degassing. Therefore, if potential drift effects or alterations in dripwater $\delta^{13}C_{DIC}$ values caused by sample processing were to be examined using a check-standard of known $\delta^{13}C_{DIC}$, this standard would have to be treated exactly as the sampled dripwater, i.e. allowed to degas prior to sampling. This would however alter the $\delta^{13}C_{DIC}$ value of the standard, depending on the varying $pCO_2$ difference between dripwater and cave atmosphere, thus inevitably hampering the use of the check-standard as a control with known $\delta^{13}C_{DIC}$. However, an aliquot of the $CO_2$-equilibrated check-standard could be sampled manually and injected into an airtight sample vial with a double-cannula syringe, shortly before another aliquot of the check-standard is collected automatically by the GUARD autosampler. Comparison of the $\delta^{13}C_{DIC}$ values of

both “samples” should enable for detecting any potential sample alterations during automatic sampling. In agreement with the comment of referee #1, we will address the issue of potential sample alterations in section 4.4.

**Changes to the manuscript**: Insert at the end of section 4.4: “We note that potential drift effects or sample alterations that might be caused by the automatic sampling process have not yet been examined in detail. Corresponding tests using check-standards of known $\delta^{13}C_{DIC}$ values will be performed in future studies.”

**Referee Comment**: Harsh conditions: You state that the GUARD system is applicable in harsh (outdoor) conditions (title, P1 L19), which should include a wide range of air temperatures. However, there is no analysis of potential evaporation effects of the samples in very warm (and dry) environments. Instead, during the only long-term experiment that focused on the gas-tightness of the sampling vials, the samples were stored in the fridge at 8°C (P6 L29). In a warm (and dry) environment, I would expect the evaporative fractionation effect to be detectable, especially if the sample sits in the sampling tube for a while before it is injected into the vials. Could you please elaborate on this?

**Authors’ response**: The statement that the GUARD autosampler is applicable under harsh conditions mainly refers to its rugged water-tight casing and its ability to prevent damage from extreme weather conditions (e.g. water or dust ingress, high humidity, etc.) and to protect the samples from any external interference, e.g. from animal activity. However, this statement can be expanded to include the samples, too: Once, the samples are injected in the airtight vials, evaporative fractionation as well as other forms of sample alteration are effectively prevented, regardless of ambient air temperature or temperature fluctuations. It is certainly true that the sample is most prone to change during pre-storage in the FKM tubing. During this phase of the sampling, evaporative fractionation is at least minimised through two mechanisms: First, the FKM tubing is highly impermeable to gases and thus impedes evaporation and/or gas exchange through its walls. Second, evaporation can only occur over a very small surface of only about 12.6 $mm^2$ thanks to the small inner diameter of the tubing of only 4 mm. Furthermore, sample pre-storage inside the tubing is not necessary if sample water is provided in sufficient quantity and continuity, for instance, when sampling water from rivers, lakes or the ocean. In these cases, the sample can be injected directly into the sample vial and is therefore almost instantly sealed from the surrounding atmosphere.

**Specific comments:**

**Referee Comment**: P6 L10 and Fig. 5: You describe that you have collected one drip sample per day over a period of 33 days, however, in Fig. 5 only 14 data points from the GUARD system are shown, and these are clearly not in daily intervals. Please correctly state the used sampling interval in the text.

**Authors’ response**: The sampling interval is correctly stated as daily, however, not all of the 33 samples were analysed for $\delta^{18}O$ values.

**Changes to the manuscript**: Add at the end of the caption to Fig. 5: “Not all of the 33 samples were analysed for isotopic composition.”

**Referee Comment**: P6 L21-26 and Fig. 5: Why don’t you show the remaining data points in Fig. 5 to support your claim that the isotopic composition in drip water can vary strongly over short periods?

In this context, I would suggest to also provide the standard deviation to the arithmetic mean value in L25. If the standard deviation is substantial (which you suggest with your statement in L21-23), your conclusion based on the arithmetic means would be invalid.

**Authors' response**: While all of the manually collected samples were analysed for $\delta^{18}O$ values, not all of the hourly samples collected by the autosampler were measured. However, the sum of 16 samples over a period of 26.5 hours is sufficient to establish that there is a certain variation in dripwater $\delta^{18}O$ values on time scales as short as 30 minutes. The (absolute) standard deviation for the 14 automatically collected samples is 0.07 ‰ and 0.06 ‰ for the 12 manually collected samples. Based on the small difference of only 0.03 ‰ between the arithmetic mean $\delta^{18}O$ values calculated for both sample types, we concluded that there is no systematic discrepancy between the automatically and the manually collected samples. This conclusion holds true even if dripwater $\delta^{18}O$ values vary on time scales as short as 30 minutes as this variation includes both positive and negative excursions from the long-term mean values.

**Referee Comment**: P6 L2-3: Why didn't you simply weight the vials before and after filling in order to quantify the sample volumes?

**Authors' response**: Weighing the vials before and after sample injection is another way of quantifying the sampled volumes. As the sample vials were almost entirely filled during the various test runs we conducted, quantifying the sampled volumes by means other than visually confirming that only small air bubbles remained after sample injected simply did not seem necessary.

**Referee Comment**: P7 L 27: Sampling for 5 days, every 4 hours would yield 30 samples, not 22. What happened to the remaining 8 samples?

**Authors' response**: The remaining 8 samples could not be successfully collected during the case study due to an imprecise positioning of the sample slide and double-cannula at the position of sample 24. We have already been able to trace this positioning error to a faulty motor driver. We have therefore installed a new motor driver and achieved both precise and reliable positioning results since this change.

**Referee Comment**: Fig. 5: Why are the error bars different for some points? Please indicate in the figure caption, what the errors pars represent (measurement uncertainty?). You should also report d2H values in Fig. 5 since they are measured anyway.

**Authors' response**: The error bars represent the precision of each individual measurement. It includes the precision of the ten internal sweeps performed by the mass spectrometer on a single sample and the precision of multiple (two to three) measurements of the same sample. The error is propagated using the formula x= (a^2+b^2)^0.5, with x being the propagated error and a and b representing the two error types outlined above.

**Changes to the manuscript**: We will include the measured δD values in Fig. 5.

**Referee Comment**: Fig. 6: In greyscale, the shading of the data points is difficult to distinguish (green versus light blue). I would suggest a different way to present these data, especially since some data points overlap with each other and the error bars.

**Authors' response**: We have changed Fig. 6 so that the data points do not overlap any more. As 7 different measurement dates need to be illustrated in this figure, indicating the different

measurement dates with different data point symbols or shadings is neither practical nor intuitive in this case. For suggest to include in the figure's caption "For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper."

**Referee Comment**: Tab. 2: The sample volume can be smaller than 12ml in the GUARD system.

**Authors' response**: That is correct. The sample volume can be defined by changing the duration of the pumping step during sampling. Headspace is minimal if the vials are filled to the maximum.

**Changes to the manuscript**: We will include a ≤ sign in Tab. 2.

## References

Ramasamy, Mohankandhasamy; Lee, Jintae (2016): Recent Nanotechnology Approaches for Prevention and Treatment of Biofilm-Associated Infections on Medical Devices. In: *BioMed research international* 2016, S. 1851242. DOI: 10.1155/2016/1851242.