

Response to Reviewer 2

We would like to thank Reviewer 2 for the comments and remarks which will significantly improve the manuscript. Please find our replies to the general and particular comments, as well as suggestions for revisions, below.

R2: Soil characteristics (Table 1): TOC is undefined. Key soil characteristics such as porosity, dry density, gravel content, organic matter should be given if they are available. Could this method be applied to other soil types?

Answer: We will define TOC and CaCO₃ and revise the table header, also following suggestions from another reviewer. The revised table header would read as follows: *“Table 1: Site and material characteristics. TOC – total organic carbon; CaCO₃ – calcium carbonite. Coarse material > 2 mm; Sand 2–0.063 mm; Silt 0.063–0.002 mm; Clay < 0.002 mm.”*.

We will add information on coarse material (> 2 mm). Unfortunately, we cannot provide data on the other requested soil characteristics.

In general, this method should be applicable to a rather broad spectrum of soil textures, which is the main criteria for application. However, we do not have information yet to add textural limits, but we suppose that soils with a very sandy texture will be difficult to handle, because of their weak structural integrity.

We will add a discussion to section 4.1., which would read as follows: *“In principal, we expect that the combination method is applicable to a large variety of soil types, but cannot yet provide textural limits. However, combining two soil blocks requires removing the bordering at one side. Thus, soils that are structurally weak and could collapse without a frame are not suitable for this method.”*.

R2: Statistical analysis: differences between experiments are analyzed using statistical metrics. The Methods section does not clearly explain how this is done. For example, H and P (Line 132) are undefined. L. 114: Significance of what? What would have been the results at the 0.01 level?

Answer: *H* and *P* are the outcomes of the Kruskal-Wallis test: *H* is the test statistic, *P* is the probability measure. These two abbreviations are commonly used and we suggest not to give definitions for these test statistic or p-value.

Statistical significance was set to 5 %, as is usually the case in virtually all studies across various scientific fields. A significance level, or α , of 0.05 refers to a risk of 5 % that we conclude we have found significant differences between two (or more) groups when in fact there are none. This value is common practice, although, of course, arbitrary and sometimes criticized. We suggest not to deviate from it. Nevertheless, in the manuscript we do not solely rely on statistical significance alone, but also discuss the meaningfulness of found differences (e.g., if there is a directional effect or trend, how substantial the variation is, etc.). Our conclusions would have been no other if α would have been set at 0.001 (or any other value, for that matter).

We hope that we did not misunderstand the intention of R2 in this comment; if so, we would kindly ask R2 to clarify the issue.

R2: Experimental design is unclear. For example, on L. 116, a “second experimental set” is mentioned. What is this? The first line of Table 2 is not complete for understanding the experimental plan. A new Table listing all experiments would be useful.

Answer: The *“second experimental set”* is the one that is described in the manuscript; data on the first experimental set is provided in the Supplement. Both are described at the beginning of section 2.2. Maybe the wording (*“set”*) is misleading? What we meant is synonymous to *“trial”* or *“round”*.

During the first set, we conducted experiments on all 6 monoliths and recorded the outflow at 2 flow pathways. After some improvement of the overall set-up, the second set started, during which we conducted another round of experiments on all 6 monoliths, but this time recording the outflow at 4 flow pathways. As the latter is more elaborate, we decided to only report on the results of the second set in the main text. However, as stated in the discussion, the first set provides similar results. Thus, as long as not explicitly stated, all results in the main text refer to the second experimental set. To clarify this, we will rephrase the paragraph, which would read as follows: *“The runoff experiments were carried out in two experimental sets, each comprising a full round of experiments (runoff and tracer measurements) on all six monoliths. During the first set, two flow pathways were recorded at the lower end of the monolith, the surface runoff and subsurface interflow. For the second set, we further sampled and distinguished between percolating water that went through the whole soil body vertically and water that infiltrated into the soil body, but left the monolith again at the side due to lateral flow pathways.”*.

Maybe we are missing the point; in this case we would kindly ask R2 for a clarification.

R2: Discussion: Section 4.2 is too long, especially for a technical note. Could be more concise.

Answer: We agree that the discussion is long, especially for a Technical Note. This is mainly due to the fact that we used a rather extensive approach (outflow measurements, chemical loading of runoff, flow velocity). We will shorten and rephrase the discussion where possible to be more concise. Parts of the Discussion will also be changed/rephrased following suggestions from other reviewers.

R2: The title of the paper could be improved.

Answer: We agree. We suggest to change the title to *“Technical Note: Combining undisturbed soil monoliths for hydrological indoor experiments”*.