We thank the reviewer for the useful comments and respond to them in blue font after the original comment in black font.

Reviewer 3

"Soil and stem xylem water isotope data from two pan European sampling campaigns" presents a very interesting collaborative dataset and provides a well-written and generally clear documentation of the data and sampling protocols. The usefulness of the data and potential for applications are clear and exciting. Upon addressing minor comments, I recommend the publication of this manuscript and data.

We thank the reviewer for the positive assessment of our manuscript. We are similarly excited to use the data for further publications and hope that others will use it as well. Please find our point-to-point response below.

Specific Comments:

Line 191: When photos were taken for both campaigns, were they averaged? Or was one selected over the other? I now see this is answered in Tables 2, but it would benefit from mentioning here too.

Okay, we have now added more details for Canopy cover in the Table 1 caption: *"*** based on the average value for all photos for each sampling site"*

Section 2.3: Where were soil samples collected at each of the sites? Was there a protocol for location relative to the trees sampled? I think this is described more in the protocols document, but it should be expanded on here, too.

In the sampling protocols we stated: "*Take a maximum of 5 soil samples at 0-10, 10-20, 20-30, 50-60, 80-90 cm at a location between the 3 (or 6) trees that you sampled*", which is similar to the description in the manuscript. We do not have more information to describe the location per site any better. We slightly modified the sentence and now state:

"In addition to the stem xylem samples, soil samples were taken with a manual soil auger at a location between the selected trees."

Line 211: It seems that samples were taken between 30-50cm based on the protocols. Why is this not a range mentioned here?

For soil samples below 30 cm, the variability of the sampled soil depth varied due to differences in the maximum soil depth (and the presence of rocks). To avoid adding too many details for individual sites, we decided on a more generic statement and refer to "other depths". We modified the sentence:

"The samples were taken from a single soil core at three to five depths, typically at 10 cm intervals (0-10, 10-20, 20-30, 50-60, and 80-90 cm below the surface). In some cases, other depths were sampled, or the sampling interval was 20 cm."

Line 215: Can you add text to clarify if these soil samples were taken at deeper depths or at separate locations?

We added more details to the sentence:

"At some sites and during certain campaigns, soil samples were also taken from two to three additional nearby locations (up to four in total), resulting in a varying number of samples and sampling depths."

If separate locations, were these averaged for subsequent analyses or are they all included as separate data points?

The data of the different locations (= auger holes) are provided as separate data points.

Line 223: Can you expand on why they were shipped without cooling? It's stated in the immediately prior sentence that refrigeration helps reduce moisture loss and evaporation/fractionation.

Indeed, cooling can mitigate evaporative water loss, which causes unwanted isotopic fractionation during storage. It will also reduce microbial growth and slow the decomposition of organic material. However, we decided against shipment of the samples in refrigerated conditions to avoid complex and costly transport logistics. That said, we minimized shipment duration using an import license. Now stated in concluding remarks:

"while the availability of an import license reduced shipping times and lowered the risk of sample loss"

Line 249: I find this sentence challenging to follow. I assume the 40.9% values is referring to all samples, but it's not immediately clear and makes the "respectively" description at the end feel disconnected from the beech and spruce.

We rephrased the sentence to be clearer:

"The gravimetric water content ("gwc") varied among sample types and averaged 41% for soil, 61% for beech xylem, and 84% for spruce xylem samples (Figure 2C)."

Table 2: exe_type Are these exe_types listed and described somewhere? I think it would be beneficial to understand the different types of containers used. If this is listed somewhere, please add a reference to the location.

Good point. We now provide various weights and details on the variable "exe_type" in the supplement section S4. This list reflects quite some work, and therefore we think that it is useful that these details are published somewhere.

Line 310: I think it would be beneficial to add a sentence somewhere in the text (probably Introduction) on how these variables are related to isotopic processes to motivate these models.

That is a good idea. We now state in the introduction:

"This includes geographic details, information on soil type, texture and maximum depth, details on forest stands, tree diameter and height, sampling information, as well as data on canopy cover/gap fractions as indicators for stand density and tree health and crown defoliation (Bussotti et al., 2024). Together, the metadata and isotope data provide a strong foundation for future research on tree water use, model testing, and isotope mapping."

Table 5/Section 3.1: Is there a reason the soil samples were grouped in this way, and the 10-20 cm and 20-30cm samples were excluded? In the text, soils are suddenly grouped and discussed

as "shallow" (0-10 cm) and "deep" (30-90cm). Please add a sentence (perhaps in Section 2.3 or Paragraph 2 of 3.1) where these groupings of the soil are clearly defined and briefly discussed.

Good point. There was no specific reason. To be consistent we now add the details for the soil depth 10-20 cm and 20-30 cm to Table 5 too. We also revised the paragraph and now avoid the groupings “shallow” and “deep” and refer to the actual soil depth. But see our response to the comments below as well.

Line 331: So δ18O values were only higher in summer compared to spring for the shallow soil depth (0-10cm), not for all soil depths? The first part of the sentence seems to say that δ18O is higher in summer for both shallow and deep ("the different depths"), but then it seems to indicate that this is only true for shallow soils.

We revised the sentence to be clearer:

“$\delta^{18}O$ values were higher in summer compared to those of the spring for stem xylem water of both species and for soil water at 0-10 cm, 10-20 cm and 20-30 cm (unpaired t-test, $P < 0.05$). The $\delta^{18}O$ values of soil water at depths of 30 to 90 cm did not differ seasonally (unpaired t-test, $P > 0.05$; Figure 5).”

Line 337-339: What is the difference between "shallow" soils and soils above 30 cm? Is it that summer deep samples are more negative than the 0-10 cm & 10-20cm & the 20-30cm soils? While the spring samples are lower than the 0-10cm samples but not the 10-20cm or 20-30cm samples? It's hard to tell if these analyses are only comparing the shallow and deep groups of samples or the 10 cm increments as described in the methods.

Indeed, that was confusing. We now avoid using “shallow” and “deeper” categories and revised the sentences to be clearer:

“Comparisons across all soil depths shows that in spring, site-level mean $\delta^{18}O$ values of soil water at 30–90 cm depth were lower (i.e., more negative) compared to those at 0–10 cm (unpaired t-test, $P < 0.05$) but not to those at 10-20 cm or 20-30 cm (unpaired t-test, both $P > 0.05$). In contrast, in summer $\delta^{18}O$ values at 30–90 cm depth were lower than those at 0-10, 10-20 and 20-30 cm (unpaired t-test, $P < 0.05$). Similar seasonal differences for stem xylem and soil water were observed for the δ2H values (Figure 5).”

Line 368: Similar to line 249. There are several sentences throughout that are tough to follow because there are two seasons, two analytes, and/or two or more subsets of samples mentioned within the same sentence, and it's not immediately clear which grouping of the samples "respectively" is referring to.

We agree and now see that the sentence structure caused confusion. We rephrased the sentences to:

“Our data also shows a clear isotopic difference in stem xylem water between the two tree species (Figure 6). The mean species difference (spruce-beech) in $\delta^{2}H$ and $\delta^{18}O$ values across all sites was 5.5‰ and 0.8‰ in spring and 9.5‰ and 1.1‰ in summer, respectively.”

In addition, we further revised the overall manuscript and rephrased overly long and complicated sentences in the manuscript.

Line 371: Please add a citation

We now cite Goldsmith *et al.* (2019) as a good example of differences in root water uptake for spruce and beech.

4 Concluding Remarks: I understand this may be out of the scope, however, I feel a short section highlighting some key "lessons learned" regarding sampling, analysis, and collaborative efforts across this impressive geographic extent would be beneficial and of interest to the community.

Good point, we added some lines to the concluding remarks: *"Establishing this data set with a geographic cover across Europe was feasible because the participants took advantage of an EU Cost Action with members in most European countries. We believe that limiting the number of samples to 6 to 8 per site contributed considerably to the success of the data collection. Centralizing the laboratory and analytical work avoided potential inter-laboratory biases, while the availability of an import license reduced shipping times and lowered the risk of sample loss."*