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

Reviewer 2

The authors present an interesting dataset of soil moisture and plant xylem water (beech and spruce) isotopic compositions measured across 40 sites in the spring and summer seasons. The dataset documentation is clear and transparent. The potential uses of this dataset are also immediately clear to me. I fully support publication of this data release. I provide some minor comments below:

We thank the reviewer for the positive assessment of our manuscript and the dataset. Please find our point-to-point response below.

Specific comments:

Line 74: The "2" and "18" should be superscript.

Thanks, corrected.

Line 99: perhaps consider "resistance and resilience". Deeper roots can help a plant recover after accumulating damage during drought, but deeper roots can also help individual trees avoid negative consequences during a drought in the first place.

We agree. We have now added "*resistance*" to the text.

Line 99: rather than pointing to the topic of "climate change" it could be more informative to list the changing hazards that impact trees where greater RWU would help to avoid or mitigate negative consequences (i.e., drought, fire). Climate change just seems to broad a term in this application. Ocean acidification falls under climate change but probably doesn't impact trees much.

Okay. To narrow the scope of the word "climate change", we now link climate change specifically to hazards such as the increasing frequency and intensity of droughts in the following sentence, which we see as the most important factor affecting RWU. Now stating:

"Understanding how tree water uptake from soils varies with species, site characteristics, time, and across climate zones is essential to assess forest resistance and resilience to climate change; particularly the response of forests to the increasing frequency and intensity of droughts"

Line 102: rather than "determination" I would suggest "estimation" as these isotope methods are constantly debated and refined. This paper and dataset highlight potential issues that still should be resolved.

Good point, we changed the text accordingly.

Line 110: and possibly (iv) that all end-members have been measured with proper consideration for the locations of uptake and the transit time of water through the plant. The article states in the discussion the possibility of water sources beyond those that were measured.

We agree. We have now modified point (i) to consider your comment and changed it to:

"Additionally, it is assumed that: (i) the sampling design captures all end-members with a proper representation of the spatiotemporal variability of their isotopic composition"

Line 111-112: Water uptake is just one dimension of plant responses to drought though, right? Stomatal closure, xylem resistance to embolism, canopy position, internal water storage, etc. etc. etc. Roots are certainly important, but they do not paint a complete picture of drought resistance and resilience. As an example, cacti generally have shallow roots relative to forest trees but are capable of existing in much more arid regions.

Yes, root water uptake is just one component of a multifaceted drought response. Physiology and hydraulic traits are also key for plant drought responses. We have slightly expanded the sentence and now state:

"the method can either independently or in combination with other measurements (e.g., in combination with assessment of physiological or hydraulic traits) be used to …"

Line 130: Possibly this is the only international network that explicitly states this as the primary goal. The LTER and CZO networks across the US measured soil and plant xylem water isotopic compositions at almost all of the sites similar to this study. The political boundary of the US makes it so that all exist in one country so they are not "international", but the US is comparatively large by land area, spanning a diverse range of ecoregions. My point is that there have been other attempts at something similar, while this sentence makes it sound like this has never been done before. While the CZO is inactive (and LTER is threatened), NEON sites remain active, and isotopes are being measured at some.

Thanks, that is highly interesting, as we were not aware of these attempts. We deleted "only" and added nuance to the statement:

"The Moisture Isotopes in Biosphere and Atmosphere (MIBA) network, initiated by the IAEA in 2003-2004, is a rare example of an international network to survey the isotopic composition of water across different ecosystem compartments…"

As a sidenote, we are fully aware the US is almost as big as Europe, but we also think sampling across multiple countries is more challenging than sampling across the US, e.g., due to issues with the shipment and import of samples, languages, etc., see also response to comment above.

Lines 174 – 175: It would be really interesting to share the lessons learned from these calls. My research group is constantly communicating with others regarding challenges in field sampling, storage, shipping, analysis, etc. This could be very helpful. Possibly this is already all documented in the Ceperley et al (2024) paper?

That's a good point, which was also mentioned by reviewer 3. While large parts on sampling and extraction procedures are covered by Ceperly *et al.* (2024), we expanded 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."

Line 189: Maximum of all sites, or the maximum at each site?

We refer to the maximum of all sites, which is now stated:

"The maximum soil depth varied between 0.3 m and > 1 m. For half of the sites was the maximum soil depth > 0.6 m."

Line 189: Maybe clarify "canopy cover" here. I work with foresters who use lidar to construct 3D models of canopy structure. They might not call photos representative of canopy cover.

Yes, there are more sophisticated methods to measure canopy cover, but we feel that the sentence already describes the method explicitly so that users can decide whether they want to use it or not. Still, we replaced "determined" by "estimated" to provide more nuance to our statement:

"Canopy cover was estimated for 30 of the 40 sampling sites from non-hemispherical photographs taken systematically at varying distances from the stem with a smartphone camera (Supplementary materials S3)."

Fig 1: The country borders are showing up at a low resolution. Is there a way to improve this? Minor issue.

The available plots (including Figure 1) have been provided in screenshot quality to avoid large word and pdf files (as we had to share the document with many co-authors). Of course, we have high resolution plots for all figures. In the higher resolution plots, the country borders are more clearly visible. We will provide them to ESSD in the final submission process to achieve the highest possible print quality.

Line 202 – 203: I'm a little confused by the terminology. I typically think of "stem" as a < 3 cm diameter branch. Here it sounds like you are coring the trunk if you used an increment borer. Is it stems or trunks?

From our experience, "stem" is often used when it comes to "trunk" or "tree-ring" samples and we therefore think that this is a meaningful term. This is also underlined by having no comments on this from any of the 70 authors of the manuscript before the submission process and a quick informal check with other colleagues. Nevertheless, we now added "tree" to the title to make it more obvious that the dataset deals with samples from trees. We also added on several occasions that the stem xylem samples were taken from the trunks of trees in the abstract, introduction and material & methods. We also added a definition in section 2.3:

"Thus, in this study, "stem" refers specifically to the trunk of the tree, excluding branches and other aboveground components."

Line 205: was heartwood removed too?

We instructed the field crew to avoid sampling heartwood. Often one can hear and feel when the heartwood is reached so we think that in most cases the field crews did this correctly. That said, we could not determine the absence or presence of heartwood from the dry wood samples after water extraction and therefore cannot fully exclude that some samples may have contained some heartwood, particularly spruce samples from small trees. Please see the detailed response to a similar comment from reviewer 1. We have added some sentences to make this more explicit in section 2.3:

"The wood cores mainly reflect sapwood as participants were instructed to avoid sampling the heartwood because there are indications of isotopic differences between sapwood and heartwood (Fabiani et al., 2022). However, we cannot fully rule out the presence of heartwood in some samples as visual determination of the heartwood after water extraction was not possible.

A heartwood correction based on mean wood core length and tree diameter could be developed. Such an adjustment may be particularly important for samples from smaller spruce trees, which are likely to have limited sapwood depth (Peters et al., 2019).”

Line 211: Why say “typically”? There was a protocol or not? What are those “other depths”?

“Typically” because most samples were taken at the indicated depth, i.e., according to the protocol. “Other depths” refer to soil samples that were taken at depths that deviate from the protocol (e.g., 65-75 cm as stated in “soil_depth” in Table 3) because the maximum soil depth varied among sites. We rephrased the sentence to be clearer:

“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. The number of soil samples and the depth of the deepest soil sample depended on the soil properties (e.g., rocky soils) and the maximum soil depth at the sampling location.”

Line 214: How close?

Unfortunately, we do not know the exact distance between the location of the soil core and the trees for each site. It was only instructed that the location should be between the selected trees as mentioned in section 2.1.

The Goldsmith et al (2019) paper and my own sampling has me convinced that we need soil samples at least in triplicate due to high spatial heterogeneity. Was this disregarded to simplify data collection?

We are aware of the paper (now cited) and the issue. We discussed it during the initial phase of the project. However, collecting more soil samples would have substantially increased the workload for water extraction and isotope analysis. To ensure feasible sampling, shipping, and participation across many sites, we reduced the number of soil samples to a practical minimum. This decision is now reflected in the concluding remarks:

“We believe that limiting the number of samples to 6 to 8 per site contributed considerably to the success of the data collection.”

Interestingly, at the few sites where multiple soil cores were collected, we observed a high similarity in the isotopic composition across soil depths, supporting the adequacy of our sampling strategy.

Line 217: was root mass with depth measured at any sites? Wouldn’t the root profile be informative for designing the soil sampling protocol of each site?

We agree that this would be very helpful information. The data exist for some sites, which can be deduced from “website_link” or “paper_x” in Table 2. However, because root traits are not easy to assess and need some expertise to establish them correctly (and are difficult to determine based on just one soil core), we did not include root traits in the sampling design and protocols, also to mitigate the risk of “losing” potential contributors.

Line 222: How fast is “back in the laboratory”? We’ve done tests where we put samples into plastic containers and leave them at room temperature and stored cold for varied amounts of time. We observe no problematic effects for the first 6 hours, but we do often observe problems for samples held at longer storage times. This depends entirely on the container and ambient

environmental conditions, but it might be good to quantify the potential issues related to sample storage and transport time.

While we know the date and time of the sampling in the field (mean ~12:30 local time), we have not noted the arrival of the samples in the laboratory. We asked and therefore must assume that all contributors stored their samples in the refrigerator in the afternoon/evening of the sampling day. Because most of the analyzed samples were kept in gas-tight "Exetainers" from Labco during storage and shipment, which we have used for several decades for various studies, we doubt that evaporative effects caused isotope fractionation during sample transport. This is supported by our finding that samples handled/shipped in different plastic or glass vials showed no signs of evaporative isotopic enrichment due to water loss (new Figure 8E). We now state this more clearly in section 4.2.:

"In contrast, vial type significantly interacted with sampling campaign ($P < 0.001$), with no effect in spring but a more depleted signal for the vial type "others" compared to "exetainer" for the summer sampling campaign. This pattern provides no indication of evaporative isotopic enrichment resulting from sample handling during the warmer summer conditions."

Still, we agree that it will be interesting to see what happens to the samples during longer periods of storage, but this is beyond the scope of this study.

Line 223: "shipped without cooling" and "4 weeks" is concerning to me. Did you perform a test to confirm that the shipping time didn't result in problems? You could take a dried plant or soil sample, spike it with a standard, hold it for the duration, and then extract to estimate the potential bias.

We decided for "Shipment without cooling" to avoid complicated and costly shipment of the packages, again to mitigate the risk of "losing" potential contributors. "4 weeks" is the maximum time it took. Most samples were handled and arrived much faster. We also minimized the shipping duration with an import license, as now stated in the concluding remarks:

"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."

We do not think that we can design an easy experiment that would be representative of all the variability that might be related to the various steps between sampling, shipment, CVD extraction, isotope analysis, and different soil types across ~70 authors and 40 sites across Europe. In other words, even if we would generate one estimate for an isotopic effect due to for example shipment period, we cannot be sure if it is representative for individual samples as we did not track the exact timing of the individual steps between sampling and isotope analysis. We therefore think that testing this uncertainty will not result in a meaningful outcome and is beyond the scope for this study.

However, the vial type used for sampling ("original_vial", Table 3) can, in a broader sense, serve as a test case for potential uncertainties related to storage and shipping. While most samples were collected in Exetainers, approximately 15% were shipped in other containers ("others"), including plastic or glass vials of differing shapes and materials. These alternative vials were not specifically certified as gas-tight and therefore posed a higher risk of evaporative isotope fractionation during storage and transport. The absence of a detectable vial-type effect for spring and depletion of heavy isotopes for samples store in the "others" compared to "exetainer" for the summer sampling campaign suggests negligible evaporative influences (i.e. water loss) during sampling, shipping, and the transfer of material into Exetainers prior to

cryogenic water extraction. This further supports the robustness of our dataset in this context. We added this information to section 4.2:

"In contrast, vial type significantly interacted with sampling campaign ($P < 0.001$), with no effect in spring but a more depleted signal for the vial type "others" compared to "exetainer" for the summer sampling campaign. This pattern provides no indication of evaporative isotopic enrichment resulting from sample handling during the warmer summer conditions. Given that the "others" vial type comprises only ~15% of samples, spread across no more than 8 of 40 sites in both campaigns, we consider this effect unlikely to confound the overall dataset, though it may warrant consideration in future analyses. Collectively, these results support the overall reliability of the dataset and its suitability for analyses of cryogenic water extraction biases and methodological evaluation"

Line 248: Does the final dataset exclude samples outside of the acceptable ranges, or does it include all values?

No data was removed. Users can use their own threshold and filter data for "tef" values outside of the optimal range, if they want to exclude those samples.

Line 371: Even if they were similar, they could still be used to study differences. It might just produce a null result, or maybe environmental conditions varied such that its interesting that the values were similar. More data is always good!

Agreed, we rephrased the last sentence of the paragraph to provide a broader and more generic perspective:

"The observed isotopic variability in stem xylem water among species and sites suggests that both species-specific differences in root water uptake depth and the environmental drivers of root water uptake across Europe can be inferred from these data."

Line 400: What if the effect was similar across all containers and transportation routes? This doesn't rule out the problem for me. A small experiment to estimate the bias attributable to containers/transport might be worthwhile to at least rule this out entirely.

While we agree that this is interesting, it is beyond the scope of the paper describing the dataset and its potential uses. See our response to similar comments above.