Associate Editor

Thank you very much for your manuscript revisions. I have a few suggestions for changes: Could you please change the color or make the symbols of the "plant samples" hollow in Fig. 3 (new), so that the soil samples become visible (as currently they are overlapped by the plant samples).

WE HAVE EXPERIMENTED WITH A VARIETY OF DIFFERENT SYMBOL TYPES. EVEN WITH OPEN OR SEMI-TRANSPARENT SYMBOLS, THE VAST NUMBER OF XYLEM DATA POINTS CAUSES MOST OF THE DISTRIBUTION OF SOIL DATA TO BE OBSCURED. THE REVISED FIGURE SUBMITTED HERE REPRESENTS THE BEST COMPROMISE WE WERE ABLE TO FIND – WE HAVE ADPOTED A SMALLER AND SIMPLER SYMBOL FOR THE XYLEM DATA IN THE TOP TWO PANELS, WHICH ALLOWS THE READER TO AT LEAST SEE THE RANGE OF DISTRIBUTION OF SOIL VALUES.

In my opinion, it would be a great benefit to show the "failure" of the ChemCorrect software (e.g. in an Appendix figure or table) as so many colleagues in our community rely on this flagging system assuming the data is "fine" after conducting the analysis.

WE HAVE ADDED A SUPPLEMENTAL FIGURE THAT WE HOPE ADDRESSES THIS SUGGESTION – IT SHOWS THE DISTRIBUTION OF OBSERVED AND MODELLED BIASES FOR EACH CHEMCORRECT CONTAMINATION FLAG VALUE (GREEN/YELLOW/RED).

Could you also please add the R2 values for the fitted models in Fig. 2.

DONE

Would it be possible to analyze a "species-effect" for a subset of your plant samples to underline your arguments in the discussion or at least show the variability across different species?

ALTHOUGH WE ARE NOT ENTIRELY SURE WE UNDERSTAND THE SUGGESTION, WE HAVE RESPONDED BASED ON OUR INTERPRETATION OF THE INTENT. WE HAVE PROVIDED MORE INFORMATION ON THE TAXON-SPECIFIC BIASES BY EXPANDING THE INFORMATION PRESENTED IN TABLE 1 OF THE MANSCRIPT. THIS TABLE NOW SHOWS THE NUBER OF SAMPLES, AVERAGE AND STANDARD DEVIATION OF THE MODELLED BIASES, AND PREVELANCE OF HIGH-MAGNITUDE BIAS FOR EACH SPECIES IN THE DATA SET.

Could you please add 1-2 sentences to your revised manuscript addressing this "L183_ instrument specific-corrections..." reviewer comment?

WE COULD HAVE BEEN MORE DIRECT IN OUR REPLY TO THIS REVIEWER COMMENT...THE MANUSCRIPT ALREADY INCLUDES A FULL PARAGRAPH (LINES 184-197) THAT RAISES AND DISCUSSES THIS CONCERN, WHICH WE COPY BELOW FOR REFERENCE. WE BELIEVE THAT THIS ACCURATELY AND COMPREHENSIVELY REPRESENTS WHAT WE ARE ABLE TO SAY ABOUT THE SUBJECT AT THIS POINT, SO WE HAVE NOT MADE ADDITIONAL MODIFICATIONS.

"A common concern that has likely limited the use of *post-hoc* correction for CRDS spectral interference is that corrections may be application and/or instrument specific. Although we cannot confidently argue that the correction approach developed here will be globally applicable, we note that the same correction equations appear, based on direct (Fig. 1) and indirect evidence (Fig. 3), to successfully correct bias for vacuum-extracted water from a large and diverse range of plants and soils. Given the inevitable wide variation in VOC composition and concentration among these samples this result suggests that a single correction framework may be applicable across most ecohydrological applications and systems. It is more difficult to judge whether the model parameter values calibrated here will be applicable to other analysers given variation in instrument optics and calibrations, but we found that a single model calibration could successfully be applied to data generated on two different L2130-i analysers: the difference in the mean model residuals for samples run on the two instruments was small relative to the dispersion of the residuals (0‰ for δ^2H and 0.16‰ for δ^{18}O) and not significantly different from zero (t-test, p = 0.99 for δ^2H; Wilcoxon rank sum test, p = 0.09 for δ^{18}O), suggesting that the same optimal models accurately corrected bias on both instruments. That said, the coefficient describing CH_4 sensitivity of δ^{18}O bias on our analysers is similar but not identical to that fit by Herbstritt et al. (2024), suggesting that some variability may exist even between analysers of the same model. Further testing and comparative calibration of bias-correction algorithms is thus warranted."