REVIEW 1 (our reply in black)

This paper applies statistical modelling techniques (combining a neural network model with seasonal trend analysis) to a comprehensive dataset of soil water contents and pressure potentials measured at 10 cm depth in lysimeters moved to two different locations, in order to identify shifts in hydrological responses to climate forcing.

The shifts in these "in situ" water retention curves (WRC) are intriguing and really quite dramatic (e.g. figures 4, 7 and 8). But I do wonder about the mechanisms and underlying processes. The authors are rather vague about the causes, suggesting that they are due to changes in soil structure tiggered by climate (lines 674-676). I'm not fully convinced about this interpretation, not least because the largest changes seem to occur in the very dry range of the WRC where structure should not play such a large role.

We thank the reviewer for raising this important point regarding the interpretation of the observed shifts in the in-situ water retention curves. We agree that our original wording could be interpreted as implying that soil structural change is a verified causal mechanism. This was not our intention.

In the manuscript, soil structure was presented as a plausible explanation based on established soil physics understanding that drought can alter pore networks and hydraulic behaviour. However, as in our response to Reviewer 2, we emphasize that the present study does not directly measure pore geometry, aggregation, wettability, or contact-angle effects. Therefore, we cannot verify specific physical mechanisms responsible for the observed SWRC shifts. Instead, the model-based analysis identifies persistent changes in the soil water content response function empirically, while the discussion of soil structural or pore-scale processes serves as a theoretical interpretation consistent with existing literature.

Regarding your comment on the relevance of soil structural effects in the dry range, the effect could be larger than expected, considering that aggregation at various scales and contact networks may change considerably by shrinkage during drying (and following swelling). Such effects (as for example measured by Berli et al., 2008 Water Resour. Res., 44, W00C09) may also explain the huge variability of unsaturated hydraulic properties predicted by pedotransfer functions in the dry range. We have added a sentence in the Introduction to clarify that.

Following the reviewer's guidance, we have revised the manuscript to clarify that references to soil structural or pore-scale change are presented as possible mechanisms rather than confirmed causal evidence. The revised text now states that the observed SWRC shifts reflect changes in soil hydraulic behaviour inferred from the data, while structural and interfacial processes are discussed as plausible explanations that require targeted future measurements for verification.

This revision aligns with the general modelling philosophy of the study, where deviations between observed and simulated behaviour are used to detect change in the soil–climate response function, while mechanistic explanations remain interpretative rather than experimentally proven.

We will replace "structural variability" with "variability in hydraulic response".

We will replace "structural changes" with "persistent changes in hydraulic behaviour" and "structural evolution" with "evolving soil hydraulic response".

(In this respect, I think the WRC curves should be plotted with matric potential on a log-axis for improved readability. On a linear scale, we can't really see what is happening close to saturation, which is where most of the structural changes would be expected).

We agree that logarithmic scaling of matric potential is commonly used to improve visualization of the near-saturated range of water retention curves.

However, in this study the purpose of presenting the SWRCs is not to analyse pore-scale behaviour close to saturation or to diagnose specific structural mechanisms. Rather, the curves are shown to demonstrate that the bias drift detected by the model corresponds to a **real, physically observable shift in the soil water retention relationship itself**, across the full range of observed conditions.

For this purpose, a linear scale is intentionally used to allow direct visual comparison of curve displacement between years and between lysimeters, and to highlight the magnitude and direction of shifts consistently with the mean-bias analysis. Using a logarithmic axis would compress large parts of the potential range and reduce the visual clarity of the overall curve displacement that is central to our change-detection framework.

We therefore prefer to retain the linear representation, as it best supports the methodological objective of linking model-detected drift to observable shifts in the empirical SWRC.

There may be alternative explanations for the observations, including (slowly reversible) swell-shrink behaviour and preferential (non-equilibrium) flow. I would encourage the authors to try to strengthen the discussion and interpretation of the data with respect to the underlying mechanisms, including the above-mentioned processes. Nevertheless, although the responses to climate of apparent WRC observed by the authors "in situ" seem stronger than I would expect (especially in the dry range), I am aware of two previous large-scale (regional-continental) statistical analyses of water retention curves measured in the laboratory that have shown significant impacts of climatic factors on the structural pore space (Hirmas, D. et al. 2018, Nature 561, 100-103; Klöffel, T., et al., 2024. Geoderma, 442, 116772). These studies could be mentioned as they would give support to the authors' inferences and interpretations.

We agree that several pore-scale and hydraulic processes could contribute to the observed shifts in the apparent water retention curves, including slowly reversible swell–shrink behaviour and preferential (non-equilibrium) flow. As also noted in our response to Reviewer 2, the present study does not directly measure pore geometry, aggregate dynamics, or flow regime, and therefore cannot isolate or verify specific physical mechanisms. Our discussion of soil structural or pore-scale processes is thus intended as theoretical interpretation rather than confirmed causal evidence.

Following the reviewer's suggestion, we have strengthened the discussion by explicitly acknowledging alternative processes that may contribute to the observed behaviour, including swell–shrink dynamics and non-equilibrium flow effects. In addition, we have incorporated the two large-scale studies highlighted by the reviewer, which demonstrate statistically significant links between climatic factors and structural pore-space organization across regional to continental scales. These studies provide independent support for our interpretation that climatic forcing can influence soil hydraulic behaviour and apparent water retention characteristics.

We add this line 604:

“This suggests that the structure and function of the soil system cannot be meaningfully decoupled from its climatic history……

Alternative pore-scale processes, including slowly reversible swell–shrink behaviour and non-equilibrium preferential flow, may also contribute to the observed shifts in apparent SWRCs behaviour. Recent large-scale analyses have demonstrated statistically significant impacts of climatic factors on soil structural pore space and laboratory-measured retention characteristics (Hirmas et al., 2018; Klöffel et al., 2024), providing independent support for the plausibility of climate-induced modification of soil hydraulic response.”

Specific comments

1. Line 40: “indicator” rather than “factor”?

Thanks, we changed it accordingly.

2. Lines 40-41: “Dominant”? This seems a bit far-fetched. I don’t think soil water content at 10 cm depth is the best hydrological indicator of soil health. It could certainly indirectly reflect risks of surface runoff, but then surely surface runoff itself would be a much better indicator of soil health? Contrary to claims made in the paper, I don’t think water content so close to the soil surface says so much about the supply of water to plants (rooting depth and water availability in deeper soil layers will be far more important). I think the authors should tone down the emphasis in the introduction of the relevance of this kind of study to soil health.

We agree that soil water content at 10 cm depth should not be interpreted as a direct indicator of plant water availability or overall soil health, as root-zone water supply and deeper soil layers play a more dominant role in these processes. Our intention was not to claim that near-surface soil moisture alone defines soil health, but rather that it reflects near-surface hydraulic conditions influencing infiltration, evaporation, and the initiation of surface runoff.

Following the reviewer’s guidance, we have toned down the wording in the Introduction. We now avoid referring to near-surface soil water content as a dominant indicator of soil health or plant water supply and instead describe it as an indicator of near-surface hydraulic functioning and land–atmosphere exchange processes, which are relevant for several soil health–related surface processes. This revision better aligns the text with the actual scope of our measurements and avoids overstatement of broader ecological implications.

3. Lines 42-43: these three papers are not cited in the reference list

We fixed it, thank you.

4. Line 50: Or (2020) is not in the reference list

We fixed it, thank you.

5. Line 51: I don’t understand how a modelling approach can be destructive? This should be clarified.

The modelling approach itself is not destructive. However, the determination of soil hydraulic parameters required to parameterize physically based models commonly relies on laboratory or field experiments involving soil sampling, excavation, or sensor installation, which can disturb the soil structure. This limits the feasibility of repeated measurements at the same location for long-term time-series analysis. We have revised the text to clarify this distinction.

Revised manuscript text (Lines 51–52):

“Additionally, soil hydraulic parameters for physically based models are usually obtained through soil sampling or sensor installation, both of which disturb the soil structure, limiting the feasibility of repeated measurements for long-term time-series analysis.”

6. Line 77: I think you could also cite Robinson et al. (2019) here (*Global Change Biology*, 25, 1895-1904).

Thanks, we added the citation.

7. Lines 79-81: It may not be related to only soil structure. What about wettability (hydrophobicity)? Could that also play a role at these time scales? Perennial vegetation can also adapt to climate change. One example comes from an earlier modelling study based on the TERENO SoilCan data (see Jarvis et al. 2022. *Hydrology and Earth System Sciences*, 26, 2277-2299).

We agree that changes in soil hydraulic behaviour over time can result not only from structural evolution but also from changes in wettability, vegetation dynamics, and land-use transitions. Our statement was intended to emphasize that, given constant soil texture over decadal time scales, structural changes are expected to be a major contributor to shifts in the soil water response function. We have revised the manuscript text to acknowledge additional mechanisms explicitly.

Revised manuscript text (Lines 79–81):

“Considering that texture remains constant at time scales of decades, the changes in the response function are likely to be related to changes in soil structure and associated hydraulic properties driven by structural reorganization, wettability effects, root activity, or land-use changes rather than to textural change.”

8. Lines 92-93: this statement is disproved by the following sentence (and also by the author’s work). I think this sentence should be deleted (nothing important is lost).

We agree that the sentence is not essential and may cause confusion in light of the following statement. We have therefore deleted it from the manuscript.

9. Line 93: The authors of this paper are incorrectly specified in the reference list.

We corrected it, thank you.

10. Line 113: No need to start a new paragraph here.

We changed it.

11. Lines 250-252: this seems very subjective. Why choose these percentiles and these scores? How were they chosen, presumably by trial and error?

The thresholds defining wet, moderate, and dry water-content situations were not chosen arbitrarily but were selected through exploratory testing on the dataset analysed in this study. We evaluated several percentile combinations and found that the 30th and 70th percentiles provided a clear separation of water-content regimes while maintaining a sufficient number of observations in each class for stable model training and evaluation.

Using narrower or more extreme thresholds (for example the 25th/75th or 95th/99th percentiles) substantially reduces the number of data points in the extreme classes. This leads to sparsely populated categories and increases the risk of overfitting to rare events rather than capturing general system dynamics. For this reason, we chose broader regime classes and allow the model to resolve variability within wet and dry ranges through the continuous input features, rather than introducing additional discrete "extreme" categories.

We have added a brief clarification in the manuscript to explain that the chosen percentile thresholds are specific to the dataset analysed in this study and were selected to balance regime separation and data availability:

Line 250-253

"These thresholds were selected to provide clear regime separation while ensuring sufficient observations per class for robust model training."

Also, there is no numerical reasoning behind assigning "wet = 1" rather than "wet = 10". The number does not represent how wet the soil is in a physical sense. It simply labels which water-content situation the soil is in at a given time step.

In the current implementation, this label enters the neural network as a numerical input. We therefore acknowledge that the chosen numeric values can influence how the model internally weights this feature. However, no physical meaning, ordering, or quantitative distance between regimes is intended, regardless of the specific numeric spacing used. The values are arbitrary identifiers applied consistently to distinguish discrete water-content situations, and no intermediate values occur in the data.

12. Line 371: write "water contents" rather than "values" (if I understood correctly)

We have replaced "values" with "water contents values" to improve clarity and specificity.

13. Line 425: "differently" not "different"

Thank you, we changed it.

14. Line 426: this seems too speculative. You don't know that it is related to structure. Maybe you could write ".... which may indicate"

We agree that the original phrasing implied a stronger causal interpretation than warranted by the data. We have revised the sentence to adopt a more cautious formulation, as suggested.

Lines 423-426:

" This indicates that for the coarse-textured soil included in this study (i) the effect of changing climatic conditions was rather small (very good NSE classification for both sites), but (ii) that these coarse-textured topsoils do not show identical response to the extreme year, as each lysimeter reacts slightly differently, which may indicate small differences in hydraulic properties."

15. Line 444: "that" not "who"

Thank you, we corrected it.

16. Lines 582-586: This is interesting, but the authors neglect the fact that WRC's measured in the laboratory do show strong correlations with texture across large regions showing

strong climate contrasts. This empirical support for texture-based estimates of the WRC is really very strong, especially in the dry region of the WRC. How can you reconcile your results with this past experience and knowledge?

We agree that laboratory-measured soil water retention curves (SWRCs) exhibit strong correlations with soil texture across large climatic gradients, particularly when compared to hydraulic properties in the wet range, and our results do not contradict this empirical relationship. Instead, our findings indicate that under natural field conditions, the effective SWRC inferred from long-term in situ soil moisture dynamics can deviate from static, laboratory-based texture estimates due to structural evolution and climatic history. In this sense, laboratory SWRCs describe intrinsic textural control, while our analysis captures additional field-scale dynamics that are not represented in texture-based estimates. We have clarified this distinction in the manuscript. In addition, as already stated in the beginning of this reply letter, the effect of soil structures in the dry range may be more relevant than expected and may also depend on the measurement method (and the use of disturbed or undisturbed samples). As was shown in Hohenbrink et al. (2023 Earth Syst. Sci. Data, 15, 4417–4432) using HYPROP method to analyse the soil hydraulic properties of undisturbed samples, there was a huge variability of conductivity and retention value also in the dry range that could not be captured by texture only.

We will add a sentence in the introduction talking about the dry range also lines were edited as follows:

Lines 584-586:

"Such context-dependent behaviour highlights the limitation of the common assumption that soils with the same texture will show comparable retention across regions, an assumption often made in the absence of better descriptors. While laboratory-measured SWRCs show strong and well-established correlations with texture across climatic gradients, particularly in the dry range, experimental evidence collected under natural field conditions indicates that this simplified description does not always hold (Hannes et al., 2016; Robinson et al., 2016; Aqel et al., 2024)."

17. Line 600: this is much too speculative. You need to write this more carefully ("may have"). There are other more plausible interpretations. It could simply be that hydraulic conductivity at and close to saturation has increased. Near-surface water contents during rainfall events of any given intensity would then be smaller. It could also be a consequence of "by-pass" water flow in shrinkage cracks.

We agree with the reviewer that the mechanistic interpretation of the observed carry-over effect was speculative and not directly supported by the analyses presented in this study. As this interpretation is not essential to the main objectives of the manuscript, we have removed the sentence to avoid over-interpretation.

Revised manuscript text (Lines 597–602)

"However, a clear carry-over effect was observed: soil water in the upper 10 cm was not fully replenished during the wet phase of autumn and winter 2019 and only reached comparable, though slightly lower, values in winter 2020. A comparable multi-year legacy across the full soil column was reported in the TERENO-SOILCan lysimeter network by Groh et al. (2020)."

18. Line 636: what is meant by "functional integrity"?

the term “functional integrity” was too vague in this context. We have replaced it with a more explicit description referring directly to soil hydraulic behaviour.

Lines 631–638:

“The classification outcomes across all lysimeters highlight the role of site memory and structural resilience in maintaining hydraulic behaviour under climatic stress. Soils assessed at their origin were more frequently classified as ‘stable’ or ‘resilient’ (e.g., Selhausen at Selhausen), while those translocated to different locations were more likely to be classified as ‘changed’ (e.g., Sauerbach at Bad Lauchstädt). These patterns indicate that soil structure, once adapted to specific climate regimes, may exhibit altered hydraulic responses when exposed to new environmental conditions. The presented methods allow detection of emerging shifts in soil hydraulic behaviour that may be relevant for soil health assessment and could serve as indicators of deteriorating soil health status.”

19. Lines 643-645: yes, especially if they are statistical models. Isn’t it reasonable to expect that physics-based models might perform better outside the training data?

Although physics-based models are commonly regarded as better suited for extrapolation beyond calibration conditions due to their explicit process representation, this assumption does not generally hold when relevant stress states are absent from the calibration data. Under such circumstances, key process parameters remain weakly constrained, and model performance can deteriorate when applied to conditions involving water or heat stress, as shown by Groh et al. (2022 Vadose Zone Journal, 21, e20202). Our statement was intended to refer primarily to data-driven or statistically trained models. We have revised the text to clarify this distinction.

Revised manuscript text (Lines 642–648)

“As remote sensing missions increasingly provide continuous global soil water content estimates, the proposed framework could be adapted for large-scale assessment of soil system stability. Furthermore, under scenarios of future climate change, where shifts in precipitation patterns and evaporative demand are expected, data-driven models trained on historical data may become progressively outdated. The presented residual-based approach (quantifying MB) enables early detection of such divergence, offering a method for identifying when model retraining or reparameterization is needed to maintain predictive reliability under non-stationary conditions.”

20. Lines 650: The extent to which the subsoil is exploited by roots should be much more important for plant water supply than the surface 10 cm.

We agree that soil moisture at 10 cm depth does not represent full root-zone water availability and therefore should not be interpreted as a direct measure of plant water supply or agricultural impact. We have revised the sentence to restrict the statement to near-surface hydraulic conditions and soil water dynamics, which are directly supported by the measurements in this study.

“Temporal variations in the water content of the topsoil control near-surface hydraulic conditions, infiltration, evaporation, and gas exchange, thereby shaping soil water dynamics and physical soil functioning. Reliable information on soil water content dynamics in response to atmospheric conditions is thus essential to detect and anticipate critical changes in soil hydraulic behaviour.”

21. Minus signs (for matric potential) are missing on all the WRC plots

Matric potential is plotted as its magnitude (|h|), following a common convention in soil physics. We have clarified this by updating the axis labels in all SWRC plots to “Matric potential |h| [m]”, so explicit minus signs are not required.