

Thanks again to the reviewers for the thorough and constructive feedback in the first round of reviews. Here, we answer the last comments of reviewer #2 and outline the performed changes.

Formatting below is indicating:

*Reviewer comments*

Authors' response

Changes in section x.x / line xxx

(where line numbers refer to the document with tracked changes)

---

## Reviewer #2

*L19-21: consider adding F1 score results.*

We added F1 scores to the abstract, as well as a mention that we evaluate the model's ability for drought event detection. **Changes: See lines 20ff.**

*L23: You state "reproduce drought indices", but provide correlation values for the entire time series, please align text and numbers.*

That is true; however, drought indices are defined for the entire time series (also neutral conditions or wet anomalies). Also, we write just before that we evaluate the model's skill to reproduce "observed anomalies" – which captures the entire time series, not only drought events. Hence, we want to keep it as it is (together with that we now also explicitly state that we evaluate drought event detection; see comment above).

*L23: Please consider removing "r " (redundant)*

Thanks for the suggestion. **Changes: removed "r" in line 22**

*Abstract: please consider adding evaluation period (length) for streamflow and groundwater, and that only 10 years and 3 stations are used for soil moisture (as these are key for your findings and lower robustness of soil moisture results).*

Thanks for the suggestion, we added "over up to 34 years evaluation period (10 years for soil moisture)". Please understand that we do not want to add all details (number of stations) in the abstract for the sake of brevity. **Changes: line 18f.**

*L304 and L315: Consider moving the sentence in line 315 (that introduces reference period) somewhere before line 304 (referring to reference period).*

We moved the sentence to the start of section 2.4.1. **Changes: Line 317 moved to line 292.**

*L607 and L724: In line with your argumentation in your response to general comment 3 (i.e. need of at least 30 years for robust drought evaluation): Consider adding "only 10 years", and briefly discuss the limited number of stations and years for soil moisture, and how that may influence the results (e.g. do you trust the results of soil moisture or could they be artifacts of the limited data in space and/or time).*

Thanks for the suggestions; indeed this is a valid addition. Changes: Line 608f: added “of 10 years” and “which potentially limits the validity of the evaluation”. Line 728: Added “from only 3 stations over 10 years”.

*Paragraph starting with L718: consider mentioning the F1 score results in the conclusions, which are backing up your statements in line 718-719 and L723-724*

We added these, in a sentence explicitly addressing drought event detection. Changes: Added “It also reproduces occurrences of moderate drought events, with F1 scores for detection of moderate drought in streamflow and groundwater between 0.56 and 0.60.” to lines 721ff.