

Second review on Peichl et al. 2021

The state-dependent clustering was removed in this version and all relevant figures have been updated accordingly. I therefore recommend the article for publication after minor revision.

Major

Table 2 and lines 180-181: In all 12 cases the test R^2 is higher than the train R^2 . In my understanding, test performance is usually below training performance. Higher test performance can occur occasionally, but it seems odd that this is the case for all your settings. Do you have an explanation for this? How did you split the train and test data set? Is there e.g. any possibility of a bias due to the splitting? Did you try various ways of splitting the data?

Related to this: You mention overfitting as an explanation of underestimation of year 2019 in line 328-329. But your better test performance compared to training performance would not indicate overfitting. (However, the following argument in line 330-334 seems reasonable.)

Minor

L. 65: What are mean average effects? Mean and average seem synonymous.

Figure 1: Isn't it 2003 instead of 2004?

Table 2: If you explain the abbreviations T and P here, you should also explain SMI and SMIa for the sake of completeness.

L. 220-221: SMI4 (5th in non-cluster); SMI8 (6th in non-cluster)

Figure 4: Why is it not completely identical for "4a) no cluster" to the submitted version? SMI4 is now on fourth place instead of SMI12.

Figure 4: Mention in the caption that this is for PAM (4).

Figure 4, A5-A9: You mention a grid size of 50, however in your answer to the reviews you state 100 (lines 139-148 of your rebuttal letter).

L. 243: "A low drought signal of soil moisture is found for April (SMI4)." Odd phrasing, consider rephrasing.

L. 268: What about cluster 1 and 3? They also exhibit drought vulnerability. Maybe write "subregion such as cluster 2..."

L. 270 and 338: crop **yield** potential

Figure 5: Missing data in gray should be indicated in the caption.

Fig. A4-A6: What is CR12?

L. 405: It is mean **absolute** error, not average. This was already mentioned in the first revision.

ALE Plots

The ALE plots are confusing because they contain a variety of small mistakes and inaccuracies. Basically I think you want to show (as indicated in lines 414-418) in these plots the 3 best cluster algorithms for cluster 2, 4 and 6 for the two soil moisture index configurations "Soil moisture for uppermost 25cm" and "Soil moisture index for both uppermost 25cm and entire soil column" (one of the plots (Fig. 4) is in the main text), right?

L. 376-378: This text was not updated. It is no longer PAM (8).

Fig. A5 is not for PAM with 2 clusters as indicated in the main text and the caption, but identical to Fig. A6.

Figs. A5-A9 all say "for both cluster". In three of these figures there are more than two. Also the plural-s is missing.

First line of the caption: Instead of "SMI" call this "soil moisture index configuration" in all ALE plots as you do in Table 2.

Fig. A5 first line of the caption: Shouldn't this be HIERARCHICAL (2) instead of PAM (2)? According to Table 2 HIERARCHICAL (2) is the best algorithm cluster size 2 and soil moisture index configuration "soil moisture for the uppermost 25 cm." Also in Line 414-415 it is mentioned as HIERARCHICAL (2).

Fig. A7 second line of the caption: 2 clusters, not 6.

Figs. A5-A9 first sentence of the caption: This is a bit misleading (in all ALE plots). It should rather say something like: best combination of cluster algorithm (PAM) and the soil moisture index configuration "soil moisture for the uppermost 25 cm" for cluster size 2. Each plot shows a different cluster size, so it is not part of the best combination, but rather it is the best combination for a given cluster size.

Typos

L. 115: extra bracket

L: 233: bracket missing

L. 240: **4d**

L. 249: **Heat8** in upper case

L 247: bracket missing