The main achievement of the analysis performed by Petrovic et al. consists of providing insights into the utility of improving spatial resolution and customizing the model setup of RCMs aimed at reproducing (and, possibly, projecting) drought characteristics. Through a simple yet straightforward experiment, they give clear answers that can drive further development of RCMs aimed at reproducing and projecting drought characteristics. I only have two main suggestions for the authors:

1) I wonder if SPEI behaviour for other aggregation scales would be the same. Would it be possible to give any information (even as supplementary material) also for SPEI-6 and SPEI-12?

2) In the Introduction, the authors declare the objective of gaining insights into drought development for Germany etc., but I can't find specific indications about that in the paper. A detailed section could help; otherwise, I suggest avoiding emphasizing this objective in the introduction.

Please find below some minor comments and typos.

L84: ERA-Interim, capital I.
L87: “concluded”. In the line below: “saw”.

Table 1: are those listed the only models available? Are there others (e.g. HIRHAM)? Please specify in the text. If some models are neglected, please explain why.

L222: to make the paper more self-consistent, I suggest providing more information about how the SPAEF metric works. On the other hand, less room can be devoted to the Mann-Kendall test, which is older and more widely known.

Section 4.1: Taylor diagrams don’t provide information about possible bias. I suggest adding this piece of information.

L296: from 50, I guess.

L314: I suggest removing Table 4 and introducing the mean correlation coefficient as an inset in the respective map (or near the title).

Section 4.3: why not use WRF@15 km here, at least in Table 5? It would further highlight the benefits of model settings.

Table 5: are the Mean SPEI values and the other statistics averaged over the German territory or the whole domain?

L347: … of this RCM …

L355: in the authors’ opinion, why do we see these results with the Tiedtke scheme? However, from Table 2, I observe that also RegCM uses Tiedtke.

Figs. 4 and 5: I would merge them. In general, I suggest always considering as two different configurations WRF@5km and WRF@15km (see my comment to Section 4.3).

Section 4.5.1: E-OBS drought frequency looks too high for some areas (up to 22/24 times over 30 years). What is the reference period on which the index is calculated? 1980-2009?
L414: so, it is mean absolute error (MAE).

Sections 4.4 and 4.5: discussion, specifically in terms of comparison to existing literature, is mostly missing.

Conclusions: maybe, the main achievements of the paper can be highlighted with bullet points (e.g., increased resolution and setup are not useful for drought characteristics; they are helpful for correlations; they are useful for trends, etc.).