The manuscript by Strandberg et al. analyses an ensemble of CORDEX simulations for a variety of climate indicator over Fennoscandia. The influence of the GCM, RCM and RCP selection on future projections is analyzed. Further, the similarity of different sub-ensembles based on the same GCM/RCM/RCP at a global warming level of +2°C is assessed. Lastly, the slope of trends is assessed with respect to the trend in global mean temperature.

Analyzing the validity of using GWLs for various metrics and the influence of GCM/RCM/RCP choices on the projected trends is of relevance.

Major concerns

1) One of your key analyses is the influence of the ensemble sub-setting on mean trends in various metrics. However, the base ensemble is too small to create sufficiently large sub-ensembles to draw robust conclusions from the comparison. For example, some of your sub-ensembles only consist of two realizations.
2) In the same way, your conclusion on the possibility to mix emission scenarios when using global warming levels, is only based on the analysis of a single global warming level (GWL2). You would have the possibility to analyze a larger range of GWLs and test whether this holds true across all GWLs or whether at some point the mixing emission scenarios becomes a problem. When you choose smaller GWL increments, then you can also include the rcp2.6 simulations.
3) In my opinion a clear discussion is missing. The discussion has partly been implemented into the results; however, I think the following key points require a dedicated discussion:
    a. how the pre-selection has influenced your sub-ensembles, meaning there have been reasons for not using specific GCMs for downscaling, i.e. for example generally bad performance of the GCM.
    b. the new insights on the aerosol problem in RCMs (see Schumacher et al 2024 (https://doi.org/10.1038/s43247-024-01332-8)); this will influence your results on local vs. global trends
    c. the influence of the bias-correction on your trends. I am not familiar with the bias adjustment scheme, so a discussion on whether the method is trend preserving or not is required. In this regard you might also want to check the raw RCM ensemble in comparison to the GCM ensemble. The bias adjustment likely reduces the spread of your ensemble.
    d. discussion on the validity to use GWLs for precipitation (e.g., maybe see Pendergrass et al. 2015 (https://doi.org/10.1002/2015GL065854) and Gampe et al 2024 (https://doi.org/10.5194/esd-15-589-2024))
    e. discuss the reasons for the large difference between the CMIP5 ensemble and the CORDEX ensemble
    f. how robust is your ANOVA analysis in the very small sub-ensembles?
    g. Generally, place your findings within the existing literature. Both, your projections and the influence of the GCM or RCM choice, as well as the use of GWLs instead of time. (e.g., Evin et al 2021 (https://doi.org/10.5194/esd-12-1543-2021), Sorland et al 2018 (https://iopscience.iop.org/article/10.1088/1748-9326/aacc77), Sobolowski et al 2025 (https://doi.org/10.1175/BAMS-D-23-0189.1), Christensen et al 2022 (https://doi.org/10.5194/esd-13-133-2022))

h. Discuss the influence of natural climate variability on your results. You discuss the uncertainty from GCM, RCM and RCP, however, the uncertainty of natural climate variability can on local scales be an important source of uncertainty.

General comments:

1) The scientific merit of your paper needs to be better presented. The paper is based on a previous iteration of analysis. While model generations and spatial resolution are mentioned as an advancement, the clear advancement over the older paper is the analysis of global warming levels and the influence of the ensemble selection. This should be mentioned more prominently.
2) Further, the introduction can profit from the following additions:
    a. At the end of the introduction extend your description of what the paper is about. Try to be more concrete. Connected to this, add a motivation for your research questions (why is it important to look at local vs. global trends? Why do we need to analyze the influence of ensemble selection?) and what the actual research gap is.
    b. a short paragraph on projected changes over Fennoscandia
3) The methods section needs to be extended to be more comprehensive and clearer.
    a. Your paper is based on both bias-adjusted RCMs and raw GCMs, however, it is not clear when and where you use which data. This could be better explained in the methods sections. Further, more information on the bias-adjustment itself is needed.
    b. Also, a better and more straightforward description of your ensemble subsetting is required. (see my suggestion further down)
    c. Several details are missing, e.g. reference period and justification why you used the old 1971-2000 period; how were the GWLs calculated
4) The figure captions and figure description in the text are often not very clear. It is often not clear what data we are looking at. The same for the figures in the supplementary.
5) The analysis in section 3.4 needs to be extended to more GWLs to back your statements.
6) You have calculated multiple indicators, however, there is no consistency throughout your analysis. You sometimes show all indicators, then focus the analysis on seasonal temperature/precipitation, in other cases on annual temperature and csu, then on all indicators again. I would suggest focusing on a few key indicators and perform all analysis on these indicators. This way we can consistently follow the influence of the different choices (e.g. rcm vs gcm, sub-ensembles, trends at different GWLs).
7) Find better section headings to clearly reflect the content of the section.

Detailed comments:

Abstract:
L 15: “The regional climate models capture the signal of the driving global models.” This statement only tells half the story. Yes, they represent the trends of their respective driving GCMs, but are under representative of the model spread in CMIP5, which means that the RCM signals here are not representative of the possible model spread of available ensembles.

L16f: “This implies that it would be safe to mix emission scenarios […]”. I am not fully convinced that you can robustly draw this statement from your analysis. You have only analysed this for a single global warming level.

Introduction:
P2, L44: “model sensitivity to …” Add model uncertainty to this list as well. For model uncertainty vs. natural variability see von Trentini et al. 2019 (https://doi.org/10.1007/s00382-019-04755-8).

P2, L47: “improvements include …” The scientific merit of this study can be strengthened more. Now it sounds more like, we have added a few models and did bias correction compared to the previous iteration.

Methods:

Table 1:
Does this list only include the official EURO-Cordex runs or did you also include additional runs, such as from Reklies-De? This would fill the GCM-RCM matrix. Or the other way around why didn't you use the additional RCM runs to fill the GCM-RCM matrix and focus on even sub-ensemble sizes?
Reklies-De: https://swift.dkrz.de/v1/dkrz_26083c6525be4627aeecde1fffc2b977/ReKliEs-De/ReKliEs-De/Internet-ReKliEs-De/startseite.html#Regio; data available at the ESGF

P5, Bias-adjustment:
Can you please clarify whether only the RCM data was bias corrected or also the driving GCMs. Also when you compare the different ensembles CMIP5-full ensemble, CMIP5-Cordex-GCM, and Cordex-RCM mention which data you compare; bias adjusted RCM trends vs. GCM non-bias adjusted trends. For me this is not clear, but this might explain some of the differences that you find. I would suggest to also include the raw RCM data in your analysis.

P5, L77: “Midas …” A bit more information on the bias adjustment method is required. For example, what is the training period; is the method trend preserving; does the day-of-year adjustment mean that for each day of the year adjustment factors were calculated.

P6, L82: “software package Climix” Include the footnote as a regular reference.

Indicator selection:
Since the indicator selection is partly based on stakeholder engagement it would be nice to learn more about the reasoning behind this selection. For example, could you give a few examples for which sectors or applications the indices are relevant for.

Selection and analysis sub-ensembles:
P7, L94-103: This entire section can be condensed. A table might convey this information more straightforward. A suggestion could be: Replace the table 1 with a matrix of GCM-RCM combinations, then either highlight for the three categories of sub-ensembles one example, e.g. highlight the row of a single GCM highlighting all RCMs that downscaled this GCM; or highlight a column with a single RCM highlighting that you are looking at the GCM spread; etc. If this is too cluttered, then have three matrices highlighting the sub-sampling method. This will make it much easier for the reader to grasp your sub setting strategy.

P7, L104: “We are here looking at GWL2, […]” Why +2°C of warming? and why didn’t you analyze multiple warming levels? Since you are discarding the rcp2.6 simulations from this analysis anyway, you could have looked at multiple warming levels. This would make your statements on “mean trends are the same at GWL2 despite the emission scenario” more robust and convincing. Also, I would here define your acronym GWL.

P7, L113: “[…] a region in northern Sweden and a region in southern Sweden (regions C and D in figure 1)” Why did you choose these two sub-regions for comparison?

P7, L113f: “The domain and time averaged data for each member together form an ensemble.” This sentence is not well connected to the previous or the following sentence.

P7, L 114: “family-wise error” Please clarify what you mean with family-wise. This is the first time you use this terminology.

Fig1 caption: “full model domain” This is misleading. The Euro-CORDEX domain is larger than the black outline. Maybe also mention on which of these domains the bias-adjustment was performed.

Results:

Figure 2:

1) Please clarify which ensemble is used for the ensemble mean trends here.
2) Please also clarify whether this is based on bias-adjusted data or the raw projections
3) This is the first time you mention your reference period. Please include this information in the methods section. Further, why is the relatively old reference period chosen instead of the recent reference period (1991-2020)? Please include the reasoning for this choice of reference period; in the methods section.

P10, L 131f: “The annual warming […]” Not clear what you mean by annual warming. Please clarify. Also, why don't you give the range of warming for rcp4.5?

P10, L 142: “more similar to” Language. maybe use "comparable to"

Figure 3:

1) Color scales are not very intuitive and are not colour-blind friendly. Maybe think about streamlining the colormaps. For each panel one needs to constantly check the colorbar to interpret the patterns. For example green can mean either a very small change (f) or very large changes (d,e,g).
2) Same comment as for Figure 1; Which ensemble? Bias-adjusted, yes/no.

Section 3.3 (RCMs compared to GCMs and the larger CMIP ensemble)
I would maybe recommend performing a bootstrapping of the large CMIP5 GCM, i.e. draw 5 (9) random GCM from the CMIP5 ensemble, then you can quantify the influence of the ensemble size on your results.

P12, L 177-181: “This could not entirely be […] the spread is much smaller” This needs to be discussed in much more detail. For example, the large spread in winter in the GCMs can in parts be explained by the presence of large internal climate variability. See von Trentini et al. 2019 (https://doi.org/10.1007/s00382-019-04755-8) on the comparison of one RCM single model large ensemble compared to the full CORDEX multi-model ensemble. Also see Maher et al 2021 (https://doi.org/10.1038/s41467-020-20635-w) showing that the CMIP5 full ensemble captures

parts of the internal climate variability given by multiple single model large ensembles. Also please clarify which "67 members are only forced by 7 unique GCMs".

P12, L181-183: "Kjellström et al (2018) […]" Are your results also pointing to this conclusion? Or is this the reason for why the spread is smaller? If this is the latter, can you please elaborate more on this. Are there other studies that have a detailed look on the conservation of GCM to RCM signals? (e.g. Taranu et al 2022, https://doi.org/10.1007/s00382-022-06540-6)
Also, what needs to discussed is the influence of bias-adjustment on the model spread.

P12, L183: "Especially, the distance between the minimum and maximum larger" Language. Please rephrase, also larger than what?

P12, L184f: "the choice of emissions scenario is of greater importance than the construction of the ensemble" Where can we see this? Please include a figure reference.

Figure 4:

1) Why not simply show boxplots with whiskers extending to min-max?
2) It is not clear what we exactly see here. Are the triangles the non-bias adjusted driving GCMs or is it the GCM-RCM used in your small ensemble, what about the squares are these also GCMs or RCMs. Maybe it would be worth to also analyse the RCM runs here once based on the raw data and once on the bias adjusted data. This would allow for a cleaner separation of the influence of the GCM ensemble selection, the influence of the RCM, and whether the bias adjustment further reduces the spread.
3) I don't know how you constructed your CMIP5 ensemble, whether it contains multiple members per GCM or whether you only used a single member per GCM. In anyway, it might make sense to only compare a single realisation of each GCM here (ensemble of unique GCMs). The reason for this is, that contributions to the CMIP5 ensemble are not evenly spread meaning that some models, e.g. the CanESM2 has contributed multiple members and they might skew the spread to one or the other side.

Section 3.4

The section needs a better title that fits the content of the section. For example, "Agreement in trends at the same GWL"

P14, L 203: "it reduces the uncertainty around the choice of emission scenarios." Where can we see this? Otherwise please add some references, for example, these references might fit (e.g., Gampe et al 2024 (https://doi.org/10.5194/esd-15-589-2024))

P14, L206: "ensemble is sensitive to how it is constructed with regards to which models" This is however true for any ensemble also the one based on time.

P14, L210: "pair-wised compared" -> "compared pairwise "

Section 3.5 and Figure 9:

1) What is the exact purpose of this section? For panel Fig 9 panel a, I can see the purpose, we see a stronger local warming than for global temperature. This fits the section header. However, for all other indicators the relationship between the x-axis and y-axis is not straightforward. We can't compare whether the local trends are stronger/weaker than global trends. If your goal is to compare the local trends vs. global trends at the same GWL, then I would suggest calculating the global mean of the respective indicator and compare the global mean trend with the local mean trend. The different dots would

then be different GWL, e.g., 1, 1.5, 2, 2.5, etc. For frost days or nzero it might make sense to calculate global trends on the same latitudes as your Scandinavian domain.

2) The figure caption and the text in the manuscript lack some description of what we actually see. Are the dots ensemble mean change in RCMs vs. the change in GCMs? Which GCMs, the CMIP5 ensemble mean GMST or the driving GCMs of the RCMs?
3) Why did you move to the time period approach again? I thought your motivation is to use GWLs instead of time.

P18, L73: "Figs 8a-c" I think this should say Fig 9a-c, right?

Summary and conclusions

P18, L290: "Trends within a GWL period" What do you mean by this? Please clarify.

P18, L291-293: "The largest difference [...] most to these variations" What about the bias correction method? Could this influence the trends as well? Did you also bias correct the GCMs or only the RCMs? especially for absolute temperature thresholds bias adjusted vs. non-bias adjusted makes a difference.

P19, L297-298: "This suggests that [...]. And that the ratio [...]" This needs to be discussed in more detail. Especially, in the light of the aerosol implementation in the RCMs.