Comment on hess-2020-673

Anonymous Referee #1

Referee comment on "Bridging the scale gap: obtaining high-resolution stochastic simulations of gridded daily precipitation in a future climate" by Qifen Yuan et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-673-RC1, 2021

Review of Bridging the scale gap: "*Obtaining high-resolution stochastic simulations of gridded daily precipitation in a future climate*" by Qifen Yuan et al. Manuscript number HESS-2020-637

This is a very meaningful study that puts forward a combination of statistical methods and approaches to produce fine scale resolution rainfall estimates for catchments (small to medium/large-ish sizes). It combines aspects of the weather-generator type approach with spatial statistics, acknowledging the need to also consider temporal dependencies and structures.

The language and graphics are clear and concise, explicit formulas are provided, accompanied by relevant references to underpinning methods and reference to used R-packages/methods.

Whilst I think that there are a one or two of issues to address method wise before it can be applied in a climate change context, I think it is a meaningful suggestion to build on. Overall, in my opinion this manuscript offers a very meaningful contribution to the peer-literature on statistical downscaling and merits publication with only minor edits.

General comments

Given the focus of this paper, I suggest a few changes that would make the paper more

accessible to a wider audience that should pay attention to this work but would find it difficult to detangle the methodology (the paper has a very strong statistical flavour and assumes that readers are familiar with statistical terminology). Firstly, I suggest that you put in a conceptual graphic, that outlines the different components in our analysis and how they fit together, this will help the reader understand how the different model parts fits together, and where the 'change' is applied (change between the two time periods). Second, I think it will help the reader if you outline when you expect the model to work well, and when you expect there to be difficulties in the methods section. Some of the performance issues discussed in the discussion section are entirely predictable and I think it will help the reader understand the method better if you have a paragraph that speaks about the expected strengths and expected weaknesses before you conduct the experiments, this will get the reader introduced to how the method works, and then the results will make more sense. I think the manuscript will be more accessible and thus more successful if it embraces a wider audience.
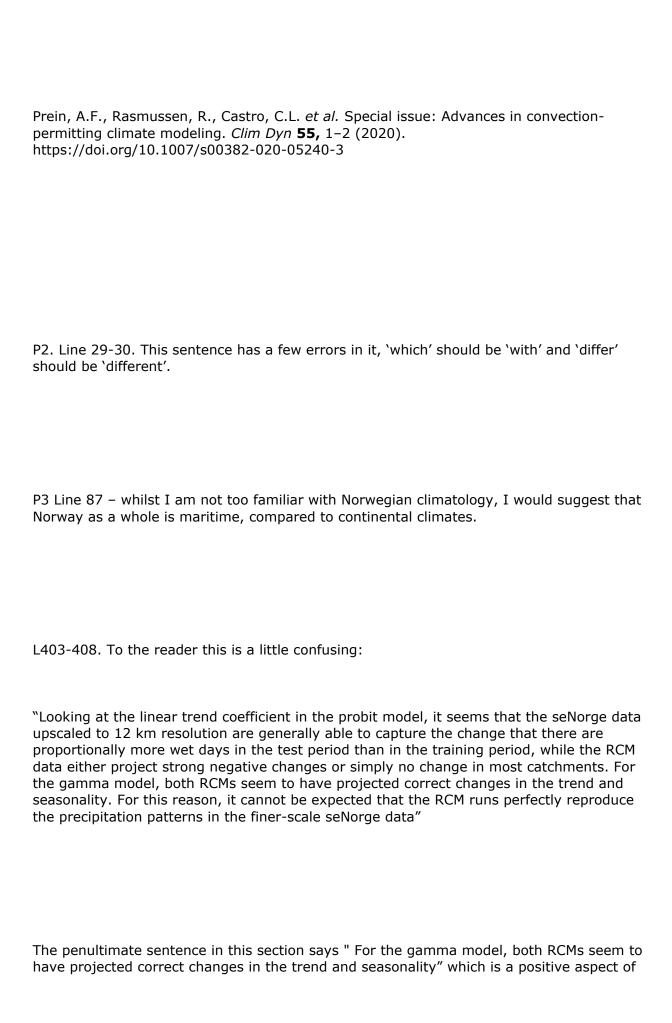
One aspect not discussed in this paper is the applicability in a climate change context. Whilst the two time periods do exhibit some climatological differences, changes that one might expect in 20-50 years time are likely to be more pronounced. In this context, you may need to include a representation of oscillatory behaviour in the spatial dependency structure (in addition to a seasonal representation). This may not be the case for Norway, but other regions with more variable rainfall climatologies (strongly influenced by e.g. ENSO or IDO … or any other large-scale climate oscillation or teleconnection pattern. I think it would be very meaningful if you could elaborate on this need in your discussion section, and perhaps propose a couple of tests to see if this is necessary, e.g. compare spatial correlation in sub-sets of RCM rainfall fields that capture certain seasonal/oscillation modes – are these still similar enough (how to judge?) to what we see in the coarse resolution observed rainfall fields?

Four smaller issues:

(1) I think it is always useful to be critical of the RCM – hence I think the authors should point out the need to demonstrate that the RCMs selected for the analysis do indeed capture the spatial and temporal patterns of the variable of interest on its resolution before downscaling (where the downscaling is drawing on a relationship derived from observed data).

(2) Using empirical scaling (or decile scaling) as a refernce is meaningful because that is a very common method. However, many water-related studies would use daily scaling rather than monthly scaling as you are draping a scaling coefficient of monthly resolution onto a variable with daily resolution, then evaluating on the daily resolution. The

comparison seems a little unfair. I see no need to redo your analysis, but I would point out how the monthly scaling could impact the comparison.

(3) Sequencing of dry days (and wet days) is very important in many regions, hence if straight forward I would contemplate including a figure that captures this information.

(4) It would be good to have a climatology map and a DEM for the catchment shown in Figure 3, it would help with interpretation of bias pattern.

Minor comments

P1. Line 17. What do you mean by 'changed runoff' (magnitude, seasonal flow, low/high flow metrics?). Perhaps qualify this a little, otherwise it looks a little 'hand-wavy'.

P2. Line 24. RCMs can explicitly resolve some process, though at 10-15 km, many processes relevant to rainfall events are parameterised. It would be more relevant here to speak about the convective permitting regional climate models that operate on 1-4 km resolution (the recent UK national projections provide output on grid resolution just over 2km). The RCM modelling community is now pushing more towards the CPM scale rather than the 10-15 km scale. There are many recent overview type publications that you could cite here e.g:

Prein, A.F., Rasmussen, R., Castro, C.L. *et al.* Special issue: Advances in convection-permitting climate modeling. *Clim Dyn* **55,** 1–2 (2020). https://doi.org/10.1007/s00382-020-05240-3

P2. Line 29-30. This sentence has a few errors in it, 'which' should be 'with' and 'differ' should be 'different'.

P3 Line 87 – whilst I am not too familiar with Norwegian climatology, I would suggest that Norway as a whole is maritime, compared to continental climates.

L403-408. To the reader this is a little confusing:

"Looking at the linear trend coefficient in the probit model, it seems that the seNorge data upscaled to 12 km resolution are generally able to capture the change that there are proportionally more wet days in the test period than in the training period, while the RCM data either project strong negative changes or simply no change in most catchments. For the gamma model, both RCMs seem to have projected correct changes in the trend and seasonality. For this reason, it cannot be expected that the RCM runs perfectly reproduce the precipitation patterns in the finer-scale seNorge data"

The penultimate sentence in this section says " For the gamma model, both RCMs seem to have projected correct changes in the trend and seasonality" which is a positive aspect of

the model. Therefore, when reading the last sentence "For this reason, it cannot be expected that the RCM runs perfectly reproduce the precipitation patterns in the finer-scale seNorge data" one is a little surprised. Presumably the performance issues relate to the first sentence in this section? To make this a little more understandable, I would replace "For this reason" in your last sentence with the actual reason, presumably the somewhat different trend coefficient in the RCMs? It is also worth noting that RCMs are not hind-casts (unless you were using reanalysis forced RCMs), rather largely free running models (following the global response to observed emissions as simulated by the driving GCM), hence you could easily end up with somewhat different temporal trends, particularly on such short time frames.