Comment on egusphere-2022-282
Anonymous Referee #1

Referee comment on "CCdownscaling: an open-source Python package for multivariable statistical climate model downscaling V1.0 " by Andrew D. Polasky et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-282-RC1, 2022

A potentially useful software package that enables comparison of different statistical downscaling results from multiple techniques with a focus on self-organising maps (SOM). As the package is adaptable to the incorporation of new and different machine learning methods, it would be useful to publish even given the limitations of the research example presented. Some general and technical comments that would improve the manuscript follow.

Line 72-73: No information is given on how or why the input variables for SOM training were selected. At least a reference to supporting work, that led to these selections, should be cited.

Line 83: In discussing Fig. 1, and within the figure itself, there is no information on how to use projected variables for climate change impact assessment. Ideally this should be included as part of the example case. For example, would variables from GCMs require bias-correction before use, so that their observed-period statistics match those of the reanalysis data?

Line 116: Regarding Fig. 2, none of the units for the colour gradient or numbers are explained.

Line 117: Regarding Fig. 3, what are the gradient plots x and y axis showing?

Line 133: Re: Table 1, where possible the units of the various metrics should be shown,
e.g. what are the units of the bias?

Line 136: You say "quantile mapping is inherently single variable", but the earlier description of SOM doesn't make it clear how it's applied jointly to multiple variables. It appears it's applied independently to precipitation and temperature, is that not the case?

Line 183: Re: Fig. 4, what do the dimensions of a SOM refer to? i.e. what are the horizontal and vertical axis representing?

Line 188: So choice of SOM size is somewhat subjective, based on expert knowledge, rather than any objective or automated optimisation? Is that a weakness?

Line 192: Qmap appears to perform best across the full distribution, is that the case and so worth mentioning?

Table 2 says "Random Forest does the best job of matching the day-to-day values, with the lowest RMSE", but actually Qmap has the lowest RMSE.