Many thanks to both the editor and the reviewers for their careful reading of the manuscript. We attach an updated version addressing the points raised.

49-51: "...ability of a model to project a certain change". It's not clear to me what this means. I assume you're talking about fitness-for-purpose here. E.g. a model's skill might vary for different regions of the world, or for different variables of interest. Please expand this sentence to make it clearer.

Thanks for the comment, we have expanded the paragraph as follows:
"The weights should also be representative of the question at hand: skill is not a property of the model *per se*, but indicative of the ability of a model to project a certain change (Parker et al 2009). In other words, a climate model is fit for purpose if it can adequately represent the response of relevant physical processes in the required range of boundary conditions. This assessment of adequacy might change based on the regions and variables in question."

126: "...contains on the ..." A typo, I think?

Thanks. Fixed.

139: should say "For each variable, v, ..."

Fixed.

144: "monthly climatology". But table 1 says "seasonal". Is there another aggregation step here? Or is one of these incorrect?

Thanks for catching this. Sorry - version control. This version is seasonal data only.

146-147: "area-weighted root mean square difference over the domain". This needs a little more explanation. Presumably, model output is gridded (but on different grids, depending on the model), and some (all?) of the observational data is gridded. I guess "area-weighted" suggests you've done something to handle the grid disparities. But I don't understand the steps involved in getting from the individual data points to an overall distance measure for each variable. Please elaborate (if only so that others could follow the same steps if they want to use the same analysis for other purposes).

Added the following:
"All observations and model data are first linearly interpolated to a common 1 by 1 degree grid and 17 vertical levels. "

147 "The matrix is then normalized". I think this should say "Each matrix...", or did I misunderstand? And is it normalized against the entire (inter-model) matrix mean, or just the mean distances of each model to the observations?

Adjusted as follows:
"Each matrix corresponding to each variable is then normalized by the mean pairwise inter-model distance, such that for each field in Table 1, there is a n+1 by n+1 matrix representing the pairwise distance between each model (and the observations)."

160: "The inter-model distance matrix is also computed from the inter-model distance matrix". This sentence doesn't make any sense.

Sorry - yes. Corrected to:
"The independence weights can be computed from the inter-model distance matrix δ. "

202: The notation here is very awkward. Is it necessary to include the 20c superscript? You don't reference it anywhere in the text.

It was for consistency with our previous paper where we used different time periods, but agreed - it makes little sense out of context here. Removed.

Figure 6: Please label the x axis.

Done

386: "tailer" -> “tailor"
Corrected