Performance of ESPO-G6 v1.0

Juliette Lavoie, Pascal Bourgault, Travis Logan, Trevor J. Smith

5th January, 2023

In this document, we analyze the performance of ESPO-G6 v1.0. Our goals are to confirm that the bias adjustment is working correctly, to find its strengths and weaknesses, and to serve as a benchmark for future versions. We use a similar framework to the VALUE project [MW15] for our diagnostics. A diagnostic is built with a property (called "indices" in the VALUE project) and a measure. Properties are evaluating a statistical characteristic of a dataset by collapsing the time axis. Measures are evaluating the difference in a property between two datasets. Properties are divided into four aspects: marginal, temporal, multivariate and spatial. We calculate the properties on the three variables of ESPO-G6 v1.0: maximal daily temperature (`tasmax`), minimum daily temperature (`tasmin`) and mean daily precipitation flux (`pr`).

For this analysis, we compute the properties on ERA5-Land, on the regridded simulations, and on the regridded and bias-adjusted simulations. We'll call these datasets, respectively: reference, simulation, and scenario. Then, we compute the measures between the reference and the simulation, as well as between the reference and the scenario. This is done over the daily timeseries of the 1991-2020 period, for each model, for each experiment and for three diagnostic regions. The three small regions, shown in Figure 1, are chosen to be representative of climates present in the full North America domain. It would have been very computationally expensive to compute the diagnostics over the whole domain. An example of one property and measure for one model and one experiment is shown in Figure 2. All the diagnostics that we compute are summarized in Table 1. The code to compute them and more details on their implementation, including sources, can be found in the modules xclim.sdba.properties and xclim.sdba.measures.

Table 1: Diagnostics computed to assess the performance of ESPO-G6 v1.0.

Property	Short name	Variables	Measure	Aspect
First percentile	q01	`tasmax`, `tasmin`	bias	marginal
95th percentile	q95	`pr`	bias	marginal
99th percentile	q99	`tasmax`, `tasmin`, `pr`	bias	marginal
Dry spell frequency	dry_spell_freq	`pr`	bias	marginal
Amplitude of the annual cycle	aca	`tasmax`, `tasmin`	bias	temporal
Relative amplitude of the annual cycle	aca	`pr`	ratio	temporal
Dry-Wet Transition	dry_wet_transition	`pr`	bias	temporal
Wet-Wet Transition	wet_wet_transition	`pr`	bias	temporal
Maximum length of dry spell	max_length_dry_spell	`pr`	bias	temporal
Maximum length of warm spell	max_length_warm_spell	`tasmax`	bias	temporal
Intervariable correlation (Spearman)	corr	`tasmin-tasmax`, `pr-tasmax`	bias	multivariate
Decorrelation length (threshold = 0.5)	decorrelation_length	`tasmax`, `tasmin`, `pr`	bias	spatial

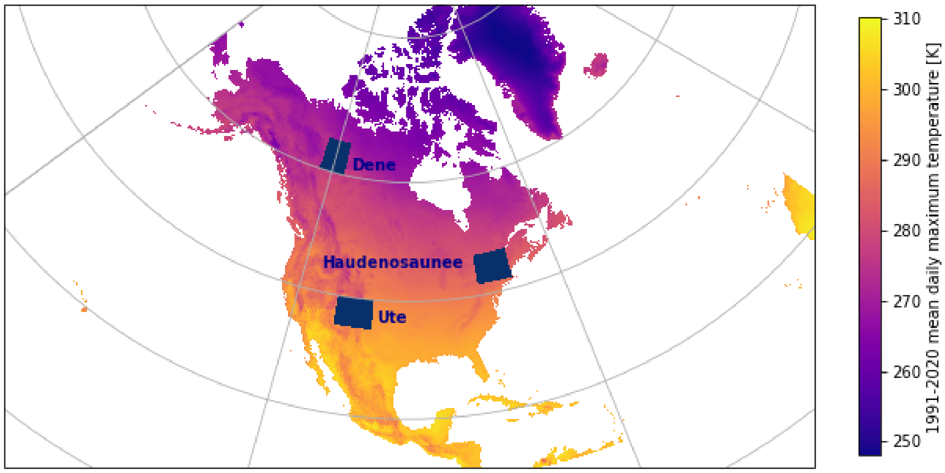

Figure 1: The domains of the three diagnostic regions are shown in dark blue over the complete domain of ESPO-G6 v1.0. The three regions have the same number of grid cells.

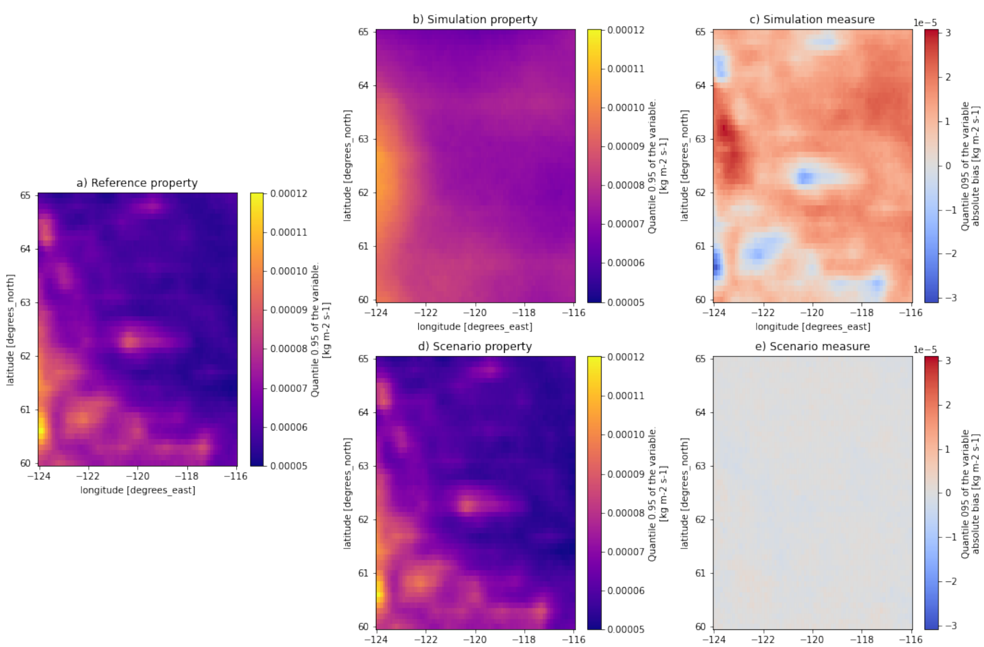

Figure 2: 95th quantile of the precipitation (property in a,b,d) and its biases (measure in c,e) during the 1991-2020 period over the Dene region using ERA5-Land reference (a), the MPI-ESM1-2-HR SSP2-4.5 simulation (b,c) and the MPI-ESM1-2-HR SSP2-4.5 scenario (d,e).

Instead of showing a similar figure to Figure 2 for each model, experiment and property, we create a summary figure per region (Figure 3) with a new metric, the fraction of improved grid cells (IMP). IMP is calculated as the fraction of grid cells of the scenario measure (e.g. Figure 2e) that are "better" than the simulation measure (e.g. Figure 2c). In the case where the measure is a bias, "better" means closer to 0. In the case where the measure is a ratio (e.g. for the relative amplitude of the annual cycle), "better" means closer to 1.

$$IMP = \frac{1}{N}\sum_{i,j} I_{i,j} \qquad \text{where} \qquad I_{i,j} = \begin{cases} \begin{cases} 1 & \text{if } |M_{i,j}^{sim}| > |M_{i,j}^{scen}| \\ 0 & \text{if } |M_{i,j}^{sim}| < |M_{i,j}^{scen}| \end{cases} & \text{if M is a bias} \\ \begin{cases} 1 & \text{if } |M_{i,j}^{sim} - 1| > |M_{i,j}^{scen} - 1| \\ 0 & \text{if } |M_{i,j}^{sim} - 1| < |M_{i,j}^{scen} - 1| \end{cases} & \text{if M is a ratio} \end{cases} \tag{1}$$

and where M is the measure of the bias between the simulation (sim) or scenario ($scen$) data and the reference (ref) and N is the number of grid cells (i,j) in the region.

Using this metric, we can see that ESPO-G6 v1.0 is generally well adjusted. Thought, there is still room for improvement in future versions of the dataset. This analysis will be useful to target where we should focus our efforts. General trends are the same for all regions. Figure 3 shows results for Haudenosaunee and the figures for Dene and Ute are shown in the appendix. Data for the diagnostics of Haudenosaunee is provided in the `data/` folder of the github repository. The rest of the data can be provided upon request.

An important caveat to bring up for this analysis is that we are assuming that the reference dataset is the "truth". Unfortunately, the ERA5-Land dataset is not a perfect reflection of reality. Still, an analysis was conducted to identify this reanalysis as the best option compared to other available datasets (see ReadMe of ESPO-G github repository). Hence, we go ahead while noting that the flaws of the reference dataset are a limitation of our analysis.

The following sections go into more detail on the performance of each aspect.

1 Marginal

As expected, the detrended quantile mapping bias adjustment method performs generally very well for the marginal aspect of `tasmax` and `pr`. Indeed, this method is made to adjust the quantile directly (Figure 4). As an example, we can see in Figure 2, there is barely any bias left in the scenario measure. This matches the corresponding IMP of 97% seen in Figure 9.

The story is a bit different for `tasmin` which was not adjusted directly. Indeed, in order to avoid temperature inversions (`tasmax<tasmin`), we adjusted the daily temperature range (`dtr`) and reconstructed `tasmin` afterward. This could partly explain that a few models get a worst performance, after the adjustment, for the 1st and 99th quantile of daily minimum temperature. Although, the performance might not be as bad as Figure 3 makes it seem. Figure 5 shows a more complete story. Indeed, in panel d, we can see that the scenario reproduces the spatial pattern of the reference much better than the simulation even if overall the correction creates an unrepresentative warming of `tasmin` (created by a low `dtr`). We note that we need to be careful with IMP, although it is a useful tool to get a quick look at the performance, it does not show the full picture.

2 Temporal

The bias adjustment method is applied to each day of the year. Hence, we are expecting the annual cycle to perform well. The average IMP of the amplitude of the annual cycle of temperatures is 97% over all models, experiment and regions. For the relative amplitude of precipitation, it is 87 %. This drop could be explained by a weaker annual cycle in some regions compared to temperature.

Comparatively, the properties looking at a sequence of days have not been explicitly corrected for, but they still performed reasonably well with an average IMP of 72% for maximum length of warm spell, 70% for maximum length of dry spell and 92 % for wet-wet transition. The weakest property is dry-wet transition with 56%.

3 Multivariate

Our bias adjustment method is univariate, in the sense that each variable is corrected separately. However, the workflow for each variable is not completely independent as `tasmin` is reconstructed from `tasmax` and `dtr`. This could explain in part the mean IMP of 92% for the correlation between `tasmax` and `tasmin`. Although, the IMP

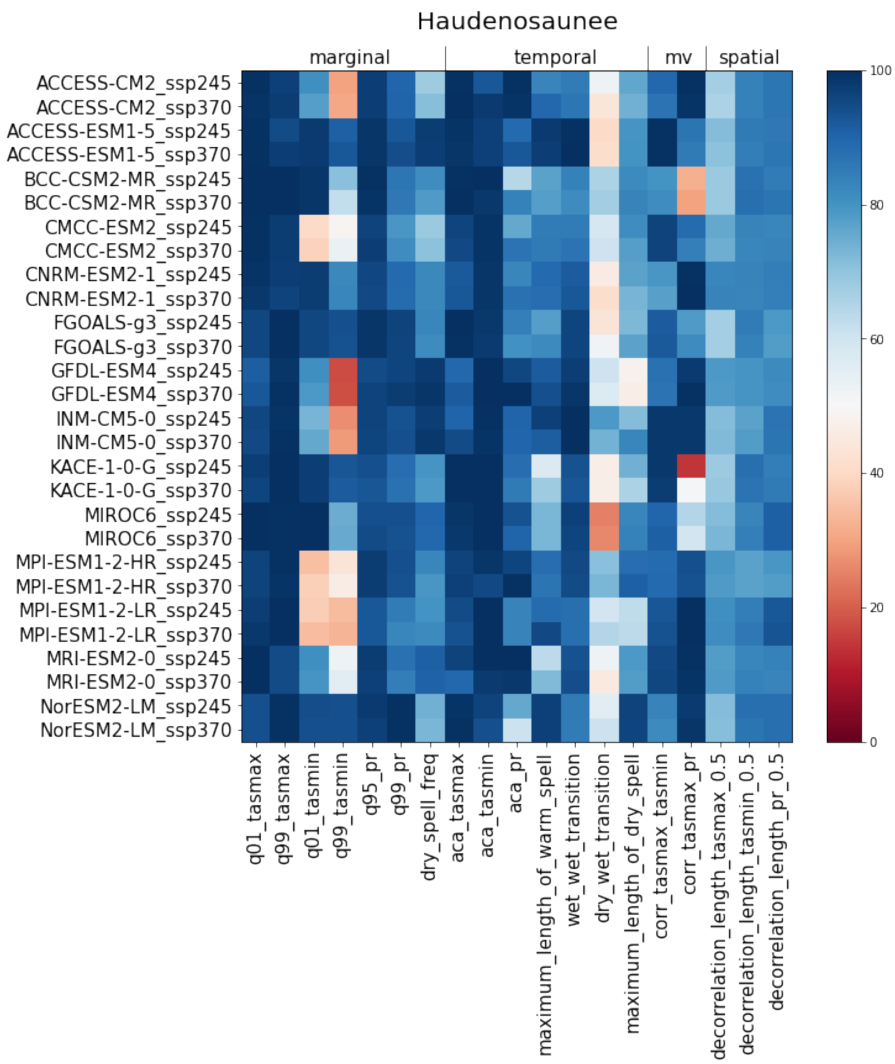

Figure 3: Heatmap of the percentage of improved grid cells between the simulation and the scenario in the region of Haudenosaunee. The columns represent the properties (identified by their short name) and the row represent the dataset (identified by "model_experiment"). When the bias adjustment worked well, the fraction should be close to 100%.

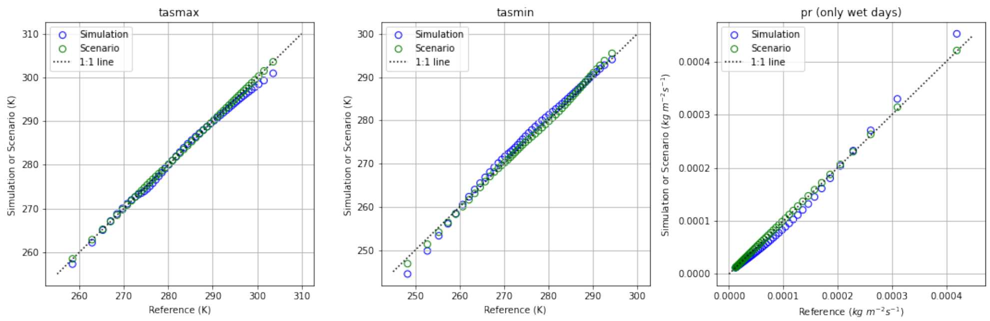

Figure 4: Q-Q plots for INM-CM5-0 SSP3.7-0 including all grid cells of the Haudenosaunee region. For precipitation, the plot is created with only wet days (more than 1 mm/d).

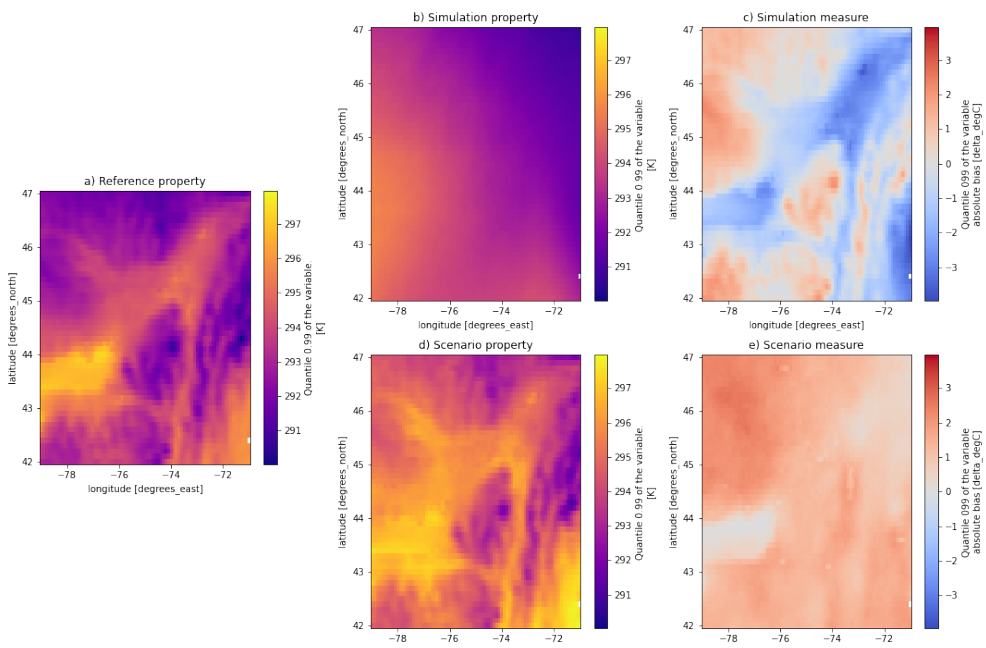

Figure 5: 99th quantile of the daily minimum temperature (property) and its bias (measure) during the 1991-2020 period over the Haudenosaunee region using ERA5-Land reference, the ACCESS-CM2 SSP2-4.5 simulation and the ACCESS-CM2 SSP2-4.5 scenario.

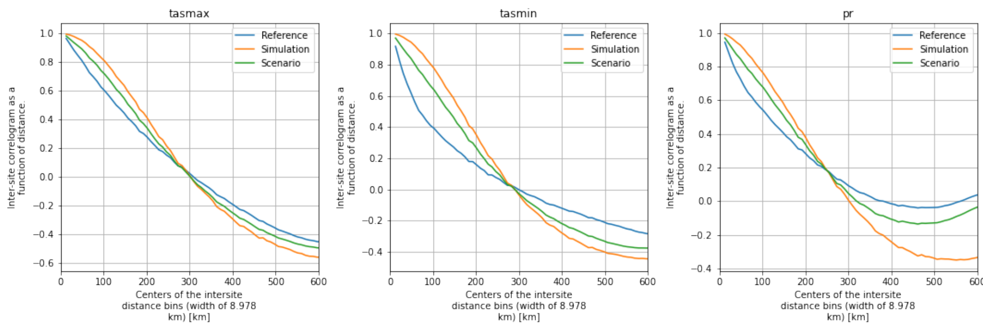

Figure 6: Correlograms `tasmax`, `tasmin` and `pr` in the Ute region during the 1991-2020 period using the ERA5-Land reference, the MRI-ESM2-0 SSP3-7.0 simulation and the MRI-ESM2-0 SSP3-7.0 scenario.

for the correlation between `tasmax` and `pr` is also high (87%) even though, they were not corrected together. The mean IMP is large, but still a few models that see a worsening in the scenario. In those cases, the correlation is generally small.

4 Spatial

Our bias adjustment method does not directly correct spatial features, each grid cell is corrected individually. However, we still see an improvement of the decorrelation length with mean IMP above 80% for all variables. This property is supplemented by the full correlograms (Figure 6). For all variables, the simulation has a higher correlation than the reference. This behaviour comes from the fact that the simulation initial resolution is much coarser than the reference grid on which it was regridded. It makes sense that the grid points of the finer grid that used to be in a single cell are strongly correlated. The scenario improves on that feature, but still stays more correlated than the reference.

5 Climate Change Signal

By removing the trend on a 30-year window before the quantile mapping and reapplying the trend afterward, we make sure that the climate change signal is preserved. To verify this, we calculated the 30-year means timeseries and subtracted the reference period 30 years. As expected, the simulation and the scenario timeseries are the same (not shown).

6 Ensemble Variability

Figure 7 shows the annual timeseries of three indices. The indices are given as change from the 1991-2010 period mean, which allows us to see that the bias-adjustment did not change the ensemble spread in a significant way. We infer this to signify that the interannual variability of each member was preserved.

References

[MW15] Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., & Wilcke, R. A. I. (2015). VALUE: A framework to validate downscaling approaches for climate change studies. Earth's Future, 3(1), 1–14. https://doi.org/10.1002/2014EF000259

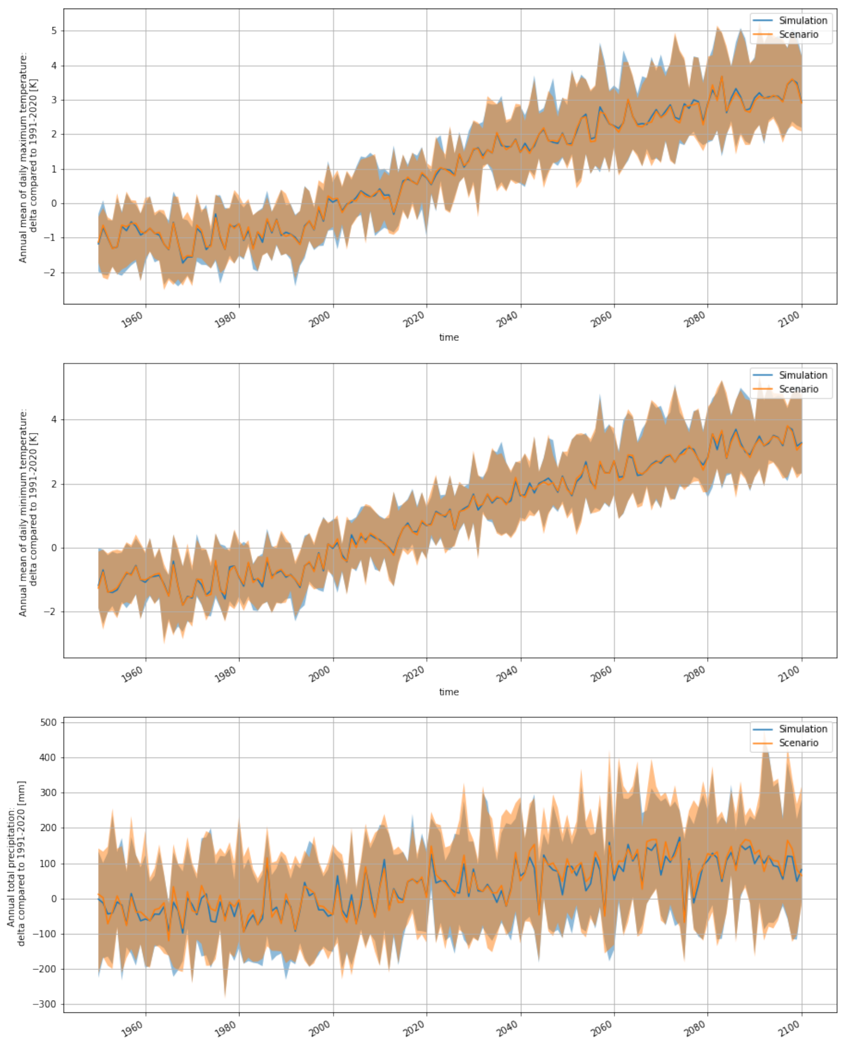

Figure 7: Simulation and scenario ensemble spread of the change in three annual indices: mean daily maximum temperature, mean daily minimum temperature and total precipitation. The change is computed with reference to the 1991-2020 period mean.

A Appendix

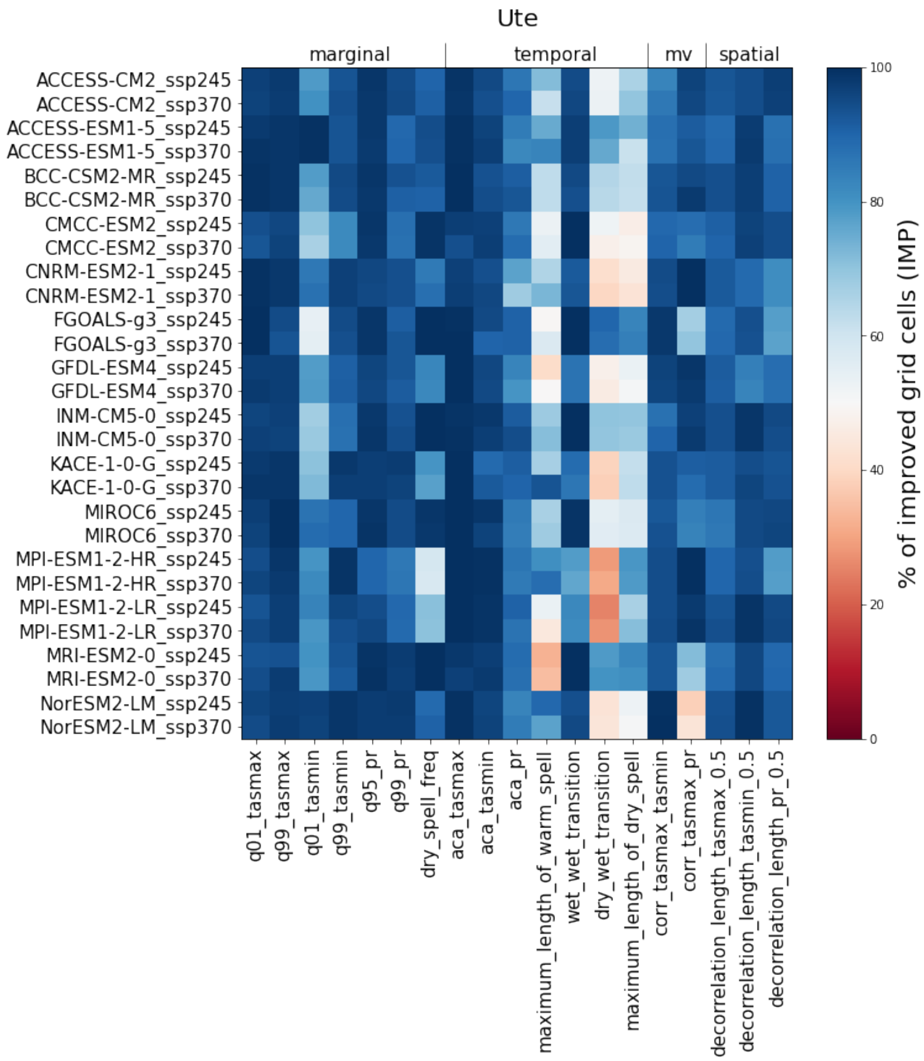

Figure 8: Heatmap of the percentage of improved grid cells between the simulation and the scenario in the region of Ute. The columns represent the properties (identified by their short name) and the row represent the dataset (identified by "model_experiment"). When the bias adjustment worked well, the fraction should be close to 100%.

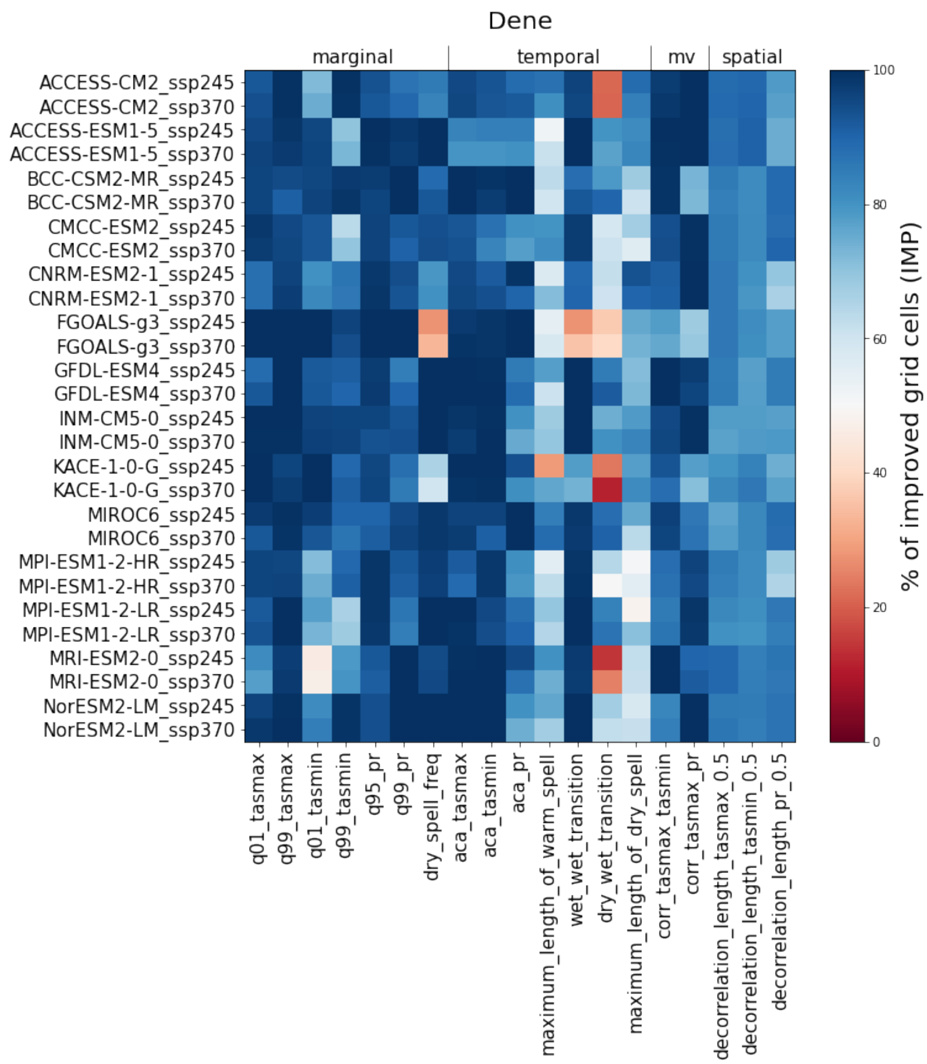

Figure 9: Heatmap of the percentage of improved grid cells between the simulation and the scenario in the region of Dene. The columns represent the properties (identified by their short name) and the row represent the dataset (identified by "model_experiment"). When the bias adjustment worked well, the fraction should be close to 100%.