Performance of ESPO-G6-R2 v1.0.0

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

4th April, 2023

1 Context

In this document, we analyze the performance of ESPO-G6-R2 v1.0. This analysis is following the same format as the one for ESPO-G6-EL v1.0 (previously name ESPO-G6 v1.0 and found in the documentation section of release v1.0.0). The main difference between ESPO-G6-R2 and ESPO-G6-EL is the reference dataset. The ESPO-G6-EL used ERA5-Land while ESPO-G6-R2 uses RDRS v2.1. For now, ESPO-G6-R2 only covers the domain of Quebec, where most of the users are located. Hence this analysis will be focused on this region. Eventually, ESPO-G6-R2 will be computed for the full NAM domain. This analysis is not meant to compare ESPO-G6-EL and ESPO-G6-R2, but rather to inspect ESPO-G6-R2 v1.0 as a stand-alone.

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. Each diagnostic is based on 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 three aspects: marginal, temporal and multivariate. We calculate these properties for the three variables of ESPO-G6-R2 v1.0: maximal daily temperature (`tasmax`), minimum daily temperature (`tasmin`) and mean daily precipitation flux (`pr`). Further, we inspect the spatial structure and verify the conservation of ensemble variability.

2 Diagnostics

For this analysis, we compute the properties on RDRS, on the regridded simulations, and on the regridded and bias-adjusted simulations, which we call the reference, simulation and scenario, respectively. The measures are then calculated between the reference and the simulation, as well as between the reference and the scenario. This is done over the daily time series of the 1989-2018 period for each model and each experiment. This is the same period used a reference in the bias adjustement. It is the most recent 30-year period available in RDRS v2.1. An example of one property and its measure for one model and one experiment is shown in Figure 1. The complete list of diagnostics computed is provided 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.

In order to summarize the analysis across all models, experiments and properties, Figure 1 shows the fraction of improved grid cells (IMP) for the three regions. IMP is calculated as the fraction of grid cells of the scenario measure (e.g. Figure 1e) that obtain a better score than the simulation measure (e.g. Figure 1c), which means either a smaller bias or a ratio closer to 1. As shown in Table 1, one type of measure is associated with each property.

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

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.

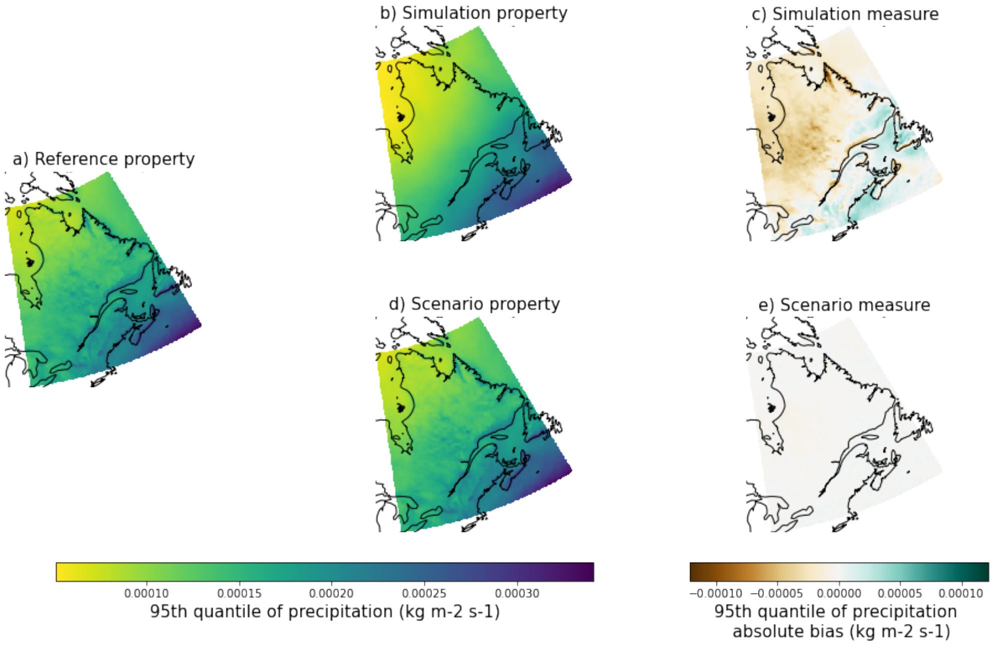

Figure 1: 95th quantile of the precipitation (property in a, b, d) and its biases (measure in c, e) during the 1989-2018 period using the RDRS reference (a), the KACE-1-0-G SSP3-7.0 simulation (b, c) and the KACE-1-0-G SSP3-7.0 scenario (d, e).

Table 1: Diagnostics computed to assess the performance of ESPO-G6-R2.

Property	Short name	Variables	Measure	Aspect
Mean	mean	`tasmax`, `tasmin`, `pr`	bias	marginal
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	`pr`	bias	temporal
Wet-Wet Transition	wet_wet	`pr`	bias	temporal
Maximum length of dry spell	max_dry_spell	`pr`	bias	temporal
Maximum length of warm spell	max_warm_spell	`tasmax`	bias	temporal
Inter-variable correlation (Spearman)	corr	`tasmin-tasmax`, `pr-tasmax`	bias	multivariate

Using this metric, we can see that ESPO-G6-R2 v1.0 is generally well-adjusted (Figure 2). 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 next. Results are shown for SSP3-7.0, but results are similar for SSP2-4.5.

An important caveat to bring up for this analysis is that we are assuming that the reference dataset is the "truth". Unfortunately, the RDRS dataset is not a perfect reflection of reality. What this analysis tells us is that the bias adjustment did a good job of bringing the model close to the reference reality. RDRS v2.1 was chosen over ERA5-Land for this version because it is closer to station data and has a better modelisation of lakes. 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.

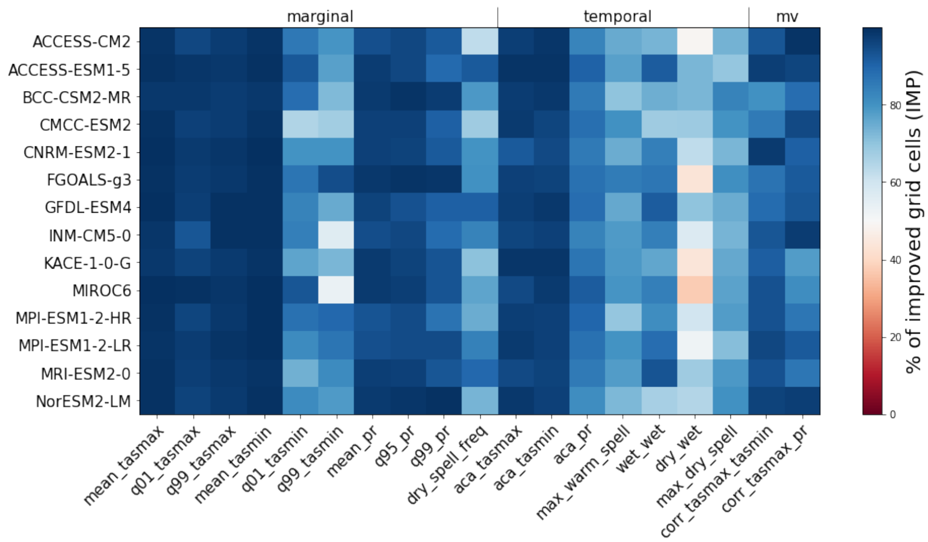

Figure 2: Heatmap of the percentage of improved grid cells between the simulation and the scenario. The columns represent the properties (identified by their short name) and the row represent the dataset (identified by "model_experiment"). When the bias adjustment worked well, the fraction should be close to 100%.

2.1 Marginal

As expected, the detrended quantile mapping bias adjustment method performs generally very well for the marginal aspects of `tasmax` and `pr`. This is not surprising since this method allows to adjust all quantiles separately (Figure 3). As an example, we can see in Figure 1 that there is barely any bias left in the scenario measure. This matches the corresponding IMP of 96% seen in Figure 2.

The story is a bit different for `tasmin` which was not adjusted directly. Indeed, in order to avoid temperature inversions (`tasmax<tasmin`), we adjusted `dtr` and reconstructed `tasmin` afterwards. Hence, an underestimation `dtr` could lead to an overestimation of `tasmin`. This could partly explain why the performance is not as good as for the other variables for the 1st and 99th quantile of daily minimum temperature. Although, the performance might not be as bad as Figure 2 makes it seem. Figure 4 shows a more complete story. Indeed, we can see that the scenario (d) reproduces the spatial pattern of the reference much better than the simulation (b) even if there is a warm bias and more grid cells of the simulation are closer to the reference than the scenario. 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 account for the spatial structure.

2.2 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 maximum temperatures is 97% over all models. 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 most of them still performed reasonably well with an average IMP of 77% for maximum length of warm spell, 76% for maximum length of dry spell and 81 % for wet-wet transition. The weakest property is dry-wet transition with 59%. Figure 5 shows that, for this property, there is very little change between the simulation and the scenario.

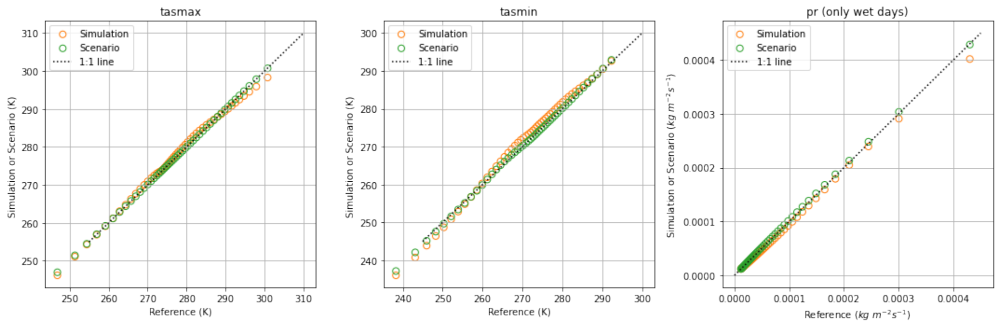

Figure 3: Q-Q plots for INM-CM5-0 SSP3.7-0 including all grid cells of the Quebec region for the 1989-2018 period. For precipitation, the plot is created with only wet days (more than 1 mm/d).

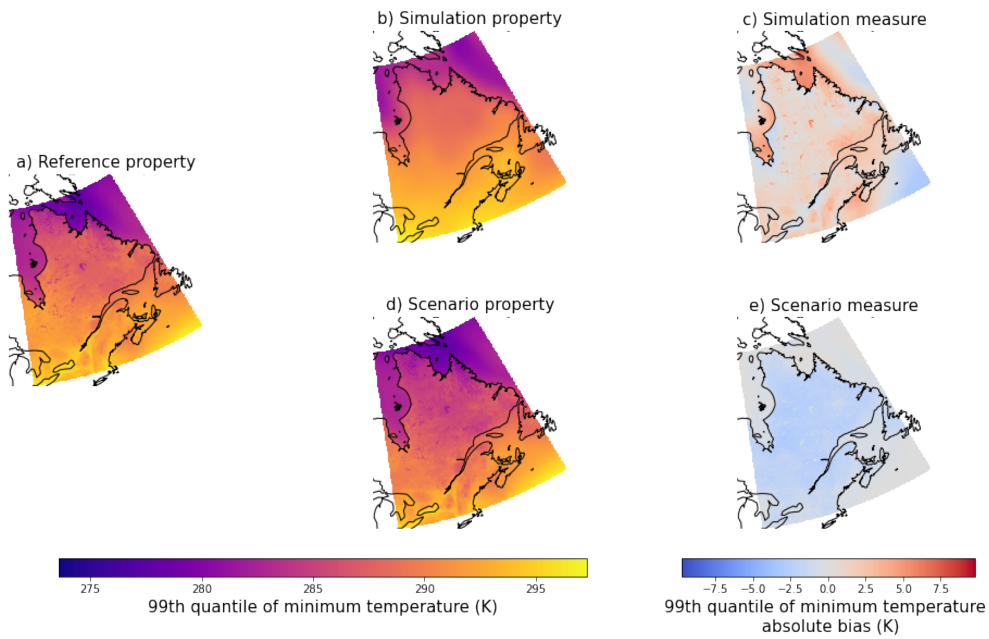

Figure 4: 99th quantile of the daily minimum temperature (property) and its bias (measure) during the 1989-2018 period using the RDRS reference, the MIROC6 SSP3-7.0 simulation and the MIROC6 SSP3-7.0 scenario.

The simulation is slightly better than the scenario, but not by a large amount. This degradation might be due to the second pre-processing step of bias adjustment which adapts the frequency of dry days. We consider this step, and accompanying low performance in the dry-wet transition property, necessary to avoid an important wet bias.

2.3 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 for the correlation between `tasmax` and `pr` is also high (91%) even though they were not corrected together.

3 Spatial

Our bias adjustment method does not directly correct spatial features, each grid cell is corrected individually. On top of a visual inspection of the maps of the diagnostics from the previous section, we compute correlograms (Figure 7) of the three variables on a smaller region (see blue contour Figure 6) to validate the spatial structure. The smaller region was chosen because computation cost of spatial correlograms are very high. For all variables, the simulation

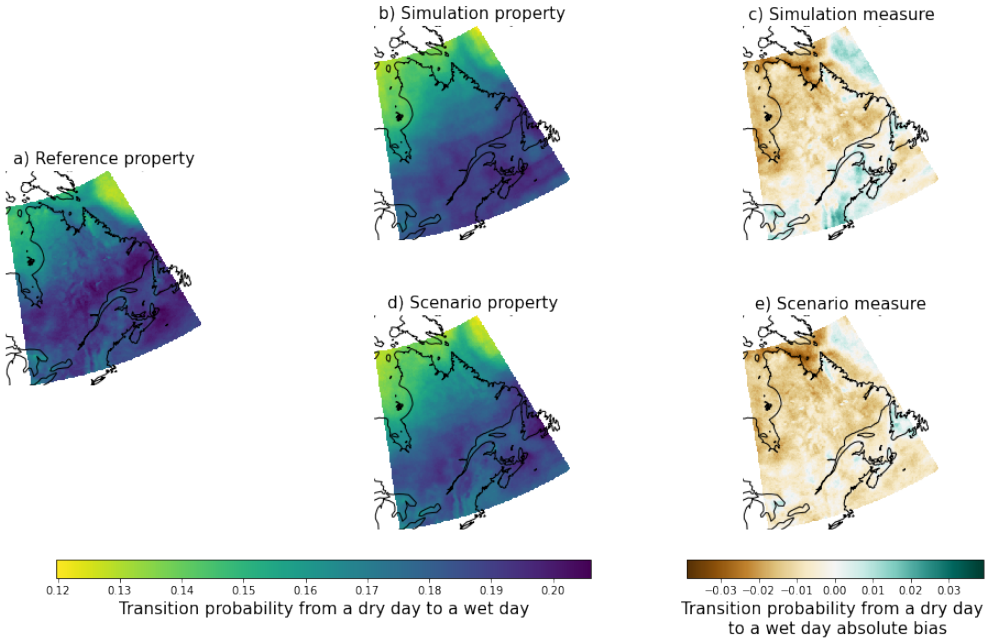

Figure 5: Transition probability of going from a dry day to a wet day (property) and its bias (measure) during the 1989-2018 period using the RDRS reference, the FGOALS-g3 SSP3-7.0 simulation and FGOALS-g3 SSP3-7.0 scenario.

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. For `tasmin` and `tasmax`, the scenario improves on that feature, but still stays more correlated than the reference. For `pr`, the correlogram of the scenario stays very close to the simulation.

4 Ensemble Variability

Figure 8 shows the ensemble spread of the annual time series of change (compared to the 1989-2018 period) of three indicators. We can say that the ensemble variability is conserved as the ensemble spread (90th percentile - 10th percentile) is similar for the scenario and the simulation. From this figure, we can also see that the median of change of the scenario and simulation are similar, confirming that the climate change signal is conserved.

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

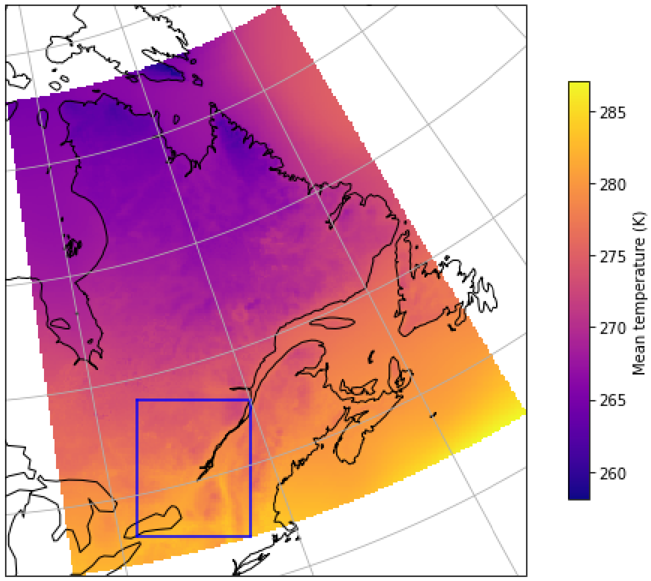

Figure 6: Mean temperature for the year 1991 for the scenario MIROC6 SSP3-7.0 to represent the full Quebec domain. The blue rectangle represents the small region on which the spatial correlogram is computed.

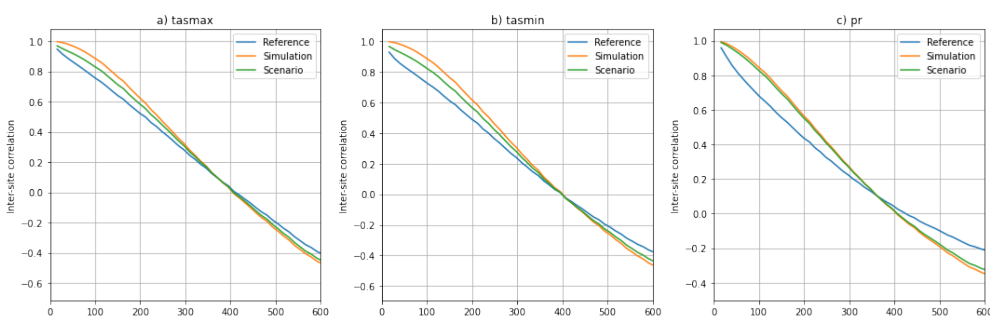

Figure 7: Correlograms `tasmax`, `tasmin` and `pr` over the small region during the 1989-2018 period using the RDRS reference, the MRI-ESM2-0 SSP3-7.0 simulation and the MRI-ESM2-0 SSP3-7.0 scenario.

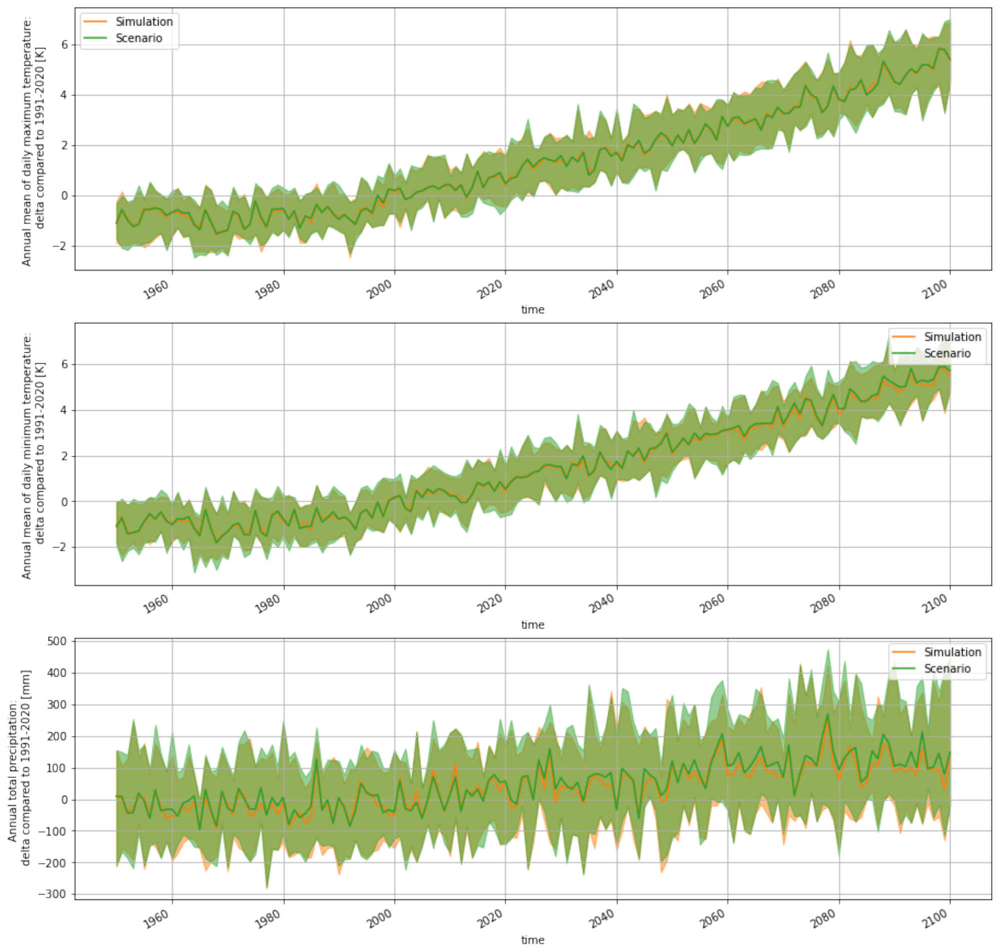

Figure 8: Simulation and scenario ensemble spread of the change in three annual indices: mean daily maximum temperature, mean daily minimum temperature and total precipitation. The change is computed with reference to the 1989-2018 period mean. This is computed for one grid cell near Montreal.