Performance of ESPO-G6-R2 v1.0.0

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

Ouranos: Consortium on Regional Climatology and Adaptation to Climate Change

26th 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-E5L 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-E5L is the reference dataset. The ESPO-G6-E5L used ERA5-Land while ESPO-G6-R2 uses RDRS v2.1. This analysis is not meant to compare ESPO-G6-E5L 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.

2 Diagnostics

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 differences 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`).

For this analysis, we compute properties on RDRS v2.1, on the regridded simulations, and on the regridded and bias adjusted simulations, which we call the reference, simulation and scenario, respectively. Measures are then calculated between the reference and the simulation, as well as between the reference and the scenario. 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.

The diagnostics are computed over the daily time series of the 1989-2018 period for each model and each experiment. This is the same period used in the bias adjustment training phase. It is the most recent 30-year period available in RDRS v2.1. The diagnostics are calculated over the Magtogoek region (green contour on Figure 1). It would have been very computationally expensive to compute the diagnostics over the whole North American domain and this subregion contains most of our users. This document presents figures for SSP3-7.0, but results are similar to SSP2-4.5.

In order to summarize the analysis across all models, experiments and properties, Figure 2 shows the fraction of improved grid cells (IMP). IMP is calculated as the fraction of grid cells of the scenario measure (e.g. Figure 3e) that obtain a better score than the simulation measure (e.g., Figure 3c), 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.

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

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

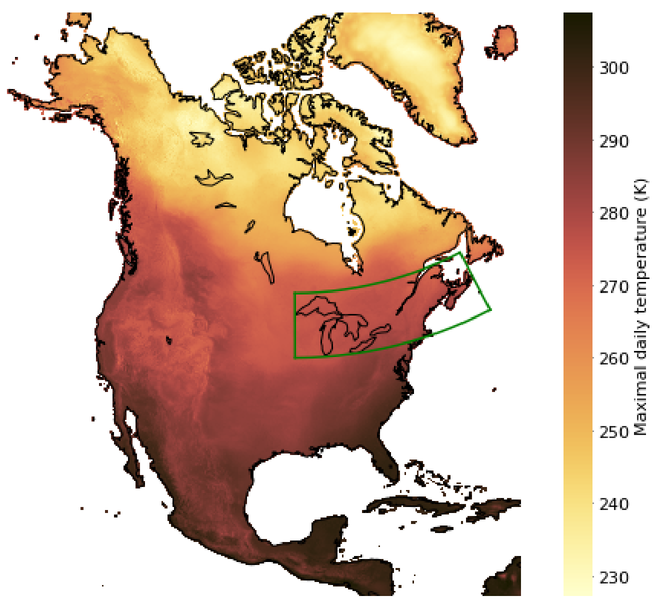

Figure 1: Maximal temperature on January 1, 1991 for the scenario MIROC6 SSP3-7.0 to represent the full North American domain. The green contour shows the Magtogoek region.

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-R2 v1.0 is generally well adjusted (Figure 2). Though, 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.

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. The closeness of RDRS v2.1 to the real world is out of the scope of this document.

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

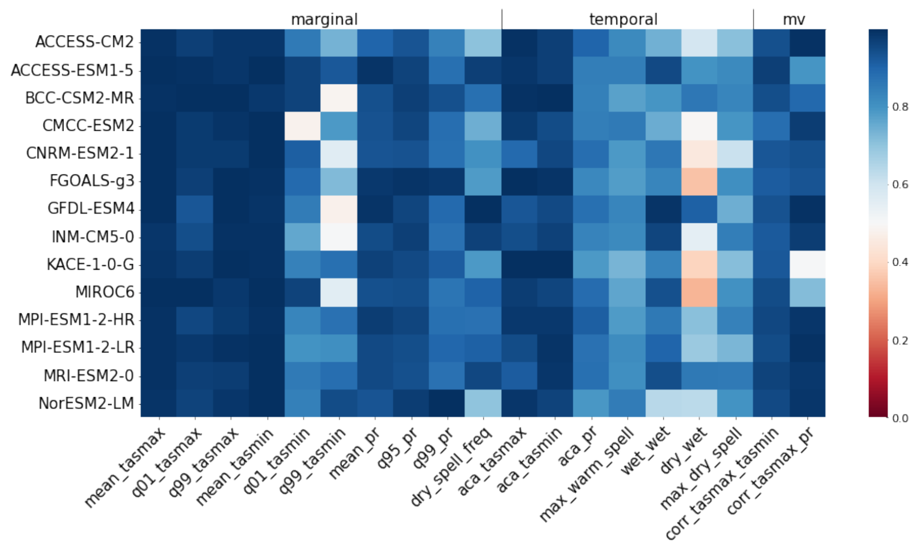

Figure 2: Heatmap of the fraction of improved grid cells between the simulation and the scenario in the Magtogoek region. The columns represent the properties (identified by their short name) and the row represent the models. When the bias adjustment worked well, the fraction should be close to 1.

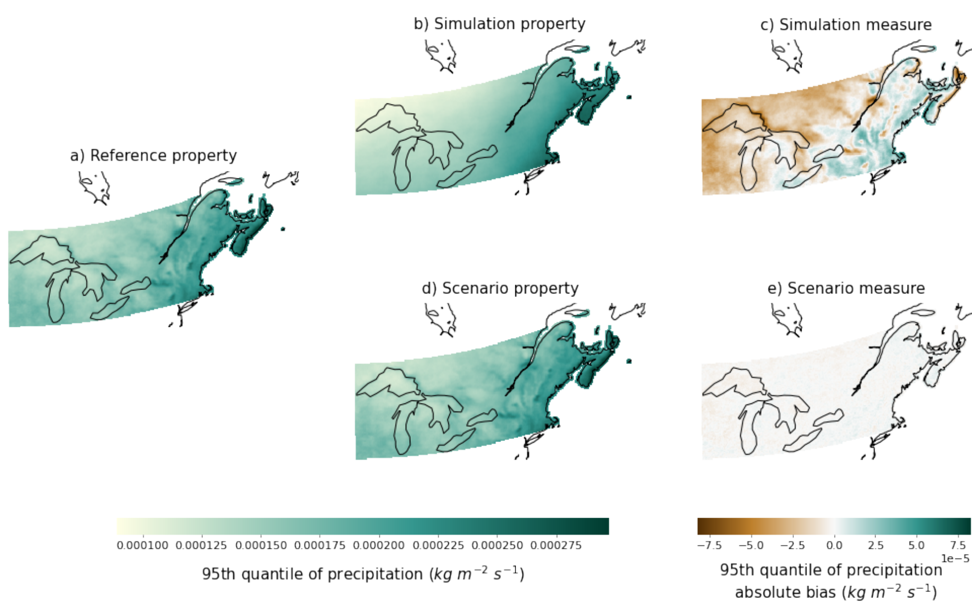

Figure 3: 95th quantile of the precipitation (property in a, b, d) and its biases (measure in c, e) during the 1989-2018 period using the RDRS v2.1 reference (a), the KACE-1-0-G SSP3-7.0 simulation (b, c) and the KACE-1-0-G SSP3-7.0 scenario (d, e).

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 adjusting all quantiles separately (Figure 4). As an example, we can see in Figure 3 that there is barely any bias left in the scenario measure. This matches the corresponding IMP of 95% 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 overestimation `dtr` could lead to an underestimation 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 5 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 cold bias and about half the 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.

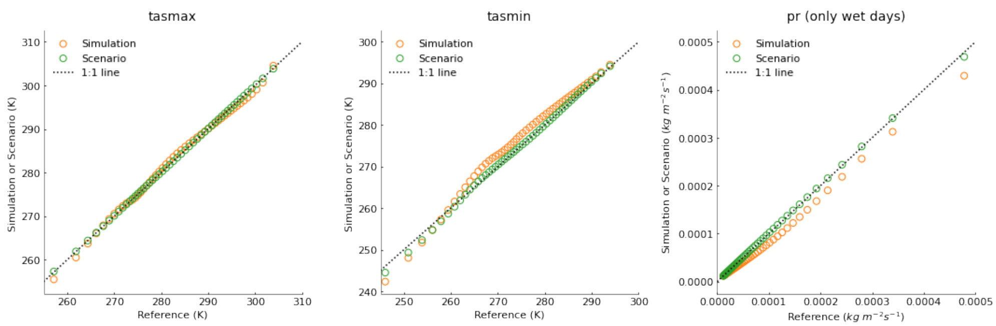

Figure 4: Q-Q plots for INM-CM5-0 SSP3.7-0 including all grid cells of the Magtogoek region for the 1989-2018 period. For precipitation, the plot is created with only wet days (more than 1 mm/d).

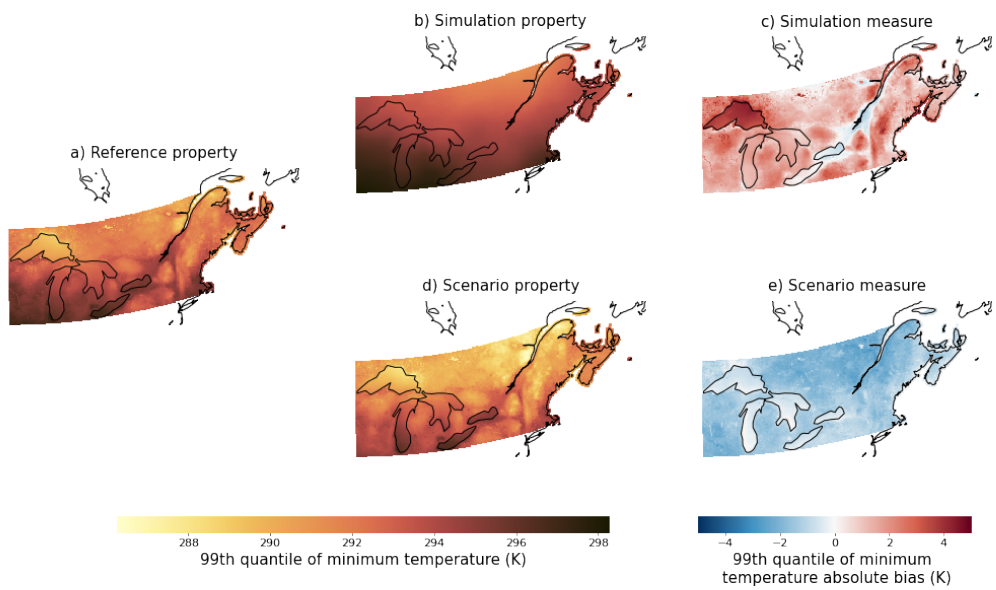

Figure 5: 99th quantile of the daily minimum temperature (property) and its bias (measure) during the 1989-2018 period using the RDRS reference, the MIROC6 SSP3-7.0 simulation and the MIROC6 SSP3-7.0 scenario.

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 96% over all models. For the relative amplitude of precipitation, it is 85 %. 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 80% for maximum length of warm spell, 78% for maximum length of dry spell and 86 % for wet-wet transition. The weakest property is dry-wet transition with 61%. Figure 6 shows that, for this property, there is very little change between the simulation and the 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 (see ESPO-G6-R2v100_adjustment.pdf). We consider this step, and accompanying low performance in the dry-wet transition property, necessary to avoid an important wet bias.

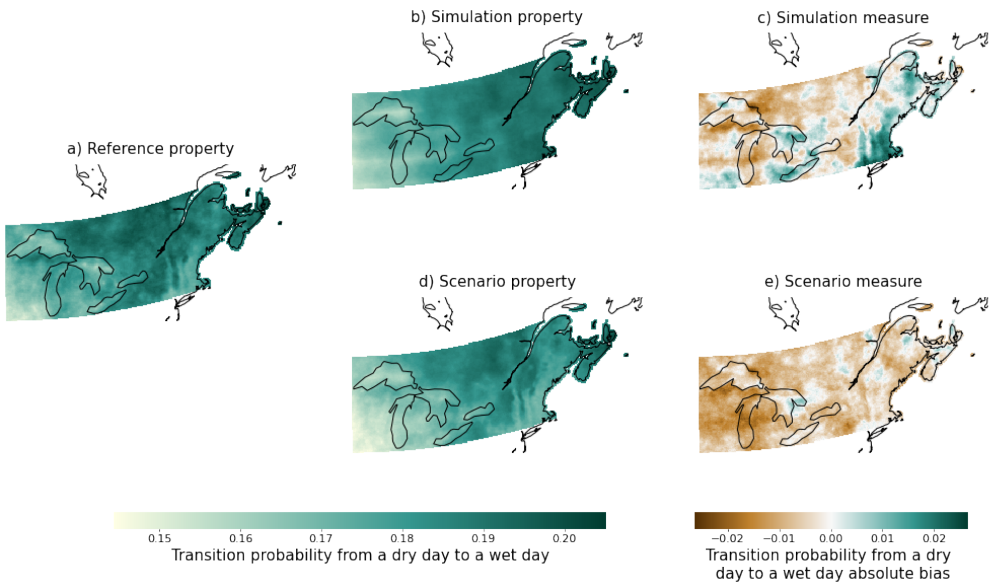

Figure 6: Transition probability of going from a dry day to a wet day (property) and its bias (measure) during the 1989-2018 period using the RDRS reference, the FGOALS-g3 SSP3-7.0 simulation and FGOALS-g3 SSP3-7.0 scenario.

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 94% for the correlation between `tasmax` and `tasmin`. Although, the IMP for the correlation between `tasmax` and `pr` is also high (90%) even though they were not corrected together. For both properties, the improvement is high, but for the correlation between `tasmax` and `tasmin` (Figure 7) the correlation is higher and the bias lower than for the correlation between `tasmax` and `pr` (Figure 8).

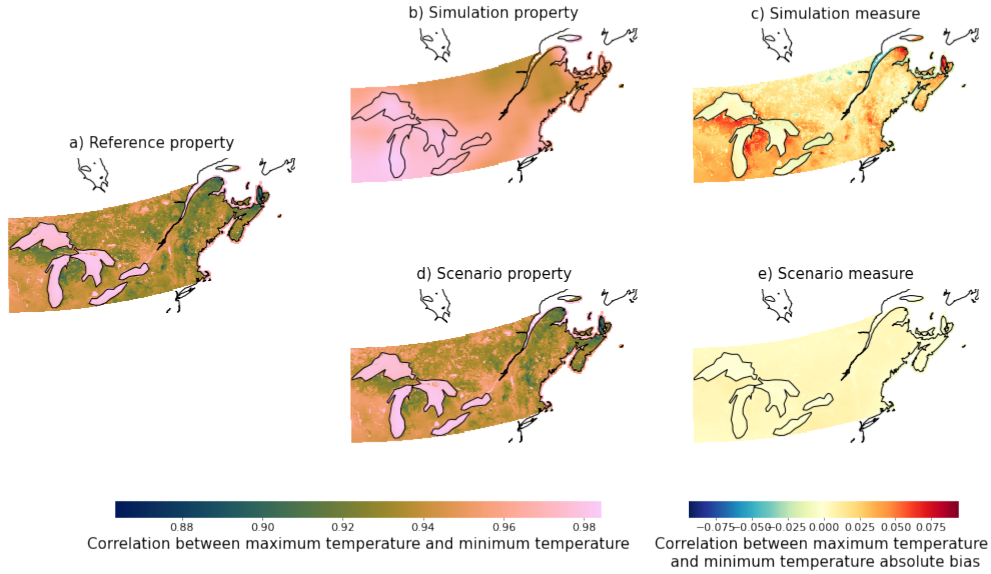

Figure 7: Correlation between maximum temperature and precipitation (property) and its bias (measure) during the 1989-2018 period using the RDRS reference, the ACCESS-CM2 SSP3-7.0 simulation and ACCESS-CM2 SSP3-7.0 scenario.

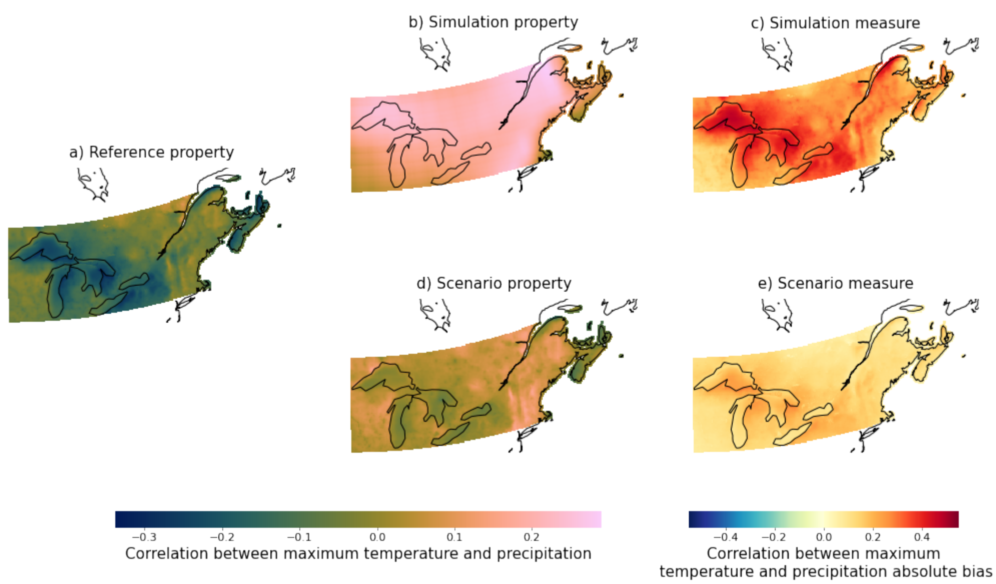

Figure 8: Correlation between maximum temperature and precipitation (property) and its bias (measure) during the 1989-2018 period using the RDRS reference, the ACCESS-CM2 SSP3-7.0 simulation and ACCESS-CM2 SSP3-7.0 scenario.

3 Ensemble Variability

Figure 9 shows the ensemble spread of the annual time series of change (compared to the 1989-2018 period) of three indicators. From this figure, we can say that the ensemble variability is conserved as the ensemble spread (90th percentile - 10th percentile) is similar for both the scenario and the simulation. Further, we can also see that the median of change of the scenario and simulation are similar, confirming that the climate change signal is conserved.

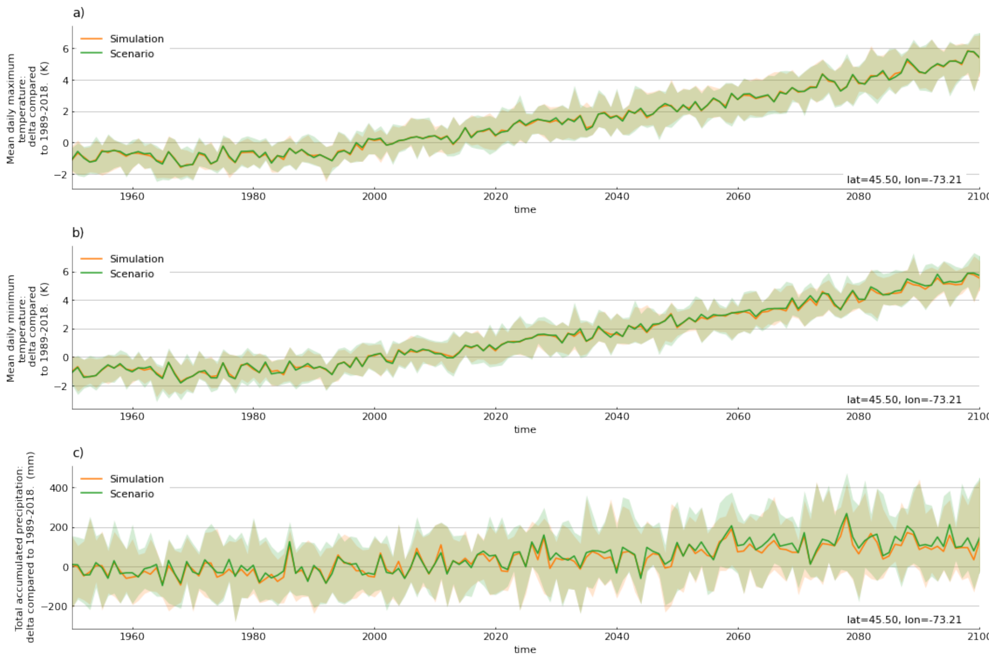

Figure 9: Simulation and scenario ensemble spread of the change in three annual indices: mean daily maximum temperature, mean daily minimum temperature and total precipitation. The change is computed with reference to the 1989-2018 period mean. This is computed for one grid cell near Montreal.

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