

# Performance of ESPO-R5 v1.0

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

November 21, 2022

The bias-adjustment performance assessment for ESPO-R5 was inspired by the "VALUE" project([MW15]), a framework for evaluating different bias-adjustment and downscaling methods, like the Detrended Quantile Mapping (DQM) method used here (see *adjustment.pdf*). This framework defines a large collection of "properties" (VALUE calls them "indices") to study different aspects of the data, which can be compared to each other with "measures". Through an implementation of a subset of these properties and measures into xclim, we looked at the changes brought about by the bias-adjustment for many different aspects of the data. Following the "VALUE" project, we define four (4) different aspects: marginal, temporal, multivariate, and spatial. Table 1 lists all properties studied for this analysis. In addition to those reference-period diagnostics, we looked at the climate change signal and the ensemble variability.

Table 1: Properties used in the performance assessment of ESPO-R5v1.0. All properties are compared with the simple "bias" measure, except the relative amplitude of the annual cycle (see footnote). All properties were computed on the full 30 years of daily data, except the means which were computed separately for each month.

Property	Short name	Variables	Aspect	Source
Mean	mean	tasmin, tasmax, pr	marginal	VALUE
First percentile	q01	tasmin, tasmax	marginal	VALUE
95th percentile	q95	pr	marginal	VALUE
99th percentile	q99	tasmin, tasmax, pr	marginal	VALUE
Dry spell frequency	dry_spell_freq	pr	marginal	VALUE
Amplitude of the annual cycle	aca	tasmin, tasmax	temporal	VALUE
Relative amplitude of the annual cycle	aca	pr	temporal	VALUE
Lag-1 autocorrelation	lag1	tasmax	temporal	VALUE
Inter-variable correlation (spearman)	corr	tasmin-tasmax, pr-tasmax	multivariate	VALUE
Spatial correlograms	correlogram	tasmin, tasmax, pr	spatial	[FV20]
Principal empirical orthogonal function	first_eof	tasmin, tasmax, pr	spatial	[V18]

The *properties* were computed for the calibration period (years 1981-2010) for the raw simulations, the bias-adjusted scenarios, and the reference (ERA5-Land). The results for the raw simulations were compared to those for the reference using a *measure* appropriate for the respective property. This way, for each member of the ensemble, one can visualize a given property and compare the measures to assess the improvement brought by the bias adjustment. Because of performance issues, the properties and measures were not computed on the full domain of ESPO-R, but rather on three smaller regions chosen to cover different climate types. These are shown in figure 1.

To complement this useful but quite exhaustive collection of maps, the proportion of improved grid points (we'll call this indicator IMP) was found for each property. The IMP score for a given member, property and region can be defined as

$$IMP = \frac{1}{N} \sum_{i,j} I_{i,j} \quad \text{where} \quad I_{i,j} = \begin{cases} 1 & \text{if } |M_{i,j}^{sim}| > |M_{i,j}^{scen}| \\ 0 & \text{if } |M_{i,j}^{sim}| < |M_{i,j}^{scen}| \end{cases} \quad (1)$$

and where  $M$  is the measure of the bias between the simulation (sim) or adjusted (scen) data and the reference (ref). For most properties  $P$ , it is a simple bias :  $M^{sim} = P^{sim} - P^{ref}$  (and similarly for  $M^{scen}$ ). A grid point  $(i, j)$  is said to have been improved if the bias between the scenario and the reference is closer to zero than the one between the raw simulation and the reference. The only exception is for the *relative amplitude of the annual cycle*, which uses a ratio :  $M^{sim} = \frac{P^{sim}}{P^{ref}}$ . In that case, the improved condition  $I_{i,j}$  becomes true if the measure on the

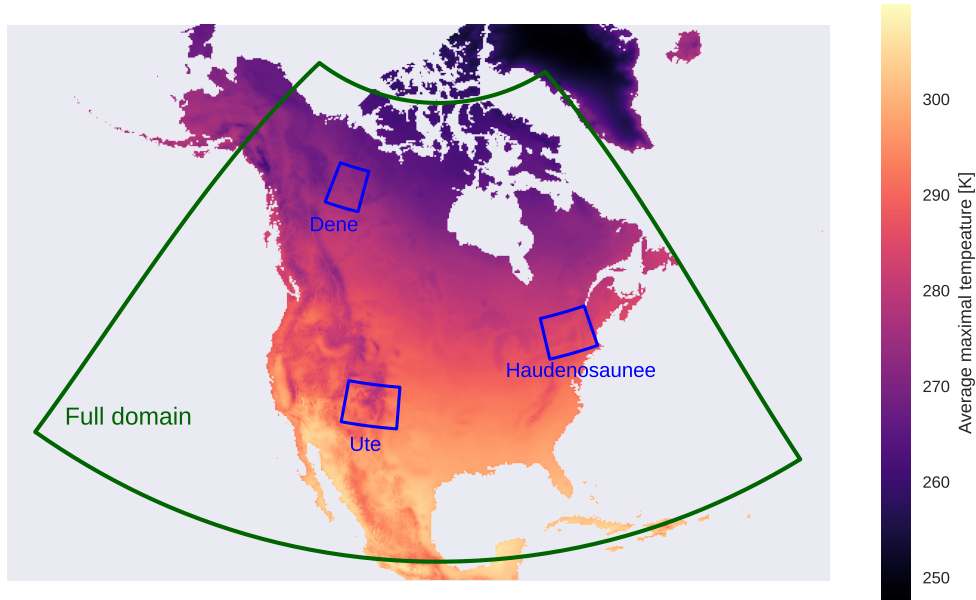


Figure 1: Map showing the domain of ESPO-R5v1.0 and the three diagnostics regions used in this study. All three regions have the same number of points, their size varies on this figure because of its rotated pole projection.

adjusted data is closer to 1:

$$|M_{i,j}^{sim} - 1| > |M_{i,j}^{scen} - 1|$$

By taking the mean across all members of the ensemble, this allowed us to have an idea of which properties were improved by the bias adjustment (those with IMP close to 100 %), which were not impacted (IMP around 50%), and which were worsened (IMP smaller than 50 %). Fortunately, we found almost no properties of the latter case in this study. Figure 2 shows a heat map of those IMP values for each member and properties.

As always when asserting the quality of a bias-adjustment, one must not forget that terms like "improvement" and "performance" refer to the change in comparison with the reference. Within the context of bias-adjustment, the reference is taken as the "Truth", as if it was an exact representation of the real climate. Thus, the validity of bias-adjusted datasets are limited by the validity of the reference dataset. A succinct analysis of ERA5-Land is detailed in the README at the root of this repository and provides the justification for using it as our reference.

This whole analysis not only serves as an assessment of the performance of ESPO-R5v1.0, but also as a benchmark for the next iterations of the ESPO datasets. Here, we aim to test our performance evaluation tools and tune the list of properties for evaluation purposes.

The complete properties and measures dataset is too large to be delivered through a git repository, but we offer a sample here. Values for most properties, and their measures, on the Haudenosaunee region are given in the `data/` folder. Other regions and missing properties can be provided to interested parties on demand.

## 1 Marginal aspects

Detrended Quantile Mapping will adjust most properties marked as "marginal" by design. Our algorithm divides the distribution of values for each day of the year into 50 bins and thus the percentile properties should look almost perfectly adjusted (see figure 3). Indeed, all members show over 90% of improved grid points for the percentiles properties of pr and tasmax. Oftentimes the non-improved grid points are those where the simulation was already comparing well with the reference. This indicates that there is noise (variability) in the scenario and shows a limit in using this IMP metric. This can also be seen when comparing results for members where the initial resolution was  $0.44^\circ$  with those for the  $0.22^\circ$  sources: the former systematically have stronger IMP scores than the latter; However, we can't really conclude anything other than the observation that bilinear regridding is an inferior downscaling technique than a regional climate model.

Scores are lower for tasmin. Recall that here tasmin is not directly adjusted, but is rather computed from the adjusted tasmax and dtr, the daily temperature range. This indirect adjustment seems to be visible when

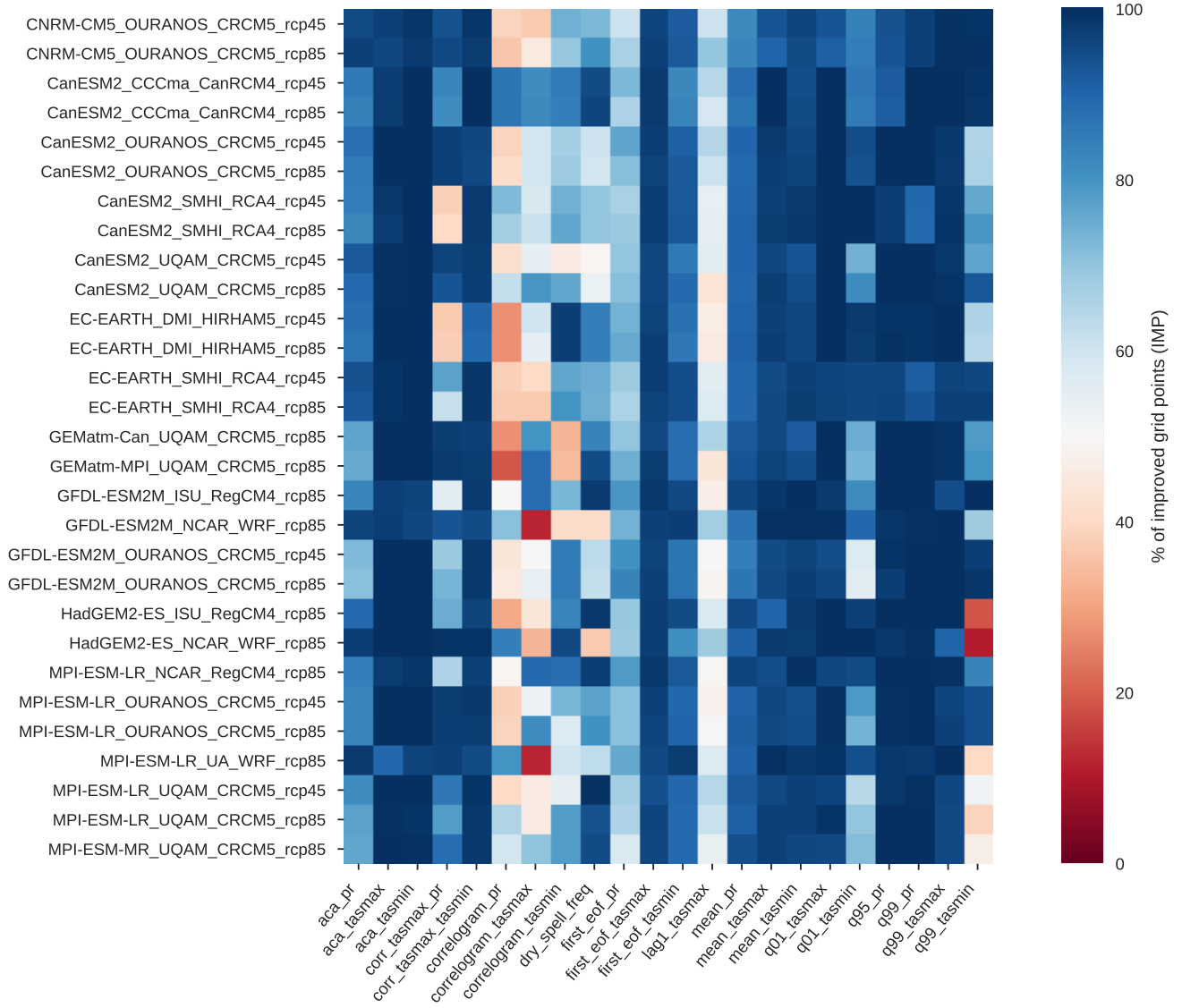


Figure 2: The proportion of improved gridpoints (IMP) for the Haudenosaunee region.

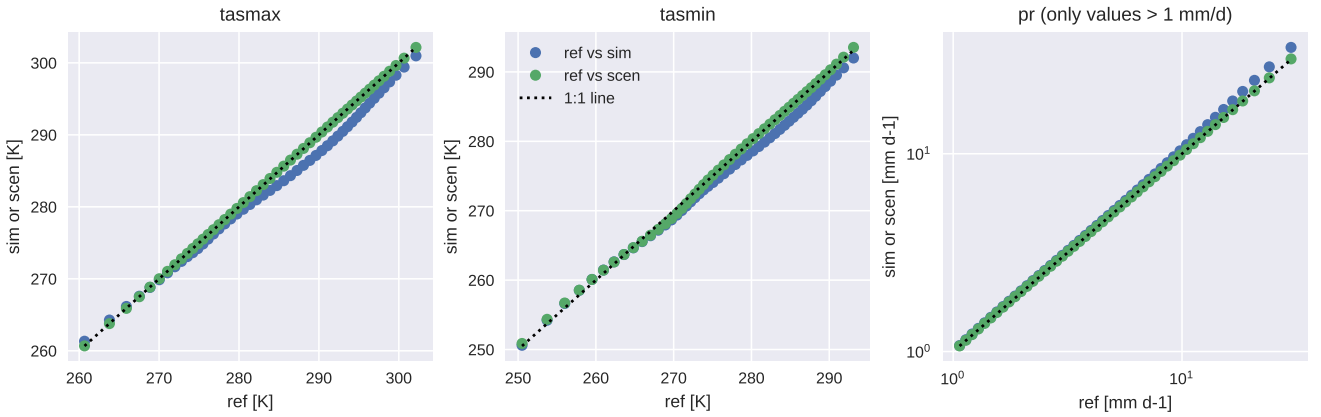


Figure 3: Q-Q plots for member MPI-M-MPI-ESM UQAM-CRCM5 rc45, including all points of the Haudenosaunee region. This member's source simulation was on the NAM-44 domain. Only wet days are included.

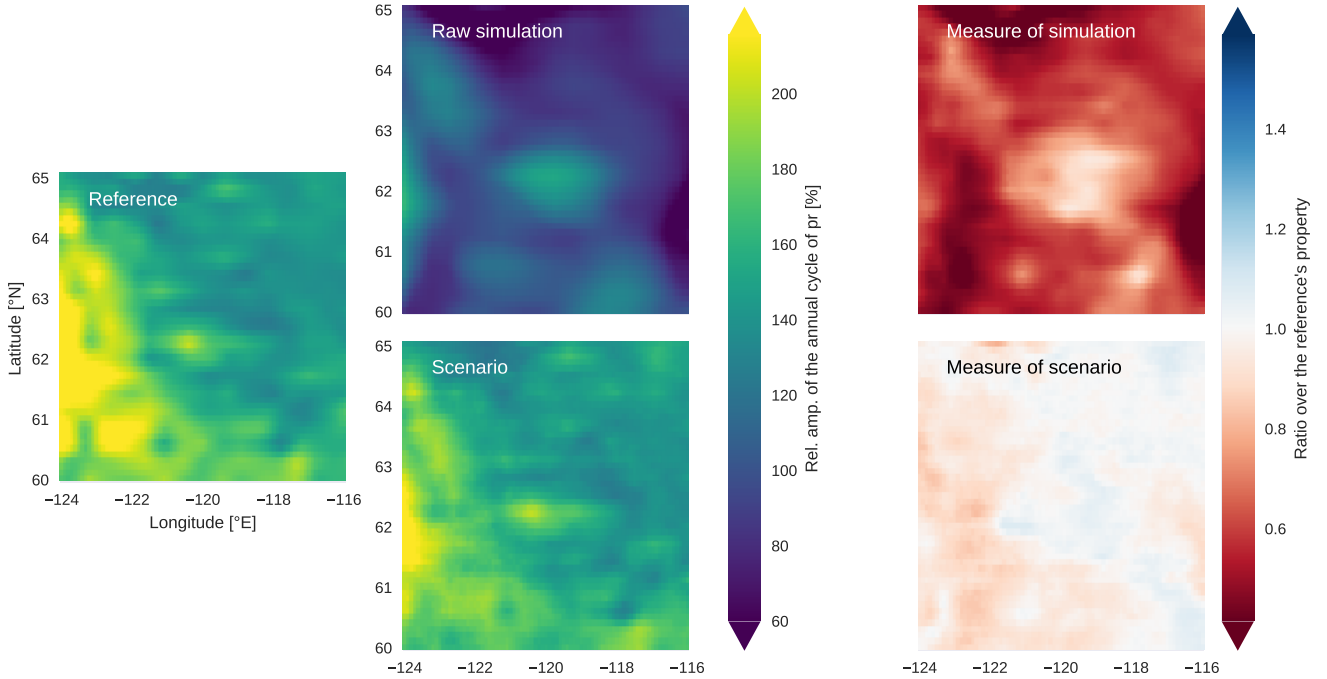


Figure 4: Relative amplitude of the annual cycle of precipitation ICHEC-EC-EARTH DMI-HIRHAM5 rcp45, 1981-2010, Dene region. This example of diagnostic maps showing a large bias between the simulation and the reference as well as a large improvements in the scenario. This case has a IMP value of 88%.

inspecting the properties and measures, as will be shown again further down. Conclusions are very similar for the "mean" properties. However, a discrepancy appears in the 99th quantile, but only for some models (see figure 1). For most members our IMP indicator seems to indicate an improvement from the adjustment, but some members show values close to 50 % and two members (those where the RCM is driven by a simulation of the HadGEM2-ES model) show an overall worsening of the performance for that property. Interestingly, these issues only show up in the Haudenosaunee region, in the other two regions, only two members show values under 50% (46% and 44%, EC-EARTH DMI-HIRHAM5, Dene region). This might be an indication of a misbehaving adjustment of dtr, but we haven't studied that problem yet.

The dry spell frequency has a similar behaviour where only members with large initial biases show strong IMP. This property is also somewhat adjusted by design during the pre-processing of precipitation data, at least for members with too many dry days, when compared to the reference. The process explained in the "Adjustment" document (section "Pre-processing of precipitation") is mostly meant to ensure valid quantile mappings, but it does have a small effect on the final scenarios. There are still some members where the frequency of dry days is nearly 6% higher in the scenarios then in the reference, especially over lakes. The bias is also quite consistently positive for all members and regions.

## 2 Temporal aspects

As the bias adjustment acts separately on each day of the year and, as we've seen, the scenario distribution matches well with the reference, we expected that the annual cycle amplitude would be well adjusted. And indeed, for both temperatures, the improvement is major for all members, most showing IMP values over 90 %. Scores are slightly weaker for precipitation, but it is important to note that in some regions precipitation doesn't have an annual cycle as strong nor as consistent as temperature. Still, IMP values are over 75 % for two-thirds of the members, and there is strong correlation between the improvement amount and the initial bias. Some members showed severely inadequate annual cycle amplitude patterns before the adjustment, which makes their high IMP scores almost trivial; So much so that it becomes less a measure of the performance of the bias adjustment than a measure of the bias between the simulation and the reference, see figure 4 for example.

Unsurprisingly, the lag-1 autocorrelation property doesn't seem to be significantly improved by the bias-adjustment.

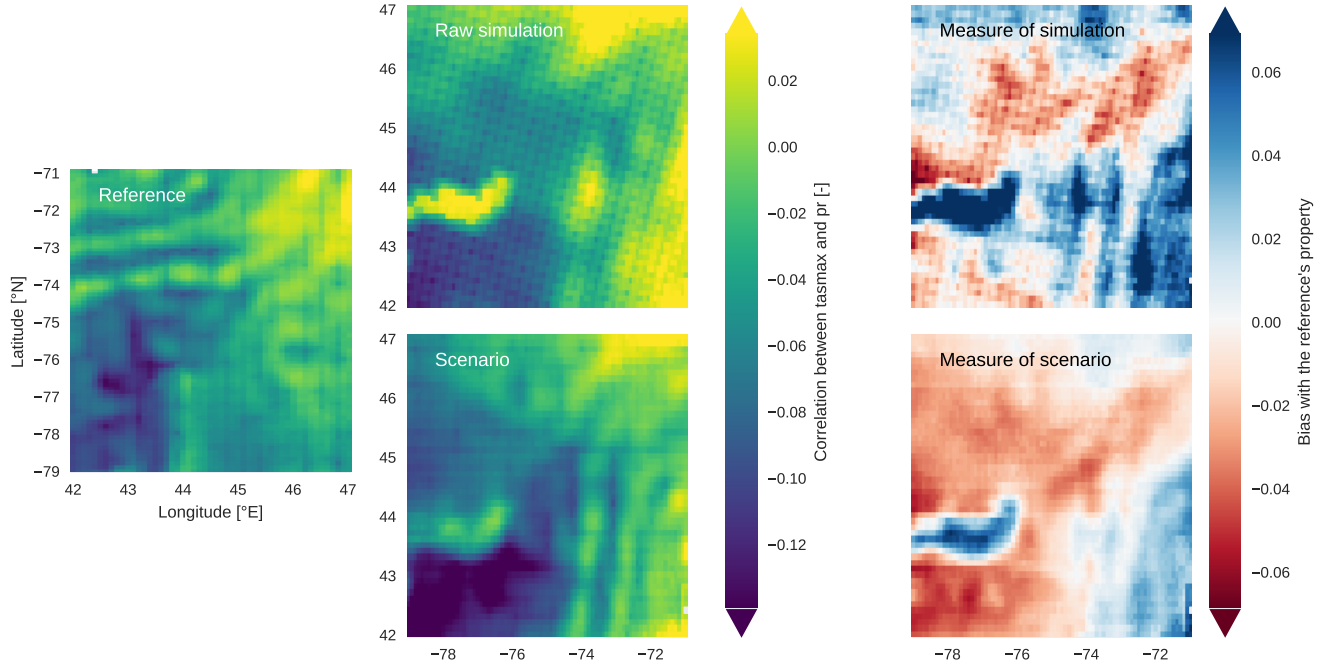


Figure 5: Interviabile correlation (Spearman rank-order) between tasmax and pr for GFDL-ESM2M ISU-RegCM4 rcp85, 1981-2010 period, Haudenosaunee region. This case has a IMP value of 55%

Generally, the lag-1 autocorrelation structure from the simulations is almost identical in the scenarios. Indeed, the algorithm does not adjust this aspect explicitly, but it may be interesting for future versions of the ESPO datasets to address this issue.

### 3 Multivariate aspects

Our algorithm is inherently univariate, but the special treatment of tasmin does seem to have a strong effect on the correlation between tasmax and this variable. Most simulations of the ESPO-R5 ensemble tend to underestimate the correlation between the two variables and this bias is drastically reduced in the scenarios. If our analysis is correct, the reconstruction of tasmin from dtr and tasmax does improve that aspect of the data, while reducing the improvement of other properties, as discussed above (section 1).

When looking at the biases in the tasmax-pr correlation, we do see large IMP values in some regions and members. In most of these cases, the initial bias between the simulations and the reference was quite large (much larger than in the tasmax-tasmin case), which leaves much room for improvement. This is visible as a strong correlation between the improvement of the measure in the scenario and its value in the simulation. In other cases, the improvement is less clear with IMP values around 50%. Figure 5 shows such a case, but it reveals a pattern created by the bilinear interpolation used when regridding the simulations to the reference grid. The pattern gets smoothed out by the bias adjustment, but it is replaced by another one inherited from the reference itself, at least in the case of this particular property.

### 4 Spatial aspects

Despite the lack of real benchmark values, the spatial correlograms seem to visually show an already good fit between the simulations and the reference, though we see no improvement in the scenarios. In general, the correlogram are very similar between the simulations and the scenarios and we see no significant change in the spatial correlation length (distance where the correlation drops below a certain threshold, taken here as 0.35 for all variables). Figure 6 shows an example where the bias is quite large (compared to those of other members of the ensemble). Although we explicitly take care of adjusting the dry days frequency, we do not see the spatial correlation improvement described by [FV20].

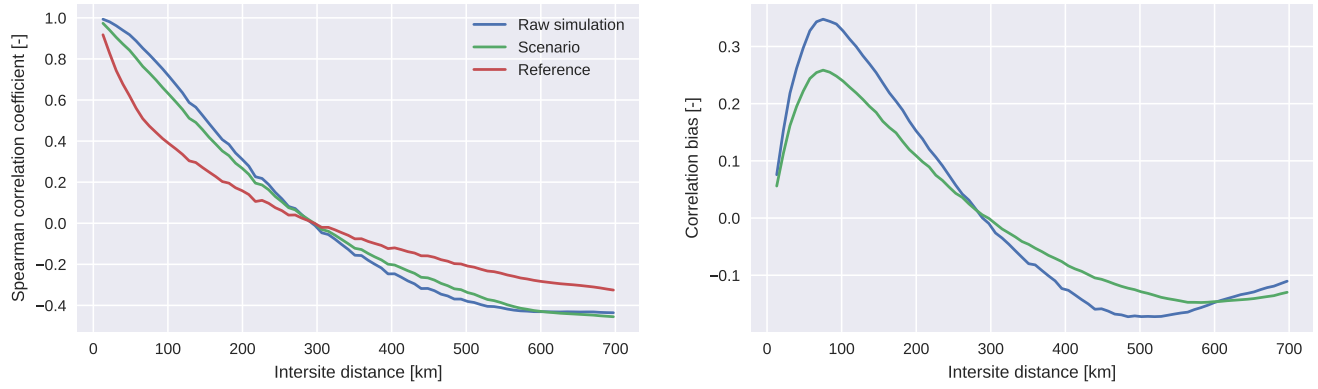


Figure 6: Correlogram for tasmin, member ICHEC-EC-EARTH DMI-HIRHAM5 rcp 4.5, Ute region, 1981-2010.

## 5 Climate change signal

Nothing much is to be said about the effect of the bias adjustment on the climate change signal since it is preserved by design. The use of a LOESS detrending with a 30-year window on each day of year preserves the mean change in the variables as simulated by the regional models. If we look at 30-year means of the variable, as is often done in climate change studies, there is no difference between the raw simulations and their adjusted scenarios (not shown).

## 6 Ensemble variability

Figure 7 shows the annual timeseries of two indices, the average daily maximal temperature and the total precipitation. The indices are given as change from the 1981-2010 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 inter-annual variability of each member was preserved.

## References

- [FV20] François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth System Dynamics*, 11(2), 537–562. <https://doi.org/10.5194/esd-11-537-2020>
- [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>
- [V18] Vrac, M. (2018). Multivariate bias adjustment of high-dimensional climate simulations: The Rank Resampling for Distributions and Dependences (R2D2) bias correction. *Hydrology and Earth System Sciences*. <https://doi.org/10.5194/HESS-22-3175-2018>

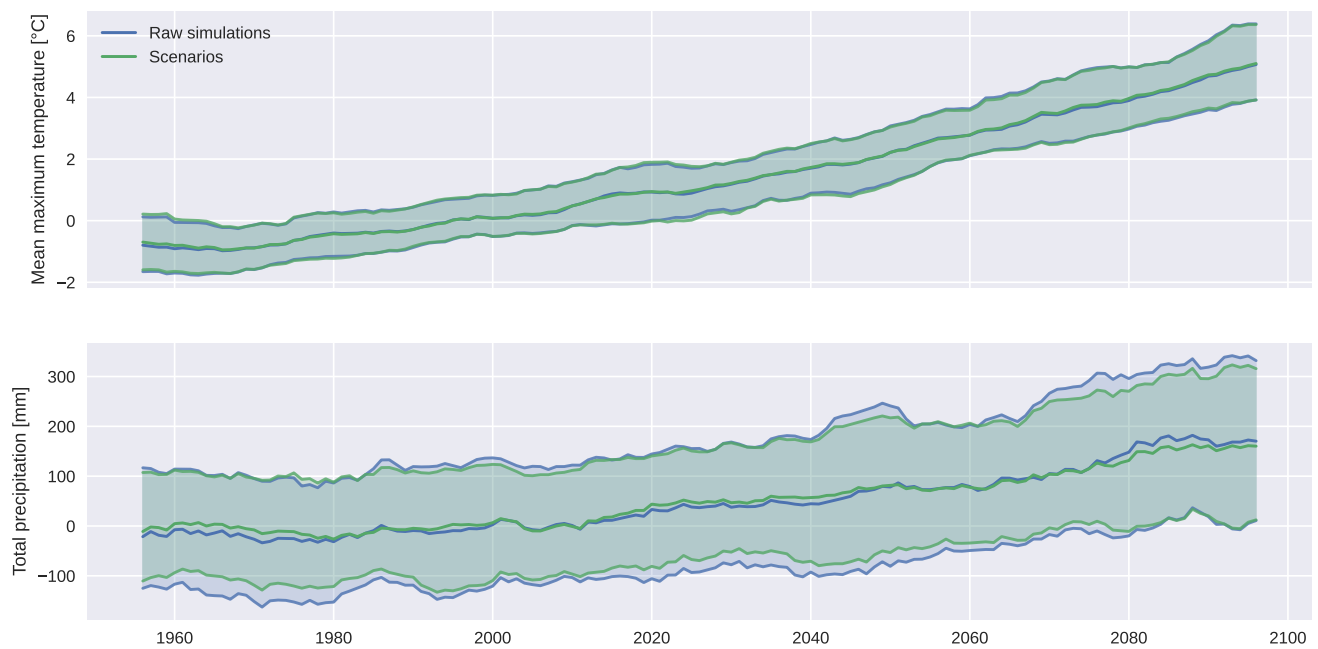


Figure 7: Ensemble spread of the change in two annual indices. Change computed with reference to the 1981-2010 period mean. The lines are smoothed with a 10-year moving window.