ESTCP Perfect Model Downscaling Evaluation Web Atlas
User's Guide

How to Use the Web Atlas

The Web atlas is located at https:www.esrl.noaa.gov/psd/models/estcp/

The atlas of evaluation results contains numerous maps that show the evaluation of downscaling methods. Step-by-step instructions are shown here. Further details about abbreviations are available in other sections of the User Guide.

Step 1: Choose Phase 1 or Phase 2 atlas from the welcome page according to the type of evaluation you are interested in.

Step 2: Choose the variable and evaluation metric you are interested in. For example, you can choose "Tasmax 95% "(95th percentile of daily maximum temperature). This choice represents very warm days. 99th percentile would be more extreme, roughly corresponding to the hottest day of the year (for the season chosen). Choices that note "KS" are a little different in that they show regions that are statistically significant difference in the entire probability distribution, accordging to the Kolmogorov-Smirnov (KS) test, between the downscaled data and the "Truth".

Step 3: Choose the season for analysis. You can choose among the four canonical climatological seasons (DJF (December – February), MAM (March – May) JJA (June – August) and SON (September – November). Only days within the chosen season will be evaluated. If you are interested in the warmest day of the year, a choice of Tasmax 99 %, and JJA would pe appropriate for most regions. For the coldest day of the year, Tasmin 1% and DJF would likely be appropriate.

Step 4: Choose the combination of downscaling method and eras (time periods) that you are interested in. For Phase 1 the only option is to choose "all eras" so that all four evaluation periods will be shown on the same graphic. For Phase 2, a greater set of options is allowed, including comparisons among downscaling methods for a single era (denotes as "KQ", short for "KDDM QDM") as well as showing all eras at once for a single downscaling method.

Step 5: Choose the GCM/RCM model dataset that is the basis for the analysis. For Phase 1, only a single GCM/RCM dataset is chosen (out of six available) as the evaluation only uses a single model. For Phase 2, only two GCM/RCM models were considered. GFDL-ESM2M/RegCM4 and GFDL-ESM2M/WRF. Evaluation was done with one model as "Truth" and the other as the "Eval" model, and vice versa.

Certain combinations of choices have no data associated with them. The main deficiency you are likely to run into is that precipitation data for Phase 2 were not available for the KDDM downscaling method. These will show as a "Data Not Available".

How to read a sample graphic (Phase 1)

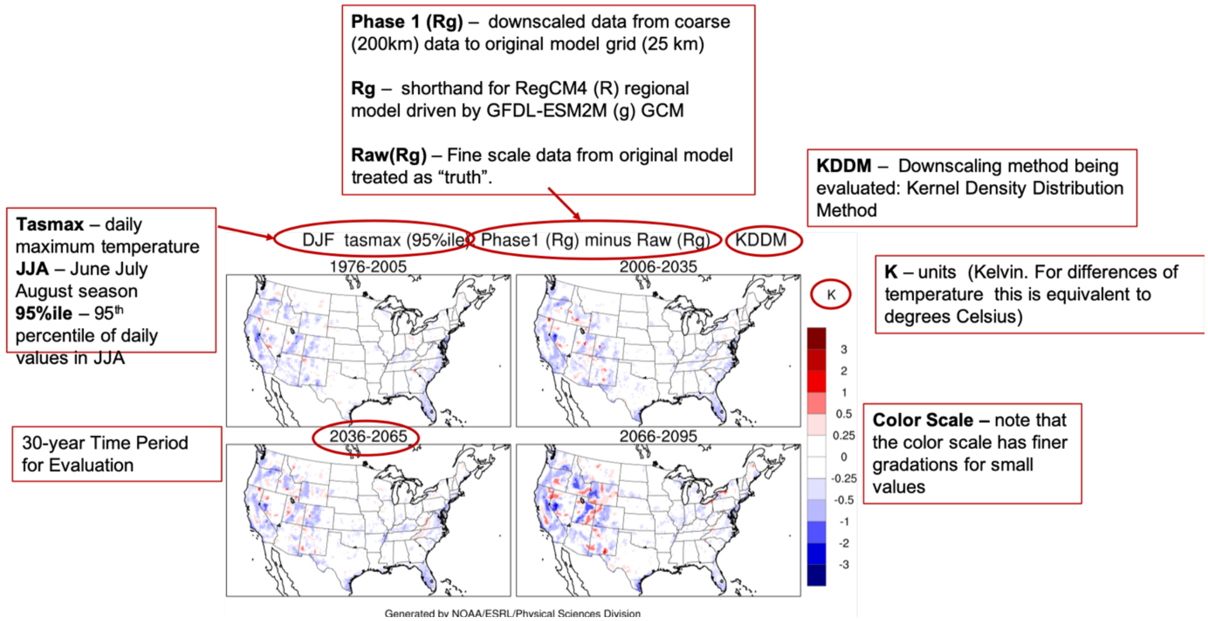

Figure 1. Features of a typical Phase 1 evaluation graphic.

How to read a sample graphic (Phase 2)

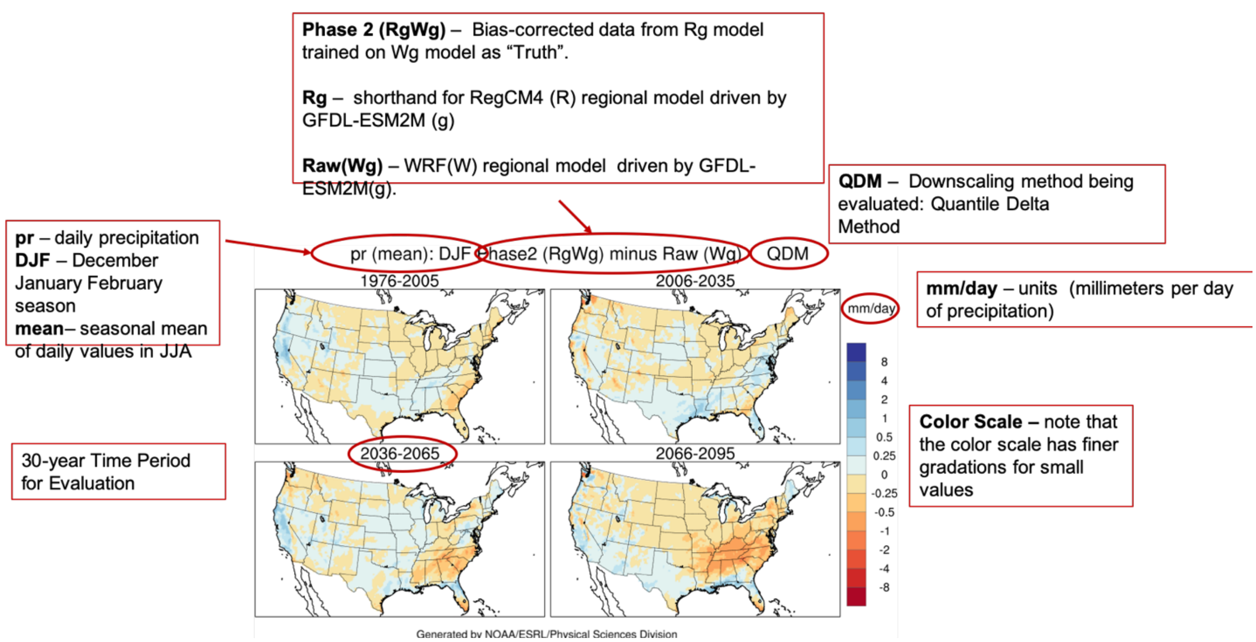

Figure 2. Features of a typical Phase 2 evaluation graphic.

How to interpret a "KS" graphic

How can we tell whether, and more importantly, where a downscaling method fails? The answer is to look at the entire probability distribution of the downscaled data compared to the "Truth". The Kolmogorov-Smirnov statistical test determines whether one can reject the null hypothesis that two samples of data come from the same underlying probability distribution. This powerful test detects not only differences in the mean values but also in the shape of the probability distribution. This is a particularly sensitive test for the Phase 1 analysis because we expect that the data should be exactly recovered if the method were "perfect". For the graphic below we note that areas along the Gulf Coast show discrepancies, even for the historical period (the line along the Canadian border is an artefact). These areas of discrepancy become

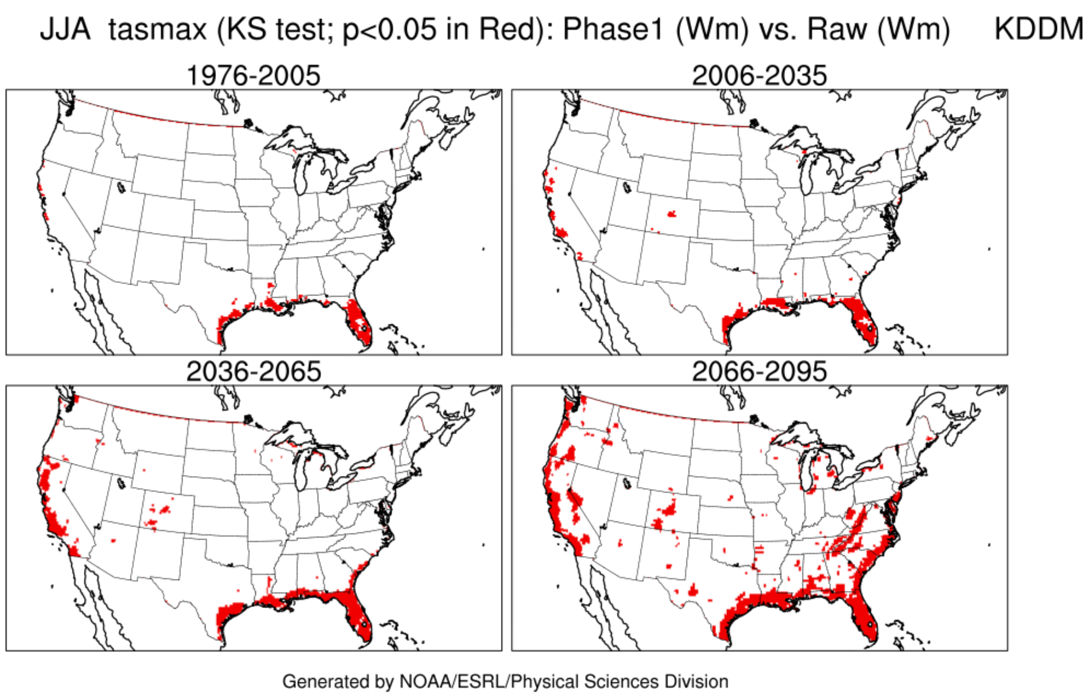

Figure 3. Example of a figure showing the regions (red) where the Phase 1 downscaled data (200km -> 22 km) differs from the 22 km resolution "truth" according to the Kolmogorov-Smirnov (KS) test with p-value < 0.05.

larger the further into the 21st century one looks, with discrepancies showing up in large areas along the West Coast as well. This analysis can guide further regional analysis. From a practical standpoint, areas where there is statistical difference (white areas on the map) the error in the downscaling method is small enough to be ignored compared to the error one would get from a finite (30-year) sample of data.

Perfect Model Methodology: Phase 1 and Phase 2

Climate model output is typically produced at too coarse a spatial scale to be usable for climate adaptation efforts at a particular site. This is true not only of the global climate models (GCMs, or general circulation models), but is also true of Regional Climate Models for many applications. In addition, climate model output has systematic biases relative to historical observations that need to be accounted for. Empirical statistical downscaling (ESD) is an umbrella term that includes many methods used to create fine-scale climate projections from the coarse-scale output of global climate models while reducing the magnitude of systematic biases. What ESD methods have in common is that they are based on statistical relationships that are developed between the model output and observational data. These relationships are then applied to future model output to generate downscaled/bias-corrected projections.

How well do the empirical/statistical methods used to correct and downscale the climate model hold up in the future? While the accuracy of a method may be tested during the historical period by using, for example, cross-validation, there are no "future observations" to test whether these relationships are valid in the future. One way to test this is to treat a fine-scale climate model as the "truth" for both the historical and future time periods and see whether the statistical models that are trained on the "historical" model output perform well when applied to the future model output. This is referred to as the "perfect model" framework. It is also referred to in the European literature as "pseudoreality", capturing the idea that we are treating a model as a proxy for reality.

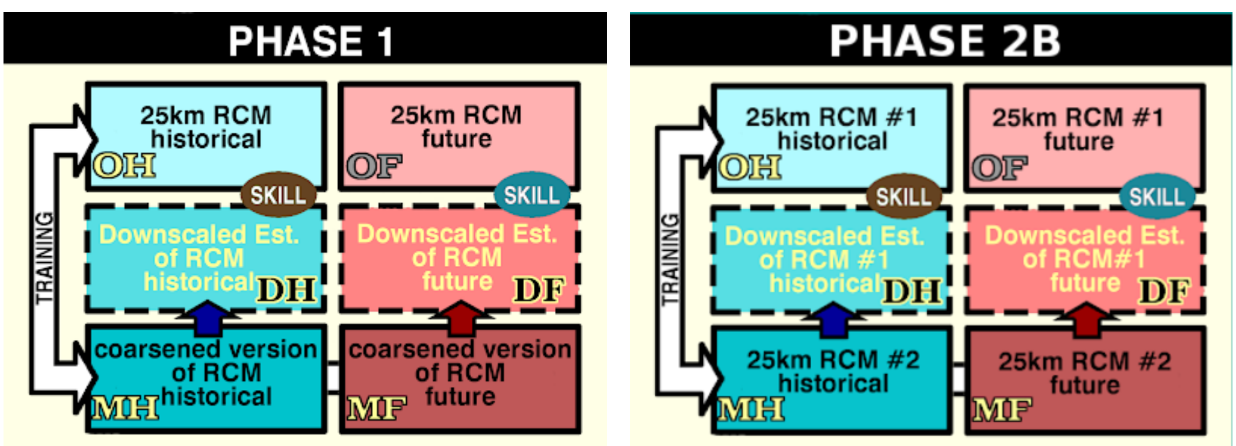

Figure 4. Schematic of perfect model approach. Phase 1 focuses on spatial refinement and uses 200-km coarsened versions of the 22 km NA-CORDEX model output as the predictors, and the native 22 km resolution as the "truth". Phase 2 focuses on bias correction at the 22 km scale by training one model on a second model as "truth", approximating the biases seen in when SD is applied to observational data. Note that Phase 2A was not used in this study.

The “perfect model” experimental protocols employed here split the evaluation into two parts: Phase 1 looks at the ability to recover fine scale details when only the coarse scale averages are known. Phase 2 looks at the ability of bias correction between two models to faithfully represent the future. The method is shown schematically in Figure 1. For Phase 1, take 22 km model output, coarsen the output to a 200 km grid by simple averaging, and then try to recover the future 22 km model by statistical methods. For Phase 2, take one 22 km model as the “Truth”, and a second 22 km model as the climate projection to evaluate (“Eval”). Try to recover the future “truth” from the future “evaluation” model by statistical methods. Phase 2 is more representative of real-world applications of climate projections where biases between models and observations can be significant. Prior work has using the Phase 1 methodology, though applied to different models and different downscaling methods, can be found in Dixon et al. (2016) and Lanzante et al. (2018).

Downscaling Methods

Two quantile-based asynchronous downscaling methods for daily minimum and maximum temperature and daily precipitation were provided for evaluation, the Kernel Density Distribution Method (KDDM, McGinnis et al. 2015, 2016) and the Quantile Delta Method (QDM, Cannon et al, 2015). These methods correct the bias not only in the mean but in all quantiles of daily values. For example, hotter temperatures may see more bias adjustment than cooler ones, or heavy precipitation amounts may receive larger adjustments than light precipitation amounts. These methods are univariate – meaning that only temperature data is used to downscale temperature data, and only precipitation data is used to downscale precipitation data. Nonetheless, the quantile-dependent adjustments may reflect underlying physical processes such as the relationship of soil moisture or snow cover to surface air temperature. Because of unforeseen technical issues with the KDDM method and precipitation, this method was not available for evaluation in time to complete the analysis of Phase 2 precipitation.

Metrics, Variables, Seasons

The primary evaluation is to quantify how the biases in the downscaled data develop over time throughout the 21st century for both means and extremes. To do this we have defined a baseline (Historical) 30-year period (1976-2005) during which the training occurred, and three future 30-year periods, early (2006-3035), mid (2036-2065), and later 21st century (2066-2095). These time periods were chosen to provide non-overlapping 30-year periods centered on roughly 1990, 2020, 2050, and 2080.

We evaluate biases for the mean seasonal climate as well as extremes represented by the 1st, 5th, 95th , and 99th percentile values. The choice of the four “meteorological” seasons: DJF, MAM, JJA, and SON.

Table 1. Evaluation choices.

Metrics	Variables	Seasons	Comparisons
Mean Error 1^{st}, 5^{th}, 95^{th} and 99^{th} percentiles Kolmogorov-Smirnov test (KS test) for differences in distribution	Daily Maximum Temperature (tasmax) Daily Minimum Temperature (tasmin) Daily Precipitation (pr)	Winter (DJF) Spring (MAM) Summer(JJA) Fall (SON)	One model/method: all 4 time periods Comparison of KDDM and QDM (Phase 2 tasmax and tasmin only)

Models

The model data were taken from the NA-CORDEX archive and regridded to a common 22km grid. The NA-CORDEX project coordinated regional climate modeling from a number of institutions that were used to “dynamically downscale” from several GCMs used in the CMIP5 experiment. For this project two RCMs were chosen: RegCM4 and WRF. For phase 1, three separate driving GCMs were used as the lateral boundary conditions, yielding 6 GCM/RCM combinations. The model output was coarsened to a 200km grid by averaging, and this was used as the predictor in the downscaling method. For Phase 2 only a single GCM was chosen, the GFDL-ESM2M model, yielding only two GCM/RCM combinations. All analysis for phase 2 was done in the 22km model grids. For more information on NA-CORDEX and on the individual GCMs and RCMs, please visit https://www.na-cordex.org

Tables 2 and 3 below summarize the primary NA-CORDEX models, downscaling methods and variables that were downscaled/bias corrected using the phase 1 and phase 2 protocols. These data were generated by NCAR and GFDL and the perfect model downscaling was generated by GFDL and made available to ESRL/PSD for evaluation. Additional datasets (such as cross-validated historical datasets) were also generated and provided but are not shown in the web atlas.

Table 2. Variables available for Phase 1 evaluation in this project from different GCM/RCM drivers and different downscaling methods

Phase 1 GCM/RCM	Downscaling Method	
Coarsened and native grid	QDM	KDDM
HadGEM2/ WRF	tasmax,tasmin,pr	tasmax,tasmin,pr
HadGEM2/ RegCM4	tasmax,tasmin,pr	tasmax,tasmin,pr
GFDL-ESM2M/WRF	tasmax,tasmin,pr	tasmax,tasmin,pr
GFDL-ESM2M/RegCM4	tasmax,tasmin,pr	tasmax,tasmin,pr
MPI-ESM-LR/WRF	tasmax,tasmin,pr	tasmax,tasmin,pr
MPI-ESM-LR/RegCM4	tasmax,tasmin,pr	tasmax,tasmin,pr

Table 3. Variables available for Phase 2 evaluation in this project from different GCM/RCM drivers and different downscaling methods

Phase 2 GCM/RCM		Downscaling method	
"predictor model"	"truth model"	QDM	KDDM
GFDL-ESM2M/WRF	GFDL-ESM2M/RegCM4	tasmax,tasmin,pr	tasmax,tasmin
GFDL-ESM2M/RegCM4	GFDL-ESM2M/WRF	tasmax,tasmin,pr	tasmax,tasmin

Collaboration

The overarching goal of this part of the project is to explore the use of the "perfect model" approach to evaluating empirical-statistical downscaling (ESD) using NA-CORDEX model results as both predictor and predictands. The primary roles of the three institutions involved are as follows. NCAR**:** provide NA-CORDEX model output to GFDL on the 22 km interpolated grid, coarsened data for Phase I (see below), as well as providing code and guidance for the KDDM downscaling/bias correction method. Climate variables include daily minimum and maximum temperature, and precipitation for the continental United States. GFDL**:** Perform the Perfect Model experiments employing different ESD methods and archive the post processed data for analysis. ESRL/PSD**:** Facilitate and supervise the overall workflow and perform the analysis of the SD data for the examination and quantification of errors. Scientists from all institutions participated in interpretation of the results.

References

Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, **28**, 6938–6959, doi:**10.1175/JCLI-D-14-00754.1**.

Dixon, K. W., J. R. Lanzante, M. J. Nath, K. Hayhoe, A. Stoner, A. Radhakrishnan, V. Balaji, and C. F. Gaitán, 2016: Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Climatic Change*, **135**, 395–408, doi:10.1007/s10584-016-1598-0.

Lanzante, J. R., K. W. Dixon, M. J. Nath, C. E. Whitlock, and D. Adams-Smith, 2018: Some Pitfalls in Statistical Downscaling of Future Climate. *Bulletin of the American Meteorological Society*, 99, 791–803, doi:**10.1175/BAMS-D-17-0046.1.**

McGinnis, S., and L. Mearns, 2016: Bias-correction of extreme temperatures and precipitation in NA-CORDEX regional climate model output. *AGU Fall Meeting 2016*, American Geophysical Union, San Francisco, CA, US.[**Poster Link**].

McGinnis, S., D. Nychka, and L. O. Mearns, 2015: A New Distribution Mapping Technique for Climate Model Bias Correction. *Machine Learning and Data Mining Approaches to Climate Science*, V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, Eds., Springer International Publishing, Cham, 91–99, doi: **10.1007/978-3-319-17220-0_9.**

Acknowledgements

This work funded by the Department of Defense Environmental Security Technology Certification Program (ESTCP) and performed by NOAA Earth System Research Lab Physical Sciences Division (PSD) in collaboration with the National Center for Atmospheric Research and the NOAA Geophysical Fluid Dynamics Laboratory. Downscaled data following the Phase 1 and Phase 2 protocols were generated at NOAA/GFDL and are available upon request from keith.dixon@gfdl.noaa.gov. Thanks to the team who have been involved in the ESTCP-SD project (alphabetically): Mike Alexander, Joe Barsugli, Candida Dewes, Keith Dixon, Seth McGinnis, Linda Mearns, Mary-Jo Nath, Imtiaz Rangwala, Lesley Smith.