PA43B-0321

# Examining the Performance of Statistical Downscaling Methods: Toward Matching Applications to Data Products

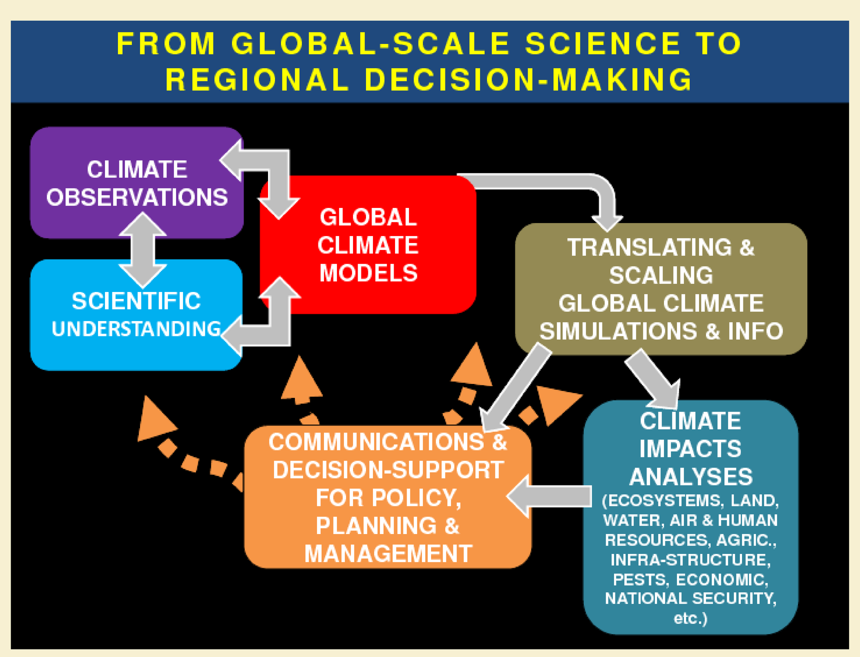

Keith W. Dixon[1], John R. Lanzante[1], & Dennis Adams-Smith[2]

(1) NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ (2) University Corporation for Atmospheric Research


## BACKGROUND:

Using future climate model projections in climate impacts work can pose several challenges. Global climate model (GCM) outputs often are deemed to be unsuitable for direct use in a climate impacts application. Statistical downscaling (SD) of climate projections – a type of post-processing that uses observations to inform the refinement of GCM projections – is commonly used in an attempt to account for GCM biases and to provide the additional spatial detail desired for some applications.

"***What downscaled climate projection is the best one to use***" is a frequently asked question, but one that often is not easy to answer with confidence. Matching statistically downscaled future climate projections to an application depends on a combination of stakeholder needs and the performance characteristics (*i.e.*, strength & weaknesses) of different SD methods. Though generally successful at producing "value-added" products, in some cases SD processing can lead to erroneous results – a factor that is not always appreciated nor extensively studied.

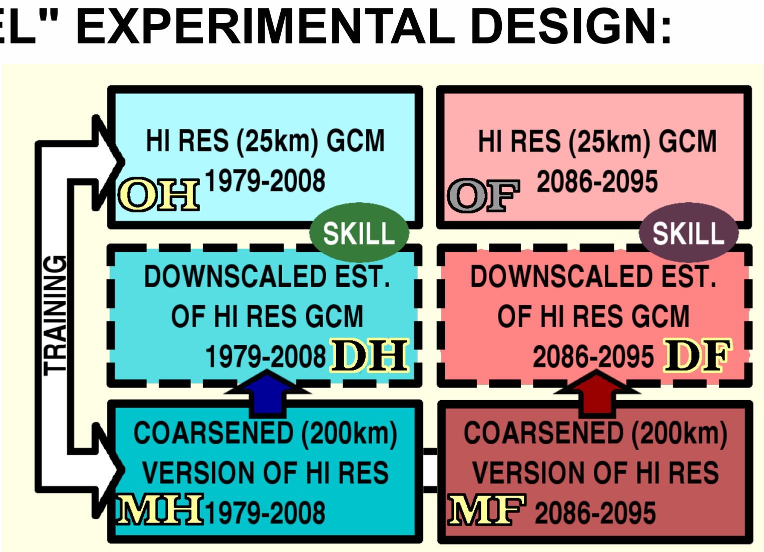



Statistical Downscaling (part of the "Translating & Scaling" step) is generally viewed as creating a "value added" data product ... a refined version of a large-scale climate model's output that is more suitable for use in climate impacts studies than the raw climate model output.

Many statistical downscaling methods exist, with all aiming to correct model biases and to add realistic spatial detail via processing that is informed by observations.

## A "PERFECT MODEL" EXPERIMENTAL DESIGN:

All SD methods assume that transfer functions computed during a training step are fully applicable to GCM simulations of the future, though the climate is changing — this is what we refer to as the "***stationarity assumption***." Lacking observations of the future, the validity of this assumption is difficult to assess.




In typical SD experiments, one has three types of input data. For a historical period, observations (**OH**) and model output (**MH**) exist. One also has model future (**MF**) data. In an SD training step, statistical methods use 2 or all 3 of the input types to develop transfer functions that quantify relationships between the input data types. The functions can be used to create downscaled results for the historical (*blue arrow,* **DH**) and future periods (*red arrow,* **DF**). Measures of a SD method's skill in the historical period can be computed directly as **DH - OH** for both PM and typical SD applications. In the PM case only, use of high resolution GCM output as a proxy for future observations (**OF**) allows computation of an SD method's skill for the future projections as **DF - OF**. Comparing SD method skill for the historical & future periods provides information on how well the stationarity assumption holds.

Our Perfect Model (PM) experimental design uses high-resolution (25km) dynamical climate model output as a proxy for both past & future observations. Via a process that includes interpolating the 25km data to a 200km grid, we create data sets that serve as proxies for coarse resolution GCM output. A key difference between this PM experimental design and typical SD applications is the availability of "future observations" (OF) in the PM design, which allows us to evalute an SD method's future skill, thereby quantifying aspects of the stationarity assumption.

## TWO CIRCUMSTANCES WITH VERY DIFFERENT ANSWERS TO THE QUESTION OF WHICH SD METHOD PERFORMS BETTER

**Variable of Interest = Daily Maximum Temperature (tasmax); Future Scenario = Late 21st Century High Greenhouse Gas Emissions; Comparing 3 Univariate SD Methods**

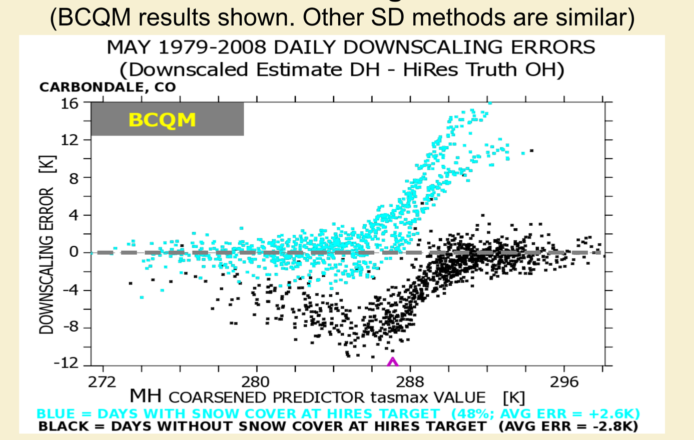
### CASE 1: ROCKY MOUNTAIN SPRING SNOWMELT SEASON

***Month:*** May. ***Location:*** near Carbondale, CO. A 25km grid cell that lies at an intermediate elevation relative to nearby 25km grid cells.
***The Statistical Downscaling Challenge:*** Daily May tasmax OH values are sensitive to the presence of snow cover at the Carbondale grid cell (for a given MH value, OH is colder when snow exists). Also, Carbondale has snow cover on ~half of historical May days, but warming leads to no snow-covered May days in the OF sample.

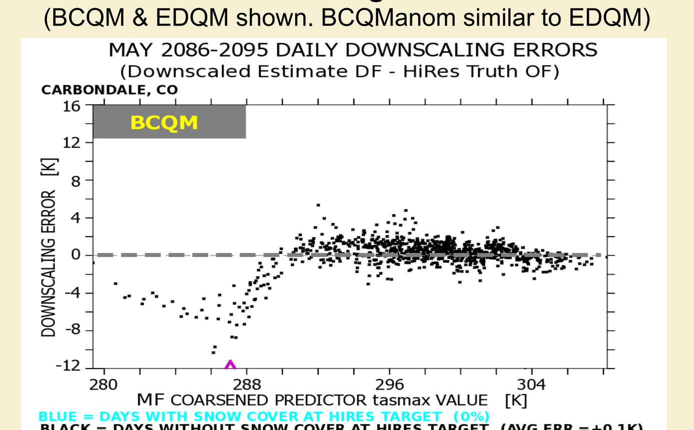
***Historical Period SD Performance:*** All 3 SD methods perform fairly well, with May daily mean downscaling errors (DH - OH) between -0.2K & +0.3K, compared to a pre-downscaling mean bias (MH - OH) of +3.6K. But individual daily downscaling errors range from +16K to -12K, due to the snow-related challenge outlined above.

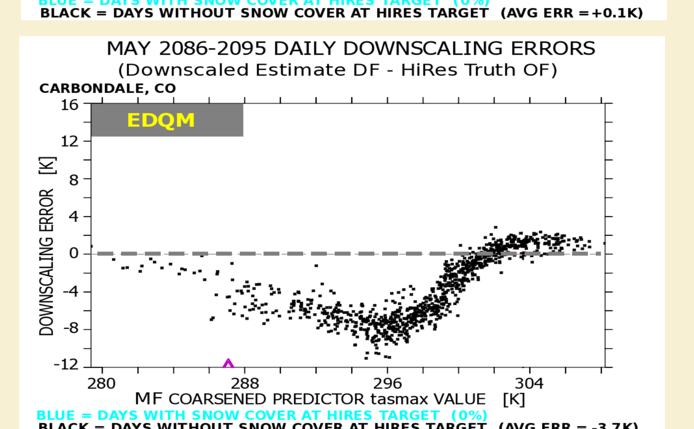
***Future Period SD Performance:*** The *BCQM* method performs markedly better than *EDQM* & *BCQManom* in the snow-free future. The largest differences occur when the large-scale MF predictor is 290K < MF < 300K. In the historical period, MH values in that range tend to occur on snow-free May days. The SD methods' future skills differ because *BCQM* transfer equations operate on the raw MF tasmax values, applying bias corrections representative of snow-free conditions when MF > 290K. Conversely, the *BCQManom* & *EDQM* methods incorporate adjustments for the projected warming trend (~7K for CONUS), so the bias corrections they make in this MF temperature range, in effect, reflect some snow effects. The intent of these two methods to adjust for the strong projected warming thus leads to an underestimation of Carbondale's true future tasmax values and multi-degree errors (DF < OF).

**Historical Period Downscaling Errors vs. MH values**
(BCQM results shown. Other SD methods are similar)




**Future Period Downscaling Errors vs. MF values**
(BCQM & EDQM shown. BCQManom similar to EDQM)




EDQM & BCQManom distort (reduce) the future May monthly mean warming signal by more than 3K. BCQM performs better.

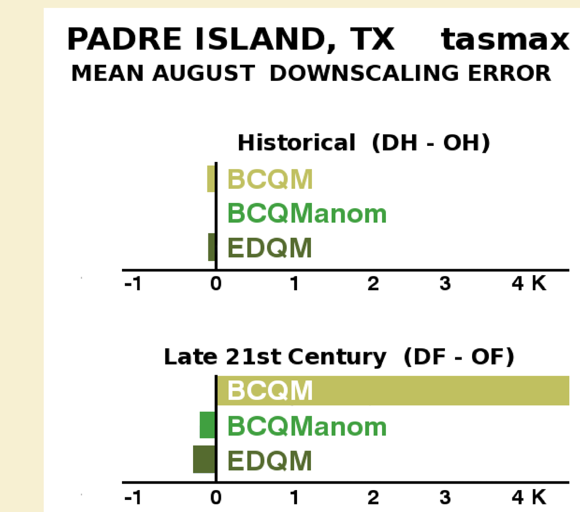
| | |
|---|---|
| EDQM<br>AVG ERR = -3.7K<br>CAUTION STATIONARITY ASSUMPTION VIOLATION | BCQManom<br>AVG ERR = -3.0K<br>CAUTION STATIONARITY ASSUMPTION VIOLATION |
| BCQM<br>AVG ERR = +0.1K | |

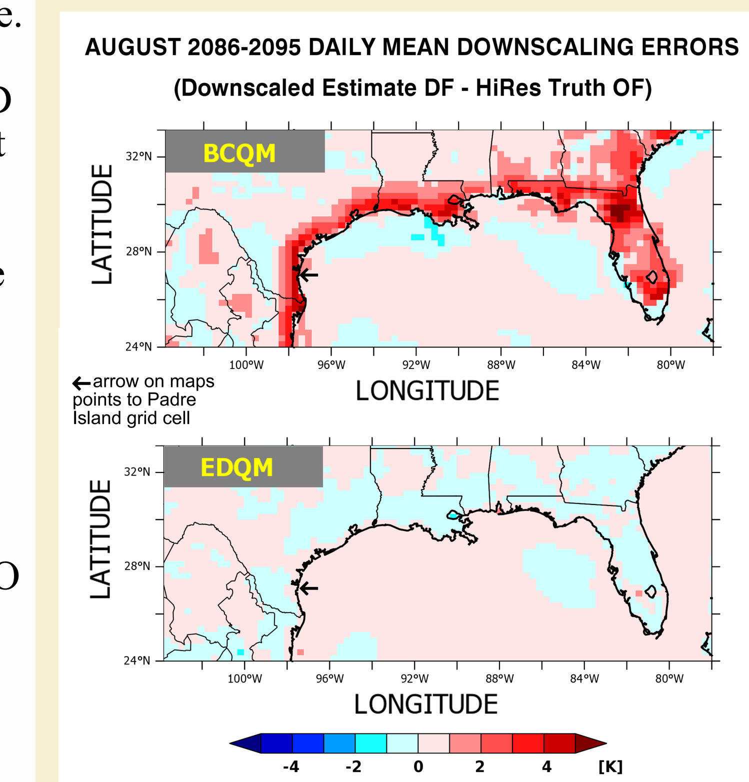
### CASE 2: U.S. SOUTHERN COAST IN SUMMER

***Month:*** August. ***Location:*** near & inland of Padre Island, Texas. A 25km grid cell similar to many coastal points in that it is mostly land, but the overlying 200km grid values are more maritime.
***The Statistical Downscaling Challenge:*** Distributions of daily August tasmax OH & OF values have larger variances, higher means, and different shapes (more positively skewed) than do the more maritime MH, MF distributions. Also, the climate change signal causes most late 21st century MF values to exceed the warmest historical MH value.

***Historical Period SD Performance:*** All three SD methods perform relatively well, yielding August monthly mean downscaling errors (DH - OH) of < 0.25K. Also, all SD methods generate DH distributions with shapes and variances that more closely resemble the OH distribution than the large-scale and more maritime MH tasmax data.

***Future Period SD Performance:*** The *EDQM* & *BCQManom* methods perform markedly better than *BCQM* does in the warmer future – the opposite of the results seen in the Carbondale, CO case. At the Padre Island 25km grid cell, *BCQM* introduces a mean bias of +4.6K for the late 21st century August case, and multi-degree warm biases occur at most U.S. southern coastal land points *(upper map)*. Such SD-created warm biases are not found in the *EDQM (lower map)* & *BCQManom* cases. Methods like *BCQM* that do not adjust for future trends can break down when posed with MF > *max*(MH), as they may rely on *ad hoc* tail adjustments to apply the correction at *max*(MH) to all warmer MF values – a poor choice when the MH and OH distribution shapes differ. SD methods that incorporate adjustments for trends guard against this particular pitfall.




BCQM distorts (amplifies) the future August monthly mean warming signal by more than 4K. EDQM & BCQManom perform better.

| | |
|---|---|
| EDQM<br>AVG ERR = +0.2 K | BCQManom<br>AVG ERR = +0.2K |
| BCQM<br>AVG ERR = +4.6K<br>CAUTION STATIONARITY ASSUMPTION VIOLATION | |

## THE 3 SD METHODS:

We show results drawn from three relatively simple SD methods exhibiting behaviors commonly found in the larger set of methods we have examined. The ***BCQM*** and ***BCQManom*** methods employ the same simple **B**ias **C**orrection **Q**uantile **M**apping process and use OH and MH in the training step. ***BCQManom*** operates on tasmax anomalies (historical & future climatologies are removed) whereas ***BCQM*** operates on raw tasmax values. The ***EDQM*** (**E**qui-**D**istant **Q**uantile **M**atching) method (Li *et al.*, 2010) trains using OH, MH, and MF, incorporating an adjustment to account for distributional differences between the future and historical model periods.

### About the GFDL Statistical Downscaling Team:

Our team seeks to evaluate the performance characteristics of SD methods and to share our findings so others can make better informed decisions about the suitability of different SD methods for various applications. To date, our collaborative work has used a combination of "Perfect Model" experiments and sensitivity studies, focusing on the CONUS region and time scales ranging from monthly forecasts to multi-decadal projections.

**Project Web Page: www.gfdl.noaa.gov/esd**


SOUTH CENTRAL CLIMATE SCIENCE CENTER

**This work was funded in part by the U.S. Dept. of the Interior's South Central Climate Science Center**


## SUMMARY:

The research presented here illustrates how statistical downscaling pitfalls vary geographically, with time of year, climate conditions, and across SD methods. Some SD methods' poor performance under certain future conditions could not have been gleened from typical evaluations based only on available historical observations. It was the availability of proxies for future observations in the Perfect Model experimental design that allowed assessments of a critical ***stationarity assumption***. From the perspective of climate impacts studies, the question of whether a specific pitfall may be a serious concern depends on the details of a study's climate data needs and sensitivities – factors that can preclude simple one-size-fits-all guidance. However, increased awareness that pitfalls and non-stationarities of the type exposed in these PM experiments exist can help promote the better-informed use of statistically downscaled climate projections.

## REFERENCES:


Dixon, K. W., *et al.*, 2016: Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Climatic Change*, 135, 395–408.

Lanzante, J. R., *et al.*, 2017: Some Pitfalls In Statistical Downscaling Of Future Climate. *Bull. Amer. Meteor. Soc., in press.*

Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res., 115, D10101.*