Adding Value *And* Uncertainty

On the role of statistical downscaling in connecting upstream climate science with applied research

Keith W. Dixon & GFDL's Empirical Statistical Downscaling Team

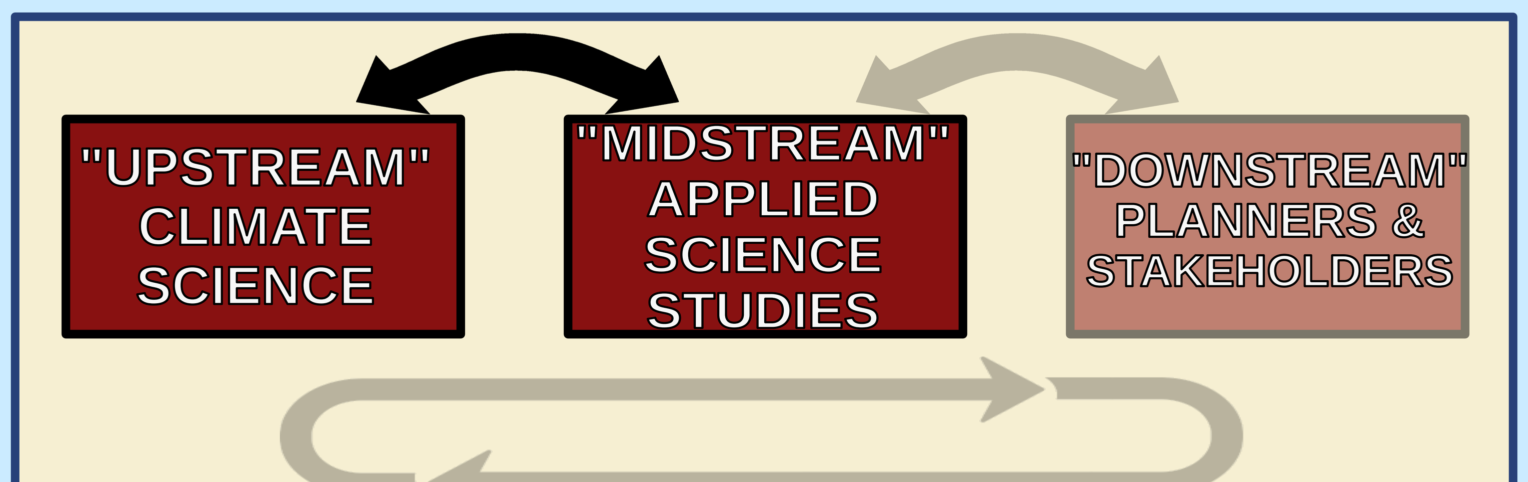
www.gfdl.noaa.gov/esd

Background:

Bias correction & statistical downscaling methods often are used to address dynamical climate model shortcomings, with an aim being to create data products more suitable for direct use as input to applied science studies. However, different statistical processing techniques and configurations yield somewhat different results, thereby revealing uncertainties in the presumably "value-added," statistically processed data products. Appreciation of this often overlooked source of uncertainty can lead to the question...

"How correct are the corrections being applied to refine climate predictions and projections?"

In recent years, NOAA GFDL's Empirical Statistical Downscaling (ESD) Team has conducted research evaluating the performance of several commonly used bias correction & statistical downscaling techniques, with a focus on United States daily surface temperatures and precipitation. Our approach can be considered somewhat analagous to Consumer Reports® testing. Here we present results that illustrate the utility of these evaluation approaches and help frame questions to ponder when seeking to match statistically refined climate predictions and projections with the data requirements and sensitivities of a given applied science study.



The development, distribution, and informed use of statistically refined climate projections involves multiple interdisciplinary exchanges. Here we focus on strengthening the link between ***"upstream"*** researchers who produce climate projection data products and ***"midstream"*** researchers who use that data as input to their applied science climate change studies (*e.g.*, research into the effects of future climate change on human health, living marine resources, agriculture, endangered species, infrastructure, etc.)

Because the data requirements and sensitivities of different applications to climate change varies greatly, there is no one-size-fits-all answer to the question about what is "the best" set of statistically refined climate projections to use as input to all climate impacts studies.

About the GFDL Empirical Statistical Downscaling Team:

Our team seeks to evaluate bias correction & statistical downscaling methods and to share our findings so others can make better-informed decisions about the suitability of different data products for various applications. To date, our work has focused on CONUS surface climate (daily variables) & time scales ranging from monthly forecasts to multi-decadal climate projections.

Two GFDL ESD Team References:

[1] Lanzante JR, *et al.* (2018): Some Pitfalls in Statistical Downscaling of Future Climate. *Bulletin of the American Meteorological Society.*

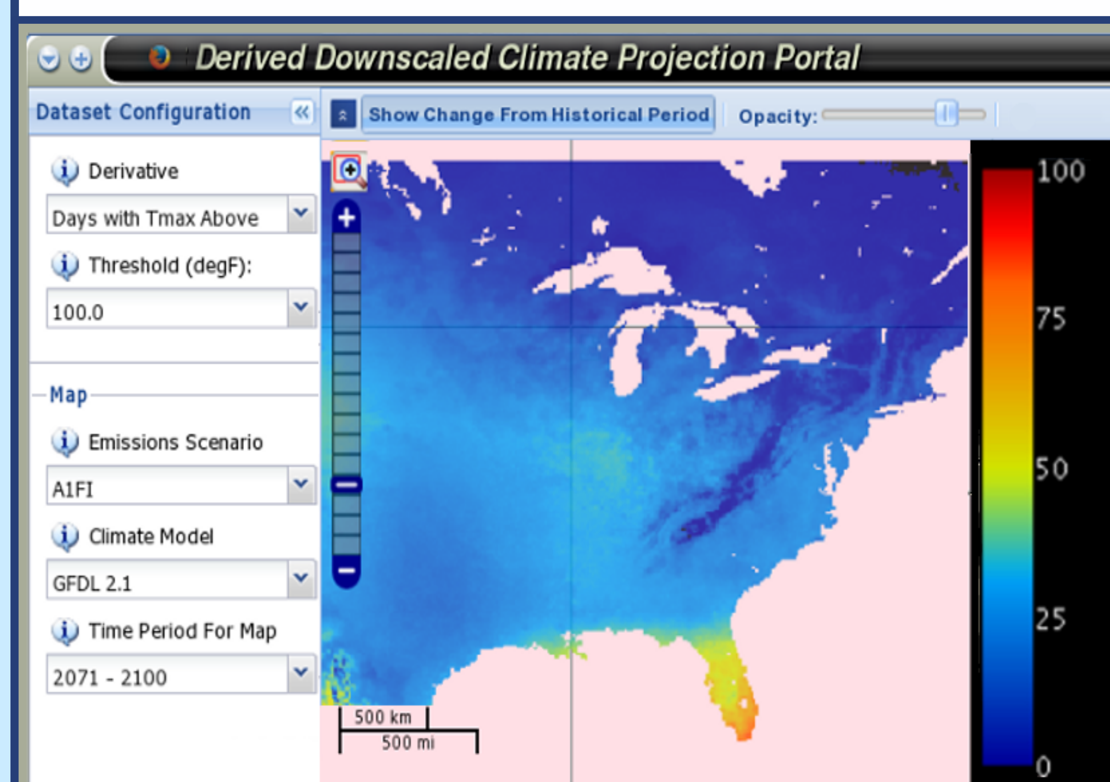
[2] Dixon KW, *et al.*, (2016): Evaluating the stationarity assumption in statistically downscaling climate projections: Is past performance an indicator of future results? *Climatic Change.*

A Perfect Model Experiment Example:

Q: How can bias correction & statistical downscaling methods alter the "true" climate change signal?

Regardless of the math details used in a bias correction / statistical downscaling technique, there is an underlying assumption that information gleaned about model biases during a historical period is equally applicable when applied to future climate model projections (a kind of ***"stationarity assumption"***). Our team's research shows that this assumption does not always hold and that conditions associated with serious violations of the stationarity assumption vary from method to method, geographically, and across climate variables.

For example, consider the following case...



This figure, showing the projected increase in the number of days per year with temperatures above 100F for the late 21st century, was shown during a discussion of data products that can be used in impacts studies. Attention was drawn to Florida, with increases of more than 50 days per year.

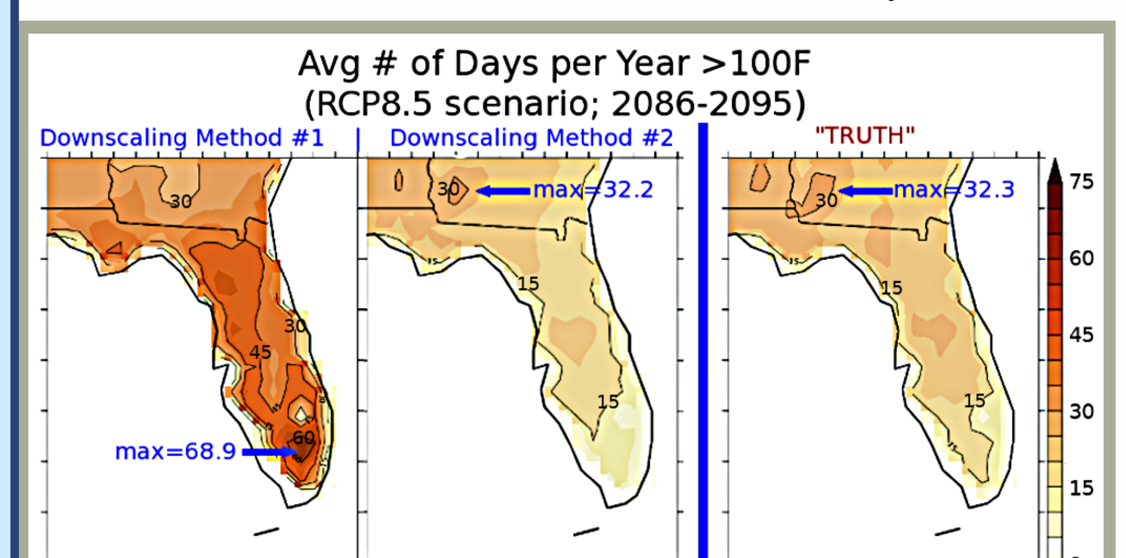
A couple people scratched their heads.

"Has Miami Beach every been above 100F?," one asked.

"I don't remember Florida being a warming signal hot spot in papers about that GCM," said another.

Q: Is this Southern Florida feature reliable?

A "Perfect Model" experimental design [1],[2] allows us to test the stationarity assumption, using a high resolution model as a proxy for future observations. The left and center figure panels below show results from two statistical methods. Both were created using the same climate model inputs and the same proxy observations - only the statistical methods differ. Method #1 is very similar to the method shown above.



The right panel shows the "true" future values, which are closer to 15 days than 60 days in southern Florida. ***Clearly, Method #1 violates the stationarity assumption in a way that amplifies the climate change warming signal.***

Analyses presented in papers [1] and [2] describe the reasons some bias correction methods exhibit this behavior in coastal regions for large climate warming projections.

So, the results above tell us that Method #2 is simply better than Method #1 when bias correcting / statistical downscaling daily temperature projections, right? ***WRONG!***

As explained in [2] and in a 2017 AGU poster (available on our home page and via the QR code to the right) in some mountainous regions during spring snowmelt months the opposite is true, with Method #1 outperforming Method #2!

These research results help illustrate a key point regarding the selection of statistically downscaled data products for use in a climate impacts study:

There is no one "best" method. It's application dependent!

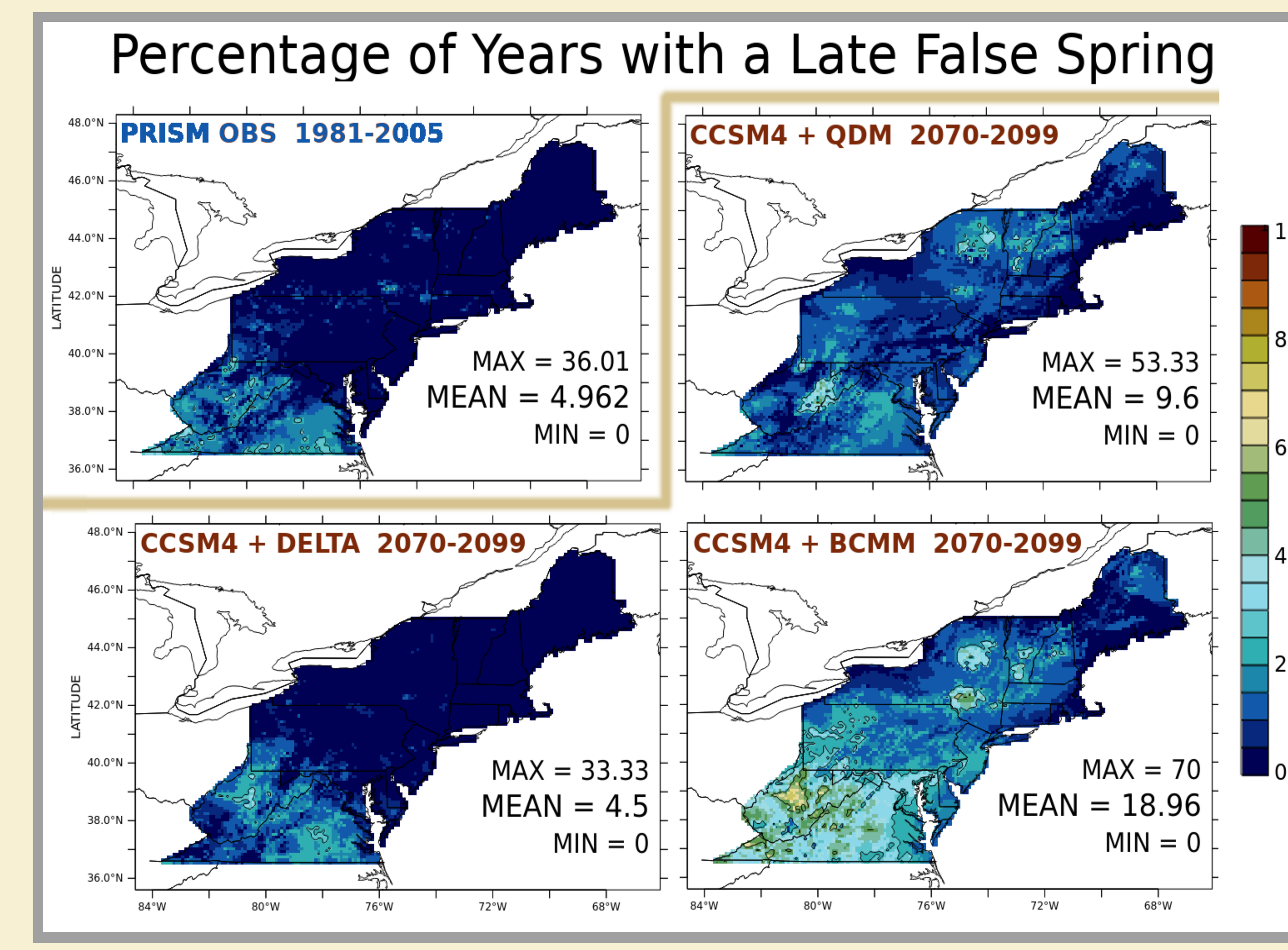
A Sensitivity Study Example:

Q: How sensitive are different climate diagnostics to the choice of bias correction / statistical downscaling method?

The short answer is simply, ***"It depends on several things."***

Determining to which climate factor(s) the application of interest is most sensitive is key. Are changes in monthly or seasonal means the driving force? Or are day-to-day fluctuations, extreme events, spells, or low frequency variations critical? Climate models simulate some factors better than others. Thus, the work one is asking a statistical refinement process to do depends on the climate factor of interest *-and-* on the climate model simulation's resemblance to observations. And not all of the many statistical refinement methods that exist address climate model shortcomings in the same way.

The "Late False Spring" (*LFS*) example below is based on a phenological index that uses daily minimum and maximum temperatures as input. An *LFS* occurs when a hard freeze occurs after the "first bloom date" in spring when flowers start to open on the plants used to develop the indices.[3]



Top Left: *LFS* frequency (1981-2005) from regridded PRISM[4] observation-based daily min & max temperature data used for subsequent statistical refinements. **Other 3 Panels:** Late 21st century *LFS* frequency based on the NCAR CCSM4 model, the PRISM data, and three different relatively simple statistical methods (*QDM*: a quantile mapping bias correction variant, *Delta:* change factor by the ~monthly mean modelled change signal, *BCMM:* bias correction by the ~monthly mean model historical bias.)

[3] Schwartz, MD, *et al.*, (2006): Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biology.*

[4] The PRISM Climate Group, Oregon State Univ. http://www.prism.oregonstate.edu/

These three statistical refinement methods produced different answers for late 21st century *LFS* projections, though each used the same input files. Do statistically downscaled CCSM4 projections show little change in the probability of an *LFS* occuring (*Delta*) or does the probability increase on average from about 1 in 20 to almost 1 in 5 (*BCMM*)? Or something in between (*QDM*)? This is another illustration of ***Statistical Downscaling Uncertainty***. Unlike the Perfect Model studies, there is no proxy "future truth" data available to allow one to clearly assess which is more correct.