Statistical downscaling using machine learning

Carlos F. Gaitán[1,2]

[1]*South Central Climate Science Center – University of Oklahoma*

[2]*National Oceanic and Atmospheric Administration – Geophysical Fluid Dynamics Laboratory*

Overview

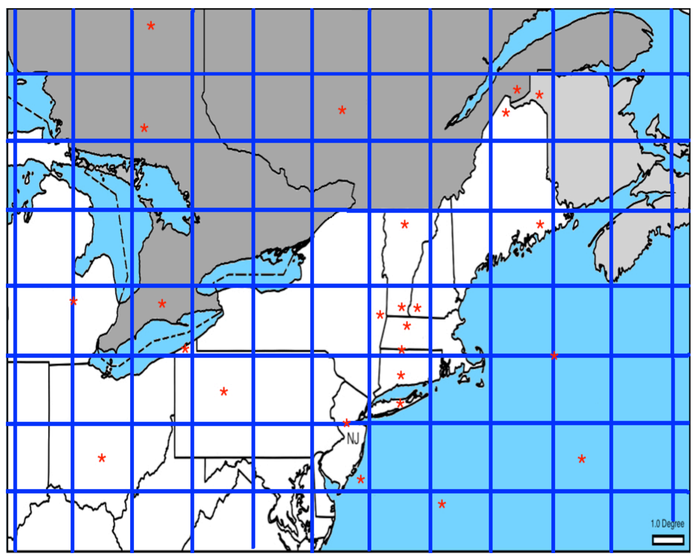
- **Statistical downscaling is used to add value to the coarse resolution model output.**
- **Statistical downscaling relies on four (4) assumptions, one of them time-invariance. Here at GFDL we are testing that assumption using a "perfect-model" evaluation framework.**
- **Machine learning methods are used to derive empirical relationships between coarse resolution data and local scale observations; and provide a viable alternative to classical linear regression techniques.**

Statistical Downscaling

What is Statistical Downscaling?

Statistical Downscaling is the process of adding value to coarse resolution model output by linking the state of a large scale predictor (X, e.g. from a GCM) and the state of a predictand (Y) representing a much smaller scale.

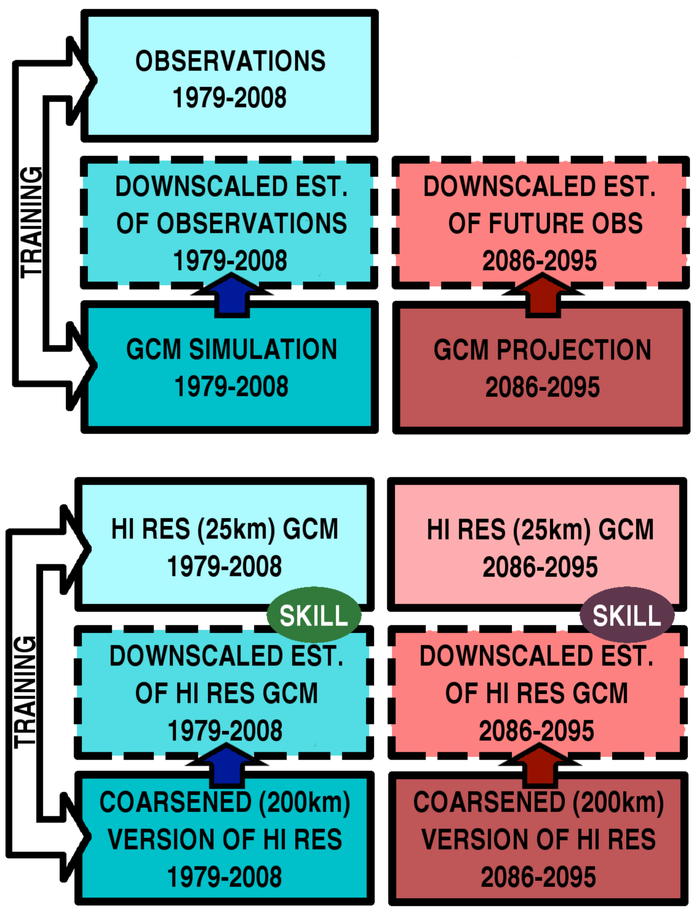
Why we need to add value to the GCM's output?



Statistical downscaling assumptions:

1) Time-invariance
2) Strong predictor-predictand relationship
3) Description of change
4) Model representation

Statistical downscaling evaluation



Top figure: Classical SD evaluation.
Bottom: GFDL's Perfect Model Evaluation Framework

Machine Learning

What is machine learning?

Machine learning is a branch of artificial intelligence. It studies systems that can learn from data.

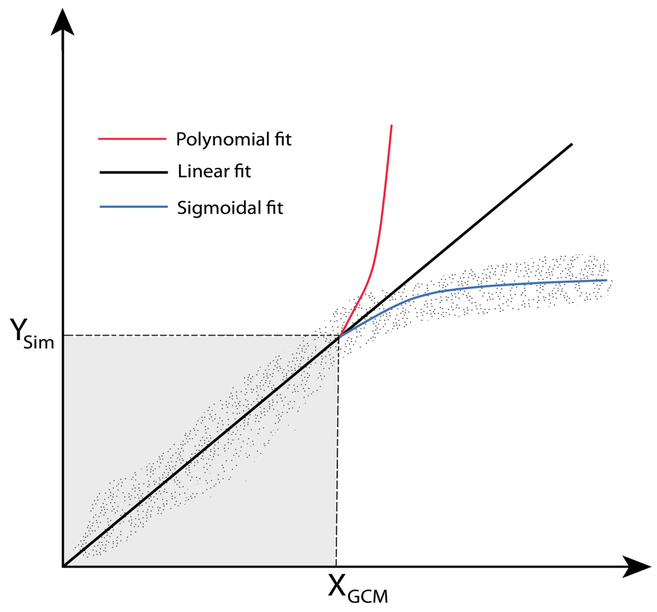
Our current research focuses on systems that are able to identify empirical relationships between the coarse resolution global climate models and finer scale observation-like datasets.

Types of machine learning algorithms used:

- Artificial Neural Networks
- Support Vector Regression
- Regression and Classification Trees
- Genetic Programming

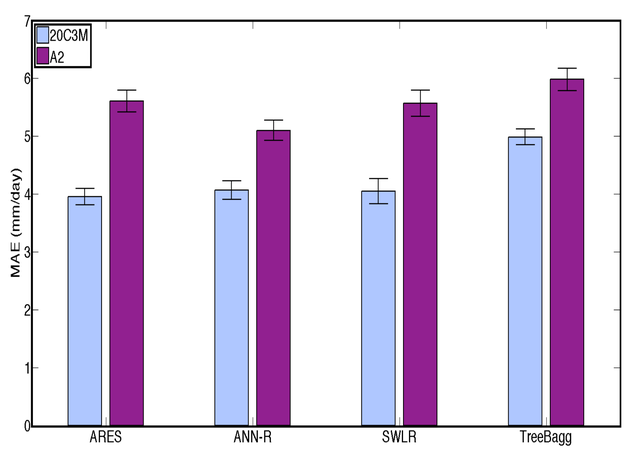
Why we use machine learning algorithms to downscale?

Extrapolation effect



The schematic diagram illustrates the types of extrapolation used by different downscaling methods. The area in gray represents the historical period domain, where $Y_{sim} = f(X_{GCM})$. The area in white represents conditions outside the training/ historical period.

Comparing linear and nonlinear downscaling methods (example: precipitation downscaling)



Linking research to stakeholders needs

Statistical downscaling techniques, like the ones implemented by GFDL's ESD Team, are often used to refine the coarse resolution outputs from GCMs or reanalysis products as the spatial resolution of these models is insufficient to resolve many local scale phenomena occurring at a much higher scale, like convective and orographic precipitation (*Salameh et al. 2008*), and accurate local estimates of meteorological variables are often needed as professionals working in the impacts and adaptation community, ecologists, biologists, engineers, and hydrologists, among others, regularly use local scale variables.

Applications

Energy

Wind Turbines Design

Engineering

Engineering Design
Hydrology
Hydraulics

Agriculture & Aquaculture

Food Security
Irrigation & Drainage
Insurance

Ecosystems & Ecology

Animal Ecology
Ecosystem Modeling
Ecosystem Services
Forest Ecosystems
Grassland Ecosystems
Wetland Ecosystems

Acknowledgements:

carlos.gaitan@noaa.gov