Statistical downscaling of daily precipitation and the stationarity assumption

Carlos F. Gaitan[1], Keith W. Dixon[2], V. Balaji[3], Berrien Moore III[1], Renee McPherson[1], and Aparna Radhakrishnan[4]

1 South Central Climate Science Center & University of Oklahoma
2 NOAA - Geophysical Fluid Dynamics Laboratory
3 Cooperative Institute for Climate Sciences, Princeton University
4 DRC/HPTG

Overview

- **We downscaled daily precipitation outputs from two 30-year long runs (1979-2008).**
- **Transform functions derived from the historical period were applied to future climate change simulations.**
- **We used a "Perfect-Model" approach to evaluate the downscaling method performance in the future.**
- **We implemented a hybrid CART-SVR downscaling method and selected the SVR hyperparameters using Evolutionary Strategies (ES).**
- **The study area includes ~22,000 gridpoints over the contiguous US. Here we show results from 16 points located in different climate regions.**

1. Introduction

Global Climate Models' (GCM) resolution is often too coarse for direct use in regional climate change impact studies (Warner 2011), and doubling their resolution generally implies 16 times the amount of computations (Coiffier 2011).

Because of the final users' need for fine-scale information at lower computational cost, various statistical techniques and higher resolution regional climate models (RCMs) have been developed for downscaling GCM simulations to regional and local scales (Denis et al. 2002). However, the RCMs can also be computationally intensive and their spatial resolution does not always provide the information required by regional climate change impact studies (Vrac and Naveau 2007).

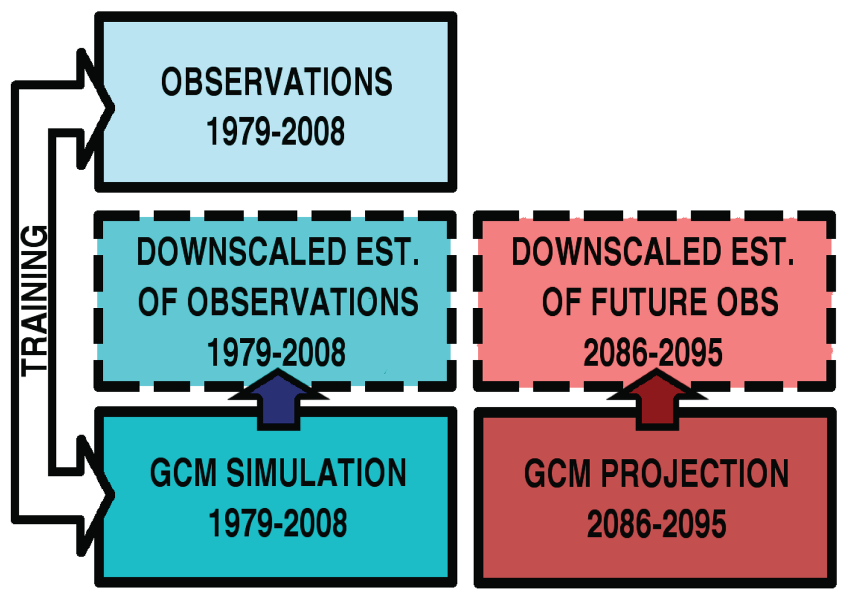

Figure 1. Statistical downscaling schematic

On the other hand, the statistical downscaling (ESD) techniques derived from the local forecasting model output statistics (MOS) method (e.g. quantile matching approaches like the CDFt method (Michelangeli et al. 2009)), or from the numerical weather prediction's perfect prog post-processing technique (Marzban et al.2006), establish a synchronous (valid at a given time) statistical relation between the coarse resolution predictors (i.e. GCMs or the global reanalysis data sets) and the small-scale predictand(s) representing the climate of a region (Warner 2011).

Statistical downscaling methods are computationally inexpensive and easier to implement compared to the RCMs, but rely on the **time invariance** of the statistical relationships.Therefore, both the RCMs and the statistical downscaling methods depend on the credibility of the coarser scale GCM to model the predictors.

In real-world applications, the process of statistical downscaling uses 3 types of data files as input (SOLID boxes in the top figure - Figure 1) in order to produce as output downscaled future projections (DASHED boxes).

In general, a transform function is calculated during the training step in which a statistical technique compares observations to GCM output representing the same period. Applying the transfer function (ARROW) to independent historical and future GCM outputs yields dowscaled projections.

Though one can assess the skill of the ESD technique during the historical period, lacking observations from the future, there is no straightforward way to quantitatively determine the future skill.

2. Perfect-Model evaluation

Our "perfect-model" experimental design seeks to isolate and quantify key aspects of the stationarity assumption. As illustrated in Figure 2, this design does not make use of observational data, Rather we substitute high resolution model outputs for observations and we substitute coarsened (smoothed by interpolation) versions of the high resolution GCM output for what would be the GCM results, in a more typical, real world application.

Specifically, the datasets we use all derive from the GFDL-HIRAM-C360 model (aka "C360") simulations. The model's domain covers the entire globe, but we will examine a region centered on the contiguous 48 United States (CONUS 48). Two time periods are considered: a 30-year long "historical" era (1 Jan. 1979 - 31 Dec. 2008 - BLUE boxes) and 10-year long "future" era (1 Jan. 2086 - 31 Dec. 2095 - RED boxes).

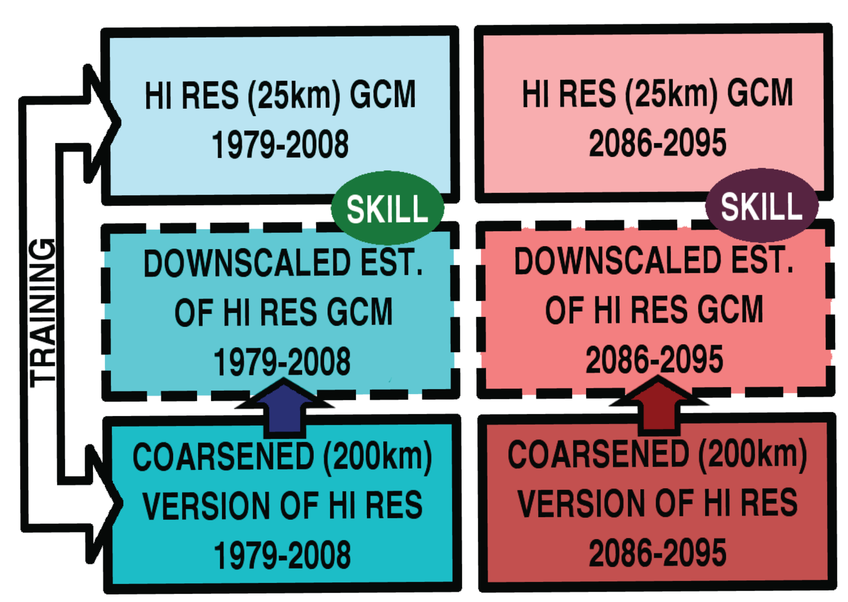

Figure 2. Perfect-model evaluation

For the historical era, there is a two member ensemble of model runs, for a total of 60 years of data. For the future era, there are two different sets of C360 experiments, each of which is comprised of three ensemble members (30 years of data).

A notable difference between the two future ensembles is that one (C) on average exhibits about 2 degrees Celsius more warming over the CONUS 48 than does the other ensemble of future experiments (E).

To obtain the transfer functions, we are testing different linear and nonlinear regression and classification methods, but, for the purposes of this poster, we will focus on the downscaled results from a support vector regression (SVR - Vapnik 1998) model -used to get precipitation amounts- and different classification methods (e.g. support vector classification, classification trees), used to determine the days with precipitation.

SVR has shown to be an effective downscaling technique when used to get finer scale local precipitation over India (Tripathi et al. 2006) and when used to get seasonal predictions of winter extreme precipitation over Canada (Zeng et al. 2010), but its model output is highly dependant on the values of multiple hyper-parameters, often optimized via an extensive grid search. Alternatively, one could use evolutionary strategies to obtain these parameters, decreasing the computing time. On the other hand, classification and regression trees (CART) ensembles have been extensively used in numerous applications (Hastie et al. 2009, Hsieh 2009) showing to be a fast and efficient classifier, often outscoring popular methods like discriminant classification, naïve-Bayes classification and k-nearest neighbours (Breiman 1996).

3. Objective

To test the stationarity assumption by determining the extent to which the selected ESD methods' skill is degraded for the future relative to the historical period.

To test if the methods are equally skillful when downscaling precipitation over different climate regions.

4. Study region

The study includes approximately 22,000 grid points over the CONUS 48 region, but here we focus on 16 points from different climate regions across North America.

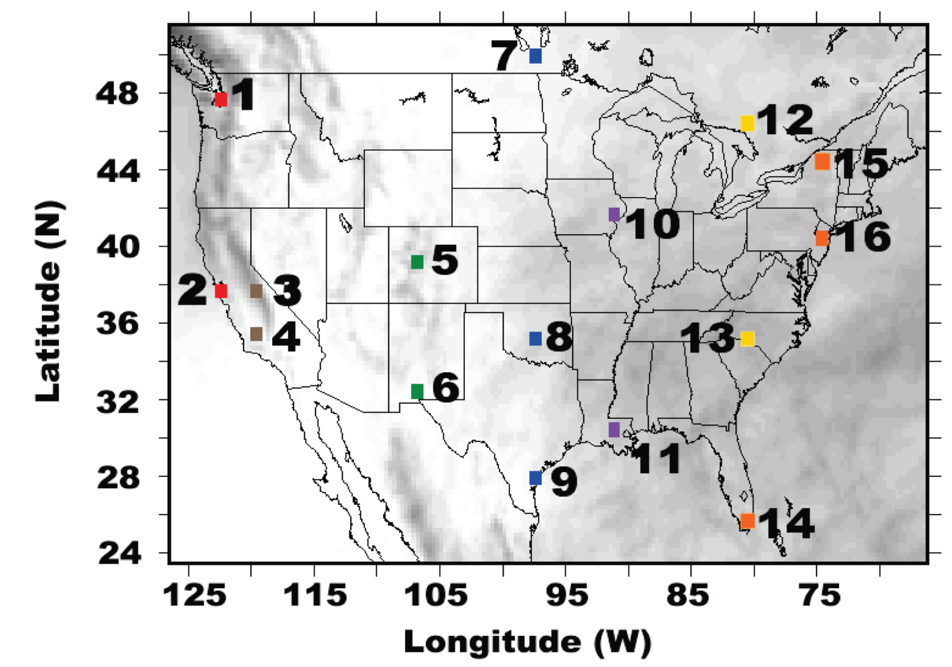

Figure 3. Study Area

Point number	Location	State/Province
1	Seattle	WA
2	San Francisco	CA
3	Yosemite	CA
4	Bakersfield	CA
5	Aspen	CO
6	Las Cruces	NM
7	Winnepeg	MB
8	Norman	OK
9	Corpus Christi	TX
10	Iowa City	IA
11	Baton Rouge	LA
12	Sudbury	ON
13	Charlotte	NC
14	Miami	FL
15	Saranac Lake	NY
16	South Brunswick	NJ

Region of interest:
***Approx. 23N - 52N, 128W - 65W**
***A subset of the C360 model's global domain.**

5. Methods

Support vector machines were extended to regression problems after their conception as nonlinear classifiers (Vapnik 1995).

In general, if **x** denotes *m* predictors and *y* the predictand, by introducing a mapping function *O,* the nonlinear regression between **x** and *y* can be converted to a linear regression problem between *O* and *y*:

$$f(\mathbf{x},\mathbf{w}) = <\mathbf{w},O(\mathbf{x})> + b,$$

where <,> is the inner product and **w** and *b* are coefficients obtained by minimizing the *e*-insensitive error norm:

$$|f(\mathbf{x},\mathbf{w}) - y| = \{ 0, \text{ if } | f - y | < e, \text{ or } | f - y | - e, \text{ otherwise.}$$

However, as $O(\mathbf{x})$ may be a very high dimensional vector solving the linear regression problem may be prohibitely expensive (Zeng et al. 2010), hence a kernel trick is used to replace the inner product $<O(\mathbf{x}), O(\mathbf{x}')>$ in the solution algorithm, and Lagrange multipliers are used to minimize the cost function.

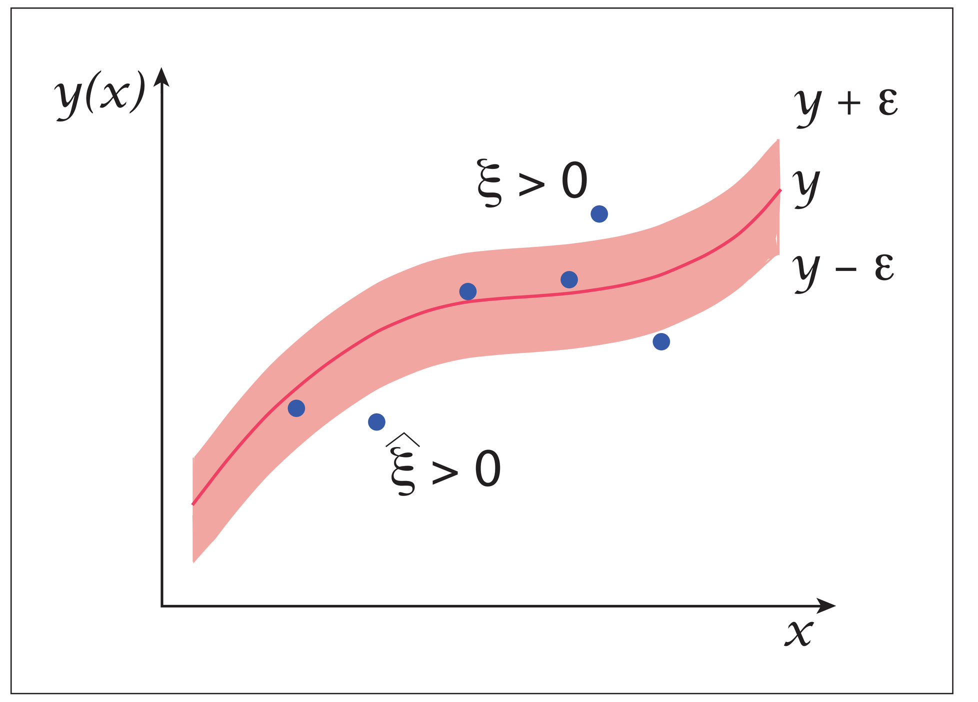

Figure 4. SVM regression (Bishop, 2006)

Here we use the Gaussian kernel with the hyperparameter Sigma controlling its width. Besides e and Sigma, we need to calculate a cost hyperparameter which controls the regularization or weight penalty term.

The 3 hyperparameters are generally calculated using an exhaustive grid search or are determined a priori using Cherkassky-Ma (2004) suggested values; nevertheless, the grid search is computationally expensive and the Cherkassky and Ma estimates are often not optimal. Alternatively, one can use evolutionary strategies (ES) to optimize the hyperparameters, which are used as chromosomes and can co-evolve with the solutions.

In particular, we implemented the uncorrelated mutation with **p** step sizes, following Eiben and Smith (2004).

To classify precipitation occurrences, we used support vector machines for classification (SVM-C) and classification and regression trees (CART). We used coarsened GCM outputs from the 9 gridpoints surrounding each of the 16 study points (Fig. 3) as predictors for the regression and classification models. The predictands were precipitation amounts (for the regression model), and a vector of zeros and ones, indicating days with and without precipitation (for the classification models).

6. Results

The historical classification results (Fig. 4) show both CART and SVM-C underpredicting the total number of days with precipitation above the 1 mm/day threshold. CART models outscored SVM-C models in terms of the Peirce skill score (not shown) and obtained more non-zero precipitation days.

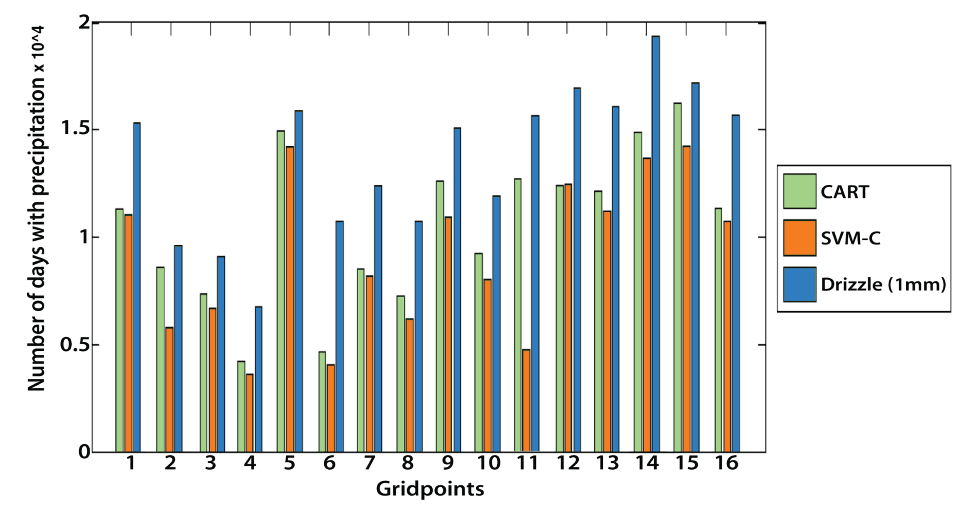

Figure 5. Number of days with precipitation

Regarding the precipitation amounts, we evaluated the downscaled results in terms of historical and future mean absoulte error skill score (MAE SS) to assess if their skills were time-invariant. The hi-resolution (~25km) model outputs were used as reference.

The results show that for 9 out 16 points the hybrid CART-SVR downscaling model had positive historical and future MAE SS. Effects of the CART underprediction include biases on the length of the wet/dry spells and on the yearly precipitation amounts. The SVR model underpredicted the precipitation amounts.

When looking at the historical and future ESD skills (Fig. 5) we found marginal skill deterioration for most of the selected 16 points, while some ESD models improved their MAE SS between periods. In particular, we noticed that the ESD models from stations located west of the Rocky Mountains presented higher skills, suggesting the models' ability to recover local scale futures unresolved by the coarsened GCM.Future implementations will test other classification methods (e.g. neural networks, k-nearest neighbor) and will expand the predictor variables aiming to improve the overall MAE SS. Ongoing work includes the development of statistical downscaling models using Bayesian neural networks, quantile regression and different types of linear and nonlinear regression.

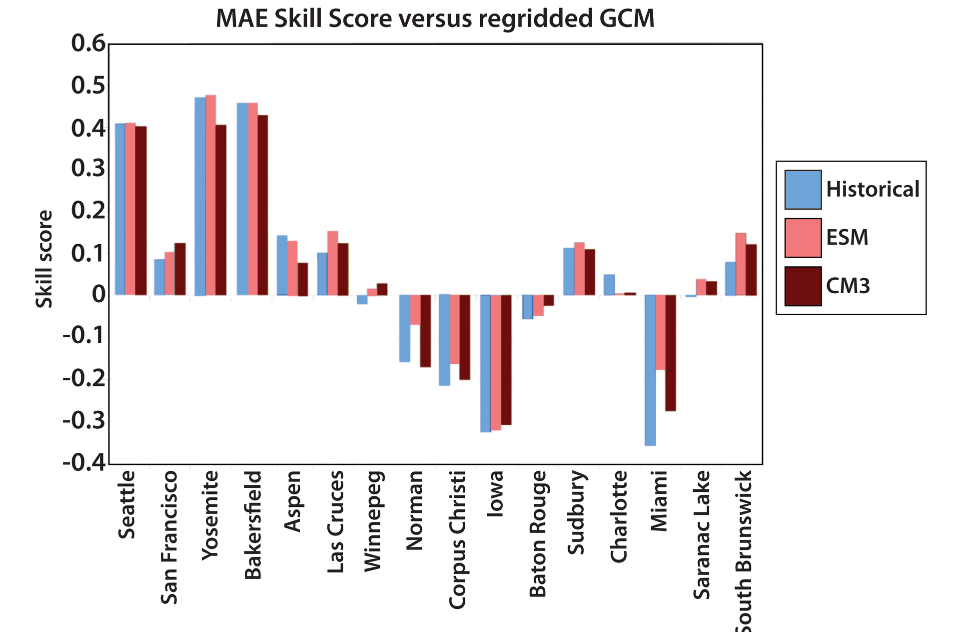

Figure 6. Downscaled MAE Skill score

Overall, we aim to learn more about the methods' strengths and limitations, and specially learn about the future behavior of different extrapolation techniques, as some of the future GCM outputs may be outside of the historical period range used during training.

Conclusions

- **For 9 out of 16 points the hybrid CART-SVR downscaling method had positive historical and future MAE skill score.**
- **The classification methods (CART and SVR-C) under-predicted the number of rainy days.**
- **Points near the Rocky Mountains had better skill score. This suggest the method's statistical refinement is more evident where the coarsened data cannot represent local scale processes.**
- **Results could be improved by adding other relevant predictors like convective precipitation and specific humidity.**

Acknowledgements

We wish to thank the National Oceanographic and Atmospheric Administration for sharing the GFDL HIRAM C360 model outputs, and Jeff Varanyak for his help with Figure 4.

References

https://www2.image.ucar.edu/sites/default/files/ci2013_submission_45Gaitan.pdf

carlos.gaitan@noaa.gov