A New Distribution Mapping Technique for Climate Model Bias Correction

Seth McGinnis, Doug Nychka, Linda Mearns

IMAGe – CISL – NCAR
mcginnis@ucar.edu
2014-09-30
Outline

• Bias correction and distribution mapping

• Overview of 4 competing methods
  – Probability mapping
  – Quantile mapping
  – ARRM
  – Empirical CDF map

• New technique: Kernel Density Distribution Mapping

• Oracle Evaluation

• Results
Bias Correction

• Models have bias
• Removing bias improves usability
  – especially for non-specialists (e.g., impacts users)
• Make model output statistically similar to observations
  – Similar mean
  – Similar variability (spread, stdev)
  – Similar distribution
Bias Correction Process

1. Moving Window
   – to handle seasonality
2. Normalize data
   – recenter (mean 0)
   – rescale (sd 1)
   – detrend
   – transform to deskew (e.g., log precip)
3. Distribution Mapping
   – reshape distribution
4. Denormalize
   – $\sigma_{\text{proj}} = \sigma_{\text{obs}} \times \sigma_{\text{fut}} / \sigma_{\text{cur}}$
   – $\mu_{\text{proj}} = \mu_{\text{obs}} + \mu_{\text{fut}} - \mu_{\text{cur}}$
   – reapply trend
   – reverse transform
Bias Correction Process

1. Moving Window
   - to handle seasonality
2. Normalize data
   - recenter (mean 0)
   - rescale (sd 1)
   - detrend
   - transform to deskew (e.g., log precip)
3. Distribution Mapping
   - reshape distribution
4. Denormalize
   - $\sigma_{\text{proj}} = \sigma_{\text{obs}} \times \sigma_{\text{fut}} / \sigma_{\text{cur}}$
   - $\mu_{\text{proj}} = \mu_{\text{obs}} + \mu_{\text{fut}} - \mu_{\text{cur}}$
   - reapply trend
   - reverse transform
Distribution Mapping

\[ x_{bc} = \text{CDF}_{\text{obs}}^{-1}(\text{CDF}_{\text{mod}}(x_{\text{raw}})) \]
Q-Q Plot

Q-Q mapping: sort datasets and plot against one another
Use to visualize distribution mapping

Any monotonic function fit to Q-Q mapping is a transfer function
bias = obs - mod
Bias Correcting Future Data

Construct transfer function using current and observed data

Reshape distribution of future data using transfer function

N.B.: Assumes bias is stationary!
PMAP
Probability Mapping

- Fit parametric distribution to each dataset (using MLE or sample moments)
- Transfer function: composition of analytic CDF and quantile f’ns
- Distribution must be specified a priori
- For normal distribution, transfer function will be linear – equivalent to fitting Q-Q plot with a straight line
QMAP
Quantile Mapping

- Estimate quantiles of each dataset
- Transfer function: piecewise linear between quantiles
- Number of quantiles is a free parameter; for evaluation, we use:
  - “few” = 5
  - “some” = \( \sqrt{N} \)
  - “many” = \( N/5 \)
ARRM
Asynchronous Regional Regression Model

• Transfer function: fit Q-Q mapping with piecewise linear model using splines
• Up to 6 breakpoints at inflection points
• Find breaks using least-squares fitting on moving window
ECDF

Empirical CDF Mapping

• Transfer function: piecewise linear interpolation between points of the Q-Q mapping.
• Requires equal numbers of points in each dataset
• Used in BCSD (Maurer, et al)
KDDM
Kernel Density Distribution Mapping

• Transfer function:
  o calculate PDFs for each dataset using KDE
  o integrate to CDFs using the trapezoid rule
  o equate CDFs

• KDE (kernel density estimation) is well-developed in statistics literature

• Simple implementation: 12 lines of R code
KDDM Algorithm

Straightforward – only 12 of code lines in R
Oracle Evaluation

• Synthetic datasets
• Oracle: perfect bias correction of future
• Compare each technique to oracle

• 3 distributions
• 1000 iterations
• 450 data points
• 6 metrics: MAE, RMSE, MaxE, L tail error, R tail error, K-S stat
Mixture Distribution

Real data is frequently not Gaussian

- Test location: Pineville, AR
- 1970-2000
Performance metrics: Normal dataset

Mean absolute error

Root mean square error

Maximum error

Left tail error

Right tail error

Kolmogorov-Smirnov test
Performance metrics: Gamma dataset

Mean absolute error

Root mean square error

Maximum error

Left tail error

Right tail error

Kolmogorov-Smirnov test
Performance metrics: Mixture dataset

Mean absolute error

Root mean square error

Maximum error

Left tail error

Right tail error

Kolmogorov-Smirnov test
Computation Speed

Note log scale

ARRM has to loop on least-squares linear fit hundreds of times

KDDM is only 2x slower than fastest methods (ECDF / PMAP-SMP)

KDDM is 100x faster than slowest method (ARRM)
Conclusions

• PMAP performs best iff data is parametric
• Otherwise, KDDM is the best performer
• KDDM also performs well when data is fittable
• Based on well-developed methods
• Fast, flexible, simple to implement