Statistical/Stochastic approaches for downscaling

Simple Quantile Mapping
Spatial Disaggregation Quantile Delta Mapping
Bias-Correction & Stochastic Analog

Oct. 2018
Syewoon Hwang
Department of Agricultural Eng.
Gyeonsang National Univ.
Outline

1. Stochastic downscaling
   - Methodology
   - Significance

2. Case study
   - BCSD (Bias-correction & spatial disaggregation)
   - SDBC (spatial disaggregation & Bias-correction)
   - BCCA (Bias-correction & constructed analogue)
   - BCSA (Bias-correction & stochastic analog)

3. Comparative Evaluation

4. Application
   - Hydrologic implications

5. Message to take home
Bias–Correction & Stochastic analog (BCSA) vs. SQM, BCSD, SDBC, BCCA, SDQDM

(How to evaluate the performance?)
General framework of climate change impact assessments

- **General circulation models**
  - Reanalysis data
  - Dynamical modeling
  - Probability mapping approach
  - BCSD
  - SDBC
  - BCCA
  - BCSA

- **RCM**
  - GCM res.
  - RCM res.
  - Climate models & Statistical downscaling

- **Bias correction**
  - <10 km resolution
  - 100-300 km resolution

- **Downscaling**
  - GCM forcing

- **Impact models**
  - Agricultural factors
    - Water, Land use, Cropping, Vegetation, Soil, Topography
  - Integrated Application Model
  - The impact assessment

- **Evaluation of climate information**

![Diagram showing the general framework of climate change impact assessments](image)
Simple Quantile Mapping (SQM)

- Sampling daily data to build ECDF for a month
  - 12 ECDFs necessary

- GCM data
  - Historical (reference) Period
  - Future Period

- Spatial downscaling
  - Inverse distance weighting method
  - Disaggregate GCM values to target points (e.g. stations)

- Bias correction
  - Create CDFs of OBS and GCM and quantile map
  - Apply the developed quantile mapping to future GCM data

GCM data

Historical (reference) Period

Future Period

Spatial downscaling

Inverse distance weighting method

Disaggregate GCM values to target points (e.g. stations)

Bias correction

Create CDFs of OBS and GCM and quantile map

Apply the developed quantile mapping to future GCM data
Bias Correction/Spatial Disaggregation (BCSD)

- Statistical bias correction of GCM simulations
  - Quantile mapping
- Spatial downscaling to fine scale (i.e. stations)
- Temporal disaggregation from monthly to daily

Source: Wood et al 2006, BAMS
Bias-Correction & Constructed Analogue (BCCA)

I) NEW PATTERN AT COARSE-RESOLUTION:
A new pattern obtained from a coarse resolution source, but the corresponding high-resolution (downscaled) pattern is unknown.

II) FITTING THE ANALOGUE (DIAGNOSIS):
A subset of patterns from a historical library is selected as contributions to a constructed analogue of $Z_{obs}$ based on spatial similarity evaluated at the 2.5 x 2.5 degree resolution.

III) DOWNSCALING THE PATTERN (PROGNOSIS):
A linear combination of the predictor patterns produces a least squares (constructed) analogue of $Z_{obs}$ at 2.5 x 2.5 degree resolution.

Where $A_{analogues 1}$, $A_{analogues 2}$, ..., $A_{analogues n}$ are regression coefficients.

Schematic of the method of constructed analogues for downscaling reanalysis fields from 2.5° x 2.5° grids to 1/8° x 1/8°

Hidalgo et al., 2008
Need to preserve spatial distribution in the future?

- **BCCA**
  - Hard to find similar weather patterns from historical data
  - Not reliable climate projections

Few analogs similar to future spatial structure
Quantile Delta Mapping (QDM)

- Preserving model-projected relative changes in quantiles
- Correcting systematic biases in quantiles

Source: Cannon et al. (2015)
Unintended consequences…

Cannon et al. (2015) J. Climate
**stochastic downscaling approach**

Bias removal

Remind CDF mapping!

Bias correction

**BCSA method**

Bias-Correction & Stochastic Analog

1. Take normal score transform of observed daily precipitation data at each station

\[ x_{i,t}^{*} = F_{obs}^{-1}(x_{i,t}) \]

2. Estimate spatial correlation of normal score transformed data over all stations

\[ \rho_{i,j} = \frac{1}{N} \sum_{t=1}^{N} (x_{i,t}^{*} - \bar{x}^{*})(x_{j,t}^{*} - \bar{x}^{*}) \]

\[ \rho = \begin{bmatrix} \rho_{1,1} & \cdots & \rho_{1,n} \\ \vdots & \ddots & \vdots \\ \rho_{n,1} & \cdots & \rho_{n,n} \end{bmatrix} \]

3. Use the decomposition method to generate an ensemble of normal-score fields that honor observed spatial correlation structure

\[ \mathbf{ \tilde{x} } = \mathbf{ L } \mathbf{ \hat{x} } \]

\[ \tilde{x}^{\mathbf{p}} = \mathbf{ L }^{\mathbf{p}} \]

4. Back transform normal-score fields to generate an ensemble of spatially distributed precipitation fields

\[ \hat{x}_{i,t} = F_{obs}^{-1} \left( F_{norm} (x_{i,t}^{\mathbf{p}}) \right) \]

5. For each daily GCM prediction, randomly select realization from ensemble with spatial mean equal to bias-corrected GCM prediction but observed small-scale spatial structure

*Hwang and Graham, 2013*
Motive for developing BCSA

GCM future information

- Temporal Statistics (Mean, Std., distribution, Auto-correl., etc.)
  - We can reproduce these statistical attributes under assumption of stationarity!

None!

- spatial Statistics (spatial correlation, covariance, etc.)
  - But cannot determine the spatial pattern of precipitation events on daily basis.
  - Generate plausible patterns and use those as scenarios!

None!

- spatial Pattern (Local characteristics)
Schematic representation of the methodology

Method 1: BCSD

- Bias correction
  - CDF mapping

Method 2: SDBC

- Spatial downscaling
  - IDW interpolation

- Bias correction
  - CDF mapping

Method 3: BCSA to generated spatially correlated precipitation field

- 172 sub-basin scale obs.
- Normal score transformation
- Estimate spatial correlation structure of obs.
- Generate random field sequences for 172 stations
- Using correlation matrix of normal score

Gridded dataset: spatial resolution of 1/8° (~12 km)

Method 2 results
SDBC GCMs

Method 1 results
BCSD GCMs

Method 3 results
BCSA GCMs

Evaluate against observation

Library of spatially distributed precipitation fields

Select field from library

Spatially correlated field

Back transformation (CDF mapping)
Bias-Correction & Stochastic Analog Analogue (BCSA)

- Day 1

  - Raw GCM
  - Bias-corrected GCM
  - Observation

  Stochastically generated precipitation fields of which spatial average is equal to the bias-corrected GCM data for a specific day and grid

- Day 2, 3, …: same procedure is repeated
## 4 GCMs (CMIP3) used in the practice

<table>
<thead>
<tr>
<th>Modeling Group, Country</th>
<th>WCRP CMIP3* I.D.</th>
<th>Primary Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bjerknes Centre for Climate Research, Norway</td>
<td>BCCR-BCM2.0</td>
<td>Furevik et al., 2003</td>
</tr>
<tr>
<td>US Dept. of Commerce/NOAA/Geophysical Fluid Dynamics Laboratory, USA</td>
<td>GFDL-CM2.0</td>
<td>Delworth et al., 2006</td>
</tr>
<tr>
<td>Canadian Centre for Climate Modeling &amp; Analysis, Canada</td>
<td>CGCM3.1</td>
<td>Flato and Boer, 2001</td>
</tr>
<tr>
<td>National Center for Atmospheric Research, USA</td>
<td>CCSM</td>
<td>Collins et al., 2006</td>
</tr>
</tbody>
</table>

*WCRP CMIP3: World Climate Research Programme's Coupled Model Inter-comparison Project phase 3*
Raw GCM data vs. bias-corrected GCM results

Comparison of monthly first-order dry to wet (TP_01, upper row) and wet to wet (TP_11, bottom row) transition probabilities for raw GCM data (first column) and bias-corrected GCM results (second column). Averaged transition probabilities for all grids over the study area (i.e., the state of Florida) were plotted for each GCM.
Spatial distribution of the daily mean precipitation

for wet season

for dry season
Spatial distribution of the temporal standard deviation

for wet season

for dry season
Spatial distribution of the $90^{th}$ and $50^{th}$ percentile daily precipitation

**the $90^{th}$ percentile**

**the $50^{th}$ percentile**
Spatial distribution of the frequency of wet spell length (> 5 days) events.

for wet season

for dry season

Units in “number of events/year. ME, RMSE, and R calculated for the downscaled predictions are reported on each map.
Number of the events for given (a) wet (≥0.1mm) and (b) dry (<0.1mm) spell lengths.

**Wet season**

(a) Number of events/year vs. wet spell length

(b) Number of events/year vs. wet spell length

**Dry season**

(a) Number of events/year vs. dry spell length

(b) Number of events/year vs. dry spell length
Daily precipitation Predictions

Averaged mean daily precipitation

Raw GCM data vs. observation

Averaged standard deviations of daily precipitation

Standard deviation of spatially averaged daily precipitation
Transition probabilities

Dry to wet day (P_01, left column) and wet to wet day (P_11, right column)
Spatial Variability of Daily Precipitation

Number of rainy sub-basins

Spatial standard deviations of daily precipitation

Variograms
observed vs. simulated mean daily spatial correlation indices

(a) Moran’s I and spatial variance indices (b) Geary’s C for each month. 4 GCMs are not separately represented but are indicated by the same marker for each downscaling method.

\[ I_r = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_{r,i} - \bar{x}_r) (x_{r,j} - \bar{x}_r)}{\sum_i (x_{r,i} - \bar{x}_r)^2} \]

\[ C_r = \frac{(N - 1)}{2 \sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_{r,i} - x_{r,j})^2}{\sum_i (x_{r,i} - \bar{x}_r)^2} \]
Hydrologic implication of BCSA
Integrated Hydrologic Model

- TBW and SWFWMD commissioned the development and application of an integrated surface water/groundwater model to gain an increased understanding of the surface and groundwater flow systems in the Tampa Bay Region.
- The Integrated Hydrologic Model (IHM) was developed which integrates the EPA Hydrologic Simulation Program–Fortran for surface–water modeling with the US Geological Survey MODFLOW96 for groundwater modeling.

Ross et al., 2004 (IHM theory manual)
Study domain

<table>
<thead>
<tr>
<th>Name (data source)</th>
<th>Watershed</th>
<th>Lat.</th>
<th>Lon.</th>
<th>Drainage area, (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alafia River at Lithia</td>
<td>Alafia</td>
<td>27.8719</td>
<td>-82.2114</td>
<td>867.3</td>
</tr>
<tr>
<td>Hillsborough River near Zephyrhills</td>
<td>Hillsborough</td>
<td>28.1497</td>
<td>-82.2325</td>
<td>569.6</td>
</tr>
<tr>
<td>Cypress Creek at Worthington Gardens</td>
<td>Hillsborough</td>
<td>28.1856</td>
<td>-82.4008</td>
<td>302.9</td>
</tr>
<tr>
<td>Ancloete River near Elfers</td>
<td>Ancloete</td>
<td>28.2139</td>
<td>-82.6667</td>
<td>187.7</td>
</tr>
<tr>
<td>CYC TMR-5 SH (SWFWMD)</td>
<td>Hillsborough</td>
<td>28.2057</td>
<td>-82.4680</td>
<td></td>
</tr>
<tr>
<td>S21-J26As (TBW)</td>
<td>Ancloete</td>
<td>28.0730</td>
<td>-82.5800</td>
<td></td>
</tr>
<tr>
<td>STK-Starkey-26s (TBW)</td>
<td>Ancloete</td>
<td>28.1956</td>
<td>-82.6953</td>
<td></td>
</tr>
<tr>
<td>CBR-SERW-s (TBW)</td>
<td>Springs coast</td>
<td>28.3151</td>
<td>-82.5146</td>
<td></td>
</tr>
<tr>
<td>CYC TMR-5d (TBW)</td>
<td>Hillsborough</td>
<td>28.2053</td>
<td>-82.4680</td>
<td></td>
</tr>
<tr>
<td>S21-Jackson26d (TBW)</td>
<td>Ancloete</td>
<td>28.0730</td>
<td>-82.5800</td>
<td></td>
</tr>
<tr>
<td>STWF 10 DP (SWFWMD)</td>
<td>Ancloete</td>
<td>28.1963</td>
<td>-82.6951</td>
<td></td>
</tr>
<tr>
<td>CBR-SERW-d (TBW)</td>
<td>Springs coast</td>
<td>28.3151</td>
<td>-82.5146</td>
<td></td>
</tr>
<tr>
<td>Masaryktown DP (SWFWMD)*</td>
<td>Springs coast</td>
<td>28.3740</td>
<td>-82.5262</td>
<td></td>
</tr>
<tr>
<td>Lithia spring</td>
<td>Alafia</td>
<td>27.5159</td>
<td>-82.1353</td>
<td></td>
</tr>
<tr>
<td>Weeki Wachee spring</td>
<td>Springs coast</td>
<td>28.3048</td>
<td>-82.4680</td>
<td></td>
</tr>
</tbody>
</table>

*: Unconfined Floridan aquifer monitoring well. Corresponding surficial aquifer, therefore, does not exist.
The BCSD method tends to underestimate streamflow for the wet season more than the SDBC and BCSA methods at all stations due to the high spatial correlation of the BCSD daily precipitation fields and higher frequency of low precipitation events resulting in higher evapotranspiration.

The monthly average streamflow predicted by the SDBC results are reasonably close to calibrated results and similar to the BCSA results.
Mean of errors in streamflow simulation

Mean of errors (simulated-calibrated) of monthly average streamflow over four GCM results for each target station.
Mean of errors (simulated–calibrated) of temporal standard deviation of daily streamflow over four GCM results for each target station

Mean of errors in streamflow simulation

Alafia River

Mean of errors (simulated–calibrated) of temporal standard deviation of daily streamflow over four GCM results for each target station
• The frequency of daily streamflow was not reproduced by SDBC as closely as for the BCSA results.
• Comparing the , it is evident that the under/overestimations of extreme events is canceled out when evaluating only the monthly mean streamflow.
No substantial difference of skill in reproducing the frequency of groundwater level was found among the various downscaling techniques.

The SDBC showed some overestimation of the frequency of lower and higher groundwater level at the CBR-SERW and S21 stations compared to the calibrated results.

This may be due to the same mechanism that resulted in overestimating the frequency of extreme streamflow events and peak flow in the wet season by the SDBC.
### Summary

<table>
<thead>
<tr>
<th>BCSD</th>
<th>SDBC</th>
<th>BCSA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precipitation prediction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Annual mean cycle</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>- Temporal mean</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>- Transition probability</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>- Frequency of events</td>
<td>Δ</td>
<td>X</td>
</tr>
<tr>
<td>- Wet &amp; dry spell length</td>
<td>X</td>
<td>Δ</td>
</tr>
<tr>
<td>- Temporal variance</td>
<td>Δ</td>
<td>Δ</td>
</tr>
<tr>
<td>- Spatial variability</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td><strong>Hydrologic simulation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Annual mean cycle</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>- Temporal variability</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>- Frequency of event</td>
<td>O</td>
<td>X</td>
</tr>
</tbody>
</table>

cf. SQM, SDQDM
Drawbacks of BCSA

- Conducted independently
  - on a daily basis and
  - At each GCM grid

- Stationary covariance

- Advantages:
  - Can downscale into any temporal (hourly, daily, monthly) and spatial scale (e.g., gridded, irregularly distributed points)
  - Generate Ensemble of possible local scale precipitation patterns
  - Uncertainty due to the downscaling process could be examined using collection of equally probably downscaled climate fields
Evaluation and selection of Downscaling methods
Screening the method to use for application?

- **GCMs**
  - Future climate info. (AR5, RCP scenario)
  - PCMDI (CMIP5)

**GCM evaluation**
- not OK
- OK

**Screening**
- OK

**Screening GCM**
- Physical probability
- Global scale
- Climatology
- Purpose:

**Model weighting**

---

**Downscaling methodologies**
- Spatially detailed climate info.
  - BCSD, BCCA, SDBC, BCSA, etc.

**Method evaluation**
- not OK
- OK

**Selecting**

**Method selection**
- OK

**intercomparison**

- Statistical estimation
- Regional/local scale
- Extreme, diurnal cycle
- Purpose:

---

**vs.**
GCM skill evaluation

Different observations
1 Maurer’s data (Maurer et al., 2002)
2 NLDAS data
3 USGS data
4 Reanalysis data?

Different metrics
1 RMSE
2 Correlation
3 Others?

Different resolutions
1 each GCM scale
2 0.5’X0.5’
3 1.0’X1.0’
4 etc.

Different indicators
1 d10TS’
2 monthlyTS’
3 4seasonalTS’
4 annualTS’
5 annualCVTS’
6 monthlyQ95’
7 monthlyQ99’
8 monthlymean’
9 monthlyCV_daily’
10 d10mean’
11 monthlymean’
12 monthlyCV_monthly’
13 monthlyCV_monthly’
14 d10std’
15 monthlystd_dailydata’
16 monthlystd_monthlydata’:

Distribution of weights for each GCM (precipitation)

How are the results for other variables? using other observation data as reference? at other spatial resolution?

What if we have different inferences for future change?

Reliable Future information?

UNCERTAINTY
How GCMs work?

Performance evaluation at different resolution with different observation
For different criteria
(Masson and Knutti, 2011) Hierarchical clustering of the CMIP3 models for (left) surface temperature and (right) precipitation in the model control state. Models from the same institution and models sharing versions of the same atmospheric model are shown in the same color. Observations also are marked by the same color. Models without obvious relationships are shown in black.
Messages to take home

Statistical downscaling

Matters to be attended…

- Why downscale?... Overdoing downscaling?

- Specify the study objectives
  - For mean values, the simpler statistical downscaling methods perform comparably to the more sophisticated methods
  - For extreme values, the more sophisticated methods (e.g., quantile mapping) are needed
    → the more sophisticated the method, the smaller the future change projected for the warmest extreme temperatures in the U.S.

- Select technique
  - The more sophisticated, the better?

- What to evaluate
  - downscaling purpose vs. downscaled results

- How to evaluate
  - Evaluations may vary by applying different observations, resolutions, index, matrix, etc.
  - a danger of getting the right result for the wrong reason by tuning the wrong end of the model
Messages to take home

Idea
• Probabilistic approach with statistically downscaled results
  ⇒ Stochastic concept
• Quantifying uncertainty of processes ⇒ credible impact assessment

Downscaling… Added value?
• Once we have more regional detail, what difference does it make in any given impacts assessment?
• What is the added value?
• Do we have more confidence in the more detailed results?

Issues
› GCM boundary conditions are a main source of uncertainty for most downscaling techniques
› Different downscaling methods can yield different scenarios even when forced with the same GCM
› Ability to downscale the current climate does not guarantee accuracy about downscaling the future

Inter-comparison of statistical downscaling techniques
Comparison of statistical & dynamical downscaling (e.g., CORDEX)
Thank you!
For more details, swhwang@gnu.ac.kr

Keeping pace with a changing environment for the future!!!