

A Statistical Model for Climate Downscaling with Uncertainty Quantification for Engineering Infrastructure Design and Adaptation

Ernst Linder (UNH) elinder@unh.edu

Collaborators:

- UNH: Meng Zhao, Yiming Liu, Jennifer Jacobs,
- Texas Tech: Anne Stoner



## Climate Change and Infrastructure Design





# Downscaling (ESD) from Climate Model Output to a Weather Scenario

Consider only one climate variable

(not a weather generator)

- Assume temporal pattern trend, seasonal (annual) cycles, short term dependence
- Method based on quantile-(quantile) matching

(or *pattern matching*, or *distribution translation*)

- Known as *Probabilistic Downscaling* or *Empirical Statistical Downscaling (ESD)*: The probability distribution (CDF) of climate is shifted/translated to that of local weather But: No attempt at trying to match actual weather: *Asynchronicity* (See Benestad et al. 2013)
- Develop method for Uncertainty Quantification: Confidence Intervals for Infrastructure "Endpoint"
- Uncertainty due to Downscaling only. Assume all climate model information as given.



# ESD of a Temporally Patterned Climate Variable

- $X_C$  current model output time series (or historic / baseline)
- $Y_C$  current **weather** time series
- $X_P$  projected (future) **model** output
- *Y<sub>P</sub>* projected **weather** scenario; *unknown*
- Assume temporal pattern (eg daily temperature) has trend, seasonal (annual) cycles, short term dependence



ESD: Marginal Distribution Translation via q-q matching

Matching Quantiles:

$$x^* \sim F_X(x) \text{ and } y^* \sim F_Y(y) \text{ if } F_Y(y^*) = F_X(x^*) \qquad F(); \text{ cdf}$$
  
then:  $y^* = T(x^*) = F_Y^{-1}(F_X(x^*))$  (1)

(Panovsky and Brier, 1958)

Downscaling: (1) Match empirical quantiles of X<sub>c</sub> and Y<sub>c</sub>
 (2) Fit a curve or interpolate plus interpolation(\*)

Note: - this permutes the series: asynchronicity - corrupts temporal dependence

(2) Then use values of  $X_P$  to predict  $y_P = T(x_P)$ 

• **Corruption of temporal structure more serious** than suboptimal choice of interpolation *Empirical quantiles of the marginal distribution are not representative of the F() for a structured time series* 

(\*) Variations: linear regression , piecewise linear, spline regression, kernel convolution, analog method... Stoner et al. 2012, McGinnis et al. Michelangeli et al. Dixon et al. 2016

(\*) Delta method is a special case



# • Empirical quantiles of the marginal distribution are not representative of the F( ) for a structured time series

Two Series: Same (marginal) mean and standard deviation (left: random, right: AR(2)









### Detour: Two possible ways to predict future scenario

- 1 Model to Station Translation (MST): (this one is used more often) Match quantiles of X<sub>c</sub> and Y<sub>c</sub>, then apply the translation to future X<sub>p</sub>. So-called "stationarity" assumption: Translation remains valid when going from present to future.
- Current to Future Translation (CFT): Match the quantiles of X<sub>c</sub> and X<sub>p</sub> (clim. model output) then apply the translation to Y<sub>c</sub> (station weather).
  Here "stationarity"; translation (from current to future) is valid when going from model output to station data. Better term would be: "(model-to-station) consistency"
- Note: CDFt (Michelangeli et al., 2009) and xCDFt (Kallache et al. 2011) is based on CFT
  - only very specialized situations give same result for MST and CFT

We will only discuss MST

 $\begin{array}{ll} (X_C,Y_C) \ \text{current (or historic / baseline)} & (model output, weather) \\ X_P & \text{projected (future) model output} \\ Y_P & \text{projected weather; (unknown)} \end{array}$ 



Multivariate (MV) Distribution for Structured Time Series

• q-q matching DS for scale-location marginal distribution (eg normal, t, Gumbel):

$$y^* = \left[ T(x^*) = \mu_Y + \sigma_Y \Phi_Y^{-1} \left( \Phi_X \left( \frac{(x^* - \mu_X)}{\sigma_X} \right) \right) = \mu_Y + \sigma_Y \left( \frac{(x^* - \mu_X)}{\sigma_X} \right);$$

 $\Phi$  is stand. cdf

• MV cdf distr. of time series:  $\mathbf{x}_{C} = (\mathbf{x}_{C,1}, \mathbf{x}_{C,2}, \dots, \mathbf{x}_{C,T}) : F_{\mathbf{x}C}(\mathbf{x}_{C})$  etc.

• Correct downscaling: 
$$\boldsymbol{y}_p = \boldsymbol{\mu}_{Yc} + \Sigma_{Yc}^{1/2} \boldsymbol{\Phi}_{Yc}^{-1} \left( \boldsymbol{\Phi}_{Xc} \left( \Sigma_{Xc}^{-1/2} (\boldsymbol{x}_p - \boldsymbol{\mu}_{Xc}) \right) \right)$$
, (2)

where  $\mu$  is the fitted mean (series), and  $\Sigma$  is the variance covariance matrix



# Downscaling (ESD) and Time Series Modeling

• TS model fitting:

Trend: cubic spline with 3 knots (defines mean) $R^2 \sim 85\%$ Seasonal cycles: cubic spline with 8 knots (defines mean) $R^2 \sim 85\%$ Time dependence: ARMA Model autoreg. order p=3 (AR(3)) (def cov) $+ \sim 8\%$ Noise (error) distribution: Not specified for DS

• Noise variable: independent realizations: DS via QQ matching possible

• Equation again: 
$$y_p = \mu_{Yc} + \Sigma_{Yc}^{1/2} \Phi_{Yc}^{-1} \left( \Phi_{Xc} \left( \Sigma_{Xc}^{-1/2} (x_p - \mu_{Xc}) \right) \right)$$
  
3 2 1

- 1: Standardized residual operation: mean subtraction and "whitening'; Model for  $X_C$  applied to  $x_P$
- 2: Downscale standardized residuals using empirical q-q matching of  $e_X_C$  and  $e_Y_C$
- 3: "Coloring" and mean structure adding; Model of  $Y_C$



# Whitening and Coloring

- AR and ARMA time series: Whitening (removing the dependence structure) requires filtering (i.e. a sequential calculation)
- Coloring (adding dependence structure) requires inverse filtering (TS simulation)
- Noise variable: independent realizations: *DS via QQ matching possible*
- Available in software (eg. R)
- Steps:
- a) Whitening with model for  $X_C$ ;
  - 1. Subtract the mean structure  $\mu_{Xc}$  of  $X_C$  (lin. Regression residuals)
  - 2. Whiten the ARMA model for the regression residuals of 1.
- b) Coloring a standardized residual with model for  $Y_C$ 
  - 1. Simulate the ARMA model using the given residuals as noise (innovation)
  - 2. Add the mean structure  $\mu_{Yc}$  to the result of 1.



### Uncertainty Quantification via Bootstrapping

- Linear Standard Error approximations not feasible
- Use Bootstrap (resampling/simulation) based calculations for standard error and confidence intervals.
- Bootstrapping time series: nonparametrically using block resampling (*difficult* to implement with annual cycles)
- For simple AR models (no trend, cycles): Can resample residuals (Efron, Politis et al ..)
- Best for us: Parametrically sample from limiting normal distribution (\*) of fitted parameters, then simulate series with randomly permuted residuals. (A hybrid parametric/nonparametric bootstrap)

(\*)Limiting normal distributions are correct for the long daily series of 30 – 50 yrs



#### Freeze-Thaw Cycle: Winter Weight Premium - Spring Load Restriction (see Jacobs et al.)

#### A Prototype Freeze-Thaw Model that only depends on daily temperature

 $FI_i = T_{ref} - T_{avg,i}$  $CFI_n = \sum_{i=1}^n FI_i$ 

FI, CFI = daily & cumulative freezing index TI, CTI = daily & cumulative thawing index  $T_{ava,i}$  = average air temperature  $T_{ref}$  = reference temperature (often taken as 32 °F, but may vary)  $TI_i = T_{avg,i} - T_{ref}$  $CTI_n = \sum_{n=1}^{n} (Daily Thawing Index - 0.5 \times Daily Freezing Index)$ 

Simplified Freeze-Thaw Model that only depends on average daily temperature

MnDOT (2009):

- Apply **WWP** When CFI > 280°F days
- Apply SLR When CTI > 25°F days



## Some Results (using CCSM4)

#### Freeze-Thaw: Madison ME: Observed WWP SLR Apr 1 Apr 1 Mar Mar 1 Feb 1 Feb 1 Jan ' Jan 1 Dec 1 Dec 1 1960 1970 1980 1990 2000 2010 1950

#### Future Downwscaled (RCP45) with 95% Conf.Band WWP SLR Apr 1 Apr 1 Mar 1 Mar 1 Feb 1 Feb 1 Jan 1 Jan 1 Dec 1 Dec 1 2050 2060 2080 2090 2040 2070 2100

#### Temperature Annual Means Station DS



#### Using simple q-q matching method



#### Freeze-Thaw: Madison ME: Observed



#### Future Downwscaled (RCP85) with 95% Conf.Band





## Summary

- We have applied time series modeling and analysis to empirical statistical downscaling (ESD) for studies where temporal (short term) structure is important and needs to be translated correctly from climate model to local weather scenarios.
- This enables the **coupling** of process **models** for infrastructure design (eg bridges, pavements, road maintenance, etc.) with climate model outputs.
- Either GCM's (global climate models) or RCM's can be used in this framework.
- We applied a hybrid Bootstrap for standard error and confidence interval calculations. It allows to correctly propagate statistical uncertainty measures through the downscaling calculations and additional process model steps.
- The bootstrapping limits the use of very computationally intensive coupled process models somewhat.



#### Next steps

- Multiple climate model outputs: Downscale first, then combine estimates.
- Could the framework be extended to multiple climate variables combine a weather generator model with this ESD model, which then can be coupled to infrastructure design (or impact) models.
- More realistic impact process model: add its uncertainty. How? (model calibration, unified statistical framework, Bayesian meta analysis...)



Ernst Linder Dept. of Math. & Statistics U. of New Hampshire elinder@unh.edu

Thanks!

