A multisite stochastic precipitation generator using truncated and power-transformed normal distribution

Oleg Makhnin
Dept. of Mathematics
New Mexico Tech

October 17, 2008

Abstract

Stochastic precipitation generators produce time series of precipitation in a region, of arbitrary length. However, the modeling process should account for a large quantity of zero values. One possible approach is presented, when we model the precipitation occurrence and amounts simultaneously, via a truncated normal distribution. This approach also allows us to fit correlations and lag-1 cross-correlations between the study sites.

Intro

Precipitation data

- Multiple variables are observed R_{jt}, sites (gauges) $j = 1, ..., N$, time $t = 1, ..., T$.
- A lot of observed values are 0, others are rounded to 0.01 inch.
- Have to deal with dependencies across time, as long as dependencies between the sites.

Use of Normal distribution

- Typical picture of the precipitation values:
  [includegraphics]
- Typical picture of the Normal distribution:
  [includegraphics]

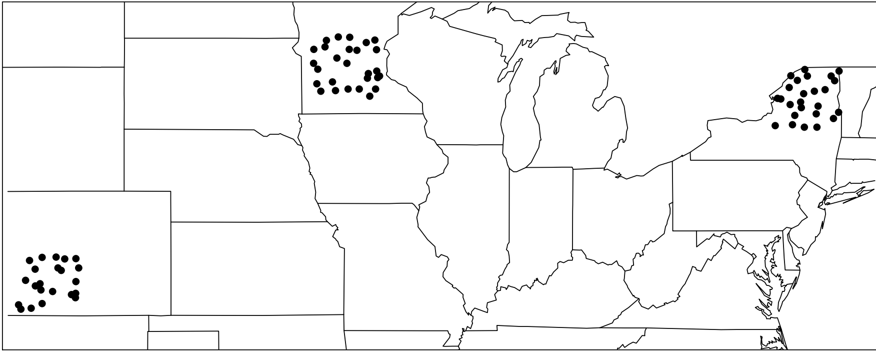
TPT: truncated and power-transformed

Let $W=$ normally distributed *precipitation potential*

- Truncated: consider the negative values of W to lead to zero precipitation
- Power-transformed: the positive values of W lead to precipitation $R = W^{1/\beta}$
- Lets you fit both the *occurrence* and *amount* of precipitation. Parameters: μ=mean, σ = st.dev., β.

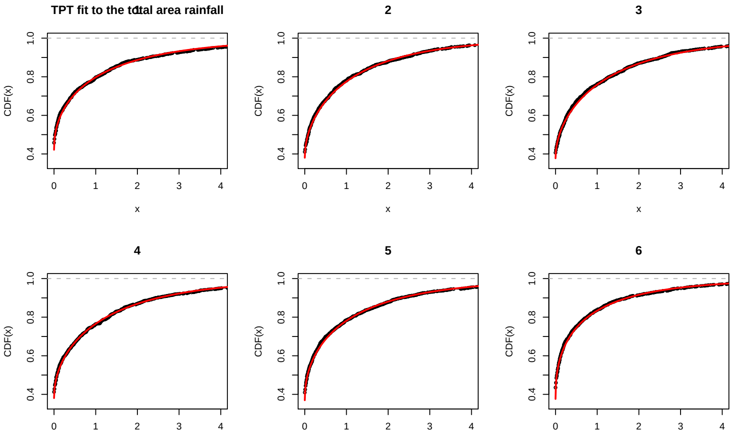
Pros: Normal distribution easy to work with (matrices)
Cons: "Missing data" whenever $R < 0$

Study areas

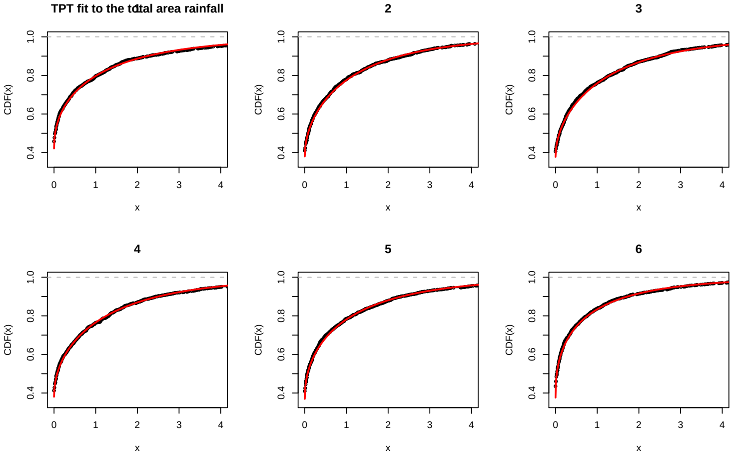


Some fit results

Plotting the CDF (cumulative distribution function = percent of observations below a value): CO daily totals, months 1-6



CO daily per station, August !!change graph !!

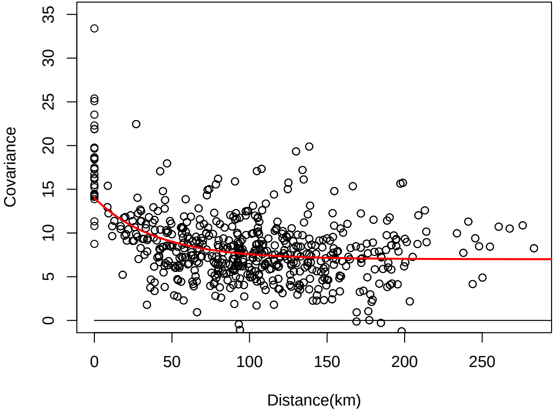


Some fit results

- Fits fairly well for each station separately, also for total of all stations,
- Estimate precipitation probability as $pr = P(W > 0)$
  mean positive: $pr > 0.5$
  mean negative: $pr < 0.5$
- Values of beta between 1.5 and 3.

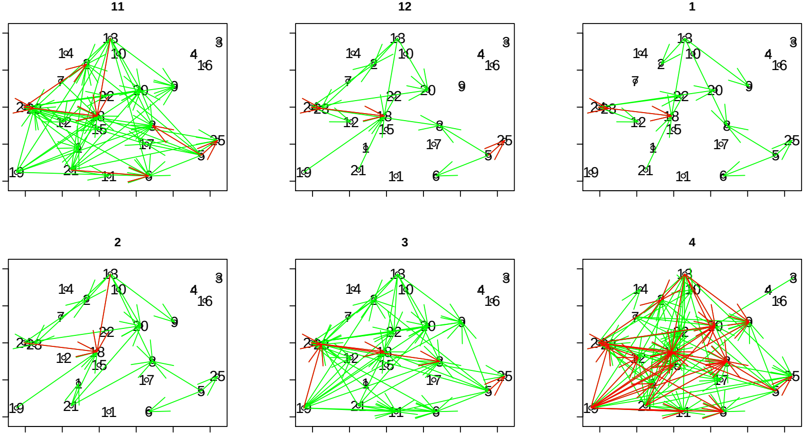
Spatial dependency

Typically, if we plot covariance of precipitation measurements versus distance we observe something like this

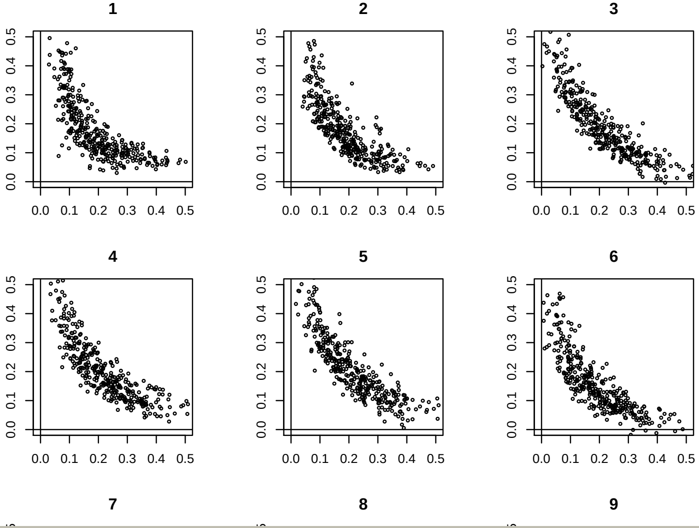


Temporal dependency

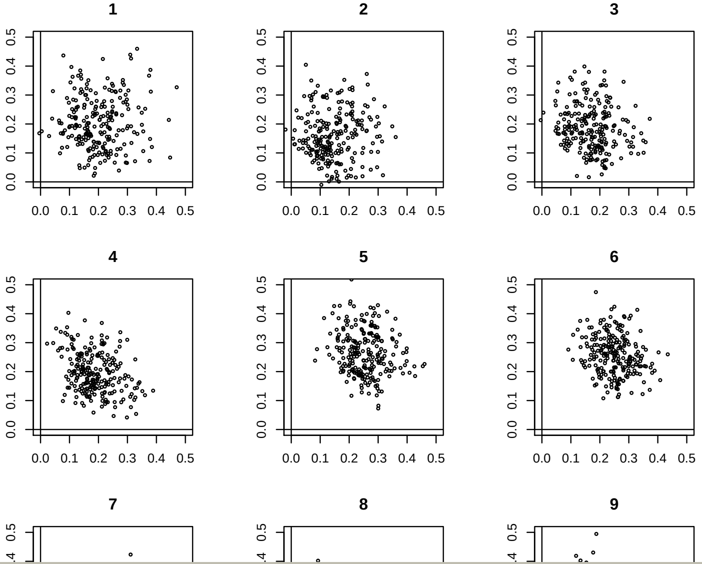
We might also be concerned with persistency in time:
correlation (or covariance) of R_{jt} with $R_{k,\mathbf{t+1}}$ – between sites j, k.



Strange but persistent patterns!
Even more strange: "correlation inversion" for MN area



But: no correlation inversion for the CO study area:



Autoregression models

Scalar autoregression model

persistency at a single site by using

$$W_{t+1} - mean = r(W_t - mean) + error$$

Values of r close to 1: very persistent climate,
Values of r close to 0: "white noise"

Multivariate autoregression (MAR)

combine all sites in a region, vector $\mathbf{W}_t$:

$$\mathbf{W}_{t+1} - mean = \mathbf{r}(\mathbf{W}_t - mean) + \boldsymbol{\xi}_t$$

Matrix-valued **r**, with eigenvalues within unit circle to insure stationarity.

Multivariate AR

Goal: mimic the lag-1 behavior we observed earlier
Correction:

$$\begin{cases} \mathbf{Z}_t = \mathbf{r}\mathbf{Z}_{t-1} + \boldsymbol{\xi}_t \\ \mathbf{W}_t = \mathbf{Z}_t + \mathbf{S} + \boldsymbol{\varepsilon}_t, \end{cases} \tag{1}$$

S: station effects (means)
$\boldsymbol{\xi}$ are correlated, variance matrix $\mathbf{V}_\xi$
$\boldsymbol{\varepsilon}$ are white noise

Multivariate AR

Number of parameters describing $\mathbf{r}$ and $\mathbf{V}_\xi$ is $N^2 + N(N+1)/2$.
On the other hand:
"lag-0" autocovariance of $\mathbf{Z}$, $\hat{\Sigma}_0$ has $N(N+1)/2$
and lag-1 autocovariance $\hat{\Sigma}_1$ has N^2 parameters

Thus, if $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ were known, we would get:

$$\mathbf{r}^T = \hat{\Sigma}_0^{-1}\hat{\Sigma}_1, \quad \mathbf{V}_\xi = \hat{\Sigma}_0 - \mathbf{r}\hat{\Sigma}_0\mathbf{r}^T \tag{2}$$

Fitting the model

Use Markov Chain Monte Carlo, with Gibbs sampler

- Things to estimate: $\mathbf{r}$, $\mathbf{V}_\xi$, Station effects, β, W_{jt} for all j, t ... etc ... etc ...
- Rename them $\{X_1, X_2, X_3, ..., X_d\}$

  Main idea: draw a parameter, say X_1 from its *full conditional posterior* $p(X_1 \,|\, X_2, X_3, ..., X_d)$, then draw
  X_2 from $p(X_2 \,|\, X_1, X_3, ..., X_d)$
  ....
  X_d from $p(X_d \,|\, X_1, X_2, ..., X_{d-1})$
  then repeat this procedure over and over again.
- After convergence, get samples from all the parameters to find the means, credible intervals etc.

QUESTIONS?

THANK YOU!

`www.nmt.edu/~olegm/talks/Multivar/`

for pdf file