A multisite stochastic precipitation generator based on switching weather states

Oleg Makhnin
Dept. of Mathematics
New Mexico Tech

April 29, 2011

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.

Precipitation generators

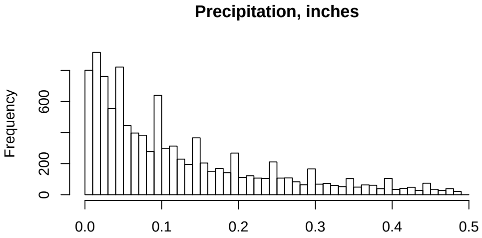
- • Get long series of synthetic precipitation
- • Statistical properties similar to real precipitation
- • Used in hydrologic modeling (water resources, floods, droughts...), agriculture, forestry etc.
- • Get some idea of precipitation variability from year to year.
- • Climate change?

Intro

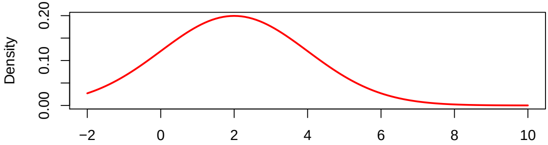
Precipitation data

- 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.
- • Most generators have a separate (discrete) process, modeling occurrence and another one (continuous) modeling amounts. Hard to make them work together.

- Typical distribution of the precipitation values:



- Typical picture of the Normal distribution:



TPT: truncated and power-transformed

Let $W=$ normally distributed *precipitation potential*

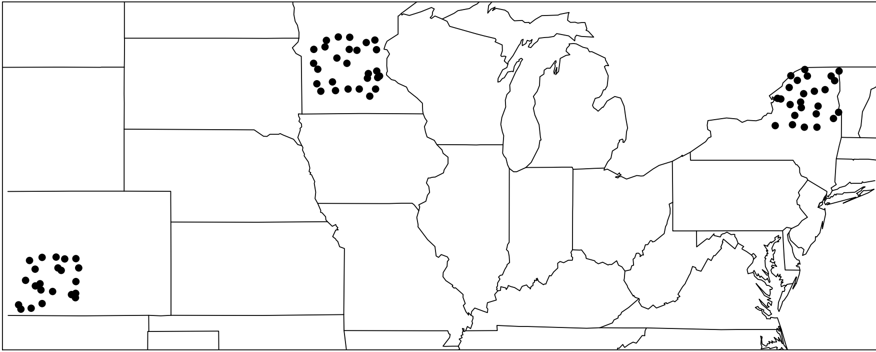
- Truncated: consider the negative values of W to lead to zero precipitation
- Power-transformed: the positive values of W lead to precipitation $R = W^{\beta}$
- Lets you fit both the *occurrence* and *amount* of precipitation. Parameters: μ=mean, σ = st.dev., β = shape

Pros: Multivariate Normal distribution is easy to work with (matrices)

Cons: "Missing data" whenever $R < 0$

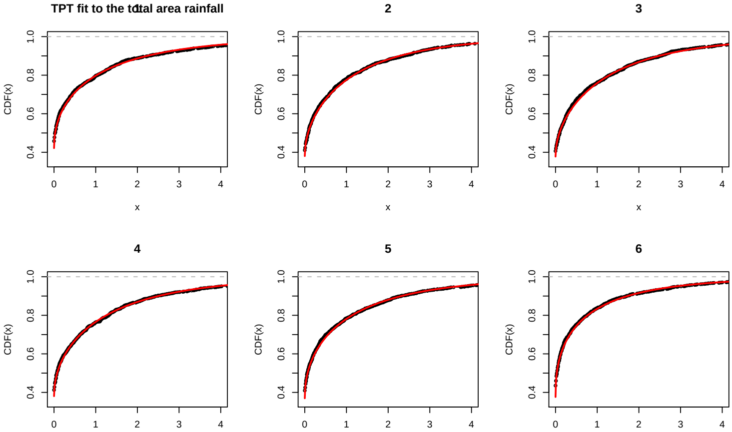
- Used by Bardossy et al, Sanso et al, ...

Study areas



Some fit results

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



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.

Seems a good fit for the static model.

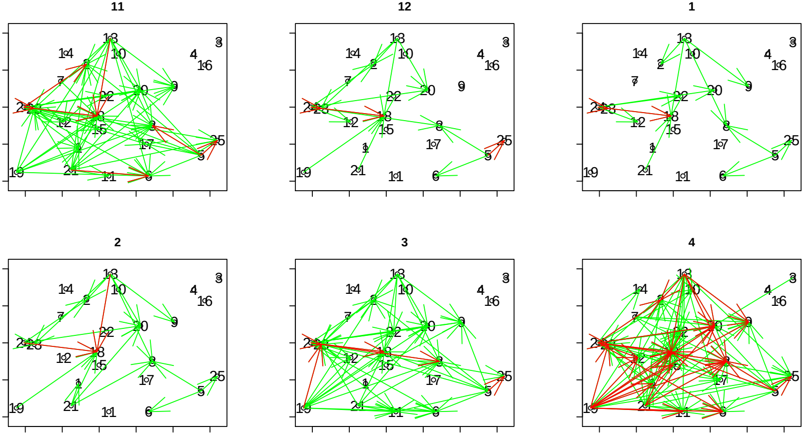
But our model is dynamic: changes with time.

Spatial dependency

Typically, covariance of precipitation measurements decreases with distance.

Temporal dependency

Persistency in time: lag-1 correlation (or covariance) of R_{jt} with $R_{k,\mathbf{t+1}}$ – between sites j, k.



Autoregression models

Scalar autoregression model

persistency at a single site by using

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

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 $\mathbf{r}$, with eigenvalues within unit circle to insure stationarity.

Multivariate AR

Goal: reproduce the lag-1 behavior

Correction: "weather states" $\mathbf{Z}_t$

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

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

Multivariate AR

$$\mathbf{Z}_t = \mathbf{r}\mathbf{Z}_{t-1} + \boldsymbol{\xi}_t$$

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 $N(N+1)/2$ (symmetric matrix)
and lag-1 autocovariance $\hat{\Sigma}_1$ has N^2 parameters

Thus, if $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ were known (estimated from $\mathbf{Z}$'s), 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}, Z_{jt} for all j, t ... ...
- Rename them $\{X_1, X_2, X_3, ..., X_d\}$

  Goal: get $p(X_1, X_2, X_3, ..., X_d \,|\, \mathbf{data})$ - hard!

- Main idea: draw a parameter, say X_1 from its *full conditional posterior* $p(X_1 \,|\, X_2, X_3, ..., X_d, \mathbf{data})$, then draw
- X_2 from $p(X_2 \,|\, X_1, X_3, ..., X_d, \mathbf{data})$
- ....

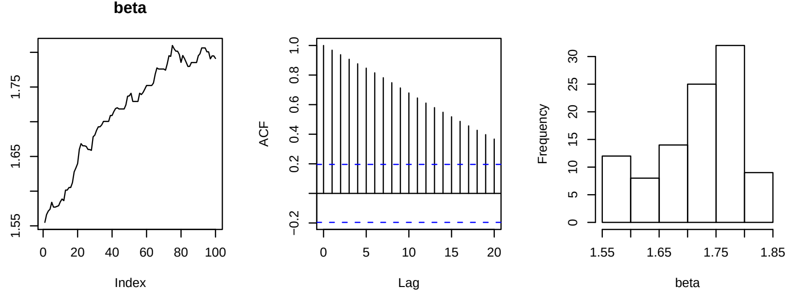
  X_d from $p(X_d \,|\, X_1, X_2, ..., X_{d-1})$

  then repeat this procedure over and over again.

Gibbs sampler

After convergence, get samples from all the parameters to find the means, credible intervals etc.

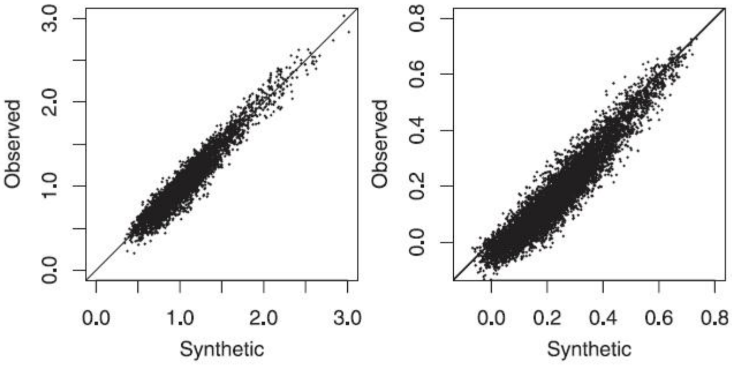
Example: output for β



Gibbs sampler

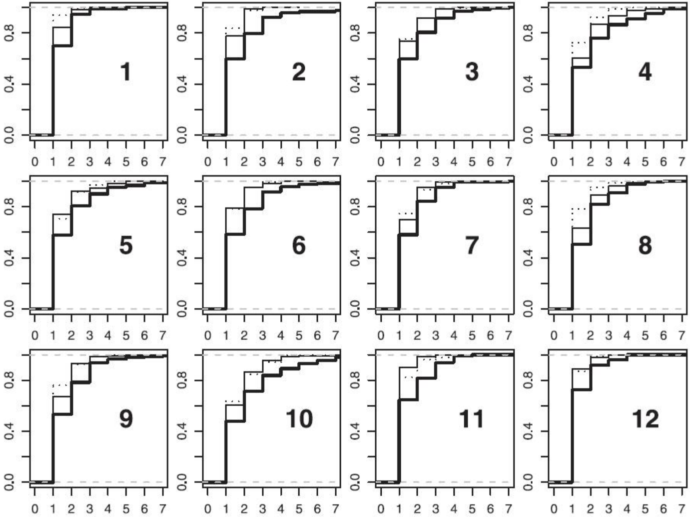
- Advantages:
  modular, allows to fully assess variability
  nice for "missing data" problems
- Computationally intensive!

Some results



Comparing observed and synthetic. Left : covariances, right: lag 1 cross-covariances between $\sqrt{precip}$ values

Not so great: area-wide dry spell lengths are consistently underestimated



Conclusion

- Models complex correlation structure between rainfall observations
- Allows for wide interannual variability
- May require tweaking for some data

Extensions:

- Weather typing: dry/wet weather, fronts moving etc.
- From sites to the whole area (gridded)

Current work: switching weather states

(HMM-like)

Make the parameters in the equation $\mathbf{Z}_t = \mathbf{r}\mathbf{Z}_{t-1} + \mathbf{S} + \boldsymbol{\xi}_t$ depend on the latent state G_t, with a finite possible number of states = "weather regimes". For example, dry or wet.

Region-wide latent states would switch according to a Markov Chain.

Simplest version: $\mathbf{Z}_t = \mathbf{r}\mathbf{Z}_{t-1} + \mathbf{S} + \boldsymbol{\mu}_{G(t)} + \boldsymbol{\xi}_t$

Work in progress

QUESTIONS?

THANK YOU!

`www.nmt.edu/~olegm/talks/TPT2/` for pdf file