STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 1

(1) Environmental Motivation

(2) Probabilistic Background

(3) Statistical Background

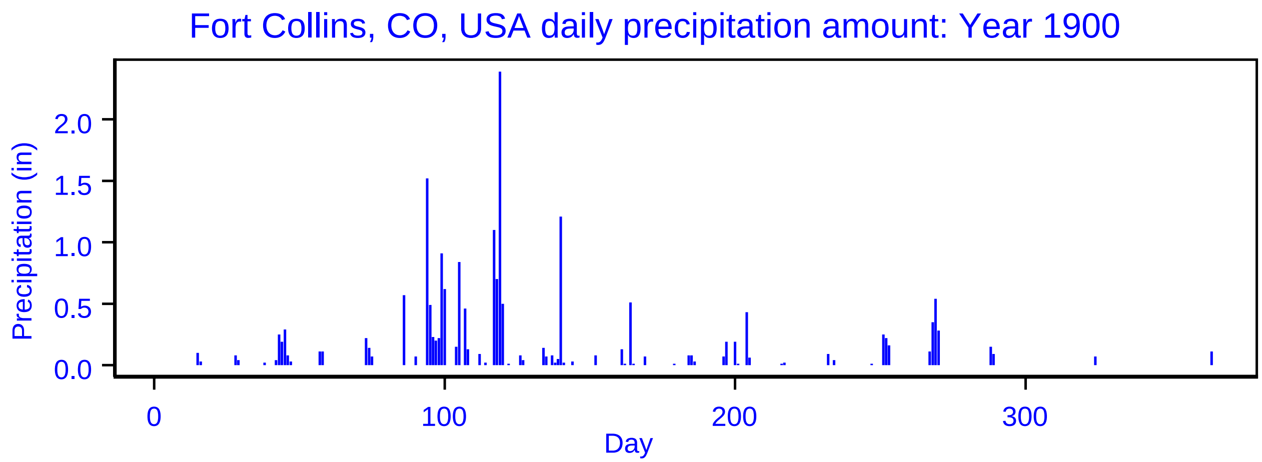
(1) Environmental Motivation

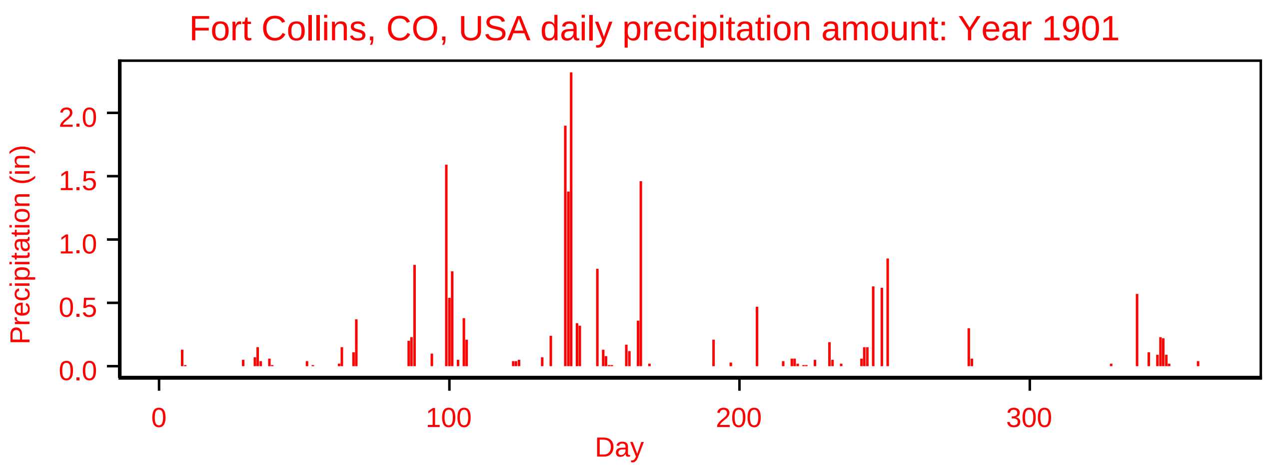
- **Example**

-- Time series of daily precipitation amount at single location (Fort Collins, CO, USA: 1900 – 1999)

- **Statistical Characteristics of Environmental Time Series**

- **Scientific Research Questions**


Fort Collins, CO, USA daily precipitation amount: Year 1900
Precipitation (in)
0.0
0.5
1.0
1.5
2.0
0
100
200
300
Day


Fort Collins, CO, USA daily precipitation amount: Year 1901
Precipitation (in)
0.0
0.5
1.0
1.5
2.0
0
100
200
300
Day

Statistical Characteristics

- **Intermittency**
- **Skewness**
- **Extremes (heavy tail?)**
- **Temporal dependence**
- **Seasonality**
- **Spatial dependence (not shown)**
- **Trend?**

Scientific Research Questions

Overdispersion

-- Stochastic models for daily precipitation tend to underestimate variance of monthly, seasonal, or annual total precipitation

Explanations

- **High Frequency Variations**

-- Inadequate stochastic model for daily precipitation?

- **Low Frequency Variations**

-- "Potential predictability" or climate change?

(2) Probabilistic Background

Preliminaries

-- To be used in statistical modeling of overdispersion phenomenon

- Scaling/Aggregation

- Conditional Moments

- Effects of Intermittency

Scaling/Aggregation

- **Stationary stochastic process** $\{X_t : t = 1, 2, \ldots\}$

-- Moments

$$\mu = \mathrm{E}(X_t), \qquad \sigma^2 = \mathrm{Var}(X_t), \qquad \rho_l = \mathrm{Corr}(X_t, X_{t+l}), \; l = 1, 2, \ldots$$

- **Sum** $S_T = X_1 + X_2 + \cdots + X_T$

-- Variance

$$(1/T)\,\mathrm{Var}(S_T) \longrightarrow \sigma^2\,[1 + 2(\rho_1 + \rho_2 + \ldots)] \equiv (\sigma^*)^2 \text{ as } T \longrightarrow \infty$$

-- Distribution $(S_T - T\mu) / (\sigma^* T^{1/2}) \longrightarrow N(0, 1)$ (in distribution)

- **Example: First-order autoregressive [AR(1)] process**

$$X_{t+1} - \mu = \varphi(X_t - \mu) + \varepsilon_t\,, \quad t = 1, 2, \ldots$$

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{E}(\varepsilon_t\, \varepsilon_{t+l}) = 0, \quad l = 1, 2, \ldots$$

Here $\mu = \mathrm{E}(X_t)$, $\sigma^2 = \mathrm{Var}(X_t)$, $\varphi = \mathrm{Corr}(X_t\,, X_{t+1})$

-- Variance of sum S_T

$$(1/T)\ \mathrm{Var}(S_T) \longrightarrow (\sigma^*)^2 = \sigma^2\ [(1 + \varphi)/[(1 - \varphi)] \ \text{ as } T \longrightarrow \infty$$

Central Limit Theorem holds with this normalizing constant σ^*

- **Example: Two-state, first-order Markov chain**

$$\{J_t: t = 1, 2, \ldots\}, \quad J_t = 0, 1$$

Transition probabilities:

$$P_{jk} = \Pr\{J_{t+1} = k \mid J_t = j\}; \; j, k = 0, 1$$

$$\pi = \Pr\{J_t = 1\} = P_{01} / (P_{10} + P_{01}), \quad d = \mathrm{Corr}(J_t, J_{t+1}) = P_{11} - P_{01}$$

Sum: $N(T) = J_1 + J_2 + \cdots + J_T$

Variance:

$$(1/T)\,\mathrm{Var}[N(T)] \longrightarrow (\sigma^*)^2 = \pi(1 - \pi)\,[(1 + d)/(1 - d)] \;\text{ as } T \longrightarrow \infty$$

Central Limit Theorem holds with this normalizing constant σ^*

Conditional Moments

- **Expected Value**

$$E(X) = E[E(X|Y)]$$

- **Variance**

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}[E(X|Y)]$$

- **Covariance**

$$\mathrm{Cov}(X, Y) = E[\mathrm{Cov}(X, Y|Z)] + \mathrm{Cov}[E(X|Z), E(Y|Z)]$$

Effects of Intermittency

$\{N(t): t = 1, 2, \ldots\}$ number of events in time interval $[0, t]$ (corresponding to point process)

$\{X_t: t = 1, 2, \ldots\}$ independent and identically distributed [$\mu = \mathrm{E}(X_t)$, $\sigma^2 = \mathrm{Var}(X_t)$]

$\{X_t\}$ independent of $\{N(t)\}$

- **Random sum** $S_{N(T)} = X_1 + X_2 + \cdots + X_{N(T)}$

Mean: $\mathrm{E}[S_{N(T)}] = \mathrm{E}[N(T)]\, \mu$

Variance: $\mathrm{Var}[S_{N(T)}] = \mathrm{E}[N(T)]\, \sigma^2 + \mathrm{Var}[N(T)]\, \mu^2$

Central Limit Theorem: Still holds for $S_{N(T)}$

- **Extremes**

-- Ordinary maximum $M_T = \max\{X_1, X_2, \ldots, X_T\}$

Extreme value theorem: M_T**, suitably normalized, has asymptotically Generalized Extreme Value (GEV) distribution function**

$$G(x) = \exp\left(-\{1 + \xi\,[(x - \mu)/\sigma]\}^{-1/\xi}\right), \quad 1 + \xi\,[(x - \mu)/\sigma] > 0$$

-- Maximum with random indices

$$M_{N(T)} = \max\{X_1, X_2, \ldots, X_{N(T)}\}$$

Extreme value theorem still holds for $M_{N(T)}$ **if**

$$(1/T)\, N(T) \rightarrow \pi \ \text{ as } T \rightarrow \infty \ \text{ (in probability)}$$

But now limiting distribution is $[G(x)]^{\pi}$ **(i.e., GEV with adjusted parameters)**

(3) Statistical Background

Mixtures:

-- Chance mechanism for overdispersion

- **Mixture of Distributions**
- **Dependence Induced by Mixtures**

Mixture of Distributions

- **Finite mixture (two-component case):**

$$\Pr\{X \le x\} = \Pr\{X \le x \mid I = 0\}\,\Pr\{I = 0\} + \Pr\{X \le x \mid I = 1\}\,\Pr\{I = 1\}$$

I two-state random variable (generally only observe X, *not* I)

-- Notation $\quad F(x) = \Pr\{X \le x\}, \quad F_i(x) = \Pr\{X \le x \mid I = i\}, \quad w = \Pr\{I = 1\},$

$$\mu_i = \mathrm{E}(X \mid I = i), \qquad \sigma_i^2 = \mathrm{Var}(X \mid I = i), \quad i = 0, 1$$

-- Moments $\quad \mathrm{E}(X) = (1 - w)\mu_0 + w\mu_1$

$$\mathrm{Var}(X) = (1 - w)\sigma_0^2 + w\sigma_1^2 + w(1 - w)(\mu_1 - \mu_0)^2$$

- **Example: Mixture of two normal distributions**

Conditional densities:

$$f_i(x) = [(2\pi)^{1/2}\sigma_i]^{-1} \exp(-\tfrac{1}{2}[(x - \mu_i)/\sigma_i]^2), \quad i = 0, 1$$

-- Unimodal or bimodal?

Only bimodal if difference between μ_0 and μ_1 large enough relative to σ_0 and σ_1

Sufficient condition for unimodal distribution:

$$|\mu_1 - \mu_0| \leq 2 \min(\sigma_0, \sigma_1)$$

- **Example: Mixture of two exponential distributions**

$$F_i(x) = 1 - (1/\sigma_i)\exp[-(x/\sigma_i)],\ \ x > 0,\ \ i = 0, 1$$

-- Longer tail than single exponential (*not* memoryless)

$$\Pr\{X \le u + x \mid X > u, I = i\} = \Pr\{X \le x \mid I = i\},\ \ i = 0, 1$$

But

$$\Pr\{X \le u + x \mid X > u\} \ne \Pr\{X \le x\}$$

-- Used to fit precipitation "intensity"

Conditional distribution of precipitation amount given occurrence:
$\Pr\{X \le x \mid X > 0\}$

- **Example: Infinite-dimensional mixture of exponentials**

-- Let X have conditional exponential distribution

$$\Pr\{X \leq x \mid \sigma\} = 1 - (1/\sigma)\exp[-(x/\sigma)],$$

where $Y = 1/\sigma$ has gamma distribution

$$\Pr\{Y \leq y\} = [\beta\Gamma(\alpha)]^{-1}(y/\beta)^{\alpha-1}\exp[-(y/\beta)],\ \ y > 0$$

Then unconditional distribution of X is Pareto

$$\Pr\{X \leq x\} = 1 - (1 + \beta x)^{-\alpha}$$

-- Chance mechanism by which heavy tail obtained from light tail

Dependence Induced by Mixtures

- **Example: Induced "spatial" dependence**

Let X & Y both have mixture distributions given same two-state conditioning variable I

Conditional means: $\mu_X(i) = \mathrm{E}(X \mid I = i), \quad \mu_Y(i) = \mathrm{E}(Y \mid I = i), \; i = 0, 1$

Conditional variances: $[\sigma_X(i)]^2 = \mathrm{Var}(X \mid I = i), \; [\sigma_Y(i)]^2 = \mathrm{Var}(Y \mid I = i)$

Assume X & Y conditionally independent given I [so $\mathrm{Cov}(X, Y|I) = 0$]

-- Unconditional covariance

$$\mathrm{Cov}(X, Y) = w(1 - w)[\mu_X(1) - \mu_X(0)][\mu_Y(1) - \mu_Y(0)]$$

- **Example: Hidden Markov model**

$\{X_t: t = 1, 2, \ldots\}$ has mixture distribution given J_t (two-state variable with conditional means μ_0 & μ_1, standard deviations σ_0 & σ_1)

$\{J_t: t = 1, 2, \ldots\}$ first-order Markov chain (with parameters π & d)

Assume $\{X_t\}$ conditionally independent given $\{J_t\}$

-- Unconditional autocorrelation function of $\{X_t\}$

$$\mathrm{Corr}(X_t, X_{t+l}) = a\,\mathrm{Corr}(J_t, J_{t+l}) = a\,d^{\,l}, \quad l = 1, 2, \ldots$$

Where $$a = [\pi(1-\pi)(\mu_1 - \mu_0)^2] \,/\, \mathrm{Var}(X_t),$$

$$\mathrm{Var}(X_t) = (1-\pi)\sigma_0^{\,2} + \pi\sigma_1^{\,2} + \pi(1-\pi)(\mu_1 - \mu_0)^2$$

- **Example: AR(1) process with shifting mean**

$\{X_t: t = 1, 2, \ldots\}$ conditional AR(1) process given $\{I_t: t = 1, 2, \ldots\}$

$$\mu_i = \mathrm{E}(X_t \mid I_t = i), \quad i = 0, 1$$

Assume $$\sigma_i^2 = \mathrm{Var}(X_t \mid I_t = i) = \sigma^2, \quad i = 0, 1,$$

$$\varphi_{ij} = \mathrm{Corr}(X_t, X_{t+1} \mid I_t = i, I_{t+1} = j) = \varphi, \quad i, j = 0, 1$$

$\{I_t\}$ independent & identically distributed ($w = \Pr\{I = 1\}$)

-- Autocorrelation function of $\{X_t\}$

$$\mathrm{Corr}(X_t, X_{t+l}) = a\,\varphi^l, \quad l = 1, 2, \ldots$$

where $$a = \sigma^2 / \mathrm{Var}(X_t), \qquad \mathrm{Var}(X_t) = \sigma^2 + w(1 - w)(\mu_1 - \mu_0)^2$$

Derivation of autocorrelation function:

Because $\mathrm{Cov}[\mathrm{E}(X_t \mid I_t, I_{t+l}), \mathrm{E}(X_{t+l} \mid I_t, I_{t+l})] = 0$,

$$\mathrm{Cov}(X_t, X_{t+l}) = \mathrm{E}[\mathrm{Cov}(X_t, X_{t+l} \mid I_t, I_{t+l})]$$

So $\mathrm{Corr}(X_t, X_{t+l}) = \mathrm{E}[\mathrm{Cov}(X_t, X_{t+l} \mid I_t, I_{t+l})] / \mathrm{Var}(X_t)$

$$= \sigma^2\varphi^l / \mathrm{Var}(X_t)$$

$$= \{\sigma^2 / [\sigma^2 + w(1 - w)(\mu_1 - \mu_0)^2]\}\varphi^l, \;\; l = 1, 2, \ldots$$