Lecture 25, November 16, 2010 (Key Points)

STATISTICAL VARIABILITY IN SPACE-TIME RAINFALL AND ITS PHYSICAL ORIGINS

25.1 Stochastic Models of Space-Time Rainfall: A brief Overview

Since a very long time, hydrologists have been interested in investigating the structure of rainfall variability using statistical methods and stochastic models. This focus developed due to availability of rainfall time series at individual rain gages. These 'Point' measurements of rainfall have been used extensively in engineering applications. In the absence of spatial rainfall observations, hydrologic engineers developed a variety of empirical method to extrapolate point rainfall to spatial rainfall. For example, we briefly covered the problem of area reduction factor (ARF) to scale point rainfall to areal rainfall in Lecture 24. Waymire and Gupta (1981) published a review of stochastic models of point rainfall that were developed till the late 1970s. Many other point rainfall stochastic models have been published since then. We will not be able to cover this body of literature in this course. Likewise, another body of literature covers spatial rainfall variability at multiple gauges using multivariate statistics, which we will not discuss here. I have chosen to give a very brief overview of stochastic models of space-time rainfall rather than only temporal or only spatial rainfall models.

The somewhat idealized hierarchical geometric structure of rainfall as depicted in Fig. 24.1 (Lecture 24) guided the development of many statistical models of space-time variability of rainfall. LeCam (1961) published the first set of mathematical results that were aimed at representing the hierarchical geometric structure of space-time rainfall as a stochastic process. Waymire and Gupta (1981) and Waymire et al. (1984) further pursued this line of inquiry. These mathematical approaches are based on the assumption that distinct spatial scales and corresponding time scales exist in rainfall fields depicted in Fig. 24.1, and Table 24.1 in Lecture 24. A substantial body of literature has developed along this line. One drawback of these models is that they have too many model parameters, which are not easy to estimate from data.

In the early 1980s, a second set of models began to appear which were based on the assumption that rainfall did not have characteristic spatial scales, and that it exhibits statistical scale invariance. A simple form of statistical self-similarity called statistical simple scaling was assumed to hold (Lovejoy and Mandelbrot, 1985). We have encountered it before in previous lectures, but let us understand it mathematically following Gupta and Waymire (1998).

25.2 Statistical Simple Scaling

Let $\{Q(A)\}$ be a random field indexed by subregions A of D, for example all the sub basins of a basin. Let $\lambda > 0$ be an arbitrary scalar. The random field is defined to exhibit *simple scaling* if for a positive *scaling function* $g(.)$ the following equality holds:

$$Q(\lambda A) \overset{d}{=} g(\lambda) Q(A) \tag{25.1}$$

where λA is the set A rescaled by the factor $\lambda > 0$. This equality is understood to mean that all of the finite dimensional joint distribution functions of the two random fields are the same (for simplicity, you may take that Eq. (25.1) holds for marginal distribution). To evaluate the scaling function $g(.)$, we take a *unit* subregion and consider two arbitrary positive scalars, λ and μ. We apply Eq.(25.1) iteratively as,

$$Q(\lambda\mu) \overset{d}{=} g(\lambda)Q(\mu) \overset{d}{=} g(\lambda)g(\mu)Q(1) \overset{d}{=} g(\lambda\mu)Q(1)$$

Comparing the last two terms produces the functional equation,

$$g(\lambda)g(\mu) = g(\lambda\mu) \tag{25.2}$$

The solution of Eq.(25.2), as explained in Lecture 7, is given by a *power law* ,

$$g(\lambda) = \lambda^{\theta} \tag{25.3}$$

The fundamental statistical parameter θ is called a *scaling exponent*. It can take either positive or negative values. This is an important result, because it shows how the power laws are connected with statistical simple scaling. Substituting Eq. (25.3) into Eq. (25.1) gives,

$$Q(\lambda A) \overset{d}{=} \lambda^{\theta} Q(A) \tag{25.4}$$

You have seen simple scaling in the empirical results for annual flood quantiles described in Lecture 4. To see this connection precisely, consider the p-th quantile of *Q(A)* defined as, $P(Q(A) > q_p(A)) = p$. Take λ=*A*, where *A* denotes drainage area in Eq. (25.4), and *A*=1. It follows from the definition of a quantile that,

$$P(Q(A) > q_p(A)) = P(A^{\theta}Q(1) > q_p(A)) = P(Q(1) > q_p(A)/A^{\theta}) = P(Q(1) > q_p(1)) = p$$

Therefore,

$$q_p(A) = q_p(1)A^{\theta} \tag{25.5}$$

where $q_p(1)$ is the p-th quantile of the discharge from a unit area. Eq. (25.5) is identical with the empirical results for flood quantiles in New Mexico regions 1 and 5. Similarly, annual flood quantiles in GCEW obey statistical simple scaling (Lecture 4).

For an application of simple scaling Eq. (25.4) to peak flows in rainfall-runoff events, take λ=*A*, and *A*=1, and rewrite Eq. (25.4) as,

$$Q(A) \overset{d}{=} A^{\theta} Q(1) \tag{25.6}$$

Taking logs on both sides,

$$\log Q(A) \overset{d}{=} \theta \log A + \log Q(1) \tag{25.7}$$

If two random variables X and Y have identical probability distributions, then all of their statistical moments are the same. This can be easily checked from the definition of moments. Therefore, taking conditional expectation given A on both sides of Eq. (25.7) gives,

$$E[\log Q(A)|A] = \theta \log A + E[\log Q(1)] \tag{25.8}$$

This by definition is a regression equation that the peak flow data obey in GCEW (Lecture 20), and in the Iowa River basin (Lecture 23).

Taking moments on both sides on Eq. (25.6), implications of simple scaling for higher-order statistical moments are obtained as,

$$E[Q^n(A)] = A^{\theta n} E[Q^n(1)] \tag{25.9a}$$

or,

$$\log E[Q^n(A)] = n\theta \log A + c_n \tag{25.9b}$$

Here, $c_n = \log E[Q^n(1)]$. This equation shows two very important features, namely, (i) log-log linearity between statistical moments and scale parameter, and (ii) a linear increase in slope growth $s(n) = \theta n$. These two features play important roles in data analysis, but care is needed in addressing the issue of bias in the estimation of slopes from data and how it might affect the interpretation (Lecture 4).

Resuming our discussion of rainfall, Kedem and Chiu (1987) showed that the assumption of statistical simple scaling contradicts the intermittency of rainfall. These difficulties led to developing another set of stochastic space-time models that exhibited statistical multiscaling. It is briefly explained below in the context of rainfall following Gupta and Waymire (1998).

25.3 Statistical Multiscaling

A formal way to construct a multiscaling stochastic process that exhibits (1) log-log linearity between the statistical moments of order n and a scale parameter such that (2) the slope $s(n)$ is a nonlinear function of n is to assume that the following equality holds:

$$Q(\lambda A) \overset{d}{=} G(\lambda) Q(A). \tag{25.10}$$

Here $G(\lambda)$ is a positive random function that is statistically independent of $\{Q(A)\}$. Unlike the case for simple scaling, we get two representations of $G(\lambda)$ corresponding to whether $\lambda < 1$ or $\lambda > 1$. Here, we consider only the case $\lambda < 1$ because it is most relevant to understanding spatial variability in rainfall and floods. Eq. (25.10) can be formally iterated with respect to the two positive scalars $\lambda_1 < 1$ and $\lambda_2 < 1$ in a similar manner as was discussed for simple scaling in the previous section. This gives a functional equation:

$$G(\lambda_1 \lambda_2) \overset{d}{=} G(\lambda_1) G(\lambda_2) \tag{25.11}$$

Because $G(\lambda)$ is a random function of λ, the equality in equation (25.11) holds in the sense of probability distributions. The random scaling function solution of $G(\lambda)$ can be represented as

$$G(\lambda) = \exp\{\pm\mu \log\lambda + Z[\log(1/\lambda)]\},\ 0 < \lambda < 1 \qquad (25.12)$$

Any stochastic process with stationary increments, such as a Brownian Motion (BM) process, has additive growth; however, the converse need not hold.

To illustrate the behavior of the moments for equation (25.12), let us take $\{Z(t): t > 0\}$ to be a BM process starting at zero, i.e., $Z(0) = 0$. This means that $Z(t)$ has statistically independent increments and has a normal distribution, with mean zero and variance $\sigma^2 t$, where $\sigma^2 > 0$. Using the well-known expression for the moment generating function of a normal random variable, it can be verified that

$$E[G^n(\lambda)] = \exp\{\mu n \log\lambda - \frac{1}{2}\sigma^2 n^2 \log\lambda\} \qquad (25.13)$$

Equation (25.13) exhibits, (1) log-log linearity between moments and the scale parameter λ, and (2) a nonlinear slope function $s(n) = \mu n - \frac{1}{2}\sigma^2 n^2$. See Gupta and Waymire (1990) for other examples of this kind. These two features of a stochastic process or a random field are being called *multiscaling* to distinguish them from the simple scaling stochastic described in Section 25.2. *Multiscaling is also a scale invariant property.*

It is not difficult to determine how the quantiles of a multiscaling stochastic process behave with respect to a scale parameter. The derivation given by Gupta et al. (1994) shows that the log-log linearity between quantiles and the scale parameter holds only approximately, and the slopes in these plots decrease as the probability of exceedance decreases or as the return period increases. This feature was noted to be present in the flood data for regions 2 and 4 in New Mexico (Lecture 4).

25.4 Spatial Generalization of Multiscaling

Kolmogorov (1941) assumed that the rate of mean energy dissipation from large to smaller scales remained constant within the inertial range of spatial scales (Eq. 24.7, Lecture 24). Subsequently, Landau pointed out that fluids undergoing turbulent motion exhibit sudden bursts of strong activity, which he called “intermittency”. In turbulence, the presence of intermittence contradicted the homogeneity assumption in Kolmogorov’s theory. Subsequently, Kolmogorov (1962) introduced the lognormal cascade model to account for such a behavior, which later developed into the mathematical theory of random cascades. In the rainfall literature, intermittence refers to the presence of zeros in rainfall.

Schertzer and Lovejoy (1987) first began to explore the theoretical framework of random cascades to describe the spatial variability of rainfall. Although rainfall is not a passive scalar advected by turbulence as they assumed for simplicity, many of the spatial features of rainfall could be modeled within the cascade framework. Therefore, the multifractal cascade framework has continued to be developed and tested on rainfall data sets. It also eliminates some of the difficulties that were encountered with the two previous sets of approaches explained in Sections 25.1 and 25.2. There are two types of cascade theories, continuous and discrete. For an introduction to discrete random cascades, the reader may consult Gupta and Waymire (1993). For the theory of continuous cascades, readers may consult Lovejoy and Schertzer (1990). A basic introduction to the literature on stochastic modeling of rainfall requires a one-semester advance course.

References

Gupta, V. K., O. Mesa, and D. R. Dawdy. Multiscaling theory of flood peaks: Regional quantile analysis, *Water Resour. Res.*, 30(12): 3405-3421, 1994.

Gupta, V. K. and E. C. Waymire. A Statistical Analysis of Mesoscale Rainfall as a Random Cascade. *J. Appl. Meteo.,* 32(2): 251-267, 1993.

Gupta, V.K. and E. Waymire. Spatial variability and scale invariance in hydrologic regionalization, In: *Scale dependence and scale invariance in hydrology*, (Ed. G. Sposito), Ch. 4, pp. 88-135, Cambridge University Press, 1998.

Kedem, B., and L. S. Chiu. Are rain-rate processes self-similar? *Water Resour. Res.*, 23(10): 1816-1818, 1987.

Kolmogorov, A.N. Local structure of turbulence in incompressible fluid at very high Reynolds numbers (in Russian) *Dok. Akad. Nauk. SSSR* 30, 299–303, 1941 (English translation) *Proc. Roy. Soc. Lond*., A 434, 9–13, 1991.

Kolmogorov, A. N. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *J. Fluid Mech.*: 13, 82-85, 1962.

LeCam, L. *4th Berkeley Symposium on Mathematical Statistics and Probability*, V. 3, Univ. of California, Berkeley, California, p. 165-186, 1961.

Lovejoy, S., and B. B. Mandelbrot. Fractal properties of rain and a fractal model. *Tellus*, 37A, 1985.

Lovejoy, S., and D. Schertzer. Multifractals, universality classes, and satellite and radar measurements of cloud and rain fields, *J. Geophys. Res.*,95(D3), 2021-2031, 1990.

Mandelbrot, B. B. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J. Fluid Mech.,* 62: 331-358, 1974.

Schertzer, D., and S. Lovejoy. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, *J. Geophys. Res.* 92(D8): 9693-9714, 1987.

Waymire, E. C., and V. K. Gupta. The mathematical structure of rainfall representations: I. A review of the stochastic rainfall models. *Water Resour. Res.,* 17(5):1261-1272, 1981.

Waymire, E. C., V. K. Gupta, and I. Rodriguez-Iturbe. A spectral theory of rainfall intensity at the meso-ß scale. *Water Resour. Res.,* 20(10): 1453-1465, 1984.