Lecture 4, September 2, 2010 (Key Points)

The problem of hydrological prediction in the presence of sparse or nonexistent stream flow data sets is a very important hydrologic engineering problem of global significance. By partitioning a large region into 'homogeneous sub regions', engineers are able to pool together gauged stream flow data sets, conduct multiple regression analyses and use them as predictors of annual floods and low flows quantiles within a homogeneous region. As explained in lecture 3, the USGS performs regional flood quantile analyses on a state-by-state basis throughout the United States.

(1) However, a purely statistical approach has serious limitations in providing reliable stream flow estimates outside the space-time domains over which data is collected. For example, stream flow records often have missing values that need to be imputed from other sources. During extreme flood conditions, gauge stream flows tend to have large errors due to malfunctioning or destruction of recording devices or errors in extrapolating the rating curve of the gauging station. Therefore, auxiliary estimates are desirable.

(2) The gauged stream flow records are of relatively short duration. They seldom exceed 50-100 years in length. Humans are altering the natural environment all over the world in substantial ways, which impact the gauged records. Therefore, "naturalized" stream flow estimates at locations with substantial human modification of the land hydrology are often needed.

(3) Recognition is growing that climate change due to global warming would exert systematic changes in the hydrologic cycle over diurnal, seasonal, interannual, interdecadal and longer time scales, because water is a key component of the global climate system. As a consequence, hydrological records can become non-stationary, which violates the traditional stationarity assumption underlying regional flood quantile analyses and a majority of other hydrological analyses.

Collectively, the above issues define the problem of predictions from ungauged basins (PUB). This is a seminal unsolved problem in engineering hydrology. Starting in 2003, the International Association of Hydrologic Sciences (IAHS) adopted a ten-year global initiative to make progress towards solving the PUB problem (Sivapalan et al. 2003). One of the key problems within PUB is to develop a scientific framework for estimating stream flow time series and its statistics consistently across gauged and ungauged spatial locations at multiple temporal intervals of interest. It is crucial to address this problem in the context of understanding and predicting how human induced changes affect water availability and hydrologic extremes, and how they differ from natural and/or climate induced changes.

Pattern of Power laws (Scaling) in regional annual flood quantiles

The quantile-regression method of the USGS consists of fitting a log-Pearson III probability distribution to flood data (Brutsaert, 2005, Section 13.4.4) at each station in a region and then using those distributions to compute the quantile discharges corresponding to return periods of 2, 5, 10, 25, 50, 100, and 500 year. The quantile regression equations derived for regional flood frequency analysis show that often, but not always, the drainage area serves as the key basin descriptor for many regions in the United States. This observation follows from the fact that the reduction in the standard error of estimate (Eq. (4.2)) is usually observed to be small by regressing on more variables over and above drainage area. Moreover, many of the variables in multiple regressions can be correlated with area. Following Gupta and Dawdy (1995), define a region to be *homogeneous* if the flood quantiles depend only on drainage area and no other descriptor.

Let $Y = a + bX + \varepsilon$ denote a simple linear regression equation. Then the least squares estimate of b is given by,

$$\hat{b} = \frac{\sigma(Y)}{\sigma(X)}\rho(X,Y) \tag{4.1}$$

Here σ is the standard deviation, and ρ is the correlation coefficient. The standard error of estimate is given by,

$$SE = \sigma(Y)[1 - \rho^2(X,Y)]^{1/2} \tag{4.2}$$

The term $[1-\rho^2(X,Y)]^{1/2}$ is called the regression effect. If $\rho = 1$, then both the regression effect and SE are zero, and $b = \sigma(Y)/\sigma(X)$ gives the slope estimate without the regression effect. In general, $\rho \neq 1$. For a fixed $\sigma(Y)$, larger SE means a smaller ρ, and therefore a smaller slope b due to the regression effect.

Above discussion implies that if the standard errors within a region are about the same for different flood frequencies, then each regression equation explains about the same amount of variance, and the regression effect does not influence the relative values of the exponents in the quantile relations. Presence of uniformly large standard errors for all the quantiles will reduce the magnitude of the exponents, but will not change their relation to each other. *It is mainly the relation of slopes to each other that we care about the most, as that will enable us to infer a fundamental pattern known as power law, or scaling, in a region.*

(1) Consider New Mexico, which is divided into eight homogeneous regions with an average area of 15,200 mi^2 (39,000 km^2). Table 4.1 below gives the computed values for the exponents (EXP) and the corresponding standard errors (SE) for four regions in New Mexico. **Region 5** is mountainous and contains the headwaters of the Canadian River. Not only are the exponents relatively constant, but also the runoff is almost directly related to the drainage area because all the exponents are greater then 0.9. In the northern mountain region, floods are mostly produced from snowmelt runoff. **Region 1** is downstream from region 5, and its mainstream stations receive flows from region 5. It also exhibits an approximate constancy of the exponents, but with a considerably smaller value than 0.9. For these two regions, the exponent is independent of the return period, 1/*p*, or $m(p) = m$ for all *p*.

(2) By contrast, **Regions 2 and 4** have generally decreasing exponents, with increasing return periods for floods. These two regions, as well as region 6 between them, which is not shown here, constitute a wide band across the middle of the state

and exhibit similar decreases in exponents. Those floods in the plains, plateaus, valleys, and deserts generally are produced by rainfall.

Quantiles	Region 1 Northeast Plains		Region 5 Northern Mountains		Region 2 Northwest Plateau		Region 4 Southeast Plains	
	EXP	SE	EXP	SE	EXP	SE	EXP	SE
Q2	.56	.35	.912	.34	.52	.38	.671	.45
Q5	.55	.31	.920	.32	.47	.33	.591	.34
Q10	.55	.30	.924	.32	.44	.31	.546	.28
Q25	.55	.30	.929	.34	.41	.30	.498	.23
Q50	.55	.31	.933	.35	.39	.30	.465	.21
Q100	.56	.32	.936	.37	.37	.30	.436	.19
Q500	.581	.32	.940	.39	.365	.32	.408	.19

Table 4.1 Exponents and the standard errors of estimate in flood-quantile-drainage-area relations for four New Mexico regions (Gupta and Dawdy, 1995)

(3) The decreasing exponents suggest that the regional relationships produced by rainfall generated flood peaks are quite different from those produced by snowmelt generated Floods. The reader may refer to Gupta and Dawdy (1995) for further details of the New Mexico study and similar analyses for other regions in the United States. Cathcart (2001) used the definition of regional homogeneity given above, and conducted a comprehensive study of regional flood frequency analysis for Oregon.

In summary, the quantile regression equations above for each homogeneous region can be written as (Brutsaert, 2005, eq. 13.9.2),

$$q_p = c(p)A^{m(p)} \tag{4.3}$$

Observed Power laws in annual flood quantiles in river basins

River basins are nested in the sense that all sub basins drain to a common outlet. USGS homogenous regions are not river basins because different gauging locations in it don't drain to a common outlet. *But, to understand how regional annual flood frequencies are connected with physical processes, it is necessary to consider river basins rather than arbitrary geographic regions*. We consider only basins from here on out.

1. We also consider basins that have stream flow gauges at many locations inside it that are drained by a wide range of drainage areas. This way annual flood

frequencies are computed for a wide range of drainage areas representing spatial scales. Such a set-up is necessary to analyze if flood quantiles exhibit power laws (scaling) with respect to drainage areas.

2. Two experimental basins of USDA/ARS satisfy the conditions in items 1 and 2. They are 21 km^2 Goodwin Creek Experimental Watershed (GCEW), MS, and 150 km^2 Walnut Gulch basin, AZ. Recently, the Iowa river basin (~35,400 km^2) has been added to this list (Gupta et al., 2010).

3. Many small to medium-sized basins in the US (and all over the world) don't have multiple stream flow gauges inside of them. The requirement of multiple gauges inside a basin can severely limit the choice of basins.

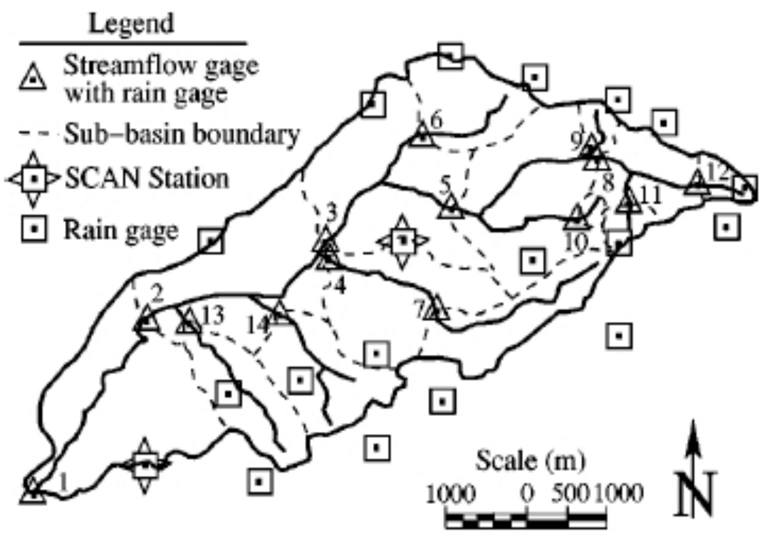

Fig. 1. USDA-ARS National Sedimentation Laboratory Goodwin Creek Experimental Watershed

4. A figure of GCEW is given above, which shows the spatial locations of rainfall and stream flow gauges.

5. Ogden and Dawdy (2003) analyzed 2-year and 20-year flood quantiles with respect to drainage areas as shown in Fig. 2. You can clearly see the power laws in both the graphs with the same exponent, 0.77.

6. Consider Walnut Gulch basin. AZ. Goodrich et al. (1997) plotted 2-year and 100-year flood quantiles against drainage areas using annual peak flow data from the eighteen nested subbasins in the Walnut Gulch basin as shown in Fig. 3 below.

7. Three features in Fig. 3 should be noted: (i) a power law or scaling relationship between peak flows and drainage areas with an exponent $m_1 \in (0.85, 0.9)$, which is close to 1 for basins smaller than 1 km^2; (ii) a change in scaling exponents around 1

km^2 (unfortunately, missing gauged observations around 1 km^2 don't allow us to say precisely whether the transition is gradual or abrupt); and (iii) a power law exponent $m_2 \in (0.55, 0.58)$, which is close to 0.5 for basins larger than 1 km^2. Goodrich et al. (1997) don't give error bounds for the exponent values.

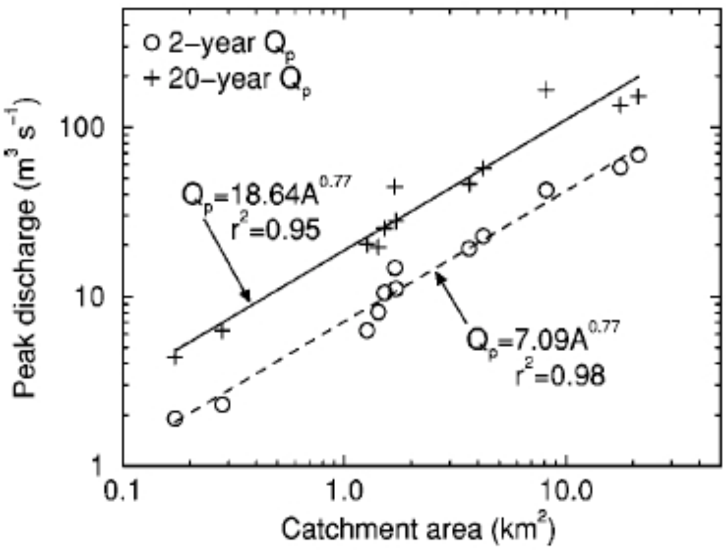

Fig. 2. Flood quantiles versus drainage area for 2- and 20-year recurrence intervals with power law regressions, excluding basins 10 and 12

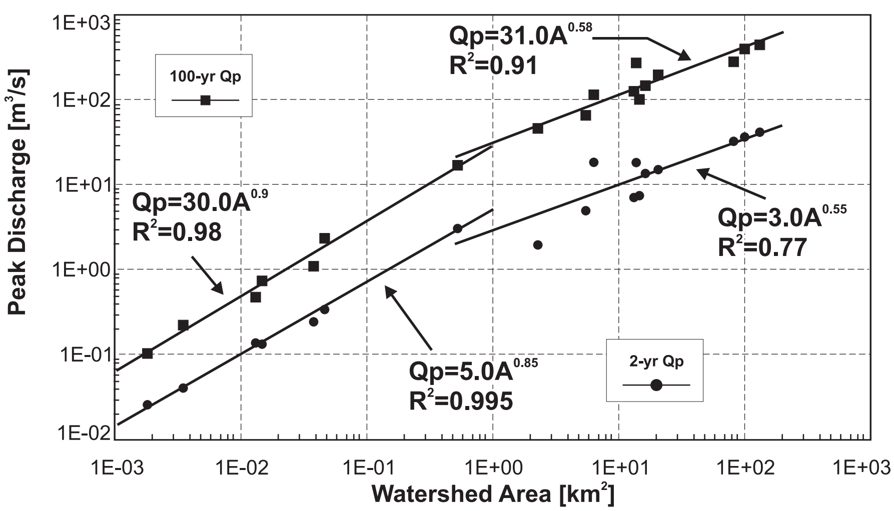

Fig. 3. A plot of annual flood quantiles for 2- and 100-year return periods versus drainage areas in the Walnut Gulch basin (Goodrich et al. 1997).

8. Do you see a correspondence between the exponents above and the values of the exponents in region 1 and 5 in New Mexico displayed in Table 4.1? Is it a

coincidence or does it have a physical interpretation? The new concepts introduced in this course would help to address such questions.

References

Cathcart, J. PhD dissertation, U. British Columbia, Vancouver, CA, 2001.

Goodrich et al., *Water Resources Research*, 33, 2951-2965,1997.

Gupta, V.K. and D.R. Dawdy, *Hydrological Process*, 9(3/4): 347–361, 1995

Gupta, V.K., R. Mantilla, B. M. Troutman, D. Dawdy, W. F. Krajewski, Generalizing a nonlinear geophysical flood theory to medium-sized river networks, *Geophysical Research Letters,* V. 37, L11402, doi:10.1029/2009GL041540, 2010.

Ogden, F. and D. Dawdy, *J. Hydrologic Eng.* (*ASCE*), March/April, 64-73, 2003

Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J., Mendiondo, E., O'Connell, P., Oki, T., Pomeroy, J., Schertzer, D., Uhlenbrook, S. and Zehe, E. (2003) IAHS decade on predictions in ungauged basins (PUB), 2003-2012: Shaping an exciting future for the hydrologic sciences. *Hydro. Sci. J.* 48, 857-880.