Presented at 15th Conference on Hydrology, AMS, January 9-14, 2000, Long Beach, CA

P1.8

CRITERIA TO SELECT BASINS FOR HYDROLOGIC MODEL DEVELOPMENT AND TESTING

John Schaake*, Qingyun Duan, Michael Smith and Victor Koren
Office of Hydrology, NOAA/NWS, 1325 East-West Highway, Silver Spring, Maryland 20910

1. Introduction

In light of a plethora of hydrologic models in existence today and continued interest in their development, there is a real need for formulating objective criteria to evaluate the models. Data for many basins are needed to support development of methods for a priori estimation of model parameters and for testing model performance. Hydrologists traditionally have approached parameter estimation by calibrating models to specific basins and by assuming that parameters for one basin may be appropriate for near-by basins that could not be calibrated. Moreover, hydrologists have generally relied on retrospective testing of models over only a few basins to evaluate model adequacy.

This paper is concerned with the question what kinds of basins and data are needed to develop a priori parameter estimation techniques and to conduct effective retrospective testing. Clearly, there should exist sufficient length of data period and adequate density of precipitation gages. Study basins should be representative of a variety of climates and of different hydrologic response characteristics in a given climate. Basins with single dominant vegetation and soil characteristics are needed to understand the effects of vegetation and soils on hydrologic processes. But basins with complex vegetation and soils are needed for testing as well. Lastly, with availability of radar data, consideration must also be given to the location of study basins in relation to radar sites. Presented below is an analysis of potential basins in the U.S. that might be used for the international Model Parameter Estimation Experiment (MOPEX) project and for retrospective testing in the GCIP Land Data Assimilation Schemes (LDAS) project. Additional information about this study may be found at http://www.nws.noaa.gov/oh/mopex.

One of the most difficult questions to answer is how many basins are needed for a particular study. In practice the answer often is limited by resources to develop data sets for many basins. Because of this, the MOPEX project is developing data sets that can be used to study a large number of basins located mainly in the U.S. but including basins worldwide as well. This study focuses on the potential availability of data sets for the U.S. and leaves open the question of how many basins should be used.

2. Unregulated Basins for Model Development and Testing

Two different subsets of USGS gaging stations that are believed to be free of significant upstream flow diversion or flow regulation have been compiled to support hydoclimatic studies. One of these is the Hydroclimatic Data Network (HCDN) (Slack, et al, 1993) compiled by the USGS. The other is a data set compiled by Wallis, Lettenmaier and Wood (1991). The total set of stations for the conterminous United States consists of 1861 basins. Some of these are nested upstream of other gages. They cover a wide range of basin size from 3 to 10,000 sq. km. The mean is 1352 sq km; the standard deviation 1826 sq km. Figure 1 shows the locations of these gages. Basin boundaries were developed as part of the MOPEX project for these basins.

3. Required Length of Historical Record

Experience with model calibration in the National Weather Service is that the record length must be long enough to include extremes of both wet and dry periods so that a full range of basin response can be observed. Generally this is about 10 years and may be longer for drier basins

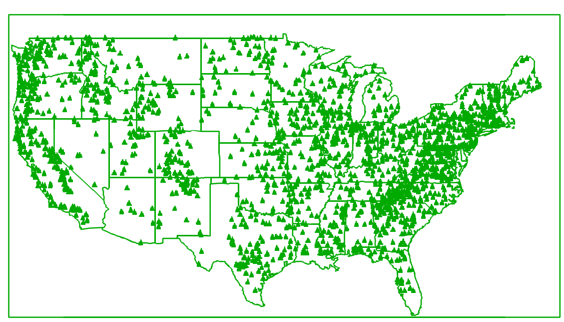

Figure 1. Location of natural flow stream gages

* Corresponding author: John Schaake
OH/NWS/NOAA, 1325 East-West Highway
Silver Spring, MD 20910

where relatively fewer precipitation events occur. Our recent diagnostic water budget study (Total water storage range of the Mississippi basin, paper 2.9) indicates that at least 10 years of data are needed to observe a reasonable range of variability of total water storage and more years are preferred. Although some model parameters may be observable from only a few years of observations, much more confidence will likely come from having 20 or more years of data for a given basin.

4. Precipitation Gage Density Requirements

The number of precipitation gages required for a given basins depends on how the data are to be used. In this study, it is assumed the data are needed to compute a mean areal precipitation over the basin. The time step is application dependent. For model calibration and parameter estimation studies, this will be daily or less. For water budget studies, it might be monthly. The accuracy of the mean areal precipitation estimate depends on the spatial decorrelation structure of the precipitation. Spatial decorrelation distance tends to be less at high latitudes for short time steps and less in summer than in winter. By far the largest number of available long period gages are from the NCDC daily and hourly precipitation networks.

A practical estimate of gage density requirements was made by Schaake (1981) for river forecasting applications. The required number of gages for a basin of area, A (sq km), is

$$N = 0.6 A^{0.3} \qquad (1)$$

The time step used to derive this equation is one-fourth of the basin lag time. The exponent 0.3 implies that the required number of gages doubles as the basin size increases by a factor of 10. The number of gages given by this equation should give mean areal precipitation estimates for each time step that are accurate to within 20 percent 80 percent of the time during thunderstorm rainfall events (in the 20,000 sq km Muskingum,OH river basin). Basins with less than average time lag may require slightly more gages to achieve this accuracy. The equation is reasonable to apply for basins between 200 and 20,000 sq km. This is a fairly conservative gage density requirement for hydrologic modeling studies, but basins that have at least this number of gages should have high quality data available. The equation was developed from a study of observations from a very dense gage network (45 sq km per gage). Below 200 sq km the equation may underestimate gage requirements because the spatial decorrelation function for precipitation (estimated from pairs of stations over many events) tends to fall quickly with distance for very short distances and then quite slowly thereafter. This means that a minimum number (about 3) gages are needed for small basins, mainly to filter the noise associated with this "nugget" of the decorrelation function.

5. Basins with Potentially Adequate Historical Data

The locations of daily climatological precipitation gages for which data are available from NCDC were used to compute gage densities in the vicinity of each of the basins in Figure 1 and to estimate the effective number of gages available for each basin. Figure 2 compares the number of available gages for each basin with the required number of gages according to equation (1). Points that lie above the solid curve representing equation (1) have potentially sufficient data. Only 16 percent of all basins have a sufficient number of gages. In fact, because not all of the stations may have been operating at the same time, the number of available gages may actually be less than shown in Figure 2. Figure 3 shows the locations of gages with potentially adequate data. Also shown in Figure 2 is a relaxed required gage criteria equal to half the equation (1) requirement. A total of 305 basins or 16 percent of the total meet this requirement.

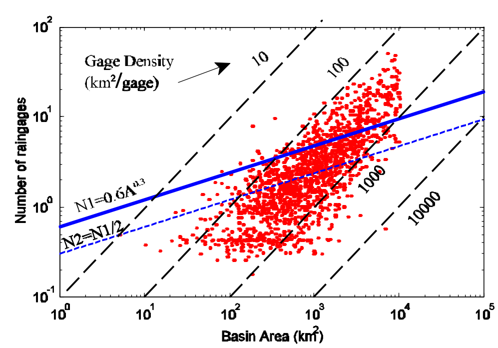

Figure 2. Required number of gages vs available number of gages for all study basins in the US

6. **Radar Coverage Requirements**

More spatially complete precipitation data are

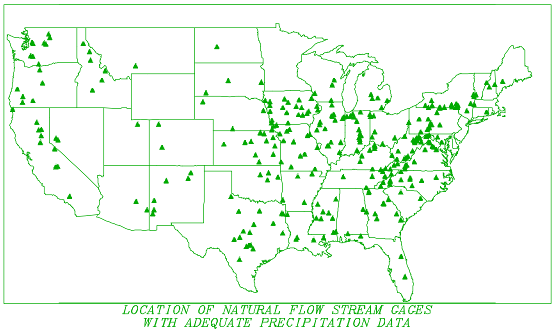

Figure 3. Location of natural flow stream gage with adequate precipitation data

needed to understand how spatial variability of precipitation influences basin response than can are generally available from operational and climatic precipitation gage networks. Multisensor precipitation estimates from combined weather radar and precipitation gage data may be sufficient for many basins. This requires that the basin be located under one or more radar umbrellas within the effective range of the radar. The upper limit of the radar range is limited by several factors. One of these is the radar beam may rise above the precipitating clouds and fail to detect precipitation. Figure 4 shows that for two radars, State College, PA (CCX) and Sterling VA (LWX) this begins to occur at about 150 km. During individual events in the cold season the limiting range may be less if the radar beam rises above the freezing level.

7. **Climate, Soils and Vegetation Characteristics**

Gridded values of climate soils and vegetation

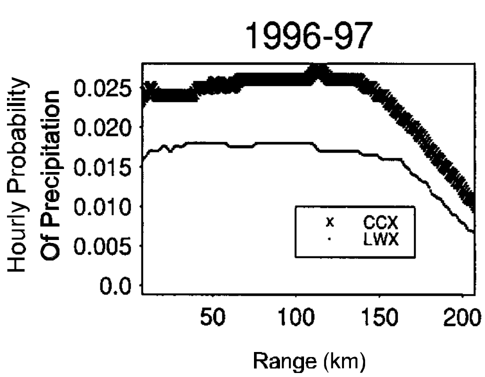

Figure 4. Hourly Probability of Precipitation vs Range

characteristics can be used to derive basin characteristics for each basin. A useful variable

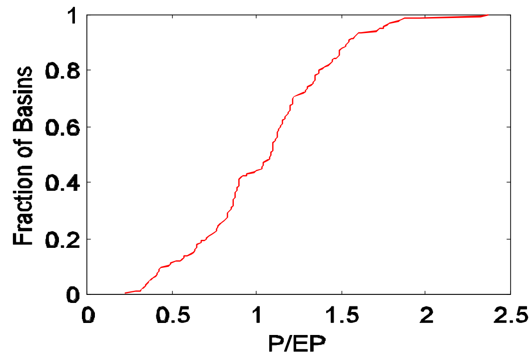

Figure 5. Distribution of P/EP in the Mississippi River basin

to characterize the climate of each basin the ratio (P/EP) of mean annual precipitation to mean annual potential evaporation. P/EP for each basin was estimated from gridded values of P from the natural Resources Conservation Service PRISM project (see http://www.ocs.orst.edu/prism/prism_new.html) and gridded values of EP from the NOAA Evaporation Atlas (Farnsworth, et.al., 1982). The distribution of P/EP values for basins within the Mississippi river basin are shown in Figure 5. A number of approaches have been developed to classify vegetation and gridded files of these are now being processed to identify the vegetation distributions in each basin. These will be used to identify basins with very large fractions (say 80 percent) of each vegetation type. Soils texture data from STATSCO have been gridded by Miller (1997). Mean soils hydraulic properties are associated with each texture class and average properties are being computed for each basin. Figure 6 shows the distribution of basin average saturated hydraulic conductivity values derived for 39 basins in the Arkansas/Red river basin.

8. **Summary**

The data sets and analysis described above are being developed to produce a complete set of basin characteristics for each of the basins illustrated in Figure 1. It is planned to use these characteristics together with alternative clustering algorithms to partition the total number of basins into various subsets depending on the clustering criteria. The status of this work will be presented.

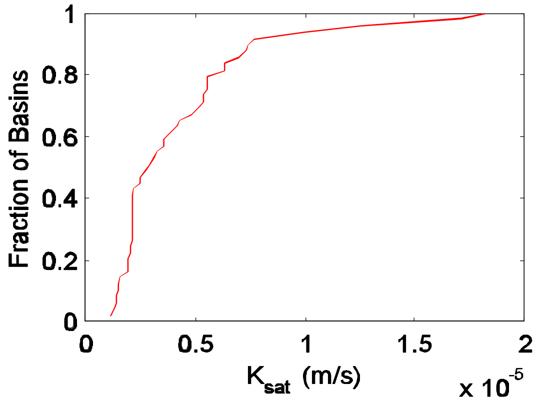

Figure 6. Distribution of basin average staturated hydraulic conductivity in the Arkansas/Red river basin

9. References

Farnsworth, R.K., E.S. Thompson and E.L. Peck, 1982, Evaporation Atlas for the Contiguous 48 United States, NOAA Technical Report NWS 33.

Miller, D.A. and R.A. White, 1997: A Conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. Submitted to Earth Interactions

Slack,J.R., A.M. Lumb and J.M. Landwehr, 1993, Hydroclimatic Data Network: Streamflow Data Set, 1874-1988, USGS Water Resources Investigations Report 93-4076.

Wallis, J.R., D.P. Lettenmaier and E.F. Wood, 1991,

Schaake, J.C., 1981, Summary of River Forecasting Raingage Network Density Requirements, unpublished, available at http://www.nws.noaa.gov/oh/mopex/raingage%20density%20requirement.htm.

Seo, D.J., R.A. Fulton and J.P. Breidenbach, 1997, Final Report for June 1, 1996 to May 31,1997 Interagency Memorandum of Understanding Among the NEXRAD Program, WSR-88D Operational Support Facility, and the National Weather Service Office of Hydrology Hydrologic Research Laboratory, Silver Spring, MD.