INTERNATIONAL SYMPOSIUM ON UNCERTAINTIES IN HYDROLOGIC
and WATER RESOURCE SYSTEMS

# UNCERTAINTIES IN ESTIMATING RUNOFF-PRODUCING RAINFALL FOR THUNDERSTORM RAINFALL-RUNOFF MODELS

H.B. Osborn, W.C. Mills, and L.J. Lane

Research Hydraulic Engineer, Research Hydraulic Engineer,
and Hydrologist, respectively, Southwest Watershed
Research Center, Soil, Water and Air Sciences,
Western Region, Agricultural Research Service,
USDA, 442 E. 7th St., Tucson, Arizona, 85705

## ABSTRACT

A stochastic model of thunderstorm rainfall is being developed at the USDA, ARS Southwest Watershed Research Center, Tucson, Arizona, for use in a runoff prediction model for arid land watersheds. Records from the 57-square-mile USDA-ARS Walnut Gulch Experimental Watershed in southeastern Arizona, and elsewhere, indicate that thunderstorm rainfall fits the definition of a stochastic process very well. A runoff prediction model based on a stochastic input (thunderstorm rainfall) and a deterministic watershed response produces a stochastic output (runoff). The validity of the runoff model depends upon the accuracy and certainty of the input and watershed response models. In this paper the uncertainties in the stochastic rainfall model are examined and the effects of these uncertainties on the stochastic output are investigated, assuming that the deterministic watershed response model is without error. Sampling errors and effects of simplification are carried through the process and compared with observed variability in actual data.

## INTRODUCTION

Records from a 95-rain-gage network on the USDA, ARS Walnut Gulch Experimental Watershed in southeastern Arizona (Figure 1) are being used to develop a stochastic model of air-mass thunderstorm rainfall. These air-mass (regional, as opposed to frontal) thunderstorms occur in the summer rainy season, June through September, and produce about 70 percent of the annual precipitation and essentially all of the runoff from rangeland watersheds in southeastern Arizona (Osborn and Hickok, 1968).

In the stochastic rainfall model, probability distributions are used to model random variables--number of cells, spatial distribution of the cells, and cell center depths--describing thunderstorm rainfall on the Walnut Gulch watershed. The occurrence or non-occurrence of a thunderstorm on the watershed for any given day during the rainy season is modeled as a Bernoulli random variable with a variable probability of occurrence based upon historical data. In general, the occurrence of air-mass thunderstorms in a particular area within a climatic region appears random, and the depths and intensities of rainfall appear, within limits, to be random. The definition of a stochastic process is one whose development in time and/or space is partly or wholly random; therefore, thunderstorm rainfall appears to fit neatly the definition of a stochastic process in hydrology.

The basic stochastic model of thunderstorm rainfall is being developed primarily for use with a watershed response model to predict runoff and peak discharge from arid land watersheds in the Southwest. The complete runoff prediction model, or method, would consist of a stochastic input (thunderstorm rainfall) and a deterministic watershed response. This would produce a stochastic output (runoff). Modeling watershed response is particularly vexing in semiarid regions because of the compounding effects of abstractions in ephemeral channels and the limited spatial precipitation distribution. The validity of the output for prediction depends upon both the accuracy of the input and of the watershed response model. In this paper, the uncertainties in the stochastic rainfall model are examined and the effects of these uncertainties on the stochastic output are investigated, assuming the deterministic watershed response model is without error. Effects of uncertainties in watershed response is a subject for subsequent analysis.

## UNCERTAINTY

In analysis, the objective is to quantify the uncertainty; that is, to group or combine the errors and assign a distribution to them so that probabilistic statements can be made. In this manner, the errors are given a relative likelihood of occurrence. In a stochastic model uncertainty must be considered for the process and for specific random variables.

### Random Variables

If X is a random variable with distribution F(X), then the uncertainty in X is quantified by F. Although normally F is not known, certain conclusions can be made based on samples or observations on X. For example, Chebyshev's inequality, assuming a second moment, can be used to construct probabilistic bounds for X (Feller, 1957). In some cases the central limit theorem holds so that an even stronger statement can be made. For these cases stronger conclusions can be drawn since the distribution of a sample statistic, the sample mean, is known.

Stochastic Process

If X(t) is the output from a stochastic process, the usual case, especially in hydrology, is a single time series of the process. That is, only one sample function is available from a process. Repeated samples are not available as is often the case for random variables. Therefore, the problem is much more complex and requires more assumptions or knowledge of the physical system. While this concept is elementary and obvious, it is quite often unsaid or overlooked. For stochastic processes the limit laws are of primary importance. The degree of uncertainty relative to random variables and stochastic processes is shown in the attached sketch (Figure 2). While this sketch is not all inclusive of the differences between random variables and stochastic processes, it does serve to illustrate the primary differences for hydrologic time series. It is obviously a simplification of the concepts of random variables and stochastic processes.

## THUNDERSTORM RAINFALL MODEL

Many assumptions and simplifications are involved in developing a model of a physical process. This is certainly true in modeling thunderstorm rainfall. The physical processes causing a thunderstorm at a certain time and place are very complex, as are the processes determining depths, duration, and areal extent of the thunderstorm rainfall. Thus, many assumptions and simplifications are necessary to make model solutions practical.

A stochastic air-mass thunderstorm rainfall model for generating runoff-producing rainfall based upon certain assumptions and simplification has been proposed (Osborn, Lane, and Kagan, 1971). This model (CELTH-5) is developed in two parts. The first part, or routine, determines whether a storm will occur, and if so, the time of occurrence. The second part generates runoff-producing rainfall through addition of individual synthetic storm cells.

More recently, CELTH-6, a model for generating total storm rainfall has been suggested. In CELTH-6, total storm rainfall is generated by adding random amounts of nonrunoff producing rainfall, as determined by a negative exponential distribution, to each runoff-producing rainfall cell generated by CELTH-5. In this paper, the assumptions and simplifications incorporated in the runoff-producing rainfall model are examined, since they are possible causes of uncertainty in the stochastic output. The assumptions and simplifications are listed and discussed as follows:

1. All runoff-producing storms for small (100-square-mile and less) watersheds in southeastern Arizona result from air-mass thunderstorms. These thunderstorms are the runoff "design" storms for small watersheds. Moist air for air-mass thunderstorms generally comes from the Gulf of Mexico.

2. Frontal activity is not important in runoff design in southeastern Arizona, although tropical storms off Baja California may move moist Pacific air into Arizona, particularly in September (Sellers, 1960). In southeastern Arizona (Walnut Gulch) storms occurring from "Pacific" air are still considered air-mass for small watershed design.

3. Storm probability of occurrence is based on 12 years of Walnut Gulch data. The process is assumed stationary, and the 12-year record is assumed to adequately represent a longer record.

4. There is no persistence between events. That is, there is no allowance for a causal relation resulting in wet-wet, dry-dry, and so on. However, there is seasonal persistence, as indicated by changing probabilities for thunderstorm occurrence during the season (May 15-Oct. 15). There is a much greater chance of occurrence on a day in late July, for example, than in June or early July, and this is included in the model.

5. Storm starting time is normally distributed about a mean of 1700 hours with a standard deviation of 3.5 hours (determined from Walnut Gulch data), corresponding to the late afternoon occurrences due to diurnal heating.

6. No two storms can occur within less than 3 hours; two or more storms can occur in one day. There is a 1/5 chance of two storms occurring on the same day, 1/25 chance of three occurring, and so on. The fractions for multiple occurrence are multiplied times the regular probabilities. For example, if the model indicates the chance of a storm occuring before 2100 hours is 0.4, there would be a .08 probability of a second storm occurring on the same day.

7. Thunderstorms are assumed to be made up of three or more circular cells.

8. Individual cell center depth varies according to a negative exponential distribution.

9. Cells have a fixed diameter at near zero rainfall (.01 inch).

10. Cell depth-area relationship is linear from the center out to a radius of $\sqrt{\frac{1}{\pi}}$ (the radius for an area of one square mile). At this radius the depth is 85% of center depth. From this "isohyet" down to .01 the relationship is logarithmic.

11. Cells within each thunderstorm develop sequentially both in time and direction, although they may occur almost simultaneously. Individual cells are temporally contiguous.

12. The model generates runoff-producing rainfall (0.5 in/hr or greater) continuously at any point, and this rainfall can be adequately described by depth, duration, and centroid.

13. The first thunderstorm cell can be centered anywhere in a specified field. Its location is random as determined by a uniform distribution.

14. The preferred direction of the second cell in respect to the first cell is random as determined with a uniform distribution.

15. The distance between successive cells is determined independently by a triangular distribution roughly representing a gamma distribution. The triangular distribution was chosen for simplicity by trial and error, because a more sophisticated distribution was not believed to be justified due to the difficulty of precisely defining limits of individual cells.

16. The third cell movement direction is determined by a truncated normal distribution about the direction established between the second and first cell. Direction of movement of successive cells is determined similarly.

17. The number of cells in a storm is determined by a Poisson distribution, truncated with a 3-cell minimum as suggested by Petterssen (1957) and by observations of Walnut Gulch data. The value used in the distribution is such that very few storms contain more than 6 or 7 cells.

## MODEL VALIDITY

In order to test the validity of the storm rainfall model and thus determine whether the assumptions and simplifications incorporated are reasonable, it is necessary to compare observed with simulated rainfall characteristics. Also, since the objective in simulating thunderstorm rainfall is to obtain peak discharge predictions through a deterministic functional relation of rainfall and runoff, model validity can be tested by comparing observed peak discharge rates with those obtained with the rainfall-runoff functional relation and simulated rainfall. In this paper both types of comparisons are made. The functional relation used to obtain peak discharge for a given simulated thunderstorm rainfall is one previously developed for the Walnut Gulch watershed.

In addition to testing validity of assumptions and simplifications in the rainfall model, a sensitivity analysis was done to evaluate the uncertainty in values of rainfall model parameters.

The runoff-producing model (CELTH-5) was used to generate several 12-year sequences of synthetic thunderstorm rainfall. In generating these sequences, four parameters--mean cell center depth, cell radius, cell separation, and standard deviation of direction of storm movement--were varied independently. The maximum 10 peak discharges on the 57-square-mile watershed were determined for a range of values for each parameter by using the functional relation of thunderstorm rainfall to peak discharge described above (Tables 1, 2, 3, and 4). This rainfall-runoff model was developed for predicting peak discharges for major events, and for this reason, only the maximum 10 peak discharges for synthetic and actual data were compared. (Because of large channel abstractions no rainfall-runoff model is yet available for accurately predicting the more numerous lesser events). By this means, the relative sensitivity of the four parameters was determined.

Crude limits on the allowable range of values for the four parameters were indicated using a K-S statistic of .05 for a small sample distribution (Birnbaum and Hall, 1964)(Table 5). Mean cell center depth was the most sensitive parameter, but cell radius and cell separation were sensitive enough to indicate they are necessary in this particular model. Change in standard deviation of direction of cell movement did not affect the K-S statistic significantly, but it can drastically change some storms as Table 4 shows. Obviously, the validity and significance of the model and its parameters cannot be determined entirely with this K-S test.

If a more sophisticated rainfall-runoff model which would be acceptable for predicting peak discharge for all events on Walnut Gulch is developed in the future, a better comparison could be made between the model, and actual data. However, the thunderstorm rainfall model was developed primarily for predicting flood peaks and estimating sediment transport, and "matching" the top ten storm peaks may be a good test of the model for these purposes.

Simultaneously to comparing peak discharges, negative exponential curves of maximum center depths of real and generated (CELTH-6) total storm rainfall were compared. Thus, there were simultaneous checks between maximum center depths through CELTH-6, and between maximum peak discharges through CELTH-5 and the Walnut Gulch rainfall-runoff model. Normal values for the 4 parameters were chosen to best satisfy both tests. As an example, occurrences of maximum storm depths for two 12-year synthetic records, and 12 years of Walnut Gulch data were compared (Figure 3). The accumulated curves are certainly similar and suggest the model and actual data are comparable with respect to maximum storm depth. Other generated 12-year series also seem comparable to the 12-year Walnut Gulch record, but a large number of 12-year records would need to be generated to establish quantitative confidence in the model.

As stated earlier, the degree of uncertainty between observations on a random variable and observations on a single observed sample function and several generated sample functions is great. Necessarily, comparisons

Table 1. Comparison of the 10 greatest peak discharges (cfs) in 12 years of simulated data and Walnut Gulch data (1960-1971) for different values of mean cell center depth.

| Event Rank | Observed Peak Discharge (1960-71) (cfs) | Simulated Peak Discharge (cfs) Mean Cell Center Depth (Inches) .30 | .35 | .425 | .45 | .50 |
|---|---|---|---|---|---|---|
| 1 | 4,700 | 1,990 | 2,820 | 5,000 | 6,770 | 11,000 |
| 2 | 4,680 | 1,570 | 2,040 | 4,400 | 5,730 | 10,000 |
| 3 | 4,030 | 1,480 | 2,020 | 3,340 | 3,950 | 6,240 |
| 4 | 3,600 | 1,420 | 1,970 | 3,100 | 3,680 | 5,480 |
| 5 | 2,980 | 1,260 | 1,710 | 2,740 | 3,230 | 4,560 |
| 6 | 2,870 | 1,170 | 1,680 | 2,380 | 2,810 | 3,950 |
| 7 | 2,770 | 1,140 | 1,520 | 2,320 | 2,760 | 3,920 |
| 8 | 2,640 | 1,130 | 1,440 | 2,130 | 2,460 | 3,830 |
| 9 | 2,120 | 1,060 | 1,340 | 2,110 | 2,410 | 3,770 |
| 10 | 2,010 | 990 | 1,300 | 2,060 | 2,360 | 3,230 |

Table 2. Comparison of the 10 greatest peak discharges (cfs) in 12 years of simulated data and Walnut Gulch data (1960-1971) for different values of cell radius.

| Event Rank | Observed Peak Discharge (1960-71) (cfs) | Simulated Peak Discharge (cfs) Cell Radius (Miles) 2.1 | 2.6 | 3.1 | 3.6 | 4.1 | 4.6 | 5.1 |
|---|---|---|---|---|---|---|---|---|
| 1 | 4,700 | 3,710 | 3,880 | 4,110 | 5,000 | 6,150 | 7,600 | 8,900 |
| 2 | 4,680 | 2,480 | 3,240 | 4,000 | 4,400 | 4,700 | 5,620 | 7,060 |
| 3 | 4,030 | 2,140 | 2,540 | 2,840 | 3,340 | 4,360 | 4,970 | 5,220 |
| 4 | 3,600 | 1,860 | 2,170 | 2,530 | 3,100 | 3,320 | 3,690 | 4,080 |
| 5 | 2,980 | 1,850 | 2,090 | 2,470 | 2,740 | 2,990 | 3,410 | 4,040 |
| 6 | 2,870 | 1,780 | 1,870 | 2,040 | 2,380 | 2,760 | 3,230 | 3,860 |
| 7 | 2,770 | 1,730 | 1,860 | 2,010 | 2,320 | 2,700 | 3,140 | 3,500 |
| 8 | 2,640 | 1,590 | 1,780 | 2,000 | 2,130 | 2,610 | 3,120 | 3,450 |
| 9 | 2,120 | 1,430 | 1,690 | 1,980 | 2,110 | 2,390 | 2,910 | 3,210 |
| 10 | 2,010 | 1,300 | 1,540 | 1,780 | 2,060 | 2,320 | 2,690 | 3,020 |

Table 3. Comparison of the 10 greatest peak discharges (cfs) in 12 years of simulated data and Walnut Gulch data (1960-1971) for different values of mode of distance between cells.

| Event Rank | Observed Peak Discharge (1960-71) | Simulated Peak Discharge (cfs) Mode of Distance Between Cells (Miles) | | | | |
|---|---|---|---|---|---|---|
| | | 0.5 | 1.0 | 2.0 | 3.0 | 4.0 |
| | (cfs) | | | | | |
| 1 | 4,700 | 7,670 | 7,220 | 5,000 | 4,380 | 4,370 |
| 2 | 4,680 | 5,100 | 4,870 | 4,400 | 3,950 | 3,650 |
| 3 | 4,030 | 4,590 | 4,450 | 3,340 | 2,830 | 2,680 |
| 4 | 3,600 | 4,360 | 3,620 | 3,100 | 2,440 | 2,360 |
| 5 | 2,980 | 4,260 | 3,450 | 2,740 | 2,440 | 2,270 |
| 6 | 2,870 | 3,330 | 3,240 | 2,380 | 2,410 | 2,120 |
| 7 | 2,770 | 3,220 | 3,020 | 2,320 | 2,240 | 2,030 |
| 8 | 2,640 | 3,110 | 2,960 | 2,120 | 2,120 | 1,940 |
| 9 | 2,120 | 2,970 | 2,720 | 2,110 | 2,060 | 1,820 |
| 10 | 2,010 | 2,850 | 2,370 | 2,060 | 1,990 | 1,730 |

Table 4. Comparison of the 10 greatest peak discharges (cfs) in 12 years of simulated data and Walnut Gulch data (1960-1971) for different values of standard deviation of change in direction of cell movement.

| Event Rank | Observed Peak Discharge (1960-71) | Simulated Peak Discharge (cfs) Standard Deviation of Direction Change (Degrees) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 15 | 30 | 45 | 60 | 90 | 120 |
| | (cfs) | | | | | | |
| 1 | 4,700 | 5,000 | 4,810 | 4,780 | 5,000 | 14,100 | 7,100 |
| 2 | 4,680 | 4,400 | 4,400 | 4,400 | 4,400 | 6,050 | 4,480 |
| 3 | 4,030 | 3,070 | 3,075 | 3,080 | 3,340 | 5,990 | 3,450 |
| 4 | 3,600 | 2,640 | 2,770 | 2,680 | 3,100 | 4,410 | 3,450 |
| 5 | 2,980 | 2,590 | 2,650 | 2,670 | 2,740 | 3,170 | 3,380 |
| 6 | 2,870 | 2,480 | 2,630 | 2,520 | 2,380 | 3,060 | 3,180 |
| 7 | 2,770 | 2,320 | 2,320 | 2,320 | 2,320 | 3,030 | 2,910 |
| 8 | 2,640 | 2,130 | 2,130 | 2,130 | 2,130 | 2,640 | 2,830 |
| 9 | 2,120 | 2,110 | 2,110 | 2,110 | 2,110 | 2,320 | 2,770 |
| 10 | 2,010 | 1,980 | 1,930 | 1,900 | 2,060 | 2,140 | 2,650 |

Table 5. K-S statistic for 10 greatest peak discharges (cfs) in 12 years for simulated and observed data with various values of rainfall model parameters.

| Mean Cell Center Depth | | Cell Radius | | Mode of Distance Between Cells | | Standard Deviation of Change in Direction of Cell Development | |
|---|---|---|---|---|---|---|---|
| Parameter Value (Inches) | K-S Statistic | Parameter Value (Miles) | K-S Statistic | Parameter Value (Miles) | K-S Statistic | Parameter Value (Degrees) | K-S Statistic |
| | | 2.1 | 0.7 | | | 15 | 0.4 |
| .30 | 1.0 | 2.6 | 0.6 | 0.5 | 0.4 | 30 | 0.3 |
| .35 | 0.8 | 3.1 | 0.5 | 1.0 | 0.3 | 45 | 0.4 |
| .425 | 0.3 | 3.6 | 0.3 | 2.0 | 0.3 | 60 | 0.4 |
| .45 | 0.2 | 4.1 | 0.2 | 3.0 | 0.5 | 90 | 0.3 |
| .50 | 0.6 | 4.6 | 0.4 | 4.0 | 0.5 | 120 | 0.3 |
| | | 5.1 | 0.6 | | | | |

NOTE: Reject at the 0.05 level with a K-S statistic of 0.5 according to Birnbaum and Hall (1960).

were between random variables. Therefore, statements about correspondence of stochastic processes based on a few sample functions must remain qualitative.

## CONCLUSIONS AND OBSERVATIONS

A model of thunderstorm rainfall is being further developed at the USDA, ARS, Southwest Watershed Research Center in Tucson, Arizona. Numerous simplifications and assumptions add uncertainties to the model, and most of these uncertainties cannot be compared directly with actual rainfall data. Testing of the model is limited primarily to comparisons of direct output; i.e., point and areal rainfall depths and volumes, and indirect output; i.e., estimates of peak discharges. Such comparisons (assuming a rainfall-runoff model) indicate that the rainfall model does generate storm depths and volumes of runoff-producing rainfall that generally are comparable to depths and volumes of runoff-producing rainfall from 12 years of record. Although there are many uncertainties in the assumptions and simplifications stated here, any model of thunderstorm rainfall will contain similar uncertainties even if the uncertainties are not spelled out.

Efforts to improve the basic air-mass thunderstorm rainfall model for Walnut Gulch continue, and at the same time the model is being regionalized. Such regionalization must incorporate meteorologic and topographic differences such as available moisture aloft and the effect of significant frontal activity. By starting with a basic model (air-mass thunderstorm rainfall), hopefully, regionalization will be more easily accomplished, since regional variables can be included without having to restructure the basic model.

REFERENCES

Birnbaum, Z. W., and Hall, R. A. Small Sample Distributions for Multi-Sample Statistics of the Smirnov Type. Ann. Math. Stat., 31, 710-720, 1960.

Feller, W. An Introduction to Probability Theory and Its Applications, Vol. 1, 2nd ed., John Wiley & Sons, Inc., New York, 1957.

Osborn, H. B., and Hickok, R. B. Variability of Rainfall Affecting Runoff from a Semiarid Rangeland Watershed. Water Resources Res., AGU, 4(1):199-203, 1968.

Osborn, H. B., and Lane, L. J. Depth-Area Relationship for Thunderstorm Rainfall in Southeastern Arizona. Trans. ASAE, Vol. 15(4): 670-673, 1972.

Osborn, H. B., Lane, L. J., and Kagan, R. S. Stochastic Models of Spatial and Temporal Distribution of Thunderstorm Rainfall. (To be pub. as USDA Bull., Proc. Symp. Stat. Hydrol., Tucson, Ariz., Sept., 1971).

Petterssen, Sverre Weather Analysis and Forecasting. McGraw-Hill, New York, Vol. 2, pp. 156-165, 1957.

Renard, K. G. The Hydrology of Semiarid Rangeland Watersheds. USDA, ARS 41-162, 26 pp. 1970.

Sellers, W. D. The Climate of Arizona. In Arizona Climate, by C. R. Greene and W. D. Sellers, University of Arizona Press, pp. 5-64, 1960.

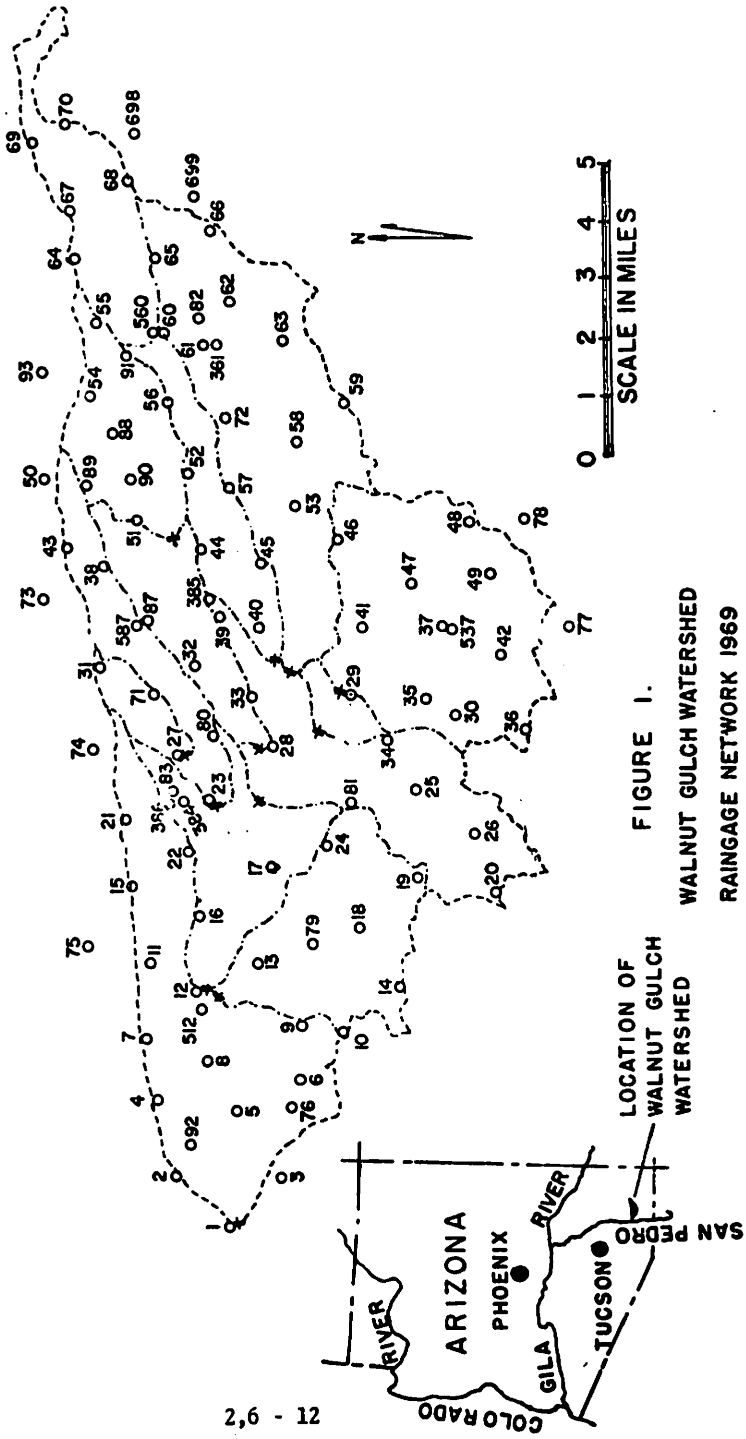
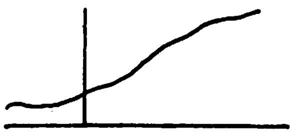



FIGURE I.
WALNUT GULCH WATERSHED
RAINGAGE NETWORK 1969

## RANDOM VARIABLE X

Distribution of X

F(X)



Variable X

PARAMETERS:

Mean ~ Location

Variance ~ Uncertainty

QUANTIFICATION::

Distribution Function?

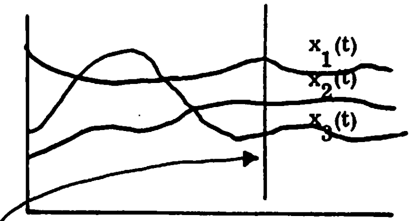
Central Limit theorem (CLT)

Chebyshev Inequality

The data here are
observations on X~
A sample of size N.

## STOCHASTIC PROCESS X

Variable
(value)

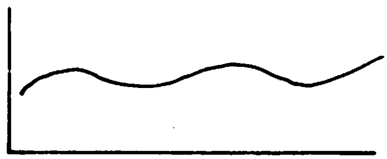



Time (Space, No. of trials)

Given Observations Across the Ensemble
We Have Observations on a Random
Variable.

The Data Here are Observed on the
Process-Sample Function, Usually
There is Only One.

Variable



Time

ASSUMPTIONS:ABOUT PROCESS

Stationary → Markovian?

Non-Stationary

Yes:

Model:
States,
Transition
Probabilities,
Limiting
Distributions
and Laws.

No:

Limit Laws:
Weak Law
Strong Law
CLT (Also applies to
non-stationary processes,
if the non-stationarily can be
quantified).

Figure 2. Sketch illustrating the degree of uncertainty using statistical models: random variables, and stochastic processes.

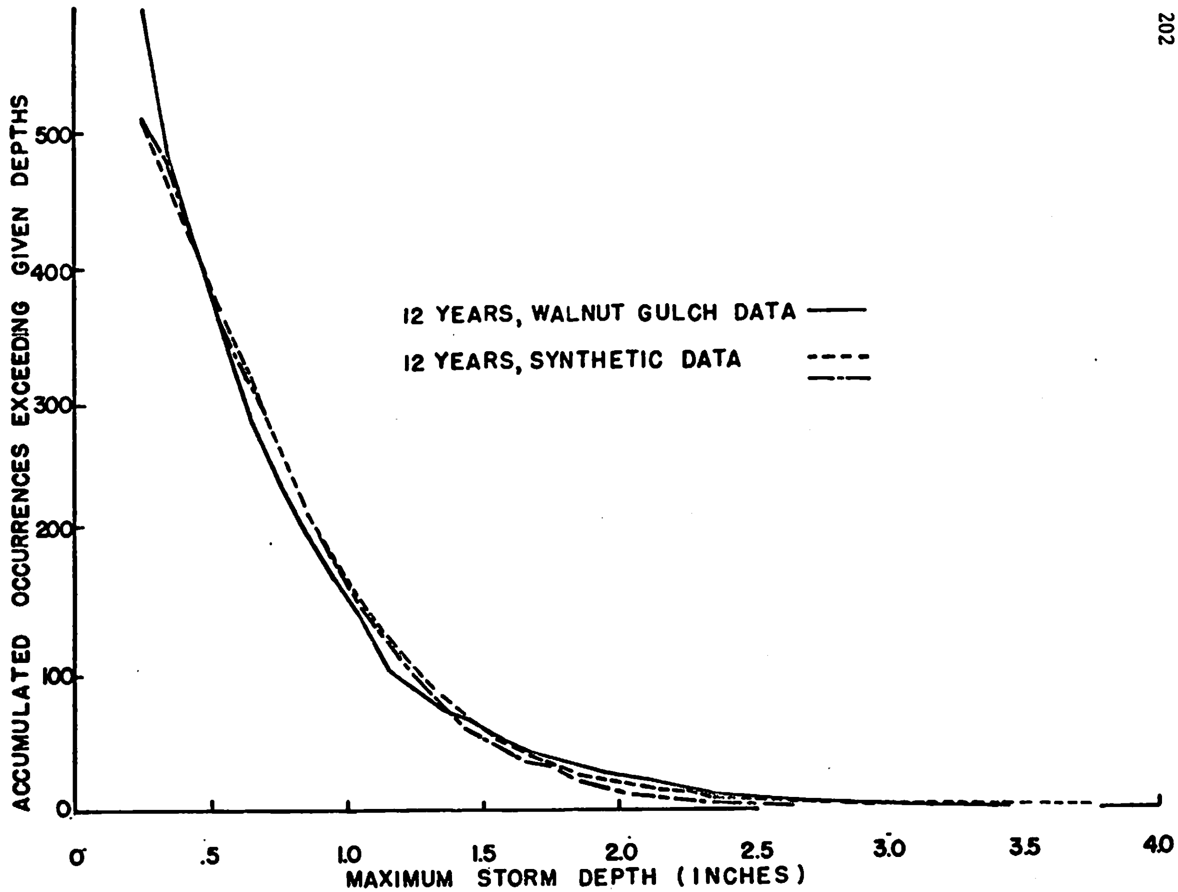



Figure 3. Accumulative occurrences based on maximum storm depths comparing 12 years of synthetic and 12 years of Walnut Gulch data.