FINAL REPORT

Improved Methods for National Water Assessment

Water Resources Contract: WR15249270

Harvard Water Resources Group

Harold A. Thomas Jr.

Principal Investigator

Harvard University

Cambridge, MA 02138

January 31, 1981

D, 141

~ TW Car

Introduction

The purpose of our research is to develop methods to make National Water Assessment more useful in estimating water availability for economic growth and more helpful in determining the effect of water resource development upon the environmental quality of related land resources. There are serious questions pertaining to the 1975 Water Assessment and these amplify the significance of decisions made as to the planning and scheduling of the next assessment.

In the first section of this report we discuss the water-supply adequacy analyis model used in the Second Assessment and outline limitations and assumptions that inhere in its structure and in the computational procedure used. The principal difficulty with the model is that in some regions its accuracy depends strongly on estimates of consumption based on inadequate data. The model provides no independent check on errors in these estimates. Another source of dissatisfaction stems from aggregation. Reported overall flow depletion at the outlet(s) of a subregion may not provide adequate indication of large distortions of the natural regime in parts of the subregion. Some alterations of land use increase stream flow; others cause a decrease. The aggregated model obscures these differences.

In the following sectionsalternative analytical methods are examined. These range from supplementary computations to provide a check on estimates of consumption and of other estimates required by the model of the second assessment, to water balance models that utilize precipitation and other data not used in the Second Assessment. The models incorporate groundwater flow and soil moisture and thus have potential for evaluation of environmental effects of stream diversion, regulation and other changes beyond that inhering in analysis based on streamflow data alone. It should be said perhaps that these water balance models are simple formulations with modest data requirements. They are much less complicated than conventional watershed models such as the Stanford and the Kentucky models. Our models are intended to provide a simple but rational framework for water balance based on monthly or weekly inputs.

Our investigations have utilized data from three river basins (Wisconsin River, Wis.; Little River, Ga.; and the San Pedro River, Ariz.) that were found to meet most of the criteria for selection stated in our research proposal. In particular they satisfy criteria pertaining to diversity of hydrological conditions and geographic location. Data from the Little River on the southeast coastal plain have been especially useful in model construction and validation. The basin of the Little River near Tifton, Georgia is small (540 km^2) but is representative of a large area on the plain with rolling topography with a mosaic of small agricultural fields bounded by forests of mixed evergreen and deciduous trees along the drainage system. An active watershed monitoring program* has been underway for the past twelve years and has produced a valuable collection of data pertaining to stream flow, groundwater elevations, precipitation and other meteorological variates.

<u>Natural Runoff: The Water Adequacy Assessment Model</u>

The problem of estimation of natural runoff has challenged generations of water supply engineers and others concerned with the availability of water for economic development. Despite the importance of the problem it has not

---

*Southeast Watershed Research Program, USDA-SEA, Athens Ga.

often been possible to obtain satisfactory estimates of water availability with which to assess the feasibility of projected demand schedules. Difficulties stem not only from inadequate data but also from the ambiguities of the meaning of natural flow. Drainage systems evolve and in most basins in the United States regimes have been more or less continuously perturbed by man's activities and structures. The building of dams, levees, drainage works, well fields and other facilities for water management, as well as land use changes associated with deforestation, urbanization, etc., alter natural hydrological regimes in a myriad of ways, each with implications for environmental quality and for further development.

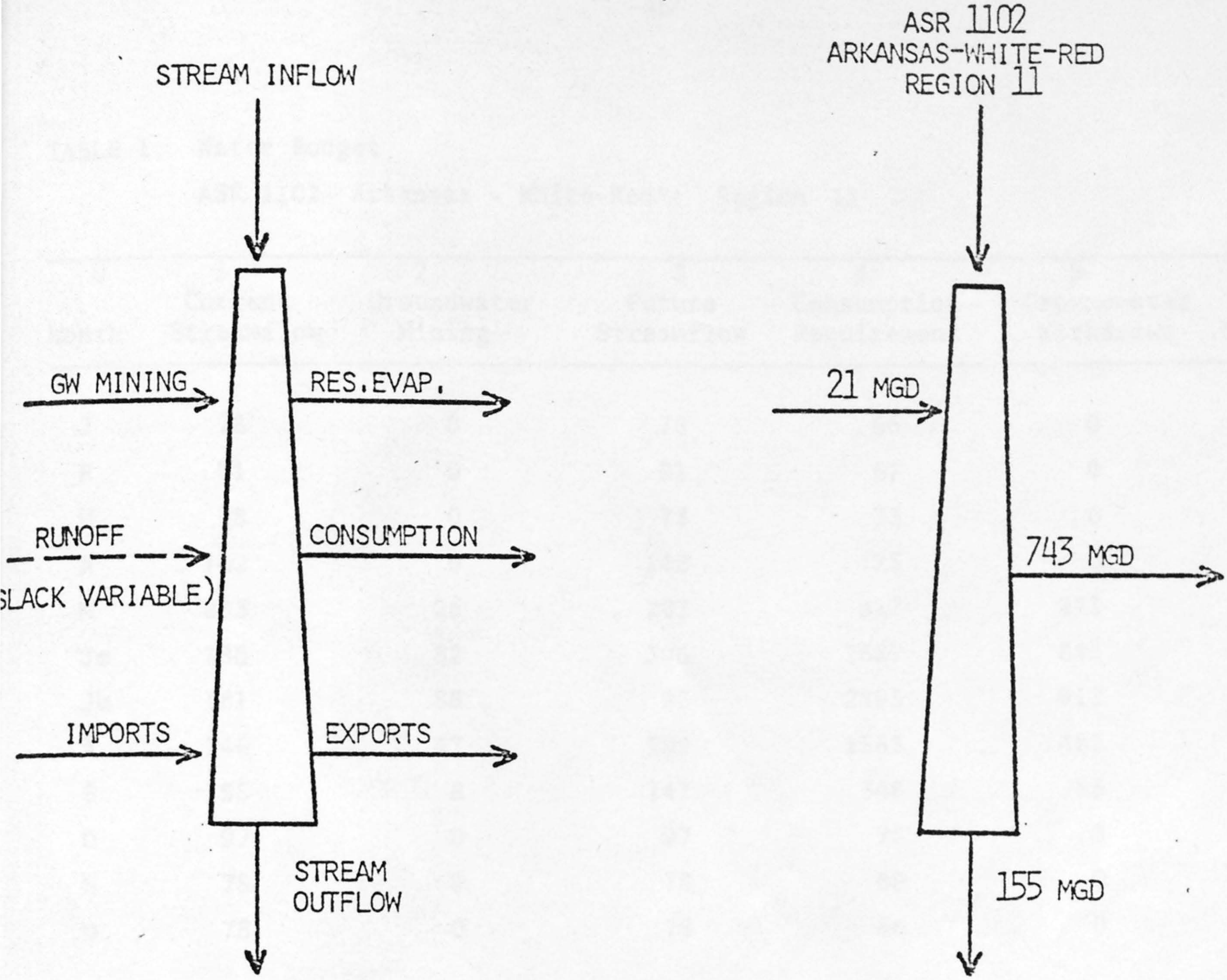
In the usual computation natural runoff at a point is calculated from monthly streamflow observed at that point plus monthly net evaporation from reservoirs and other man-made water surfaces plus transbasin exports minus transbasin imports plus water mined from aquifer plus change in consumptive use from agricultural lands. Also estimates may be adjusted to take account of changes in reservoir storage during the month.

The water-supply adequacy analysis model of the 1975 assessment which was used in the 106 water resources subregions of the nation (Figure 1a) accords closely with the conventional approach. The model balances supply and use and estimates deficiency in natural flow (potential supply). Water input to the subregion consists of inflow from subregions upstream, runoff from rainfall in the subregion, imports to the basin and groundwater mining. Water uses include consumptive losses from diverted streamflow and pumped groundwater, exports, and net evaporation from reservoirs and artificial lakes. It should be noted that runoff within the subregion, which is defined as precipitation minus natural evapotranspiration minus groundwater recharge is not estimated explicitly in the calculations - rather it is treated as a slack variable in the equation to honor the mass balance. Thus while the water-supply adequacy analysis model equates supply and use, it does not involve a water balance.

To illustrate a typical application, data for Subregion 1102 of the Arkansas-White-Red Region are shown in Figure 1 and Table 1. The subregion is in the southeast quadrant of Colorado and has a drainage area of 66,000 km^2. The Subregion has no inflow from upstream subregions, the assessment includes no import, export or evaporation from man-made lakes. The water budget for the subregion shown in Table 1 lists average streamflows by month (1951-76) from the record of the USGS gage on the Arkansas River near Coolidge, Kansas. Thus the outflows represent averages of observed runoff over a 25-year period when irrigation use increased rapidly. Mined groundwater (Column 2 of Table 1) is estimated at about 10 percent of groundwater withdrawn (Column 5) during the growing season. And groundwater withdrawn during May through September is in turn taken to be about 31% of the total consumptive requirement (Column 4). This requirement is based presumably on estimated irrigated area in the base year with estimated consumptive loss coefficients appropriate to crops and region. Estimates from the water budget are used in computations of overall depletion of the subregion as shown on Figure 1. Before turning to a discussion of new analytical methods that may be useful in water availability assessment, it will be helpful perhaps to comment briefly on the components of the supply-use model of the 1975 assessment and also to discuss some factors and processes that it does not include.

<u>Stream Outflow</u>

Base-year flow is usually estimated from average (arithmetic mean) monthly and annual flows of the record at a gage(s) adjusted by a coefficient based on drainage areas to apply at the outflow point(s). The use of observed mean flows in the computation will cause depletion to be



(1A) (1B)

$$\text{TOTAL STREAM FLOW} = 155 + 743 - 21 = 877\,\text{MGD}$$

$$\text{CURRENT DEPLETION} = 100\,\frac{743}{898} = \underline{\underline{83\%}}$$

$$\text{FUTURE DEPLETION} = 100\,\frac{743}{877} = \underline{\underline{85\%}} \quad \text{(CESSATION OF MINING)}$$

FIGURE 1. WATER SUPPLY ADEQUACY ANALYSIS MODEL. CALCULATIONS FOR AGGREGATED SUBREGION 1102.

TABLE 1: Water Budget

ASR 1102 Arkansas - White-Red*: Region 11

0 Month	1 Current Streamflow	2 Groundwater Mining	3 Future Streamflow	4 Consumption Requirement	5 Groundwater Withdrawn	6 Total Streamflo
J	78	0	78	66	0	144
F	91	0	91	67	0	158
M	78	0	78	73	0	151
A	142	0	142	75	0	217
M	233	26	207	917	273	1124
Je	388	82	306	2687	845	2993
Ju	181	88	93	2895	912	2988
A	246	47	199	1583	483	1782
S	155	8	147	348	86	495
O	97	0	97	75	0	172
N	78	0	78	69	0	147
D	78	0	78	66	0	144
Average	155	21	134	743	217	877

Column 3 = Column 1 - Column 2

Column 2 ≑ 0.1 Column 5 (May-Sept)

Column 4 Exogenous

Column 5 ≑ 0.31 x Columm 4 (May-Sept)

Column 6 Column 3 + Column 4

---

* Southeastern Colorado

Drainage Area Arkansas River near Coolidge, Kansas, 25,410 mi^2.

Mean Annual runnoff 1951-1976, 153.6 mgd.

underestimated if, as often is the case, consumption has increased during the period of record. Beard[1] has recommended a computational method that in principle would eliminate this source of error. His method requires that each monthly flow of the record be adjusted to a value that would obtain with the levels of regulation and diversion of the base year of the assessment. Beard's proposal applied to the 106 subregions would require a substantial effort even if implemented to include only the major effects of increments of diversion and regulation on the spatial and temporal distribution of flow. The practicality of this scheme hinges on the quality of data on the historical growth of storage and diversion and on the quality of estimates of consumptive loss (the difference between water diverted from the stream and return flow above the gage). The quality of the data base differs widely from region to region and this militates against use of uniform assessment methods. The merit of Beard's proposal should be examined in relation to the merits of other proposals for improving water availability assessment suggested in this report.

Evaporation

In the water supply adequacy analysis model of the 1975 Assessment net monthly evaporation from artificial lakes is usually estimated as reservoir evaporation minus precipitation. The former is estimated from pan evaporation observations applying an appropriate coefficient; rainfall is measured near the shoreline of the impoundment. This procedure introduces error since conceptually net evaporation should be defined as the difference between reservoir evaporation and evapotranspiration losses that occurred before dam construction.

Groundwater Mining

This component cannot be estimated satisfactorily in many regions without an extensive investigation of the entire aquifer system - which may underlie several adjoining subregions. While many of the large aquifers of the nation have been investigated and records of withdrawal are available for major well fields, most withdrawal estimates must be based on general land-use data. Groundwater depression cones about well fields spread slowly for many years over wide areas and pumping lifts may become large before a reasonably accurate assessment can be made of the rate of mining (withdrawal minus natural recharge). Assessment is difficult in some regimes because recharge parameters change significantly as depression cones spread.

Consumptive Use

Estimates of consumptive losses over large areas are subject to considerable error. The common procedure is to use regional meteorological and climatological data as input to regression equations for evapotranspiration derived from lysimeter tests and other small scale experiments. Accuracy depends not only on the validity of the extrapolation but also on correct information pertaining to land use, cropping patterns and irrigation practices. Losses from spray irrigation exceed those of ditch irrigation. And losses from day-time spraying exceed those of night-time spraying. In regimes where supplementary irrigation is important and also in metropolitan areas consumptive losses are large during extended hot and dry weather not only because evaporation rates are high but also because water-use increases. In some cases such as large power generating plants with evaporative cooling records of withdrawal and return flow provide data for reasonably accurate estimates of loss. However, in several parts of the country errors in estimating consumption over the entire subregion are so large as to impair the validity of water availability calculations. Where these errors are large refinements in the measurement of other model components may be unwarranted.

Processes not included in Assessment Model

(1) Changes in land use such as urbanization, deforestation, reforestation, dry land farming, channel straightening and swamp drainage modify runoff characteristics but these effects are not explicitly taken into account in estimation of natural flow or assessment of future water availability.

(2) With increases of diversion and regulation of the streams within a subregion, channel losses are altered. Loss alteration occurs also in stream reaches below hydroplants used for peaking. In some situations it is possible to obtain good estimates of water losses concomitant with these changes but the effect is usually ignored.

(3) Water transfer rates at the groundwater-surface water interface may be changed by man-made works. Rates of percolation (influent streams) and of infiltration and seepage (effluent streams) can be altered considerably by facilities such as dams, levees, and well fields. Evaluation of these perturbations is difficult, and their effects are usually not taken into account.

(4) The pattern of local precipitation may be changed substantially with the growth of cities and large reservoirs and irrigation projects but the effect is omitted in the calculations of natural flow.

In summarizing the discussion of the water availability analysis model of the 1975 assessment it should not be concluded that because of the potential for error in estimation of several of its components, and because of the factors and processes not explicitly included in the formulation, it is incapable of yielding useful evaluation of water availability. The model has the capability but successful application requires accurate and detailed information relating to trends in land use and water management in the subregion as well as extensive data pertaining to topography, aquifers, soils, vegetative cover and climate. Moreover, reliable water availability projections are contingent upon a substantial effort by competent hydrologists and other experts with experience in the region. The principal sources of dissatisfaction with the analyses are (i) the lack of uniform methods of error evaluation; and (ii) the aggregated nature of the analytical results. From the tabulated results of the Second Assessment it is difficult to distinguish high quality analysis in some subregions from perfunctory work in others. The National Assessment was not designed to generate detailed information relevant to the design and management of water projects. However, beyond the project level responsible officials in some states have found the aggregated data of the Assessment to be not very helpful in planning adaptive and remedial programs and in allocating investment for more efficient water use and enhancement of environmental quality.

Our research has been directed to surmounting or at least mitigating these imperfections of the Second Assessment. Our approach has been to try to find ways of using the data of conventional hydrological and meteorological times series more effectively so as to provide information for error evaluation and for describing hydrological trends in a more meaningful way. We have also been concerned with the potential utility of non-conventional hydrological data sources and instrumentation - remote sensing, radar, neutron soil-moisture probes, etc.

We believe that the number of water accounting units used in the 1975 Assessment (106 subregions) provides a satisfactory level of disaggregation for national assessment of water availability and our models have been adapted to this scale. For areas of this size we now examine tentative methods for improved analyses of hydrological data to assay possibilities for better assessment - methods that rely more on objective physical measurements and less on subjective judgment and opinion.

Trend Analysis of Annual Flows

Gaging station records used to determine subregion outflows were routinely examined for discernible trends in the 1975 Assessment. The standard technique of regressing the annual flows (or log flows) on time did not offer much information of interest or value. This outcome was not unexpected. Variances of annual flows in most streams of the United States are large in relation to common rates of flow depletion or augmentation associated with development, and as a result estimated trendline slopes have a high level of uncertainty. The inability to measure directly the impact of development on streamflow has been noted in countless reports on water supply and local investigations of water availability. Indeed the high degree of uncertainty associated with results of trend analysis is the main reason for use of models such as that of the 1975 Assessment which are heavily dependent upon general land use information and expert opinion.

All effects of man's activities and watershed changes that perturb natural regimes are enciphered in changes in the spatial and temporal distribution of runoff. The difficulty lies in deciphering the message. It is pertinent to look into the matter more closely and to ask whether forms of statistical analysis other than conventional trend analysis can be deployed to extract information from stream-gaging records to check or otherwise supplement results obtained with the water adequacy analysis model of the 1975 Assessment. We start by turning to an example.

Annual flows (1940-62)* of the Blue River in the Kansas City metropolitan area are presented in Table 2 and are plotted in Figure 2. The slope of the trend line obtained by method of least squares is - 0.26 cfs/yr, but the computed 90% confidence limits for the population slope are broad ($\beta = 0.26 \neq 4.88$) as shown by the dashed lines of Figure 2. As usual the conventional analysis of annual flow data does not yield any significant indication of changes in land use. Yet this basin was exposed to major impacts -- intensive urbanization and population growth**, massive investment in social overhead capital, and diversified industrialization -- during two decades of sustained postwar economic development. In examining the data plotted on Figure 2 and the residuals of the last column of Table 2, it may be noted that much of the uncertainty pertaining to the trend line slope derives from high flows such as those of 1945, 1951 and 1961 and to low flows such as occurred in the mid 1950's. It is likely that other streams in the region outside the metropolitan area would exhibit the same effects from the same set of high and low runoff years. The question arises as to whether a form of multivariate statistical analysis of a set of regional records can be devised to eliminate or reduce regional fluctuations (hydrological "noise") and whether the filtered series for the stations would reveal statistically significant patterns indicative of the scale and character of impacts of development patterns on the individual basins.

Much information relating to the cross-correlation of streamflows is available in nearly all parts of the nation. However, these data were developed for other purposes. Regional analysis of streamflow has been used to establish generalized regional runoff characteristics for

---

* Gage Lat. 38° 57' 25"; long 94° 33' 32"; 1.7 mi. southeast of Kansas City, Drainage area, 188 mi^2; slope 12.4 ft/mi.

** Kansas City, SMSA increased by about 15% during the intercensal years 1960-70.

TABLE 2: Annual Flows: Blue River Basin: Kansas City SMSA. (1940-1962)

t	Water Year	Flow cfs	Residual* cfs
1	1940	77	-65
2	1941	125	-17
3	1942	189	47
4	1943	133	-8
5	1944	165	24
6	1945	270	129
7	1946	76	-64
8	1947	220	80
9	1948	137	-3
10	1949	148	8
11	1950	162	23
12	1951	290	151
13	1952	133	-6
14	1953	22	-117
15	1954	23	-115
16	1955	92	-46
17	1956	12	-126
18	1957	56	-82
19	1958	231	94
20	1959	65	-72
21	1960	98	-39
22	1961	285	148
23	1962	194	58
	Mean	139	
	Standard dev.	83	

* Residuals from linear trend line: $x = 142.3 - 0.26\ t$, cfs.

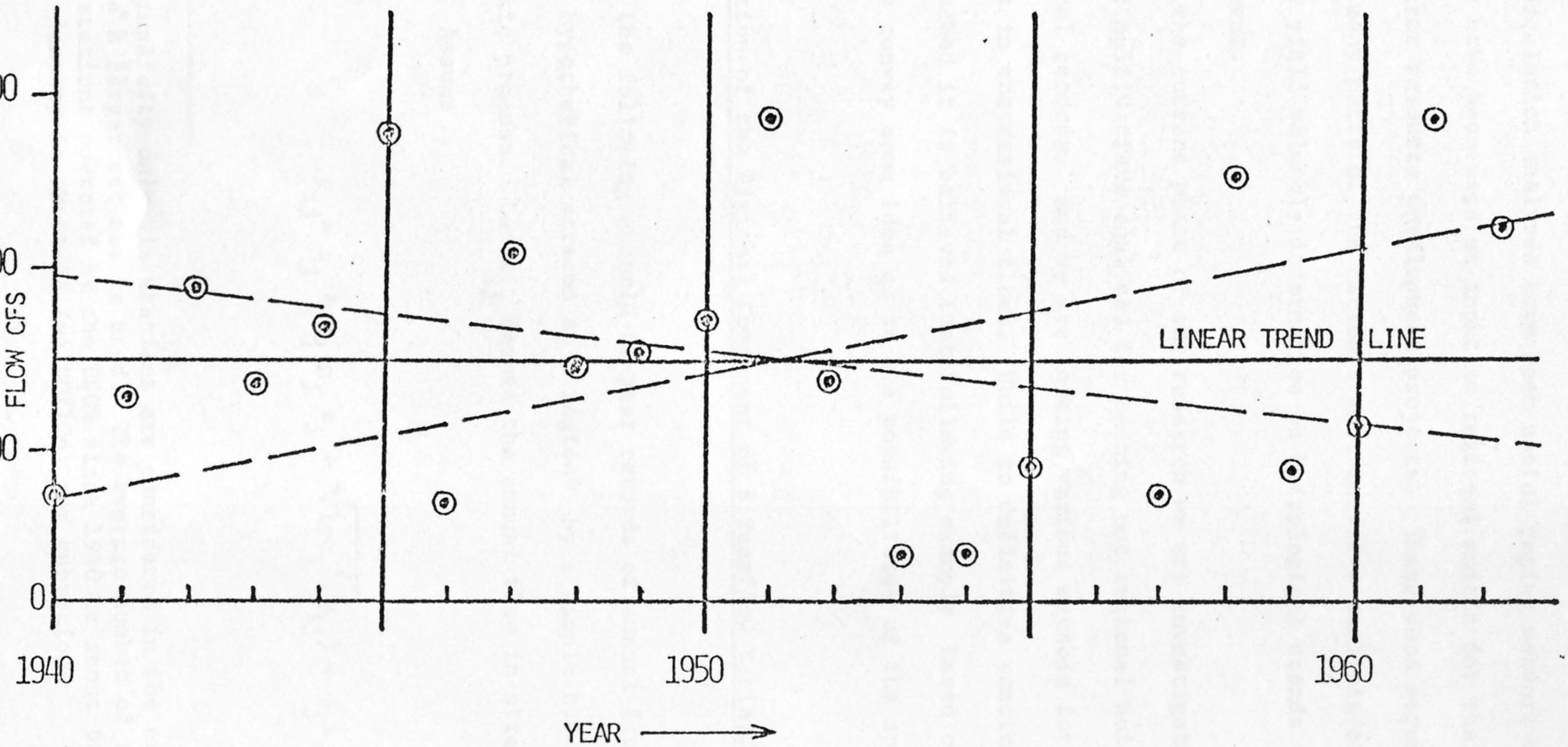


FIGURE 2 ANNUAL RUNOFF: BLUE RIVER NEAR KANSAS CITY 1940-62

LINEAR TREND LINE $X = 142.3 - .260\ (T-1939)$

$R^2 = 0.0004$

DASHED LINES: 90% CONFIDENCE LIMS FOR TRUE SLOPE

$\beta = -.26 \pm 4.88$, CFS/YR.

application to ungaged areas or areas with short or incomplete records; cross correlation analyses have been useful gaging metwork design studies; and they have been used as input to regional models for the design of large water resource development projects. These data strongly suggest that in many parts of the nation a more incisive analysis of streamflow data may yield valuable information on hydrological trends caused by development.

In the current phase of our research we are investigating different types of multivariate analyses for netting out regional noise from individual records. And we are testing various methods for detecting patterns in the residual flows. While no definitive conclusions have been reached it is believed that following example based on synthetic data may convey some idea as to the possibilities of the approach.

Elimination of the Regional Component of Streamflow Variation

In the following example 25-year records of annual flows are generated for six hypothetical streams of a region* by a simple but realistic stochastic process. Let x_{ij} denote the annual flow in stream j during year i. Assume

$$x_{ij} = \mu_j + \sigma_j(\alpha_j z_i + \sqrt{1-\alpha_j^2}\, u_{ij}) + \delta_j i \tag{1}$$

---

* For simplicity only six stations are considered in the example. In practice a larger set may be used. The average number of active complete gaging stations operated by the USGS since 1940 is about 6100. Thus station density is about 58 (=6100/106) per subregion.

where μ_j and σ_j are the mean and standard deviations of natural annual flows at station j, and z_i and u_{ij} are standard normal variates. The last term of the equation represents the effect of man-induced regime modification and land use changes. The parameter δ_j is the annual rate of augmentation or depletion of the natural flow of streams. Variate z_i which is the same for all streams during the year i, represents the regional contribution to the variance of stream flows. It is presumed to be a Markov lag-one variate. That is

$$z_i = \rho z_{i-1} + \sqrt{1-\rho^2}\, e_i . \qquad \text{(1a)}$$

Here ρ is a serial correlation coefficient and e_i is a standard normal variate independent of z_{i-1}. The variate u_{ij} represents local variation particular to each stream each year. It is taken to be independent of z_i and t_i. Further it is assumed that $E(u_{ij}\, u_{i+1,j}) = 0$ and $E(u_{ij}u_{ik}) = 0$. This source of variance has no serial correlation or cross-correlation with other streams. The parameter α_j accounts for the partition at each station between regional and local sources of flow variation. The cross-correlation coefficient between annual flows of streams j and k, $\rho(x_j x_k)$, is $\alpha_j\alpha_k$. In the example the following parameter set is assumed:

Stream No.	μ mgd	σ mgd	α	δ mgd/yr	
1	1000	300	0.97	0	
2	200	56	.95	0	
3	600	174	.96	-12	$\rho=\rho(z_i,z_{i-1})=0.25$
4	800	248	.97	0	
5	1200	360	.94	0	
6	400	128	.96	0	

Stream #3 is used as a test case. Its flow is reduced at a fixed rate of 12 mgd each year to represent the effect of a linear growth of consumptive loss due to withdrawal use. With the consumptive loss the population mean of stream #3 is

$$\mu_3(t) = 600 - 12t \text{ mgd} \qquad 1 \leq t \leq 25$$

The mean flow is depleted to fifty percent of the initial value over the twenty-five year period. Table 3 shows a typical simulation of the six streamflow together with means and standard deviations. It may be noted that years 4 and 14 were unusually wet and years 9 and 11 unusually dry.

The objective of the analysis is to determine how accurately the sustained depletion of Stream #3 can be detected in relation to trends in the other streams which are modeled as natural flows. The following is a simple method of separating (or netting out) regional variation from each of the individual series.

(i) The first step is to fit linear trend lines to the data for each station. Results of least squares regression are shown in Table 3a. The coefficients of determination of all regressions are low indicating that as usual no useful information pertaining to time trends has been extracted by ordinary trend analysis. The uncertainty pertaining to the true slopes is indicated by the large differences between the upper and lower 95% confidence limits shown in the table. In this simulation the random number generator produced a z-sequence that caused a rather high trend of increasing flows in all streams. Such trends - positive and negative - occur in natural streams. The point is that from this conventional step of analysis it cannot be ascertained whether the indicated trend is due to natural fluctuation in stream flow or to systematic depletion or

TABLE 3 Six Streamflow Series: 25 yr. Simulation (flows in mgd)

Yr	x_1	x_2	x_3	x_4	x_5	x_6
1	884	147	426	659	815	308
2	718	141	402	532	865	213
3	743	153	444	617	1124	309
4	1428	292	781	1150	1643	644
5	972	191	493	820	1237	463
6	908	226	560	827	1522	472
7	684	156	391	629	879	296
8	1114	236	686	947	1410	533
9	531	87	163	340	351	120
10	664	151	328	460	937	268
11	579	113	180	355	405	162
12	823	174	351	527	1104	314
13	1038	176	326	686	1177	381
14	1389	252	741	1121	1895	566
15	927	229	388	829	1186	321
16	687	132	209	587	876	348
17	750	153	383	662	990	316
18	638	149	247	684	850	326
19	1116	232	359	962	1184	471
20	1313	251	544	1016	1631	512
21	865	166	274	818	1224	398
22	1181	208	453	854	1148	480
23	1066	177	290	713	1327	390
24	1121	249	362	865	1616	423
25	1343	278	511	1162	1555	578
Mean	936.8	1888.8	411.6	754.9	1158.0	384.5
Std.Dev.	266.9	53.5	161.0	227.1	374.0	130.5

Table 3a. Regression Results

ariate	Coef. Determ. r^2	Intercept	Slope	95% Conf. Lims. for slope*, mgd/yr lower	upper
x_1	.108	782.0	11.9	-3	27
x_2	.084	161.4	2.1	-1	5
x_3	.058	480.1	-5.3	-14	4
x_4	.108	623.4	10.1	-2	23
x_5	.084	966.6	14.7	-6	36
x_6	.074	322.0	4.8	-3	12

or augmentation. Many undisturbed streams exhibit upward or downward flow fluctuations over 25-year or 50-year periods. For such periods these random oscillations cannot be distinguished from permanent trends. For example, if the five low-flow years (1953 to 1957) of the flow record of the Blue River shown in Table 2 had chanced to come at the beginning or end of the record the trend line slope would differ greatly from that calculated. Indeed even with records of 50 years the location of the largest annual flow may have a significant effect on the tilt of the trend line. This holds for streams both in developed and undeveloped areas. In comparing flow trends between stations within a subregion it is useful to adjust for regional "tilt" by rotating each flow sequence about its centroid so as to make the average slope zero. Because of differences in stream size the slope of each trend line is normalized on the standard deviation of the annual flow.

(ii) The second step of the algorithm, the tilt adjustment, therefore is to compute the average value of b_j/s_j for the streams and to rotate each

---

* $\beta = b \pm t_{95}\, b \sqrt{\frac{1-r^2}{r^2(n-2)}} = b\left[1 \pm .431\sqrt{r^{-2}-1}\right]$

where b is the estimated trend line slope, and t_{95} (=2.069) is Student's t with 23 degrees of freedom.

sequence according to the following equation,

$$x_{ij} = x_{ij} - s_j \bar{b}'(i-13) \qquad (2)$$

where $$\bar{b}' = \frac{1}{6}\sum(b_j/s_j) \; * \qquad (2a)$$

The rotated flows are then regressed on time. The fitted parameters of these regressions are as follows:

Table 3b. Regression Results

Variate	Coef. Determ. r^2	Intercept mgd, a	Slope mgd/yr b	Std. error of estimate mgd $s\sqrt{1-r^2}$
X_1	.015	881.6	4.25	252.1
X_2	.007	181.3	.58	51.2
X_3	.178	540.2	-9.89	156.2
X_4	.015	708.1	3.60	214.6
X_5	.007	1106.3	3.98	358.3
X_6	.004	370.7	1.06	125.6

* A comparison of the numerical values of the tilt, $\bar{b}'$, of the subregions of a region and a comparison of the average regional tilt of the regions of the nation would provide information of value in evaluating results of national assessment. Tilt reflects both man-induced flow changes and long random oscillations about the mean. The serial correlation coefficient, ρ, of equation (1a) is intended to model the latter phenomena. A statistical analysis of a set of 251 representative stream flow records for the period 1942 to 1971 is presented in Appendix A. The basic assumption of the Water-Supply Adequacy Analysis Model of the 1975 assessment was that past streamflow at the subregion outlet with appropriate adjustments would represent future streamflow potential. The validity of this assumption can be assessed by reference to statistical summaries for the nation such as that of Appendix A. An analysis by Langbein and Slack (2) of long-term records is of considerable interest in this connection.

The purpose of the "tilt" adjustment in our algorithm for the elimination of regional hydrological noise is to make a conceptual separation between oscillations about the mean and those caused by permanent regime alterations with development

The residuals of the regressions normalized on the standard error of estimate,

$$v_{ij} = \frac{x_{ij} - a - bi}{s_j\sqrt{1-r_{ij}^2}}, \qquad (3)$$

are tabulated in Table 4.

(iii) The third step of the algorithm is the estimation of the coefficients, α_j, from cross-correlation analysis. The cross-correlation coefficients, $\hat{\rho}(v_{ij}, v_{ik})$, for the six streams are shown in the following matrix.

6 Streams: Matrix of Zero Order Correlation Coefficients

Stream No.	1	2	3	4	5	6
1		.9091	.9090	.9200	.8709	.9112
2			.9119	.9369	.8955	.9000
3				.9277	.9006	.9171
4					.8777	.9511
5						.8830

With N streamflow records there N(N-1)/2 cross correlations and for each of these the relation $\rho_{jk} = \alpha_j\alpha_k$ obtains. The α_j may be estimated by a least squares solution of the M = N(N-1)/2 linear equations,

$$ln\ \hat{\alpha}_j + ln\ \hat{\alpha} = ln\ \hat{r}_{jk}$$

With N=6 the normal equations associated with minimization of

$$\sum (ln\ \hat{\alpha}_j + ln\ \hat{\alpha}_k - ln\ r_{jk})^2$$

TABLE 4: Six Streamflow Series: Residuals of Linear Regression

Yr	v_1	v_2	v_3	v_4	v_5	v_6	Z
1	.36	-.31	-.31	.12	-.46	-.15	-.14
2	-.35	-.47	-.43	-.52	-.37	-.95	-.54
3	-.30	-.29	-.13	-.17	.32	-.22	-.13
4	2.37	2.39	2.06	2.27	1.72	2.41	2.31
5	.52	.37	.25	.68	.55	.93	.58
6	.22	1.02	.71	.66	1.30	.97	.86
7	-.72	-.39	-.33	-.30	-.53	-.47	-.48
8	.94	1.13	1.59	1.13	.91	1.37	1.23
9	-1.42	-1.83	-1.73	-1.75	-2.09	-1.95	-1.88
10	-.94	-.62	-.64	-1.23	-.49	-.81	-.83
11	-1.56	-1.40	-1.55	-1.77	-2.02	-1.69	-1.75
12	-.40	-.24	-.42	-.78	-.11	-.52	-.43
13	.40	-.25	-.55	-.32	.05	-.03	-.12
14	1.75	1.20	2.14	1.66	2.02	1.41	1.78
15	-.13	.71	-.08	.25	-.01	-.58	.03
16	-1.13	-1.23	-1.19	-.92	-.91	-.40	-1.01
17	-.93	-.87	-.05	-.62	-.63	-.70	-.67
18	-1.42	-.98	-.89	-.57	-1.06	-.66	-.98
19	.43	.60	-.14	.68	-.17	.46	.32
20	1.16	.93	1.08	.89	1.03	.75	1.02
21	-.66	-.78	-.61	-.08	-.14	-.20	-.43
22	.54	.00	.57	.04	-.40	.41	.20
23	.04	-.63	-.44	-.67	.06	-.34	-.34
24	.21	.72	.05	-.01	.83	-.12	.30
25	1.05	1.25	1.04	1.33	.62	1.08	1.11

are

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 5 & 1 & 1 & 1 & 1 \\ 1 & 1 & 5 & 1 & 1 & 1 \\ 1 & 1 & 1 & 5 & 1 & 1 \\ 1 & 1 & 1 & 1 & 5 & 1 \\ 1 & 1 & 1 & 1 & 1 & 5 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\alpha}_5 \\ \hat{\alpha}_6 \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{pmatrix} \tag{4}$$

where $S_j = \sum' \ln r_{jk}$ and where Σ' denotes summation over all $k \neq j$. Solution of equation (4) yields the following estimates of the α_j.

Stream :	1	2	3	4	5	6
$\hat{\alpha}_j$:	.9476	.9563	.9598	.9718	.9235	.9585

(iv) The fourth step is the estimation of z_i and the filtering of regional noise.

$$\hat{z}_i = \frac{1}{6} \sum v_{ij}/\hat{\alpha}_j \tag{5}$$

Calculated values are displayed in the last column of Table 4. Filtered annual flows are calculated using the following equation:

$$\tilde{x}_{ij} = x_{ij} - S_j \sqrt{1-r_{ij}^2}\ \hat{\alpha}_j \hat{z}_i \tag{6}$$

(v) The last step of the algorithm is regression of the filtered streamflow and calculation of confidence limits for trend line slope. Results are shown in the following tabulation:

Variate	Coef.Determ. $\tilde{r}^2$	Intercept, mgd $\tilde{a}$	Slope, mgd/yr $\tilde{b}$	95% Conf.Lims for slope mgd/yr Lower	Upper
$\tilde{x}_1$	.156	881.6	4.2	0	9
$\tilde{x}_2$	.086	181.3	.6	-.2	1.4
$\tilde{x}_3$	.761	540.2	-9.9	-12	-8
$\tilde{x}_4$	.224	708.1	3.6	1	6
$\tilde{x}_5$	.056	1106.3	4.0	-3	11
$\tilde{x}_6$	.053	370.0	1.1	-1	3

The computations show all confidence bands have been narrowed substantially with removal of the regional component of flow variation and the depletion of Stream #3 is clearly identified. It is pertinent to note that the difference between the trend line for Stream #3 and the average slope of the other five streams is -12.6 mgd/yr, which is close to the 12 mgd/yr used in the simulation. Results of the analysis are shown in Table 4a and Figures 3 and 4, which indicate flow traces for Stream #3 before and after removal of regional noise and tilt. The values of $\tilde{x}_{3j}$ for Stream #3 plotted in Figure 4 show much reduced scatter about the least squares line. The remaining dispersion of the points about the line is due (i) to local variation (last term of Eq. (1)); and Iii) to approximations associated with the method used for netting out regional noise. The dashed lines of Figure 4 depicting the 95% confidence limits for the true slope show close adherence to the calculated (solid) line.

While no definitive conclusions can be drawn from the analysis of this didactic example, it would appear that the technique of eliminating regional

TABLE 4a: Streamflows with Removal of Regional Tilt and Noise

Yr	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$	$\tilde{x}_6$
1	1008	173	502	765	989	369
2	930	185	534	716	1161	319
3	832	175	511	710	1276	363
4	945	193	476	727	975	400
5	895	175	444	752	1132	424
6	757	195	464	694	1314	395
7	845	189	492	769	1103	377
8	857	183	524	722	1055	403
9	1012	185	464	759	1017	362
10	885	196	466	652	1243	379
11	952	202	451	733	1006	381
12	934	197	421	673	1258	370
13	1066	182	344	711	1217	395
14	956	163	469	743	1295	348
15	906	225	375	811	1155	310
16	906	177	348	779	1180	459
17	879	179	464	775	1167	381
18	834	189	371	856	1120	425
19	993	207	283	857	1013	410
20	1015	190	358	757	1217	363
21	907	175	302	856	1281	420
22	1064	184	381	753	985	422
23	1072	179	295	719	1333	394
24	965	217	266	731	1399	346
25	986	205	289	852	1059	400
Mean	936.8	188.8	411.6	754.9	1158.0	384.5
S.Dev.	79.2	14.4	83.4	56.0	123.7	34.1

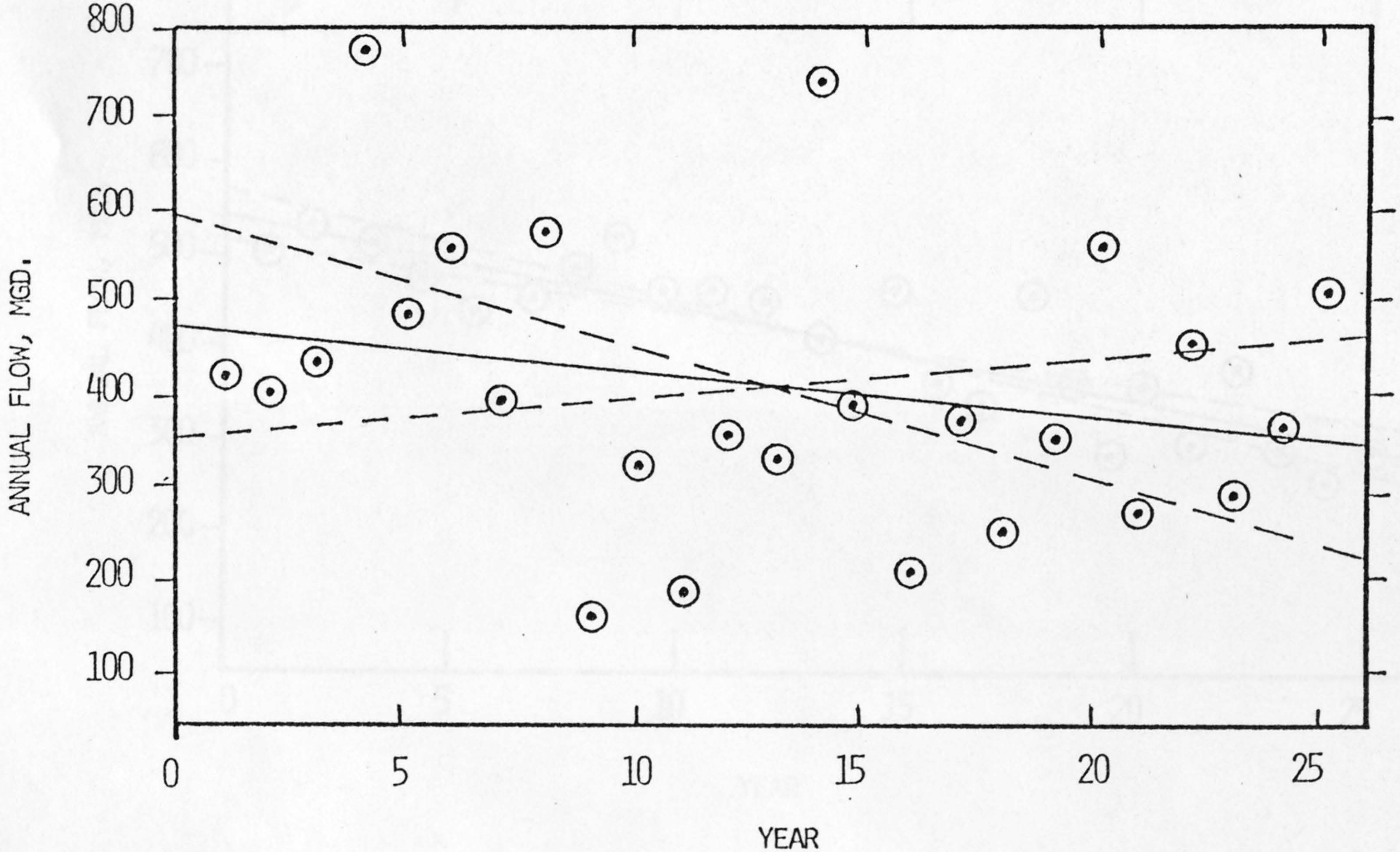


FIGURE 3 STREAM #3, ANNUAL FLOWS

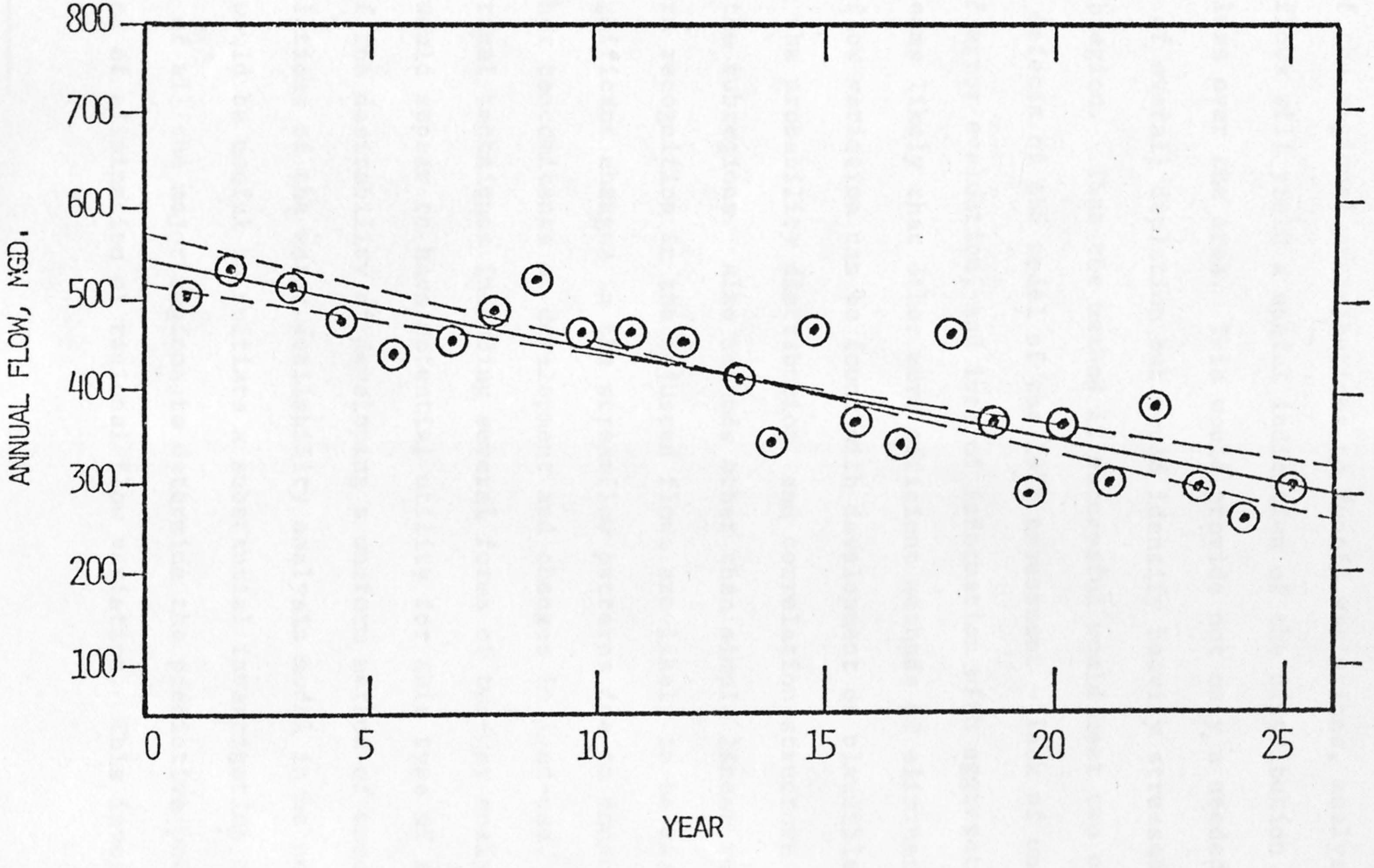


FIGURE 4 STREAM #3. ANNUAL FLOWS. REGIONAL NOISE AND TILT REMOVED

contributions to flow variation of the streams of a subregion may provide a useful external check on estimates required by the water availability analysis model regarding consumptive losses in different parts of the area. Hopefully with removal of the regional contributions to runoff variations, analysis of adjusted flows will yield a useful indication of the distribution of consumptive loss over the area. This would provide not only a needed check on estimates of overall depletion but would identify heavily stressed parts of the subregion. Thus the method if successful would meet two of the principal defects of the model of the 1975 assessment - lack of uniform methods of error evaluation, and loss of information with aggregation.

It seems likely that other more efficient methods of eliminating regional flow variation can be found with development of plausible assumptions regarding the probability distribution and correlation structure of monthly flows in the subregions. Also methods other than simple linear regression for pattern recognition in the adjusted flows are likely to be adapted to detect significant changes in the streamflow patterns due to consumption and to other concomitants of development and changes in land-use. A number of statistical techniques including several forms of two-way analysis of variance would appear to have potential utility for this type of analysis.* In view of the desirability of developing a uniform method of checking the calculations of the water availability analysis model in the next assessment, it would be useful to initiate a substantial investigation on streamflow data of all the major regions to determine the predictive power of the best method of elimination of regional flow variation. This investigation

---

*A theoretical conceptualization based on an empirical Bayesian approach is presented in Appendix C. Computational procedures are proposed that appear to be practicable and useful.

would not be labor intensive since standardized computer programs would operate from taped records. We believe that the results would be helpful in making a less-expensive and more reliable Third National Water Assessment. The technique may also be useful in reducing the time needed for obtaining useful information regarding water quality changes at the stations of the NASQAN program.

Water Balance Models

In the past assessment has meant comparison of the adequacy of water supply with projected demands. And demands have related primarily to off stream uses - water for homes, farms and factories. The comparison is necessary in the context of economic growth, but past national assessments have not produced information relevant to society's interest in the quality of the environment and protection of natural systems. The effort made in the Second Assessment to measure adequacy of supply in relation to instream uses, though based on oversimplified criteria, was a significant improvement over previous assessments. We believe that this aspect of national water assessment merits increased attention due to the acceleration of offstream use. It should include not only consideration of instream flow needs but also the water needs of related land resources where changes in the hydrologic regime may impinge on environmental quality and the health of natural systems. This broadened connotation of "adequacy of supply" would necessitate new forms of assessment analysis and more diversified input data. The current assessment accounts for only about one-third of precipitation falling on the coterminous states. The ultimate goal of assessment we believe should be a scientifically valid accounting system of all stocks and flows of the hydrological cycle within the national borders. A more detailed and informative water accounting system, analogous to the national income accounts, would be of great and lasting importance for wise decisions in development of our water resource and for the rational evolution of our water law.

Changing the present system of assessment to a national water account system is a formidable undertaking. Even with improved surveys and monitoring data from remote sensing and other new means for data collection, development of a satisfactory national water accounting system will require several years and much effort.

Our motivation in investigating simple water balance models for application to assessment subregions is to begin to bridge the gap between present assessment techniques and those needed for an operational system of national water accounts.

An important principle in developing aggregated (or "lumped") watershed models is to use as few parameters as possible and in so far as possible to define the parameters so that each will reflect a regime characteristic that is subject to change with land use changes and installation of facilities for water management.*

Water Balance Model for Subregions

The following model has four parameters - two pertaining to runoff characteristics, and two relating to groundwater. Inputs to the model are monthly precipitation and potential evapotranspiration. Outputs include monthly runoff (direct and indirect), soil moisture, and groundwater storage. The following symbols are used

X_i = precipitation during month i, time span from t=i-1 to t=i

E_i = evapotranspiration during month i

V_i = potential evapotranspiration during month i

S_{i-1} = soil moisture at beginning of month i; t=i-1

G_{i-1} = groundwater storage at beginning of month i

---

* Our models use a much smaller number of parameters than conventional runoff models such as the Stanford Watershed Model. Although additional parameters and model components after intensive tuning can improve the predictive power of a model for special purposes such as forecasting runoff at specific sites, they are not useful for national water assessment because (i) they require intensive calibration for each application and (ii) the parameters cannot be accurately determined from the results of calibration owing to the high level of collinearity obtaining between many of the parameters.

With reference to Figure 5 the rationale of the model is as follows: the allocation of monthly precipitation x_i, to (i) runoff (direct and indirect), (ii) evaporation and transpiration, (iii) storage in soil moisture, and (iv) storage in groundwater aquifers is determined by (1) the magnitude of precipitation, x_i, (2) potential evapotranspiration V_i, and (3) the initial storages in soil and groundwater, S_{i-1} and G_{i-1}. The following equation controls the allocation:

$$y_i = y_i(X_i) = \frac{X_i+b}{2a} - \sqrt{\left(\frac{X_i+b}{2a}\right)^2 - \frac{X_i b}{a}} \tag{4}$$

Here y_i is the sum of monthly evapotranspiration and soil moisture storage at the end of the month. X_i is the sum of monthly precipitation and initial soil moisture. The difference X_i-y_i represents the sum of direct runoff and groundwater recharge (discharge to streams and aquifers). As X_i increases y_i approaches b, the upper bound of storage in the unsaturated zone above the goundwater table. Further increments of rainfall are discharged to streams and groundwater. As X_i decreases the derivative dy/dX approaches unity with runoff and groundwater recharge decreasing to zero. Runoff produced by a given depth of rainfall will vary depending upon antecedent soil moisture.** The parameter a, which is related to b by the following equation

$$a = 2b/y(b) - (b/y(b))^2$$

models the propensity in many catchments for runoff to occur well before the soils are saturated to capacity. The value of a is close to unity in flat topography with low drainage density. In topography with greater relief and drainage density the value of a is reduced. Urbanization and deforestation tend to decrease a. No special significance is attached to

---

** Antecedents of the model appear in publications of Thornthwaite (3)(4) and Sanderson (5).

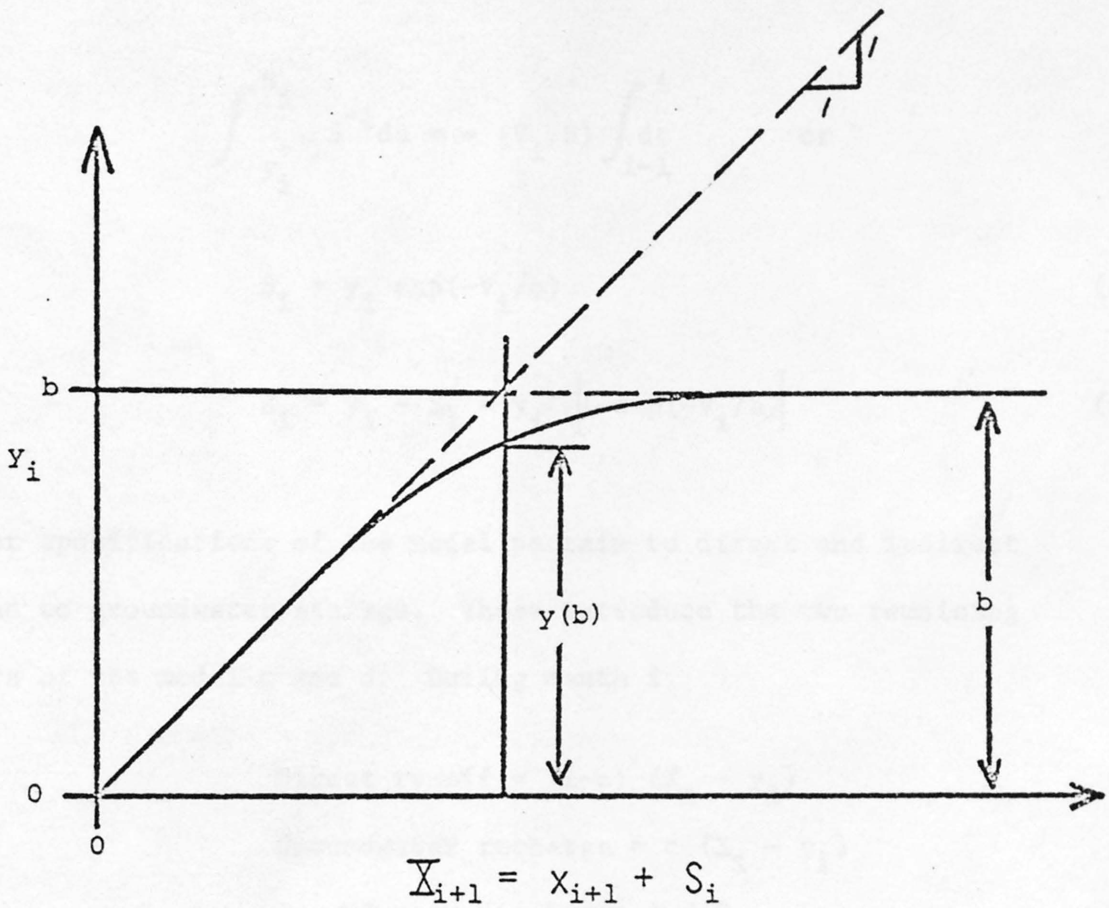


FIGURE 5 WATER BALANCE MODEL

the particular functional form of equation (4). Other two-parameter functions for which $y'(0) = 1$, and $y'(\infty) = 0$ might be used.

The model assumes that after a rainfall the rate of loss of soil moisture by evapotranspiration will be proportional to potential evapotranspiration for the month, and further that the constant of proportionality is the ratio S/b reflecting the degree of soil moisture saturation. That is during the period after a rainfall,

$$\frac{dS}{dt} = -\frac{V_i}{b} S$$

$$\int_{y_i}^{S_i} S^{-1} dS = -(V_i/b) \int_{i-1}^{i} dt \qquad \text{or}$$

$$S_i = y_i \exp(-V_i/b) \tag{5}$$

and

$$E_i = y_i - S_i = y_i \left[1-\exp(-V_i/b)\right] \tag{6}$$

Other specifications of the model pertain to direct and indirect runoff and to groundwater storage. These introduce the two remaining parameters of the model c and d. During month i,

$$\text{Direct runoff} = (1-c)\ (\overline{X}_i - y_i) \tag{7}$$

$$\text{Groundwater recharge} = c\ (\overline{X}_i - y_i)$$

$$\text{Groundwater discharge} = d\ G_i \tag{8}$$

From the mass balance on groundwater

$$c(\overline{X}_i - y_i) + G_{i-1} = d\ G_i + G_i$$

current groundwater storage may be calculated

$$G_i = (1+d)^{-1} \left[c(\overline{X}_i - y_i) + G_{i-1}\right] \tag{9}$$

From the foregoing equations it is evident that the parameter c controls the water input to the aquifers, and d controls the mean residence time of water in groundwater storage. Parameter c partitions runoff into its direct and indirect components. The reciprocal of parameter d is the mean residence time of water in aquifers. These parameters together with a and b provide quantitative measures of the dominant response characteristics of the drainage of the subregion.

For a given set of parameters (a, b, c and d) the computation is simple:

From monthly inputs x_i, V_i and initial soil moisture, S_{i-I}, calculate

(1) y_i by eq. (1)

(2) S_i and E_i by eqs. (5) and (6)

(3) direct runoff by eq. (7)

(4) final groundwater storage, G_i, and indirect runoff by equations (8) and (9).

Before discussing methods for fitting the model it will be useful perhaps to comment on two important features of the model that derive from its basic simplicity.

(1) Although the model includes variables pertaining to storage in aquifers and in the soil (and subsoil) it is not acutally necessary to have observations of these quantities to fit and use the model. Nor is it necessary to separate direct and indirect runoff. As will be seen the model can readily be fitted from monthly data on precipitation, potential evapotranspiration (or pan evaporation) and total runoff. The fitted model outputs the same variates as it would if it had been calibrated with a more complete set of observations including soil moisture and groundwater storage. Availability of data on soil moisture and groundwater of course would make possible a more accurate determination of the parameters, but the model can be fitted and used without them .

(2) From an overall mass balance of the inputs and outputs of the model the following relation may be derived:

$$\frac{S_i}{b} = \frac{x_i - r_i - \Delta S_i - \Delta G_i}{V_i} \; * \tag{10}$$

where r_i = total runoff (direct plus indirect) and

$\Delta S_i = S_i - S_{i-1}$

$\Delta G_i = G_i - G_{i-1}$

It may be noted that the right hand side of equation (10) involves observable quantities; it does not include estimated parameters. Moreover, it does not require estimates of storage in soil and aquifers; only <u>changes</u> in these storages are needed. The left hand side of equation (10) is a dimensionless measure of soil moisture. This ratio like the runoff coefficient r_i/x_i, (which also is calculated from observables) varies from month to month and year to year. Effects of regulation, diversion and land use changes are manifest in trends or significant shifts in these two ratios, which are important state variables for the subregion. The vector of runoff coefficients provides a macro-measure of the state of the aquatic environment, and the soil moisture vector provide an analogous state parameter for the related land resources and hence for environmental quality.

The significant point here is that the model, without using fitted parameters or estimates of storage in the soil and aquifers, serves to define a useful state variable relating to soil moisture and therefore to productivity.

It is interesting to note that climatologists have long used the first term on the right hand side of equation (10), x_i/V_i, as the basis for the classification of climates of various regions of the earth. The vectors of the monthly rates rainfall/potential evaporation, for example, are the primary determinants of Thornthwaite's classification of world

---

* With small values of the ratio V_i/b, equations (5) and (6) yield the approximate relationship: $E_i \doteq V_i S_i / b$.

climates (5). In equation (10) these meteorological variables are supplemented by hydrological variables - the last three terms in the numerator of the right hand side of the equation (10). These adapt the formula for use in classifying watersheds. It would appear that the model with its two state variables, r_i/X_i and S_i/b, can provide a useful macro-description of both natural and man-perturbed basins on the scale of subregions. The model also provides monthly estimates of the ratio of direct to indirect runoff. The model also provides monthly estimates of the ratio of direct to indirect runoff. These three ratios calculated as monthly and annual vectors would appear to have potential utility in establishing the carrying capacity of different environments.

Application to Data of Little River, Georgia

The model was tested with data from a 44.4 sq.mi. subarea of the basin of the Little River near Tifton, Georgia. The site is in the southern coastal plain about halfway between the Gulf of Mexico and the piedmont region. Nearly a third of the area is covered with forests. These lie in the shallow valleys of the drainage system. Water surface of streams and small inpoundments cover about 1.5% of the area. Most of the remaining area is used for growing corn, soybeans and peanuts in small fields mostly less than 0.5 ha. Slopes in the rolling topography are generally less than 8%. The agricultural fields have deep well drained soils of medium to high permeability. In the low-lying forested areas drainage is poor although the loamy subsoil is moderately permeable. Underneath the soil mantle throughout the entire area is a dense impermeable stratum (Hawthorne formation) that inhibits loss to deep seepage.

The test area is intensively instrumented. Runoff is gaged at 5-minute intervals with V-notch weirs and digital stage recorders. Precipitation is measured from twelve rain gages on a grid with digital recording equipment.

Tables 5,6 and 7 show monthly precipitation, runoff and pan evaporation for the period from January 1967 through November 1974. Lowest rainfall occurs during the months of September, October and November. During this period runoff recedes to very low levels. Mean annual rainfall for the 1967-1974 period was about 51 inches. This is about 5 inches less than average annual pan evaporation. Monthly averages of pan evaporation, based on records from 1938 to 1971 are as follows:

TABLE 5 : Little River, Georgia — Runoff, Inches — Station F (44.4 sq mi)
January 1969 to November 1976

Mo \ Yr	1969	1970	1971	1972	1973	1974	1975	1976	Mean
J	.22	.98	1.89	2.70	1.66	.80	2.29	1.38	1.49
F	.69	1.82	2.78	3.74	4.39	4.36	1.88	1.28	2.62
M	2.42	3.68	4.09	1.74	1.88	2.15	3.73	1.21	2.61
A	.70	2.41	1.90	1.17	5.38	2.45	4.33	.37	2.34
M	1.03	2.17	1.60	.04	1.10	.24	1.33	3.76	1.41
Je	.18	2.94	.37	1.36	1.10	.20	.44	.42	.88
Ju	.06	1.27	1.87	.56	.77	.12	.89	.73	.78
A	1.27	2.81	1.55	.05	.61	1.08	.80	.72	1.11
S	.94	.58	.22	.00	.15	1.30	.02	.99	.53
O	.17	.27	.02	.00	.00	.03	.06	.86	.18
N	.01	.20	.08	.00	.00	.01	.03	1.89	.38
D	.43	.45	2.39	.01	.03	.27	.20	—	.54
Annual	8.12	19.58	18.76	11.37	17.07	13.01	16.00	— (13.61)	14.87

TABLE 6 : Little River, Georgia — Precipitation, Inches
January 1969 to November 1976

Mo \ Yr	1969	1970	1971	1972	1973	1974	1975	1976	Mean
J	.65	2.71	2.96	5.07	5.74	4.42	6.39	4.05	4.00
F	3.49	4.16	6.16	5.88	6.23	7.19	2.90	1.82	4.73
M	6.13	9.52	6.45	4.53	6.11	4.09	6.93	3.90	5.96
A	1.43	1.27	4.81	.64	8.68	4.70	9.10	2.57	4.15
M	6.15	8.19	3.38	2.00	3.29	3.77	3.83	9.64	5.03
Je	1.91	3.94	5.07	9.10	5.72	5.11	3.46	3.53	4.73
Ju	6.77	6.51	7.86	4.18	5.52	5.29	6.19	4.95	5.91
A	6.49	9.81	6.48	2.08	4.19	5.74	4.64	4.64	5.51
S	5.72	1.21	.89	.83	1.43	6.22	2.57	3.96	2.85
O	.23	3.71	2.08	2.14	.66	.64	2.87	5.03	2.17
N	.69	1.02	3.34	2.52	1.21	2.01	1.67	6.03	2.31
D	4.19	3.54	5.91	5.24	3.64	2.05	3.85	—	3.92
Annual	43.85	55.59	55.36	44.21	52.42	51.23	54.40	— (50.12)	51.27

TABLE 7: Little River, Georgia: Pan Evaporation, Inches

Yr / Mo	1969	1970	1971	1972	1973	1974	1975	1976
J	2.30	2.06	1.74	1.93	1.79	2.09	1.99	2.60
F	2.38	3.02	3.02	2.50	2.41	2.72	2.35	3.59
M	4.45	4.02	4.23	4.56	4.02	4.72	4.67	4.20
A	6.07	5.63	5.67	6.36	4.94	6.31	5.50	6.69
M	6.24	6.98	6.98	5.91	6.87	5.78	6.25	5.61
Je	8.28	6.56	6.72	7.80	5.69	6.25	7.48	6.53
Ju	6.71	6.57	5.79	8.00	7.49	7.12	6.30	7.22
A	5.83	6.00	5.74	6.54	6.11	5.90	6.46	6.71
S	4.60	5.72	5.57	6.44	5.32	4.66	5.07	4.40
O	4.33	4.38	3.84	4.58	5.33	5.28	3.67	3.87
N	2.65	2.49	3.23	2.39	3.04	3.25	2.75	2.37
D	2.28	2.51	1.92	1.91	2.01	2.29	2.01	-
Total	56.12	55.94	54.45	58.92	55.02	56.37	54.46	- (53.79)

Month	Evaporation in/day	Month	Evaporation in/day
Jan.	2.14	July	6.43
Feb.	2.94	Aug.	6.08
Mar	4.44	Sept.	5.11
Apr.	5.90	Oct.	4.07
May	6.92	Nov.	2.81
June	6.90	Dec.	2.05

Preliminary Fitting of the Model

Input data consisted of monthly precipitation and evaporation. A simple fitting procedure was used in this initial test -- a systematic search to identify the parameter set (a,b,c and d) for which the mean of the square of the residuals (observed minus fitted runoff) was minimized. This method is particularly appropriate in this case since groundwater residence times are less than a month, and the value of parameter c of the model is small. In fact for the best fit, c was found to be zero. Thus only two parameters, a and b, needed to be estimated.

In lumped watershed models with monthly rainfall as input, errors in prediction arise in months in which a substantial part of the rain occurs late in the month with much of the induced runoff occurring in the next month. Unless an adjustment is made for such months the model is likely to overestimate runoff in the current month and underestimate it in the subsequent month. A method of adjustment is described in Appendix B.

In essence this adjustment defines an "effective" rainfall for each month from a weighted average of the observed rainfall for that month and the antecedent month. The weighting factors are calculated from the distribution of daily precipitation during the months as described in Appendix B.

No observations relating to groundwater elevation or soil moisture were used in this fitting. The procedure to estimate them was simply to assume trial values of the initial values S_{00} and G_{00}; then with a tentative parameter set (a,b,c, and d) to route water through the system over several eight-year cycles until the time traces of G_{ij} and S_{ij} attained a quasi-steady state. The values of G and S at the start of the year were averaged to provide initial values for the final 8-year run from which the mean square error of runoff residuals was calculated. The computations were simplified in this case by the fact that the best fits were obtained by setting parameter c to zero ($G_{ij} = 0$). This reduced the fitting to a two-dimensional search on a and b. Results of the search yielded the following estimates:

$$a = .979, \quad b = 13.1 \text{ inches.}$$

The root mean square of the residuals of runoff was 0.43 inches. Since the standard deviation of monthly runoff over the test period was 1.26 inches, the coefficient of determination for the fit with this parameter set is:

$$r^2 = 1 - \left(\frac{.43}{1.26}\right)^2 = 0.88$$

Thus 88% of the variance of the runoff is explained by the model. The residuals are shown in Table 8.

In this simple version of the model the effect of interception of rainfall by forests and crops was not incorporated. It may be noted that the model tends to underestimate runoff slightly during the first half of the year and to over estimate it during the latter half of the year. This pattern is indicative of the larger effect of interception during the warm months.

*A theoretical discussion of fitting the model based on a generalized Bayesian approach is presented in Appendix C. The method leads computational procedures that appear to be practicable and useful.

Interceptions may readily be modelled by making b a function of the time of the year rather than constant throughout the year. This elaboration is also needed to account for snow cover.

From the pattern of residuals in Table 8 there is an indication of the effect of supplemental irrigation during the summer months from wells and impoundments in the basin. During years when summer runoff is lower than normal the model tends to overestimate runoff. Low runoff is correlated with low soil moisture and the need for supplemental irrigation during long intervals between rains. A regression analysis of runoff and the sum of residuals for the months of June, July, August and September yield a coefficient of determination of 0.52 and a regression coefficient of 0.33 inches. That is for each inch of deficiency of summer rainfall, supplemental irrigation appears to increase by 0.33 inches. In future applications the model will be elaborated to include explicitly the effects of pumping and diversion as well as seasonal changes of parameters pertaining to runoff soil moisture and groundwater. Other methods of fitting and parameter estimation will also be investigated.

It is pertinent to note that conventional regression methods of predicting runoff from rainfall yield fairly high r^2 values for this basin. For example, the regression:

$$Q_i = \sum_{j=1}^{12} a_j \delta_{ij} + b_1 P_i + b_2 P_i^2 + b_3 P_{i-1}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & 0 \text{ otherwise} \end{cases}$$

yielded an r^2 of 0.84, somewhat lower than that found with the model. This

TABLE 8: Little River Georgia: Residuals of Runoff, $Q_{ij} - \hat{Q}_{ij}$, Inches

Mo \ Yr	1969	1970	1971	1972	1973	1974	1975	1976	Mean
J	-.17	-.39	1.13	-.08	-.78	.21	.90	.22	.23
F	.01	.52	.51	-.15	-.35	.75	.51	.38	.27
M	.05	.07	.20	-.08	-.07	.45	.01	.43	.13
A	.12	.75	.58	.51	-.57	.23	-.54	-.03	.13
M	.13	1.01	.11	-.20	-.32	-.46	.09	.26	.08
Je	.19	1.08	-.30	.23	-.20	-.74	-.19	-.39	-.09
Ju	-.46	-.19	.01	-.37	-.74	-.67	-.09	-.38	-.36
A	-.45	-1.23	-.71	-.21	-.08	-.18	.06	.08	-.34
S	-.61	-.02	-.29	-.09	-.13	-.62	-.42	.32	-.23
O	-.20	-.15	-.23	-.07	-.09	-.34	-.35	-.24	-.21
N	-.14	-.07	-.15	-.11	-.06	-.15	-.25	.73	-.03
D	.06	.13	.82	-.45	-.13	.03	-.19	-	.04
Sum	-1.85	2.29	1.68	-1.07	-3.52	-1.49	-.46	1.38	

regression, however, requires fifteen parameters whereas the model used only two. Moreover the model provides monthly estimates of soil moisture, and the frequency distribution of soil moisture. This information as well as that pertaining to supplementary irrigation cannot be obtained from conventional rainfall-runoff analyses.

Conclusion

The calculations with the Little River data show that it is possible to model the major characteristics of basin drainage in a simple analytical framework using ony a few descriptive parameters. In future research with other data sets minor elaborations of the model and alternative methods of fitting will be tested.

REFERENCES

1. Beard, L.R., Needs of National Assessment. Report Phase 3 to U.S. Water Resources Council, Program for Continuing Assessment: Center for Research in Water Resources, University of Texas, Austin, July 26, 1978.

2. Langbein, W.B., and Slack, J.R., Variations in National Water Supply - 1979 Update Water Resources Review Jan. 1980. (A preliminary memorandum of a statistical analysis of the year to year variations of average flow of U.S. rivers since 1911.)

3. Thornthwaite, C.W., The Moisture Factor in Climate Trans Am. Beophys. Union 27 pp.41-48, Feb. 1948.

4. Thornthwaite, C.W., and Mather, J.R., the Water Balance, Publ. Climatol., 8 (1) Laboratory of Climatology, Elmer, New Jersey, 1955.

5. Sanderson, Marie, Variability of Climate in the Lake Ontario Basin. Wat. Res. Research I pp.554-565, June 1971.

6. Thornthwaite, C.W., Climates of the Earth, Geogr. Rev. 23 pp.433-500 (1933). See also 33 pp.232-255 (1943); 38 pp.55-95 (1948).

Trend Line Slopes in Different Regions

Results of statistical analyses of 30-year records of mean annual flow (1942-1971) at 251 gaging stations are summarized in the following table:

Region	No. Stations	Regression line slope * Mean	S.Dev.
1 Maine	15	.006	.010
2 Susq. Basin	29	-.037	.011
3 Potomac Basin	19	-.020	.008
4 N.Carolina	11	.002	.010
5 Georgia	14	.005	.010
6 Michigan	17	-.026	.026
7 Ill+Ind+Ky	13	-.014	.015
8 AR + LA	10	-.036	.010
9 Wisconsin	21	-.013	.019
10 Missouri	14	-.032	.015
11 Mississippi	19	-.018	.016
12 MT + ID	12	.013	.016
13 Washington	27	.020	.020
14 Oregon	20	-.001	.011
15 California	10	.010	.033
All Regions	251	-.010	.025

* Normalized on the standard deviation of observed annual flows

The average slope for all regions is negative, indicating increasing consumptive use in the nation. However, the variation in slopes is large both between regions and within regions. It may be noted that some of the largest negative slopes occur in the humid eastern part of the USA and some of the largest positive slopes occur in the west. This would suggest that a substantial part of the variation is due to a small temporary nation-wide downward trend in precipitation and runoff during this particular period. However, it cannot be concluded from the sample statistical analysis how much the slopes are determined by man induced change and how much is due to natural fluctuations in weather patterns. The data do however provide an indication of the range of uncertainty inhering in thirty-year records. A more precise indication is given by an analysis by Langbein & Slack (2) based on larger records of several of the major basins of the country.

APPENDIX B

Method of Calculating Monthly Effective Rainfall from Daily Rainfall

When monthly precipitation is used as input to aggregated watershed models errors in predicted outputs occur in months in which much of the rain occurs in the last few days of the month since some of the induced runoff will occur during the next month. Thus models will tend to overestimate runoff for the current month and to underestimate it in the next month. A simple adjustment to be described will eliminate much of this error.

An "effective" precipitation for month i is calculated as a weighted average of observed monthly precipitations for months i and i-1

$$\tilde{P}_i = \eta_i P_i + (1-\eta_{i-1}) P_{i-1}$$

where P_i = precipitation in month i

η_i = weighting factor for month i $0<\eta_i<1$

The weighting factors for each month are calculated from daily rainfall by the formula,

$$\eta_i = P_i^{-1} \sum_{t=1}^{N} p_t \, (1-\theta^t)$$

Here p_t is rainfall occurring on the t^{th} day counted from the end of the month*, N is the number of days in the month, and θ is a parameter reflecting the time response of runoff to precipitation. Depending on runoff characteristics of the catchment, θ ranges from 0.60 to 0.95.

---

* For example for the 5th of May, t=27.

Assume that the runoff unit hydrograph may be fitted approximately by a geometric frequency distribution with parameter θ. Let p_1, denote the proportion of runoff occuring on the day of the rainfall, p_2 runoff on the second day, etc. With a geometric distribution these proportions are simple functions of θ:

t	p_t	Σp_t
1	$1-\theta$	$1-\theta$
2	$(1-\theta)\theta$	$1-\theta^2$
3	$(1-\theta)\theta^2$	$1-\theta^3$
t	$(1-\theta)\theta^{t-1}$	$1-\theta^t$

The time interval between the centroid of the rainfall distribution and the centroid of the runoff hydrograph, say $\bar{t}$, is related to θ by the equation

$$\theta = 1 - 1/\bar{t}$$

Thus θ can be estimated from observed hydrographs. However, in most situations it is simpler to treat θ as a minor variable of the system to be fitted with the major parameters, a,b,c,d. In fitting the Little River data runs with several trial values showed that $\theta = 0.85$ ($\bar{t}$=6.7days) gave the best fit. This value was used in defining effective rain successive months. The correction made a substantial improvement to the coefficient of determination for predicted and observed runoffs. It eliminated several large residuals when heavy rain occurred near the end of the month.

APPENDIX C

Multiple Response Model Estimation

Hydrologic models typically involve complex interactions among multiple dependent responses. Parameter estimation techniques that ignore these couplings or inter-equation dependencies are inefficient and can lead to incorrect inferences. Fiering and Kuczera (1979) proposed a fitting approach that accounts for inter-equation dependencies in linear models. This section presents an extension, based on the work of Box and Draper (1965), to the non-linear case and its application to the water balance model presented in the main body of this Report.

A multiresponse hydrologic model can be written in the following general form:

$$\underset{m\times 1}{\underset{\sim}{z}_j} = \underset{m\times 1}{\underset{\sim}{\eta}_j(\theta, \underset{\sim}{x}_j)} + \underset{m\times 1}{\underset{\sim}{\varepsilon}_j} \qquad j = 1,\ldots,n \tag{1}$$

where

$\underset{\sim}{z}_j \equiv$ response vector at time j

$\underset{\sim}{\eta}_j(\underset{\sim}{\theta}, \underset{\sim}{x}_j) \equiv$ predictive vector equation at time j

$\underset{\sim}{\varepsilon}_j \equiv$ error vector at time j

$\underset{\sim}{x}_j \equiv$ known input vector at time j

$\underset{\sim}{\theta} \equiv$ parameter vector

assuming that the errors follow a multivariate normal distribution with zero expectation and covariance Σ,

$$\underset{\sim}{\varepsilon}_j \ \underset{\sim}{\text{iid}}\ N_m(\underset{\sim}{0}, \Sigma). \tag{2}$$

permits the likelihood for $\underset{\sim}{\theta}$ and Σ to be written as

$$\ell(\underset{\sim}{\theta},\Sigma|\underset{\sim}{\varepsilon}) = p(\underset{\sim}{\varepsilon}|\underset{\sim}{\theta},\Sigma) = \prod_{j=1}^{n} p(\underset{\sim}{\varepsilon}_j|\underset{\sim}{\theta},\Sigma)$$

$$\propto |\Sigma|^{\frac{-n}{2}} \text{ etr } \frac{-1}{2}(\Sigma^{-1}V(\underset{\sim}{\theta})). \tag{3}$$

where:

$$V(\underset{\sim}{\theta}) = \sum_{j=1}^{n} \underset{\sim}{\varepsilon}_j \underset{\sim}{\varepsilon}_j^T$$

$$= \sum_{j=1}^{n} (\underset{\sim}{z}_j - \underset{\sim}{\eta}_j(\underset{\sim}{\theta},\underset{\sim}{x}_j))(\underset{\sim}{z}_j - \underset{\sim}{\eta}_j(\underset{\sim}{\theta},\underset{\sim}{x}_j))^T$$

Combining (3) with non-informative priors and intergrating over the marginal psoterior of Σ yields the marginal posterior of $\underset{\sim}{\theta}$ inversely proportional to the determinant of $V(\underset{\sim}{\theta})$:

$$p(\underset{\sim}{\theta}|\underset{\sim}{\varepsilon}) \propto |V(\underset{\sim}{\theta})|^{\frac{-n}{2}} \tag{4}$$

taking

$$\hat{\underset{\sim}{\theta}} = \text{mode of } p(\underset{\sim}{\theta}|\underset{\sim}{\varepsilon})$$

is equivalent to

$$\hat{\underset{\sim}{\theta}} = \inf_{\underset{\sim}{\theta}} |V(\underset{\sim}{\theta})|.$$

Thus the estimator based on the modal value of equation (4) results in the least squares estimate when there is only one equation (i.e., m=1).

For the case in which only a subset of $\underset{\sim}{z}_j$, say z_{ij} and z_{3j}, is observed, wherein

$$\underset{\sim}{z}_j = (z_{1j}, z_{2j}, \dots z_{mj})^T,$$

the estimation of $\underset{\sim}{\theta}$ is based on the appropriate partition of $V(\theta)$, i.e.

$$p(\theta|\varepsilon 1,\varepsilon 3) \propto \begin{vmatrix} v_{11}(\underset{\sim}{\theta}) & V_{13}(\underset{\sim}{\theta}) \\ v_{31}(\underset{\sim}{\theta}) & V_{33}(\underset{\sim}{\theta}) \end{vmatrix}^{-\frac{n}{2}}.$$

Further inferences on $\underset{\sim}{\theta}$ can be made from the relation

$$-2 \log \frac{p(\underset{\sim}{\theta}|\underset{\sim}{z})}{p(\hat{\underset{\sim}{\theta}}|\underset{\sim}{z})} \mathrel{\dot\sim} \chi_2^2$$

from which it follows that the contour defined by

$$\log p(\underset{\sim}{\theta}|\underset{\sim}{z}) = \log p(\hat{\underset{\sim}{\theta}}|\underset{\sim}{z}) - \frac{1}{2} \chi^2(2,\alpha)$$

where $\chi^2(2,\alpha)$ is the upper 100 α% point of a χ^2 with 2 degrees of freedom, encloses a region whose probability content is approximately $(1-\alpha)$.

The water balance model presented in the main text has five potentially observeable responses,

(i) Direct Runnoff (DR_j)

$$DR_j = \eta_{j,1} + \varepsilon_{j,1} = (1-c)(X_j+S_{j-1}-Y_j) + \varepsilon_{j,1}$$

(ii) Groundwater Discharge (GD_j)

$$GD_j = \eta_{j,2} + \varepsilon_{j,2} = \frac{d}{(1+d)} \left[c(X_j+S_{j-1}-Y_j) + G_{j-1}\right] + \varepsilon_{j,2}$$

(iii) Soil moisture content (S_j)

$$S_j = \eta_{j,3} + \varepsilon_{j,3} = Y_j \exp(-Vj/b) + \varepsilon_{j,3}$$

(iv) Groundwater Storage (G_j)

$$G_j = \eta_{j,4},\ \varepsilon_{j,4} = (1+d)^{-1} \left[c(X_j+S_{j-1}-Y_j) + \right] G_{j-1} + \varepsilon_{j,4}$$

(v) Evaporation (E_j)

$$E_j = \eta_{j,5} + \varepsilon_{j,5} = Y_j(1 - \exp\ (Vj/b))$$

with

$$Y_j = \frac{X_j+S_{j-1}+b}{2a} - \left[\left(\frac{X_j+S_{j-1}+b}{2a}\right)^2 - \left(\frac{X_j+S_{j-1}}{a}\right)b\right]^{\frac{1}{2}}$$

and can be put in the form of (1) by letting

$$\underset{\sim}{\theta}^T = (a,b,c,d)$$

$$\underset{\sim}{z}_j^T = (DR_j,\ GD_j,\ S_j,\ G_j,\ E_j)$$

$$\underset{\sim}{\varepsilon}_j^T = (\varepsilon_{j,1},\ \varepsilon_{j,2},\ \varepsilon_{j,4},\ \varepsilon_{j,5})$$

$$\underset{\sim}{\eta}_j^T(\underset{\sim}{\theta},\ \underset{\sim}{x}_j) = (\eta_{j,1},\ \eta_{j,2},\ \eta_{j,3},\ \eta_{j,4},\ \eta_{j,5})$$

Application of the water balance model to the Little River in Georgia was based on monthly data on precipitation, pan evaporation and total runnoff from 1/69 to 11/76.

Since data was available on only one observable (total runnoff) the model was reduced to one equation. A search over the parameter space yielded

$$\hat{\theta} = (\hat{a},\ \hat{b}) = (0.979,\ 13.1)^T = \text{mode}\ p(\underset{\sim}{\theta}|\underset{\sim}{\varepsilon})$$

with

$$\sum_{j=1}^{n} \varepsilon_{j,1}^{2} = 17.51.$$

The resulting 50% and 99% probability contours for a and b are shown in Figure C.1.

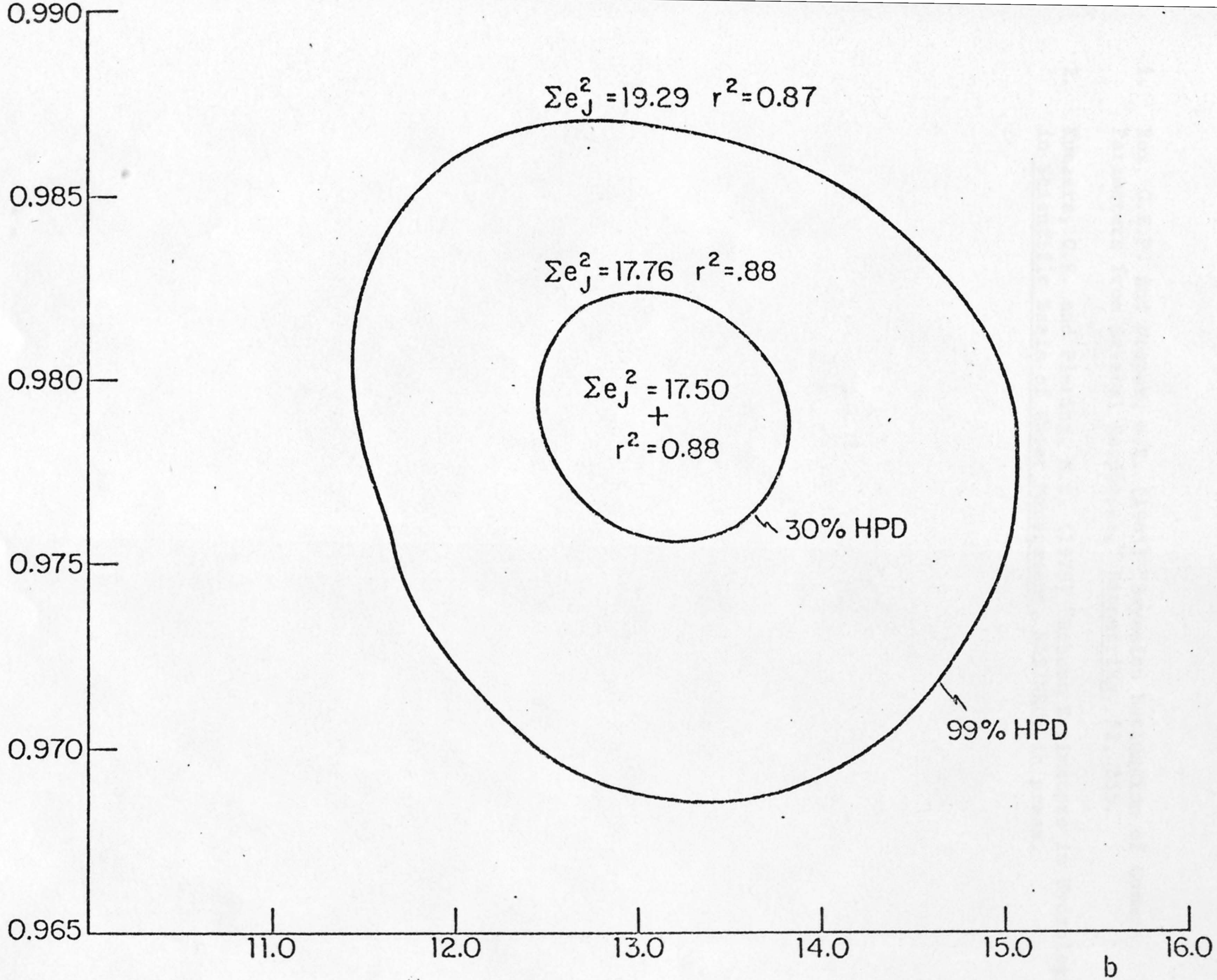


FIG. C.1 POSTERIOR PROBABILITY COUNTERS FOR PARAMETERS a, b.

REFERENCES

1. Box, G.E.P. and Draper, N.R. (1965) "Bayesian Estimation of Common Parameters from Several Responses," Biometrika, 52, 355.

2. Kuczera, G.K. and Fiering, M.B. (1979) "Robust Estimators in Hydrology," in Scientific Basis of Water Management, NAS/NRC, in press.