January 1981

Update on Estimation of Water Surface Elevation Probabilities for the Great Salt Lake

L. Douglas James

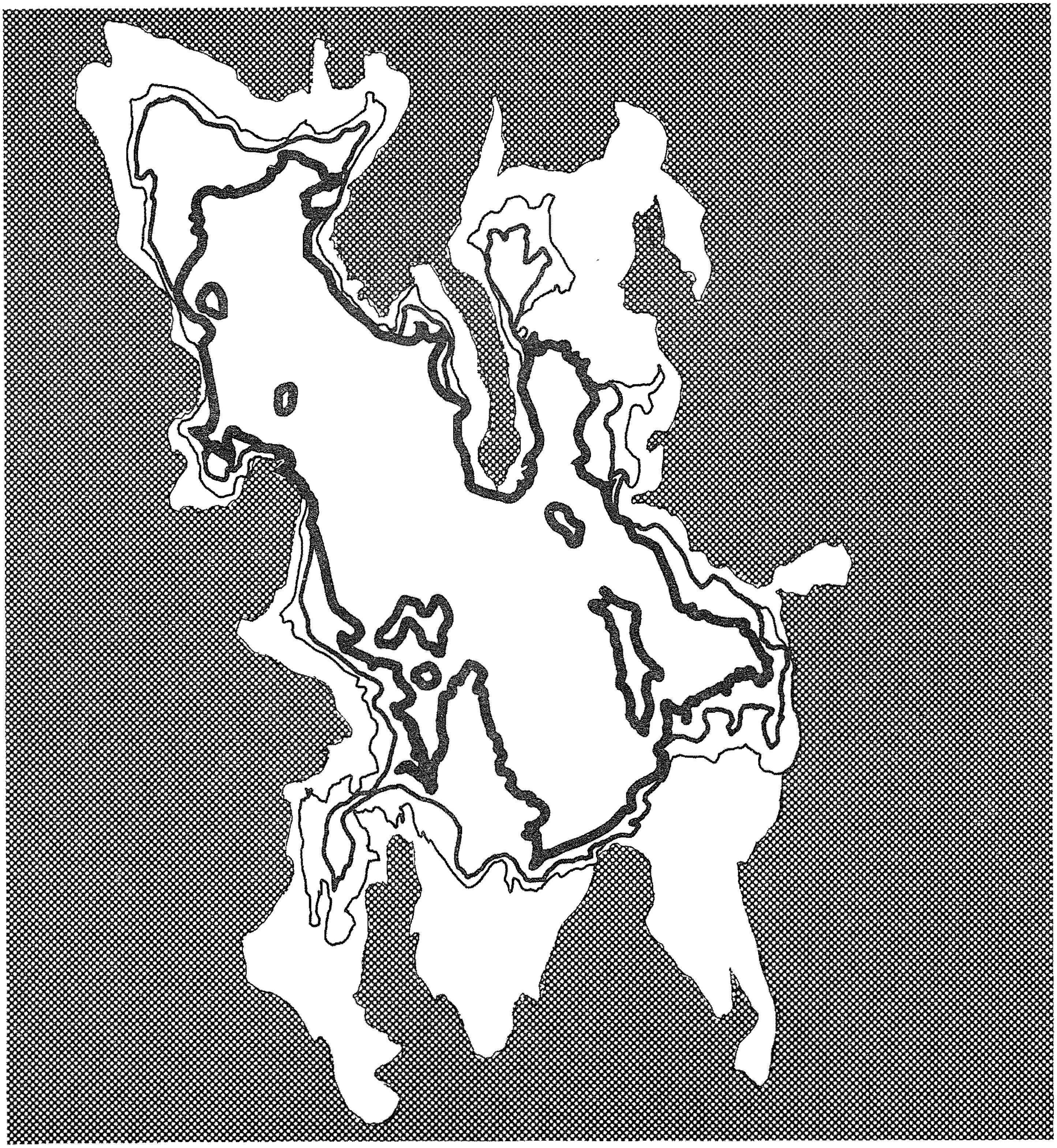
David S. Bowles

W. Robert James

Ronald V. Canfield

Update on Estimation of Water Surface Elevation Probabilities for the Great Salt Lake

L. Douglas James David S. Bowles W. Robert James Ronald V. Canfield



Utah Water Research Laboratory
Utah State University
Logan, Utah 84322 February 1981

Water Resources Planning Series
UWRL/P-81/01

Update on Estimation of Water Surface Elevation Probabilities For the Great Salt Lake

L. Douglas James
David S. Bowles
W. Robert James
Ronald V. Canfield

WATER RESOURCES PLANNING SERIES
UWRL/P-81/01

Utah Water Research Laboratory
Utah State University
Logan, Utah 84322

February 1981

ABSTRACT

The techniques of operational hydrology, employing an autoregressive moving average (ARMA (1,0) model were used to replicate historical patterns of streamflow into, precipitation on, and evaporation from the Great Salt Lake. The results were combined with a lake water balance model to simulate lake stage sequences beginning with known initial conditions and extending up to 125 years into the future and used to generate probability distributions for future lake stages. Starting with a spring 1980 high stage of 4200.45 ft msl, the best estimate is that the 1981 spring high will be 4200.19, but there is one chance in four that it will reach 4200.74 and one in ten that it will reach 4201.24. Over the long run, an average spring high of 4195.20 is forecast with one year in a hundred reaching as high as 4205.21 and one year in a hundred with a spring high of only 4185.19 and dropping as low as 4183.5. Followup annual forecast updates will be published about each July 1.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Update on Estimation of Water Surface Elevation Probabilities For the Great Salt Lake

INTRODUCTION

Industries, railroad companies, highway agencies, and natural resource user groups suffer when water surface levels in the Great Salt Lake rise higher or fall lower than their customary range. High water can flood the ponds and plants of lakeshore mineral industries, close recreation facilities and reduce visitation to shoreline areas, inundate feeding and nesting areas for migratory waterfowl with salt brine that over time destroys marsh productivity, and erode embankments and eventually close shoreline or causeway crossing railroads and highways. Low water increases the cost to the mineral industry of pumping lake brines to evaporation ponds and leaves wide unsightly beaches that reduce recreation visitation and require moving facilities toward the receded shoreline.

While slow rates of rise and fall provide preparation time to reduce sudden unexpected damages, the managers of the above facilities could do a better job of maximizing the economic advantage of lakeshore locations and minimizing losses from lake level change with better information on the probabilities of lake levels rising or falling to their levels of concern within their planning horizons. In order to determine whether any governmental action is warranted and better plan justified measures, public agencies could also profit from lake level forecasts.

Standard methods (U.S. Water Resources Council 1976) for estimating probabilities for riverine flood stages are not appropriate for terminal lakes because they do not account for the dependence of lake levels on the levels of immediately previous years. Great Salt Lake levels vary over many feet, rising over series of years when cumulative inflows exceed cumulative evaporation and falling when opposite conditions prevail. These cycles last decades and make the level during the previous year the major parameter in estimating the level for the following year.

The methods of operational hydrology, originally applied to storage in water supply reservoirs, provide a tool that can be used to derive terminal lake level probability distributions. The method calculates observed historical relationships among inflow (surface, subsurface, and precipitation) and outflow (evaporation) and of the patterns in all these time series. Estimation of the parameters of these relationships requires time series that are homogeneous over time.

The primary homogeneity problem on the Great Salt Lake is that land use changes and water supply development since 1847 have substantially reduced streamflow. Streamflow data series have been compiled on two bases. One, "natural streamflows," estimates annual flows such as would occur from natural watersheds. The other, "present modified streamflows," estimates annual flows such as would occur from watersheds with 1965 practices of land and water use (Utah Division of Water Resources 1977). The hydrologic effect of these practices has not changed much in the years since.

James et al. (1979) found "present modified streamflows" to provide the more satisfactory data base for stochastic generation of flow sequences. Natural flow sequences were less satisfactory because of higher cross correlations and the lack of a reliable relationship for converting generated natural flows into the flows which now actually occur and govern lake levels. These "present modified flows" for the Bear, Weber, and Jordan Rivers into the Great Salt Lake provide three of the needed five homogeneous data series. For the other two, average precipitation on the lake was estimated from seven gages around the lake (James et al. 1979, p. 38); and a time series of freshwater equivalent lake evaporations was constructed from available pan data (p. 40). The period beginning with water year 1937 was selected for the data base because 1) it utilizes the full length of evaporation record (the shortest series duration and largest flow quantity involved) and 2) the streamflow and precipitation data for the period have statistical characteristics representative of those measured over the period of record beginning in 1890 and of those estimated from tree-ring data beginning in 1700 (p. 51).

Autoregressive moving average (ARMA) models were applied to replicate the serial and cross correlation characteristics of these five data series over the 41-year data base.

James et al. (1979) applied both 5-variate ARMA (1,1) and ARMA (1,0) models to these five series and concluded that with the parameter estimating procedures available that the ARMA (1,0) generation based on 1937-1977 present modified flows best matched most flow statistics. The exception was that this model is not able to replicate the significant persistence observed in some of the five series. Thus, the decision to proceed with the ARMA (1,0) model was tempered by a feeling that the ARMA (1,1), or perhaps even a higher ordered model, should theoretically give better results in that it can replicate persistence and that the difficulty may be overcome by developing improved techniques for model parameter estimation. Trial of a homogeneous ARMA (p,q) model for which the method of moments for parameter estimation could be replaced by the method of maximum likelihood was recommended (pp. 65-67).

For their immediate application, however, the ARMA (1,0) was used to generate 100 sets of 125-year streamflow, precipitation, and evaporation sequences. These became inputs to a lake water balance model used to generate an equal number of 125-year lake stage sequences. These 100 equally probable sequences beginning from the known water surface elevation of 4198.6 feet msl on October 1, 1978, provide a distribution of the possible lake levels at the beginning of each subsequent water year through 2103.

The level for October 1978 is a known amount. The distribution of possible lake levels for October 1979 (as estimated in 1978) varies over a range, and as uncertainties increase into the future, this range becomes progressively larger in subsequent years. James et al. (1979) found the distribution to stabilize after about 35 years upon reaching a pattern that turned out to be approximately a normal distribution with a mean of 4196.42 feet msl and a standard deviation of 4.56 feet. Because the estimate of expected lake level is influenced by known initial conditions for about 35 years; industrial and land use management decisions around the lake shore, generally based on a 10 to 20 year time frame, should vary considerably with current lake levels.

GOALS OF THIS STUDY

While the above results provided reasonable probability distributions for Great Salt Lake water surface elevations, they left several issues unresolved. The purpose of this study was to consider these residual problems and refine the lake level estimates as necessary.

The most challenging possibilities for modeling improvement were those of using a homogeneous approach to stochastic modeling (James et al. 1979, pp. 65-67) and maximum likelihood parameter estimation techniques to move from an ARMA (1,0) to a higher order model that might do a better job of preserving observed patterns in the historical flow sequences. As long as this attempt was to be made, several other refinements were also included:

1. The error in using a 1977 Weber River flow of 770,580 AF rather than the actual 77,580 AF (p. 43) was corrected.

2. The precipitation, evaporation, and streamflow data series were updated to include 1978.

3. The original time series scaling had used a three parameter log normal distribution (3PLN) for three of the five series and a normal distribution for the precipitation and the Weber River series because of their low skew (p. 49). Following the recommendation of Burges and Hoshi (1978), the 3PLN was used for all five cases.

4. The Box-Jenkins identification procedure (Box and Jenkins 1970) was employed to determine trial values for p and q in fitting a general ARMA (p,q) model. In the previous study, only ARMA (1,0) and ARMA (1,1) models were considered.

5. The homogeneous modeling method and maximum likelihood estimation of model parameters were tried in attempts to improve upon the ARMA (1,0) match of flow statistics.

6. The best performing model would be used to revise the previously derived probability distributions of annual peak Great Salt Lake levels (James et al. 1979, Figure 11).

7. A simple method would be devised for annually updating the estimated probability distributions. The estimated distribution would begin with the recorded springtime annual peak and cover the probabilities for the range of possible levels over the subsequent 35 years.

DATA CORRECTION AND UPDATE

The 1977 Weber River present modified flow was corrected, and the 1978 data were added to all five present modified sequences. The data for the five series from 1890 through 1976 are recorded in James et al. (1979, pp. 42-43); the 1977 and 1978 information and revised statistics for the complete series are on Table 1. In comparison with the previously published statistics, the mean of the Weber River flows dropped 0.4 percent, the standard deviation dropped by 0.9 percent, and the skew increased by 4.3

Table 1. Data for complete recorded streamflow, precipitation and evaporation time series.

		Evap. in.	Prec. in.	Bear R. AF	Weber R. AF	Jordan R. AF
Beginning		1937	1890	1890	1890	1890
1977		48.7	9.90	689300	77580	218600
1978		44.3	13.10	945100	305100	284720
	Mean	51.8	9.99	1179035	452655	275311
	Std. Dev.	4.2	2.42	481860	233737	66826
	Skew	0.536	0.046	0.434	0.555	0.533
	Hurst	0.68	0.557	0.330	0.748	0.736
	Lag One Coeff.	0.23	0.176	0.622	0.425	0.715

percent. Changes for the other four series were of the same general order of magnitude.

APPLICATION OF BOX-JENKINS IDENTIFICATION PROCEDURE

The results of the Box-Jenkins identification procedure (Box and Jenkins 1970) used to select appropriate autoregressive and moving average model orders (p and q respectively) for the five time series are shown in Table 2. Each box contains parameter estimates at the top and chi-square statistics at the bottom.

The parameters shown are the estimates from the Box-Jenkins program for the p autoregressive terms and the q moving average terms. Therefore, for an ARMA (1,0) model, only one parameter is estimated; and for an ARMA (2,1), there are two autoregressive parameters followed by the one moving average parameter.

Recorded under the parameter estimates are the chi-square statistics needed to test whether the residuals between the fitted model and the original data time series are white noise, i.e., a purely random and thus serially uncorrelated time series. A numerator smaller than the denominator implies a random residual series and thus no need for the complexity of a higher order model.

The chi-square statistic for the residual time series is calculated from K orders of serial correlations of the residuals with the formula:

$$Q = n \sum_{k=1}^{K} \rho_k^2 (\hat{a}) \quad \cdots\cdots\cdots\cdots \quad (1)$$

where $\rho_k(\hat{a})$ is the serial correlation coefficient for the series with k lags, K is the maximum number of lags used and consequently the number of serial correlation coefficients calculated, and n is the length of the time series. The values of Q approximately follow a chi-square distribution with K-p-q degrees of freedom. According to the test, a Q less than χ^2(K-p-q) indicates that the null hypothesis of nonsignificant serial correlation cannot be rejected. By implication, the ARMA model being examined explains all but a white noise component of the data series at the chosen significance level.

The Box-Jenkins model identification procedure begins by fitting a model of higher order than necessary to the data and proceeds by trying models of lower order to see if the residual can still all be explained as white noise. The lowest order model (minimum p+q with minimum p used in comparing models with the same p+q because autoregressive components are easier to apply than are moving average components) for which the residuals are determined by the above chi-square test to be white noise is selected. In a case such as this one where several time series are involved, the lowest order which reduces all the series to white noise should be selected to avoid problems caused by using different models of different series.

From the data on Table 2, this test selects the ARMA (1,0) model. This first order autoregressive model is also suggested by a partial autocorrelation function which drops below a 95 percent significance level after the first lag for the three streamflow series. The evaporation and precipitation data have serial correlation coefficients which are just statistically significant at the 95 percent level for the first lag. These results verify use of the ARMA (1,0) model, but it should be remembered that none of these tests deal explicitly with persistence statistics, the property initially suggesting investigation of the higher order models.

ARMA (1,0) MODEL RESULTS

In a separate evaluation of model alternatives, James et al. (1979) compared

Table 2. Parameter estimates for Box-Jenkins ARMA (p,q) models.

MODEL	GSL EVAP. 1937-1978	GSL EVAP. 1875-1978	GSL EVAP. 1937-1978	PMF BEAR RIVER 1890-1978	PMF BEAR RIVER 1937-1978	PMF WEBER RIVER 1890-1978	PMF WEBER RIVER 1937-1978	PMF JORDAN RIVER 1890-1978	PMF JORDAN RIVER 1937-1978
K =	10	25	10	22	10	22	10	22	10
ARMA(1,0)		0.245 25.8/35.2	0.025 12.0/15.5	0.611 26.9/31/4	0.623 4.3/15.5	0.419 15.5/31.4	0.561 7.7/15.5	0.703 15.7/31.4	0.701 8.9/15.5
ARMA(2,0)		0.206 0.164 23.7/33.9		0.507 0.189 23.0/30.1	0.587 0.055 3.7/14.1	0.370 0.127 14.4/30.1	0.601 -0.081 7.9/14.1	0.531 0.254 14.6/30.1	0.487 0.302 8.4/14.1
ARMA(3,0)		0.185,0.111 0.233 18.3/32.7		0.486,0.117 0.124 22.7/28.9		0.349,0.053 0.175 12.0/28.9		0.555,0.322 -0.115 14.6/28.9	
ARMA(4,0)				0.49,0.12 0.16,-0.06 21.6/27.6		0.3,0.4 0.1,0.2 13.6/27.6			
ARMA(5,0)				0.5,0.1,0.17 -0.15,0.09 15.3/26.3		0.3,0.2,0.12 0.1,0.11 10.6/26.3			
ARMA(0,1)	-0.167 8.1/15.5	-0.2 29.1/35.2 -0.145* 25.3/35.2	-0.019 12.2/15.5						
ARMA(0,2)		-0.074* -0.142 22.9/30.1	0.040 -0.199 9.6/14.1						
ARMA(0,3)		-0.099,-0.096* -0.229 20.7/28.9	0.007,-0.113 -0.116 10.1/12.6						
ARMA(1,1)		0.934 0.783 18.1/33.9		0.791 0.307 21.6/30.1	0.674 0.085 3.8/14.1	0.853 0.607 12.2/30.1	0.498 -0.094 7.7/14.1	0.828 0.261 14.1/30.1	0.851 0.306 7.7/14.1
ARMA(2,1)									-0.125,0.766 -0.642 8.3/12.6

*Test run with 1890-1978 data, K = 22.
Chi squared goodness of fit test on residuals of each model compared at 95 percent confidence level; $Q/\chi^2.95)$ to not being significantly different from white noise.

statistics from the present modified series with those from sets of 20 125-year flow series generated with various models, and found the ARMA (1,0) to give the most satisfactory match for the Great Salt Lake data (p. 61). After revising the 1977 data and adding the 1978 data shown on Table 1, the comparative statistics were recompiled and are shown on Table 3.

Comparison of these results with those previously obtained show the generation with 20 125-year sequences to have a somewhat poorer match. A second comparison, with 40 42-year sequences, was somewhat better, particularly as seen in the ratios of the differences to the values of the four means and in the M_0 matrix. The M_1 and M_2 matrices are less satisfactory, but an ARMA (1,0) model would not be expected to perform very well in preserving the M_1 matrix and would not be expected to preserve M_2 at all. The two comparisons also suggest that validity checks require more than 20 generated sequences to determine how well a model is really doing.

The comparison of 20 125-year sequences and 40 42-year sequences encompasses the effects of both 1) longer versus shorter sequences and 2) more versus fewer sequences. Preserving series statistical properties as the sequence becomes longer is called Type A resemblance while preserving the property as the number of sequences becomes more is called Type B resemblance. The usual application of synthetic hydrology is concerned with Type B resemblance because the sequence length is fixed at the economic life of the project. For comparing data and generated sequences, it makes sense to use the same period so that any biases which depend on sequence length are equivalent for both series; and the period of record is the most convenient for the data series because it provides the best estimate of the statistics.

RESULTS WITH HIGHER ORDER MODELS

The studies with higher order models involved estimation of parameters for a homogeneous model of the multivariate, multilag system and Monte Carlo testing of the sensitivity of the model to nonhomogeneous data. They are reported in Appendix A.

The conclusion with respect to the immediate Great Salt Lake application was to continue with the ARMA (1,0) model for generating flows for estimating lake stage probability distributions. The fundamental problem in applying higher order multivariate models to the Great Salt Lake data is that the high cross correlations in the lagged data among the historical time series violated the assumptions of the homogeneous model.

Table 3. Comparison of generated with original data correlation matrices for 1937-1978 series (see also James et al. (1970), Table 30).

Model	ARMA(1,0)	ARMA(1,0)
Data type	Pres. Mod. F.	Pres. Mod. F.
Number of series	20	40
Length of series (years)	125	42
Mean $\Delta\hat{\mu}/\hat{\mu}$	0.00	0.00
Mean $\Delta\hat{\sigma}/\hat{\sigma}$	0.06	0.02
Mean $\Delta\hat{\rho}(1)/\hat{\rho}(1)$	0.08	0.03
Mean $\Delta\hat{h}/\hat{h}$	0.05	0.06
M0 MATRIX		
within 0.25σ	3	8
0.25 to 0.5σ	3	2
0.5 to 0.75σ	2	0
0.75 to 1.0σ	2	0
M1 MATRIX		
within 1.0σ	1	4
0.1 to 0.15σ	2	7
0.15 to 0.25σ	1	5
0.25 to 0.5σ	10	7
0.5 to 1.0σ	11	2
M2 MATRIX		
within 0.5σ	10	11
0.5 to 1.0σ	9	5
1.0 to 1.5σ	4	5
1.5 to 2.0σ	2	2
over 2.0σ	0	2

GENERATED FLOW STATISTICS

The means, standard deviations, lag-one autocorrelation coefficients, and Hurst coefficients for the 1937-1978 flow sequences are given on Table 4 along with transformed statistics for the 3PLN distribution and three correlation matrices. The M_0 correlation matrix shows the cross correlations between pairs of simultaneous values for the various combinations of the five series and is symmetrical. The M_1 matrix shows the correlations for the column data lagged 1 year behind the row data. The M_2 matrix shows correlations for a 2-year lag. Coefficients in the correlation matrices are significantly different than zero (5 percent level) when they exceed 0.320 (James et al., p. 50). Thus 8 of the 10 cross correlation coefficients are statistically significant, 12 of the 25 lag-one coefficients, and 6 of the 25 lag-two coefficients.

The ARMA (1,0) model was used to generate 40 sequence sets for the five variables of length 42 years and also to generate 20 sets of length 125 years. These results with the 40 sequence sets are summarized on Table 5 in a format that can

Table 4. Parameters and correlation matrices for present modified flows (1937-1978).

Variable	Raw Data Statistics: Mean $\hat{\mu}$	Standard Deviation $\hat{\sigma}$	Lag-One Auto-correlation $\hat{\rho}(1)$	Hurst h
Evap.	51.82	4.20	0.244	0.68
Precip.	10.91	2.20	0.025	0.71
Bear	1028690	440990	0.623	0.53
Weber	340210	167270	0.561	0.72
Jordan	266430	79530	0.703	0.66

Variable	Transformed 3PLN Statistics for Stochastic Generation: a	$\hat{\mu}_y'$	$\hat{\sigma}_y'$
Evap.	3.33	0.07	0.34
Precip.	-4.55	1.70	0.03
Bear	0.19	-0.38	0.50
Weber	-1.58	0.63	0.08
Jordan	0.06	-1.75	0.38

M_0 Correlation Matrix

	Evap.	Precip.	Bear	Weber	Jordan
Evap.	1.0	-0.466	-0.335	-0.098	-0.074
Precip.		1.0	0.391	0.378	0.385
Bear			1.0	0.756	0.690
Weber				1.0	0.515
Jordan					1.0

M_1 Correlation Matrix

	Evap.	Precip.	Bear	Weber	Jordan
Evap.	0.234	0.228	-0.176	0.025	-0.099
Precip.	0.185	0.025	0.195	0.090	0.269
Bear	-0.480	0.482	0.652	0.493	0.497
Weber	-0.210	0.474	0.441	0.562	0.250
Jordan	-0.198	0.600	0.557	0.476	0.718

M_2 Correlation Matrix

	Evap.	Precip.	Bear	Weber	Jordan
Evap.	0.243	-0.044	-0.082	0.080	-0.196
Precip.	0.122	0.178	0.154	-0.020	0.335
Bear	-0.163	0.128	0.452	0.299	0.516
Weber	-0.081	0.241	0.299	0.268	0.291
Jordan	0.006	0.368	0.554	0.312	0.664

easily be compared with Table 4. The mean values for the generated sequences (marked Mn on Table 5) can be compared directly with the values on Table 4; the standard deviations (marked SD on Table 5) indicate the variability in results among the sequence sets. The results with the 20 long sequence sets are summarized on Table 6. As was seen from Table 3, the match using fewer but longer sets of generated sequences was generally not so good.

A comparison of Tables 4 and 5 shows that the correlation coefficients for the generated sequences run smaller than those for the "present modified" data. In the M_0 matrix, three of the eight coefficients significant in the data are not significant in the generated sequences. For the M_1 matrix, the same 12 lag-one coefficients were still significant even though all 12 were reduced by an average of 13.5 percent. For the M_2 matrix, five of the six coefficients larger than 0.32 in the data dropped below that figure with the average reduction being 49.7 percent. The ARMA (1,0) model, as would be expected, does not preserve these lagged statistics, and this is the primary reason for continuing to explore methods using the higher order models.

LAKE STAGE PROBABILITIES FROM SIMULATED DATA

The multivariate ARMA (1,0) model calibrated to match present modified flows for the period of 1937 to 1978 was used to generate 100 125-year sets of the five sequences. These intermediate data were then used to generate 100 125-year lake stage sequences beginning from 1978 conditions. This initial condition was a lake stage of 4198.60 ft msl on October 1, 1978, after

Table 5. Means and standard deviations of the values for the parameters and correlation matrices computed for the 40 42-year present modified flow sequences generated using an ARMA (1,0) model.

Variable	Mean $\hat{\mu}$		Standard Deviation $\hat{\sigma}$		Lag One Autocorrelation $\hat{\rho}(1)$		Hurst h	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	51.29	0.98	3.98	0.71	0.174	0.158	0.56	0.08
Precip.	10.86	0.44	2.10	0.24	-0.091	0.146	0.49	0.13
Bear	1,042,400	130,020	400,950	129,260	0.528	0.136	0.69	0.14
Weber	334,810	26,700	165,450	30,130	0.515	0.111	0.69	0.15
Jordan	265,350	22,940	69,590	17,790	0.599	0.152	0.63	0.13

M_0 Correlation Matrix

	Evap.		Precip.		Bear		Weber		Jordan	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	1.00	0.00	-0.486	0.134	-0.229	0.190	-0.018	0.167	-0.009	0.205
Precip.			1.00	0.00	0.273	0.137	0.323	0.141	0.254	0.137
Bear					1.00	0.00	0.733	0.108	0.625	0.130
Weber							1.00	0.00	0.478	0.182
Jordan									1.00	0.00

M_1 Correlation Matrix

	Evap.		Precip.		Bear		Weber		Jordan	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	0.174	0.158	0.281	0.153	-0.108	0.214	0.098	0.177	-0.080	0.222
Precip.	0.285	0.162	-0.091	0.146	0.125	0.168	0.030	0.162	0.174	0.137
Bear	-0.456	0.164	0.429	0.101	0.528	0.136	0.416	0.149	0.397	0.173
Weber	-0.163	0.152	0.462	0.111	0.357	0.166	0.515	0.111	0.167	0.225
Jordan	-0.165	0.210	0.544	0.099	0.438	0.186	0.411	0.211	0.599	0.152

M_2 Correlation Matrix

	Evap.		Precip.		Bear		Weber		Jordan	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	0.286	0.157	-0.058	0.182	-0.087	0.184	-0.020	0.162	-0.094	0.213
Precip.	-0.049	0.187	0.160	0.153	0.060	0.174	0.028	0.177	0.095	0.159
Bear	-0.188	0.206	0.182	0.150	0.245	0.179	0.162	0.202	0.254	0.205
Weber	0.036	0.162	0.119	0.156	0.124	0.181	0.132	0.169	0.030	0.242
Jordan	0.047	0.226	0.280	0.158	0.235	0.239	0.214	0.269	0.329	0.208

receding from a high of 4200.25 ft msl on the previous June 1. Points on the probability distributions for selected years of the annual peak lake stages are shown on Table 7. Annual lows occur about October 1 of each year and average 1.70 feet below the peaks.

Three plots of this probability data were developed from these generated sequences to meet various user needs. First, Figure 1 provides curves from which one can read lake stage probabilities as estimated in 1978 for any desired calendar year from 1979 through 2030. The solid lines crossing any year define the distribution for the high stages for the year. The dashed lines define the lower end of the distribution of low stages for the year. The probabilities are labeled to indicate chance of rising that high (for points higher than the mean) or dropping that low (for points below the mean) rather than as a cumulative probability distribution.

Figure 1 shows a mean long term expected lake level of 4196.0 feet msl and is drawn to smooth out the noise in the simulated estimates for individual years. It also assumes no signficant increase in consumptive water use in the basin from 1965 amounts, and thus the estimated levels would need to be adjusted should significant changes in water use practice occur.

The long-term future mean annual peak lake level estimate reported in Table 7 of 4195.2 ft msl is 2.49 feet lower than the 4197.69 average annual peak over the last 20 years (1961-1980) and 1.22 feet lower than the estimate reported by James et al. (1979). High recent lake levels seem to be associated

Table 6. Means and standard deviations of the values for the parameters and correlation matrices computed for the 21 125-year present modified flow sequences generated using an ARMA (1,0) model.

Variable	Mean $\hat{\mu}$		Standard Deviation $\hat{\sigma}$		Lag One Autocorrelation $\hat{\rho}(1)$		Hurst h	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	51.80	0.53	4.14	0.40	0.210	0.121	0.50	0.05
Precip.	10.91	0.31	2.14	0.11	-0.012	0.137	0.51	0.07
Bear	1,031,410	88,280	427,490	59,470	0.619	0.068	0.69	0.07
Weber	336,190	28,080	168,100	11,180	0.556	0.066	0.69	0.08
Jordan	269,990	17,440	76,050	9,740	0.681	0.069	0.66	0.08

M_0 Correlation Matrix

	Evap.		Precip.		Bear		Weber		Jordan	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	1.00	0.00	-0.494	0.083	-0.284	0.120	-0.077	0.114	-0.065	0.118
Precip.			1.00	0.00	0.381	0.091	0.374	0.092	0.340	0.112
Bear					1.00	0.00	0.750	0.063	0.705	0.076
Weber							1.00	0.00	0.522	0.113
Jordan									1.00	0.00

M_1 Correlation Matrix

	Evap.		Precip.		Bear		Weber		Jordan	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	0.210	0.121	0.242	0.113	-0.095	0.130	0.073	0.104	-0.084	0.120
Precip.	0.208	0.110	-0.012	0.137	0.172	0.092	0.068	0.098	0.231	0.083
Bear	-0.466	0.109	0.487	0.077	0.619	0.068	0.482	0.076	0.485	0.092
Weber	-0.177	0.118	0.468	0.068	0.416	0.087	0.556	0.066	0.234	0.124
Jordan	-0.219	0.105	0.593	0.058	0.561	0.091	0.497	0.103	0.681	0.069

M_2 Correlation Matrix

	Evap.		Precip.		Bear		Weber		Jordan	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Evap.	0.277	0.110	-0.032	0.090	-0.098	0.107	-0.045	0.097	-0.087	0.108
Precip.	-0.053	0.067	0.188	0.072	0.151	0.105	0.122	0.098	0.176	0.101
Bear	-0.227	0.129	0.207	0.093	0.369	0.114	0.278	0.108	0.328	0.126
Weber	0.014	0.106	0.148	0.078	0.197	0.110	0.237	0.102	0.101	0.123
Jordan	-0.040	0.120	0.329	0.095	0.388	0.130	0.330	0.110	0.442	0.109

with the relatively wet period from 1971 through 1976 in which Bear River flows averaged 33 percent and the Jordan 51 percent above normal, and evaporation was about 2 percent below normal.

Figure 2 presents the results of the 100 generated lake stage sequences in terms of the probabilities of the lake level rising to elevations 4201, 4202, 4203, 4204, and 4205 feet msl at least once by various dates through 2030. For example, the simulations indicate a 24 percent chance of a level reaching 4202 sometime before 1990 and a 48 percent chance of reaching that level by 2030.

Figure 3 presents analogous information on the probabilities of the lake dropping to low levels of 4198, 4196, 4194, 4192, 4190, and 4188 feet msl. For example, the simulations indicate an 11 percent chance of the

Table 7. Statistics of the distributions of peak lake stages simulated for various years (1937-1978 data, 100 generated sequences).

Year	Lowest percentile	Lowest decile	Mean	Highest decile	Highest percentile	Standard deviation
1978	-	-	4200.25	-	-	0.00
1983	4193.46	4195.85	4198.91	4202.73	4207.68	2.56
1988	4188.45	4193.14	4197.31	4201.64	4204.84	3.09
1983	4187.86	4192.17	4196.49	4201.05	4205.95	3.65
2003	4183.18	4188.81	4194.75	4199.97	4203.13	3.96
2013	4184.11	4188.34	4194.35	4199.56	4205.10	4.18
Long term*	4185.19	4189.69	4195.20	4200.71	4205.21	4.30

*Estimated by averaging means and standard deviations for levels generated for years 2014 through 2028 and noting a mean skew of close to zero. The percentile and decile values are computed assuming a normal distribution.

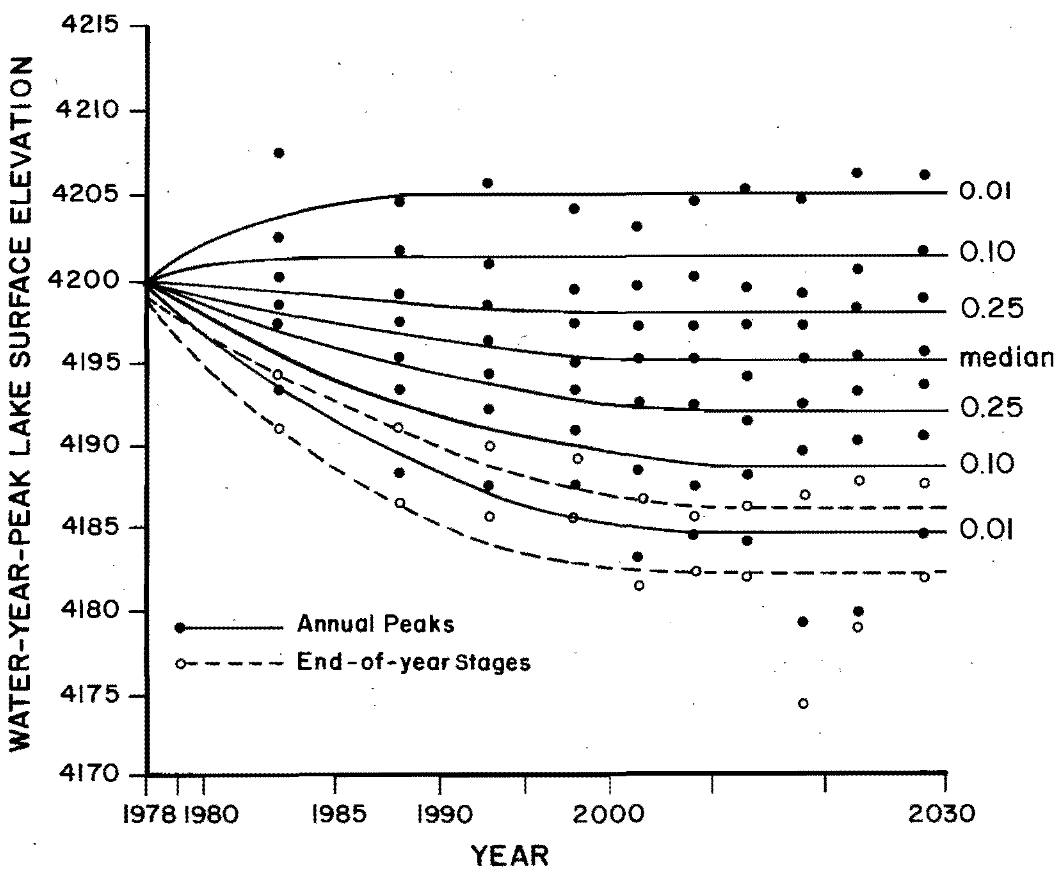


Figure 1. Probabilities of future annual Great Salt Lake levels, given initial conditions as of October 1, 1979.

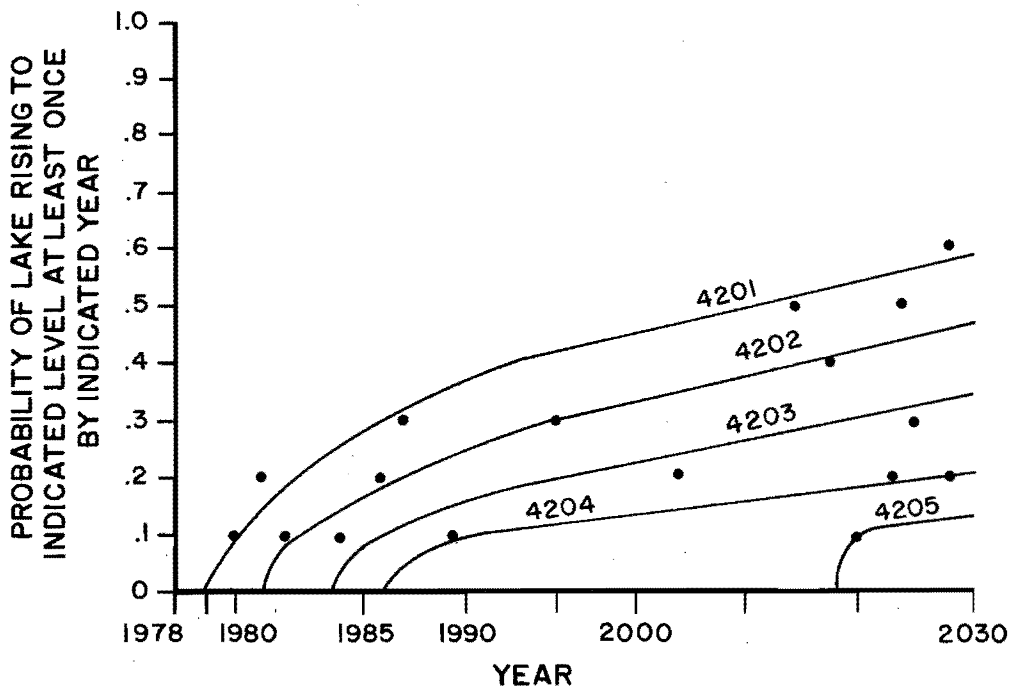


Figure 2. Probabilities of lake rising to various levels by date, given initial conditions as of October 1, 1979.

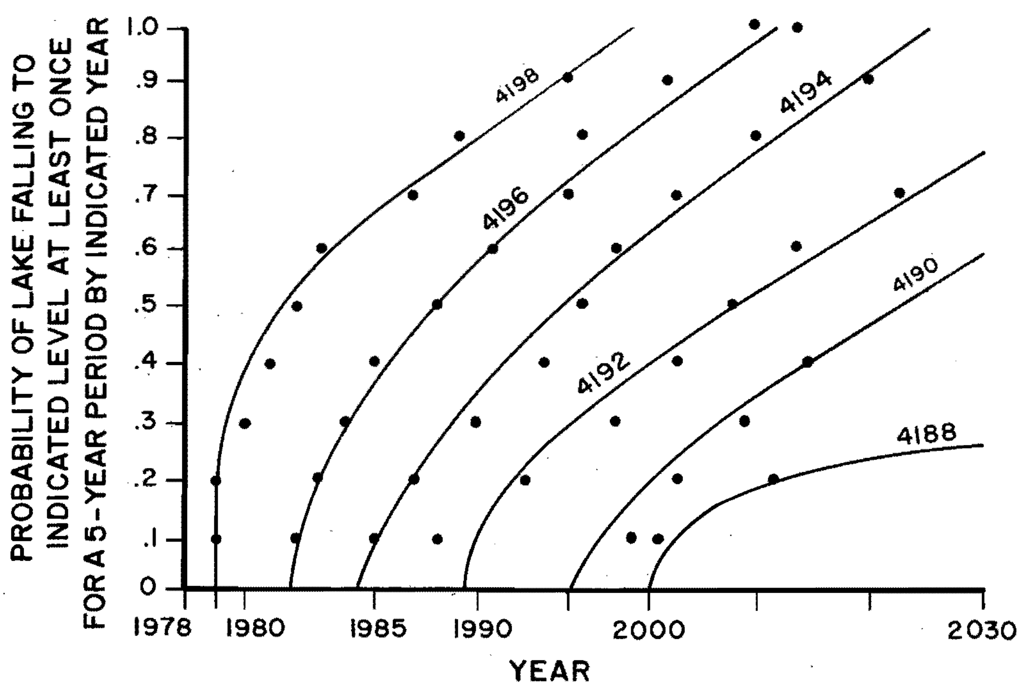


Figure 3. Probabilities of lake falling to various levels by date, given initial conditions as of October 1, 1979.

level dropping to 4192 at least once on or before 1990 and a 75 percent chance of dropping to that level once by 2030.

UPDATING PROBABILITY CURVES

Every year, a new set of streamflow, precipitation, and evaporation data is recorded, and the new information could be used to reestimate the distributions used to estimate future lake stage probabilities. Also, every year the lake has a new annual peak and hence a new starting point along the vertical axis on probability curves having the form of Figure 1. Adjustments for the first sort of change are complex because they require recalculation of the statistics for the five series, refitting of the ARMA model parameters (including the possibility of changing the order for the model), and regenerating and reanalyzing lake stage sequences. Adjustments for the change in initial lake stage alone can be approximated much more simply by curve fitting to the data on Table 7.

Since the process of updating the flow series data is costly and unlikely to have a major effect on the results, repeating the generating process should not be necessary more often than once every 5 to 10 years unless 1) unusually high or low lake inflows occur, 2) implemented lake level control efforts substantially alter the lake water balance computations or 3) significant changes occur in upstream consumptive use or transbasin diversions. Since the process of updating to account for a new initial lake stage can be made relatively simple and will have a major effect on lake level probabilities (particularly in the short term), annual updating is recommended. A simple procedure for doing so is derived below and used to update for known 1980 lake levels.

Austin (1979) developed a method of fitting exponential curves to the data of Figure 1 that could be used for the annual updating. An alternate method is to derive curves from the complete data abridged on Table 7. Both methods could be refined by generating stage sequences from other initial lake stages, but the effort was not considered to be justified given the cost of stage simulation. The computer program presented in James et al. (1979) could be used for this purpose should that be desired.

Table 7 shows the means of the probability distributions of the lake stages generated for various years to decline fairly rapidly toward the long term mean in the early years and then progressively more slowly. The generated means for years 25, 30, and 35 dropped below the long term mean, taken as the average of those means generated for years 35, 40, 45, and 50. This drop and subsequent rise was not considered to be statistically significant. The standard deviations of the distributions increased from zero to 4.3, again averaged for years 35, 40, 45, and 50. The skewness coefficients did not seem to follow any particular pattern but rather to scatter randomly with small values around zero. Hence, the simplified lake stage estimation procedure was based on the assumption that the distribution is normal at all dates.

The simplified estimation procedure uses the normalized plots of the Table 7 data on Figure 4. The means are plotted as fractions of the difference between the long term mean and the initial mean generated for the year in question. Thus the 4198.91 shown in Table 7 for 1983 (year 5) is 73 percent of the departure of 5.05 feet of the 1978 (year 0) from the long term mean. The standard deviations are plotted as fractions of the long term value. Both curves are fitted by eye, and points are listed in Table 8.

ESTIMATES BASED ON 1980 INITIAL CONDITIONS

The 1980 peak lake stage was 4200.45 on June 15, and the October 1 level was 4199.10. The seasonal decline was thus 1.35 feet to October 1, and the lake level dropped another 0.10 feet.

Given the 1980 peak and the fractions in Table 8, lake level probability estimates were made, listed in Table 9, and plotted in Figure 5. A summary leaflet of these findings for distribution to managers of lakeshore property and facilities is presented in Appendix B.

All tabulated stages are annual peaks at Boat Harbor on the South Arm and generally occur in May or June. Annual lows average 1.7 feet below the peak (James et al. 1979, p. 81) and generally occur in October or November. Levels in the North Arm of the Great Salt Lake average 1.5 feet below South Arm highs and 1.0 feet below South Arm lows. The material in Appendix B will be updated and disseminated shortly after the annual peak or about July 1 of each year beginning in 1981.

FURTHER MODEL DEVELOPMENT

Three needs stand out as future directions for Great Salt Lake levels modeling:

1. The process of model identification and application still falls short of the capabilities needs for terminal lake levels modeling. Long term persistence could be a major factor for estimating the levels of a water surface which fluctuates over a large range and over cycles lasting many years. Practical methods are needed for fitting multivariate stochastic models of higher orders.

Table 8. Variation of the expected means and standard deviations of the Great Salt Lake level probability distributions at various numbers of years into the future.

Years into the Future	Fraction of Initial Departure from Long Term Mean Remaining	Standard Deviation as a Fraction of its Long Term Value
0	1.00	0.00
1	0.95	0.19
2	0.89	0.30
3	0.84	0.38
4	0.78	0.44
5	0.73	0.50
7	0.61	0.59
10	0.42	0.70
15	0.23	0.82
20	0.11	0.90
25	0.05	0.95
30	0.01	0.98
35	0.00	1.00

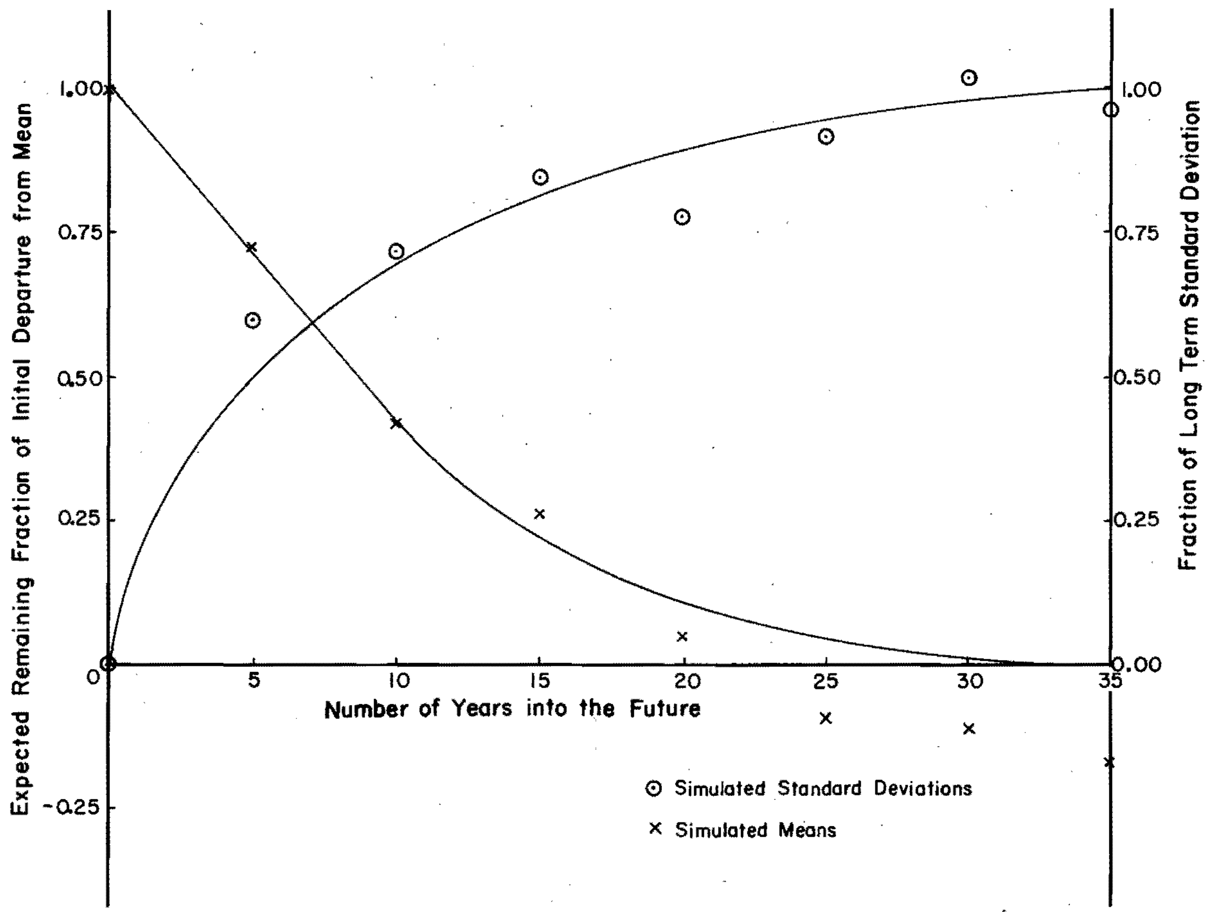


Figure 4. Variations of the means and standard deviations of the lake level probability distributions over the 35-year period in which Great Salt Lake level estimates are influenced by known initial conditions.

Table 9. Probability distribution of annual high levels given 1937-1978 data adjusted to reflect 1965 land use and the 1980 lake level high of 4200.45 feet msl.

Year	Probabilities of Exceeding			Mean	Probabilities of Dropping Below		
	0.01	0.10	0.25		0.25	0.10	0.01
1980	-	-	-	4200.45	-	-	-
1981	4202.09	4201.24	4200.74	4200.19	4199.64	4199.14	4198.29
1982	4202.87	4201.52	4200.74	4199.87	4199.00	4198.22	4196.87
1983	4203.41	4201.70	4200.71	4199.61	4198.51	4197.52	4195.81
1984	4203.70	4201.72	4200.58	4199.30	4198.02	4196.88	4194.90
1985	4204.03	4201.79	4200.48	4199.03	4197.58	4196.27	4194.03
1987	4204.30	4201.65	4200.11	4198.40	4196.69	4195.15	4192.50
1990	4204.41	4201.27	4199.44	4197.41	4195.38	4193.55	4190.41
1995	4204.62	4200.93	4198.79	4196.41	4194.03	4191.99	4188.20
2000	4204.79	4200.72	4198.39	4195.78	4193.17	4190.82	4186.77
2005	4204.97	4200.70	4198.22	4195.46	4192.70	4190.22	4185.95
2010	4205.06	4200.65	4198.19	4195.25	4192.41	4189.85	4185.44
2015+	4205.21	4200.71	4198.10	4195.20	4192.30	4189.69	4185.19

All elevations in feet, mean sea level

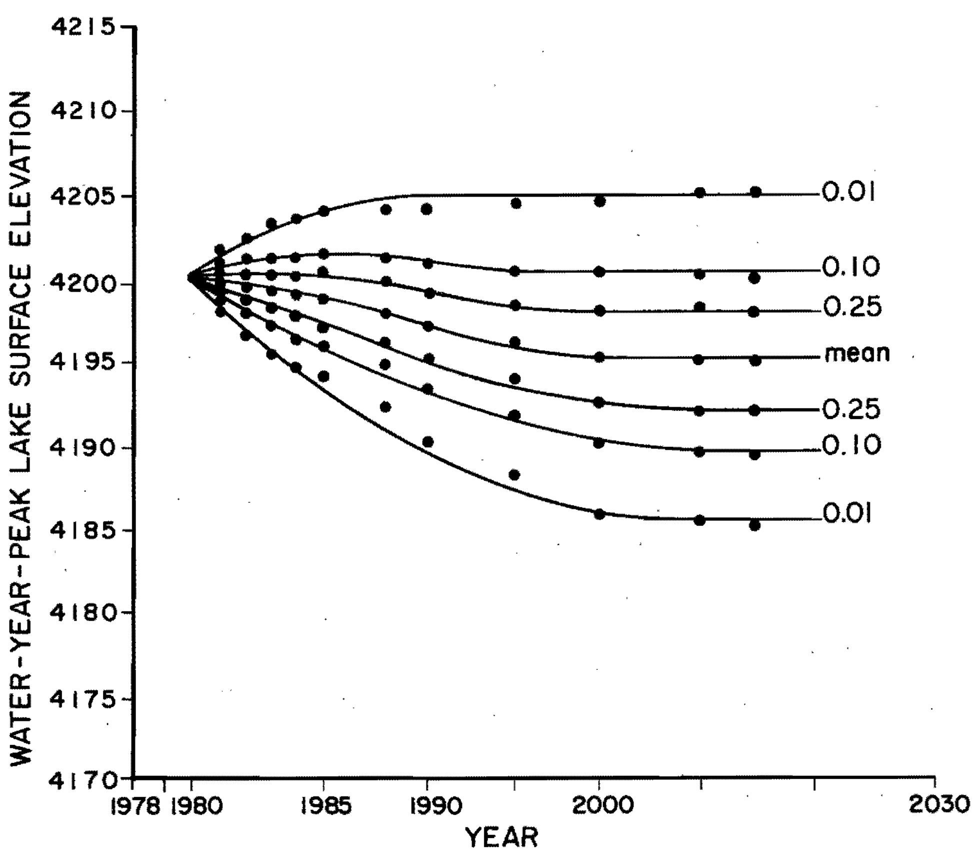


Figure 5. Probability curves of annual Great Salt Lake level highs given 1980 high of 4200.45 feet msl.

2. New hydrologic data (additional years of record) and hydrologic change (major urbanization, water project development, lake level control efforts, etc.) will change the model used here for lake level probability estimation. At about every 10 years, the new data should be collected, checked for homogeneity with the old; and the stochastic lake level generating model should be recalibrated accordingly. Thereafter, the recalibrated model should be used for estimating lake level probabilities.

3. Appendix B provides basic lake level forecasting information. As user needs become better understood, it will be possible to change the content and form of the presentation to satisfy user needs more efficiently.

BIBLIOGRAPHY

Arnow, Ted. 1978. Water budget and water-surface fluctuations, Great Salt Lake, Utah. U. S. Geological Survey, Salt Lake City, Utah. Open File Report 78-912. 21 p.

Austin, Lloyd H. 1979. Unpublished notes on Great Salt Lake level probability forecasting. Utah Division of Water Resources, Salt Lake City, Utah. 7 p.

Box, G. E. P., and G. M. Jenkins. 1970. Time series analysis: Forecasting and control. Holden-Day, Inc. San Francisco, 553 p.

Burges, S. J., and Kiyoshi Hoshi. 1978. Approximation of a normal distribution by a three-parameter log normal distribution. Water Resources Research 14(4):620-622.

James, L. Douglas, David S. Bowles, W. Robert James, and Ronald V. Canfield. 1979. Estimation of water surface elevation probabilities and associated damages for the Great Salt Lake. UWRL/P-79/03. 182 p.

U. S. Water Resources Council. 1976. Guidelines for determining flood flow frequency. Bulletin 17. Washington, D. C. 197 p.

Utah Division of Water Resources. 1977. Great Salt Lake hydrologic system management alternatives report. Salt Lake City, Utah. 92 p.

Waddell, K. M., and F. K. Fields. 1977. Model for evaluating the effects of dikes on the water and salt balance of Great Salt Lake, Utah. Utah Geological and Mineral Survey. Water Resources Bulletin 21. 54 p.

APPENDIX A

HOMOGENEOUS MULTIVARIATE ARMA PROCESSES

by

R. V. Canfield, C. H. Tseng, W. R. James

Univariate time series have been found useful in hydrologic modeling. However, modeling of water systems requires multivariate (MV) methods which brings additional complexity to an already difficult problem. Interpretation of the model in terms of physical phenomena is often intractable. However, if a method works, it is hard to avoid its use, even if no physical interpretation exists. Thus, analysts are accused of "black box hydrology."

As more complex models become available to the analyst, the need for some direction in choice increases. This direction can only come from an understanding of the relationships between and within the physical system and the mathematical model. Box-Jenkins identification procedures (1970) aid in the univariate case, but as will be seen, this procedure is of little or no use with general multivariate (MV) systems.

In this paper, the general MV ARMA(p,q) process is viewed in light of some reasonable assumptions with respect to many water systems. The resulting mathematical model is considerably simpler as compared with the general case. This simplification permits estimation in cases previously impractical.

The simplified model is denoted a "homogeneous model" for reasons which will be apparent. Its simplicity suggests that it has undoubtedly been used before (e.g. Nelson 1976). The primary contribution here lies not in the model, but in the hypotheses and their interpretation with respect to physical systems.

The general MV ARMA(p,q) process is reviewed and critically examined. The problems of estimation are considered. The homogeneous process is defined heuristically and then the analytical description is derived. Estimation is considered next, and the results of application to real and Monte Carlo data are given in the last section.

General MV ARMA(p,q) Processes

The model for the general ARMA(p,q) processes may be written

$$\underline{Z}_t = A_1\underline{Z}_{t-1} + A_2\underline{Z}_{t-2} + \dots + A_p\underline{Z}_{t-p} + \underline{\varepsilon}_t - B_1\underline{\varepsilon}_{t-1} - B_2\underline{\varepsilon}_{t-2} - \dots - B_q\underline{\varepsilon}_{t-q} \qquad \dots\dots\dots\dots (1)$$

where $\underline{Z}_{t-i}$ is an n component random vector with unit variances and zero means. The error vector $\underline{\varepsilon}_{t-1}$ is assumed to be independent in time, have zero means and variance-covariance matrix Σ. The $A_{i,\, i=2,3,\dots,p}$, and $B_{j,\, j=1,2,\dots,q}$ are n x n matrix parameters. The value p represents the order of the autoregressive portion of the model and q the order of the moving average part. The number of individual parameters of this model is $n^2(p+1+\frac{1}{2})+\frac{n}{2}$.

In the application of multivariate time series models to real problems, a major obstacle seems to be estimation of parameters from data. Two methods have been used, the method of moments (O'Connell 1974) and the maximum likelihood method (Ledolter 1978). The method of moments has two difficulties. First, it is restricted to multivariate ARMA models with p=1 and q=1. Second, matrix equations necessary to estimate parameters may not have real solutions.

Lack of a real solution may denote an improper choice of model. However, Monte Carlo experience in which an ARMA (1,0) model was used to generate 100 data sets provides an insight into the problem. In 7 of the 100 data sets, the method of moments failed to provide solutions (James et al. 1979). This experiment indicates that the problem is at least occasional.

The method of moments for univariate series has one feature which can be important. It permits the inclusion of the Hurst coefficient as a parameter. This feature is not available with maximum likelihood estimation.

Maximum likelihood estimation is shown to be more efficient than the method of moments for ARMA parameters (Ledolter 1978). However, a solution requires numerical optimization routines on $n^2(p+q)$ parameters. Consider for example a five component ARMA (1,1) model. Estimation requires simultaneous optimization of 50 parameters. Even if a solution could be guaranteed, the computer bill could be very high. It seems that neither method of estimation is satisfactory.

There is an inconsistency of multivariate ARMA models noted by Ledolter (1978). He has shown that it is not necessary to have the individual series in a multivariate autoregressive process follow the univariate process of the same order. Isolating a single component of (1) easily shows that the univariate model has no ARMA representation in the classical form (e.g. Box and Jenkins 1970). There is a natural tendency to build or generalize from the simple to the complex multivariate system. It seems logical to examine the simpler univariate components first (e.g., Parzen 1969, Haugh 1972). This procedure is apparently invalid for general ARMA processes.

In the next section the conditions which insure that the characteristics of the individual univariate processes (or any subset of the components) are preserved in the multivariate framework are defined. ARMA models which satisfy this condition have been termed homogeneous.

Homogeneous ARMA Processes

As noted in the previous section, univariate components of multivariate ARMA models are not necessarily univariate ARMA processes. In this section, a special case of the ARMA(p,q) model is defined which possesses the property of preserving the ARMA order of every subset of components of the system. Estimation of parameters for this homogeneous model is then discussed.

Because it is possible that the components of a system may not all be of the same order in p and q, there may be some ambiguity in the order of the multivariate model.

The order of a homogeneous ARMA model is defined as follows. Let p_i and q_i $i=1,2,\ldots,n$ be the order in AR and MA parts respectively of the i^{th} component of the system. The AR order of the system is defined as $p=\max_i\{p_i, i=1,2,\ldots,n\}$ and the MA order of the system is defined as $q=\max_i\{q_i, i=1,2,\ldots,n\}$.

Mathematical definition of the homogeneous model is provided in the following theorem.

The theorem states the conditions under which an ARMA (p,q) model is homogeneous. Note that the conditions are "if and only if," i.e., no further simplifications nor generalizations can be made and still preserve the homogeneous character of the model.

Theorem. The ARMA (p,q) model (Equation 1) is homogeneous if and only if the matrices A_i, $i=1,2,\ldots,p$ and B_i, $i=1,2,\ldots,q$ are diagonal.

Proof. The "if" portion of the proof is immediate. Every vector subset of the original variable has the model formed by taking the corresponding subset of the rows of the A_i and B_i matrices. Since these are diagonal, the subset matrix of the rows can be redefined to be diagonal and the $\underline{\varepsilon}$ vector shortened to include only those elements necessary.

The "only if" portion of the proof is most easily accomplished by assuming that one (or more) of the A_i and/or B_i have one or more off-diagonal elements that are nonzero. If there is one off-diagonal element of an A_i that is nonzero, then it follows directly from the proof of the inconsistency referred to in Ledolter (1978) that there exists a marginal univariate model with autoregressive order greater than p.

Suppose there is one nonzero off-diagonal element of B_i, $i=1,2\ldots,q$. Without loss of generality, suppose it is the 1,2 element of B_1. The first element of $\underline{Z}_t$ may be written

$$Z_{1,t}=\underline{a}'_{1,1}\underline{Z}_{t-1}+\underline{a}'_{2,1}\underline{Z}_{t-2}+\ldots+\underline{a}'_{p,1}\underline{Z}_{t-p}+\varepsilon_{1,t}-\underline{b}'_{1,1}\underline{\varepsilon}_{t-1}$$
$$-\underline{b}'_{2,1}\underline{\varepsilon}_{t-2}-\ldots-\underline{b}'_{q-1}\underline{\varepsilon}_{t-q} \quad (2)$$

where $Z_{1,t-i}$ represents the first element of $\underline{Z}_{t-i}$, $\underline{a}'_{i,1}$ and $\underline{b}'_{i,1}$ represent the first row of A_i and B_i respectively and $\underline{\varepsilon}_{1,t}$ the first element of ε_t.

Thus, the vector $\underline{b}'_{1,1}$ in Equation 2 results in the inclusion of the second element of $\underline{\varepsilon}_{t-1}$ into the model for $\underline{Z}_{1,t}$. Since this element cannot be incorporated in the "error" term at time t (i.e., $\underline{\varepsilon}_{1,t}$) the error structure does not have the univariate moving average form.

It has been shown that no off-diagonal elements of A_i, $i=1,2,\ldots,p$ nor B_i, $i=1,2,\ldots,q$ can be nonzero or the ARMA (p,q) model will not preserve at least one univariate subset. The theorem is proved.

Maximum likelihood estimation for the nonhomogeneous model inherits an operational difficulty. Simultaneous numerical optimization is required which involves $n^2(p+q)$ parameters. Thus, a system with five variables requires numerical optimization for 50 parameters. By contrast, the homogeneous model involves n(p+q+n) parameters and more important, simultaneous numerical optimization is required on only p+q parameters at a time. Thus, the five variable homogeneous ARMA (1,1) case requires five separate optimizations each with two variables.

The most notable effect due to the diagonal nature of the A_i and B_i matrices is that the historical contributions to the present value of a given element of the random vector $\underline{Z}_t$ is limited to the historic values of that same element. This does not at all imply independence because the present "error" (i.e., $\underline{\varepsilon}_t$) can be correlated random vector. As time progresses, e.g., as time t becomes t+1, the values in $\underline{\varepsilon}_t$, as they contribute to the then present values of $\underline{Z}_t$, contribute to the new present value only as modified by a constant specific to each element in $\underline{\varepsilon}_t$. For example, suppose the homogeneous model were used to model yearly volume of several streams in a large basin. The restrictions imply that for a given stream, the present flow volume is due to the present and past "errors" (precipitation) and past streamflow in that stream drainage area and no other areas. The "errors" may be correlated, but what actually happened in the past in the given area is what influences streamflow in that area. This implies that if, for example, one stream in the system receives considerable recharge from ground-water originating in the area of another stream in the system, the homogeneous model would not be applicable.

Parameter Estimation

Although maximum likelihood estimation is more efficient than the method of moments for multivariate models, it is not necessarily best for the univariate case. The method of moments with the bias correction given by O'Connell (1974) has the distinct advantage of preserving observed long term persistence in synthetic sequences. This persistence is usually measured by the Hurst coefficient. In O'Connell's formulation, the Hurst coefficient is a parameter in the univariate case. It is possible using the homogeneous model to take advantage of O'Connell's method to estimate the n(p+1) parameters which are the diagonal elements of the A_i, $i=1,2,\ldots,p$ and B_i, $i=1,2,\ldots,q$. Then the conditional maximum likelihood method, or Equation 4, can be used to estimate the n^2 elements of the variance-covariance matrix of the $\underline{\varepsilon}(t)$ random variable. This estimation technique has the advantage of preserving a measured persistence in the univariate series and of requiring no numerical optimization procedures. However, it requires information which may be difficult to obtain in many cases. O'Connell (1974) has prepared tables for the ARMA (1,1) case only.

Estimation of the parameters of the homogeneous model can be based upon standard univariate methods. The n x n matrices A_i and B_j may be represented,

$$A_i=\mathrm{diag}\{a_{i,1},a_{i,2},\ldots,a_{i,n}\}$$

$$i=1,2,\ldots,p$$

and

$$B_j=\mathrm{diag}\{b_{j,1},b_{j,2},\ldots,b_{j,n}\}$$

$$j=1,2,\ldots,q.$$

According to the definitions of p and q it follows that $a_{i,k}=0$ for $i>p_k$ and $B_{j,k}=0$ for $j>q_k$. At least one value of $a_{p,k}$ and one value of $b_{q,k}$ is greater than zero for $k=1,2,\ldots,n$. The k^{th} component of (1) is written

$$z_{k,t}=\sum_{i=1}^{p_k} a_{i,k}z_{k,t-1}+\varepsilon_{k,t}-\sum_{j=1}^{q_k} b_{k,j}\varepsilon_{k,t-j}$$

The only dependency between $z_{k,t}$ and $z_{k',t}$ is the result of the dependency between $\varepsilon_{k,t}$ and $\varepsilon_{k',t}$ where $k\neq k'$. Thus all of the presently available univariate procedures for estimating and identifying forms used in the estimation of p_k, q_k, $a_{i,k}$ and $b_{j,k}$, $k=1,2,\ldots,n$ may be used in the homogeneous case. There remains to estimate the variance-covariance matrix of ε_t. Conditional maximum likelihood estimation (Ledolter 1978) may be used.

However, another more direct method which is convenient is demonstrated here for the ARMA(1,1) case.

It follows from (1) with p=q=1 that the lag zero correlation matrix (M_o) can be written:

$$M_o=\Sigma+(A-B)\Sigma(A-B)'+A(A-B)\Sigma(A-B)'A'+A^2(A-B)\Sigma(A-B)'A'^2+\ldots \qquad (3)$$

Let $A = \mathrm{diag}\,(a_1, a_2,\ldots,a_n)$

$B = \mathrm{diag}(b_1,b_2,\ldots,b_n)$

$\Sigma = (s_{ij})$

Then the ij^{th} element of M_o (m_{oij}) may be written

$$m_{oij} = s_{ij}+\left(\sum_{k=0}^{\infty} a_i^k a_j^k\right)(a_i-b_i)(a_j-b_j)s_{ij}$$

$$= \left(1+\frac{(a_i-b_i)(a_j-b_j)}{1-a_ia_j}\right)s_{ij}$$

$$= \left(\frac{1-a_ib_j-a_jb_i+b_ib_j}{1-a_ia_j}\right)s_{ij}$$

It follows that,

$$s_{ij} = \frac{m_{oij}(1-a_ia_j)}{(1-a_ib_j-a_jb_i+b_ib_j)} \qquad (4)$$

The elements of Σ are estimated by substituting appropriate estimated values of a_i,a_j,b_i,b_j and m_{oij} into (4).

It is evident that the estimated homogeneous model will preserve the characteristics of the univariate systems in accordance with the method of univariate estimation used. It is also evident that the lag zero correlation matrix as estimated by M_0 is preserved in the homogeneous model.

Monte Carlo Study

Sensitivity of the homogeneous model to data which originate from a nonhomogeneous source is investigated in this section. The method used here to measure the performance of a model is to compare the M_0, M_1 and M_2 moments (calculated from the estimated parameters) with those of the model used to generate data. Since it is possible for very different appearing models to have similar moments, comparison of estimated parameters with the population values is less informative. However, tabulated results of the estimated model parameters are given for completeness.

For simplicity, this study is restricted to specific two dimensional multivariate ARMA(1,1) models. Thus, it is not possible to draw broad inferences. However, the sensitivity of the cases studied certainly provide insight.

Two models were used to generate data. The first is homogeneous (i.e., diagonal parameter matrices). The other model has nonzero off-diagonal parameter values which represent increasing nonhomogeneity. The model parameters, (A,B,Σ) and the lag 0,1 and 2 correlation matrices (M_0,M_1,M_2) for the two cases are given in Table A-1. For each model, 400 series each of length 80 were generated. Estimates of A,B and Σ were computed for each series by maximum likelihood, assuming a homogeneous model. Then M_0,M_1 and M_2 were computed using the estimates of A, B and Σ. The mean values of these parameters and the mean squared error (i.e., the mean squared deviation of the estimated parameter values from the generating model values) are shown in Table A-2. The mean value of the estimate is used to judge bias in the method. The MSE measures precision of estimation.

In general, it is evident that MSE of the homogeneous model works well if the data are from a homogeneous (or nearly homogeneous) system. Model 2 seems to show a slight increase in MSE for the off-diagonal elements of M_2.

The generating models used here were only two dimensional ARMA(1,1). Therefore, generalization to higher order models is not valid from this study. However, univariate methods are used in the homogeneous model. Since there is a great deal of accumulated experience in univariate estimation, it seems safe to extend these results to the more general homogeneous models.

Application

Data from the three major rivers flowing into the Great Salt Lake are used in this section to illustrate the method and some of its characteristics. Yearly river volume for each stream is recorded for the years 1937 to 1978 and has been adjusted to represent present management conditions (Table A-3). Data from each stream were subjected to standard Box-Jenkins identification and estimation procedures. The ARMA(1,1) model was selected in all cases. Estimated A and B matrices from this procedure and Σ using (4) are shown in Table A-4.

It is evident from the method of estimation of A and B that the lag 1 and 2 correlations of the univariate components (i.e., the diagonal elements of M_1 and M_2) should be close to the observed values. It also follows from (4) that the observed M_0 will be preserved exactly in the mathematical model. The M_1 and M_2 matrices as computed from the estimated model parameters and as observed from the raw data are shown in Table A-5.

As expected, the diagonal elments of the M_1 and M_2 computed from model parameters agree closely with the observed values with the possible exception of the Weber River. However, the computed value of the lag 1 correlation for the Weber River is within a 95 percent confidence interval for correlation as determined from the raw data.

The off diagonal elements of M_1 and M_2 (computed and observed) can be compared to evaluate the conformity of the physical process to the homogeneous assumption of the model. The 95 percent confidence interval for the lag 2 cross correlation between the Bear and Jordan Rivers is approximately $-0.30<\rho_2<0.74$. The computed value of 0.291 is outside of this range, which indicates a possible violation of the homogeneous assumption.

Conclusions

The problems of estimation inherent with both MLE and the method of moments for the general multivariate model make the homogeneous model attractive. It appears to be the only model that is computationally feasible for multivariate, multilag systems. Multiple lags are required to reproduce series where lag 2 (or higher) correlations exceed lag 1 correlations because of long aquifer travel time or long carryover storage periods in large reservoirs.

The restrictions imposed by the homogeneous model can be characterized from two viewpoints. First, what are the physical

Table A-1. Parameters of models used for generating series.

Parameters	Model 1		Model 2	
A_1	0.9	0.0	0.9	0.1
	0.0	0.8	0.1	0.8
B_1	0.4	0.0	0.4	0.1
	0.0	0.2	0.1	0.2
Σ	.432	.232	.432	.232
	.232	.500	.232	.500
M_o	1.0	.302	1.0	.303
	.302	1.0	.303	1.0
M_1	.726	.116	.727	.123
	.139	.763	.147	.700
M_2	.654	.105	.669	.181
	.112	.560	.190	.572

Table A-2. Monte Carlo results MSE.

Statistic/ Parameter	Model 1		Model 2	
Average Estimated A_1	.814	0	.879	0
	0	.737	0	.806
MSE A_1	.024	0	.009	.01
	0	.022	.01	.015
Average Estimated B_1	.326	0	.369	0
	0	.158	0	.253
MSE B_1	.031	0	.021	.01
	0	.034	.01	.034
Average Estimated Σ	.433	.225	.439	.218
	.225	.502	.218	.489
MSE Σ	.004	.003	.005	.003
	.003	.006	.003	.006
Estimated Average M_o	1.0	.332	1.0	.273
	.332	1.0	.273	1.0
MSE M_0	0	.007	0	.008
	.007	0	.008	0
Average Estimated M_1	.638	.123	.718	.105
	.148	.649	.116	.679
MSE M_1	.026	.001	.018	.001
	.002	.013	.003	.015
Average Estimated M_2	.530	.100	.639	.091
	.107	.486	.090	.556
MSE M_2	.041	.001	.027	.001
	.001	.024	.011	.023

characteristics of a system which is homogeneous, and second, what characteristics of a nonhomogeneous system are preserved (or not preserved) when the homogeneous model is used to describe the system.

In the first case, correlation of stream flows are determined only from the yearly shocks, ε_t (e.g., precipitation) and not from events or processes which may alter the relationships at a later time.

The second characterization is that the lag zero correlation structure of a nonhomogeneous system is preserved by the homogeneous model. In addition, the characteristics of the individual univariate components of the system are preserved in accordance with the method of univariate estimation used.

References

Box, G.E.P., and G.M. Jenkins. 1970. Time series analysis: forecasting and control. Holden-Day, Inc., San Francisco, CA p. 553.

Haugh, L.D. 1972. The identification of time series interrelationship with special reference to dynamic regression. Ph.D. Thesis, University of Wisconsin, Madison, WI.

James, L. D., D. S. Bowles, W. R. James, and R. V. Canfield. 1979. Estimation of water surface elevation probabilities and associated damages for the Great Salt Lake. Water Resources Planning Series, UWRL/P-79/03. Utah Water Research Laboratory, Utah State University, Logan, Utah, p. 182.

Ledolter, J. 1978. The analysis of multivariate time series applied to problems in hydrology. Journal of Hydrology, 36:327-352.

Nelson, C. R. 1976. Gains in efficiency from joint estimation of systems of autoregressive-moving average processes. Journal of Econometrics, 4:331-348, Nov.

Table A-3. Flow Data 1937-1978.

River								
Bear					767800	811800	600200	468600
	524800	707800	875400	697400	812000	1041400	1070600	1167800
	1020000	1741000	1630200	1886400	1043800	539200	611200	879800
	954400	1058800	602200	569800	405200	874800	629400	916400
	1091000	1154400	1054200	1059200	1215400	875800	2067800	2070600
	1485000	1505000	1457000	1827000	689300	945100		
Weber					412200	438400	283100	166900
	180100	411500	436100	344600	370400	483600	322900	447200
	512600	633400	602000	659100	444300	151600	119200	301300
	327100	425600	112500	112100	60500	210400	145900	312300
	337700	116000	175100	211100	492400	233500	496800	522400
	451000	529900	560000	353546	77580	305100		
Jordan					219200	236200	239100	225500
	235200	280000	264000	263800	235100	282100	236500	284000
	270300	258100	252600	235200	280000	264500	191400	211700
	224000	252200	220800	180900	131900	168100	157900	199400
	239700	231200	240200	278100	373400	389200	379900	374400
	360000	430400	395000	496411	218600	284720		

Table A-4. Estimates of A, B and Σ.

Univariate Maximum Likelihood, ARMA(1,1)

PARAMETER	RIVER	Bear	Weber	Jordan
A	Bear	.674	0	0
	Weber	0	.498	0
	Jordan	0	0	.851
B	Bear	.085	0	0
	Weber	0	.094	0
	Jordan	0	0	.306
Σ	Bear	.611	.557	.394
	Weber	.557	.822	.373
	Jordan	.394	.373	.481

Nelson, C. R. 1976. Gains in efficiency from joint estimation of systems of autoregressive-moving average processes. Journal of Econometrics, 4:331-348, Nov.

O'Connell, P. E. 1974. Stochastic modeling of long-term persistence in streamflow sequences. Civil Engineering Department, Imperial College, London, England. p. 309.

Parzen, E. 1969. Multiple time series modeling. Multivariate Analysis II (P. Krishnaiah, Editor), Academic Press, New York.

Table A-5. Computed and observed M_1 and M_2.

		Computed			Observed		
Parameter	River	Bear	Weber	Jordon	Bear	Weber	Jordan
M_1	Bear	.622	.324	.467	.652	.493	.497
	Weber	.462	.421	.324	.441	.562	.250
	Jordan	.432	.221	.704	.557	.467	.718
M_2	Bear	.419	.161	.397	.452	.299	.516
	Weber	.312	.210	.276	.299	.268	.281
	Jordan	.291	.110	.599	.559	.312	.664

APPENDIX B

1980 FORECAST ON GREAT SALT LAKE LEVELS

Utah Water Research Laboratory
Utah State University
Logan, Utah 84322

Over the years since the first records in 1851, the surface level of the Great Salt Lake has varied from a low of 4191.5 feet msl in 1964 to 4211.8 feet msl in 1873. In any given year, lake levels rise to a high toward the end of the spring snowmelt period and then fall an average of about 1.7 feet to fall lows.

The recent South Arm high was 4202.25 feet msl in the spring of 1976. Since then the level dropped to 4200.30 (November 1976), rose to 4200.75 (April 1977), dropped to 4198.60 (December 1977), rose to 4200.25 (June (1978), dropped to 4198.40 (November 1978), rose to 4199.90 (May 1979), dropped to 4197.50 (October 1979), rose to 4200.45 (June 1980), and dropped to 4199.00 (October 1980). Thus a general downward trend from 1976 to 1979 was reversed with a rising trend in 1980.

Industrial plants processing minerals from the lake can use lake level forecasts in deciding whether they should be moving operations to higher ground, raising protective levees, continuing present operations unchanged, or possibly extending lines to collect brines from lower levels. Decisions on the operation and maintenance of railroads and highways, waterfowl areas, recreation facilities, and other properties close to the lake would similarly profit from reliable forecasts.

We cannot know the future, but we can estimate probable lake levels by date from the lake level trends which have occurred in the past. The Utah Water Research Laboratory, in a project sponsored by the Office of Water Research and Technology, employed advanced techniques in stochastic hydrology to establish probabilities for possible high and low lake levels in coming years. We cannot forecast an exact level for any date in the future, but available models provide numerical estimates of what levels are so high or so low that they have, for example, only one chance in 100 of occurring.

These forecasts are based on assumptions of water use in the tributary drainage basin continuing at current levels, no major climatic changes, and no construction of levees or pumping schemes for lake level control. Given the June 1980 lake level at Boat Harbor near Saltair on the lake's south end of 4200.45 feet msl, the distribution of probable future annual lake level highs follows the data or curves on the back of this sheet. For an example as to how to read this information, there is one chance in 100 of the lake high in 1985 reaching 4204.03 feet, the most likely level is 4199.03 feet, and one chance in 100 exists of the lake high being as low as 4194.03 feet.

These lake levels assume a level water surface on the south arm of the lake. Winds on the lake induce seiches which can raise the level on the downwind shore by 2 or 3 feet. Wind blown waves can cause water to splash another 3 or 4 feet higher along gradual shoreline slopes of 1 foot vertical rise in a horizontal distance of 30 feet. Splash levels can exceed 20 feet on the causeway or more steeply sloped levees near the lake.

The data on the attached table and figure are annual highs for the south arm of the Great Salt Lake. These highs usually occur between March and June with the latest peaks coming in the wetter years when the greater snowpack takes longer to melt and run off. The annual lows usually occur between September and November and average 1.70 feet lower than the tabulated values. Annual peaks on the north arm average about 1.50 feet lower than those on the south arm, and north arm annual lows average 1.00 feet lower.

The probability forecasts for Great Salt Lake levels contained herein are based on the best data available as of October 1, 1980. About June 1981, data on another annual high will become available, and a revised forecast will be issued shortly thereafter.

Table 1. Probability distribution of annual high levels given 1937-1978 data adjusted to reflect 1965 land use and the 1980 lake level high of 4200.45 feet msl.

Year	Probabilities of Exceeding			Mean	Probabilities of Dropping Below		
	0.01	0.10	0.25		0.25	0.10	0.01
1980	-	-	-	4200.45	-	-	-
1981	4202.09	4201.24	4200.74	4200.19	4199.64	4199.14	4198.29
1982	4202.87	4201.52	4200.74	4199.87	4199.00	4198.22	4196.87
1983	4203.41	4201.70	4200.71	4199.61	4198.51	4197.52	4195.81
1984	4203.70	4201.72	4200.58	4199.30	4198.02	4196.88	4194.90
1985	4204.03	4201.79	4200.48	4199.03	4197.58	4196.27	4194.03
1987	4204.30	4201.65	4200.11	4198.40	4196.69	4195.15	4192.50
1990	4204.41	4201.27	4199.44	4197.41	4195.38	4193.55	4190.41
1995	4204.62	4200.93	4198.79	4196.41	4194.03	4191.99	4188.20
2000	4204.79	4200.72	4198.39	4195.78	4193.17	4190.82	4186.77
2005	4204.97	4200.70	4198.22	4195.46	4192.70	4190.22	4185.95
2010	4205.06	4200.65	4198.19	4195.25	4192.41	4189.85	4185.44
2015+	4205.21	4200.71	4198.10	4195.20	4192.30	4189.69	4185.19

All elevations in feet, mean sea level

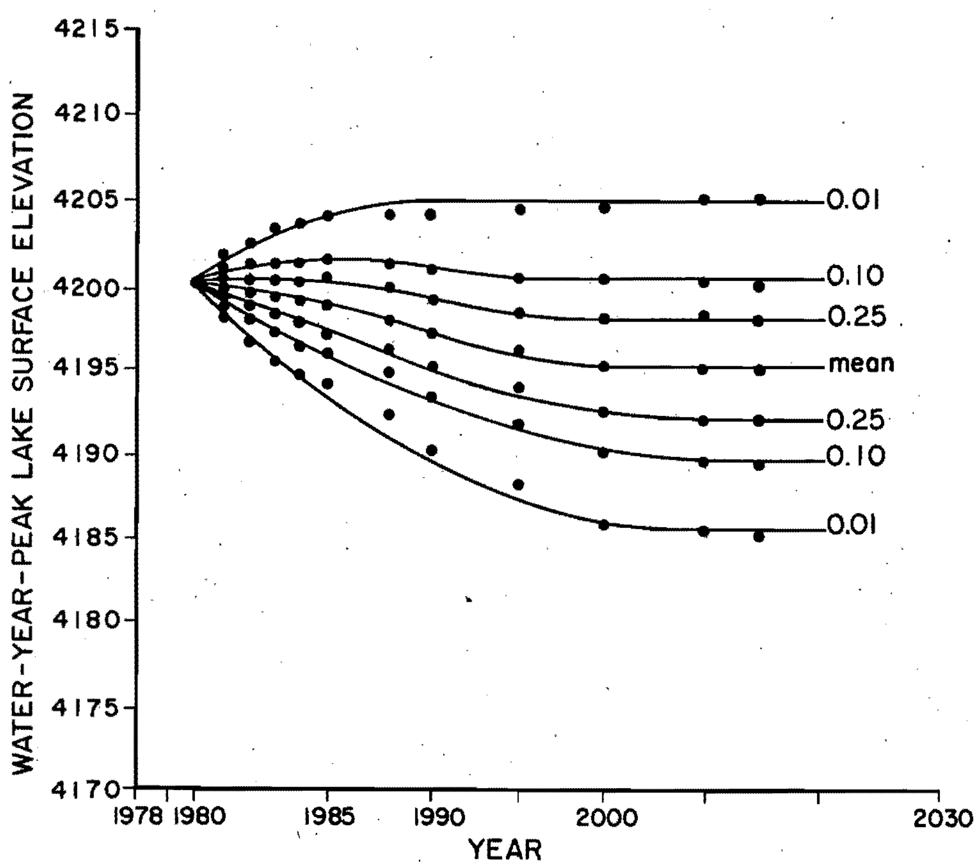


Figure 1. Probability curves of annual Great Salt Lake level highs given 1980 high of 4200.45 feet msl.