188
11-28-78

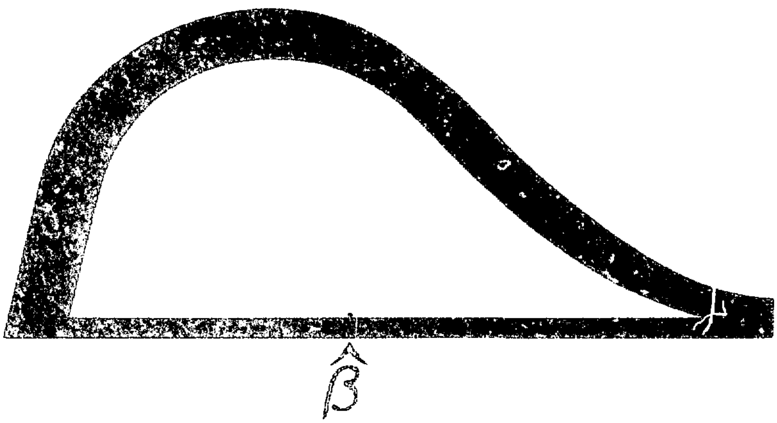
DR-9

CONF-771042

Proceedings of the

1977 DOE STATISTICAL SYMPOSIUM

MASTER



October 26-28, 1977
Pacific Northwest Laboratories
Richland, Washington

Sponsored by the
Department of Energy

Prepared by
Oak Ridge National Laboratory
Oak Ridge, Tennessee
Operated by Union Carbide Corporation

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

CONF-771042
Distribution Category UC-32

Proceedings of the

1977 DOE STATISTICAL SYMPOSIUM

October 26–28, 1977
Pacific Northwest Laboratories
Richland, Washington

Compiled and Edited by
Donald A. Gardiner
Tykey Truett

Sponsored by the
Department of Energy

Date Published: March 1978

Prepared by
• Mathematics and Statistics Research Department
Computer Sciences Division
• Oak Ridge National Laboratory
Oak Ridge, Tennessee
Operated by
Union Carbide Corporation










CONTENTS

Preface

The publication of these proceedings completes the plans conceived and originated in early February of 1975. At that time, statisticians from Los Alamos, Oak Ridge, and Pacific Northwest met to discuss the feasibility of identifying what was called the "ERDA Statistical Community" and to discuss means of bringing the members of that community together to meet one another and to share their experiences in helping to solve the nation's energy problems.

With the encouragement of ERDA's Division of Physical Research, a three-year plan was undertaken in which meetings would be held at national laboratory sites and the responsibilities shared by the three instigating laboratories. The program committee would be chaired by the host laboratory. (To this committee were added, subsequently, representatives from Sandia Laboratories and Princeton University.) The proceedings would be published by one of the nonhost laboratories.

In accordance with the plan, the First ERDA Statistical Symposium was held at Los Alamos, New Mexico, in 1975. The proceedings were published by Pacific Northwest (BNWL-1986). The Second ERDA Statistical Symposium was held in 1976 at Oak Ridge, Tennessee, with proceedings published by Los Alamos Scientific Laboratory (LA-6758-C). The third symposium was held at Richland, Washington, on October 26 28, 1977. That symposium, of which these are the proceedings, was named the 1977 DOE Statistical Symposium because ERDA had been dissolved and reorganized into the Department of Energy just four weeks previously.

The Program Committee, chaired by Wesley L. Nicholson, consisted of Donald A. Gardiner, Ronald K. Lohrding, George P. Steck, and Thomas W. Woteki. It is a pleasure to express appreciation for the excellent efforts of the Local Arrangement Committee, Ethel S. Gilbert, Richard L. Hooper, James W. Johnston, Anthony R. Olsen, and Donald Stevens under the able leadership of Pamela G. Doctor, and to acknowledge the assistance of Charles K. Bayne, Thomas L. Hebble, William E. Lever, and Deborah E. Shepherd in preparing these proceedings.

Thus the original plans have been successfully implemented and the cycle is complete. The ERDA Statistical Community, now the DOE Statistical Community, has been rather completely identified, we think. The first mailing list consisted of just a few names recalled from the tops of our heads; the list now contains more than 300 names. The interaction among the statisticians at the DOE laboratories, the academic community, and industries engaged in energy-related enterprises has increased to a surprising and gratifying degree. The organizers and their sponsors should be well pleased.

At the conclusion of the symposium in Richland, the participants met for a critique and to discuss plans, if any, for the future. They were of the mind that the symposia should continue on an annual basis and that the organizers should be recruited from a wider base. Sol Rubinstein of Rockwell International volunteered to lead the Program Committee, and representatives from the national laboratories, universities, and industry volunteered to serve. Richard Prairie of Sandia Laboratories offered the auspices of Sandia Laboratories as host for a symposium in 1978, and Nicholson and Gardiner offered to work out a plan for the publication of proceedings.

The contributions of all those who supported and participated in the first three statistical symposia are greatly appreciated.

Donald A. Gardiner

Welcome

Tommy Ambrose

Pacific Northwest Laboratories
Richland, Washington

INTRODUCTION OF T. AMBROSE—*Wes Nicholson*

The formal welcome to the 1977 Department of Energy Statistical Symposium will be given by Dr. Tommy Ambrose, the Director of the Pacific Northwest Division of Battelle Memorial Institute. This division of Battelle consists of the Pacific Northwest Laboratory in Richland, which Battelle operates for the Department of Energy, the Seattle Research Center and the Human Affairs Research Center, both located adjacent to the University of Washington campus, and a marine sciences laboratory at Sequim on the northwest coast of Washington State. Dr. Ambrose represents DOE and the various contractors as he formally welcomes you. His remarks will include a description of the rather unique situation here resulting from the fact that Battelle operates a dual laboratory in Richland, being a DOE contractor and a broad-spectrum, nonprofit research organization. It is a pleasure to introduce our laboratory director, Tommy Ambrose.

It is my pleasure to welcome you to Richland and the Tri-Cities on behalf of the Department of Energy and its contractors. We are particularly pleased to host the 1977 DOE Statistical Symposium for those who are interested in the nation's energy problems. I understand our audience is made up of people from other DOE laboratories, the university community, and industry.

For the individuals who are first-time visitors to the Tri-City area, you now know that the State of Washington is not entirely covered by green trees and lush vegetation. In fact, we are nearly in the center of the remaining three-quarters of the State. Also, you have no doubt learned that Richland is next to impossible to reach, and one really must work hard to get here. We thank each of you for the extra effort.

The Richland operation is made up of a group of contractors who operate the entire complex for DOE. This arrangement is different from most DOE sites and perhaps warrants an explanation as to how it came into existence.

In the dim, dark past before ERDA and even before AEC, the U.S. Government contracted with the Du Pont Company to construct and operate a facility known as the Hanford Works whose purpose was to produce plutonium for weapons. In 1946, Du Pont turned over the operation of the plant to the General Electric Company, and the AEC was created. The operation grew to the point in the mid-1950s that eight reactors were in operation along with a reprocessing plant and a work force of about 8000.

In 1963, Hanford's mission to make plutonium for nuclear weapons was almost completed. The nation's stockpile of plutonium was sufficient. Richland, the one-payroll town created just 20 years earlier, was headed for a major reduction in employment. At that point the creative ingenuity of the AEC, its contractor (General Electric), and community leaders went to work, and a far-sighted program was developed in 1964 based on the concepts of "segmentation"—dividing the single operating contract for the total Hanford Project into parts to be performed by other industrial organizations, and "diversification" the new Hanford contractors strengthening the economic base of the community by new programs and investments.

Today the operating contractors for DOE include Battelle Memorial Institute, Boeing Computer Services, Hanford Engineering Development Laboratory (Westinghouse), J. A. Jones, Rockwell Hanford Company, United Nuclear Industries, and Vitro. All eight of the reactors have been shut down as well as the reprocessing plant, and the labor force now exceeds 8000.

Two additional companies that have located in the area because of our strong nuclear technical base are Exxon Nuclear Company, Inc., and Washington Public Power Supply System. Exxon has established a light-water fuel fabrication plant and has undertaken research and pilot plant work in uranium enrichment technology. The Washington Public Power Supply System is a group of public and private utilities in the process of building several nuclear power plant units in the area.

The symposium is sponsored by the Department of Energy as part of its ongoing effort to solve the nation's energy problems. Participants from DOE and the various contractors are your symposium hosts. Speaking for DOE and the contractor group, we are pleased with the scope of the statistical program and are delighted with the caliber of people who are attending and expressing their keen interest in the nation's energy problems. Thank you for your participation, and welcome to Richland.

Invited Papers

Harmonic Regression*†

F. J. Anscombe

Yale University
New Haven, Connecticut

ABSTRACT

Ordinary linear regression, by the method of least squares, is used to determine a linear relation between given independent observations of two or more variables. An analogous problem for time series is to determine a linear relation between two or more given stationary series. The linear relation may take the form that one series is a linear filtering of the other series, plus a stationary error process. The coefficients of the filter can be determined directly by multiple regression in the time domain, but there are various difficulties. An easier procedure, leading to more intelligible results, is to estimate the Fourier transform of the coefficients of the filter, which can be expressed in terms of a gain function and a phase-shift function, by a simpler regression calculation in the frequency domain.

The procedures are illustrated by a study of the interrelationship of an annual series of output of U.S. copper mines from 1860 to 1975 and two annual economic series relating to the same years, namely a series of copper prices at New York and a series of the total dollar value of general imports of merchandise into the United States.

The calculations of ordinary regression analysis linear regression by the method of least squares have been done correctly for a century and a half. However, there have been changes in the computational methods used. There is plenty to discuss about regression—for example, is it appropriate for the data, and what do the results mean? No doubt the calculations are sometimes of little value, but sometimes they are appropriate and lead to new understanding.

Regression analysis of time series has a much shorter history. Although there is a good deal of literature about it, the literature often has the air of arm-chair meditation by a nonparticipant. My concern has been to implement principles that are in the literature, and devise a working procedure. Various practical difficulties have been encountered that do not seem to be discussed in the literature.

Does anyone need to do regression analysis of time series? Conflicting opinions are heard. Great amounts of time-series material are being collected and stored relating to the environment (weather, pollution), the observations being made daily or even more frequently. Many economic series are developed for monthly, weekly, or daily activities. I have worked with annual series, which are probably the least satisfactory material for this kind of study.

Some broad generalities are presented below, and an example is given. The details are vital, but as they have been fully described elsewhere they are not given here.[1]

---

*Invited address.

†Prepared in connection with research supported by the Army, Navy, Air Force, and NASA under a contract administered by the Office of Naval Research.

1. The detailed study on which this paper is based is found in the following: F. J. Anscombe, "Time Series: Yale Enrollment," Chap. 10 in *Statistical Computing with APL* (in preparation).

FORMULATION

We consider regression of one "dependent" variable on just one "independent" or predictor variable. (Methods extend, of course, but not without some difficulties, to several predictor variables.) All means will be supposed zero. Then ordinary linear regression can be formulated. We are given observations on pairs of variables, (x_i, y_i) for $i = 1, 2, \ldots, m$. We suppose that for all i,

$$y_i = \beta x_i + \epsilon_i \,, \tag{1}$$

where the errors $\{\epsilon_i\}$ are considered to be (in some sense) *independent* of each other and of the predictor variable $\{x_i\}$. The method of least squares can be equated to the method of maximum likelihood when we suppose that the $\{\epsilon_i\}$ are independent random variables identically distributed $N(0, \sigma^2)$.

How should regression of time series be formulated? We are given series $\{x_t\}$, $\{y_t\}$, where $t = 1, 2, \ldots, n$. We shall not suppose these series, nor the error series $\{\epsilon_t\}$ when we introduce it, to consist of independent elements. We shall instead suppose the series to be *stationary*, that is, realizations of some kind of stationary stochastic process. (In practice the appearance of stationarity with zero mean is encouraged by subtracting a linear or other trend, usually after taking logarithms.) To correspond to Eq. (1), one might suggest

$$y_t = \beta x_t + \epsilon_t \,.$$

But if the series are related, the relation may be not simultaneous. One might have

$$y_t = \beta x_{t-j} + \epsilon_t \,,$$

for some integer lag j. But then one might as well postulate

$$y_t = \Sigma_j\, \beta_j x_{t-j} + \epsilon_t \,, \tag{2}$$

where j runs over some suitable set of integer values. Equation (2) seems to be the appropriate formulation for stationary processes, to correspond to Eq. (1) for independent processes. The first member of the right side of Eq. (2) represents a linear filtering of $\{x_t\}$.

There are two main approaches to trying to estimate the parameters $\{\beta_j\}$ of the filter in Eq. (2).

Time-domain Methods

One can try direct multiple regression of $\{y_t\}$ on $\{x_t\}$ and on lagged versions of it, $\{x_{t-j}\}$ for various j. There is a difficulty about deciding how many lags should be considered. If $\{x_t\}$ is strongly autocorrelated, conditioning will be poor. An accurate representation of the relation between two stationary stochastic processes could easily involve a large number of nonzero coefficients $\{\beta_j\}$.

If our reason for trying to fit a relation like Eq. (2) is to be able to forecast y_t from past values of $\{x_t\}$, possibly a very crude estimate of the $\{\beta_j\}$ will be good enough. The precision of a forecast is limited by the variance of the error term. The greater precision that would be attained if the $\{\beta_j\}$ were known exactly may be only negligibly greater.[2] Box and Jenkins[3] have presented a set of practical procedures for estimating the structures of time series well enough for forecasting. If our purpose is not forecasting, but understanding as well as we can the relation between the series, the Box-Jenkins methods may be less satisfactory.

It will be argued that some of these difficulties are mitigated or avoided by frequency-domain methods. However, we must usually be alert to temporal instability or change in a relation like Eq. (2), and that will be detected by time-domain methods.

Frequency-domain Methods

The idea is to Fourier-transform Eq. (2) and to estimate the transform of $\{\beta_j\}$. It will be suggested that (i) this procedure is easier to carry out than multiple regression in the time domain and that (ii) the results are easier to understand. Claim (i) is derived from the fact that the first member on the right side of Eq. (2), the filtering of $\{x_t\}$, is a convolution of $\{\beta_j\}$ and $\{x_t\}$ and transforms to the product of the separate transforms of $\{\beta_j\}$ and $\{x_t\}$. Thus Eq. (2) becomes

$$\mathrm{FT}\{y_t\} = (\mathrm{FT}\{\beta_j\})(\mathrm{FT}\{x_t\}) + \mathrm{FT}\{\epsilon_t\} \,.$$

---

2. W. S. Cleveland, *Time Series Projection: Theory and Practice*, Ph.D. dissertation, Yale University, New Haven, Conn., 1967.

3. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970, 1976.

These Fourier transforms are complex-valued functions of a real variable λ representing frequency. Consider a narrow frequency band (interval for λ). Suppose that in this interval the transforms of $\{\beta_t\}$ were (near enough) constant. Then in this interval the relation between FT$\{y_t\}$ and FT$\{x_t\}$ would be exactly like Eq. (1) between $\{y_t\}$ and $\{x_t\}$, with the exception that the variables and the regression coefficient are complex-valued. In real terms, FT$\{\beta_t\}$ is conveniently expressed as an amplitude, the gain function $G(\lambda)$, and an angle, the phase-shift function $\phi(\lambda)$. Thus if $G(\lambda)$ and $\phi(\lambda)$ could be regarded as constant over the frequency band, they could be estimated from the transforms of $\{y_t\}$ and $\{x_t\}$ by a slight modification of the usual procedure for the linear regression relation Eq. (1)—expressed in real terms it looks a bit different, but the procedure is really ordinary linear least squares with two real coefficients to be estimated. The least-squares procedure is particularly appropriate if the error process $\{\epsilon_t\}$ is a stationary Gaussian process whose spectral density is nearly constant over the band.

However, it has been generally recognized (refs. 4 and 5, and others) that treating $\phi(\lambda)$ as constant is not satisfactory when its derivative is much different from 0 and that it is better to approximate the behavior of the transform of $\{\beta_t\}$ in the narrow frequency band by three real parameters, the average values of $G(\lambda)$, $\phi(\lambda)$, and $\phi'(\lambda)$ in the band—that is, treat $G(\lambda)$ as constant and $\phi(\lambda)$ as linear in λ. Now the regression procedure is further modified, becoming in fact nonlinear and requiring an iterative solution, but still computationally rather easy.

Thus the complete procedure involves examining the frequency range of λ in bands, using a moving "window," and in each band doing a small computation to determine three real parameters, representing average values of $G(\lambda)$, $\phi(\lambda)$, and $\phi'(\lambda)$. Upon putting the solutions together we see the whole behavior of $G(\lambda)$ and $\phi(\lambda)$. With $G(\lambda)$ and $\phi(\lambda)$ estimated, $\{\beta_t\}$ could be inferred by making the inverse Fourier transform.

Intelligibility

Claim (ii) is that $G(\lambda)$ and $\phi(\lambda)$ are what we need, in order to understand the relation between $\{y_t\}$ and $\{x_t\}$, rather than the $\{\beta_j\}$. If the latter were given, we should have to Fourier-transform them to see qualitatively the effect of the filter. Compare this with the usual commercial description of performance of an amplifier in a sound-reproduction system.

EXAMPLE

As an example of methods, we try interrelating an annual series of total copper mine output for the United States and two economic annual series, one giving the New York price of copper, the other the total dollar value of imports of merchandise into the United States. The copper price series is thought to reflect the world supply and demand for copper. Changes in price might be expected to lead to similar changes in production, possibly a little later. The imports series is taken as an indicator of the U.S. economy. The copper production series is N235, and the price series is N241, in *Historical Statistics of the United States*;[6] the production figures run from 1845 to 1970, the prices from 1850 to 1970. The figures have been taken exactly as published, except that to smooth a change in price definition in 1968 the average of two definitions has been used for 1967. The price figures for 1850–1859 are of uncertain meaning, and the production figures for before 1860 show a more rapid proportional rate of growth than for later time periods. For present pusposes it has seemed wise to ignore the pre-1860 data. Continuation of the series from 1970 to 1975 has been obtained from the *Statistical Abstract of the United States*.[7] The two series are reproduced in Fig. 1, except that the last two digits of the production entries have been dropped for ease of reading. The imports series has been given in ref. 1 and is not reproduced here; only the portion from 1860 to 1975 is used.

Figure 2 shows a plot against the date of the logarithm of the production series, with the linear regression on date subtracted. Figure 3 is a similar plot for the price series. A plot for the imports series has been given in ref. 1.

The three given series each have 116 entries (for 1860–1975). To prepare them for Fourier analysis they have been prewhitened by these three steps: (i) take logarithms, (ii) subtract the linear regression on

4. H. Akaike and Y. Yamanouchi, "On the Statistical Estimation of Frequency Response Function," *Ann. Inst. Stat. Math.* 14: 23–56 (1962).

5. W. S. Cleveland and E. Parzen, "The Estimation of Coherence, Frequency Response, and Envelope Delay," *Technometrics* 17: 167–72 (1975).

6. U.S. Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition*, U.S. Government Printing Office, Washington, D.C., 1975.

7. U.S. Bureau of the Census, *Statistical Abstract of the United States*, U.S. Government Printing Office, Washington, D.C., 1975, 1976.

+	0	1	2	3	4	5	6	7	8	9	
U.S. COPPER PRODUCTION (MINE OUTPUT, HUNDREDS OF SHORT TONS)											
1860	81	84	106	95	90	95	100	112	130	140	*A*
1870	141	146	140	174	196	202	213	235	241	258	*B*
1880	302	358	453	578	725	829	789	907	1132	1134	*C*
1890	1299	1421	1725	1647	1771	1303	2300	2470	2633	2843	*D*
1900	3031	3010	3298	3490	4063	4444	4585	4236	4784	5633	*E*
1910	5441	5574	6245	6178	5742	7440	10029	9477	9550	6062	*F*
1920	6123	2331	4823	7389	8031	8391	8626	8250	9049	9976	*G*
1930	7051	5289	2381	1906	2374	3865	6145	8420	5578	7283	*H*
1940	8781	9581	10801	10908	9725	7729	6087	8476	8348	7528	*I*
1950	9093	9283	9254	9264	8355	9986	11042	10869	9793	8248	*J*
1960	10802	11652	12284	12132	12468	13517	14292	9541	12046	15446	*K*
1970	17197	15220	16650	17180	15970	14110					*L*
PRICE OF REFINED COPPER AT NEW YORK (CENTS PER POUND)											
1860	22.88	22.25	21.88	33.88	47.00	39.25	34.25	25.38	23.00	24.25	*A*
1870	21.19	24.12	35.56	28.00	22.00	22.69	21.00	19.00	16.56	18.62	*B*
1880	21.50	18.25	18.50	15.88	13.75	11.10	11.00	11.25	16.80	13.75	*C*
1890	15.75	12.88	11.50	10.65	9.43	10.70	10.92	11.30	12.01	17.75	*D*
1900	16.54	16.40	11.96	13.62	13.11	15.98	19.77	20.86	13.39	13.11	*E*
1910	12.88	12.55	16.48	15.52	13.31	17.47	28.46	29.19	29.19	18.90	*F*
1920	17.50	12.65	13.56	14.61	13.16	14.16	13.95	13.05	14.68	18.23	*G*
1930	13.11	8.24	5.67	7.15	8.53	8.76	9.58	13.27	10.10	11.07	*H*
1940	11.40	11.87	11.87	11.87	11.87	11.87	13.92	21.15	22.20	19.36	*I*
1950	21.46	24.37	24.37	28.92	29.82	37.39	41.88	29.99	26.13	30.82	*J*
1960	32.16	30.14	30.82	30.82	32.17	35.19	35.82	38.01	41.17	47.43	*K*
1970	58.07	52.00	51.20	59.50	77.30	64.20					*L*

Fig. 1. The data.

date, (iii) filter by the two-point filter with weights (−0.9, 1). The last operation reduces the length of each series to 115. Then the series have been circularized (tapered) by linearly splicing the first seven and the last seven entries, so that the length of each series becomes 108. The Fourier transform is made at frequencies (0, 1, 2, ..., 54)/108 cycles per year; the transform is expressed as a set of (real) coefficients of cosine and sine terms, or alternatively as a set of squared amplitudes and phase angles. The frequencies are referred to as harmonics, numbered 0 through 54.

The first step to perceiving an interrelation between any pair of series is to plot the difference of phase angles at each harmonic against the harmonic number. Figure 4 shows this for the production and price series, and Fig. 5 for the production and imports series. At each harmonic, the product of the amplitudes is classified by size into one of six categories and represented by one of the plotting symbols:

. ∘ ○ ⊖ □ ⊟

Each phase difference is plotted with this symbol twice over in the interval from 0 to 8 right angles. In looking for trends, the viewer's eye should be guided by the heavier symbols.

Figure 5 shows a fairly strong relation between production and imports, especially at the higher frequencies the phase differences are mostly rather close to 4 (or 0 or 8) right angles and show no trend with frequency. A simultaneous positive correlation between these two series is indicated. Figure 4 shows a less clear relation between production and price. At lower frequencies there is some suggestion of trend in the phase differences, implying that production follows price, possibly by two years, possibly by four.

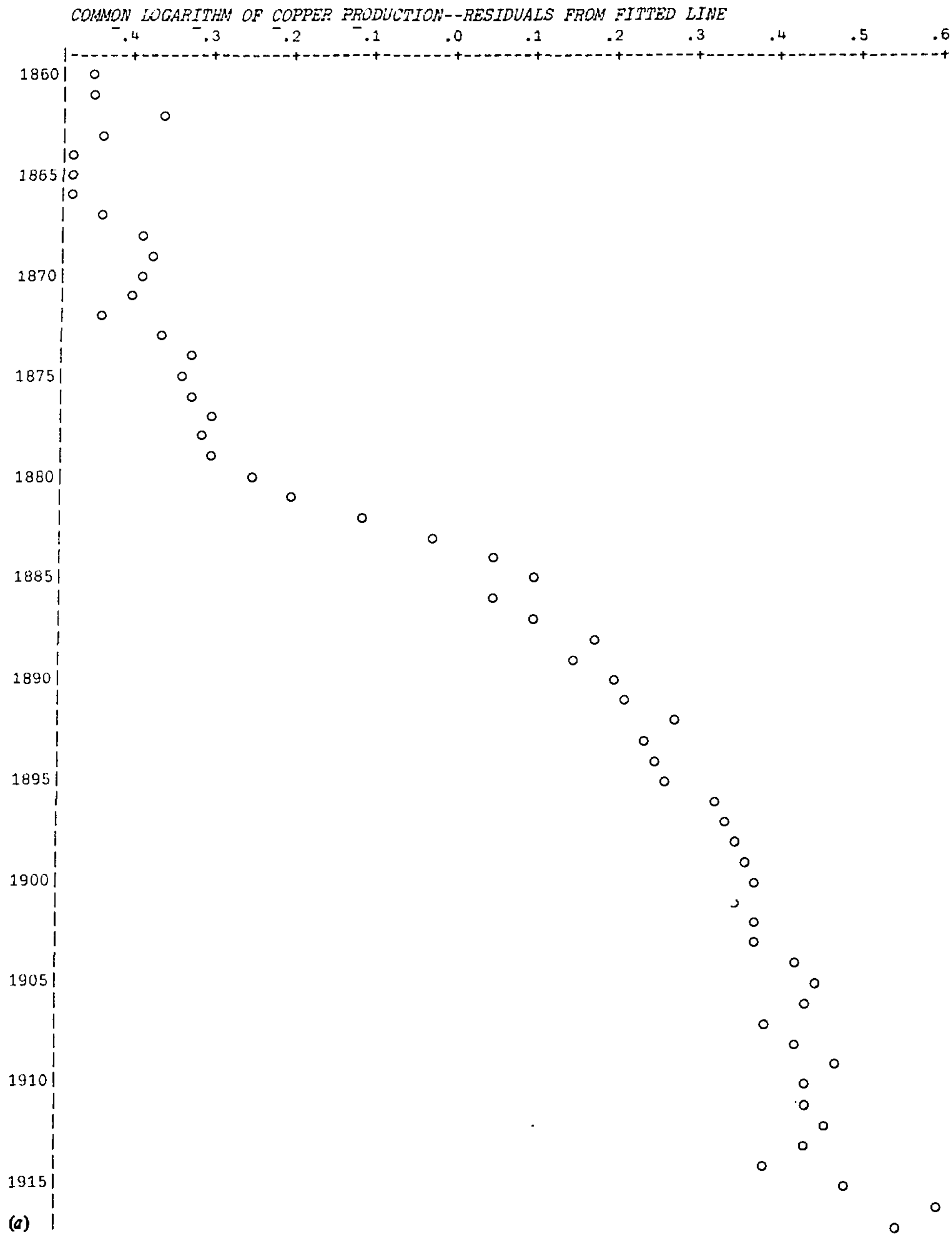


Fig. 2. Plot of the copper production series: (a) 1860–1915; (b) 1920–1975.

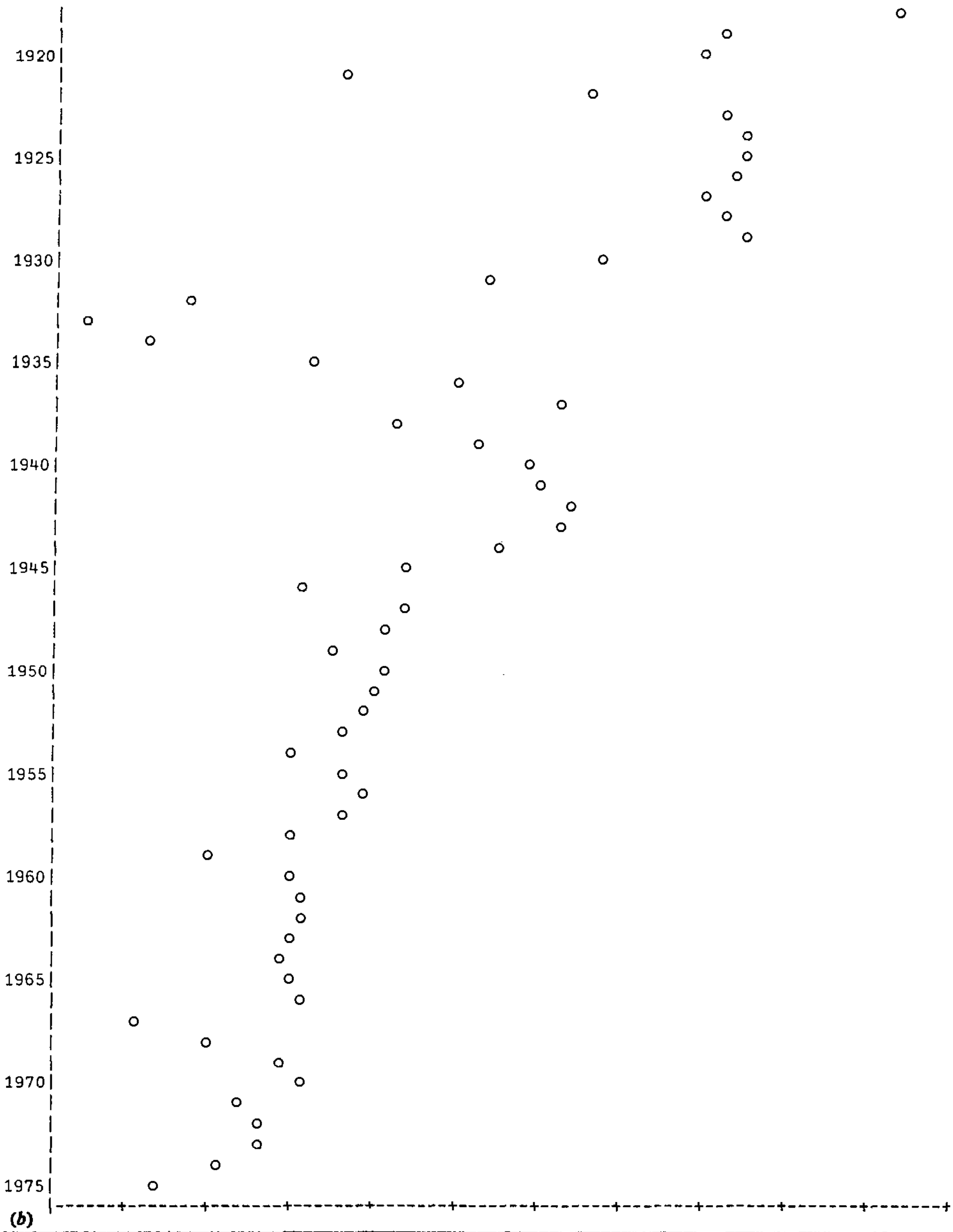


Fig. 2 (continued).

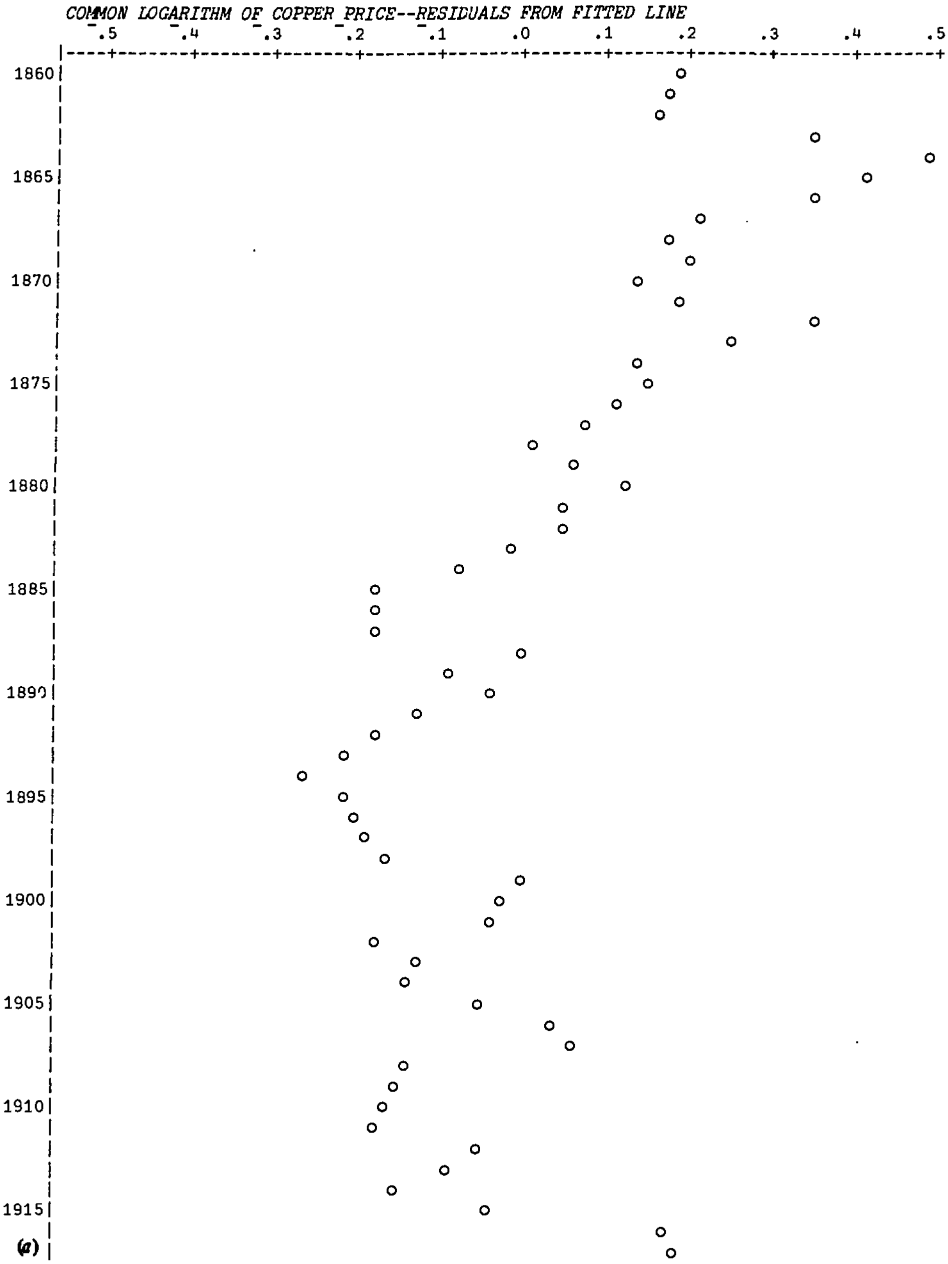


Fig. 3. Plot of the copper price series: (*a*) 1860–1915; (*b*) 1920–1975.

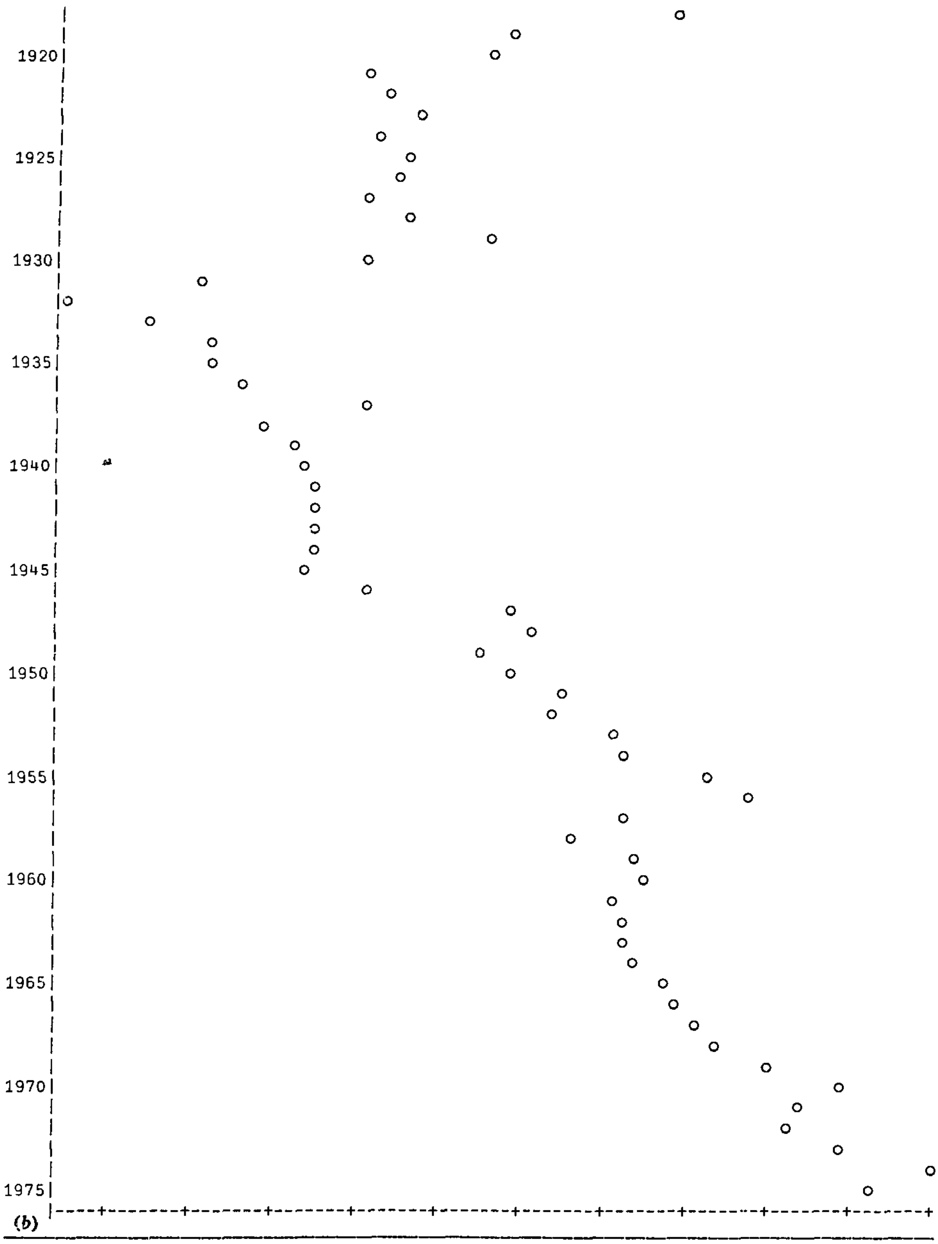


Fig. 3 (continued).

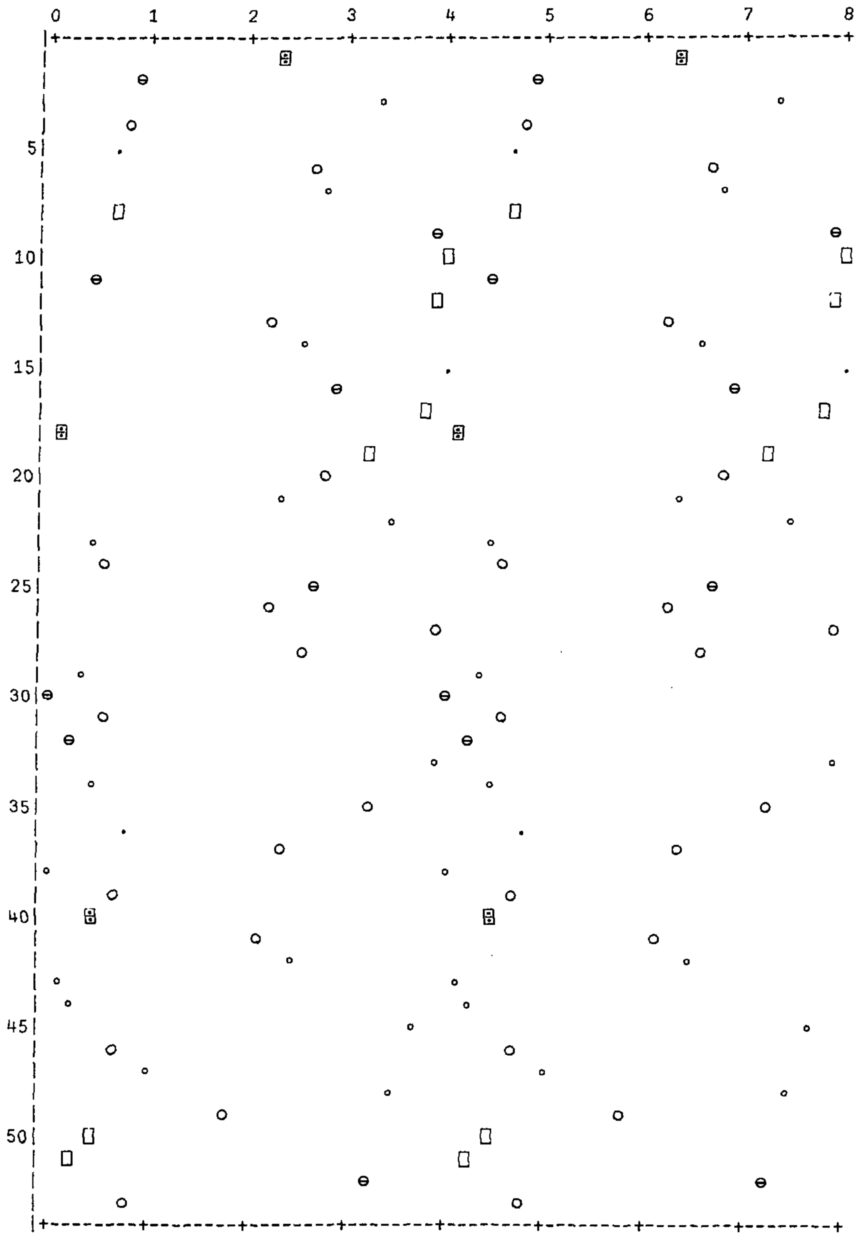


Fig. 4. Phase-difference plot for copper production and copper price.

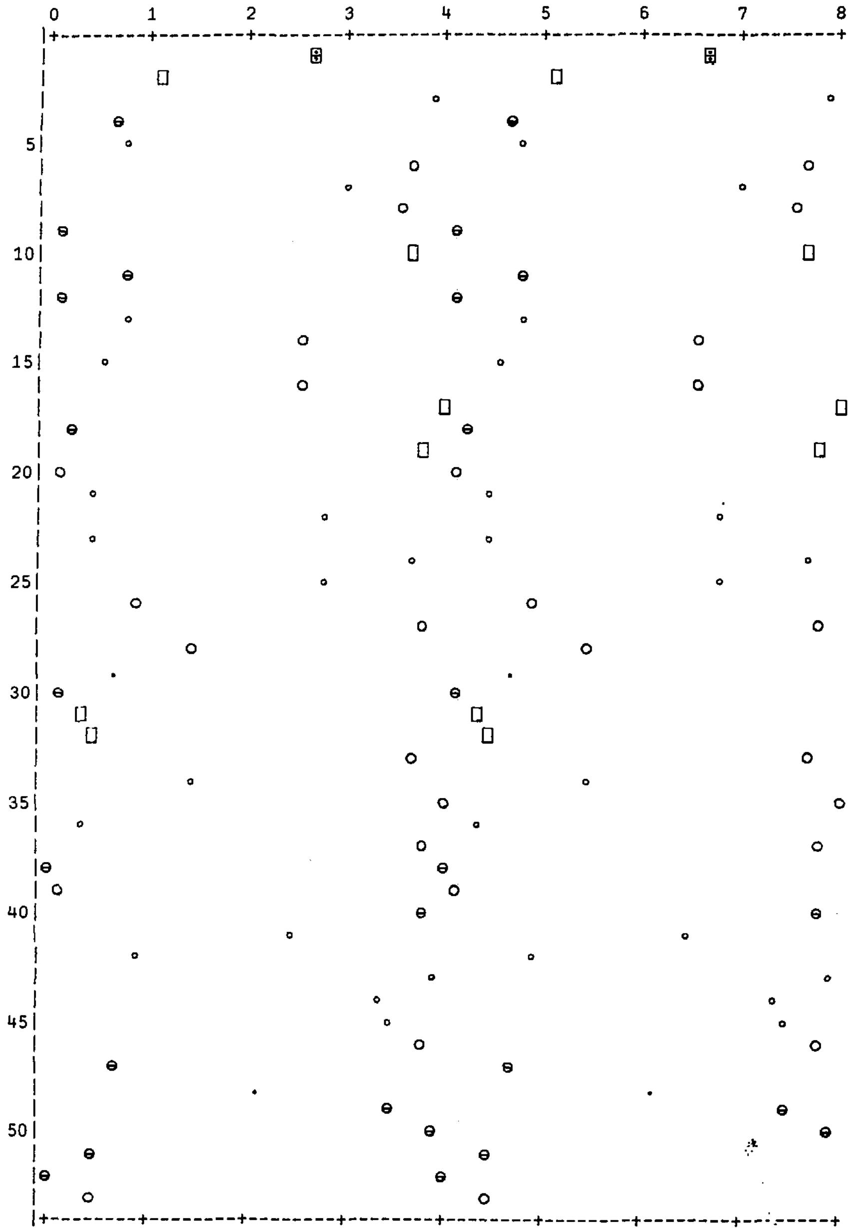


Fig. 5. Phase-difference plot for copper production and general imports.

	Smoothed spectral estimates			Regression coefficients of copper prodn, with R*2, (i) on copper price				(ii) on imports			
11	1274	853	583	.635	6.10	15.1	.270	.625	6.26	1.1	.179
12	1272	860	566	.666	.01	12.2	.299	.687	6.26	2.0	.210
13	1252	861	542	.695	.15	9.8	.332	.756	6.26	2.4	.248
14	1259	868	519	.692	.25	9.6	.330	.787	.02	1.0	.255
15	1258	879	503	.686	.36	10.6	.328	.796	.03	.3	.253
16	1240	883	486	.674	.46	10.7	.323	.804	.03	¯.4	.253
17	1213	880	472	.665	.57	11.3	.321	.795	.02	¯.6	.246
18	1166	867	455	.652	.69	11.9	.316	.779	6.28	¯.5	.237
19	1109	852	439	.621	.79	11.5	.296	.768	6.26	¯.8	.233
20	1058	824	427	.580	.86	10.5	.262	.766	6.21	¯1.8	.237
21	1014	805	427	.525	.84	7.5	.219	.762	6.18	¯3.3	.244
22	972	777	427	.487	.70	1.8	.190	.769	6.15	¯3.8	.259
23	936	745	425	.473	.60	¯1.8	.178	.790	6.14	¯4.8	.284
24	893	719	427	.457	.56	¯3.0	.168	.801	6.12	¯4.9	.307
25	842	706	439	.441	.51	¯3.3	.163	.798	6.08	¯4.6	.332
26	782	694	446	.409	.48	¯3.6	.149	.793	6.04	¯3.9	.359
27	716	672	471	.378	.46	¯4.2	.134	.761	6.02	¯3.1	.381
28	646	641	491	.349	.42	¯5.7	.121	.736	5.99	¯2.0	.412
29	617	620	507	.355	.42	¯10.5	.127	.717	5.97	¯.3	.422
30	614	597	517	.391	.44	¯14.9	.149	.715	5.94	2.0	.431
31	620	589	525	.400	.36	¯16.4	.152	.729	5.90	4.9	.450
32	634	589	531	.413	.21	¯15.1	.158	.743	5.93	6.0	.463
33	641	582	537	.430	.07	¯13.7	.168	.753	5.98	6.7	.475
34	644	567	536	.450	6.22	¯12.3	.178	.766	6.04	7.2	.488
35	642	545	533	.477	6.11	¯10.5	.193	.774	6.10	7.4	.497
36	637	515	526	.513	6.00	¯8.7	.213	.775	6.15	6.7	.496
37	627	486	512	.550	5.90	¯5.7	.234	.775	6.19	5.9	.491
38	625	458	492	.578	5.83	¯5.5	.245	.781	6.24	5.5	.480
39	630	438	471	.600	5.79	¯4.8	.250	.795	6.28	5.1	.473
40	640	432	450	.633	5.77	¯3.5	.270	.810	.00	3.8	.462
41	659	425	428	.641	5.78	¯2.4	.266	.834	.01	2.6	.452
42	669	421	402	.645	5.77	¯1.6	.262	.861	.02	1.4	.446
43	675	417	375	.652	5.76	¯1.2	.262	.895	.04	.0	.445

Fig. 6. Tabulation of regression calculations in frequency bands.

At higher frequencies the phase differences seem very scattered. Not reproduced is a phase-difference plot for the price and imports series, suggesting quite a strong simultaneous correlation at the lower frequencies, and not much at higher frequencies.

Now the regression calculation in frequency bands, to estimate $G(\lambda)$, $\phi(\lambda)$, and $\phi'(\lambda)$, can be performed. The window chosen is 23 harmonics wide, and sine weights have been used. The results are tabulated in Fig. 6. The first column lists the harmonic number of the central frequency in the band; we have stepped the central harmonic number from the lowest possible value, 11, by unit steps to the greatest possible value, 43. (Had there been many more harmonics and a greater bandwidth, greater steps would have been convenient.) The next three columns list estimates of spectral density for, respectively, copper production, copper price, and imports (prewhitened as explained above), obtained from the raw line spectra by the 23-point, sine-weighted moving average. The next four columns refer to regression of copper production on copper price. They list average values in the band of $G(\lambda)$, $\phi(\lambda)$, and $\phi'(\lambda)$, and (in the fourth of these columns) multiple R^2 (the coherency). The behavior of $\phi(\lambda)$ and of R^2 gives a numerical measure of the trend seen in Fig. 4. The last four columns of Fig. 6 give similar information for regression of copper production on imports and relate to Fig. 5. (Simultaneous regression of copper production on both copper price and imports is not considered at this point.)

To test a null hypothesis of no association between series, 5%, 1%, and 0.1% values for R^2 for any given frequency band are estimated (by a crude argument) at 0.19, 0.27, and 0.36, respectively; these values probably err in being a little too low. The tabulated values are very highly correlated, as one reads down the column. So for regression of production on price, it seems reasonable to claim a substantial correlation at low frequencies, in the bands centered between the 11th and 19th harmonics. For regression of production on imports, the correlation is substantial in bands centered between the 23rd and 43rd harmonics—R^2 is close to 0.5 in many of these bands.

Of our two predictor variables, copper price and general imports, the latter has on the whole the greater correlation with copper production. But the two predictor series have some correlation with each other. How useful is the price series as a predictor in conjunction with the imports series? Residual Fourier transforms of the production series and of the price series, after regression on the imports series, can be obtained, and a phase-difference plot can be made, analogous to Fig. 4 for the original Fourier transforms. This plot is shown in Fig. 7. The phase trend seems rather similar to that in Fig. 4 at lower frequencies and weaker at higher frequencies.

Figure 8 shows a calculation like that in Fig. 6, but relating to simultaneous regression of production on both price and imports, instead of to separate regressions. The R^2 in the final column is always greater than either value of R^2 (for the same frequency band) given in Fig. 6. The most striking increase over the R^2 for regression on imports only occurs for bands centered between the 25th and 28th harmonics—for example, 0.479 instead of 0.332 at the 25th harmonic, 0.511 instead of 0.359 at the 26th harmonic. The same sort of crude argument as before indicates that these four increases (but none of the others) can be regarded as significant at the 5% level. The increases, on the whole, are larger at lower frequencies than at higher frequencies.

The two phase-shift functions estimated in Fig. 8 can be fairly well approximated at most frequencies by saying that production is correlated positively with imports of the same year and negatively with prices of four years before.

Figures 9 and 10 are time-domain plots intended to show whether the relations between the series perceived in the harmonic analysis pervade the whole series or are special to particular epochs. For both plots, the original series have been transformed to logarithms, and a linear trend has been subtracted. Then for Fig. 9, low frequencies have been suppressed by taking the second difference of the series, and the resulting production values are plotted against the imports values. The correlation coefficient is 0.60. The decade of each plotted point is shown by the letters appearing on the right side of Fig. 1; a star means that two or more points have coincided. For Fig. 10, the spectra have been roughly whitened by taking the first difference of each series, and then high frequencies have been suppressed by three simple two-point averagings. The first four values of the resulting production series and imports series have been dropped, as well as the last four values of the resulting price series, then the production values are plotted against the linear combination of the imports values (for the same year) minus 0.8 times the price values (for four years earlier). The correlation coefficient is 0.53. The decade of the production values is shown as before.

The pronounced correlation in both Figs. 9 and 10 is due to a few extreme points labeled G or H, representing the two decades from 1920 to 1939. If all points for these decades were omitted, the correlation would become 0.04 for Fig. 9 and −0.08 for Fig. 10. That is, the correlation would disappear.

DISCUSSION

Harmonic regression is a systematic way of looking for association between series in all parts of the frequency range. It is unlikely to reveal anything that cannot be found by careful visual comparison of plots such as those in Figs. 2 and 3, at least when only two or three series are under consideration. (A similar remark can be made about ordinary regression.)

We have found clear evidence of association between copper production and general imports, at higher frequencies, and some suggestion of predictive value for copper price also, at middle-to-low frequencies. What associations there are seem to be inherent in the economically turbulent years of the twenties and thirties. We do not see similar associations in the other decades. Possibly relations between these series are changing; possibly the phenomena are highly nonlinear.

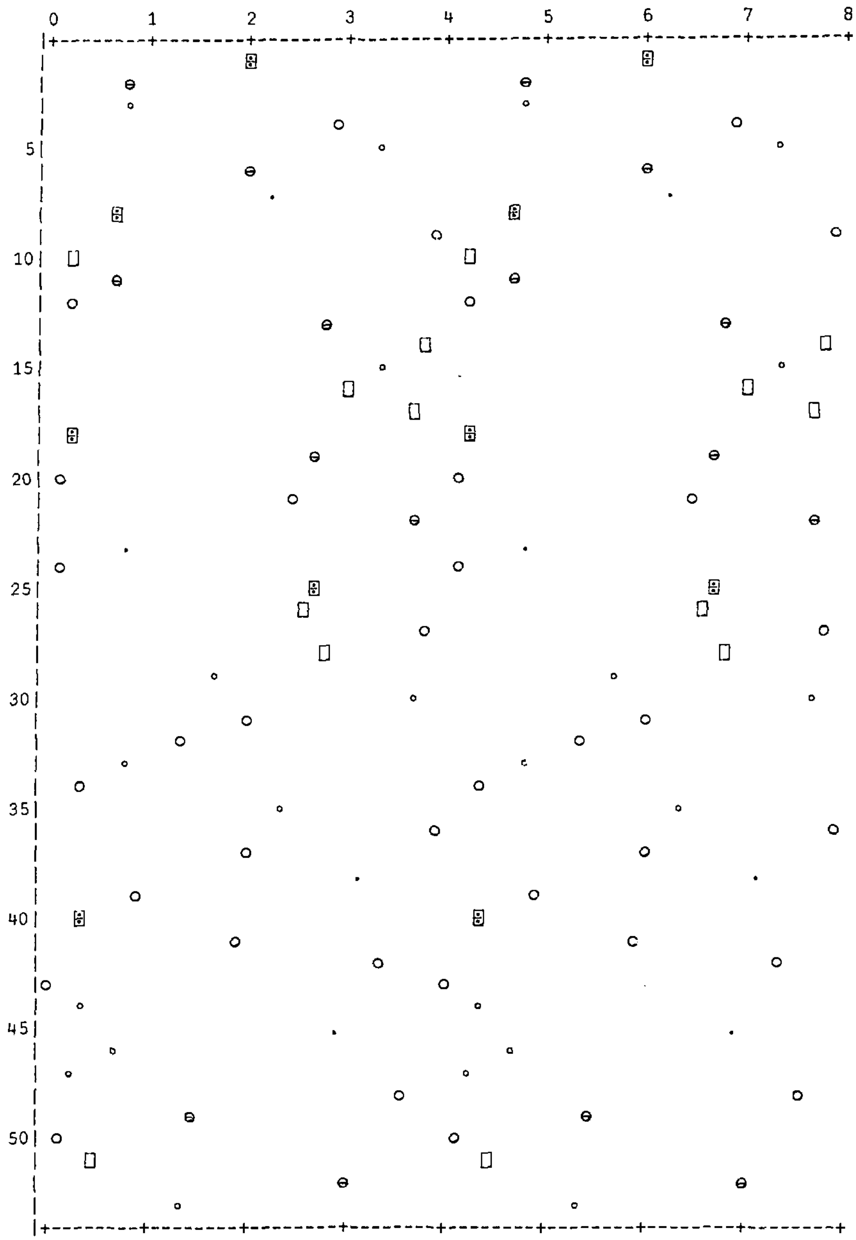


Fig. 7. Phase-difference plot for copper production residuals and copper price residuals, after regression on general imports.

	Smoothed spectral estimates			Regression coefficients of copper prodn on copper price			and on imports			Mult. R*2
11	1274	853	583	.542	5.92	18.2	.439	.34	¯6.2	.345
12	1272	860	566	.542	6.09	15.9	.424	.33	¯4.6	.359
13	1252	861	542	.537	6.26	12.5	.411	.23	¯1.3	.379
14	1259	868	519	.520	.12	12.2	.431	.24	¯2.0	.380
15	1258	879	503	.515	.26	13.2	.439	.18	¯3.1	.380
16	1240	883	486	.504	.39	13.4	.461	.15	¯3.[illegible]	.381
17	1213	880	472	.506	.52	13.9	.466	.11	¯4.4	.382
18	1166	867	455	.505	.66	14.4	.469	.05	¯5.0	.362
19	1109	852	439	.482	.78	14.7	.504	6.28	¯5.3	.377
20	1058	824	427	.450	.90	15.6	.572	6.26	¯5.5	.372
21	1014	805	427	.429	1.01	18.0	.660	.00	¯6.3	.378
22	972	777	427	.416	1.18	20.0	.734	6.27	¯5.9	.391
23	936	745	425	.414	1.35	20.7	.807	6.25	¯6.1	.415
24	893	719	427	.421	1.53	20.9	.871	6.20	¯5.9	.444
25	842	706	439	.434	1.72	21.0	.919	6.15	¯5.7	.479
26	782	694	446	.434	1.89	18.5	.957	6.04	¯4.5	.511
27	716	672	471	.425	2.09	17.9	.952	5.99	¯3.2	.525
28	646	641	491	.414	2.34	18.9	.950	5.98	¯1.7	.548
29	617	620	507	.394	2.81	24.6	.923	6.06	¯.0	.539
30	614	597	517	.391	3.05	25.0	.925	6.04	1.8	.537
31	620	589	525	.394	3.18	27.6	.943	5.99	4.4	.554
32	634	589	531	.385	3.39	27.8	.941	6.01	5.7	.563
33	641	582	537	.363	3.65	28.3	.925	6.05	6.0	.567
34	644	567	536	.344	3.91	28.8	.913	6.10	6.4	.572
35	642	545	533	.325	4.18	29.0	.896	6.15	6.3	.572
36	637	515	526	.302	4.51	29.5	.872	6.17	5.2	.559
37	627	486	512	.284	4.78	30.7	.857	6.19	4.5	.546
38	625	458	492	.275	5.10	30.9	.847	6.22	4.1	.530
39	630	438	471	.271	5.41	23.8	.823	6.28	2.8	.516
40	640	432	450	.315	5.56	64.2	.735	6.24	6.0	.517
41	659	425	428	.352	6.19	62.5	.766	6.26	5.4	.517
42	669	421	402	.359	.47	62.7	.784	.01	4.2	.513
43	675	417	375	.357	1.05	63.3	.804	.05	2.4	.513

Fig. 8. Tabulation of further regression calculations in frequency bands.

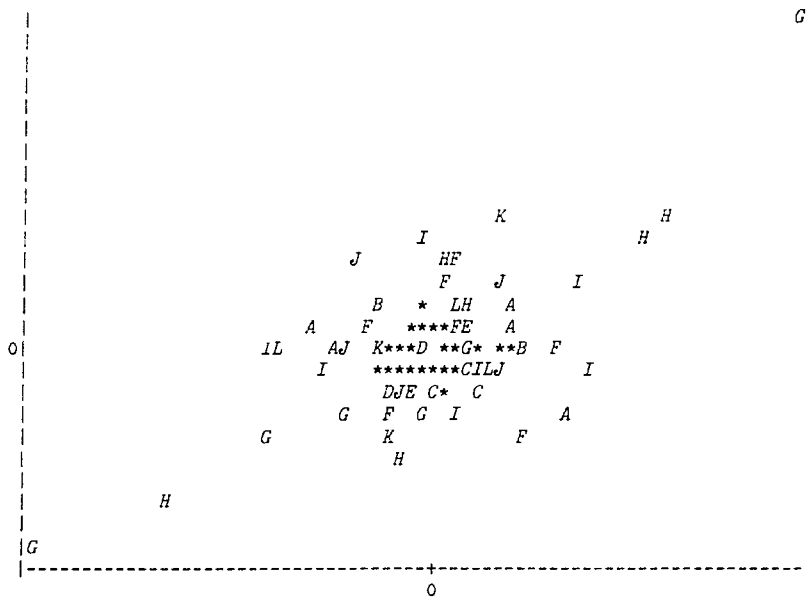


Fig. 9. Copper production vs imports: high-frequency filtering.

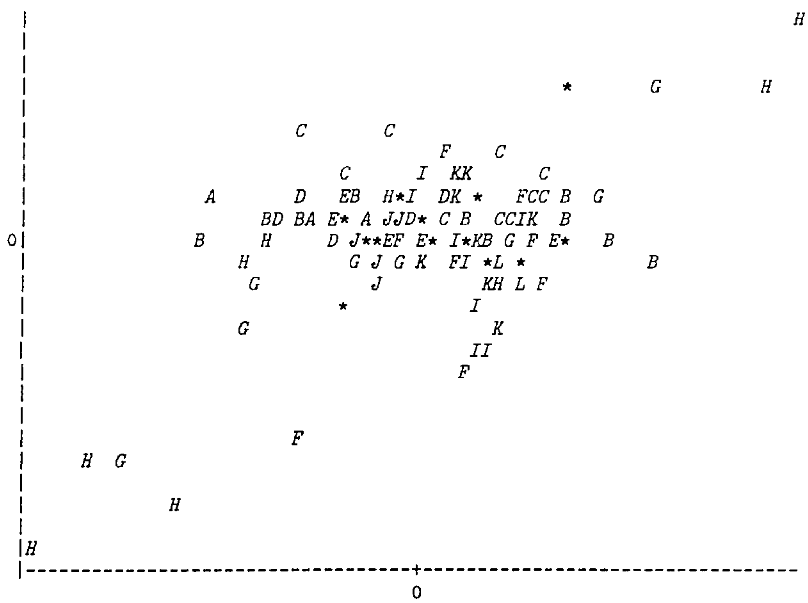


Fig. 10. Copper production vs imports and price: low-frequency filtering.

Would You Want *Your* Child to Be a Statistician?*

J. L. Jaech

Exxon Nuclear
Richland, Washington

First, let me welcome you out-of-towners to Richland. I feel that I can speak as a native since I have lived here for almost 25 years, excluding three years spent in California. When we moved to this (then) desolate spot in 1953, our intention along with everyone else who first moved here was to stay for about a year and then return to civilization. But they say this place grows on you. I can now truthfully say that if I had it to do over again, I don't believe I would.

You often hear it said about New York, or Washington, D.C., or even Chicago: "It's a great place to visit, but I wouldn't want to live there." You hear a similar thing said about our Tri-City area. "It's a great place to live, but I wouldn't want to visit there."

Actually, there are compensations. Take the scenery, for example. Up in the Horse Heaven Hills, south of town, it's simply beautiful among the wheat fields and wide open spaces. And the nicest part about the whole thing is that you don't even have to drive out there to see the Horse Heaven Hills. You just wait for one of our dust storms and watch the Hills blow right by your living room window. (It doesn't all go by; a few cubic yards filters in and settles on the furniture. Visiting our homes after one of these storms, you get the impression from the decorating schemes that everyone's favorite color is a dirty beige.) The chambers of commerce had a contest to pick a slogan to promote the Tri-Cities some years ago. The best and most descriptive entry, although not the winner, was "Wheeze and Sneeze in the Tri-City Breeze."

There are a lot of misconceptions about life here, just as I'm sure there must be about living in Los Alamos or Oak Ridge. Judging from the sensational headlines that occasionally occur in newspapers around the country, this is a very dangerous place to live. Our fellow Americans are concerned about us. In the aftermath of the minor chemical explosion that occurred in the outer area several months ago, we got a phone call from someone in the Midwest asking if the people of Richland were able to leave their houses yet and wander around the streets. A television crew was sent out in a helicopter to take pictures of the large crater that supposedly existed. By the time I got here in the fifties, this was a settled community by most standards. But an acquaintance of mine in Portland was surprised to learn that we had running water and a sewer system. And so it goes.

I had better get to the main subject of my address. Would you want *your* child to be a statistician? I ask this question because as much as we hate to face up to it, the fact remains that statisticians as a group are often regarded with suspicion, with distrust, with wariness, or at least with a vague feeling of discomfort. In short, we are much maligned. Do you want your child to go through what you have endured? Although there are in the world between 20,000 and 30,000 statisticians according to a recent estimate by Kendall,[1] the awareness of what a statistician is and does, of how he occupies his time, is often not there. (Parenthetically I might also quote the following from Kendall's article, "Not all these people are working statisticians." As a former manager in an organization that I'll not name, I can attest to that.)

---

*Banquet address.

1. Maurice Kendall, "Statisticians -- Production and Consumption," *Am. Stat.* 30(2): 49-53 (1976).

Before citing evidences that our image as a profession leaves something to be desired, let me dwell on some positive aspects. What are the desirable features of our profession? I'll not speak of the monetary rewards because, of course, we are above that sort of thing. None of us here tonight, I would venture to say, are in any sense of the word motivated by money.

What, then, are the advantages of being a statistician? For one thing, the job is, generally speaking, not hazardous. In my early days at Hanford, we were required to have monthly safety meetings, and it was a real struggle to plan an agenda for these. You can spend just so many hours dwelling on how to avoid paper cuts. The accidents and near-accidents that I'm personally aware of were not covered in these meetings anyway. I recall, for example, the time my boss in those early days (whom I won't embarrass by naming, but whose initials are the same as those of the Civil Aeronautics Board), was deeply reflecting on a problem and fell over backwards in his chair—only his built-in padding saved him from serious injury. Then, on another occasion, I was standing in the Fred (some call it a John; I prefer calling it a Fred) when, with no warning, a violent sneeze racked my body. My head jerked forward spasmodically and smashed into the top of the urinal, causing a momentary blackout followed by intense pain and suffering. My boss wouldn't let me fill out an accident report, because he had no suggestions on what actions to take to prevent a reoccurrence of the accident. Of course, this particular accident is not peculiar to our profession. It could happen to anyone with a level of intelligence required to be a statistician.

To go on with the positive aspects of being a statistician, there is the inner satisfaction that comes from tackling a tough problem and carrying it to a successful completion. I continue to be amazed at the beauty of mathematics and, in particular, of mathematical statistics—at how a hopelessly complex formulation often reduces to the essence of simplicity. This enjoyment is dampened somewhat, I must admit, when the work is presented with humble pride to a group of peers and, at the conclusion, someone rises and says, "I'm surprised that you are not aware of the paper on this subject in the 1937 issue of the Journal of Outer Mongolian Anthropologists where your key result was developed in just two steps." That tends to deflate, and can be listed as one of the most emotionally damaging hazards of our profession.

Then, there is the joy that comes from sharing our skill and knowledge with others and, in particular, with school groups in the hope of helping to mold their lives. I remember the thrill an associate of mine had some years back when he received a letter of appreciation from a school group he had addressed. I forget the details, but the talk dealt with biological experiments with rats. The letter said, "Dear Dr. N: Thank you very much for speaking to our class last Wednesday. Until you came to visit us, we didn't even know what a rat looked like."

Having disposed of the rewards of our profession, let me develop the theme mentioned earlier, that is, that we as a group have a poor public image.

I offer the following pieces of evidence. How many times have you heard, when being introduced as a statistician, something like the following: "So you're a statistician" (said with an attempt at suppressing mirth that threatens to tear the person apart). "Tell me, is it true that if you put your head in an oven and your feet in ice water you'll be comfortable on the average?" By estimated count, I've heard this, or a variation thereof, about 869 times since 1973.

Or, as another piece of evidence, you've completed a round of golf and someone suggests, "Give it to Harry to add up; he's the statistician." That's bad enough, but then someone checks your addition to find you've made a mistake. Although it may shock some to know that we can't add, we all know that's the reason we chose to be statisticians in the first place. We don't have to get the exact answer—just getting within the confidence interval is close enough. If I could add, I'd have become an accountant or a bookkeeper—not a statistician.

It's difficult to have your children respect you when they don't understand what you do. Here's a typical conversation between youngsters. "My dad's a doctor; what does your dad do?" "He works for Battelle." "But what does he do?" "He works in an office." "But what kind of work does he do?" "He's a kind of an engineer, or an accountant, or something." "My mom says your dad is a statistician; that's not the same as an engineer or an accountant, is it?" "I think so, but his work is so secret, I'm not supposed to talk about it." In case you haven't perceived the undertones there, your child tends to be ashamed of you. Perhaps that's stated too strongly; let's just say he'd rather that you were something else. I won't even bring up how your spouse may feel about your profession.

Those among our acquaintances who are truly interested in learning more about statisticians may turn to the dictionary. We read, "Statistician—one versed in or engaged in compiling statistics."[2] That doesn't help a great deal.

This rather negative attitude toward our noble profession is not restricted to the United States. To quote from Kendall, "Statisticians are still regarded as living in a world of their own and possessing very few human attributes. Nothing is more devastating in a social gathering than to be introduced to a stranger as a statistician and to watch the dismay with which he, or worse still she, wonders what you can possibly discuss on the ordinary social plane. That is not, I think, merely the layman's natural distrust of numerical information. People appear to talk quite happily to actuaries and accountants or even to numerical analysts and mathematicians."[1] I dare not quote further or I shall be accused of plagiarism. I refer you to Kendall's article for a delightful 15 minutes of reading.

As a final example to illustrate how statisticians are, at best, misunderstood, I would guess that many of you have experienced the situation in which you are regarded as a miracle worker. This opinion is not meant to be complimentary, but rather refers to the belief that the statistician can perform some hocus pocus statistical ritual that can turn an unacceptable conclusion based on a set of data into something acceptable. "Here is a set of data for lot such and such where the values for eight out of ten samples exceed the specifications. Will you please analyze the data so that the lot is acceptable?" They would like you to find some way of throwing out the eight outliers.

I could go on and on, and I'm sure you in the audience can cite other evidences to show that our profession often suffers from poor public understanding and, hence, acceptance. We, of course, recognize our true worth. We can become quite self-opinionated and are a bit puzzled that others cannot appreciate us at our true worth. It seems to me that at this point we have three choices: (1) let the public be bleeped; (2) attribute our tarnished image to the impressions created by others practicing our profession who are not nearly so capable or conscientious as we are; or (3) examine ourselves individually to see if, by some stretch of the imagination, we might have contributed to this poor image.

Some of us like alternative (1); most of us, judging from what I read on this subject, embrace alternative (2). The blame lies elsewhere. Perhaps that's why we've made so little progress in improving our image. Is it perhaps time to heed the Biblical injunction, "And why beholdest thou the mote that is in thy brother's eye, but considerest not the beam that is in thine own eye?"[3] Let us spend a few minutes in examining ourselves to see if maybe, just maybe, we have done, or have failed to do, certain things such that part of the blame lies at our own doorstep. (I know in advance that this will be a fruitless exercise. You will probably leave here with the same feeling that churchgoers often have as they leave the church after a particularly damning sermon and comment to the minister, "You certainly told *them* off—that was a fine sermon.")

I will mention some areas of concern to me, but not in any particular order of importance. First, consider the communications problem. It does no one any good if our findings, important though they may be, and representing an excellent analysis, are not communicated in an understandable way. Kendall maintains that one of the reasons that we are undervalued is the inability of many of us to get across our ideas, particularly in writing. I think it would be most helpful if each of us were to subscribe to Kendall's philosophy on communications, and I quote, "If someone fails to understand me I regard the fault as mine, not his."[1]

Problems in communications, to be fair, work both ways. The consultee is not blameless either, but it's ultimately up to us to make sure that we are attempting to solve the right problem. In a paper published 20 years ago, Kimball pointed out that poor communication can easily lead to committing an error of the third kind—giving the right answer to the wrong problem. This is a potentially serious error because it often goes unrecognized.[4]

Perhaps I simplify too much, but in a 1966 paper,[5] I attributed problems in communication to laziness. When we discuss problems with our clients, do we take enough time to make sure we understand the problem—to make sure that we've been given all the pertinent information and not just that which the client thinks we need to solve his problem? That is *our* responsibility, not his. Also, when we communicate in writing, do we take the time and trouble to make it understandable? I suspect that at times we may even take the opposite tact, that is, be purposely obscure in order to impress. Admittedly, it is a great temptation

2. *Webster's New Collegiate Dictionary*, G.&C. Merriam Co., Springfield, Mass., 1973.

3. Matthew 7:3, King James Version.

4. A. W. Kimball, "Errors of the Third Kind in Statistical Consulting," *J. Am. Stat. Assoc.* **52**: 133–42 (1957).

5. J. L. Jaech, "Problems of Consulting Statisticians—The Statistician in Industry," *1966 Joint Statistical Meetings in Los Angeles.*

to include complicated derivations and equations in a report, but it is a temptation to be avoided, in general.

Before leaving the subject of communication, I recall an early incident in my career that impressed upon me the importance of leaving nothing unsaid in dealings with the client. I had designed a fractional factorial experiment for a corrosion engineer and, to save him the trouble, I randomized the order of the 128 (I believe it was) trials when listing the experimental combinations to run. I discussed the proposed experiment with him, gave him the listing, and said goodbye. The next contact was about a month later when he came in with the data for me to analyze. I asked him if things went well. "Oh yes, no problems, except that you gave me the sets of conditions in such a jumbled up order, I had to unscramble them before running the e periment."

Turning to a second problem area, we are faced with the temptation to be too academic, or perhaps too mathematical. This subject is difficult to deal with because it is also dangerous to base results on analyses that are mathematically unsound. Yet, there is one thing to deal with practical problems on a sound mathematical basis, keeping in mind that as a practicing statistician, it is the *problem* that is important, and quite another to dwell on the mathematics, regarding the problem itself as something to be endured but of no interest.

Those of us assembled here have had all kinds of formal training in statistics. I myself attended a school where it was considered a mortal sin for any mathematics professor to even imply that there might be something useful in what was being taught. As a result, when I was turned loose, I knew how to prove Cochran's theorem, but not how to fit a straight line through a set of data. I used to keep a copy of Brownlee's old paperback hidden in my drawer so I could handle my assignments.[6] I'm convinced that my one professor, teaching matrix theory, who was supposed to tell us how to solve the problems in life that he had avoided by becoming a professor, had no idea of the practical importance to statistics of matrix theory. If he did, he managed to hide it from us.

But I am supposed to fix the blame in ourselves, and not in others. We, each of us, as supposedly practical statisticians, cannot afford to expend a significant portion of our energy on pursuing the intellectual pleasures of pure mathematics—except as a hobby. Kendall feels very strongly on this. No one can accuse Kendall of being nonmathematical. He writes, and I quote in part, "Nowadays there is a brand of mathematician who is a danger to our subject, or at least, to the acceptance of our subject in the worlds of science and business... there *is* a place in the world, even in the world of experimental science, for the scientist who is mainly interested in studying his own mind. Where we have gone too far, I think, is in allowing him to acquire pecking order over the scientist who is interested in dissecting and reducing to order the external world. The intricacies and austerities of mathematical statistics are such as to encourage intellectual arrogance on the part of their practitioners. I do not think we should let them get away with it."[1]

Turning to another, but related, subject, we consider the poor reputation we've acquired because of improper modeling. We all have pet techniques, which change over the years, and we continually seek to find problems that fit these techniques. If they don't fit exactly, no bother; I'll change the assumptions to make them fit.

In a paper that I had the occasion to reread lately, Professor Anscombe, who has honored this Symposium with his attendance, gets at the core of the problem and at the solution as well. He writes, "What is important is that we realize what the problem really is and solve that problem as well as we can, instead of inventing a substitute problem that can be solved exactly, but is irrelevant."[7]

One challenge to the practicing statistician is that reality hardly ever corresponds exactly to models on which available techniques of statistical analyses are based. On the one hand, this opens up exciting areas of potential research but, on the other hand, it can lead to time-wasting activities if we carry the problem of equating the model to reality to the extreme. How close a correspondence is needed? What are the consequences of failures in the assumptions inherent in the model? These are the questions of importance.

Criticism leveled at our profession in the area of model building is sometimes justified, and sometimes not. "You statisticians, you are so unrealistic, you assume everything is normally distributed." This is not true, of course, but some of our critics are convinced of this. We need a new image. By following Anscombe's advice, we should be able to create this new image and be regarded as realists in future years.

I must touch on one other point before leaving this subject, and that is the extent to which we are personally responsible in our spheres of influence for misapplications of statistical technique because the

6. K. A. Brownlee, *Industrial Experimentation*, 4th ed., Her Majesty's Stationery Office, London, 1949.

7. F. J. Anscombe, "Rectifying Inspection of a Continuous Output," *J. Am. Stat. Assoc.* 53: 702-19 (1958).

models are just not appropriate. Standard techniques are being applied by others on a routine basis: t tests, tests for outliers, calculation of tolerance intervals, etc. Do you ever check into the structure of the data to see that these common techniques are properly applied? Does statistics get a bad reputation, unbeknownst to you, when ridiculous answers are occasionally reported because the model simply does not fit? Too often we hear about this after the fact. As a case in point, one auditor auditing our plant took us to task because certain of our data, for which tolerance intervals are routinely calculated, were not normally distributed, according to his application of the W test for normality with which he was familiar. Closer inspection revealed that large relative rounding errors were responsible. This negative audit finding could have been avoided had I maintained closer contact with the application.

I touch briefly on another area where we deservedly have earned a bad reputation on occasion - timeliness of response. If we cannot provide answers when needed, then we are of little use to our clients. This is a tough problem area, because I suspect that most of us operate under time constraints. It's a real temptation to give priority to the more interesting problems and neglect the others, regardless of their importance. Most of us are hesitant to give advice without careful study of a problem, but well-thought-out advice after some action has already been taken is obviously worth less than timely advice based on available resources. I suspect that we all have drawers full of problems that we fully intend to give more thought to when we have time. If we wait to complete a project until we're 100% satisfied with all aspects of it, very little would be completed.

We are in the computer age, and this introduces a whole new set of problems. In balance, of course, computers have been a great boon to our professions. I recall in my early weeks at Hanford when a programmer and I struggled to invert three 8 by 8 matrices. It was easier to use the inverted Doolittle method on my mechanical Marchant calculator than to get the right answer out of that existing generation or computers.

However, the computer age also creates problems for us. I am not anti-computer by any means, although I am probably the only statistician in the country who doesn't know how to program (I and Carl Bennett). Yet, we have to face up to the dangers inherent in the misuse of computers. I identify two such dangers. First, the existence of so many package routines often replaces the thought process and can lead, if we are not careful, to the problems of poor correspondence between the model and reality. Secondly, and here is a real danger in my opinion, we lose the "feel" of the data through over-reliance on the computer. I recall a very expensive corrosion experiment with which I was peripherally associated some years ago in which the effect of iron on corrosion rate was reported to be dominant, and the strong quadratic nature of the effect puzzled the lead experimenter. He called me in to see if I concurred with the computer analysis. The problem was quickly detected. One observation was way out of line, having been incorrectly keypunched, and this dominated the results. Hopefully, we've become more sophisticated in routine processing and analysis of data in the intervening years, but we dare not completely lose contact with the raw data. I might emphasize, in view of our topic, that the statistical analysis had been blamed for this puzzling result. Was the blame deserved?

While on the subject of computers, I am also troubled at the overuse of computer simulation in solving problems. Granted that simulation is often needed, it is used on occasion in my opinion to solve problems that can easily be handled in less expensive and more exact ways. We are still able to think; let's not let the computer get all the credit.

I touch on another subject for which there is no solution, but which, unfortunately, contributes to our tarnished reputation. I refer to the fact that statisticians don't always agree with one another. Our opponents capitalize on this. "You statisticians! You can't even agree among yourselves; why should I accept what you say?" You and I realize, of course, that that is the beauty of statistics over mathematics. I am more comfortable in a situation in which the "right" answer is mostly a matter of opinion.

As an example, I have a good friend, whom I shall not name, but who is known to many of you. We have had public disagreement concerned with biases and systematic errors in nuclear materials safeguards applications for a number of years now, much to the glee of those individuals who want nothing to do with statistics. This disagreement doesn't particularly bother us. There is little chance, in my opinion, that these disagreements will ever be resolved. On the one hand, my friend is too stubborn to admit he's wrong. I, on the other hand, am not wrong, so we are at an impasse.

I think we must get the message across to our clients that disagreements among members of the statistical profession are to be expected. May I again quote Kendall. "The statistician... is rarely sure about anything. Ours is a logic of uncertainty. We make almost all our statements in terms of doubt, of expectation, of chances in favor. And rightly so, because that is what life is like. But businessmen do not care for uncertainty in the advice they receive or the statements which they are given. The function of the statistician, as they see it, is to give them accurate information. They will do their own doubting."[1]

I am reminded of a phone conversation I had many years ago, which is pertinent because of the analogy. "Hello, is this Mr. Jaech?" "Yes it is." "This is Bill at the analytical lab. I was told to talk to you about your request that we make duplicate analyses on such and such samples. I'm opposed to this practice." The conversation continued as I tried patiently to explain the reasons for the request, in my usual diplomatic way. This sound reasoning failed to convince him. Finally, in exasperation, and sensing that I was above my conversee on the organizational totem pole (which placed him pretty low), I in essence directed him to do the duplicate analyses. "Well, okay, I'll make both analyses, but I warn you, you're going to get two different answers." The point is that if two analytical results don't agree, why should we expect two statisticians to agree? I admit that this analogy is somewhat far-fetched, but I wanted to work in that telephone conversation somehow.

I have left a very important area until the end of my list of reasons why we as a group are often maligned. I refer to our lack of interest in the quality of the data base, to our tendency to want to get on with the problem and apply our methodology without being overly concerned about how good are the data. Some say that this is not the responsibility of the statistician, and in some sense, I suppose I agree. However, when we deliver what turns out to be poor advice, or give faulty conclusions because we've not dug into the data, we're the ones who tend to be discredited.

My message is, *always* be suspicious of the data base! Questioning the data base quality can be time consuming, can be boring, and can, initially at least, raise the hackles of the client if he feels the need to be defensive about his data. (One hesitates to deal with a client who has raised hackles; given the choice, one would rather find a way to warm the cockles of his heart since it's much more pleasant to deal with a client with warm heart cockles than with raised hackles.)

In referring to checking the quality of the data base, I am not speaking merely of performing such actions as running outlier tests or testing for normality, or of things of this nature. More often than not, some kinds of data are deficient in having been "cleaned up" or "massaged" too much by the client before you see them, and outlier tests are superfluous. Rather than performing these tests (although they certainly have their place), I refer to the whole process of checking the data for internal consistency, of questioning how the individual numbers might be related, or making sure you understand how the numbers were derived, of not accepting data as God's truth just because they are in the form of computer printout. (If anything, this triggers my suspicions rather than sets them at rest.) Some random examples may help to clarify the point I'm trying to make.

I recall an incident that happened many years ago in which a colleague of mine was given what appeared to be a beautiful set of data which formed a nice smooth curve when plotted. The client wanted a curve fit through the data. Several models were tried, each one getting more complicated, but although one could get close, there was always obvious nonrandomness in the residuals, indicating an inadequate model choice. Finally, in desperation, the statistician met with the client to discuss the results and express his puzzlement at why, with such a beautiful set of data, no reasonable model seemed to give a good fit. The client's frown progressively deepened as he searched within himself to find an explanation. "Well, perhaps if we look at the raw data we'll get a clue." In a distressed voice, my colleague responded, "I thought that was the raw data." "Oh, no. I used a French curve to eyeball fit the data and gave you some points off the curve." Conclusion number one is be suspicious also of beautiful data exhibiting little random scatter.

As a second example, don't always accept obvious "facts" stated by the client. A set of data was given me some months ago in which concern was expressed over the large unexplained lot-to-lot variation in the oxygen-to-metal ratio of uranium pellets. I questioned whether this actually was lot-to-lot variation or rather might be, in part at least, reflecting day-to-day analytical variation in the lab. It was agreed that a small experiment would be run to explore this possibility, and in fact, almost all of the variation turned out to be attributed to analytical difficulties. This discovery led to a totally different set of actions from what had originally been suggested. It is also further evidence of the truth of the famous corollary, "Measurement errors are always at least an order of magnitude larger than claimed by the person responsible for the measurement."

At times, initial results of an analysis will point a finger of suspicion at the data. In a 1962 JASA article, "The Case of the Indians and the Teen-Age Widows,"[8] keypunch errors in some of the 1950 census cards were uncovered as a result of what the authors called a "statistical detective story." They questioned the fact that there were more 14-year-old widowers than there were at ages 15, 16, etc; that there were similarly more 14-year-old divorcees; and that the number of young Indians seemed excessive. After a thorough investigation, circumstantial evidence pointed to the probable causes of these peculiar results.

My last example concerns an event of 18 years ago. I was traveling in a car toward Seattle with a friend and my six-year-old son. To keep him occupied (the son), I had him keep a tally of the last digit of license plates, wanting to show him that all ten digits would tend to appear the same number of times. We had agreed to stop collecting data upon entering the city of Yakima. Just after we passed the city limit sign, he saw one more car, with the digit 4. Although he had set down the pad of paper, he picked it up to mark in the 4, "because I'm short on 4s." The point is that the tendency to bias data begins at a very early age.

We've covered a number of reasons why our profession suffers from a poor public image and, by implication at least, have indicated some things that we as individuals can do about it. This has taken a good deal of your time and mine. It would have been far simpler, and also just as meaningful in my mind, if not more so, had I only had 3 minutes to fill rather than 30 or whatever. Then my message would have been simply, but completely: In order to achieve respectability in our profession, we must, each of us, live by a code of ethics. Granted, we do not have a *formal* code of ethics but if we, as individuals, do not know what it means to follow a code of ethics in carrying out our responsibilities, then we're in worse shape than I thought.

In closing, may I offer this bit of advice. We may as well learn to laugh at ourselves; everybody else does.

8. Ansley J. Coale, and Frederick F. Stephan, "The Case of the Indians and the Teen-Age Widows," *J. Am. Stat. Assoc.* 57: 338–47 (1962).

Research Papers

Exploratory Data Analysis and Classical Statistics: Their Abilities to Shed Light on Energy Issues*

Lawrence S. Mayer

Princeton University
Princeton, New Jersey

ABSTRACT

Exploratory data analysis and classical statistics are compared in terms of their potency for the analysis of major energy issues. First, the positions of the two approaches in the contemporary philosophy of science are considered. The relevance of these positions to the major energy issues is developed. Then a set of guidelines for the use of exploratory data analysis on energy data is presented. Finally, the exploratory and classical approaches are compared by using each to estimate the price elasticity of residential natural gas. The analyses suggest those areas in which each approach is most useful.

INTRODUCTION

For even the casual observer of our nation's energy situation, it should be clear that many of the difficult decisions which must be made in planning the course of our nation's energy future require accurate information about key energy parameters. For example, a reasonable choice of whether or not to implement a conservation program in the commercial sector requires a knowledge of the degree to which the program in question will impact demand as well as a knowledge of the hardships that the program might force upon the targets of its provisions. Because little is known about the potential impact of such programs, debates such as the one about the potency of the marketplace vs public law as the major conservation mechanism in the commercial sector tend to be nothing but expressions of prior political positions. Similarly, decisions regarding the optimal rate of utilization of our domestic reserves of crude oil obviously require reasonably accurate estimates of the size of such reserves as well as estimates of the response of the discovery-recovery mechanism to variables such as price and changing life styles. As evidence of the absence of such accuracy, the extremes of current published estimates of domestic crude oil reserves have a ratio of 18 to 1.

To obtain knowledge about key energy parameters such as current energy utilization rates and patterns, size of reserves, energy embodied in products, the impact of person/machine interactions on demand, response of demand to changes in price or culture, and the potential impact of various conservation programs, federal and state policy makers need to draw on the talents of people versed in the subjects of data gathering, data storage and management, data exploration, and data analysis. I contend that both classical statistics and exploratory data analysis are tools which are of considerable importance in trying to unravel our energy past in order to assess our energy present and plan our energy future. I will argue that classical statistics is being under-utilized by the policy-making process while exploratory methods are almost totally ignored. Furthermore, in those few studies in which classical inferential procedures are used, the problems often would be more amenable to analysis by the informal approach

*The author acknowledges the support of the National Science Foundation, Grant No. 72-03516-A04; the Energy Research and Development Agency, Grant No. E(11-1)2789; and the Office of Energy Programs, Department of Commerce, Contract No. 6-35599.

of exploratory data analysis. I will describe several important characteristics of empirical research in the energy policy area and then proceed to outline what I consider to be the proper role for formal and informal statistical methods to play in this area.

My feelings about the appropriate role for statistics in the analysis of energy problems have arisen, in part, as a function of three of my research efforts—efforts which have spanned the past several years. First is my involvement as one principal investigator on an interdisciplinary multiyear study of residential energy demand, a study which focused on analysis of the components of demand in a planned residential community in central New Jersey.[1] My involvement and the involvement of other statisticians led to a study of the effects of the onset of the energy crisis on residential demand,[2] studies of the effects of changes in price on the demand for natural gas and electricity,[3,4] development of a simple two-parameter model of the demand for natural gas as a function of an indicator of the coldness of a month,[5] a statistical analysis of the effects of physical modifications to the dwelling on demand for energy,[6] and several more technical collaborative modeling efforts involving statisticians, physicists, and engineers. This involvement convinced me that a careful data analyst armed with both classical statistics and exploratory methods and blessed with a little good data can make a significant contribution to understanding issues such as the consumer reaction to changes in price, issues which are central to our understanding of our energy environment.

The second effort has been an 18-month evaluation of econometric models for forecasting our energy future.[7] This effort has convinced me of the limitations of classical statistical methodology as a tool for developing basic understanding of energy-related behavior in areas which lack both well-developed empirical theory and reliable, valid data. I am particularly skeptical of using classical statistical methods to make longitudinal inferences from non-experimental cross-sectional data, a common energy econometric practice.

The third effort has been an attempt to provide a philosophical base for the use of exploratory data analysis as an alternative to classical statistics.[8] My approach involves the construction of a set of "meta-theorems" which act as informal guides to deciding whether exploratory methods are appropriate for the analysis of the problem under investigation and, if they are appropriate, act as informal guides to the selection of a particular exploratory technique. The emphasis of this work has been to argue that once cast into a sound framework in terms of philosophy of science, exploratory methods are easily viewed as a supplement to classical methods within the repertoire of the competent data analyst. The only competition that should arise between the use of classical methods and exploratory methods in policy analysis in general and in energy policy analysis in particular should arise because policy problems are often so poorly defined that the statistician has no way to decide which of the approaches is more appropriate. If the problem of interest is well defined, then the choice between informal and formal methods of inference is a rather easy one, and consequently, exploratory data analysis and classical statistics are complementary and not competing tools.

These efforts have led me to formulate a firm position regarding the optimal interaction between data, statistics, and energy policy.

CHARACTERISTICS OF THE ENERGY POLICY AREA

Several characteristics of the energy policy area make it a suitable arena for the use of statistical methods in general and for the use of both exploratory data methods and classical statistical methods in particular. First, the area is important for the future of our society and thus is worthy of the statistician's

---

1. R. Socolow, *The Twin Rivers Program on Energy Conservation in Housing: A Summary for Policymakers*, Report 51, Center for Environmental Studies, Princeton University, Princeton, N.J., 1977.

2. L. S. Mayer, "Estimating the Effects of the Onset of the Energy Crisis on Residential Energy Demand," *Energy and Resources* (to be published).

3. L. S. Mayer and C. Horowitz, "The Relationship Between the Price and Demand for Natural Gas: A Partially Controlled Study," *Energy Res.* 1: 193-222 (1977).

4. L. S. Mayer and C. Horowitz, "The Effects of Price on the Residential Demand for Electricity: A Statistician's Estimate" (unpublished manuscript).

5. L. S. Mayer, "Modeling Residential Demand for Natural Gas as a Function of the Coldness of the Month," *Energy and Buildings* (to be published).

6. T. H. Woteki, *The Princeton Omnibus Experiment: Some Effects of Retrofits on Space Heating Requirements*, Report 43, Center for Environmental Studies, Princeton University, Princeton, Mass., 1976.

7. L. S. Mayer, *The Value of Econometric Models for Forecasting Our Energy Future*, a report to the Office of Energy Programs, Department of Commerce, 1977. [A condensed version appears in *Proceedings of the International Conference on Energy Management, October, 1977, Tucson, Arizona*, Pergamon, New York (to be published).]

8. L. S. Mayer, "Exploratory Data Analysis and Classical Statistics: A Data Analyst's Perspective" (in preparation).

attention. Unlike areas of more academic interest, the energy area involves the making (or postponing) of difficult decisions regarding our energy future. Because the growth of energy utilization cannot, and will not, follow historical patterns, the question of how to reduce the growth in demand is paramount. To whatever degree the statistician can help unravel the patterns of current end-use statistics, test the potential of conservation programs, test the potential of new supply technologies, and monitor the performance of implemented conservation or supply programs, he/she can make a significant input into our future.

Second, the energy area involves considerable underlying physical theory and the policy maker must make sure that programs are consistent with the physical realities of energy processes. For example, the notion that "if enough economic incentives are given, the supply of domestic oil reserves will expand indefinitely" appears to be part of the energy economic folklore; it is a claim that is totally ignorant of the physical realities of energy supply. If the laws of economics contradict the laws of physics, I choose to have faith in the latter. I believe that the statistician is, or should be, trained to deal with the complex interaction between physical laws, data, and public policy problems. The complications introduced by physical laws make the energy area ripe for analysis by statisticians who are well versed in the fundamentals of physics and chemistry and yet are oriented toward policy problems.

Third, the energy area is plagued by severe data problems. Often the data needed to test a given theory, such as a theory of consumer demand, is partially missing and partially so unreliable that it might as well be missing. To whatever degree the statistician is capable of analyzing "weak" data and encouraging the collection of better data, the energy area appears an excellent domain for his/her efforts.

Fourth, energy policy analysis often involves subtle interactions between variables which at first glance appear either unrelated or related in a way different from their empirical relationship. Often policy initiatives which have direct positive benefits locally may exacerbate the overall energy picture. For example, a program which directly encourages residential consumers to reduce their thermostat settings may indirectly encourage consumers to use their stoves for supplemental heating of the kitchen. The overall effect of such a program might be to increase the residential demand for energy. Clearly statisticians trained to dissect data to obtain estimates of the nature of such interactions would be most helpful in such areas. Consideration of these interactions brings us to the final characteristic of the energy policy area: the need for large-scale experimentation with conservation strategies. Unlike areas such as health policy in which experiments usually involve serious moral and ethical problems, almost all strategies for reducing energy demand, such as the provision of tax incentives for home insulation or the provision of peak-load pricing structures for commercial consumers, can be tested without serious moral or ethical complications. The statistician should be involved in the design and analysis of such experiments. The design of such experiments is complicated by legal and regulatory practices and by the fact that the effect produced by a successful policy program is often small relative to the effects of uncontrolled determinants of demand. For example, it would take a carefully designed experiment using a controlled sample to demonstrate the effects of a few percent rise in energy prices on residential demand, because energy demand is heavily influenced by uncontrolled variables such as structural defects of the dwellings and variations in outside temperature. The furnace responds to temperature more than it does to price.

THE APPROPRIATE ROLE OF CLASSICAL STATISTICS

Delineation of the proper role for classical statistics to play in the formulation and analysis of energy policy begins most easily with criticism of the current use of such methods in energy analysis. The majority of formal statistical analyses found in the energy area are contained within the body of literature which develops either econometric models of a single energy variable such as the residential demand for a single fuel or econometric models of the entire energy/economic system. The level of sophistication of the statistical procedures found in such models varies enormously, as does the degree to which the principles of classical statistics are correctly applied. I will not deal with these issues. The goal of these models is to forecast one or more energy variables into the future; even where the principles of statistics are correctly applied, I am highly suspicious of these forecasting devices since they give little evidence of their accuracy and model validation is ignored. I see little reason why, in the absence of scientific scrutiny, the policy maker should treat forecasts based on formal statistical models as any more accurate than the predictions of seers.[7] In addition to the validation issue, I have doubts about the wisdom of using formal statistical inference in the development of energy-forecasting models. My doubts stem from the following problems:

First, the empirical theory regarding the behavior of energy actors is usually so weak that the necessary scientific framework within which statistical inference makes sense is rendered inoperative. (Often in an econometric model the residential consumer is assumed to shop for a fuel type the way he or she shops for a pair of shoes in the market, an assumption which is given no empirical support and virtually no theoretic support.)

Second, if the estimates of the variance of parameters in the models and estimates of the variance of the forecasts generated were reported they would render the parameters and forecasts to be "ball park" estimates and virtually useless from the policy perspective. For example, the success of an energy program such as the gasoline tax proposed in the 1977 National Energy Plan requires the elasticity of demand to be in such a narrow range that the question of whether or not the elasticity falls in this range would easily be seen to be beyond the resolution power of classical inferential procedures provided the variances of estimates were reported.

Third, even if the forecasts could be made more accurate, classical inference places inordinate concentration on the concept of explained variance. Often from the policy perspective an important component of the variance of a particular energy variable is too small to be significant in the statistical sense. Suppose one is trying to assess whether customers of a particular utility will reduce their demand in response to a few cents increase in the price of electricity. The variance due to the price change is going to be small; for this variance to be detected, very large samples and accurate models of demand are needed. Without such models the effects of small changes in price are important but are easily lost in the variation contained in the response of demand to physical factors such as outside temperature and length of daylight period. Simple classical methods such as analysis of covariance usually do not provide adequate controls for the effects of such variables on demand, because they leave too much error variation uncontrolled.

The fourth problem with the current use of classical statistics is that the units of analysis and the time frame found in such analyses are often not the ones of interest to policy makers. It is common practice, for example, to use the formal theory of linear models with cross-sectional data at the state level to fit a model of the relationship between energy prices and residential demand. The model is then used to forecast the response of individual consumers to a change in the price of fuel delivered to their household. While such articles may be statistically sophisticated, the logical foundation of this type of inference is at best questionable.

The final problem is one of emphasis. Due to the lack of empirical theory in the energy area, one of the goals of any data analysis should be to assess the form of an effect or relationship and not to be satisfied to test for statistical significance. It is not particularly instructive to know that a community responded to the onset of the energy crisis by reducing its demand for energy unless we know the form of the response. Did the high consumers respond more, or less, than the low consumers? Are there any indications that people responded more in the cold months than in the marginal or summer months? Is the form and degree of the response related to simple design characteristics of the dwelling? Does the response appear to be a response to changes in price? Without answers to these questions, the fact that consumers responded to the crisis is a piece of information which may be of academic interest but is not of policy relevance. Knowing that a nonzero effect exists may soothe the academic mind, but the policy analyst needs to know the form of the effect as well as its magnitude. Classical statistics is partially unsuited for the assessment of such form due to its emphasis on estimation and testing.

I feel that the strongest role that classical statistics can play in the formulation and analysis of energy policy is in the design and analysis of large-scale experiments for testing the potency of various strategies for increasing energy supply and strategies for reducing energy demand. For example, if utility companies are to be forced to change their rate structures to ones with marginally increasing prices or to change their billing practices so that the consumer receives accurate and complete statistics regarding his demand, then there ought to be full-scale experiments which demonstrate the value of such changes. Similarly, the claims of the emerging industry which markets energy-conserving materials and equipment should be tested statistically "in the field." Classical statistics should be involved in the design, conduction, and analysis of such experiments. These experiments need to be designed to eliminate confounding sources of variation; classical statistics could then be used to give accurate estimates of the effects of various "treatments" on demand as well as to give probabilistic statements about the probability that such treatments have an effect.

The second role that classical statistics can play in the energy-policy area is in the development of statistical models of the thermal performance of various pieces of energy equipment. For example, classical modeling techniques are probably well suited for generating first-order approximations to the thermal response of the dwelling to changes in the outside temperature. Such models could be used to assess the impact of a community retrofit program on the total energy demand for the community. The methodology would require (1) using simple experimentation to estimate the effect of the program on the parameters in the statistical model and (2) using estimates of the degree of saturation of the program to estimate the total impact on demand. Most current models of the thermal performance of the residential structure are complicated computer models which rely on engineering principles and have never been validated by field data. Such models are usually too detailed, cumbersome, and possibly inaccurate to be used to monitor a conservation program in the field.

The third role for classical statistics is in the development of simple microlevel models of the process which can be used to make optimal energy decisions. For example, many industrial decision makers face the choice of whether or not to replace natural-gas-consuming pieces of equipment by coal-fired equipment. Although it is clear that the decision depends on the relative price and availability of the two fuels in the future, it is not clear how projections of price and availability are best used by the decision maker. Similarly, a state legislator might want to utilize a simple decision model in determining whether to alter the process by which local building codes are developed. Again, empirically based classical statistical analyses are well suited for such tasks. We believe that classical statistics is well suited for these tasks because the nature of the problem forces the decision maker to adopt a relatively strong empirical theory.

THE APPROPRIATE ROLE OF EXPLORATORY DATA ANALYSIS

I do not see exploratory data analysis as a replacement for classical statistics in the analysis of energy-policy problems. The output of classical statistics includes probabilistic statements, statements which allow the policy person to attach an uncertainty to his/her claims. The realities of the policy process are such that the policy maker is probably better off using such statements than using more informal inferential statements provided he/she has reason to believe that the formal statements are reasonably accurate. Probabilistic claims appear to "sit well" with legislators and citizens. The accuracy of such statements depends on the validity of the inferential framework within which they are generated, and in turn, the validity of the inferential framework depends on the validity of the empirical theory adopted. Thus, we suggest that the policy maker lean toward the use of classical statistics provided the statistical theory underlying an analysis is understandable and palatable. Unfortunately, in most areas of energy policy so little is known about the behavior of energy actors that we are hard pressed to suggest even plausible behavioral theories. Although some economists tell us that we can understand both the individual and aggregate energy consumer by having faith in neoclassical economic principles, the economist gives us little reason to believe this claim.

I propose that exploratory data analysis is the optimal tool to use in developing an understanding of energy phenomena in problem areas where "understanding" involves the generation as opposed to the testing of empirical theory. Furthermore, I propose that there is such a scantiness of empirically scrutinized theories in the energy policy arena that exploratory data analysis is ideally suited for analysis of most energy data.

Although it can be argued that much of the data gathered in the energy area is not worthy of any analysis, it is important to note that the need for reliable, valid data is probably much less for use of the exploratory analysis method than for use of classical statistics. An exploratory data analyst can peruse even bad data to try to assess general patterns and trends. Because the exploratory analyst makes fewer formal statements about data than does the classical analyst, these statements tend to be much less influenced by missing data or the presence of bad observations.

Two recent efforts[3,4] illustrate some of the advantages of using exploratory methods in the analysis of energy data. In the first analysis, the typical consumer in the community under study responded to the onset of the crisis by significantly reducing his/her demand for both natural gas and electricity. Close exploration of the data reveals that both of these responses vary as a function of the month and weather. Furthermore, there is little evidence that the reduction in demand is in response to changes in price, the major increases in price occurring months after the major decreases in demand. The analysis was repeated using the standard econometric approach and religious application of the principles of classical statistics. The second analysis gave strong evidence to the conclusion that consumers responded to changes in price and gave no indication of the differential response as a function of the month. We prefer the exploratory analysis because its claims are based on far simpler assumptions than are the econometric claims and its methods are very straightforward.

In a related effort[5] we have developed a simple indicator of the coldness of the month and then modeled monthly demand for natural gas for the gas-heated units under study as a nonlinear function of the two parameters in the indicator. The first parameter is a reference temperature which estimates the interior temperature of the dwelling plus the free-heat contribution of the appliances, occupants, and sun. The second parameter is a slope parameter which estimates the response of the space heating system to a 1° change in the exterior temperature. Using robust methods, the model was fit to monthly demand for each dwelling for all the months since the onset of the Arab oil embargo. The model predicts monthly demand very well yielding an average product correlation of over 0.98. We also fit this model to monthly demand for each dwelling for all months prior to the embargo. We then examined the two models to see whether the response of the consumers to the onset of the energy crisis would be reflected in the parameters of the model. If most observers of the energy world are correct, then consumers responded to the onset of the crisis by reducing their thermostat settings a few degrees over the entire heating season. Such a change would be reflected by a lower reference temperature in the post-embargo model than in the pre-embargo model with equal slopes and correlations. Our analysis shows that the response to the crisis almost uniformly reduced the slope parameter and did not affect the distribution of reference temperatures or the quality of fit. At first glance this result seems to indicate that the consumers responded to the crisis by altering the thermal characteristics of the dwelling, perhaps by adding insulation to the attic or by installing storm windows. However, our experience in the community as well as our exploratory analysis of detailed hourly data from 28 of the dwellings indicate that such modifications did not occur. We thus conjecture that the change in the slope parameter reflects the fact that consumers only responded to the crisis by lowering their thermostats in the very cold months, a behavior which is reflected in the slope parameter. This conjecture has important policy implications. First, it suggests that, contrary to economic theory, consumers respond directly to the total cost of energy and only indirectly to its price. Secondly, it suggests that the mechanisms which affect demand in the severe part of the winter may not affect demand in the milder months. Although demand for energy is relatively small in these mild months, it may be an important policy goal to reduce such demand since the reduction in space heating in these months probably does not cause the discomfort it does in the coldest months. These hypotheses, which are generated by exploratory analysis, should be the objects of further statistical study. I suggest that proper testing of these hypotheses will require large-scale experimentation involving the use of classical statistics.

Energy analysis would benefit from the use of exploratory methods to uncover the empirical regularities in complex data, regularities which are then put to scientific scrutiny using classical statistical designs and analysis. Under this scenario both approaches can contribute significantly to understanding our energy environment and planning our energy future.

On a Method for Detecting Clusters of Possible Uranium Deposits*†

W. J. Conover

Department of Mathematics
Texas Tech University
Lubbock, Texas

Thomas R. Bement

Los Alamos Scientific Laboratory
Los Alamos, New Mexico

Ronald L. Iman

Sandia Laboratories
Albuquerque, New Mexico

ABSTRACT

When a two-dimensional map contains points that appear to be scattered somewhat at random, a question that often arises is whether groups of points that appear to cluster are merely exhibiting ordinary behavior, which one can expect with any random distribution of points, or whether the clusters are too pronounced to be attributable to chance alone. A method for detecting clusters along a straight line is applied to the two-dimensional map of ^{214}Bi anomalies observed as part of the National Uranium Resource Evaluation Program in the Lubbock, Texas, region. Some exact probabilities associated with this method are computed and compared with two approximate methods. The two methods for approximating probabilities work well in the cases examined and can be used when it is not feasible to obtain the exact probabilities.

INTRODUCTION

Whenever several points (occurrences) are located on a two-dimensional map, a natural question that sometimes arises is whether the points tend to cluster together in some parts of the map or whether such apparent clustering is merely the result of the chance clustering one can expect under a random distribution of the points. Examples include the location of rocket bomb hits on London in World War II, craters on the moon, or high radiation readings in an aerial reconnaissance search over a certain area. The latter application prompted this study.

The method for determining clusters is quite simple. A rectangular window of some predetermined size, much smaller than the entire area of the map, is moved across the map. Whenever the number of points within the window equals or exceeds some predetermined number k, the entire area covered by the window is considered a cluster area, and that region on the map is shaded in or otherwise marked. The size of the window is determined from practical considerations, for example, the smallest area that has some real interest to the investigator. The value for k is related more to probablity considerations and is the primary subject of this paper. The entire method is merely an application to a two-dimensional map of a procedure that has been widely used in one-dimensional situations.

Let us first examine the one-dimensional situation. Suppose events in the time period $(0, t)$, or along a line $(0, t)$, occur according to a Poisson process with parameter λ / r. That is, the times between successive events are independent exponentially distributed random variables with mean r/λ. We define a cluster to be the occurrence of k or more events within a time period of length r, for some predetermined values of k and r. Let $P_k(\lambda, r, t) = P_k$ be the probability that no clusters occur in the interval $(0, t)$. Then choosing k so that P_k is large, say 0.95, furnishes a most powerful

*The work reported in this paper was performed under the auspices of the Energy Research and Development Administration, Contract No. W-7405-ENG. 36, Grand Junction, Colorado, Office.

†Report LA-UR-77-769, Los Alamos Scientific Laboratory, Los Alamos, New Mexico.

procedure for detecting certain types of cluster-causing phenomena.[1]

Applications of this model abound in the literature. In the spare parts problem,[2] items fail according to a Poisson process and are immediately replaced by one of k available spare parts. The failed item is repaired in time r. The probability that there is always a spare (or repaired) part available when needed in the time period $(0, t)$ is represented by P_{k+1}. In human physiology one theory holds that an impulse is sent from the eye to the brain if k or more photons strike the same area in less than a fixed period of time r.[3] In physics the impurities along a line in a crystal are distributed as in a Poisson process, and a certain phenomenon occurs when k or more impurities are found in an interval of length r.[4]

In the two-dimensional application a single pass consists of moving the window horizontally the complete width of the map. The number k is selected so that the probability P_k associated with one pass is close to one. If the map represents a two-dimensional Poisson process, the number of points within the moving window at any time has the same distribution as the number of points in a moving interval in the one-dimensional case. If the map results from a system of parallel one-dimensional Poisson processes as in aerial reconnaissance, then the window may include several lines at one time as it moves parallel to them and the number of points in the window at one time again behaves as in a one-dimensional Poisson process.

Although k is selected from the P_k computed on the basis of one single pass, the same value of k is used for repeated passes as the window is moved slightly in a vertical direction prior to each pass. Admittedly, this increases the chances of signaling false clusters, but the clusters formed in this way will tend to identify entire areas of interest on the map, which is highly desirable in most applications.

The notation used in the remainder of this paper is defined as follows:

λ = the expected number of points in the window at any one time;

I = the total number of window widths necessary to traverse the map once;

k = the minimum number of points in the window at any one time that defines a cluster;

P_k = the probability that there are always fewer than k points in the window as it traverses the map once.

In the following section, the exact expression for P_k is given and some computed values are given. In later sections, an empirical approximation for P_k is introduced, and an approximation for P_k using computer simulation is described. The application of this method to data from the National Uranium Resource Evaluation Program (NURE) is described last.

THE EXACT VALUE OF P_k

An exact expression for the probability P_k of having no clusters is given by Naus[5] as

$$P_k = e^{-\lambda I} \sum_S \lambda^n \det \left| \frac{1}{c_{ij}!} \right| , \tag{1}$$

where λ and I are as defined earlier and

$$S = \{n_i\}_{i=1}^{I} \text{ such that all } n_i < k ,$$

$$n = \sum_{i=1}^{I} n_i ,$$

$$\det|\cdot| = \text{determinant of an } I \times I \text{ matrix} ,$$

$$c_{ij} = k(j-i) - \sum_{\alpha=i}^{j-1} n_\alpha + n_i , \text{ for } i < j ,$$

$$c_{ij} = k(j-i) + \sum_{\alpha=j}^{i} n_\alpha , \text{ for } i \geq j ,$$

$$\frac{1}{c_{ij}!} = 0 , \text{ if } c_{ij} < 0 \text{ or if } c_{ij} > n .$$

This formula is quite cumbersome to calculate, since there are k^I terms in the summation and theoretically each term involves the determinant of an $I \times I$ matrix.

1. Joseph I. Naus, "A Power Comparison of Two Tests on Nonrandom Clustering," *Technometrics* 8: 493–517 (1966).
2. Michikazu Kumagai, "Availability of an n-spare System with a Single Repair Facility," *IEEE Trans. Reliab.* R-24(3): 216–17 (1975).
3. R. T. Leslie, "Recurrence Times of Clusters of Poisson Points," *J. Appl. Probab.* 6: 372–88 (1969).
4. G. F. Newell, "Some Statistical Problems Encountered in a Theory of Pinning and Breakaway of Dislocations," *Q. Appl. Math.* 16: 155–68 (1958).
5. Joseph I. Naus, "Some Probabilities, Expectations, and Variances for the Size of Largest Clusters and Smallest Intervals," *J. Am. Stat. Assoc.* 61: 1191–99 (1966).

For the special case of $k = 2$, an alternative method of derivation[6] yields the form

$$P_2 = e^{-I\lambda} \sum_{i=0}^{I} \lambda^i (I + 1 - i)^i / i! \,. \tag{2}$$

A simple equation for P_3 also exists,[7] but we have not seen it. An exact expression for P_k, given in ref. 8, is not correct because of an error in the derivation. Still the formula is useful; it is discussed later.

For purposes of making comparisons the exact probabilities were computed for most combinations of k from 2 to 10 and for values of I from 3 to 6. Larger values of I were not considered because of the amount of computer time required. In each case, numerous λ values were considered so that P_k ranged from near 1.0 to small values not usually of interest. A typical set of values is presented in Table 1 and compared with the sets of approximations obtained from methods described in the next two sections. The approximation methods appear to work quite well, but more discussion of this point is deferred until later in the paper.

Table 1. Exact values of P_7 for $I = 6$ as compared with two approximate calculations

λ	Exact P_7	Approximate formula	Simulation approximation
0.83	0.999	0.999	0.999
1.00	0.998	0.998	0.998
1.17	0.994	0.995	0.994
1.33	0.988	0.989	0.988
1.50	0.978	0.979	0.977
1.67	0.963	0.964	0.962
1.83	0.940	0.942	0.944
2.00	0.911	0.913	0.910
2.17	0.873	0.876	0.880
2.33	0.827	0.829	0.829
2.50	0.773	0.773	0.771

AN EMPIRICAL APPROXIMATION OF P_k

The empirical approximation of P_k is given by

$$P_k' = \frac{3}{2} p_{k-1} \exp\left\{-\lambda(I - 1)\left(1 - \frac{p_{k-2}}{p_{k-1}}\right)\right\} - \frac{1}{2}\exp\{-\lambda I(1 - p_{k-2})\}\,, \tag{3}$$

where p_k is the cumulative Poisson probability; that is,

$$p_k = e^{-\lambda} \sum_{j=0}^{k} \lambda^j / j! \,. \tag{4}$$

The values of P_k', which are relatively easy to compute, were compared with the exact values of P_k; the agreement varied from roughly ± 0.005 for $P_k > 0.9$, and ± 0.01 for $0.9 > P_k > 0.8$, to ± 0.05 for P_k near 0.2. The approximation is worse for larger k, say $k = 10$, but does not seem to get worse as I gets large, when exact values are costly.

The approximation, Eq. (3), is the weighted average of two other approximate formulas F_1 and F_2,

$$P_k' = \frac{3}{2} F_1 - \frac{1}{2} F_2 \,, \tag{5}$$

where F_1 is the erroneous result derived in ref. 8 and F_2 is obtained as follows. The probability of obtaining a total of i points along the path of the window in one complete pass is $\exp\{-\lambda I\}(\lambda I)^i / i!$, and the probability of a given point having fewer than $k - 1$ neighbors to its right is approximately given by the Poisson probability p_{k-2}. By treating all of the events as independent (which they aren't) the approximation

$$F_2 = \sum_{i \geq 0} e^{-\lambda I} \frac{(\lambda I)^i}{i!} (p_{k-2})^i \tag{6}$$

is obtained, which converts easily into the form for F_2 defined by Eqs. (3) and (5). The particular weights used in Eq. (5) were selected because they worked for $k = 2$, $I = 10$. Further investigation indicated that results were satisfactory for all of the values of k and I investigated in this study, as reported earlier.

6. Ronald L. Iman, Personal communication, 1976.
7. M. V. Menon, "Clusters in a Poisson Process" [abstract]. *Ann. Math. Stat.* 35: 1395 (1964).
8. James M. Goodwin and Erich W. Giese, "Reliability of Spare Part Support for a Complex System with Repair," *Oper. Res.* 13: 413–23 (1965).

APPROXIMATION OF P_k BY COMPUTER SIMULATION

For our purposes the most satisfactory method of obtaining a reasonable estimate of P_k is by computer simulation. For $l = 10$, a reasonable value in our application, it is impractical to obtain the exact values of P_4, P_5, and P_6. The approximation given by Eq. (3) appears to be satisfactory for the cases examined, but without exact values to compare it with, one has no way of knowing for sure how well Eq. (3) will work for other values of l and k. So a computer simulation program was written and run. The program takes l and λ as inputs, and in a few seconds furnishes the outputs $P_2, P_3, \ldots, P_m$ for any designated integer m (we used $m = 20$) so that the choice of k can be made. Briefly, the method is as follows.

Uniform (0, 1) random variables $U_1, U_2, \ldots$, are generated using a standard computer random number generator and are transformed to exponential random numbers with mean $1/\lambda$ using $X_i = -(1/\lambda) \ln U_i$. Enough values of X are obtained so their sum barely exceeds l. If the moving sum of k consecutive X's is less than 1 anywhere along the sequence of X's, not counting the last X, then a cluster has occurred. Otherwise, no cluster has occurred. This process is repeated 10,000 times to see how many of the 10,000 repetitions result in no clusters, and this proportion is used to estimate P_k. Obviously, it is easy to keep track of several values of k at the same time.

Table 2. A comparison of the approximate formula [Eq. (3)] and the simulation approximation (S. A.) for $l = 10$ and $k = 4, 5$, and 6

λ	$k = 4$		$k = 5$		$k = 6$	
	S. A.	Eq. (3)	S. A.	Eq. (3)	S. A.	Eq. (3)
0.1	1.000	1.000	1.000	1.000	1.000	1.000
0.2	0.998	0.998	1.000	1.000	1.000	1.000
0.3	0.991	0.992	0.999	0.999	1.000	1.000
0.4	0.975	0.976	0.997	0.998	0.999	1.000
0.5	0.948	0.949	0.994	0.993	1.000	0.999
0.6	0.909	0.908	0.984	0.986	0.998	0.998
0.7	0.854	0.852	0.971	0.972	0.996	0.996
0.8	0.791	0.783	0.952	0.952	0.991	0.992
0.9	0.712	0.702	0.930	0.924	0.986	0.986
1.0	0.630	0.614	0.886	0.888	0.975	0.976
1.1	0.538	0.523	0.847	0.842	0.965	0.963
1.2	0.445	0.432	0.792	0.788	0.943	0.944
1.3	0.383	0.346	0.741	0.726	0.926	0.920
1.4	0.313	0.268	0.670	0.657	0.892	0.891
1.5	0.242	0.200	0.602	0.584	0.856	0.855

In all cases where exact values of P_k were obtained, as described earlier, an approximation by simulation was also made. The simulated value was within 0.005 of the true value more than two-thirds of the time, as one would expect. With this assurance that the method works well, simulated values were obtained for values of l up to 10 and for k up to 9. The approximation defined by Eq. (3) agreed reasonably well with the simulation value in almost all cases, as indicated in Table 2. The cases with the greatest disagreements were those where P_k was small, and those cases are generally not of interest.
The standard deviation of such an estimate is less than 0.005, and much less if P_k is close to 0 or 1.

AN APPLICATION TO NURE DATA

As part of the NURE project, airplanes have flown east-west patterns over a region near Lubbock, Texas, recording radioactivity attributable to ^{214}Bi (ref. 9). The presence of ^{214}Bi in abnormally high quantities may indicate potential for uranium mineralization. A certain amount of ^{214}Bi is present almost every where, and occasional high counts may occur even though the concentration is low, simply due to chance. Clusters of such high readings are of interest however, particularly if such clusters are unlikely to occur by chance.

The aerial reconnaissance data were analyzed by the Los Alamos Scientific Laboratory, and outliers, or anomalies, were located. Figure 1 shows 77 such anomalies in the Lubbock region. These anomalies are located on 23 "map lines," east-west flight lines, flown by radiometric reconnaissance airplanes. A window of width 0.2 degrees longitude (representing about 6 miles) and of height sufficient to include three map lines (representing about 8.5 miles) was moved horizontally across the map 21 times, for the 21 different groups of three adjacent map lines. Each time the window included five points (anomalies), that area was marked with dots. When the window contained six points, that area was shaded with diagonal lines and when the window contained seven or more points, the area was marked with cross-hatching. These results are also shown in Figure 1.

The probability P_k of obtaining no clusters in a single pass of the window was not obtained for $l = 10$

9. Geodata International, Inc., *Aerial Radiometric and Magnetic Survey of Lubbock and Plainview National Topographic Maps, NW Texas*, vol I, prepared for the U.S. Energy Research and Development Administration, Grand Junction, Colorado, office under contract number AT(05-1)-1654, 1975.

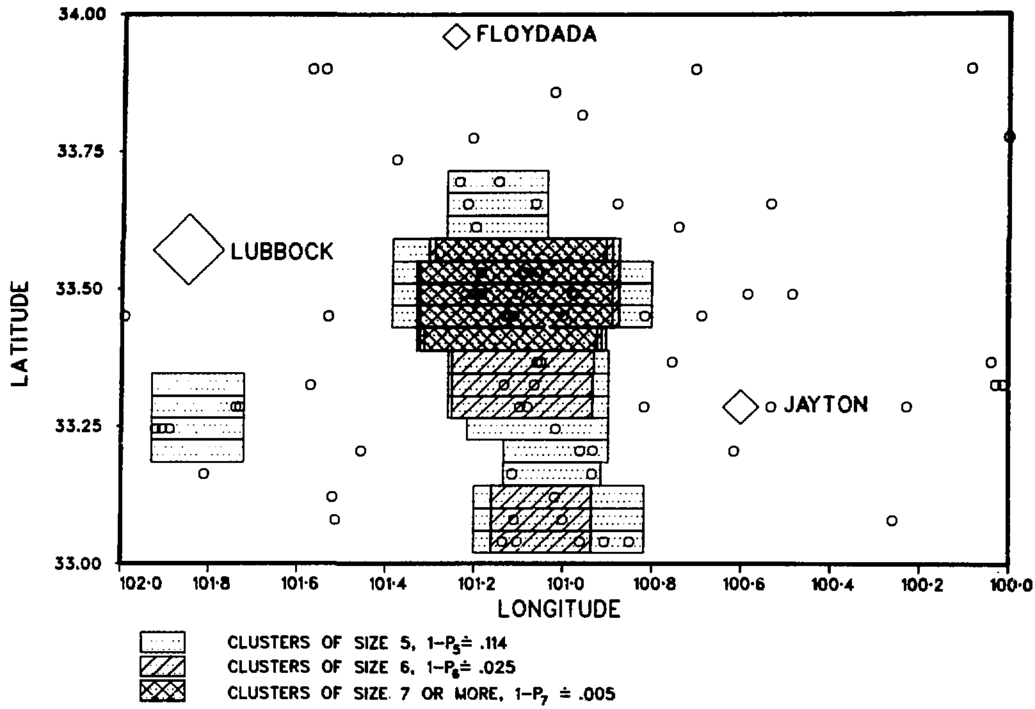


Fig. 1. Clusters of ^{214}Bi anomalies.

and $k = 5$, 6, and 7 as in this example. So the simulation program described earlier was used to estimate these probabilities. The parameter λ was estimated from the data using

$$\hat{\lambda} = \frac{xy}{zl},$$

where

x = number of points on the map,
y = number of lines in the window,
z = total number of map lines.

Thus,

$$\hat{\lambda} = \frac{(77)\,(3)}{(23)\,(10)} = 1.00\,.$$

The results where $\hat{P}_5 = 0.886$, $\hat{P}_6 = 0.975$ and $\hat{P}_7 = 0.995$ with standard deviations of 0.0032, 0.0016 and 0.0007 respectively. The estimates using Eq. (3) were in reasonable agreement, being 0.888, 0.976 and 0.996 respectively; all were within one standard deviation of the simulation results.

The results of identifying clusters in this way are interesting. Certain geographical areas stand out clearly as areas deserving further investigation using ground survey techniques.

Computer Graphics for Extracting Information from Data*†

Ronald K. Lohrding
Myrle M. Johnson
David E. Whiteman

Los Alamos Scientific Laboratory
Los Alamos, New Mexico

ABSTRACT

This paper presents computer graphics which are useful for displaying and analyzing data. Many classical and several newly developed graphical techniques in statistical data analysis are presented for small univariate and multivariate data sets. These include histograms, empirical density functions, pie charts, contour plots, a discriminant analysis display, cluster analysis, Chernoff "faces," and Andrews' sine curves.

Recent advances in data collection technology and computer data base management systems have made it imperative to utilize computer graphics for large data sets. Several innovative graphical techniques are presented to handle this situation.

Spatial relationships among the data (particularly geographic data) are difficult to conceptualize. Several cartographic techniques are presented which enhance the understanding of these spatial relationships within the data.

INTRODUCTION

The Energy Systems and Statistics Group at the Los Alamos Scientific Laboratory (LASL) is involved in several projects with energy-related data. Some of these projects have small univariate or multivariate data sets, while others have large data sets that require data management systems for efficient data manipulation. A statistically oriented graphics package is presently under development; numerous modules have been completed. The purpose of this package is to provide graphical techniques for the initial examination of the data. This paper uses data from several projects to demonstrate some of these techniques.

In the following sections, we discuss graphical methods useful for a preliminary analysis of small data sets, graphical techniques which are appropriate for large data sets, and finally, spatial relationships in geographic data sets. Throughout this paper, examples of computer graphics are used to illustrate the techniques. (The 35-mm color slides of computer-generated graphics shown at the conference are reproduced in black and white for this paper.)

PRELIMINARY DATA ANALYSIS OF SMALL UNIVARIATE AND MULTIVARIATE DATA SETS

Computer graphics for a preliminary raw data analysis may include histograms, empirical distribution function plots, and probability plots. The data used in this section were collected on 17 variables for each of the 50 states plus the District of Columbia. The variables and their means and standard deviations are listed in Table 1. Of particular interest is the average household Btu consumption per capita (HHBTU). The histogram in Fig. 1 shows that the assumption of normality may be questionable. Two graphical tests of normality are shown in Figs. 2 and 3. One test uses Lilliefors' test statistic; the other uses a test statistic developed by Lohrding. In the former, the normality assumption is tested by placing

*Work performed under the auspices of the Energy Research and Development Administration, Contract No. W-7405-Eng. 36.

†Report LA-UR-77-2456, Los Alamos Scientific Laboratory, Los Alamos, New Mexico.

Table 1. Sample statistics for 17 measurements collected on 50 states

Variable	Definition of variable	Mean	Standard deviation
1. HHBTU	Household Btu per capita (10^6)	87.33	21.30
2. DEGD	Heating degree day loads $\frac{\Sigma(65 - Y)}{365}$ °F (10^3 °F) where $Y = \begin{cases} \text{average daily temperature if } Y \leq 65°\text{F} \\ 65°\text{F if } Y > 65°\text{F} \end{cases}$	5.00	2.23
3. MAXT	Normal July maximum temperature (°F)	86.41	5.96
4. PCAIR	Percent of households with air conditioning	33.73	18.44
5. POP	1971 population (10^6)	4.04	4.36
6. FRZR	Percent of population with freezers	32.90	10.94
7. ONEP	*Single individuals per housing unit*	218.63	271.30
8. PCURB	Percent urban population	66.47	15.11
9. COML	Percent commercial sector, commercial/(residential & commercial)	36.71	3.31
10. MEDIN	Median income (10^3)	9.17	1.45
11. LOWIN	Percent of family incomes below government poverty levels	11.67	5.18
12. SINGLE	Percent of single family houses	71.72	11.19
13. NEWHS	Percent of houses built since 1960	25.91	7.92
14. OLDHS	Percent of houses built before 1950	53.42	12.34
15. AVEIN	Average income per capita (10^3)	3.96	0.63
16. LAT	Latitude of center of the state	39.48	6.44
17. LONG	Longitude of center of the state	93.59	19.50

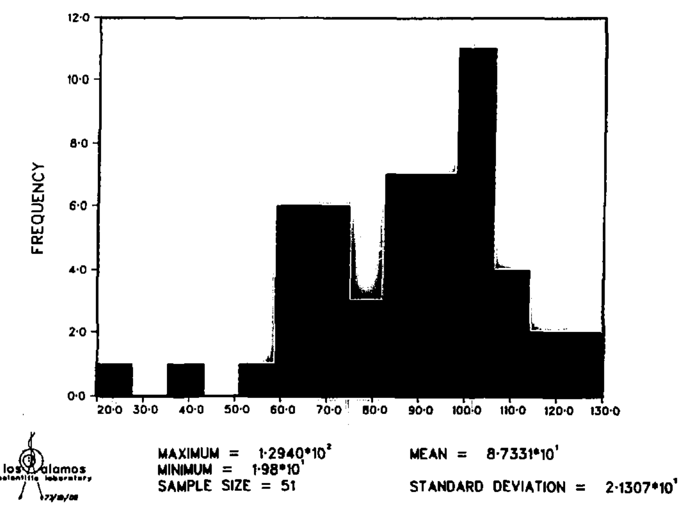


Fig. 1. Histogram of household Btu (HHBTU)—all states and Washington, D.C.

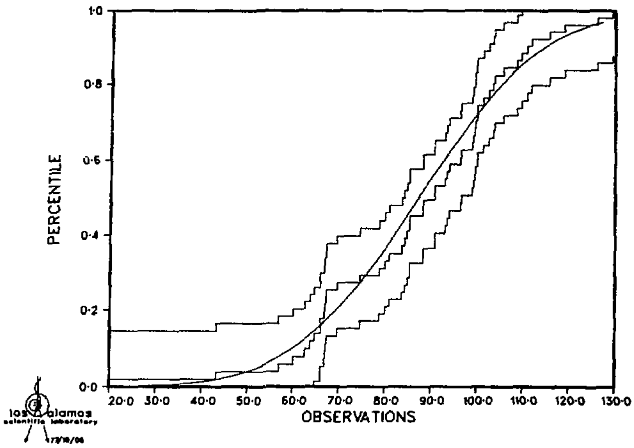


Fig. 2. Lilliefors' 95% confidence bounds on the empirical distribution function of HHBTU for all states and Washington, D.C.

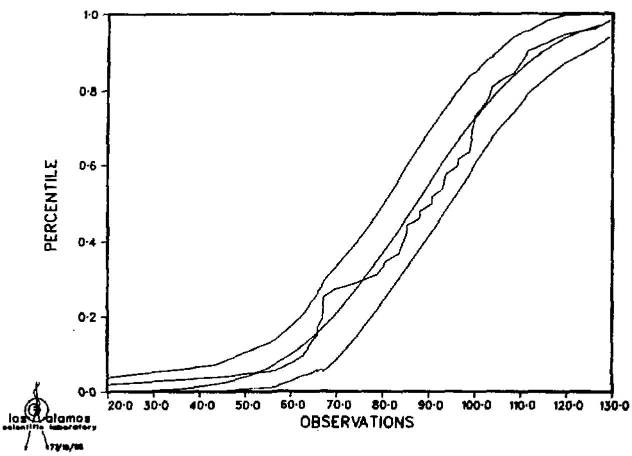


Fig. 3. Lohrding's 95% confidence bounds on the empirical distribution function of HHBTU for all states and Washington, D.C.

$(1 - \alpha)$ 100% confidence bounds on the empirical distribution function (edf). The normal cumulative distribution function (cdf) with mean and variance estimated by the sample mean and sample variance is plotted. If the cdf falls outside the bounds placed on the edf, the assumption of normality is rejected at the α level of significance. In the latter, the normality assumption is tested by placing $(1 - \alpha)$ 100% confidence bounds on the normal cdf with mean and variance estimated by the sample mean and sample variance. If the edf falls outside the bounds placed on the cdf, the assumption of normality is rejected at the α level of significance. In neither test is normality rejected at the 95% level of significance. A normal probability plot and a lognormal probability plot, two additional graphical techiques which may give further insight to the structure of the data, are given in Figs. 4 and 5.

To describe the joint relationship of HHBTU to 26 other variables (including transformations of some of the variables), a linear multiple stepwise regression procedure is used. Seventy-five percent of the vari-

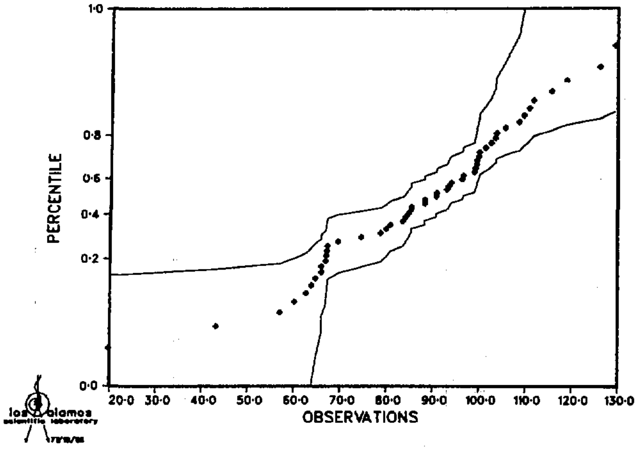


Fig. 4. Normal probability plot of HHBTU for all states and Washington, D.C.

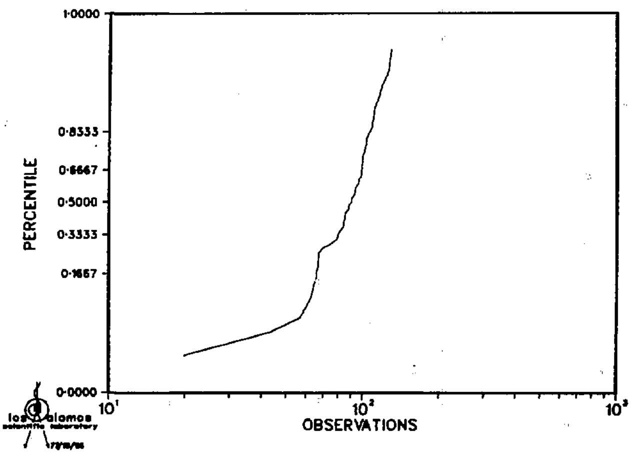


Fig. 5. Lognormal probability plot of HHBTU for all states and Washington, D.C.

ance is accounted for by the variables degree days (DEGD) and percent urban population (PCURB). The equation of the fitted linear multiple regression model is

$$Y_i = 22.155 + 8.657X_{2,i} + 0.328X_{8,i} \,,$$

where

$$i = 1, 2, \ldots, 51 \,,$$

Y_i = HHBTU for the ith state (z axis) ,

$X_{2,i}$ = DEGD for the ith state (x axis) ,

$X_{8,i}$ = PCURB for the ith state (y axis) .

Figure 6 shows a three-dimensional graphical representation where the fitted plane and the data points are plotted. Lines are drawn from the data points to the surface to give some indication of the deviations. In a nonlinear regression analysis, the equation of the fitted model is

$$Y_i = 33.835 - 77.607\left(\frac{X_{2,i}^2}{X_{3,i}}\right) + 1374.90\left(\frac{X_{2,i}}{X_{3,i}}\right) ,$$

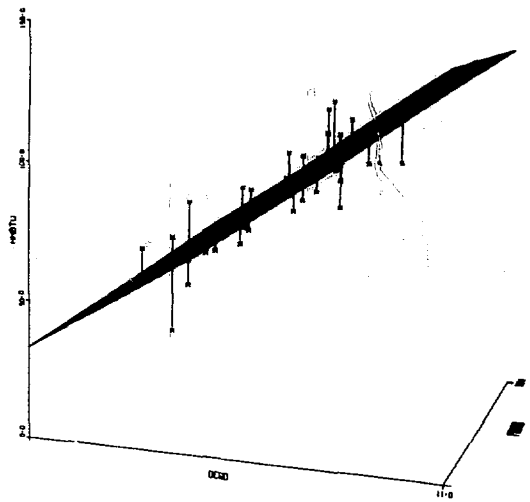


Fig. 6. 3-D representation of a linear multiple regression HHBTU model.

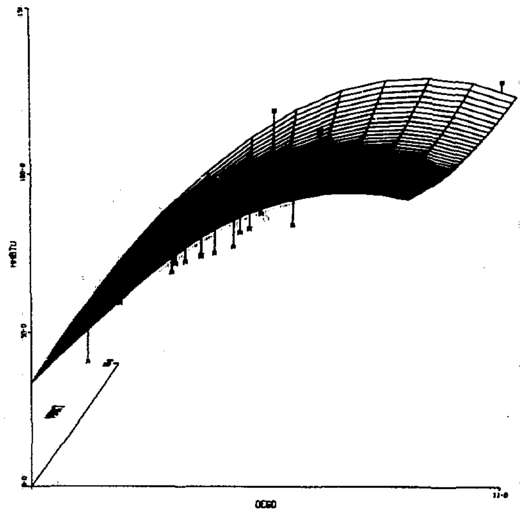


Fig. 7. 3-D representation of a nonlinear multiple regression HHBTU model.

where

$i = 1, 2, \ldots, 51$,

Y_i = HHBTU for the ith state (z axis) ,

$X_{2,i}$ = DEGD for the ith state (x axis) ,

$X_{3,i}$ = MAXT (maximum temperature) for the ith state (y axis) .

The fit of the data to the surface is shown in Fig. 7. The two extreme points are Alaska and Hawaii.

Several techniques are available for displaying multivariate data. We first discuss a gray-level coded correlation matrix which displays the pair-wise correlations between variables. The gray-level scale ranges from positive to zero to negative correlations. Frequently, such a display may be useful in directing attention to interesting variable relationships. In Fig. 8, note that HHBTU is positively correlated with DEGD, LAT, OLDHS, MEDIN, and AVEIN; negatively correlated with MAXT, PCAIR, LOWIN, SINGLE, and NEWHS; and not correlated with POP, FRZR, ONEP, PCURB, COML, and LONG.

Another technique called Andrews' sine curves uses the standardized data as coefficients of a function involving sines and cosines of t in the range $(-\pi, \pi)$. A function involving the 17 variables was plotted for each of the 50 states plus the District of Columbia to visually cluster similar states. Relatively tight bands suggest clusters. When the original data are used, it is very difficult to separate clusters, as shown in Fig. 9. However, a plot of the factor coefficients from a principal components analysis in Fig. 10 shows three possible clusters of states.

Figure 11 shows the so-called Chernoff faces for the 50 states plus the District of Columbia. Here, a facial characteristic is associated with a variable as indicated in Table 2. For example, wide noses correspond to large single populations and long noses correspond to large populations. The faces for New York and California are striking because of this feature. Similarly, Alaska has a wide face because of the large HHBTU consumption per capita, whereas Hawaii has a thin face.

The dendogram, a tree-like graph of nonoverlapping hierarchical partitions, is another visual technique used in cluster analysis. A computer program containing eight clustering techniques (nearest neighbor, furthest neighbor, simple average, group average, median, centroid, Lance and Williams' flexible strategy, and Ward's method) is used. Initially, the data are standardized; both classical and robust standardization techniques are used. When the data are standardized by the sample mean and sample standard deviation, no noticeable peculiarities in the data structure are present. However, when the data are standardized by the trimmed mean and trimmed standard deviation, four states (New York, California, Alaska and Hawaii) are distinct from the main cluster regardless of the algorithm used.

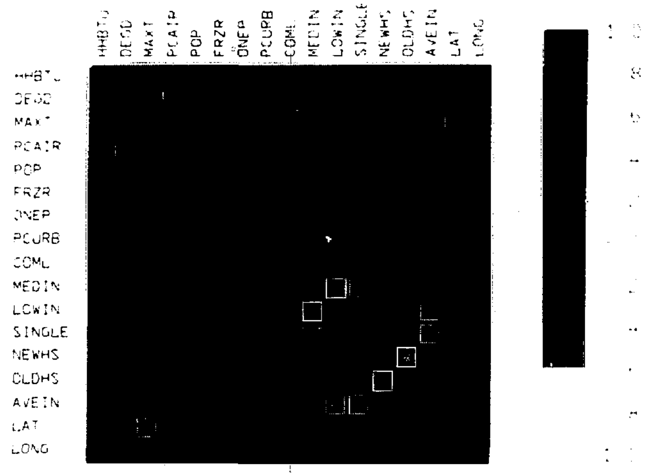


Fig. 8. Gray-level coded correlation matrix of HHBTU data.

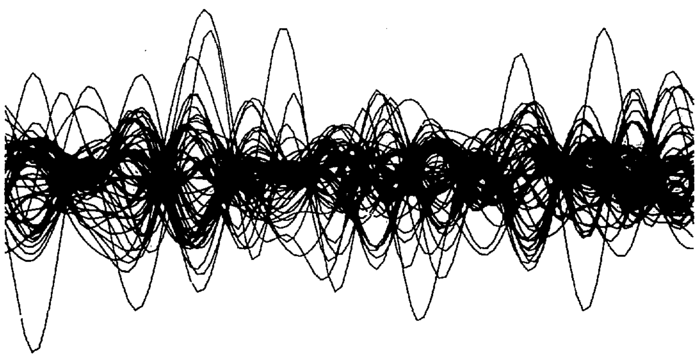


Fig. 9. Andrews' sine curves of HHBTU data.

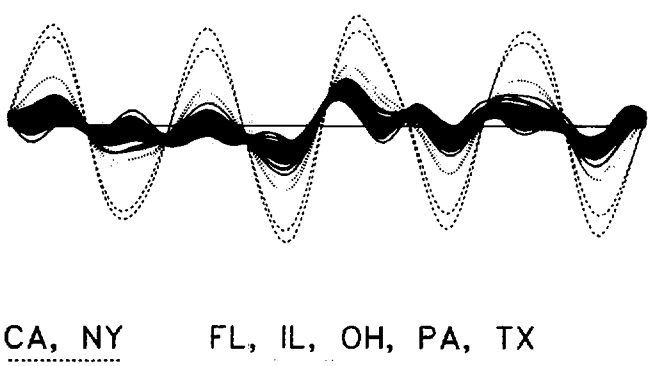
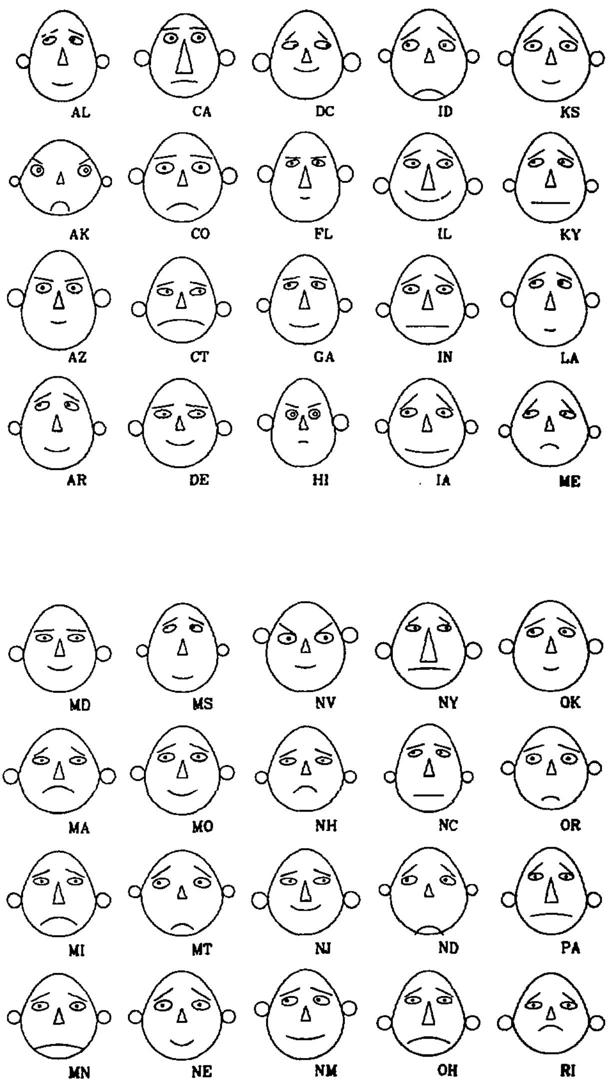


Fig. 10. Andrews' sine curves of the factor coefficients from a principal components analysis.



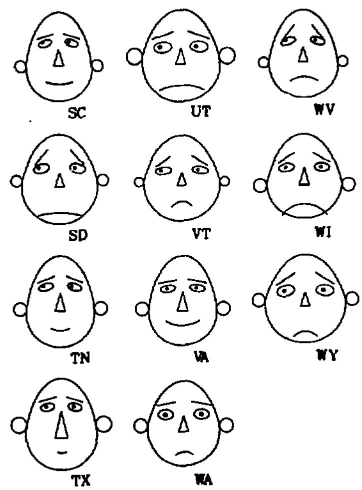


Fig. 11. Chernoff faces of HHBTU data for all states and Washington, D.C.

Table 2. Variables associated with specific facial characteristic for Chernoff faces representing all 50 states and the District of Columbia

Facial characteristic	Variable
1. Face width	HHBTU
2. Brow length	SINGLE
3. Face height	MAXT
4. Eye separation	LAT
5. Pupil position	AVEIN
6. Nose length	POP
7. Nose width	ONEP
8. Ear diameter	PCURB
9. Ear level	COML
10. Mouth length	DEGD
11. Eye slant	MEDIN
12. Mouth curvature	PCAIR
13. Mouth level	FRZR
14. Eye level	LOWIN
15. Brow height	OLDHS
16. Eye eccentricity	LONG
17. Eyebrow angle	NEWHS

LARGE DATA SETS

In data analysis many of the ensuing problems can be attributed to the data itself—perhaps inaccurate, missing, too little, and recently too much. These large data sets not only create a tremendous storage problem, but challenge computer graphics for effective display techniques.

The analyses considered here deal with National Uranium Resource Evaluation (NURE) data. The objective of this nationwide airborne and stream sediment reconnaissance survey is to classify regions with respect to their potential mineralization. For example, in the stream sediment survey, LASL analyzes the data from five states: Wyoming, Colorado, Montana, New Mexico, and Alaska. In the second year of a five-year study, LASL data bases already contain seven million words.

The probability distributions of certain random variables such as thallium signals over a given geological formation of a flight line are thought to indicate uranium concentration. A technique for computing an empirical density function (edf) used to estimate a probability density function has been developed. As many as 100 of these densities, each representing a map line or transect, can be displayed simultaneously as shown in Fig. 12. Since some of the edf's may be visually obscured by other edf's, the 3-D plots have been compressed into a 2-D grid plane in a lightness-darkness plot shown in Fig. 13.

Figure 14 is a scattergram of bismuth vs thallium for all geological formations on one map line in the Lubbock-Plainview area in Texas. The data in the lower left-hand corner represent recent geological formations and most of the formations follow a linear trend except for the data on the right-hand side of the plot where thallium becomes constant with bismuth increasing. These data belong to two older formations with known uranium mineralization. Figure 15 shows data for one geologic formation. Scattergrams such as this one are useful in identifying clusters representing misclassified geological formations data.

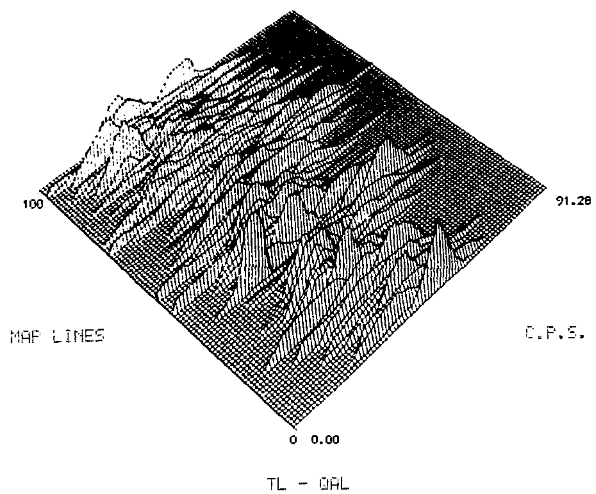


Fig. 12. 3-D plot of empirical density functions.

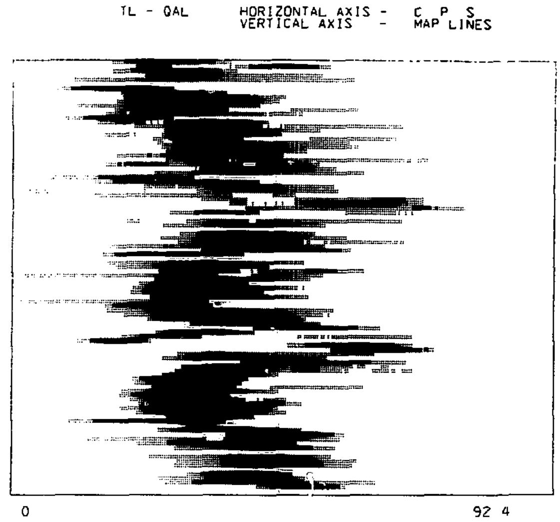


Fig. 13. Lightness-darkness plot.

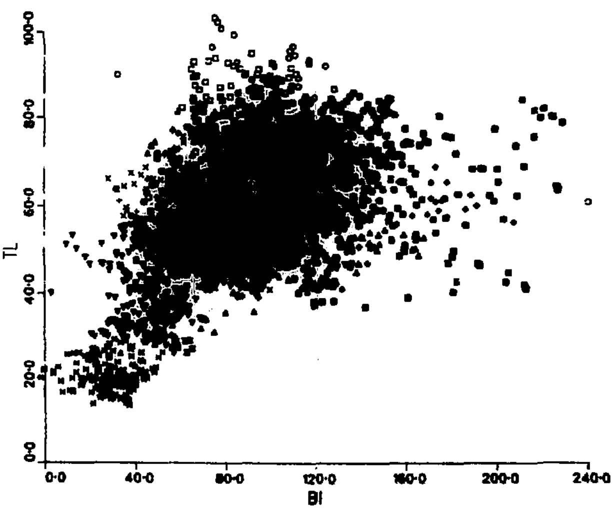


Fig. 14. Scattergram of bismuth vs thallium for all geologic formations, Lubbock-Plainview area.

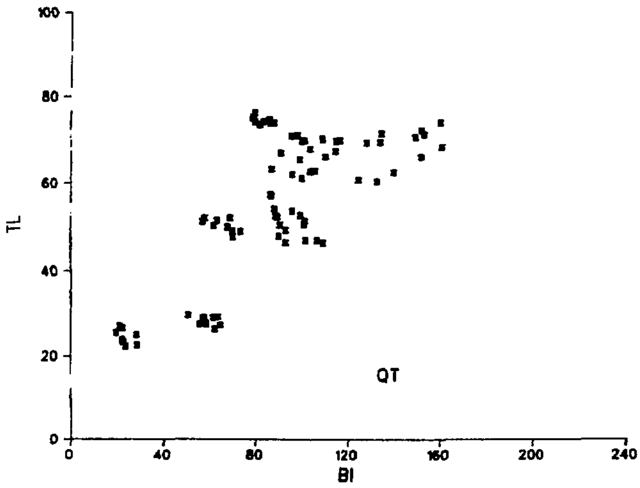


Fig. 15. Scattergram of bismuth vs thallium for one mapline, Lubbock-Plainview area.

Figure 16 shows a linear discriminant analysis displayed as a gray-level matrix useful in delineating between favorable and unfavorable regions of uranium mineralization. Each square represents 100 records (i.e., 100 sec of gamma-ray signals on a map line) in the Lubbock area. The 23 rows represent 23 map lines. There are eight gray levels which are linearly spaced from light to dark over the interval [0, 1]. The lighter shades represent low probability of favorable uranium mineralization while darker shades represent high probability of favorable uranium mineralization.

Contour maps of the Lubbock-Plainview area also indicate regions where the probability of finding uranium is high. An example is shown in Fig. 17.

CARTOGRAPHIC DATA SETS

Maps are very useful in displaying and communicating information contained in data with spatial/geographic relationships. The figures shown are applications of cartographic techniques and have been extracted from various on-going projects.

Figure 18 summarizes U.S. offshore oil and gas lease data from October 1954 through November 1976. The number of leases, the leasing years, the acreage and the producing acres through 1974 are given for individual states and regions as well as totals for all the leases. Of the total 1,940,000 producing acres, Louisiana has 1,824,000 acres and Texas has 103,000 acres.

Figures 19–21 are for a study of the impacts of electric power generation in the West. The location of existing and proposed power plants by type for the Western and Rocky Mountain regions are shown in Fig. 19. The letters represent the type of plant, that is, coal, oil, gas, and nuclear. The size of the letters indicate three levels of power generation: small, 500-999 MWe; medium, 1000–1999 MWe; and large,

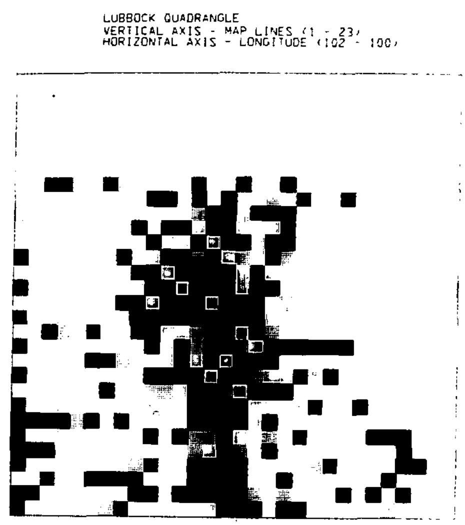


Fig. 16. Linear discriminant analysis display.

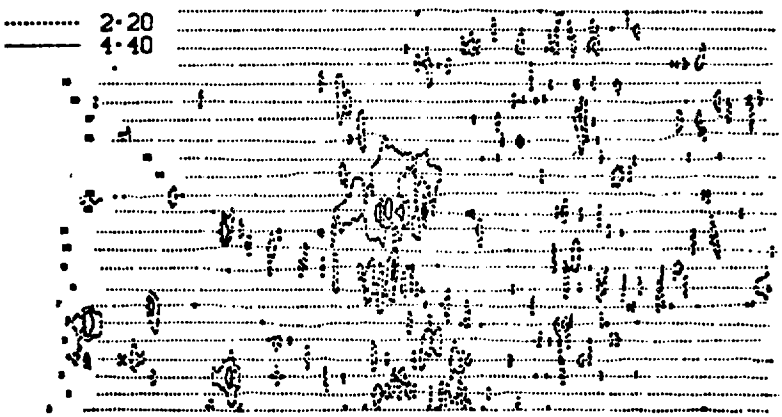


Fig. 17. Contour map of bismuth-thallium ratio.

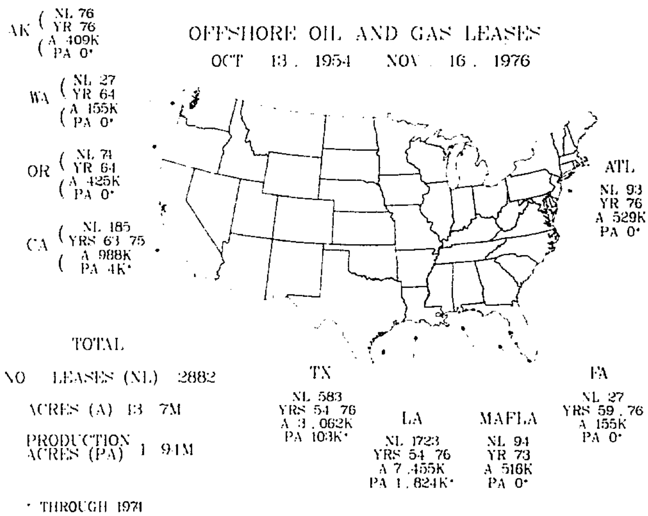


Fig. 18. Offshore oil and gas lease map.

2000' MWe. The Los Angeles and San Francisco areas have a number of oil-fired plants, and these areas are simply shaded. Figures 20 and 21 are maps to study pollution dispersion patterns. Figure 20 shows SO_2 concentration in southwest Wyoming for 1985 with pollution contours drawn every 0.25 $\mu g\ m^3$. Figure 21 shows change in length of life due to pollution in days per person. Similar graphical displays were done for exposure to suspended particulates, additional restricted activity days due to pollution and annual morbidity costs per person and per town.

Figures 22 24 are from solar feasibility studies. The first map shows heating degree days which is the average of the high and low temperatures subtracted from a 65°F base temperature for the 48 contiguous states. Simply, the colder the climate, the higher the number of heating degree days. Contrast Florida with 214 and Maine with 7511. The second map shows 1977 residential gas prices in dollars per

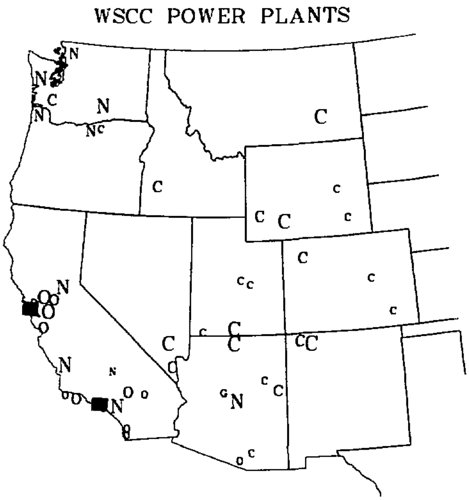


Fig. 19. Map of power plants in the West.

thousand cubic feet by state. Gas is generally cheaper in the southern, central, and Rocky Mountain regions. Note that Maine has higher prices than nearby Vermont and New Hampshire. Figure 24 shows the pattern of economic feasibility for domestic hot water under incentives provided by the National Energy Plan of April 1977 and the House Modification of that plan.

Figures 25-32 are maps displaying energy-related data from the regional studies program. Figure 25 shows the five coal export-import regions, and Fig. 26 is a flow map for the export of Rocky Mountain coal. The circle represents the within region total, and the thickness of the arrows represents relative amounts of export to the other four regions. Bar charts and pie charts are useful in displaying energy totals for regions or states. Production, consumption, export, and negative export (import) figures are displayed in Figs. 29 and 30 using shaded bars. Figure 29 uses varying sized circles to indicate production levels by region.

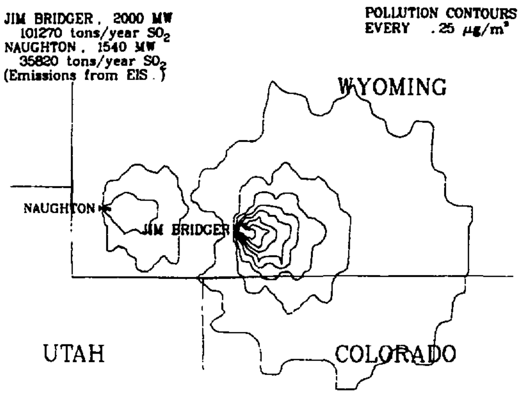


Fig. 20. SO_2 concentration in southwest Wyoming, 1985.

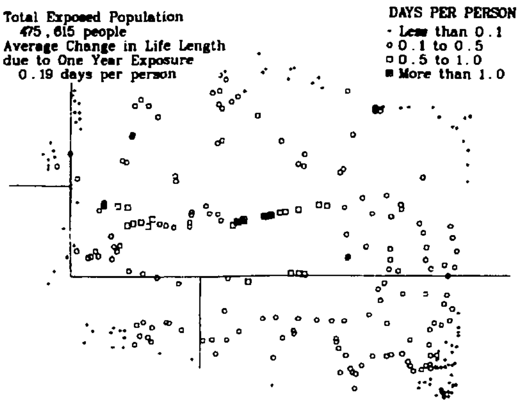


Fig. 21. Change in length of life.

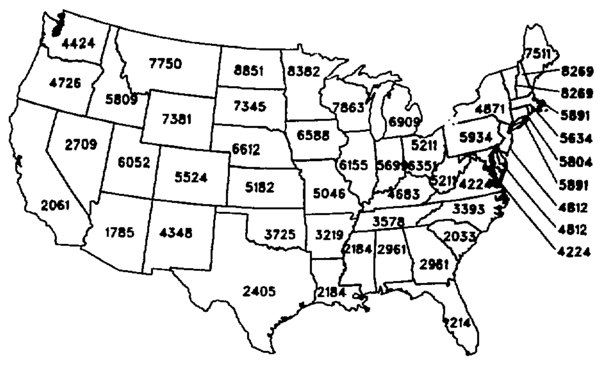


Fig. 22. Heating degree days.

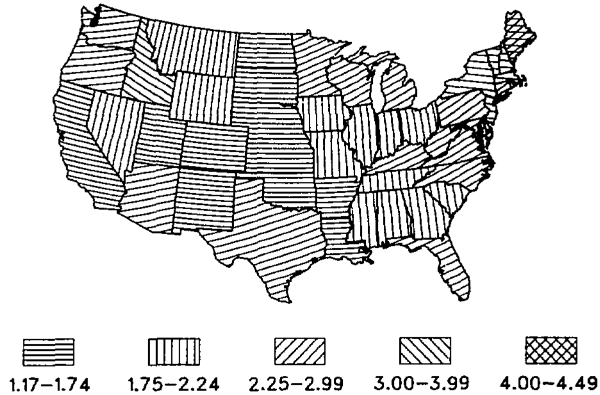


Fig. 23. 1977 residential gas prices, dollars per mcf.

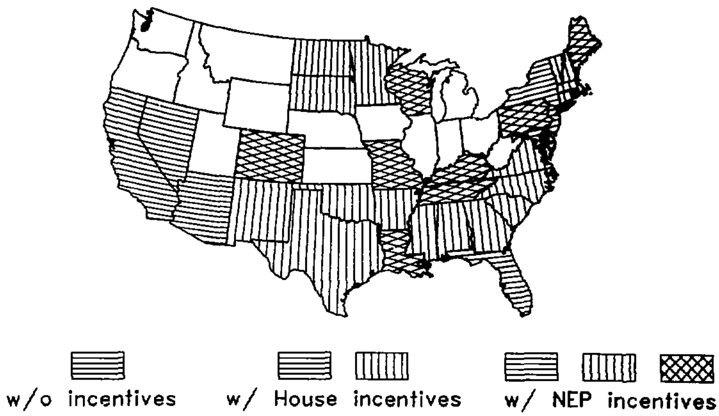


Fig. 24. Solar feasibility—domestic hot water. Alternative system electric resistance; 10-year life cycle costing.

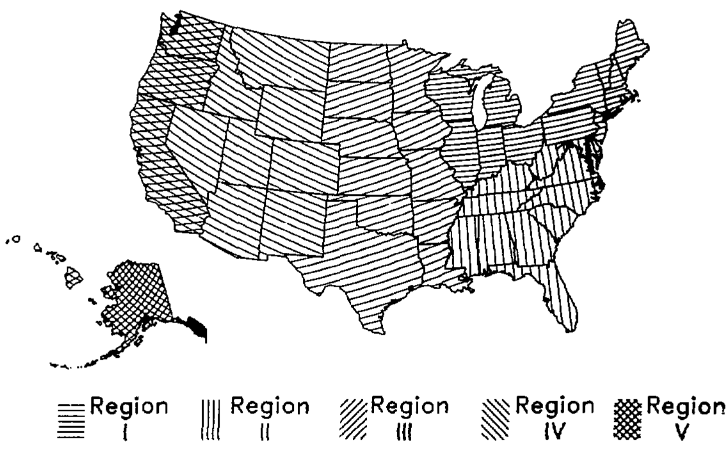


Fig. 25. Five coal export-import regions.

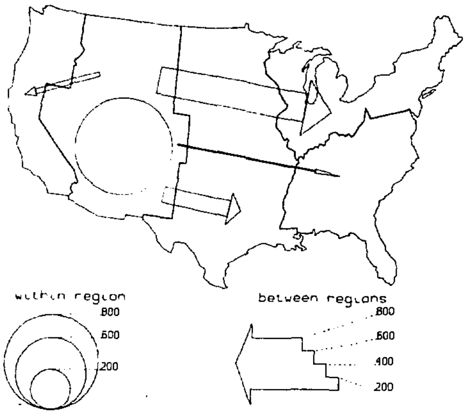


Fig. 26. Flow map of regional coal.

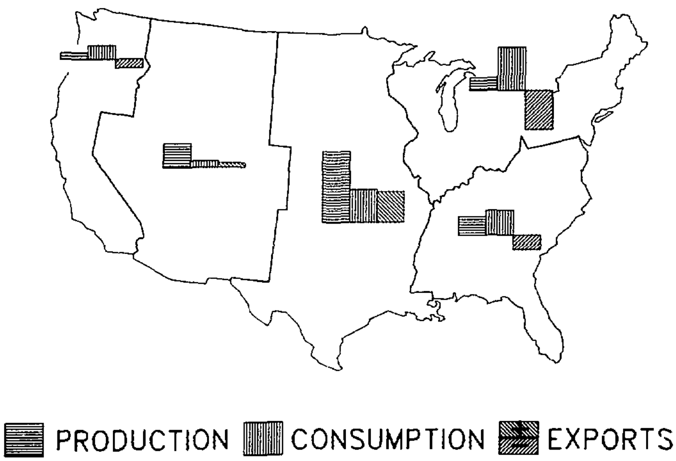


Fig. 27. 1975 regional energy totals.

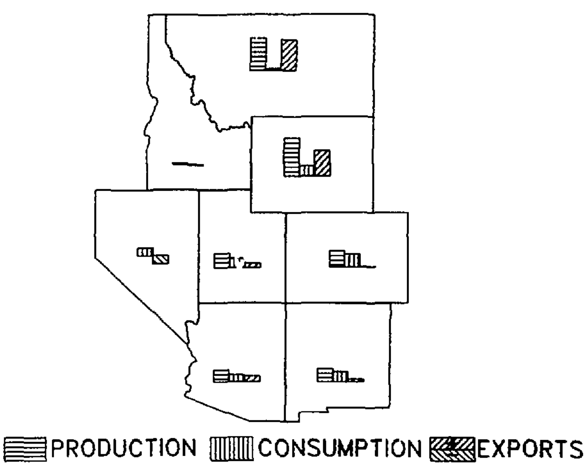


Fig. 28. Rocky Mountain coal.

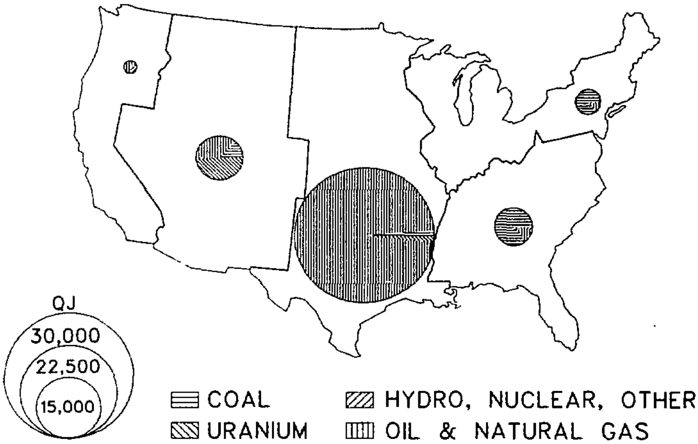


Fig. 29. 1975 regional energy production.

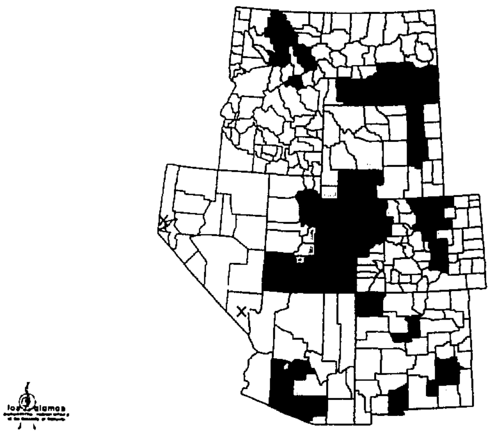


Fig. 30. County air quality maintenance area data map—Rocky Mountain states.

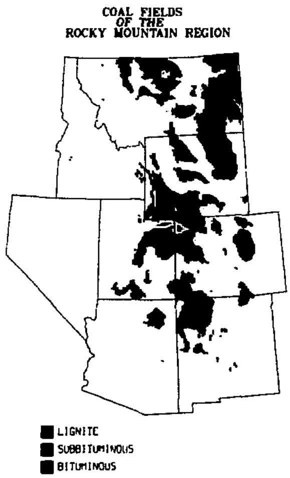


Fig. 31. Coal fields of the Rocky Mountain region.

Sections of a circle are shaded differently to indicate coal; oil and natural gas; hydro, nuclear, and other; and uranium production. Figure 30 shows county air quality maintenance area data for the Rocky Mountain region. An interactive composite geo-information mapping system known as GMAPS provides map data on such items as wilderness areas, ecosystem trends, locations of natural resources, etc., for selected regions in the United States. Figure 31 shows types of coal fields, and Fig. 32 is a composite of coal fields with oil shale basins in the Rocky Mountain region.

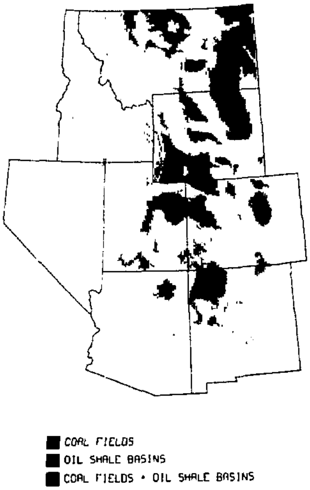


Fig. 32. Coal fields and oil shale basins.

SUMMARY

The computer-generated graphic products described in this paper represent a variety of techniques for displaying and analyzing small univariate and multivariate data sets, large data sets, and cartographic data sets. Computer graphics are useful tools for communicating information efficiently and effectively.

ACKNOWLEDGMENTS

We wish to thank Katherine Campbell and Mona Wecksung for the use of some of their slides, and Thomas Bement for assisting us on the cluster analysis. Melvin Prueitt supplied the 3-D plotting program, PICTURE, and Richard Bagley made the black and white hardcopy prints that are reproduced in this paper.

BIBLIOGRAPHY

D. F. Andrews, "Plots of High Dimensional Data," *Biometrics* **28**: 125–36 (March 1972).

T. R. Bement, D. V. Susco, D. E. Whiteman, and R. K. Zeigler, *National Uranium Resource Evaluation Program*, Los Alamos Scientific Laboratory Report LA-6804-PR, Los Alamos, N. Mex., May 1977.

L. A. Bruckner, M. M. Johnson, and G. L. Tietjen, "The Analysis of Lease, Production, and Revenue Data from Offshore Oil and Gas Leases," pp. 184–197 in *Proceedings of the Second ERDA Statistical Symposium*, [Oak Ridge, Tenn., October 1976], Los Alamos Scientific Laboratory Report LA-6758-C, Los Alamos, N. Mex., April 1977.

K. Campbell, "IMGPROC—An Image Processing Program for CDC 6600 and 7600 Computers," Los Alamos, N. Mex., November 1974.

H. Chernoff, "The Use of Faces to Represent Points in K-Dimensional Space Graphically," *J. Am. Stat. Assoc.* **68**: 361–68 (June 1973).

W. J. Conover, *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York, 1974.

A. Ford and H. W. Lorber, *Methodology for the Analysis of the Impacts of Electric Power Production in the West*, Los Alamos Scientific Laboratory Report LA-6720-PR, Los Alamos, N. Mex., January 1977.

R. Gnanadesikan and J. R. Kettenring, "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data," *Biometrics*, **28**: 81–124 (March 1972).

G. N. Lance and W. T. Williams, "A General Theory of Classificatory Sorting Strategies. I. Hierarchical Systems," *Comput. J.* **9**: 373–80 (1967).

R. K. Lohrding, "Statistical Analysis and Display of Energy-Related Data," pp. 327–55 in *Proceedings, Conference on Computer Support of Environmental Science and Analysis*, [Albuquerque, N. Mex., July 1975], NTIS, Springfield, Va., 1975.

R. K. Lohrding, *Comparative Power Studies of Some Tests of Normality*, Los Alamos Scientific Laboratory Report LA-5101-MS, Los Alamos, N. Mex., November 1972.

R. A. Waller, E. A. Monash, and J. Lohrenz, *Some Computerized Graphic Technical Applications for Federal Mineral Lease Management Support*, Los Alamos Scientific Laboratory paper presented at the First Computer Symposium, Reston, Va., March 1977.

J. H. Ward, Jr., "Hierarchical Grouping to Optimize an Objective Function," *J. Am. Stat. Assoc.* **58**(301): 236–44 (1963).

M. Wecksung, R. Wiley, and K. Turner, "GMAPS User's Manual," Los Alamos Scientific Laboratory.

J. W. Wood, "A Computer Program for Hierarchical Cluster Analysis," *Newsl. Comput. Archaeol.* **9**(4): 1–11 (June 1974).

Estimation of a Model for Electric Utility Demand in the Presence of Missing Observations*

P. M. Robinson

Harvard University
Cambridge, Massachusetts

ABSTRACT

The estimation of models for demand of coal and oil by electric utilities is complicated by the presence of many zero observations, which seem to preclude the use of standard methods, such as regression. Also, the time series characteristics of the data render inappropriate the estimates of "limited dependent variables" models, recently proposed by econometricians. Thus we suggest estimates of open-loop time series models, involving autoregressive structure in the dependent variable or the residual. The maximum likelihood and related methods we consider may prove computationally too onerous, so somewhat simpler types of estimate are proposed also. Applications to the data are described.

INTRODUCTION

This paper was motivated by some monthly time series of demand for, and price of, coal and oil to electric utilities in the United States. Series were available for each of the 50 states plus the District of Columbia, for each of the nine census regions, and for the nation. Coal and oil were classified by percentage of sulfur content. The time period covered was January 1974 to August 1976 although some series contained later observations, up to December 1976.

One is interested in modeling these data and using the model for forecasting. On the one hand a multiple equation econometric model might be constructed. This equation might treat current and lagged values of demand as the dependent variables; current and lagged values of price, along with other variables (such as scrub cost, Environmental Protection Agency standards, etc.), might be treated as the predetermined variables. Alternatively, the price series might be included among the dependent variables. Another approach would involve models of a much simpler form, involving only one or two variables. For example, one might attempt to model demand in terms of its own past history, possibly by means of an autoregressive moving average model. Such "closed loop" models, containing no predetermined variables, are easier to use in forecasting than the "open-loop" econometric models just described, for the latter require forecasts of the predetermined variables to forecast the dependent variables. Moreover, recent evidence suggests that the closed-loop models, despite their simplicity, tend to provide the better forecasts.

Whichever of these approaches is adopted, there are difficult decisions about model specification to be made. However, if the resulting model is one of several standard forms, a number of estimation procedures, many of which are available in computer packages, can be used. Unfortunately, the data in question do not lend

*This research was supported by NSF Grant SOC75-13436. The author is grateful to Data Resources Incorporated of Lexington, Massachusetts, for providing the data and to M. C. Ferrara for preparing it for use. The computations were carried out on the Massachusetts Institute of Technology's IBM 370/168 computer.

themselves to a model of "standard form," because observed demand is sometimes zero, possibly because of differences between spot and contract prices. To see what problems this causes, consider the simple regression model

$$y_t^* = \sum_{i=1}^{q} b_i z_{it} + x_t \,, \qquad t = 1, 2, \ldots,$$

where y_t^* is the dependent variable (e.g., demand), the z_{it} are predetermined (e.g., price), x_t is an unobserved residual, and the b_i are unobserved parameters, which one wishes to estimate. In the standard regression theory, x_t is a normal random variable, so in theory y_t^* is capable of taking negative values even if the z_{it} are not. The fact that observed demand cannot be negative does not really disqualify it from being y_t^*, because if $\Sigma_{i=1}^{q} b_i z_{it}$ tends to be large and positive, Prob($y_t^* < 0$) will be small. However, if y_t^* is sometimes or often *exactly* zero, it seems that it should not be modeled as a continuous random variable. Rather, some positive probability weight should be assigned to the event Prob($y_t^* = 0$), while allowing y_t to vary continuously over the positive real line.

One way of solving the problem is to introduce an underlying variable y_t such that

$$y_t = \sum_{i=1}^{q} b_i z_{it} + x_t \,, \qquad t = 1, 2, \ldots, \tag{1}$$

where

$$y_t^* = y_t \,, \qquad \text{if } y_t > 0 \,, \tag{2}$$

$$y_t^* = 0 \,, \qquad \text{if } y_t \leq 0 \,. \tag{3}$$

Thus we say that y_t has been "censored." The set-up, Eqs. (1)–(3), is often known as a "Tobit" model;[1] y_t is called a "limited dependent variable." The statistical methodology of such models is now quite well developed,[2] and Amemiya[3] has extended Eqs. (1)–(3) to multivariate regressions and simultaneous equations models, wherein y_t^* is a vector. The estimates suggested require more complicated computations than does standard multiple regression, but they are usually quite feasible on modern-day computers, at least when y_t is scalar.

Unfortunately, this work seems only of limited relevance to our data because it is motivated to deal mainly with cross-sectional data—in particular, large microdata sets on families or firms. For such data the assumption that the x_t are uncorrelated over t is usually reasonable. This assumption seems essential for the estimates suggested by Tobin, Amemiya, and others to have desirable properties. However, in applying Eqs. (1)–(3) to our time series, we would take t to represent time. The assumption that x_t is uncorrelated over time is no more reasonable than the assumption that the observed variables are correlated over time, unless the z_{it} manage to account for all the serial correlation in y_t. Because a model can seldom be perfectly specified, there are usually predetermined variables that should have been included in Eq. (1) but weren't. These go to make up x_t, so if they are serially correlated, we would expect the same of x_t. A typical model for x_t in Eq. (1) would be the autoregression

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + e_t \,, \tag{4}$$

1. J. Tobin, "Estimation of Relationships for Limited Dependent Variables," *Econometrica* **26**: 24–36 (1958).

2. T. Amemiya, "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica* **41**: 997–1016 (1973).

3. T. Amemiya, "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," *Econometrica* **43**: 999–1012 (1974).

where the e_t are serially independent with zero mean and

$$1 - \sum_{j=1}^{p} a_j s^j \neq 0 , \qquad |s| \leqslant 1 . \tag{5}$$

A closely related modification of Eq. (1) would be the dynamic model

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + \sum_{i=1}^{q} b_i z_{it} + e_t , \tag{6}$$

where we again assume Eq. (5). Chern[4] considers a model of this form for the demand for electricity. Note that Eqs. (1) and (4) can also be written in a similar way to Eq. (6):

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + \sum_{i=1}^{q} b_i \left(z_{it} - \sum_{j=1}^{q} a_j z_{i,t-j} \right) + e_t . \tag{7}$$

The right-hand sides of Eqs. (6) and (7) contain variables that are themselves subject to censoring, so the theory in Tobin[1] and Amemiya[2] does not apply.

Because they do not use information on the dependence of the variables over time, estimates based on the incorrect assumption that x_t is serially independent will not be asymptotically efficient. Moreover, it appears that they may not be consistent. When there is no censoring, ordinary least-squares (LS) estimates, although inefficient, are usually consistent, because the limit as $T \to \infty$ of the quadratic objective function will still be at the true parameter point. But in the case of censored data, the likelihoods assumed in Tobin[1] and Amemiya[2] are products of multivariate normal density functions and univariate normal probability integrals. It is not at all clear that the value maximizing this function asymptotically will be identical to the one maximizing the "true" likelihood, which as we see below, involves one or more *multivariate* normal integrals.

In the next three sections, we propose parameter estimates for models such as Eqs. (6) and (7). (In principle, it would be possible to extend our work to multivariate models of the type mentioned earlier, although the complications are then even greater than the ones we encounter here.) The last section is an application to our data set.

MAXIMUM LIKELIHOOD ESTIMATES

Let y_t^* be recorded for $t = 1, \ldots, T$. Let τ be the set of these t values for which Eq. (2) occurs, and let $\bar{\tau}$ be the set of t values for which Eq. (3) occurs.

Denote by $\Phi_d(\mathbf{z}; \mathbf{m}, S)$, the distribution function (d.f.) of the d-dimensional normal distribution, with mean vector $\mathbf{m}$ and covariance matrix S. Let y_t be generated by

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + w_t , \tag{8}$$

in which $\mathbf{w} = (w_1, \ldots, w_T)'$ has d.f. $\Phi_T(\mathbf{w}; \mathbf{f}, \sigma^2 I_T)$, where

$$\mathbf{f} = [f_1(\theta), \ldots, f_T(\theta)]'$$

4. W. S. Chern, "Estimating Industrial Demand for Electricity: Methodology and Empirical Evidence," pp. 103-20 in *Energy: Mathematics and Models*, ed. by F. S. Roberts, SIAM, Philadelphia, 1976.

and I_T is the T-rowed unit matrix. The $f_t(\boldsymbol{\theta})$ are known functions of t and of predetermined variables (reference to which is suppressed) and a vector of unknown parameters $\boldsymbol{\theta}$, of which $a_1, \ldots, a_p$, but not σ^2, are a subset. For example,

$$f_t(\boldsymbol{\theta}) = \sum_{i=1}^{q} b_i z_{it} \tag{9}$$

as in Eq. (6), or

$$f_t(\boldsymbol{\theta}) = \sum_{i=1}^{q} b_i\left(z_{it} - \sum_{j=1}^{p} a_j z_{i,t-j}\right) \tag{10}$$

as in Eq. (7). Our work applies aiso to cases where $f_t(\boldsymbol{\theta})$ is constant over t, so that we have a closed-loop model, possibly with nonzero mean, of the type already discussed.

The d.f. of $\mathbf{y} = (y_1, \ldots, y_T)'$, conditional on the predetermined variables and with $y_{1-p} = \ldots = y_0 = 0$, is

$$\Phi_T\left(\mathbf{y};\ P^{-1}\mathbf{f},\ \sigma^2(P'P)^{-1}\right),$$

where

$$P = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ -a_1 & 1 & & & & \\ \vdots & \cdot & \ddots & & & \vdots \\ -a_p & & & \ddots & & 0 \\ 0 & -a_p & & \cdots & -a_1 & 1 \end{bmatrix} \tag{11}$$

The likelihood is then

$$\underset{y_t<0,\, t\in\bar{\tau}}{\int\cdots\int} d\Phi_T\left(\mathbf{y};\ P^{-1}\mathbf{f},\ (P'P)^{-1}\right). \tag{12}$$

This equation can be written as the product of the joint probability density of the T_1 (say) y_t, $t \in \tau$, and the joint d.f. of the $T_2 = T - T_1$, y_t at 0, $t \in \bar{\tau}$, conditional on the y_t, $t \in \tau$. Evaluation of Eq. (12), or of an iterative step in the solution of the first-order conditions for a maximum, thus seems to require numerical evaluation of a multiple integral of dimension T_2.

Actually the autoregressive structure of Eq. (8) may lead to a substantial simplifying of these computations. From Eq. (11), it follows that $P'P$ has a "band" form, with the (i,j)th element zero for $|i-j| > p$. In other words the partial autocorrelations of observations more than p units apart are zero. Wecker[5] has considered prediction and estimation for closed-loop autoregressions in which censored observations are spaced more than p units apart. To be somewhat more general, suppose there is a block of p consecutive observations $y_{h+1}, \ldots, y_{h+p}$, which are uncensored. It follows that the d.f. of the y_t, $t \in \bar{\tau}$, conditional on the y_t, $t \in \tau$, can be factored into two multivariate normal d.f.'s for the y_t, $t \in \bar{\tau}$, $t \leqslant h$, and for the y_t, $t \in \bar{\tau}$, $t > h+p$. If there are other blocks of at least p consecutive uncensored observations, then the d.f. can be further decomposed. Thus, ultimately Eq. (12) contains the product of n d.f.'s, of dimension $T_{21}, \ldots, T_{2n}$; computationally it is much easier to handle these than

5. W. E. Wecker, "Prediction Methods for Censored Time Series," pp. 627–32 in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, American Statistical Association, Washington, D.C., 1974.

it is a single one of dimension $T_{21} + \ldots + T_{2n} = T_2$. (Dutt[6] gives apparently accurate and efficient formulas for computing multivariate normal probabilities of dimension less than or equal to six.)

In principal the maximum likelihood (ML) estimators can be found by one of a number of numerical algorithms. The Newton-Raphson procedure has the reputation of converging rapidly; if it is initiated with a consistent estimate, it produces an asymptotically efficient one in a single step. Unfortunately, we can see no simple way of getting an initial consistent estimate. Moreover, Newton-Raphson requires first and second derivations of Eq. (12). These are difficult to obtain and complicated to program.

An alternative approach seems rather well-suited to the problem at hand. This is an algorithm for ML estimation recently studied in depth by Dempster, Laird, and Rubin[7] and called by them the "EM algorithm." Each iteration of the EM algorithm contains two steps: in the E (expectation) step we find expectations of the sufficient statistics pertaining to the "complete" data $y_1, \ldots, y_T$, conditional on Eqs. (8) and (9) and the parameter estimates obtained on the preceding M step; in the M (maximization) step we insert these expectations into the likelihood for $y_1, \ldots, y_T$, and maximize it, to obtain new estimates.

The EM algorithm seems useful here because the M step is very easy to carry out when $f_t(\boldsymbol{\theta})$ has the form of either Eq. (9) or (10). If we observed $y_1, \ldots, y_T$, the log likelihood would be

$$-\frac{(T-p)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=p+1}^{T}\left[y_t - \sum_{j=1}^{p} a_j y_{t-j} - f_t(\boldsymbol{\theta})\right]^2, \tag{13}$$

(ignoring an asymptotically negligible term). Maximizing Eq. (13) with respect to $\boldsymbol{\theta}$ is equivalent to minimizing

$$S_T(\boldsymbol{\theta}) = \frac{1}{T-p}\sum_{t=p+1}^{T}\left[y_t - \sum_{j=1}^{p} a_j y_{t-j} - f_t(\boldsymbol{\theta})\right]^2, \tag{14}$$

and maximization of Eq. (13) with respect to σ^2 is achieved when $\sigma^2 = \min_{\theta} S_T(\boldsymbol{\theta})$. When $f_t(\boldsymbol{\theta})$ is given by Eq. (9), therefore, the a_j and b_i are estimated simply by linear LS regression of y_t on the y_{t-j} and z_{it}. In Eq. (10), things are only slightly more complicated. As an alternative to minimizing Eq. (13) by nonlinear LS, one could find approximate estimates as follows. First estimate the b_i consistently by linear LS regression of y_t on the z_{it} [see Eq. (2)]. Denoting these estimates $\tilde{b}_i$, one then estimates the a_i in an asymptotically efficient fashion by linear LS regression of the

$$y_t - \sum_{i=1}^{p} \tilde{b}_i z_{it}$$

on the

$$y_{t-j} - \sum_{i=1}^{p} \tilde{b}_i z_{i,t-j}\,.$$

This is often called the Cochrane-Orcutt procedure.

6. J. E. Dutt, "Numerical Aspects of Multivariate Normal Probabilities in Econometric Models," *Ann. Econ. Soc. Meas.* 5: 547–61 (1976). [Appendix available from author.]

7. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. R. Stat. Soc. Ser. B* 39: 1–38 (1977).

The LS estimates described in the preceding paragraph depend in simple ways on sufficient statistics of the form

$$\sum_{t=p+1}^{T} y_{t-j}y_{t-k}\,, \quad \sum_{t=p+1}^{T} y_{t-j}z_{i,t-k}\,, \quad \sum_{t=p+1}^{T} z_{i,t-j}z_{l,t-k}\,. \tag{15}$$

Unfortunately the conditional expectations of the sufficient statistics involving the y_t required for the E step are likely to be complicated. They involve first and (if there are some intervals of p or less units between censorings) second moments of the truncated multivariate normal distribution. The distribution in question is the T_2-dimensional distribution of the y_t, $t \in \bar{\tau}$ conditional on the y_t, $t \in \tau$, although as earlier described, this may sometimes be factored, in which case the expectations can be taken with respect to distributions of smaller dimensions. Even so, we have to compute T_{2j} first moments, and $\frac{1}{2}T_{2j}(T_{2j}+1)$ second moments, from distributions of dimension T_{2j}, for all $j = 1, \ldots, n$. Each moment involves a multiple integral of dimension T_{2j}, along with integrals of smaller dimension, and the formulas for second moments are particularly complicated. As a result, we shall now explore alternatives to the E step.

AN ALTERNATIVE ESTIMATOR

Instead of finding conditional expectations of the sufficient statistics, Eq. (15), we could instead find conditional expectations of the y_t, $t \in \bar{\tau}$, themselves, and use these in Eq. (14). Thus the computations described in the previous section would be limited to finding *first* conditional moments. However, unless censorings are infrequent (which is not the case with much of our demand data), this procedure may still lead to evaluation of integrals of high dimension. We propose to reduce the dimensions involved by conditioning on only part of the available information. Therefore, ultimate convergence to the ML estimate cannot be expected; it is to be hoped that the resulting iterations will converge, and to a value closely approximating the ML.

The method described below depends on the existence of at least one block of p consecutive uncensored observations. As with the ML method, the more such blocks the data contain, the easier the computations will be. Denote such a block $y_{t-p}, \ldots, y_{t-1}$, $t > p$, $t < T+1$.

We wish first to predict a censored value y_{j+h}, $h \geqslant 0$. It is convenient to represent Eq. (8) as a first-order vector-difference equation. Define

$$\mathbf{y}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad A = \begin{bmatrix} a_1 & a_2 & \ldots & a_p \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & & \vdots \\ 0 & \ldots & 1 & 0 \end{bmatrix}, \quad \mathbf{w}_t = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Thus Eq. (8) can be written

$$\mathbf{y}_t = A\mathbf{y}_{t-1} + \mathbf{w}_t\,. \tag{16}$$

Then we can recursively generate

$$\mathbf{y}_{t+l} = A^{l+1}\mathbf{y}_{t-1} + \sum_{k=0}^{l} A^k\,\mathbf{w}_{t+l-k}\,, \qquad 0 \leqslant l \leqslant h\,. \tag{17}$$

Denote by $a_{ij}^{(k)}$ the (i, j)th element of A^k. Then from Eq. (17),

$$y_{t+l} = \sum_{j=1}^{p} a_{1j}^{(l+1)} y_{t-j} + u_l\,, \quad 0 \leqslant l \leqslant h\,, \tag{18}$$

$$u_l = \sum_{k=0}^{l} a_{11}^{(k)} w_{t+l-k}\,. \tag{19}$$

We propose to estimate y_{t+h} by its expectation, conditional on y_{t-j}, $1 \leqslant j \leqslant p$, and on the events $y_{t+l} \leqslant 0$, $0 \leqslant l \leqslant h$ (assuming, for simplicity, that all these latter y_{t+l} are censored); namely,

$$\hat{y}_{t+h} = E(y_{t+h} \mid y_{t-j},\ 1 \leqslant j \leqslant p\,,\ y_{t+l} \leqslant 0,\ 0 \leqslant l \leqslant h) = \sum_{j=1}^{p} a_{1j}^{(h+1)} y_{t-j} + \hat{u}_h\,, \tag{20}$$

$$\hat{u}_h = E\left(u_h \mid u_l \leqslant -\sum_{j=1}^{p} a_{1j}^{(l+1)} y_{t-j}\,,\ 0 \leqslant l \leqslant h\right), \tag{21}$$

since the events $y_{t+l} \leqslant 0$ and

$$u_l \leqslant -\sum_{j=1}^{p} a_{1j}^{(l+1)} y_{t-j}$$

are identical, from Eq. (18). To find the distribution of the u_l, write

$$\mathbf{u} = \begin{bmatrix} u_0 \\ \\ \vdots \\ u_h \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & & \cdots & 0 \\ a_{11}^{(1)} & 1 & & & \\ \vdots & & \ddots & & \\ \vdots & & & 1 & 0 \\ a_{11}^{(h)} & \cdots & & a_{11}^{(1)} & 1 \end{bmatrix}, \quad \mathbf{f}_t = \begin{bmatrix} f_t(\boldsymbol{\theta}) \\ \\ \vdots \\ f_{t+h}(\boldsymbol{\theta}) \end{bmatrix}$$

Since the w_t are $\mathrm{NID}(f_t(\boldsymbol{\theta}), \sigma^2)$, it follows from Eq. (19) that $\mathbf{u} \sim \mathrm{N}(\mathbf{D}\mathbf{f}_t, \sigma^2 DD')$. Thus Eq. (21) and thence Eq. (20) may be deduced from formula (3) of Tallis,[8] for the mean of a truncated multivariate normal distribution. (Tallis' formula is expressed in terms of the correlation matrix, not the covariance matrix.)

The computation of $\hat{y}_{t+h}$ requires evaluation of an $(h + 1)$-variate normal probability, along with h-variate normal integrals. If there is a long run of censored observations, the prediction of the later ones will therefore be expensive. It is therefore suggested that one predict these by working *back* from the *subsequent* block of p consecutive uncensored observations, if such there be. This method will also enable us to predict any censored observations *before* the first block of p uncensored values.

8. G. M. Tallis, "The Moment Generating Function of the Truncated Multinormal Distribution," *J. R. Stat. Soc. Ser. B* 23: 223–29 (1961).

To see how this may be done, write Eq. (16) as

$$\mathbf{y}_{t-1} = A^{-1}(\mathbf{y}_t - \mathbf{w}_t)\ ,$$

$$A^{-1} = \begin{bmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ \dfrac{1}{a_p} & \dfrac{-a_1}{a_p} & \cdots & \dfrac{-a_{p-1}}{a_p} \end{bmatrix}$$

Then to predict $y_{t-p-h-1}$, $h \geqslant 0$, from $y_{t-1}, \ldots, y_{t-p}$, we first generate

$$\mathbf{y}_{t-l-2} = A^{-l-1}\mathbf{y}_{t-1} - \sum_{k=0}^{l} A^{k-l-1}\mathbf{w}_{t-k-1}\ , \qquad 0 \leqslant l \leqslant h\ . \tag{22}$$

Let $a_{ij}^{(-k)}$ be the (i, j)th element of A^{-k}. Then from Eq. (22),

$$y_{t-p-l-1} = \sum_{j=1}^{p} a_{pj}^{(-l-1)} y_{t-j} - v_l\ , \qquad 0 \leqslant l \leqslant h\ ,$$

$$v_l = \sum_{k=0}^{l} a_{p1}^{(k-l-1)} w_{t-k-1}\ .$$

Now consider

$$E(y_{t-p-h-1} \mid y_{t-j}\ ,\ 1 \leqslant j \leqslant p\ ,\ y_{t-p-l-1} \leqslant 0\ ,\ \ 0 \leqslant l \leqslant h)$$

$$= \sum_{j=1}^{p} a_{pj}^{(-h-1)} y_{t-j} - E\left(v_h \mid v_l \geqslant \sum_{j=1}^{p} a_{pj}^{(-l-1)} y_{t-j}\ ,\ 0 \leqslant l \leqslant h\right)\ . \tag{23}$$

Define

$$\mathbf{v} = \begin{bmatrix} v_0 \\ \vdots \\ v_h \end{bmatrix},\ E = \begin{bmatrix} a_{p1}^{(-1)} & & & 0 \\ a_{p1}^{(-2)} & \ddots & & \\ \vdots & & \ddots & \\ a_{p1}^{(-h-1)} & \ldots & a_{p1}^{(-2)} & a_{p1}^{(-1)} \end{bmatrix},\ \mathbf{f}_{t-1} = \begin{bmatrix} f_{t-1}(\theta) \\ \vdots \\ f_{t-h-1}(\theta) \end{bmatrix}$$

Because the w_t are NID($f_t(\boldsymbol{\theta})$, σ^2), we have

$$\mathbf{v} \sim \mathrm{N}(E\mathbf{f}_{t-1}(\boldsymbol{\theta}), \sigma^2 EE')\ .$$

Thus we can again compute Eq. (23) by again using Tallis' formula (3).[8]

Note that by using both the forward and backward predictions in concert, to fill in a gap of T_0 censored observations, the multidimensional integrals involved are of order at most $[T_0 + \frac{1}{2}]$. (The IMSL library includes a package for computing bivariate normal probabilities.) Moreover, most of the computations, such as the recursive generation of the $\boldsymbol{A}^k$, $\boldsymbol{A}^{-k}$, and the formation of $\boldsymbol{D}$ and $\boldsymbol{E}$, are straightforward to program. Only the eventual formation of the moment quantities is complicated.

A SIMPLE ESTIMATOR

The prediction method proposed in the previous section, while simpler than that proposed earlier, still does not lend itself to routine calculation and may be expensive if censorings are not sparse. Thus we suggest a general predictor that leads to simple calculations, involving only the univariate normal integral.

It is suggested by the usual one-step predictor of y_t when there is no censoring:

$$\tilde{y}_t = \sum_{j=1}^{p} a_j \tilde{y}_{t-j} + f_t(\theta) \quad ,$$

where the $\tilde{y}_{t-j}$, $j \geqslant 1$, are either observed or predictors themselves. To modify this scheme to our circumstances, first replace Eq. (8) by

$$y_t = \sum_{j=1}^{p} a_j \tilde{y}_{t-j} + w_t \quad . \tag{24}$$

Suppose y_t is censored, so $y_t \leqslant 0$. Then from Eq. (24), we define $\tilde{y}_t$ as

$$\tilde{y}_t = E(y_t \mid \tilde{y}_{t-j},\ 1 \leqslant j \leqslant p,\ y_t \leqslant 0) = \sum_{j=1}^{p} a_j \tilde{y}_{t-j} + f_t(\theta) - \sigma \frac{\phi(\alpha_t)}{\Phi(\alpha_t)} \quad , \tag{25}$$

wherein ϕ and Φ are the standard normal density and d.f., respectively, and

$$\alpha_t = -\left[\sum_{j=1}^{p} a_j \tilde{y}_{t-j} + f_t(\theta) \right] \Big/ \sigma \quad .$$

Since Φ can be computed by a simple transformation of the error function, which is one of the FORTRAN functions, the computations will be simple to program, and inexpensive. This will be so even if there are long blocks of censored observations, so long as one has p values to start with. Notice that when $(t-1, \ldots, t-p) \in \tau$, Eq. (25) is identical to Eq. (20), for $h = 0$.

APPLICATION

To illustrate the alternatives to ML in the previous two sections, the simple model

$$y_t = a y_{t-1} + b_1 + b_2 z_t + e_t \tag{26}$$

was estimated on some of the data described in the Introduction. We took $y_t = \log(1 + D_t)$, $z_t = \log P_t$, where D_t is "demand for" and P_t is "price of" oil with sulfur content between 2.01% and 3.00%. (D_t is measured in thousands of barrels and P_t is measured in dollars per barrel.) Thus $y_t \geqslant 0$ because $D_t \geqslant 0$, and $y_t = 0$ if and only if $D_t = 0$. The regional time series selected was that for the West North Central region. One would expect this model to be much too crude to explain reality, and ideally one would wish to construct a model which used the relationships between the various demand series, and information on additional predetermined variables, and involved additional lagged y_t. However, our concern is primarily to assess differences in the results of the estimation procedures, and these are likely to be easier to detect in the context of a simple model such as Eq. (26). Chern[4] also estimates a first-order dynamic model for demand for electricity.

The 32-observation time series $\{y_t^*\}$ contained seven zero values; these were numbers 11, 13, 14, 15, 29, 30, and 32. To use the procedures described in "An Alternative Estimator," we must first classify the latter observations into three groups. Group I (11, 13, 29, and 32) can be predicted using information on the immediately preceding demand. Group II (15 and 30) can be predicted using information on the immediately *following* demand. Group III (14) can be predicted using the demand of two periods back, because 13 is also zero. The prediction of Groups I and II will involve univariate integrals, and the prediction of 14 involves a bivariate one. ML estimates, on the other hand, necessitate computing a trivariate integral.

The computational formulas for the various predictions are as follows.

Group I

We must find

$$\hat{y}_t = E(y_t \mid y_{t-1}, y_t \leqslant 0)$$

which, as we may immediately infer from Eq. (25), is given by

$$\hat{y}_t = ay_{t-1} + b_1 + b_2 z_t - \sigma \frac{\phi(\alpha_t)}{\Phi(\alpha_t)},$$

where $\alpha_t = -(ay_{t-1} + b_1 + b_2 z_t)/\sigma$.

Group II

In this case we need

$$\hat{y}_{t-2} = E(y_{t-2} \mid y_{t-1}, \ y_{t-2} \leqslant 0) .$$

From Eq. (23), this is

$$\hat{y}_{t-2} = \frac{y_{t-1}}{a} - E\left(v_0 \mid v_0 \geqslant \frac{y_{t-1}}{a}\right), \ v_0 \sim N\left(\frac{b_1 + b_2 z_{t-1}}{a}, \frac{\sigma^2}{a^2}\right).$$

Thus

$$\hat{y}_{t-2} = \frac{y_{t-1} - b_1 - b_2 z_{t-1}}{a} - \frac{\sigma}{a} \frac{\phi(\beta_t)}{1 - \Phi(\beta_t)},$$

where $\beta_t = (y_{t-1} - b_1 - b_2 z_{t-1})/\sigma$.

Group III

Here we need to find

$$\begin{aligned} \hat{y}_{t-1} &= E(y_{t+1} \mid y_{t-1}, y_t \leqslant 0, y_{t+1} \leqslant 0) \\ &= a^2 y_{t-1} + \hat{u}_1 , \end{aligned}$$

$$\hat{u}_1 = E(u_1 \mid u_0 \leqslant -ay_{t-1}, u_1 \leqslant -a^2 y_{t-1}),$$

where

$$\mathbf{u} = \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \sim N\left(\begin{bmatrix} b_1 + b_2 z_t \\ a(b_1 + b_2 z_t) + (b_1 + b_2 z_{t+1}) \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & a \\ a & 1 + a^2 \end{bmatrix}\right).$$

Putting

$$\gamma_t = \left(\frac{ay_{t-1} + b_1 + b_2 z_t}{\sigma}\right), \quad \delta_t = \left[\frac{a^2 y_{t-1} + a(b_1 + b_2 z_t) + (b_1 + b_2 z_{t+1})}{\sigma\sqrt{1 + a^2}}\right].$$

Then from formula (3) of Tallis,[7]

$$u_t = a(b_1 + b_2 z_t) + (b_1 + b_2 z_{t+1})$$

$$- \frac{\sigma\{a\phi(\gamma_t)\Phi(\gamma_t a - \delta_t\sqrt{1 + a^2}) + \sqrt{1 + a^2}\,\phi(\delta_t)\Phi(\delta_t a - \gamma_t\sqrt{1 + a^2})\}}{\Phi_2\left(-\begin{pmatrix}\gamma_t \\ \delta_t\end{pmatrix}, \begin{bmatrix} 1 & a\sqrt{1 + a^2} \\ a\sqrt{1 + a^2} & 1 \end{bmatrix}\right)}.$$

The standardized bivariate normal d.f. in the denominator can be computed by means of the IMSL library program MDBNOR.

The "Simple" Predictor

From Eq. (25) we have

$$y_t = ay_{t-1} + b_1 + b_2 z_t - \sigma\frac{\phi(\alpha_t)}{\Phi(\alpha_t)},$$

$$\alpha_t = -(ay_{t-1} + b_1 + b_2 z_t)/\sigma.$$

An iterative procedure of the type described in "Maximum Likelihood Estimates" was used, alternating between computing LS estimates and predictions. The estimator of σ^2 computed on each step using the current data and estimates was

$$\hat{\sigma}^2 = \frac{1}{T}\sum_t (y_t - \hat{a}y_{t-1} - \hat{b}_1 - \hat{b}_2 z_t)^2.$$

We commenced each iterative sequence by computing the LS estimates for the data y_t^*, $t = 1, \ldots, 32$, that is, taking "zero" values to be zero. In practice better staring values would be negative. However, we wished to compare our results with the naive approach of doing LS regression on y_t^*.

In Table 1 we give the estimates and standard errors using y_t^*, the "alternative" predictors, and the "simple" predictor. Convergence to four decimal places was obtained after nine interative steps in both cases. The most noticeable aspect of the results was the difference between the estimates (particularly $\hat{b}$) using y_t^* and those estimates using the alternative and simple predictors. (We also fitted a model with $y_t = D_t$, $z_t = P_t$, but in this case the estimates of a showed the greater variability.)

The predictors of the censored values are given in Table 2. The main differences here are in the y_{15} and y_{30} values, where predicting from the future and past give very different results.

Table 1. Parameter estimates

	$\hat{a}$	$\hat{b}$	$\hat{\sigma}^2$
y_t^*	0.4224 (0.0286)	−1.5788 (0.4591)	1.5626
Alternative	0.5442 (0.0276)	−5.0370 (0.5955)	1.9766
Simple	0.5274 (0.0275)	−4.5197 (0.5280)	1.7696

Table 2. Predicted values of censored y_t

	$-y_{11}$	$-y_{13}$	$-y_{14}$	$-y_{15}$	$-y_{29}$	$-y_{30}$	$-y_{32}$
Alternative	1.301	0.923	1.569	2.715	0.792	3.154	1.096
Simple	1.107	0.772	1.286	1.360	0.665	1.056	0.931

Heat Balance in Housing: Theory and Results of a Retrofit Experiment*

Thomas H. Woteki

Department of Statistics
Center for Environmental Studies
Princeton University
Princeton, New Jersey

ABSTRACT

An experiment was designed to assess the effects of retrofits on the heat balance in a sample of occupied houses. In particular, the theoretically expected vs empirically determined effects on heat loss of installing additional insulation in attics were examined, as were the implications for policy analysis. The statistical analyses featured use of robust methods applied to large sets of primary data collected in the field.

INTRODUCTION

In this paper we briefly outline the need for systematic evaluation of changes in energy policies as they occur. Paying special attention to the residential sector, we point out the value of field experiments and the need for certain types of data in this sector. We conclude with a brief review of a field experiment, which was conducted to serve as a guide for future experimentation.

ENERGY CONSUMPTION AND PUBLIC POLICY: THE NEED FOR SYSTEMATIC ASSESSMENT

Many people agree that the depletion of our conventional energy resources and the general uncertainty about our energy future is putting a strain on our environment and social institutions and that we are faced with critical decisions about the future. However, there is not so much agreement on how to deal with this energy crisis. Several approaches have been suggested, each having its proponents and opponents: (1) develop nuclear power; (2) develop solar power; (3) develop other sources—geothermal, tidal, hydro, shale oil, coal gasification; (4) deregulate to spur exploration; and (5) switch to more abundant fuels, such as coal.

All of the above suggestions are in the nature of supply strategies or policies. Alternatively we may consider various consumption or conservation policies and strategies:

Economic—(1) realistic pricing and deregulation to control consumption, (2) rate structure reforms to encourage load shifting, and (3) tax incentives/disincentives to encourage conservation.

Behavioral—(1) alert consumers to the need for conservation, (2) inform consumers as to how they can conserve, and (3) provide consumers the opportunity to conserve energy and encourage them to do so.

Physical—(1) install retrofits—thereby altering the current stock of energy-consuming devices, (2) replace old devices with new, more efficient devices, (3) encourage cogeneration wherever possible, and (4) encourage approaches that match production of energy to end uses.

Whichever policies are adopted, the need to assess the effects of policy decisions is apparent. With

*This work was supported in part by the National Science Foundation (RANN) under Grant No. SIA03516A04 and the Energy Research and Development Administration (now the Department of Energy) under Grant No. EC-77-02-4288.

respect to conservation policies this need calls for improved or, in most cases, new data on energy consumption. The nation's energy consumption is the result of millions of individual decisions about energy use against the background of its social and institutional structures; therefore, to determine the effects of a new policy, we will need to know how the policy affects the individual decisions that determine aggregate consumption. But data on energy use are scarce and inherently hard to come by: decisions about use are decentralized, records are usually not kept, and energy costs may be only a small part of the total cost of an activity. Thus, although data presently exist to describe energy use at the aggregate level, there is very little data available at the level of the individual, either residential or industrial consumer, and no systematic effort is being made to monitor the effects of policy decisions on individuals' energy uses.

One potentially effective means of obtaining such microlevel data is the field experiment. Field experimentation affords three advantages when assessing the effects of policy decisions:

1. Field experimentation allows the relative effectiveness of alternative policies to be assessed on a scale that makes failure of the policies to achieve desired effects tolerable.

2. Partially controlled field experiments are a step toward providing precise estimates of the effects of policies; they permit a systematic assessment of the factors that help determine the results of new policies.

3. Field experimentation allows the policymaker to experience some of the problems that are bound to occur when a new policy is implemented.

The advantages of field experimentation outlined above are taken for granted as first principles by statisticians but not necessarily by policymakers, especially if the expense, difficulty of execution, and ethical problems attendant to field experiments are considered. Thus the single most important task for statisticians interested in becoming involved in problems related to energy consumption and public policy may be that of consciousness raising. We must alert policymakers to the need for data adequate to the critical decisions at hand and the singular value of data and experiences gained as part of a field experiment.

THE RESIDENTIAL SECTOR

Approximately one-third of the nation's energy consumption is attributable to fuel and electricity use in the residential sector. Of that one-third, slightly less than half is used for personal transportation, and about one-third is used for space conditioning. Both aspects of residential energy use provide scope for contributions by statisticians, especially with respect to planning field experiments and surveys and collecting, evaluating, and analyzing data bearing on policy questions. In determining how energy is used for space conditioning, for example, we need data to describe and model (1) the physical characteristics of the nation's housing stock; (2) the type, age, and typical rate of use of various heating devices and other appliances; (3) the demographic and socioeconomic characteristics of users; (4) the relationship between the energy used in space conditioning structures, the physical characteristics of the structures, and the behavior of the occupants; and (5) the expected return on investments in specific retrofits on a structure-by-structure basis. *Also*, we need measuring devices and instruments that can collect this data.

All of the tasks implied above will require a great deal of effort if they are done. We believe they must be done if the effects of conservation policies on the consumption patterns of residential consumers and aggregate energy consumption in the residential sector, the ultimate target, are to be systematically and accurately determined. We also believe that statisticians should take the initiative in pointing out the need for systematic evaluation of energy policies and that they should not hesitate to claim their rightful role in helping to determine the effects of such policies.

THE TWIN RIVERS PROJECT

Since 1972, members of Princeton University's Center for Environmental Studies have been investigating various aspects of energy consumption in the residential sector. We have been involved in a series of experiments at the planned housing development of Twin Rivers in East Windsor, New Jersey. About 12,000 people live in approximately 3000 houses in Twin Rivers. Our group has monitored the construction at the site, interviewed many of those responsible for energy-related decisions in the planning and construction phase, formally surveyed and informally interacted with the residents, obtained a complete record of monthly gas and electric utility meter readings, built an onsite weather station, placed data acquisition instruments in more than 25 townhouses, and rented and occupied another of these townhouses, turning it into a field laboratory. Both the National Science Foundation (RANN) and the Energy Research and Development Administration (now Department of Energy) have supported our efforts.

Our basic observational units have been two-story, three-bedroom, attached townhouses and their occupants. The townhouses were conventionally built with masonry walls and wood framing for floors and roofs. They sold for approximately $30,000 when built in 1971, and they enclose about 1500 ft^2 of living space. Typically, about 15 ft^3 of natural gas per Farenheit degree day are required to heat a Twin Rivers townhouse. Thus, over a six-month, 5000-degree-day winter, the total requirement is 75,000 ft^3 of gas, which at $0.25 per hundred cubic feet results in a heating bill of about $190 for the year. Electricity is used at an average rate of about 1500 kWhr per month from May to September resulting in an average monthly electricity bill of $60 during this period. Most families in Twin Rivers have roots in New York City. Their Twin Rivers townhouses represent their first home ownership experience. About half the residents are Jewish, 96% are white, and most heads of households are white-collar workers in New York. Twin Rivers is about 50 miles from New York and 15 miles from Princeton.

Our goals in working at Twin Rivers have been to

1. establish that field experiments can be carried out and provide a basis in experience for further work;
2. examine the role of the resident in conserving energy, the physical characteristics of a dwelling which determine energy consumption, and the relationship between resident and dwelling;
3. develop exportable diagnostic tools, both physical and data analytic, for evaluating conservation strategies and policies; and
4. identify some effective retrofits.

We believe we have been very successful in pursuit of these goals, especially in establishing that generally useful field experiments can be carried out. Our introduction to this section of the report summarizes how far we have gone toward achieving this particular goal. Our achievements with respect to our other goals are documented in the reports cited in the bibliography. Some highlights of these reports are

- Twin Rivers residents reacted to the onset of the 1973–1974 oil embargo by reducing their space heating energy use by 15% (Mayer, 1976).
- Simple two-parameter models provide useful descriptions of the variation in monthly and hour-to-hour consumption of energy used to heat Twin Rivers dwellings. The two parameters are the conductivity of the shell of the dwelling and the temperature at which the furnace first comes on. The models are related so that microlevel measurements can be related to estimates of monthly consumption (Mayer, 1976; Woteki, 1976).
- Residents will reduce their consumption when provided with feedback information on their consumption in relationship to their peers and to weather conditions (Seligman and Darley, 1976; 1976).
- Simple experiments employing portable electric area heaters can be done to determine the efficiency of any residential furnace (Sonderegger, 1977).
- Simple models analogous to electric circuit models with two resistances should prove very useful in diagnosing how much heat is lost by conduction from the living space of a house to its attic. Such a model would be useful in diagnosing the need for and effects of installing additional attic insulation in a wide variety of housing types (Beyea, Dutt, and Woteki, 1977; Woteki, Dutt, and Beyea, 1977).

BIBLIOGRAPHY

Field Experiments

J. Fox, H. Fraker, R. Grot, D. Harrje, E. Schorske, and R. H. Socolow, *Energy Conservation in Housing*, Report No. 6, Center for Environmental Studies, Princeton University, Princeton, N.J., 1973.

S. Hall and D. Harrje, *Instrumentation for the Omnibus Experiment*, Report No. 21, Center for Environmental Studies, Princeton University, Princeton, N.J., 1975.

D. Harrje, *A Summary of Instrumentation in the Twin Rivers Program*, Report No. 54, Center for Environmental Studies, Princeton University, Princeton, N.J., 1977.

National Research Council Committee on Measurement of Energy Consumption, *Energy Consumption Measurement: Data Needs for Public Policy*, 1977.

R. H. Socolow, *The Twin Rivers Program on Energy Conservation in Housing: A Summary for Policymakers*, Report No. 51, Center for Environmental Studies, Princeton University, Princeton, N.J., 1977.

Residents and Dwellings

R. A. Alpert, *Electricity: Residential Consumption and Conservation*, senior thesis, Princeton University, Princeton, N.J., 1976.

C. E. Horowitz and L. S. Mayer, *The Relationship Between the Price of Natural Gas and Consumption Levels: A Partially Controlled Study*, Center for Environmental Studies report, Princeton University, Princeton, N.J., 1976.

L. S. Mayer, *Estimating the Effects of the Energy Crisis on Residential Energy Consumption*, Center for Environmental Studies report, Princeton University, Princeton, N.J., 1976.

A. Pollack, *Modelling Attic Temperature in Three Highly Instrumented Townhouses in Twin Rivers, N.J.*, Report No. 28, Center for Environmental Studies, Princeton University, Princeton, N.J., 1976.

C. Seligman and J. Darley, *Feedback as a Means of Decreasing Residential Energy Consumption*, Report No. 34, Center for Environmental Studies, Princeton University, Princeton, N.J., 1976.

C. Seligman and J. Darley, *Psychological Strategies to Reduce Energy Consumption*, Report No. 41, Center for Environmental Studies, Princeton University, Princeton, N.J., 1976.

Diagnostic Tools

J. Beyea, G. S. Dutt, and T. H. Woteki, *Critical Significance of Attics and Basements in the Energy Balance of Twin Rivers Townhouses*, Center for Environmental Studies report, Princeton University, Princeton, N.J., 1977.

D. Harrje and R. Grot, "Automated Air Infiltration Measurements and Implications for Energy Conservation," *Proceedings of the International Conference on Energy Use Management*, Tucson, Arizona, 1977.

R. C. Sonderegger, *Modelling Residential Space Heating*, Ph.D. thesis, Princeton University, Princeton, N.J., 1977.

Effective Retrofits

A. Pollack, *Residential Energy Consumption: The Effects of Retrofits on Summer Electricity Demand*, senior thesis, Princeton University, Princeton, N.J., 1977.

T. H. Woteki, *The Princeton Omnibus Experiment: Some Effects of Retrofits on Space Heating Requirements*, Report No. 43, Center for Environmental Studies, Princeton University, Princeton, N.J., 1976.

T. H. Woteki, *Some Effects of Retrofits on Interior Temperatures in a Sample of Houses*, Working Paper No. 31, Center for Environmental Studies, Princeton University, Princeton, N.J., 1977.

T. H. Woteki, G. S. Dutt, and I. Beyea, *The Two-Resistance Model for Attic Heat Flow: Implications for Conservation Policy*, Report No. 51, Center for Environmental Studies, Princeton University, Princeton, N.J., 1977.

Heat Shock Threshold Estimation for Fish Eggs and Larvae in Power Plant Cooling Systems*

Allan H. Marcus

Washington State University
Pullman, Washington

ABSTRACT

Although mortality and hatching success data are often analyzed by bioassay methods (e.g., probits or logits), it is suspected that actual thresholds exist for the adaptive response of a biological system to multiple environmental stresses. These thresholds can be estimated using a general linear model. The optimal estimation of the response thresholds is a nonlinear least-squares problem, however, and the derivatives of the residual sum of squares surface will have cusps. Approximate confidence regions for thresholds are readily calculated. The separate effects of temperature shock and cumulative temperature dose in electric power plant cooling systems are shown for hatching success of striped bass eggs and for mortality of the larvae of the American shad.

INTRODUCTION AND DISCUSSION OF RESULTS

The Chesapeake Bay is the most productive marine estuary in the world. There is intense competition for use of its waters for shipping, recreational and commercial fishing, and power plant cooling (among others). The requirements, however, for efficient operation and economical design of electricity generating plants (whose condenser cooling waters should be discharged at a sufficiently high temperature to control biological fouling of condenser screens without adding biocides such as chlorine and ozone) may be directly opposed to maximal production of finfish and shellfish. Eggs and larvae of these animals may be drawn into the power plant cooling system and, as they pass through the heat exchange system, be subjected to an abrupt increase in temperature that overwhelms their adaptive capabilities (i.e., heat shock). They may then spend an extended period of time in the heated waters of the discharge where, even if they have not succumbed to heat shock, the cumulative exposure to temperatures exceeding their acclimation limit will result in increased larval mortality and in decreased hatching success of eggs.

There is sufficient evidence to believe that there are, indeed, actual physiological thresholds for heat stresses to aquatic and other organisms.[1] It is thus reasonable to look for distinct threshold temperature values for heat shock and cumulative exposure, as they represent distinct short-term and long-term phenomena. In any case, the specification of temperature limits for cooling water discharges are often given separately. A case in point is the water discharge permit of Calvert Cliffs Nuclear Power Plant owned and operated by the Baltimore Gas and Electric Company. This very efficient steam electric power generating plant has a once-through cooling system. The original operating permit required a maximum temperature increase (DELT) across the condenser of 10°F (5.6°C) and a maximal temperature in the discharge canal of 90°F (32.2°C).

---

*This work supported by the Maryland Power Plant Siting Program.

1. V. H. Hutchison, "Factors Influencing Thermal Tolerances of Individual Organisms," pp. 10 26 in *Proceedings of Thermal Ecology, II*, ed. by G. W. Esch and R. W. McFarlane, A.E.C. Report CONF-750425, U.S. Government Printing Office, Washington, D.C., 1976.

The normal summer surface water temperature is 80 F but may be as high as 83 F for 100 hr per year, so the maximal discharge temperature could be as high as 93 F (33.9 C). The plant has been operated experimentally at DELT = 12°F (6.7°C) for 316(*b*) studies, and higher values up to DELT = 14 F (7.8 C) have been considered.[2] We believe these values are likely to be typical design values for future power plants.

Numerous temperature stress studies have been carried out at the Chesapeake Biological Laboratory of the University of Maryland.[1] We will discuss some results for the mortality of larvae of the striped bass (*Morone saxatilis*) and for the hatching success of eggs of the American shad (*Alosa sapidissima*). The eggs or larvae are first acclimated to a BASE temperature, then subjected to an instantaneous temperature increase (DELT), which is maintained for TIME minutes. Here, BASE was 18°C for shad eggs and 24 C for bass larvae. DELT ranged from 10 C to 14.5 C, and TIME varied from 5 to 60 min.

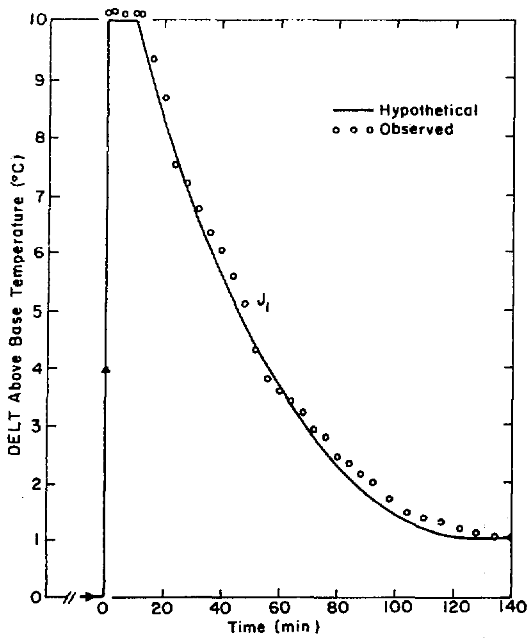


Fig. 1. Total percentage mortality for nine experiments using 50 eggs each from striped bass 1. For an explanation of letter codes, see Table 1. *Source:* T. S. Y. Koo, C. F. Smith, M. L. Johnston, G. E. Balog, Jr., and H. L. Mathers, *Effects of Heat Shocks on Fish Eggs and Larvae.* Report CEES No. 76-112-CBL, Chesapeake Biological Laboratory, University of Maryland, July 1976, adapted from Appendix I.

Table 1. Experimental runs for all fish eggs and larvae[a]

Letter designation	Excess temperature (C)	Time of exposure (minutes)	Cooling-down time (minutes)
J1	10	10	120
J2	10	60	160
K1	11.5	5	140
K2	11.5	30	160
L1	13.0	[illegible]	160
L2	13.0	20	180
M1	14.5	5	180
M2	14.5	15	200
Control	0	0	0

[a]About 450 eggs from each fish were fertilized at a base temperature, then divided into the nine groups listed above with about 50 fertilized eggs in each experimental group for each fish.

Source: T. S. Y. Koo, C. F. Smith, M. L. Johnston, G. E. Balog, Jr., and H. L. Mathers, *Effects of Heat Shocks on Fish Eggs and Larvae.* Report CEES No. 76-112-CBL, Chesapeake Biological Laboratory, University of Maryland, July 1976, adapted from Appendix I.

The waters were then allowed to cool naturally to BASE, as shown in Fig. 1.

About 450 eggs from each fish were divided into nine groups of 50 each. The nine experimental runs included a control run and the eight combinations listed in Table 1. Larvae from striped bass 6 were 2 days, 4 hr old; her larvae had the yolk-sac partially absorbed and the eyes pigmented. Larvae from striped bass 7 were only 2 days old, and although the yolk-sac was partially absorbed, the eyes were not pigmented. Because the BASE temperature was high (24°C), we believed that these data were particularly relevant to the conditions of power plant discharge.

We also evaluated eggs of four American shad at age 24 hr (late gastrula stage) to study the variation among individuals of the same species at the same developmental stage. For each of these shad, 450 eggs were allocated among nine treatments as above. The BASE here was 18°C. Eggs in the tail-free embryo stage (age 41 hr) at BASE 17°C showed practically no mortality, while those in the early gastrula stage at BASE 24°C showed almost 100% mortality.

Striped bass are a particularly important environmental indicator species for citizens of Maryland. The status of the striped bass population plays a role in their gastronomic and recreational preferences

2. Decision of Hearing Officer, NPDES Permit =MD002399, Jan. 15, 1977. In the matter of Calvert Cliffs Nuclear Power Plant, Baltimore Gas and Electric Company, Request for Adjudication.

that is scarcely captured by calculation of the economic benefits of commercial and sport fishing for this species. While neither bass nor shad are spawned near the Calvert Cliffs plant, we have for the sake of definiteness evaluated the Calvert Cliffs discharge limits for bass larvae and shad eggs.

The basic data are percent mortality of larvae (e.g., Fig. 2) or percent hatching success of eggs. These data were arc sine transformed to achieve stable variance and then were treated by the methods described in the next section. Confidence contours for larval mortality are sketched in Fig. 3. Note that for striped bass 6, the thresholds for the experimental condenser cleaning trials of 1976 12°F (6.7 C) for heat shock and a maximum discharge canal temperature of 95°F (35°C) *do not lie below* the 99% lower confidence bound for heat damage, although they do lie below the 95% lower confidence bound for heat damage. There is thus a remote possibility that these extreme conditions could start to damage some striped bass larvae. However, the striped bass 7 larvae lie well below the detectable thresholds of mortality.

The assumed exposure, TIME = 1 min, is an extremely short time for heat dose. Longer TIME

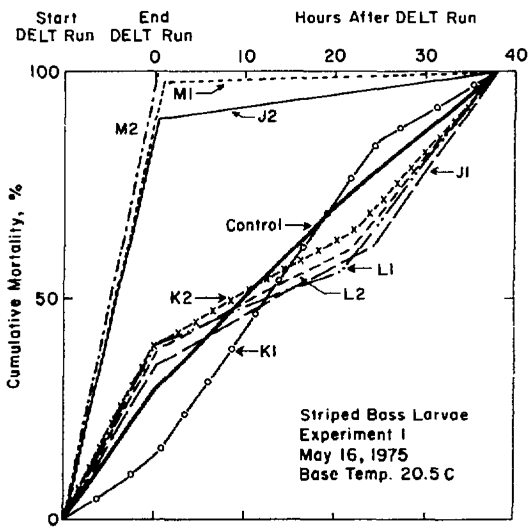


Fig. 2. Example of typical agreement between experimental time–excess temperature history, D_1, and its hypothetical analogue. *Source:* T. S. Y. Koo, C. F. Smith, M. L. Johnston, G. E. Balog, Jr., and H. L. Mathers, *Effects of Heat Shocks on Fish Eggs and Larvae*, Report CEES No. 76-112-CBL, Chesapeake Biological Laboratory, University of Maryland, July 1976, Fig. 8.

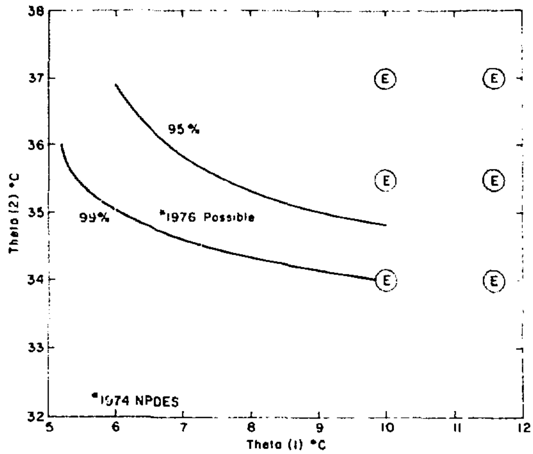


Fig. 3. Outer confidence contours at levels of 95 and 99% for temperature shock thresholds, THETA(1), and heat dose thresholds, THETA(2), for striped bass 6. Also shown are the conditions specified in the 1974 NPDES Permit and the 1976 worst-case operating condition. E experimental points

contours are being evaluated by a computer program not yet fully validated; these results will be reported in detail later. Longer TIMES will not change the center or approximate shape of the contours but will compress them somewhat in the heat dose threshold, THETA(2), direction.

For the shad egg hatching success data we evaluated the hypothesis that the threshold for heat shock fell below 14°F (7.8°C) and the threshold for test temperature dose effect fell below 90°F (32.2°C) with an assumed exposure TIME = 1 min. Application of a formal *F* test of significance is reported in Table 2. For shad 1, the thresholds are very probably below these values ($P = 0.4$); for shad 2 ($P = 0.026$) and shad 4 ($P = 0.01$), there is a small possibility that their actual damage thresholds are below these values; for shad 3 ($P = 0.0006$), there is practically no chance that her damage thresholds are below these values. We thus conclude that limits of 14°F (7.8°C) for DELT and 90°F (32.2°C) in the discharge canal will allow damage to the egg hatching success of *some* American shad. The population effects on the Chesapeake Bay fishery can then be evaluated (in principle) by an ecological model. Contours similar to those for striped bass will be presented elsewhere.

The present model *assumes* the existence of thresholds. Unlike chemical carcinogens or ionizing radiation, there is evidence that physiological

Table 2. Test of hypothesis that THETA(1) = 7.8°C, THETA(2) = 32.2°C, for TIME = 1 min

Fish	F value[a]	Significance probability
Shad 1	1.23	0.4
Shad 2	8.92	0.026
Shad 3	47.37	0.0006
Shad 4	13.34	0.01

Degrees of freedom 2, 4.

thresholds for temperature are real and represent limiting conditions of physiological homeostatic adjustment. The question of thresholds is controversial, however, in that this assumption often leads to more conservative assessments of damage and higher allowable levels of environmental insult than other methods. For example, we could estimate the proportion of bass larvae dying, or shad eggs hatching, using probit or logistic regression methods, and express the uncertainties of the analysis using a confidence interval for that proportion. Our threshold estimation procedure could thus be described as a "horizontal window" method, and the probit or logit method is a "vertical window" method. Clearly, both methods are of use in deriving temperature limits for cooling water discharge with laboratory bioassays.

METHODS

Introduction

The effects of temperature stress on aquatic organisms may depend on at least two threshold-type phenomena. In a typical bioassay experiment, organisms (e.g., larvae, for which mortality is measured, or eggs, for which hatching success is measured) are exposed to an instantaneous temperature increase DELT (°C) over a long-standing base temperature BASE (°C). This temperature increase to test temperature

$$\text{TEST} = \text{DELT} + \text{BASE} \tag{1}$$

is maintained for TIME minutes. The first effect is that of heat shock or more properly, temperature shock:

$$\begin{aligned} &\text{no effect} && \text{if} && \text{DELT} < \text{THETA}(1)\,,\\ &\text{adverse effect} && \text{if} && \text{DELT} \geqslant \text{THETA}(1)\,. \end{aligned}$$

The second effect is a dosage effect which depends on TEST temperature:

$$\begin{aligned} &\text{no effect} && \text{if} && \text{TEST} < \text{THETA}(2)\,,\\ &\text{adverse effect} && \text{if} && \text{TEST} \geqslant \text{THETA}(2)\,. \end{aligned}$$

It is by no means obvious that THETA(1) + BASE = THETA(2), although this could be tested.

This problem can be viewed as a multivariate "hockey stick" regression.[3] Suppose that N experiments are carried out at levels DELT(I), TIME(I), BASE, so that

$$\text{TEST(I)} = \text{BASE} + \text{DELT(I)}\,. \tag{2}$$

The response is the number of "successes," S(I), out of K(I) organisms. We can define possible responses:

$$\text{P(I)} = \text{S(I)}/\text{K(I)}\,,$$

$$\text{ARCSIN(I)} = \arcsin\left(\{\text{S(I)}/[\text{K(I)}+1]\}^{1/2}\right) + \arcsin\left(\{[\text{S(I)}+1]/[\text{K(I)}+1]\}^{1/2}\right), \tag{3}$$

$$\text{LOGIT(I)} = \ln\left([\text{S(I)}+0.5]/[\text{K(I)}+1-\text{S(I)}]\right),$$

$$\text{LOG(I)} = \ln\left([\text{S(I)}+0.5]/[\text{K(I)}+1]\right).$$

Let Y(I) be some one of these. [Hasselblad used LOGIT(I)]. Our model is

$$\begin{aligned} \text{Y(I)} &= \text{A(0)} + \text{A(1)}\,[\text{DELT(I)} - \text{THETA(1)}] + \text{A(2)TIME(I)}[\text{TEST(I)} - \text{THETA(2)}] + \text{error}\\ &= \underset{J=0}{\overset{p}{\text{SUM}}}[\text{BETA(J)X(I, J)}] + \text{error}\,, \end{aligned} \tag{4}$$

where

$$p = 4\,,$$

$$\text{BETA(0)} = \text{A(0)}\,,\quad \text{X(I, 0)} \equiv 1\,, \tag{5}$$

3. V. Hasselblad, J. P. Creason, and W. C. Nelson, *Regression Using Hockey Stick Functions*, Report EPA-600/1-76-024, U.S. Environmental Protection Agency, Washington, D.C., June 1976.

$$\text{BETA}(1) = \text{A}(1)\,,\quad \text{X}(\text{I},1) = \text{DELT}(\text{I})\text{U}(\text{I})\,. \tag{6}$$

$$\text{BETA}(3) = \text{A}(1)\text{THETA}(1)\,,\quad \text{X}(\text{I},3) = \text{U}(\text{I})\,. \tag{7}$$

$$\text{BETA}(2) = \text{A}(2)\,,\quad \text{X}(\text{I},2) = [\text{TIME}(\text{I})][\text{TEST}(\text{I})]\text{V}(\text{I})\,. \tag{8}$$

$$\text{BETA}(4) = -\,\text{A}(2)\text{THETA}(2)\,,\quad \text{X}(\text{I},4) = \text{TIME}(\text{I})\text{V}(\text{I})\,. \tag{9}$$

$$\text{U}(\text{I}) = \begin{cases} 0 & \text{if}\quad \text{DELT}(\text{I}) < \text{THETA}(1)\,, \\ 1 & \text{if}\quad \text{DELT}(\text{I}) \geqslant \text{THETA}(1)\,, \end{cases} \tag{10}$$

$$\text{V}(\text{I}) = \begin{cases} 0 & \text{if}\quad \text{TEST}(\text{I}) < \text{THETA}(2)\,, \\ 1 & \text{if}\quad \text{TEST}(\text{I}) \geqslant \text{THETA}(2)\,, \end{cases} \tag{11}$$

where

$$\text{Z}^{*} = \begin{cases} 0 & \text{if}\quad \text{Z} < 0\,, \\ \text{Z} & \text{if}\quad \text{Z} \geqslant 0\,. \end{cases}$$

This is a key point. Note that the "predictor" variables X(I, *J*) depend on the currently believed values of THETA(J). The ordinary least-squares estimate of the regression vector BETA is B,

$$\underset{(p+1\times 1)}{\text{B}} = \underset{(p+1\times p+1)}{\text{C}^{-1}} \times \underset{(p+1\times N)}{\text{X}^{1}} \times \underset{(N\times 1)}{\text{Y}}\,, \tag{12}$$

where

$$\text{C} = \underset{(p+1\times N)}{\text{X}^{1}} \times \underset{(N\times p+1)}{\text{X}}\,, \tag{13}$$

$$\text{X} = [\text{X}(\text{I},\text{J})]\,,\ \text{I} = 1,\ldots,N,\ \ \text{J} = 0,\ldots,p\,,$$

$$\text{Y} = [\text{Y}(\text{I})]\,,\ \text{I} = 1,\ldots,N\,,$$

$$\text{B} = [\text{B}(\text{J})]\,,\ \text{J} = 0,\ldots,p\,.$$

The predicted vector is YHAT = XB. The residuals are given by a vector RESID = Y − XB. The sums of squares are

$$\text{SSY} = \sum_{\text{I}=1}^{N}[\text{Y}(\text{I}) - \text{YBAR}]^2\,, \tag{14}$$

$$\begin{aligned}\text{SSRESID} &= \sum_{\text{I}}[\text{RESID}(\text{I})^2] \\ &= \sum_{\text{I}}[\text{Y}(\text{I}) - \text{YHAT}(\text{I})]^2 \\ &= (\text{Y} - \text{XB})^{1}(\text{Y} - \text{XB})\,. \end{aligned} \tag{15}$$

Goodness of fit is measured by

$$\text{RSQR} = 1 - (\text{SSRESID}/\text{SSY}) \tag{16}$$

and is tested by

$$\text{F} = (N - 1 - p)(\text{SSY} - \text{SSRESID})/[(p)\text{SSRESID}]\,;$$

that is,

$$\text{F} = (N - 1 - p)\text{RSQR}/(p)(1 - \text{RSQR})\,.$$

The "best" values of THETA(1) and THETA(2) may not be found. As has been noted, the function SSRESID will jump every time THETA(1) crosses DELT(I) and every time THETA(2) crosses another TEST(I). Between these values, SSRESID is a quadratic function of THETA. A derivative-free nonlinear regression program such as the BMD.PAR program to be released shortly may be of use here.

Confidence Contours

It may be preferable to present the results of analysis by a confidence region or contour for the thresholds. To test a null hypothesis about

$$\text{THETA}^{1} = [\text{THETA}(1)\,,\ \text{THETA}(2)]\,, \tag{17}$$

we note that this is equivalent to

$$\text{B}(1)\text{THETA}(1) + \text{B}(3) = 0\,,$$

$$\text{B}(2)\text{THETA}(2) + \text{B}(4) = 0\,,$$

so that the null hypothesis will now read

$$\underset{(2\times p+1)}{\text{H}} \times \underset{(p+1\times 1)}{\text{BETA}} = 0\,,$$

where the matrix **H** is

$$\begin{bmatrix} 0 & \text{THETA}(1) & 0 & 1 & 0 \\ 0 & 0 & \text{THETA}(2)\times\text{TIME} & 0 & \text{TIME} \end{bmatrix}.$$

Define the test statistic

$$T = HB ,$$

so for the usual model in which the Y(I) are independent with variance $SIGMA^2$ it has mean value and covariance matrix

$$E(T) = H(BETA) ,$$

$$VAR(T) = E(TT')$$

$$= (HC^{-1}H')^{-1}(SIGMA^2) .$$

This produces statistics

$$F(THETA) = (N - 1 - p)B'H'(HC^{-1}H')^{-1}HB/(2SSRES) ,$$

$$P(THETA) = Prob[F(2, N - 1 - p) > F(THETA)] .$$

Note that the value of THETA enters into both B and C since the values U, V depend on THETA. The function SSRESID does not now have to be minimized, merely contoured for fixed values covering a range of THETA.

Optimal Estimates: Explicit Solution

As Hasselblad, Creason, and Nelson observed,[1] it is only necessary to calculate SSRESID a total of $(2N - 1)$ times to obtain the least-squares estimate of THETA in the single-threshold case, because the optimal THETA must lie either between two adjacent values of X(I) or coincident with a value of X(I). An explicit solution is thus possible for one, two, or any number of thresholds, because in addition to the "normal" equations resulting from minimizing SSRESID in Eq. (15) [subject to the linear constraint in Eq. (19)], we also have the requirement that the regression plane is continuously joined at the threshold values. Their solution[1] for the single threshold case can then be applied repeatedly so long as there is at least one value of Y which allows us to discriminate between the two (or more) threshold effects. These equations are given explicitly in the Appendix.

The present method extends in a fairly obvious way the well-known univariate two-phase regression problem.[4-6] In our case there are two (or more) distinct but not necessarily independent thresholds for multiple predictor variables. Hinkley has established the asymptotic normality of the estimated intercept THETA but also has shown that the asymptotic approximation is poor for small samples.[4,6] The joint asymptotic distribution of THETA(1), THETA(2) estimates should then be bivariate normal, but the possibility of obtaining useful small-sample results suitable for these data appears remote.

APPENDIX: LEAST-SQUARES ESTIMATES FOR DOUBLE THRESHOLD MODEL

Indices

$$K = 1, 2 ,$$

$$L = 1, \ldots, M \text{ or } L = M + 1, \ldots, N ,$$

$$M = 1, \ldots, N .$$

Variables

$$X(1, I) = DELT(I) ,$$

ordered so that

$$DELT(I) \leqslant DELT(I + 1) , \tag{A-1}$$

$$X(2, I) = TIME(I)[DELT(I) + BASE] . \tag{A-2}$$

Statistics

$$SY(L, M) = \sum_{I=L}^{M} Y(I) , \tag{A-3}$$

$$SX(K, L, M) = \sum_{I=L}^{M} X(K, I) , \tag{A-4}$$

$$SSX(K, L, M) = \sum_{I=L}^{M} X(K, I)^2 , \tag{A-5}$$

$$SYX(K, L, M) = \sum_{I=L}^{M} Y(I)X(K, I) . \tag{A-6}$$

4. D. V. Hinkley, "Inference About the Intersection in Two-Phase Regression," *Biometrika* **56**: 495-504 (1969).

5. David V. Hinkley, "Inference About the Change-Point in a Sequence of Random Variables," *Biometrika* **57**: 1-17 (1970).

6. David V. Hinkley, "Inference in Two-Phase Regression," *J. Am. Stat. Assoc.* **66**: 736-43 (1971).

Deviations

$$\mathrm{SDEV}(K, L, M) = (M - L)\mathrm{SSX}(K, L, M) - \mathrm{SX}(K, L, M)^2 \,. \qquad \text{(A-7)}$$

$$\mathrm{SDEU}(K, L, M) = (M - L)\mathrm{SYX}(K, L, M) - \mathrm{SY}(L, M)\mathrm{SX}(K, L, M) \,. \qquad \text{(A-8)}$$

$$\mathrm{SDEW}(K, L, M) = \mathrm{SX}(K, L, M)\mathrm{SYX}(K, L, M) - \mathrm{SSX}(K, L, M)\mathrm{SY}(L, M) \,. \qquad \text{(A-9)}$$

First Case: $L < M$

If $X(1, L) < \mathrm{THETA}(1) < X(1, L + 1)$,

$$A(0) = \mathrm{SY}(1, L)\ L \,, \qquad \text{(A-10)}$$

$$\mathrm{THETA}(1) = [A(0)\mathrm{SDEU}(1, L + 1, M) + \mathrm{SDEW}(1, L + 1, M)] \div \mathrm{SDEV}(1, L + 1, M) \,, \qquad \text{(A-11)}$$

$$A(1) = [\mathrm{SY}(L + 1, M) - (M - L)A(0)] \div [\mathrm{SX}(1, L + 1, M) - (M - L)\mathrm{THETA}(1)] \,. \qquad \text{(A-12)}$$

Having estimated THETA(1), we will now estimate THETA(2) using partial residuals that are based on the estimated THETA(1). Replace Y(1) in Eqs. (A-3), (A-6), and (A-7) through (A-12) by

$$Y^*(1) = Y(1) - A(0) - A(1)X(1, 1) \,; \qquad \text{(A-13)}$$

thus,

$$A^*(0) = \mathrm{SY}^*(1, M)\ M \,. \qquad \text{(A-14)}$$

$$\mathrm{THETA}(2) = [A^*(0)\mathrm{SDEU}^*(2, M + 1, N) + \mathrm{SDEW}^*(2, M + 1, N)] \div \mathrm{SDEV}^*(2, M + 1, N) \,. \qquad \text{(A-15)}$$

$$A(2) = [\mathrm{SY}^*M + 1, N) - (N - M)A^*(0)] \div [\mathrm{SX}(2, M + 1, N) - (N - M)\mathrm{THETA}(2)] \,. \qquad \text{(A-16)}$$

If $\mathrm{THETA}(1) = X(1, L)$,

$$A(0) = -\mathrm{SDEW}(1, L + 1, M) \div \mathrm{SDEV}(L, L + 1, M) \,. \qquad \text{(A-17)}$$

$$A(1) = [\mathrm{SY}(L + 1, M) - (M - L)A(0)] \div \mathrm{SX}(1, L = 1, M) \,. \qquad \text{(A-18)}$$

Second Case: $L > M$

In equations (A-10) through (A-16) replace $K = 1$ by $K = 2$.

Statistical Analysis of Reactor Core Operating Limits

Rubin Goldstein and Raymond Krisciokaitis Krisst
Combustion Engineering
Windsor, Connecticut

ABSTRACT

Monitoring and protection systems are included in a modern pressurized water reactor (PWR) to observe the nuclear and thermal characteristics of the reactor core. These systems either alert the operator as core operating limits are approached or initiate a reactor trip before fuel design limits are exceeded. To prevent damage to the nuclear fuel, Combustion Engineering (C-E) specifies fuel design limits on the departure from nucleate boiling ratio (DNBR) and the peak linear heat rate (PLHR) or local power density. Calculational approximations, uncertainties in the design parameters, measurement inaccuracies, and calibration and processing errors all have an effect on the on-line inferred values of DNBR and PLHR.

To demonstrate the potential gains in available thermal margin by using a statistical approach, a stochastic simulation of the basic input variables to the Core Protection Calculator in the C-E reactor protection system is carried out. The results obtained from the output distribution for the DNBR uncertainty are compared with corresponding multiplicative estimates ("worst case conditions"). For a typical set of operating conditions, it is specifically demonstrated that the margin to the limit on DNBR is increased by approximately 10.5%.

Computational efficiency is improved by incorporating concepts of experimental design in the stochastic simulation. In particular, significant variance reduction of the estimator of the mean is achieved by using Latin Hypercube Sampling, as compared to simple random sampling.

The root-sum-square expression with sensitivity coefficients is investigated as an estimate of the variance of a composite parameter. For the calculational range of interest, it is found to give a reasonable approximation to the relative standard deviation of the output parameter (DNBR).

INTRODUCTION

In a modern pressurized water reactor (PWR), instrumentation and control systems are provided for the surveillance of both reactor systems and variables over their anticipated ranges of normal operation, for moderate frequency events (MFEs) and for accident conditions as appropriate to ensure adequate safety. The protection system is designed so that specified acceptable fuel design limits (SAFDLs) are not exceeded as a result of a MFE, by automatically initiating the operation of appropriate systems, including the reactivity control systems. The protection system is also designed to sense accident conditions and to initiate the operation of systems and components important to safety. An MFE corresponds to those conditions of normal operation that are expected to occur one or more times during the life of a nuclear steam supply system (NSSS).

The linear power density and the thermal-hydraulic conditions of the reactor core are physical characteristics which are important to evaluate in order to prevent damage to the fuel. Restrictions on these characteristics are placed by Combustion Engineering (C-E) on their PWRs by specifying fuel design limits on peak linear heat rate (PLHR) and departure from nucleate boiling ratio (DNBR). The PLHR in the limiting fuel pin shall not be greater than the value corresponding to the centerline fuel melting temperature, and the DNBR limit is specified at a value such that the probability of a departure from nucleate boiling event is acceptably small.

The design bases for a C-E PWR require that sufficient thermal margin be maintained under conditions of normal operation to preclude the violation of specified fuel design limits in case of an MFE. When considering such events, initial process conditions are assumed to be within the limits designated in the plant specifications. Safety analyses must demonstrate that anicipated transients initiated within process limits at any time during the core life will not violate the minimum DNBR and the PLHR limit.

Two surveillance functions are performed in a reactor. The first function is called protection, which is primarily to determine the operational status of the reactor core and provide a trip input to the reactor trip system whenever the DNBR or the PLHR reaches a calculated set point. These trips are designed to prevent fuel damage during an MFE and normally play no role in the prevention of postulated accidents, although they do provide the initial response to mitigate the consequences of some design basis accidents.

The second function is termed monitoring. For protective systems to function as intended in the design, it is necessary that NSSS parameters be maintained within established operating limits (OLs). For example, inferred DNBR and PLHR values are compared to their respective OLs. The OLs, in turn, are taken as initial conditions for various transient events. Analyses from these initial conditions are used to establish that acceptable consequences of the event occur. Monitoring systems are also provided to advise the operator of current margins to operating limits.

In the C-E PWR, these two key protection functions are performed by Core Protection Calculators (CPCs).[1] Monitoring is performed by the operator with the assistance of a Core Operating Limit Supervisory System (COLSS).[2] The CPCs are four redundant digital computers which acquire data from plant process sensors and control element assembly (CEA) position sensors and perform the required calculations. Each CPC provides trip inputs to the reactor trip system when the trip set points are exceeded. COLSS is a software system provided in the plant monitoring computer to assist the operator in maintaining normal operation within the process limits assumed for the CPC system protective functions.

An example is the operating limit on the maximum PLHR during normal operation of the reactor core. The OL is generally the maximum PLHR that can be allowed prior to the postulated initiation of a loss-of-coolant accident (LOCA) so that analysis of the latter will still show acceptable peak clad temperatures and other consequences.

The monitoring and protection systems observe the nuclear and thermal characteristics of the reactor core and have the mission either to alert the operator as core operating limits are approached or to initiate trip before fuel design limits are exceeded.

The digital protection system provides on-line routines for synthesis of the power distribution in the core and evaluation of the DNBR using measured inputs from (1) ex-core nuclear flux monitors, (2) CEA position indicators, and (3) other sensor data, such as core inlet temperature, primary system pressure, and coolant pump speed. The COLSS system employs available in-core detector signals to synthesize a hot pin power distribution for the reactor core. Inputs to both the reactor core monitoring (COLSS) and protection (CPC) systems consist of both analog and digital signals. The analog signals, consisting of sensor signals, are converted to digital signals by means of an analog-to-digital converter.

The calculations performed by COLSS and the CPCs are carried out at the plant. The on-line algorithms are a simplified version of the off-line design procedures. The simplification provides the reduced running time required for on-line processing but results in some loss in accuracy relative to the design procedures. The calculated thermal margin results are compensated for this loss in accuracy through the use of penalty factors, whose magnitudes are sufficient to result in conservative thermal margin calculations relative to more rigorous calculations used in the design.

In addition to modeling inaccuracies and analog-to-digital conversions errors, there are a variety of uncertainties arising from various sources which are associated with the inferred DNBR and PLHR thermal output parameters. For example, they may arise from measurement inaccuracies, calibrations errors, stochastic events, or signal processing errors. The calculational, measurement, and processing uncertainties must be factored into any thermal margin assessment of reactor operation.

A common practice has been to combine these uncertainties, be they random or systematic, in a multiplicative fashion, to produce an overall conservative result. It is clear, however, that this approach produces a new result which is overly conservative and that it is possible to demonstrate significant gains in available thermal margin from the application of statistical techniques.

1. Combustion Engineering, Inc., *CPC, Assessment of the Accuracy of PWR Safety System Actuation as Performed by the Core Protection Calculators*, Report CENPD-170, Windsor, Conn., July 1975.

2. Combustion Engineering, Inc., *COLSS, Assessment of the Accuracy of PWR Operating Limits as Determined by the Core Operating Limit Supervisory System*, Report CENPD-169, Windsor, Conn., July 1975.

To ensure that the design objectives of reactor trips for high local power density and low DNBR are achieved, the trip set points must account for uncertainties associated with modeling and calculational approximations, in addition to those due to sensor measurement and calibration errors. By treating these uncertainties statistically, it is possible to produce an overall uncertainty factor that is less restrictive than the multiplicative factor when evaluating the trip set points but still conservative in the determination of the thermal margins. A reduction in the net uncertainty is what will provide the future gain in available thermal margin, which can be used to improve the performance capabilities of an NSSS.

THERMAL HYDRAULICS AND NEUTRONICS

For currently operating C-E plants, the thermal margin design code is COSMO,[3] which is an open hot channel code. The CPCs currently use a simplified fast-running version of this code, called CPCTH. The latter uses a closed-channel model that does not explicity take into consideration the divergent cross-flow between the hot channel and the neighboring channels. To account for this in the CPC input, an adjustment is made to the mass velocity input to the applicable alogrithm, such that when all other system conditions are the same, the DNBR predicted by the closed-channel calculation (CPCTH) is equal to the minimum DNBR predicted by COSMO.

The CPCs compute thermal-hydraulic conditions in the hot channel using a snapshot of both directly monitored and calculated input values. The procedure used to assess the core minimum DNBR involves the synthesis of a hot pin and hot channel power distribution, which is used in conjunction with values of primary system process parameters to calculate the DNBR. The DNBR calculation is done at steady-state conditions and is corrected for changes between calculations using dynamic updates. The steady-state calculation uses COSMO with the W-3 correlation or TORC with the CE-1 correlation for the calculation of DNBR. A DNBR limit of 1.3 is used for plants whose thermal design basis was the W-3 correlation, while a corresponding value of 1.19 is used in the case of the CE-1 DNB correlation. The DNBR is the ratio of the critical heat flux to the actual heat flux in the reactor core.

The CPCs compute the hot channel minimum DNBR and the limiting void fraction using the following inputs to the CPCTH algorithm:

1. core power,
2. coolant temperature at the core inlet,
3. primary system pressure,
4. core average coolant mass velocity,
5. integrated radial peaking factor, and
6. normalized hot pin axial power distribution.

The CPCTH code also includes tables of correction factors which are applied to the input to force the output to have adequate agreement with the results from the design code COSMO.

The calculational uncertainty associated with the CPC synthesized local power density is determined with reference to design calculations. A large number of power distributions are generated with three-dimensional core simulators. Ex-core detector responses for a variety of static and transient core power distributions are simulated. For each case, the CEA positions assumed in generating the power distributions of interest, together with the simulated detector signals, are then processed by a FORTRAN version of the CPC algorithms to produce a value of the maximum peaking factor. The CPC synthesized peaking factor (to which the PLHR is proportional) is compared with the corresponding value produced by the simulator to yield an estimate of the calculational accuracy.

UNCERTAINTY ANALYSIS

The objective of the analysis is to obtain the statistical distribution of important output variables, such as DNBR and PLHR, which are functions of many input variables. To accomplish this goal, the systematic errors have to be separated from the random processes. The latter are then treated by statistical models which appropriately describe the input variable uncertainties. Ultimately, the probability of exceeding a tolerance limit is used as the means of determining the available thermal margin.

The random inputs are treated as variates whose means are taken as the nominal design values and whose variances and probability densities are either known or assumed. The aim of the uncertainty analysis is to determine the composite probability density function (pdf) of the output variable and use it to predict the variability of the system output.

Because the thermal-hydraulic and neutronics codes used in predicting thermal margins and reactor power performance are quite complex, it is convenient to compute the output (response) for arbitrary

3. Combustion Engineering, Inc., *TORC, Computer Code for Determining the Thermal Margin of a Reactor Core*, Report CENPD-161, Windsor, Conn., July 1975. [COSMO is described in this document.]

sets of input values. The response surface of the output is generated using randomly selected inputs. Since the on-line codes used by C-E for monitoring and protection are fast running (e.g., CPCTH takes approximately 4 ms per case on the CDC 7600), it is possible to use them directly to generate the response surface.

Appropriate sampling techniques must be adopted to ensure that the response surface corresponding to an important output variable is adequately covered. An experimental design is chosen to permit an efficient empirical exploration of the response surface, one which uses as few computer runs as is practical to calculate the output as a function of the input.

Although moment generating techniques and other analytical shortcuts can be used to simplify certain aspects of the problem, a direct use of Monte Carlo or stochastic simulation experiments is particularly convenient in that the output distributions and other results may be interpreted with minimum ambiguity.

Crude Monte Carlo (CMC) or simple random sampling is a straightforward technique of sampling a set of input values according to the cumulative distribution function (cdf) of the uncertainty density of the input variable. This is carried out with the aid of a pseudo random number generator. The response is computed for each set of input values, and by repeating the process many times, the distribution of the output is obtained. Any desired precision in the result can be obtained by conducting sufficient trials.

For a given number of trials, stratified sampling techniques offer an improved means of covering the response surface. They are less likely than simple random sampling to miss statistical fluctuations in the output distribution. Factorial Stratified Sampling (FSS) and Latin Hypercube Sampling (LHS) are two techniques that have improved response calculations at reduced computer cost.[4] If I is the number of statistically independent input variables and k is the number of levels or intervals associated with each input, the classical factorial design involves $n = k^I$ computer runs. In the sequence of computer runs, a systematic selection of the nominal high and low values of each input variable is made. The FSS corresponding to the k^I factorial design can involve the same number of computer runs, but each interval of the input variable is sampled according to its appropriate pdf. The principal difference between a fractional factorial design and FSS is that in the latter, intervals are sampled to produce different values of the input variables for each computer run. In LHS the same number of intervals for each input variable as the number of computer runs is used. In this case, however, the sampling is carried out so that each interval for each input variable appears exactly once in the total design.

At C-E, the code JAIALAI has been developed to perform the stochastic simulation and the uncertainty analysis. Figure 1 contains a block diagram of the code. The inputs can assume any of the standard

4. M. D. McKay et al., *Report on the Application of Statistical Techniques to the Analysis of Computer Codes*, Report LA-NUREG-6526-MS, Los Alamos Scientific Laboratory, Los Alamos, N. Mex., October 1976.

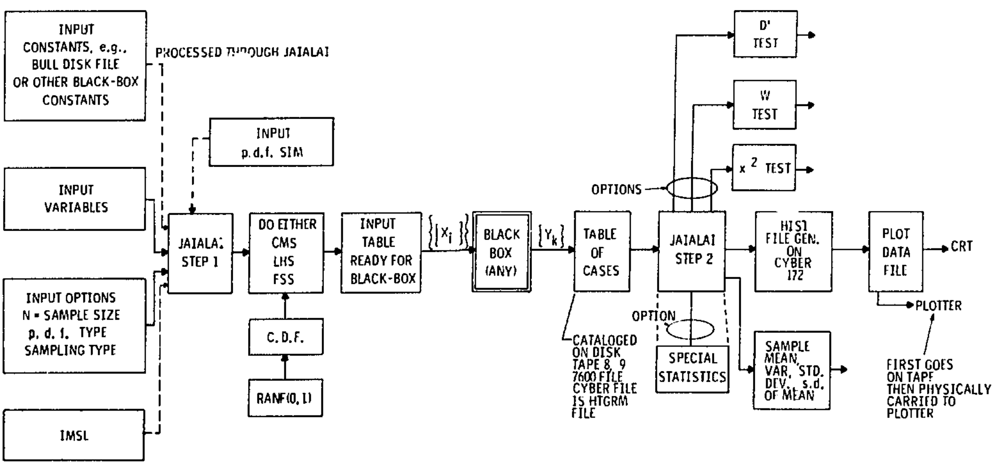


Fig. 1. JAIALAI stochastic simulation and uncertainty analysis code.

discrete or continuous distributions (uniform, Gaussian, Poisson, binomial, etc.) for their uncertainty descriptions. The International Mathematical and Statistical Library[5] (IMSL) has been attached recently to JAIALAI so that all of the distributions, routines, and tests in the IMSL are available for use in the analysis. The inputs are sampled by means of the random number generator RANF (0, 1) and the cumulative distribution function of the random variable. Three optional sampling schemes CMC, FSS, and LHS can be used in the experimental design.

The set of inputs is fed into the code of interest (black box) and the response is calculated. The output distribution is plotted on a histogram and sample statistics (mean, variance, etc.) are calculated. Various distributional tests (chi-squared goodness of fit and D' and W tests for normality), as well as other hypothesis and statistical tests, can be performed on the response. In "Special Statistics" the variance of the estimators is used to compare the efficiency of the sampling schemes in determining the output variables of interest.

A successful simulation depends to a large extent on how well an assumed pdf represents a physical input variable. A basic input variable, such as cold-leg temperature, can, itself, be dependent on many factors. For example, variations can result from

1. process noise or prompt-fluid-temperature fluctuations due to temperature eddies and other time-dependent effects;
2. dependence on the particular values of other basic input variables, such as pressure, temperature, and flow;
3. sensor- and measurement-related uncertainties (Systematic errors, e.g., those due to radiation damage or aging of insulation, are not included in the pdf construction and have to be taken into account outside the stochastic simulation.); and
4. signal processing uncertainties due to electrical noise pick-up, internal hum, and analog-to-digital conversion.

The composite pdf for the cold-leg temperature can be obtained from a stochastic simulation of its component parts. If there is a dependence on other input variables, then either conditional probabilities have to be used, or an effective temperature distribution has to be constructed. A similar procedure is used for the treatment of the other input variables.

Figure 2 contains a functional diagram of the CPC. The left-hand side of the diagram indicates the basic input variables and the right-hand side denotes the

5. International Mathematical and Statistical Libraries, Inc., *IMSL Library 3*, 6th ed., IMSL LIB03-0006VI, Houston, Tex., July 1977.

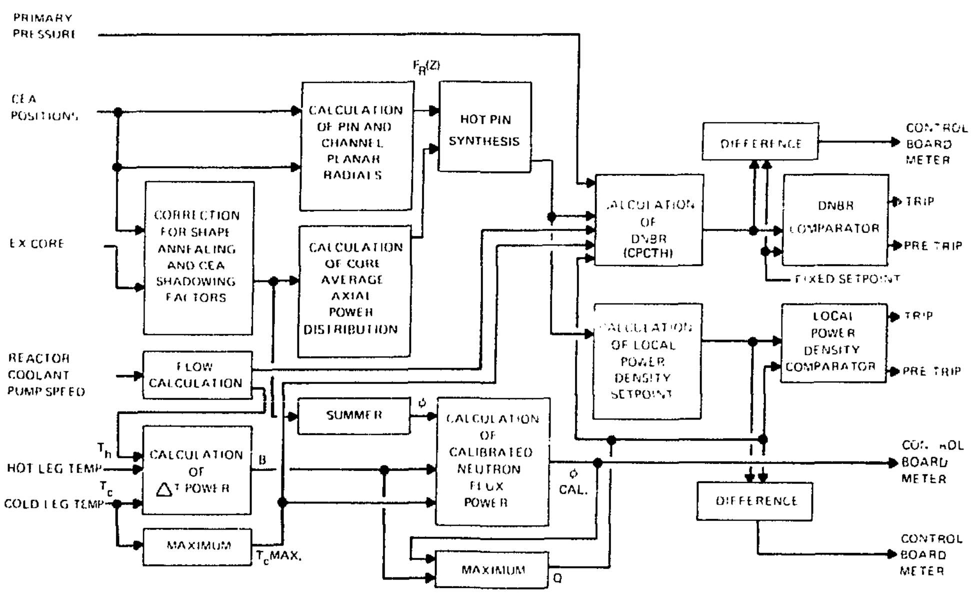


Fig. 2. Core protection calculator functional block diagram.

trip set points for DNBR and PLHR (or local power density, in this case). In between are the required neutronic and thermal-hydraulic calculations of power distributions and heat fluxes. Once the distributions of the basic inputs in Fig. 2 are specified, the stochastic simulation of the CPC can proceed and the margins to the set points can be determined.

STATISTICAL RESULTS

To demonstrate the potential gains in available thermal margin by using a statistical approach relative to a conservative multiplicative approach, consider the CPC depicted in Fig. 2. The basic six inputs, primary pressure, CEA positions, ex-core detector readings, speeds of the coolant pumps, hot-leg temperature, and cold-leg temperature, are subjected to independent random variations. The pdf used for each of the inputs, together with the corresponding mean and standard deviation, is given in Table I.

Table I. Probability distributions, means, and standard deviations for the basic input variables used in the uncertainty analysis of the CPC calculation of DNBR

Input variable	Distribution	Mean	Standard deviation
Cold-leg temperature	Gaussian	553.5°F	0.48°F
Hot-leg temperature	Gaussian	614.6°F	0.52 F
Primary pressure	Gaussian	2250 psi	6.5 psi
Ex-core detectors (three)	Gaussian	0.3171	0.0045
		0.4195	0.0060
		0.3006	0.0043
Pump speed (four pumps)	Uniform	1.0000	0.0058
CEA position	Uniform	Unrodded	1.79 in.

The static DNBR is calculated for the given set of steady-state conditions indicated by the mean values of input variables. The variations are then stochastically simulated to yield the output distribution for minimum static DNBR. For simple random sampling (CMC) based on a sample size of 1000, the result is given in Fig. 3. The output distribution for the uncertainty in the static DNBR is Gaussian in appearance and passes the standard normality tests. As indicated in Fig. 3, this sample size of 1000 yields estimates of 1.635 and 0.0564, respectively, for the mean and standard deviation of the minimum static DNBR.

Essentially equivalent results were obtained using LHS sampling. In this case, the range of each variable was partitioned into 100 intervals of equal probability content and only 100 computer runs were made. This indicates a factor of 10 in computer time can be saved with the use of LHS relative to CMC, which is consistent with an investigation that was made on the following simple algorithm:

$$Y = x_1 + x_2 + x_3 \; .$$

Each variate, x_i, was assumed to be normal and independent. The three optional sampling schemes in JAIALAI were applied in the stochastic simulation of the distribution of Y. The total sample size was 4000, which meant 40 replications of a stratified sample with 100 levels for LHS. Significant variance reduction was achieved in using LHS relative to CMC. The precision, as measured by the variance of the estimator of the mean, improved by a factor of over 200. The FSS scheme produced intermediate gains relative to CMC, but not nearly as good as the LHS did.

The absolute estimate, 1.635, for the mean DNBR should not be regarded as significant, because penalty factors already exist in the current version of the CPC code. These penalties are deterministic in nature and account for both calculational and systematic errors, as well as for instrumentation uncertainties. Therefore, this code version is being used for demonstrational convenience, but an examination of these penalty factors will eventually be undertaken.

The results obtained from the stochastic simulation of the basic input variables can be used to demonstrate significant gains in DNBR margin. Using the sample mean and standard deviation of the output distribution, a lower tolerance limit, t, for the simulation can be calculated for the minimum static DNBR:

$$t = x - ks \; .$$

At the 95% confidence level, the k factor for a sample size of 1000, which provides a lower limit that is exceeded by at least 95% of a normal population, is 1.727.[6] The corresponding lower tolerance limit, therefore, is

$$t_{95/95} = 1.635 - 1.727(0.0564) = 1.538 \; .$$

6. D. B. Owen, *Factors for One-Sided Tolerance Limits and for Variables Sampling Plans*, Sandia Corporation Monograph SCR-607, Albuquerque, N. Mex., March 1963.

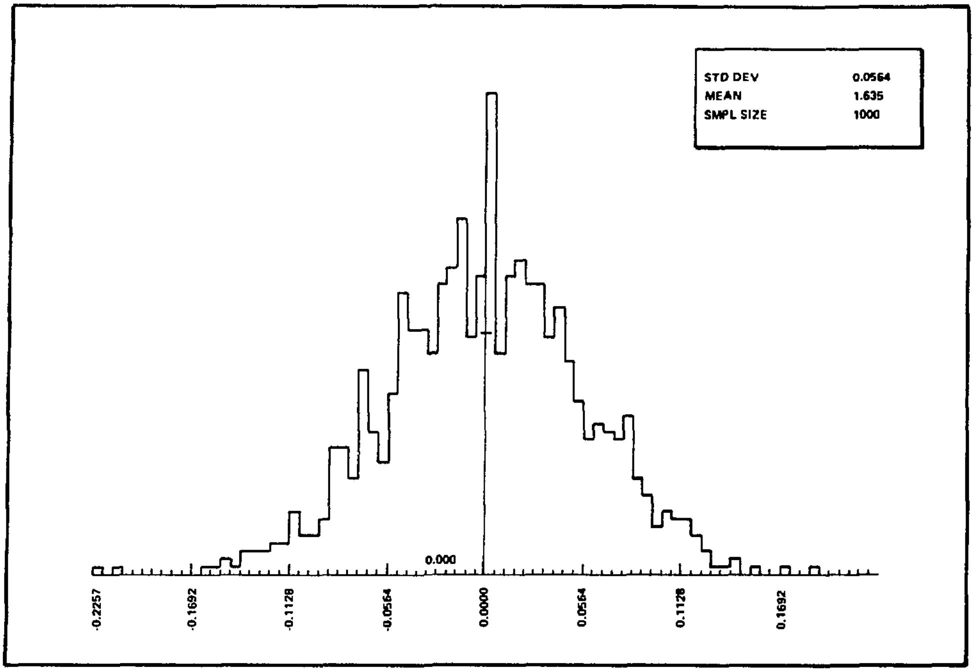


Fig. 3. DNBRST—Minimum static DNBR.

The statistical analysis of the input uncertainty distributions thus yields a lower limit, such that at a 95% confidence level, there is at least a 95% probability that the minimum static DNBR will exceed 1.538.

On the other hand, if the conservative approach of evaluating the change in DNBR due to a two-standard-deviation change in each of the inputs is adopted and if the total change in DNBR is taken as the cumulative sum of the individual-change magnitudes, then the lower limit is reduced to 1.367. This reduction is equivalent essentially to multiplying individual penalty factors together, also known as a "worst-case combination."

For example, the percent change in DNBR per percent change in the hot-leg temperature (T_h) is evaluated to be 27.7. Using this sensitivity coefficient in conjunction with a percent change of 0.169 in hot-leg temperature (corresponding to a two-standard-deviation change in the hot-leg temperature relative to its mean), this component contributes 27.7(0.169) = 4.69% to the change in DNBR. Summing all similar effects with their corresponding sensitivities yields a total change in DNBR of 16.4%. This is equivalent to the percent reduction of the DNBR from the original 1.635 to the conservative lower limit of 1.367.

To allow for the uncertainties in these basic inputs, the trip set points in the CPC are multiplied by penalty factors. The conservative estimate of the overall penalty factor to be used in the DNBR calculation is 1.164 (i.e., 1 + 16.4%). The statistical estimate of the penalty factor is 1.059, which corresponds to a (1.635 − 1.538)100 1.635 = 5.9% change in DNBR. The statistical approach, therefore, provides a 10.5% gain in margin to the DNBR limit. This improvement in the core operating limit to the specified acceptable fuel design provides a direct increase in the power margin available for reactor operation.

It is interesting to note that the relative standard deviation of the output distribution for the minimum static DNBR, $\delta\sigma_D$, is close to the value calculated from a root-sum-square expression with sensitivity coefficients. A first-order Taylor's series expansion yields

$$\delta\sigma_D^2 = \sum_i s_i^2 \delta\sigma_i^2 ,$$

where covariance and higher-order terms have been neglected. The sensitivity coefficients, s_i, are evaluated for each of the input variables. For example,

$$s_{T_h} = \Delta D \;\; \Delta T_h = 27.7$$

is the relative change in DNBR with respect to the relative change in the hot-leg temperature. The relative standard deviations, $\delta\sigma_i$, are the ratios of the standard deviations to the means of the input variables i.

Using the appropriate sensitivity coefficients and relative standard deviations of the inputs, the root-sum-square value is $\delta\sigma_D = 3.39\%$. The corresponding value from the stochastic simulation (see Fig. 3) is $100(0.0564)\ 1.635 = 3.45\%$. The closeness of these two results is probably due to covariances and higher-order effects being negligibly small in the parameter space where the calculations are being performed. In this kind of situation, the simple Taylor's expression can give comparatively good estimates of the relative standard deviation of the composite parameter.

CONCLUSIONS

A statistical treatment of uncertainties can produce significant gains in thermal margin relative to currently used multiplicative approaches, and the power performance capability of a reactor can be improved by virtue of these gains.

A stochastic simulation of the basic input variables to the Core Protection Calculator in the C-E reactor protection system was carried out, and the results obtained from the response distribution were compared with conservative estimates of the DNBR uncertainty. Specifically, it was demonstrated that the margin to the limit on DNBR was increased by approximately 10.5% for a typical set of reactor operating conditions.

The analyis emphasized uncertainties in temperature, pressure and flow, and in CEA position and excore detector readings. Uncertainties in the design parameters, calculational methods, modeling, and fabrication were not included. A proper statistical treatment of these additional aspects should produce further gains in thermal margin.

Computational efficiency was improved by incorporating concepts of experimental design. In particular, significant variance reduction of the estimator of the mean was achieved by using LHS, when it was compared to simple random sampling.

The root-sum-square estimate of the variance of a composite parameter was investigated. For the calculational range of interest, it was found that it provided a reasonable approximation to the relative standard deviation of the output parameter (DNBR).

These methods and procedures are similarly being applied to the monitoring system COLSS, with equivalent margin gains indicated and expected.

ACKNOWLEDGMENT

The computational assistance provided by R. A. Pauze, D. E. Uhl, and K. C. Wei in carrying out this analysis is gratefully acknowledged.

Problem Presentations and Discussions

Assessment of Oil-Shale Development—A Problem in Statistical Design

Donald R. Dietz

U.S. Fish and Wildlife Service
Grand Junction, Colorado

Lawrence K. Barker, Eric G. Hoffman, and Robert L. Elderkin

U.S. Geological Survey
Grand Junction, Colorado

ABSTRACT

The purpose of the Federal prototype oil-shale leasing program is to initiate a carefully controlled, environmentally sensitive effort to determine the economical, technical, and environmental feasibility of developing the billions of barrels of oil locked in the kerogen-rich marlstone of the Green River Formation located in the northeastern part of Utah and northwestern part of Colorado. These goals depend on the complete and accurate assessment of related environmental impacts, and the design and evaluation of mitigating technology. A major problem is the development of a statistically sound monitoring program that will permit early prediction of significant impacts.

INTRODUCTION

In 1973, the U.S. Department of the Interior leased four 5000-acre tracts for commercial oil-shale development under the Prototype Oil Shale Program. This program was designed to test the feasibility of producing shale oil commercially and to determine the associated environmental costs and impacts accompanying commercial operation. To date, operators of the lease tracts have completed collection of two years of intensive baseline environmental data covering both biotic and abiotic parameters. The lessees have also submitted a detailed development plan describing procedures for bringing each tract to commercial production. The cost of this effort has exceeded $25 million. Plans for the Colorado tracts have been approved but actions on the Utah tracts are currently in litigation.

Presented herein are some of the problems encountered in developing effective environmental monitoring and data management programs. Solution of these problems is critical to the design of currently developing monitoring programs to determine the degree of stress and impact occurring on biotic and abiotic aspects of the tracts during commercial development. This problem is complicated by the need to compare baseline data that were gathered in a multiplicity of formats in the past with data to be collected during development monitoring.

ENVIRONMENTAL SETTING

Physiography

Federal oil-shale lease tracts C-a and C-b are in the Piceance basin of northwestern Colorado; tracts U-a and U-b are in the adjoining Uinta basin of eastern Utah (Fig. 1). Both basins are geologic structural features that have been eroded into arid upland plateaus, intricately dissected by intermittent streams. Broad, flat, soil-covered divides give way sharply, over float and outcrop-

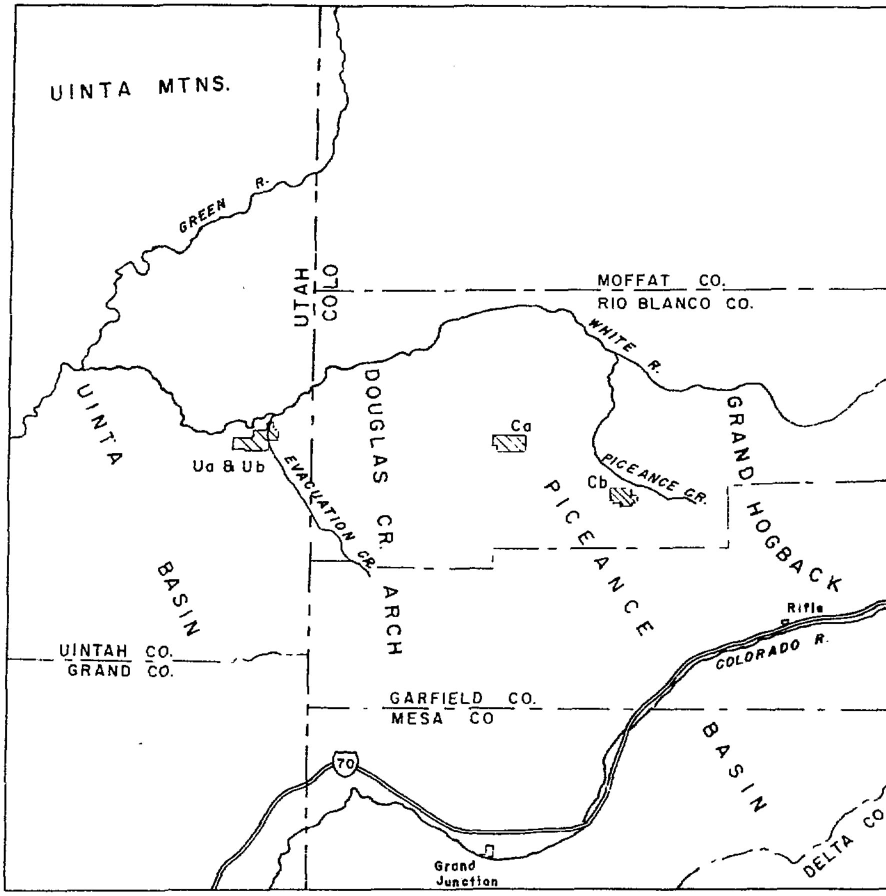


Fig. 1. Location of prototype oil-shale lease tracts C-a, C-b, U-a, and U-b.

strewn slopes, to narrow alluvium-filled drainage bottoms that terminate in distinct alluvial fans along the principal trunk streams. Relief typically ranges from a few hundred feet to 600 ft.

Both the Uinta and Piceance basins terminate to the south in cliffs along the Colorado River, cut by narrow canyons that have incised the several thousand feet of sedimentary rocks of the Upper Cretaceous Mesaverde Group and the Wasatch, Green River, and Uinta formations of Eocene age. The Green River and Uinta formations contain kerogen-rich marlstone,[1] which is a potential source of billions of barrels of shale oil. The north limits of the basins are defined roughly by the White River in Colorado and the Uinta Mountains in Utah. The eastern edge of the Piceance basin is marked by the Grand Hogback, while the western edge of the Uinta basin, for economic (oil-shale) considerations, is generally delineated by the Green River. The two basins are separated along the Colorado-Utah border by the deeply eroded Douglas Creek arch.

Climate

Climatic conditions of the Piceance and Uinta basins are semiarid. The influence of the Cascade

1. John B. Weeks, George H. Leavesley, Frank A. Welder, and George J. Saulnier, Jr., *Simulated Effects of Oil-Shale Development on the Hydrology of Piceance Basin, Colorado*, U.S. Geol. Survey Prof. Paper 908, 1974.

and Sierra Nevada mountains to the west and the Rocky Mountains to the east creates conditions characterized by abundant sunshine, hot summers, cold winters, low relative humidity, light precipitation, and large diurnal temperature variations. Temperatures range from 40°C in the summer to −40°C in the winter, with a mean annual temperature of 7°C. Very strong vertical temperature differences of more than 46°C have been observed during winter between the valley bottoms and the surrounding plateaus. The frost-free season varies from 120 days in the basins to about 50 days in the bordering mountains.[1] Annual precipitation ranges from less than 12 in. in the basins to as much as 25 in. above 8000 ft on the surrounding highlands.[1] Snowfall accounts for about 40% of this precipitation, and the remainder is from rainfall during the summer when intense thunderstorms frequent the area. Sixty percent of the days are either cloud-free or only partly cloudy; winds are generally southwesterly and average 7 mph. Air quality in the basins is generally excellent, and acute perception of distant objects is commonly limited only by terrain. Photographic measurement of visual range has been as much as 100 miles.

GEOLOGY

The Piceance and Uinta basins are broad, asymmetric, northwest-trending structural basins filled with deposits of sandstone, siltstone, and marlstone laid down mainly in shallow, warm, alkaline lakes teeming with algal and fish life. The lakes inundated the contiguous corners of Utah, Wyoming, and Colorado 70 to 40 million years ago. The deposits have since been uplifted and gently folded into broad physiographic basins. Stratigraphy within the basins is uniquely uniform, disrupted generally along the basin margins by widely separated, near-vertical, northwest-trending graben faults with as much as 200 ft of throw. The broad structure is cut by basinwide joint and fracture systems of little displacement.

A vertical slice through this basin section (Fig. 2) reveals strata dipping gently toward the basin centers. On top is a shallow mantle of moderately alkaline alluvium and colluvium generally light-colored, flaggy, and loamy. This "topsoil" ranges from a few inches thick on the drainage divides to several tens of feet thick in the valley bottoms. Immediately under the soil mantle are several hundred to a few thousand feet of massive, yellowish-brown to light-gray sandstone and siltstone of the Uinta Formation. Of principal interest to the oil-shale industry is the underlying (1500 ft or more) kerogen-rich, dark-gray marlstone of the upper part of the Green River Formation from which oil can be extracted by heat.

HYDROLOGY

The Piceance and Uinta basins are drained by creeks tributary to the Colorado River or to the White River, a major tributary to the Colorado River. Annual discharge from the White River basin averages 510,000 acre-ft: about 16,000 acre-ft from the Piceance basin and a much lesser amount from the Uinta basin. Most of the runoff occurs in late spring and early summer from melting snowpack. Summer thunderstorms can also generate violent, but short-term runoff. Approximately 120,000 acre-ft is used within the basins for irrigation; an additional 500 acre-ft is diverted for domestic use in nearby communities.

BIOLOGY

The distribution of flora and fauna of both the Piceance and Uinta basins, which is typical of the intermountain region, is particularly apparent at higher elevations on the rims that border the basins on the south and in the highlands of the Douglas Creek arch that separates the basins.

Subalpine and montane forests are common at about 8000 ft. At intermediate elevations, mountain shrub and pinyon-juniper dominate. At lower elevations in both basins, sagebrush, desert shrub, greasewood, meadow, and riparian species predominate. In the higher and wetter portions of the Piceance basin, predominant sagebrush is interspersed with pinyon-juniper and mountain shrub. Desert and salt-desert shrub in the Uinta basin are normally interspersed with greasewood along the drainages.

The Piceance basin is world famous for its large migratory herd of mule deer. Wild (feral) horses roam both basins, while domestic livestock production is one of the main endeavors of rural communities.

Many medium-to-small wildlife species, including birds, utilize the region for migration, nesting, and winter habitat. Fish are not abundant in the area's rivers, but several threatened and/or endangered species inhabit the Colorado, White, and Green River drainages.

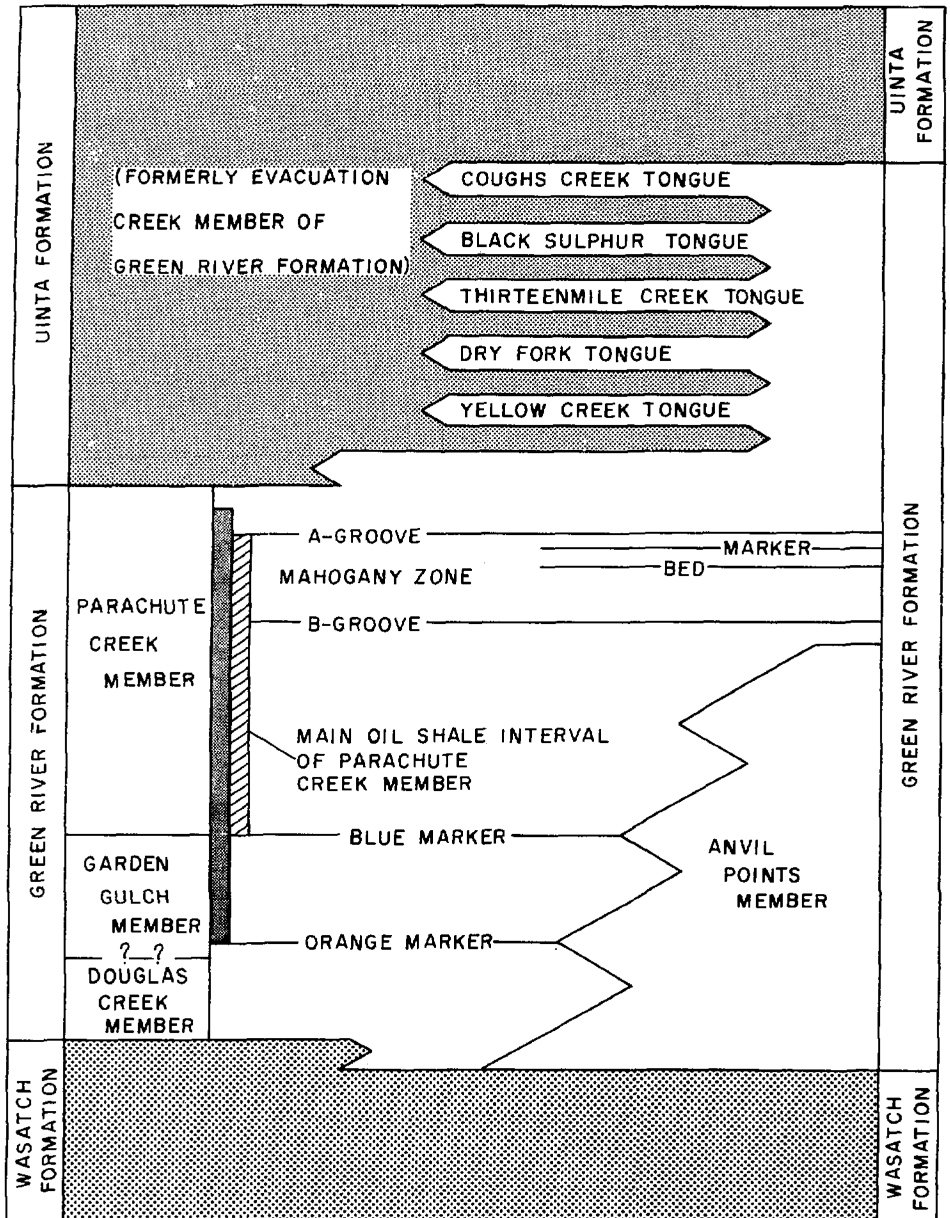


Fig. 2. Generalized stratigraphic column of Eocene Formations in the Piceance basin. Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project—Terrestrial Ecology Baseline Studies, Annual Report for Tract C-a,* March 1976.

DEVELOPMENT

Present plans call for development of the two Colorado tracts by modified in-situ retorting processes, and for development of the two Utah tracts by room-and-pillar mining with above-ground retorting. At this writing (October 1977), the Utah leases have been temporarily constrained from development by the Federal District Court because of legal questions of land ownership and existence of conflicting overlying mining claims.

The modified in-situ process involves underground mining and underground retorting to produce shale oil (Figs. 3 and 4). Approximately 20% of the oil shale within a given retort is directly mined, and the remainder is converted to rubble in place to create a bulked-full retort. In this manner the permeability needed for flow-through of the injected gases required to maintain combustion and for removal of the products formed is achieved. The rate of retorting is controlled by regulating the volume, pressure, and oxygen content of the injected air and diluting steam or recycled gases and by varying the back pressure on the gas outlet shaft.

To initiate combustion, burners are placed on top of the rubble. Air is either pulled through from the top by exhaust blowers or fed by the use of air compressors. When reaction temperature (932°F) is reached, the burners are turned off. Steam or other gases are then introduced along with air to maintain burning at a desired temperature and to control the rate of flame-front advance (Fig. 5).

Product oil and water are condensed on the cooler unretorted shale at the bottom of the retort chamber and pumped to the surface. Off-gases are exhausted through blowers to a scrubber system above ground where the gas is contacted with a circulating water stream to remove entrained dust and oil particles. The scrubbed gas is purified by removing the oil and water by compression and the sulfur compounds by a Stretford or similar process. Purified gas is then used to fuel low-Btu/lb boilers for steam production and possibly gas turbine electric power generators.

BASELINE AND DEVELOPMENT MONITORING

Many of the statistical problems in developing and operating the production monitoring program can be anticipated from review of the data gathered during the baseline monitoring program. The lease required the lessees to conduct a two-year baseline program before beginning any construction. While the two-year baseline program was intended to establish "baseline conditions" in the natural environment from which significant perturbations could be measured, it has become evident that treatment-control designs are necessary to separate development effects from random natural changes. This necessity is especially valid for dynamic parameters such as faunal populations. The premise during baseline data collection was to ensure sufficiently complete and accurate parameter evaluation so that valid statistical comparison could be made with data gathered during development.

Conducting a baseline program where many disciplines (air quality, meteorology, hydrology, geology, biology, etc.) must be interrelated proved to be a major and expensive undertaking.

The environmental stipulations of the oil-shale lease state: "The lessee shall conduct the monitoring program to provide a record of changes from conditions existing prior to development operations, as established by the collection of baseline data."[2] Conditions for approval of the lease-required detailed development plan also state: "The environmental monitoring plan shall be revised as needed, based on the analysis of the final baseline report—submitted for review and approval by the Mining Supervisor prior to commencement of commercial development."[3] The lessees, in conjunction with the Area Oil Shale Office, are attempting to develop an effective environmental monitoring program based on interpretation of baseline data. One approach being used for design of the development monitoring program is a series of matrices. This procedure compares individual engineering actions against specific biotic and abiotic parameters in consideration of four main criteria: (1) magnitude, direction, and duration of impact; (2) importance (ecological, political, and economic) of the parameter impact; (3) measurability of the impact; and (4) cost effectiveness of the measurement effort required. Comparisons are ranked from high to low based on baseline data analyses and on existing literature and professional judgment. Table 1 illustrates a representative part of one matrix.

2. U.S. Bureau of Land Management. *Oil Shale Lease, Tract C-a*, Serial No. C-20046, 1975.

3. Peter A. Rutledge, Area Oil Shale Supervisor, written communication, August 1977 and September 1977.

Fig. 3. Idealized above-ground view of the modified in-situ operation. Source: Ashland Colorado, Inc., and Occidental Oil Shale, Inc., *Mining Plan for Ancillary Development*, Ralph M. Parsons Co., Parsons-Jurden Division, June 1977.

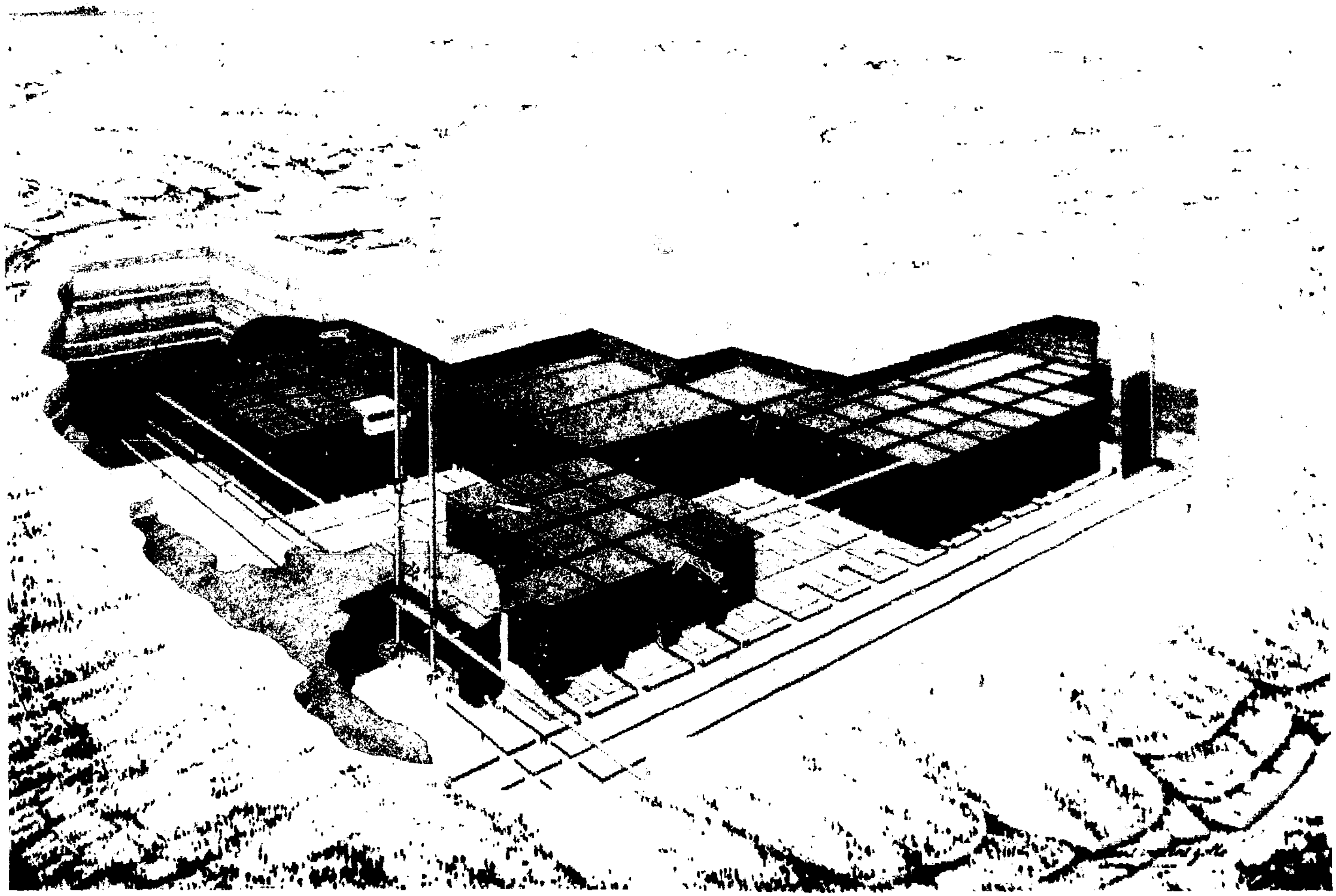


Fig. 4. Idealized below-ground view of the modified in-situ operation. Source: Ashland Colorado, Inc., and Occidental Oil Shale, Inc., *Master Plan for Ancillary Development*, Ralph M. Parsons Co., Parsons Jurden Division, June 1977.

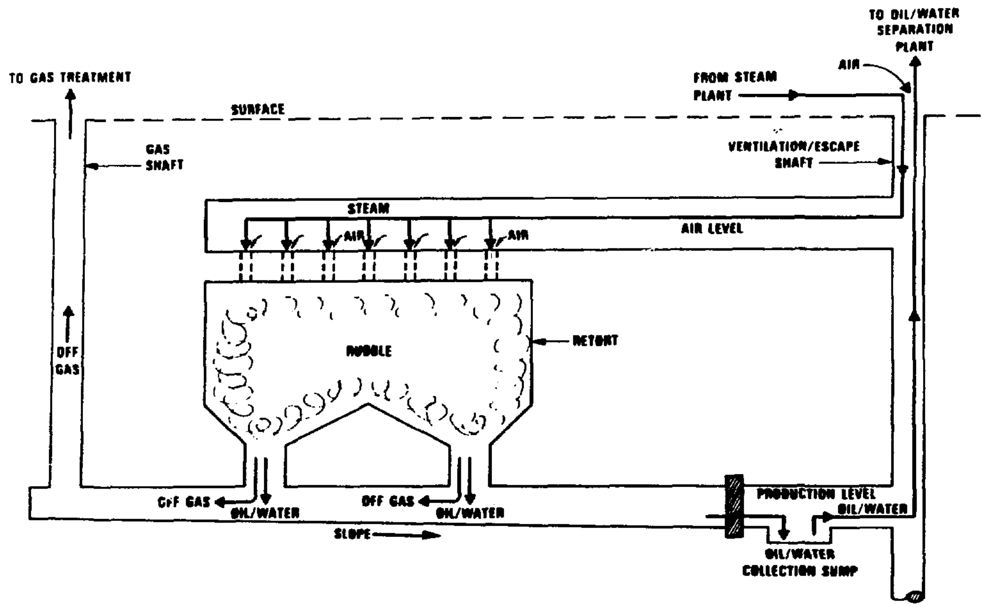


Fig. 5. Schematic diagram of a modified in-situ retort and related air and production mine levels. Source: Ashland Colorado, Inc., and Occidental Oil Shale, Inc., *Mining Plan for Ancillary Development,* Ralph M. Parsons Co., Parsons-Jurden Division, June 1977.

Table 1. A representative part of the commercial stage "cause-effect" matrix performed for the Rio Blanco Oil Shale Project for oil-shale development on tract C-a

Affected factor: vegetation	Construction[a]									
	Mine	Surface retorts	Underground retorts	Support facilities	Conveyor belts	Impoundments	Roads	Compressors	Disposal pile	Cumulative construction
Productivity	1/2	2/2	1/1	1/2	1/2	1/2	1/2	1/2	3/3	5/3
Range condition	1/2	1/2	1/1	1/2	1/2	1/2	1/2	1/2	2/3	2/3
Community composition	1/2	2/2	1/1	1/2	1/2	1/2	1/2	1/2	3/3	3/3
Distribution	1/2	2/2	1/1	1/2	1/2	1/2	1/2	1/2	3/3	3/4
Trace metal content	1/1	1/1	1/1	1/2	1/2	1/1	1/1	1/1	1/1	1/1
Cover density	1/2	2/2	1/1	1/1	1/2	1/2	1/2	1/2	3/3	4/5
Browse condition	1/2	2/2	1/1	1/2	1/2	1/2	1/2	1/2	3/3	3/3

[a]Construction rankings were defined as follows:

Importance

1. None, not applicable
2. Slightly important
3. Moderately important
4. Very important
5. Extremely important

Severity and magnitude

1. None, not applicable
2. Slightly severe, small
3. Moderately severe, medium
4. Very severe, large
5. Extremely severe, quite large

Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project – Revised Detailed Development Plan for Tract C-a,* vol. 3, May 1977.

SUMMARY OUTLINE

The above approach resulted in the identification of some specific parameters having statistical problems which are outlined below:

1. Biotic
   A. Flora
      1. Vegetation type distribution
         a. Methology—Color infrared; photos repeated annually.
         b. Baseline data and analysis—To precede development.
         c. Statistical tests—None proposed to date.
         d. Problems—Should changes be measured by visual comparison of photos; what constitutes a significant change; and what statistical methods are applicable?
         e. Possible solutions—Not yet developed.
      2. Range productivity and utilization
         a. Methodology—Tract C-a will use double-sampling method (estimate and clipped plots) on ten plots on ten transects on three vegetation types annually. Plot is 9.6 ft^2; two of each ten plots are caged from April to September. Plant types are separated by species and weighed; correction factors are applied to the green weight estimates.
         b. Baseline data and analysis—C-a baseline data for forage production, which combined all herbage weights for each vegetation type, are shown in Table 2.
         c. Statistical tests—H_0: no significant difference in vegetative productivity for

Table 2. Standing-crop estimates for major shrub species in each of the intensive study plots, 1976[a]

Shrub species	Plot number 1	2	3	4	5	6
Amelanchier sp.						
April	111 ± 24	73 ± 16	41 ± 9		101 ± 22	118 ± 25
September	121 ± 34	80 ± 23	44 ± 13		110 ± 31	128 ± 36
Artemisia tridentata						
April	1541 ± 352	214 ± 49	2026 ± 375	9015 ± 1871	54 ± 10	419 ± 77
September	1653 ± 325	229 ± 45	2429 ± 749	15926 ± 3559	65 ± 20	502 ± 155
Ceratoides lanata						
April				34 ± 5		
September				33 ± 6		
Cercocarpus montanus						
April	211 ± 59	49 ± 14			85 ± 24	
September	186 ± 53	56 ± 12			96 ± 21	
Chrysothamnus nauseosus						
April	116 ± 20	287 ± 51	11 ± 2	29 ± 5	4 ± 1	43 ± 8
September	186 ± 40	462 ± 98	17 ± 4	46 ± 10	7 ± 2	69 ± 15
Juniperus osteosperma						
April	25 ± 3	90 ± 12	4 ± 1		9 ± 1	22 ± 3
September	19 ± 3	67 ± 11	3 ± 1		6 ± 1	17 ± 3
Pinus edulis						
April	46 ± 9	70 ± 14	12 ± 3		40 ± 8	20 ± 4
September	54 ± 9	84 ± 14	15 ± 3		47 ± 8	24 ± 4
Purshia tridentata						
April	288 ± 68	88 ± 21		1 ± 0.2	33 ± 8	2 ± 0.5
September	430 ± 88	132 ± 27		2 ± 0.3	50 ± 10	3 ± 1
Total						
April	2338 ± 535	871 ± 177	2094 ± 390	9079 ± 1881	326 ± 74	624 ± 117
September	2702 ± 552	1110 ± 230	2508 ± 770	16007 ± 3575	381 ± 93	743 ± 214

[a]Plus and minus values are equal to the standard error of the mean. Values in kilograms per hectare; 1 hectare = 2.471 acres.

Source: Ashland Oil, Inc., and Occidental Oil Shale, Inc., *Oil Shale Tract C-b–Environmental Baseline Program Final Report (November 1974 through October 1976)*, vol. 4, 1976.

forage use within a vegetation type before or during initial development. Analysis by analysis of variance (AOV). Accuracy to be such that sample means will be within ±25% of population mean 90% of the time.

d. Problems—During initial development significant differences in forage use cannot be determined because of the many unknown or uncontrollable variables affecting animal use of the various habitats.

e. Possible solutions—Design the analysis to determine adequate sample size and presence of significantly different strata among plant production areas.

3. Browse condition and utilization

a. Methodology—Browse condition and utilization are estimated by the Cole Method[4] on randomly selected transects consisting of 25 individual shrubs. Transects are established in each vegetation type.

b. Baseline data and analysis—Examples of data summaries from the C-a baseline are shown in Table 3.

c. Statistical tests—H_0: no significant difference in browse utilization within a given vegetation type before or during development. Analysis proposed is an AOV. Accuracy sought is ±25% of the mean 90% of the time.

d. Problems—Determining statistical tests most useful for comparing effect of a disturbance on form and age-class percentage.

e. Possible solutions—Generally same as for productivity. Browse condition classes may be compared by chi square.

4. G. P. Cole, *Range Survey Guide*, U.S. National Park Service, Washington, D.C., 1963.

Table 3. Condition of seven principal browse species sampled in two predominant vegetation types during April 1976 for the Rio Blanco Oil Shale Project[a]

Species	Vegetation type	No. of plants sampled	Average percent utilized	Average percent available	Form class percentages[b]							Age class			
					1	2	3	4	5	6	7	Seedling	Young	Mature	Decadent
Service-berry	MB	643	11.1	98.4	76	11	2	11	1	ND	ND	ND	1	95	4
	PJ	202	69.1	99.8	19	31	48	1	ND	0.4	ND	ND	6	88	6
Sage-brush	MB	352	4.9	100	100	ND	ND	ND	ND	ND	ND	0.2	4	93	2
	PJ	328	7.2	100	93	7	ND	ND	ND	ND	ND	0.3	12	69	18
Pinyon pine	MB	17	6.8	94	88	ND	ND	11	ND	ND	ND	ND	23	76	ND
	PJ	259	12.4	67.8	42	6	1	37	8	3	ND	4	20	72	2
Rabbit brush	MB	22	9.3	100	95	5	ND	ND	ND	ND	ND	ND	ND	100	ND
	PJ	94	6.2	100	97	3	ND	ND	ND	ND	ND	ND	8	83	8
Bitter brush	MB	45	46.9	100	22	44	24	9	ND	ND	ND	ND	ND	100	ND
	PJ	153	73.8	100	3	40	56	ND	ND	ND	ND	ND	0.6	93	6
Mountain mahogany	MB	58	47.1	99.4	38	34	21	3	3	ND	ND	ND	7	90	2
	PJ	62	52.4	93.3	13	48	31	5	3	ND	ND	ND	ND	98	2
Juniper	MB	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	PJ	225	6.0	49.6	29	0.9	ND	65	2	ND	2	1	12	77	10

[a]MB, Mixed brush; PJ, Pinyon-juniper; ND, no data.

[b]Form classes:
1. All available; little or no hedging.
2. All available; moderately hedged.
3. All available; severely hedged.
4. Partially available; little or no hedging.
5. Partially available; moderately hedged.
6. Partially available; severely hedged.
7. Unavailable.

Source: Gulf Oil Corporation and Standard Oil Company. *Rio Blanco Oil Shale Project – Revised Detailed Development Plan for Tract C-a*, vol. 2, May 1977.

4. Revegetation of disposal sites—Herbaceous standing crop
   a. Methodology—Double-sampling method of Wilm, Costello, and Klipple[5] will be used on each site to estimate and correct standing crop production on fifty 1-m^2 plots (one of each ten plots is randomly selected for clipping, drying, and weighing). Results will be given in kilograms per hectare (0.89256 lb/acre).
   b. Baseline data and analysis—Lessees have ten years to demonstrate their capability of restoring disturbed sites to "like conditions" as determined during the baseline period. Current studies provide an example of field plot design (Fig. 6) and parameter methodology (Table 4).
   c. Statistical tests—H_0: no change in standing crop among years at any given site; no difference in standing crop in given year between sites; and no difference in standing crop among years between sites. To determine significance of changes in standing crop among years, sites, and year-site interactions, AOV will be used.
   d. Problems—Enough sites need to be selected on various macrosites to determine effects of slope, aspect, elevation, etc. No follow-up testing procedures are proposed if the null hypotheses are rejected.

5. H. G. Wilm, David F. Costello, and G. E. Klipple, "Estimating Forage Yield by the Double-Sampling Method," *Am. Soc. Agron. J.* 36: 194-203 (1944).

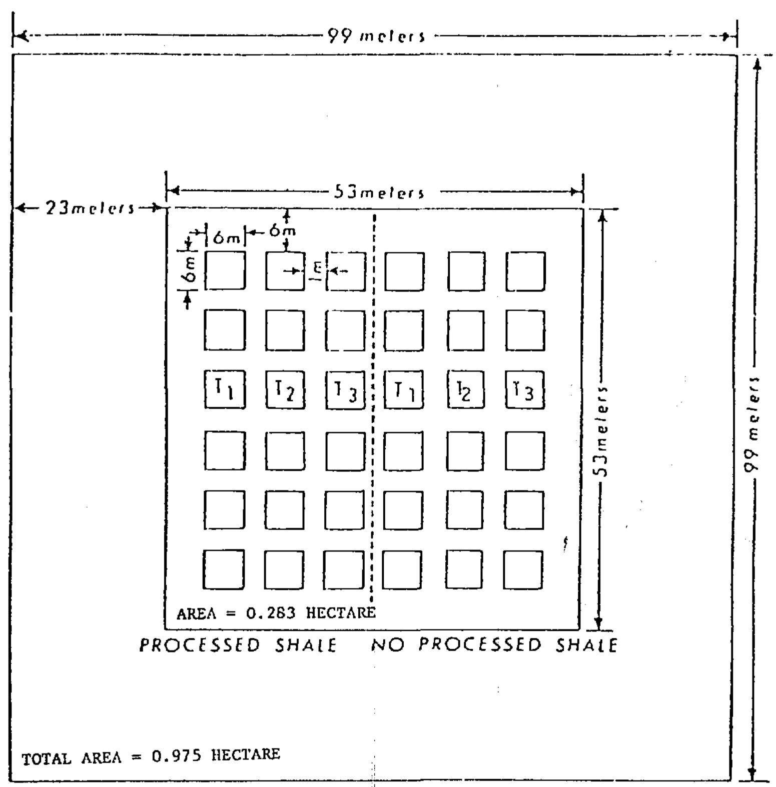


Fig. 6. Generalized layout for revegetation plot (R_3) on oil shale tract C-a, Rio Blanco County, Colorado (initiated in 1976). Applications of three mulch treatments (T_1, T_2, T_3) were applied to two conditions for a total of six treatments. These treatments were replicated six times. Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project—Revised Detailed Development Plan for Tract C-a,* vol. 2, May 1977.

Table 4. **Plant response parameters measured in revegetation experiments on oil-shale tract C-a, Rio Blanco County, Colorado, 1976–1977**

Parameter	Time of measurement	Taxa involved
Number of emerged seedlings per plot	First spring following fall planting (i.e., beginning of first growing season)	Each planted species
Number of surviving seedlings	End of first growing season	Each planted species
Above-ground biomass (dry weight)	End of third growing season	Total seeded species, total alien species, and individual seeded species contributing bulk of biomass
Percent cover	End of each growing season	Each species

Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project – Revised Detailed Development Plan for Tract C-a*, vol. 2, May 1977.

e. Possible solutions—Show complete AOV design with all components of variance delineated. Set testing level to at least ±25% of the mean at the 0.10 alpha level. Stratify sampling sites into similar physiographic units.

B. Fauna

1. Mule deer density

a. Methodology—Pellet-group counts for deer on C-a will be made semi-annually on 25 plots (100 ft^2) in at least 60 randomly selected quarter-section sampling units for each of four blocks (9 sq miles). Sampling for spring and fall deer use will be done for five consecutive years.

b. Baseline data and analysis—Pellet-group data for the C-a tract and vicinity disclosed large differences in use between the summer and winter periods and also large standard errors (Table 5).

c. Statistical tests—H_0: mule deer density estimates for the C-a study area are not significantly different from the surrounding area (DOW game management unit 22) on a per-unit basis. H_0: mule deer numbers and distribution within the C-a study area are not significantly different before or during oil-shale development activities. Analysis will be an AOV, using initial studies to establish the sample size required to provide mule deer density estimates from pellet-group data to within 10% of the mean 90% of the time.

d. Problems—The baseline data analysis showed a wide variation in pellet groups found on different study units. Pellet groups are not randomly distributed over space, which complicates testing procedures.

e. Possible solutions—If AOV does not prove suitable because of nonrandom distribution of pellet groups, some type of nonparametric analysis may be usable to detect differences due to development.

2. Feral horse abundance

a. Methodology—Feral horses on and within 3 miles of the tract boundary are counted annually from flights along designated transects. Aerial census data are supplemented by opportunistic ground observations.

b. Baseline data and analysis—Data collected from aerial surveys disclosed a wide variability in numbers of horses observed among flight dates and between adults and juveniles. During late fall and winter, separation into distinct age classes was difficult (Table 6).

c. Statistical tests—None proposed.

d. Problems—Counts vary widely among transects and between sampling dates. No valid statistical procedures are apparent.

Table 5. Mule deer pellet groups accumulated over winter and summer, from 1974 through 1975, on transects located on tract C-a for the Rio Blanco Oil Shale Project

Transect number	Pellet groups recorded	Pellet groups per acre	Period of accumulation (days)	Pellet-group index[a]
		Winter		
1	4	69.7	209	0.33
5	14	243.9	211	1.16
6	11	191.7	208	0.92
7	7	122	209	0.58
		Summer		
1	0	0	125	0
2	4	69.7	123	0.57
3	2	34.8	125	0.28
4	1	17.4	125	0.4
5	0	0	125	0
6	0	0	125	0
7	0	0	125	0
8	3	52.3	123	0.43
9	1	17.4	127	0.14
10	6	104.5	125	0.81
11	5	87.1	125	0.70
12	2	34.8	128	0.27
13	0	0	129	0

[a]Pellet-group index equals the pellet groups/acre divided by the period of accumulation.

Winter

x (mean) = 9.0.
n (number of units) = 4.
SD (standard deviation) = 4.40.
Sx (standard deviation of mean) = 2.2.
90% confidence interval = 9 ∓ (2.353 X 2.2) = 9 ∓ 5.2.

Summer

x = 1.85.
n = 13
SD = 2.08.
Sx = 0.58.
90% confidence interval = 1.85 ∓ (1.782 X 0.58) = 1.85 ∓ 1.03.

Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project – Terrestrial Ecology Baseline Studies, Annual Report for Tract C-a,* March 1976.

Table 6. Number of feral horses observed during six aerial surveys conducted for Rio Blanco Oil Shale Project from November 1974 through August 1975

Date	Number observed			
	Total	Adult	Juvenile	Unidentified
Nov. 8, 1974	108	24	15	69
Dec. 30, 1974	86	8	2	76
March 4, 1975	41	16	4	21
April 14, 1975	74	69	5	0
June 26, 1975	93	69	24	0
Aug. 18, 1975	63	55	8	0

Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project – Terrestrial Ecology Baseline Studies, Annual Report for Tract C-a,* March 1976.

e. Potential solutions None apparent.

3. Small mammal abundance
 a. Methodology Live trapping on C-a will be done with baited Sherman traps, set out in five trap groups consisting of two lines of ten traps per sampling area. Approximately ten pit-traps will be established in each mammal sampling area. An index of abundance will be calculated from numbers of small mammals trapped per 100 trap days.
 b. Baseline data and analysis Shannon-Weiner Diversity Indices (H') are shown in Table 7 for seven vegetation types on Tract C-a for nine sampling periods (October 1974 through September 1976).
 c. Statistical tests H_0: small mammal population levels are not significantly changed by habitat modification or revegetation. Analysis of variance will be determined on an index of abundance. Regression analysis is also under consideration.
 d. Problems The index of abundance may be valid only for those most abundant species. Small mammal populations are very dynamic and within treatment variance may be very large. For small mammal studies, it is difficult to determine reasonable precision and accuracy and how much replication in time and space is adequate.

4. Breeding-songbird densities
 a. Methodology The study area will be mapped and divided into 2.47-acre subunits. Study units will be replicated in control and treated areas.
 b. Baseline data and analysis No data were collected on breeding-bird activity during the two-year baseline monitoring because the Emlen Strip census technique was used.
 c. Statistical tests H_0: breeding-songbird densities are not significantly changed by habitat modification revegetation. An AOV will be used to compare control and treated areas. Territory size and reproductive effort of breeding birds will be assessed qualitatively.
 d. Problems Sampling intensity must be determined for the number of subunits and replicates before the final design can be done.
 e. Potential solutions Possibly an accuracy and precision of ±25% of the mean 90% of the time will be a reasonable and achievable objective. Possibly the

Table 7. Shannon-Weiner diversity indices (H') for all small mammal grids during nine sampling periods, from October 1974 through September 1976, for the Rio Blanco Oil Shale Project

Vegetation type	Sampling period								
	1974		1975				1976		
	Oct.	Dec.	May	July	Sept.	Dec.	May	July	Sept.
Bottomland meadow	0.349	0.803	0.908	0.500	0.000	0.000	0.000	0.892	1.114
Upland sagebrush	0.687	0.000	0.967	1.047	0.718	0.693	0.401	1.185	0.986
Rabbitbrush	0.745	0.440	0.894	0.619	0.455	0.000	0.625	0.678	0.692
Pinyon-juniper mixed brush	0.980	0.000	1.001	1.038	0.971	0.000	0.860	1.141	0.871
Mixed brush	0.665	0.693	0.642	0.683	0.655	0.000	0.540	0.942	0.469
Pinyon-juniper sagebrush	0.673	0.000	0.935	1.133	0.295	0.451	0.730	1.189	1.143
Bald	0.349	0.000	0.500	0.520	0.000	0.000	0.000	0.510	0.826

Source: Gulf Oil Corporation and Standard Oil Company, *Rio Blanco Oil Shale Project – Final Environmental Baseline Report for Tract C-a and Vicinity*, vol. 2, May 1977.

chi-square test can be used for comparison of territory size and reproductive dynamics.

5. Fish populations
   a. Methodology Estimates will be made of fish population size, length, weight, age, condition, and reproductive condition. Capture will be by electrofishing on selected stations on Piceance Creek.
   b. Baseline data and analysis Data collected on fish in the vicinity of tract C-b disclosed a considerable difference among sampling stations and sampling periods in both fish numbers and species (Table 8).
   c. Statistical tests H_0: no difference occurs between fish populations during the baseline period and during development owing to mine operations. Tests proposed are chi-square, AOV, and correlation. Specific details have not been provided.
   d. Problems Detailed statistical testing procedures are still to be designed.

Table 8. Numbers and species of fish captured, marked, and recaptured at Piceance basin stations from September 1974 through July 1975[a]

Station	Brook trout			Rainbow trout			Brown trout			Mountain sucker			Flannelmouth sucker			Speckled dace			Totals		
	C	M	R	C	M	R	C	M	R	C	M	R	C	M	R	C	M	R	C	M	R
September 1974																					
P-1	ND	ND	ND	ND	ND	ND	ND	ND	ND	103	59	ND	ND	ND	ND	28	27	ND	131	86	ND
P-2	ND	ND	ND	ND	ND	ND	ND	ND	ND	24	19	ND	ND	ND	ND	ND	ND	ND	24	19	ND
P-3	ND	ND	ND	ND	ND	ND	ND	ND	ND	4	4	ND	ND	ND	ND	ND	ND	ND	4	4	ND
P-4	ND	ND	ND	ND	ND	ND	ND	ND	ND	1	ND	ND	ND	ND	ND	ND	ND	ND	1	ND	ND
P-5	ND	ND	ND	1	1	ND	ND	ND	ND	3	3	ND	ND	ND	ND	ND	ND	ND	4	4	ND
P-6	ND	ND	ND	ND	ND	ND	ND	ND	ND	18	16	ND	ND	ND	ND	6	5	ND	24	21	ND
P-7	ND	ND	ND	ND	ND	ND	ND	ND	ND	2	2	ND	ND	ND	ND	8	8	ND	10	10	ND
W-3	1	1	ND	ND	ND	ND	ND	ND	ND	1	1	ND	ND	ND	ND	1	1	ND	3	3	ND
L.S.L.	78	54	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	78	54	ND
Totals	79	55	ND	1	1	ND	ND	ND	ND	156	104	ND	ND	ND	ND	43	41	ND	279	201	ND
November 1974																					
Totals	18	17	1	ND	ND	ND	1	1	ND	114	111	3	1	1	ND	9	4	ND	143	134	4
January 1975																					
Totals	79	ND	ND	1	ND	ND	ND	ND	ND	89	ND	ND	ND	ND	ND	17	ND	ND	186	ND	ND
March 1975																					
Totals	88	3	3	1	1	ND	ND	ND	ND	36	6	ND	6	6	ND	17	ND	ND	148	16	3
May 1975																					
Totals	19	12	0	ND	ND	ND	ND	ND	ND	6	1	ND	ND	ND	ND	5	ND	ND	30	13	ND
July 1975																					
Totals	23	19	2	1	1	ND	ND	ND	ND	75	8	1	ND	ND	ND	52	ND	ND	151	28	3
Grand totals	306	106	6	4	3	ND	1	1	ND	476	230	4	7	7	ND	143	45	ND	937	392	10

[a]C, captured; M, marked; R, recaptured; ND, no data.

Source: Ashland Oil, Inc., and Shell Oil Company, *Oil Shale Tract C-b – Detailed Development Plan and Related Materials*, vol. 2, February 1976.

Establishing a reasonable accuracy/precision level for fish sampling in small streams is desirable.

e. Potential solutions—A detailed statistical design is required before development monitoring is implemented.

6. Benthos and periphyton dynamics
   a. Methodology—The Surber sampler will be used to sample benthos. Species diversity and abundance will be compared for long-term fluctuations and seasonal changes. Periphyton will be collected from artificial substrates to determine productivity and species diversity.
   b. Baseline data and analysis—Diversity indices for benthic invertebrate species in Piceance Creek just north of tract C-b are given in Table 9.
   c. Statistical tests—H_0: no change in benthos and/or periphyton communities will occur as the result of development. Statistical tests proposed by the lessees include analysis of productivity during monitoring vs baseline by analysis of variance or covariance, by correlation, by diversity indices, and by some unspecified nonparametric tests.
   d. Problems—A study of the baseline data disclosed the need for good quality assurance programs in using aquatic sampling instruments and procedures. All statistical tests must be carefully and fully detailed prior to collection of data.
   e. Possible solutions—Suggestions are solicited for achieving a reasonable monitoring program.

II. Abiotic

A. Air

1. Gaseous constituents (sulfur dioxide, oxides of nitrogen, nitric oxide, hydrogen sulfide, and carbon monoxide)
   a. Methodology—Monitoring is done continually with automated instruments, both intermittent samplers and continuous analyzers, in environmentally controlled shelters.
   b. Baseline data and analysis—Typical average monthly and ambient air constituent concentrations of gases and particulates were monitored during the two-year baseline period (Table 10).
   c. Statistical tests[6]—The more commonly used statistics are those which describe

6. Charles E. Zimmer, "Air Quality Data Handling and Analysis," pp. 453-84 in *Air Pollution*, Arthur C. Stern, ed., Academic Press, New York, 1976.

Table 9. Benthic invertebrate species diversity indices for Piceance Creek from September 1974 through November 1975

Month	Station							
	P-1	P-2	P-3	P-4[a]	P-5	P-5A	P-6	P-7
				1974				
September	1.66	2.08	2.59	2.75	3.26		1.44	1.00
October	1.49	1.59	2.05	1.36	2.40		1 59	1.19
November	2.21	1.40	2.47	2.46	2.20	1.29	0.44	1.42
December	0.67	1.47	2.37		2.12	1.16	2.02	1.84
				1975				
January	1.76	1.82	1.58		2.35	1.06	1.27	1.76
March	1.75	1.77	1.65		2.24	2.08	1.16	1.49
May	1.62	1.99	2.00		1.65	2.80	1.44	0.97
July	1.74	1.76	2.01		1.84	1.61	1.82	0.86
September	2.31	2.40	1.96		2.18	1.55	1.33	1.52
November	1.55	1.10	2.86		2.65	0.84	0.70	0.35

[a]Station P-4 was relocated to P-5A in November 1974.

Source: Ashland Oil, Inc., and Occidental Oil Shale, Inc., *Oil Shale Tract C-b – Environmental Baseline Program Final Report (November 1974 through October 1976)*, vol. 4, 1976.

Table 10. Monthly and annual average ambient air constituent concentrations of gases and particulates

$\mu g/m^3 = 6.244 \times 10^{-11}$ lb/ft^3

1974–1975

Trailer	Item	Month												Annual
		Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	[illegible]
020	NO ($\mu g/m^3$)	1 9	0 7	3 4	*0 4	*0 3	9 6	17 8	*3 2	*0 4	9 1	*3 2	3 5	[illegible]
023		4 4	*12 8	14 7	0 4	*1 2	0 6	0 1	0 3	0 6	1 4	0 6	0 3	[illegible]
020	NO_2 ($\mu g/m^3$)	2 8	6 8	4 5	*2 5	*1 2	2 7	4 1	*0 3	*1 7	4 7	*3 0	7 9	[illegible]
023		2 4	*4 7	7 4	0 2	*0 4	0 9	0 5	0 0	1 5	1 0	0 1	[illegible]	[illegible]
020	O_3 ($\mu g/m^3$)	58 0	69 3	93 7	105 1	68 0	77 1	71 5	69 1	74 6	[illegible]	41 1	32 8	[illegible]
023		*31 4	28 0	42 4	85 8	85 6	90 5	87 3	84 4	86 1	61 0	53 4	43 3	[illegible]
020	Non Methane H C ($\mu g/m^3$)	73 4	97 4	*75 7	23 4	38 6	327 6(2)	38 9	50 6	25 8	27 9	58 9	[illegible]	[illegible]
023		933 0(1)	[illegible]	*667 8(1)	22 1	*17 1	49 1	43 3	196 5	220 2	145 6	92 2	331 8	[illegible]
020	CH_4 ($\mu g/m^3$)	876 1	918 6	*920 1	679 4	829 8	590 6(2)	633 6	821 2	625 9	[illegible]	949 3	[illegible]	[illegible]
023		825 5	1053 8(1)	*942 6(1)	*833 7	*859 8	836 3	834 3	814 7	730 7	922 7	933 4	814 1	[illegible]
020	CO ($\mu g/m^3$)	553 8	676 9	903 2	1278 5(2)	1498 0(2)	*853 0(2)	1815 5(2)	1092 1(2)	206 5(2)	[illegible]	[illegible]	[illegible]	[illegible]
023		3703 6(1)	2439 3(1)	1786 2(1)	391 4	504 5	485 8	699 9	437 0	466 5	[illegible]	[illegible]	[illegible]	[illegible]
020	SO_2 ($\mu g/m^3$)	0	1 8	0	0 1	1.1	0.1	0	1 0	1 6	0 9	0 7	1 4	[illegible]
021		1.3	1 7	0 8	0 8	0 6	1.2	1.1	2 1	1 3	1 9	1 7	1 8	[illegible]
022		2.6	0.1	0	0 2	0	0	0.5	1 7	0 4	0 6	0	0 6	[illegible]
023		0	6 4	3 3	0 1	0.6	0.5	0	0 0	0 6	2 0	0 9	0 2	[illegible]
024		0 2	1 7	1 3	5 5	0 9	1 1	0.7	0 6	0 1	0 9	0 7	2 0	[illegible]
020	H_2S ($\mu g/m^3$)	0	0	0	0	0 4	0	0	0 2	0	0	0 1	0 1	[illegible]
021		1 6	0	0.2	0 8	0 4	0.2	0 4	0 7	0 2	0 4	0 7	0 8	[illegible]
022		0 2	1.2	0	0 2	0.3	0.7	0	0 4	0 2	0 3	0 4	0 2	[illegible]
023		0	5.2	2 5	0 7	2 7	0.5	0 3	0.3	4 8	2 4	2 1	0	[illegible]
024		0	0.1	0	0 2	0 4	1 0	0 3	1.1	0 1	0 1	0 6	1 0	[illegible]
020	Particulate ($\mu g/m^3$)	*46 7(3)	4 3	3 3	3 8	6 5	11.6	12 4	10 7	14 7	[illegible]	12 8	11 2	[illegible]
021		*20 4	5 4	4 0	4 5	6 9	13 7	13 2	12 3	15 6	[illegible]	12 4	12 4	11
022		*35 3	4.2	2 9	3.2	5 3	11.9	11.2	9 5	14 6	11.5	12 6	9 7	11
023		*18 0	*6 8	*2 5	4 2	11 5	15 4	19.3	18 3	14 4	[illegible]	11 1	13 7	[illegible]
024		*117 0	2 9	2 3	3 8	4 9	10.2	11 4	8 7	11 3	13.1	9 8	12 4	

1975–1976

Trailer	Item	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Annual Average
020	NO ($\mu g/m^3$)	2.6	3.4	12.5*	5.7	(3)	(3)	(3)	(3)	4.9	(3)	55.1	(3)	[illegible]
023		.0**	2.5	0.0*	(3)	2.0	.4	1.5	2.2	1.2	2.6	3 9	4.5	[illegible]
020	NO_2 ($\mu g/m^3$)	8.5	2.2	4.8	2.1	(3)	(3)	(3)	(3)	(3)	(3)	7.4	(3)	5.0
023		.8	2.6	(3)	(3)	.1	0.3	1.3	1.5	.2	1.6	2.2	4.4	1.5
020	O_3 ($\mu g/m^3$)	30.9	32.7	35.7	30.7	40.2	66.5	72.7	66.3	69.8	74.4	55.9	55.0	52.6
023		39.2	38.2	37.1	41.2	45.2	68.0	78.4	74.5	81.5	58 8	55.5	57.4	56 3
020	Non Methane H.C. ($\mu g/m^3$)	No Data	26.4	20.5	(3)	(3)	(3)	(3)	60 3	[illegible]	26.5	[illegible]	[illegible]	[illegible]
023		303.8	151.3	171.2	389.7	388.7	818.3	1194.6	[illegible]	(3)	(3)	452.2	153.2	571.1
020	CH_4 ($\mu g/m^3$)	No Data	579.5	592.5	(3)	(3)	(3)	533.1	928.9	958.4	[illegible]	954.7	1006 0	[illegible]
023		920.6	909.6	862.5	1046.6	1046.6	1007.2	948.0	974.5	(3)	(3)	953.7	942.2	[illegible]
020	CO ($\mu g/m^3$)	No Data	(3)	408.5*	(3)	(3)	(3)	1294.8	562.6	444.0	439.9	661.0	526.4	605.4
023		1847.3	(3)	1787.3*	1161.8	1271.5	657.0	(3)	1521.5	(3)	(3)	(3)	(3)	[illegible]
020	SO_2 ($\mu g/m^3$)	.0	.0	.0	.0	.0	.0	1.1	1.1	.4	.0	.9	.6	[illegible]
021		.9	.0	1.7	.2	.6	.8	1.3	(3)	(3)	(3)	(3)	.4	[illegible]
022		.1	1.4	1.9	1.7	1.4	1.0	0.5	.0	.7	.1	.2	(3)	.6
023		.0	.2	1.1(4)	.1(4)	2.6(4)	.5	0.6	.3	.1	.1	.7	.6	.6
024		1.1	.8	.2	.1	.2	.3	0.5	1.1	.8	6 2	(3)	1 5	1 2
020	H_2S ($\mu g/m^3$)	.1	.0	.0	.0	.1	.3	.0	.0	.0	.0	.0	0	[illegible]
021		.1	.2	0.2(4)	.1(4)	.1(4)	.7	.6	.8	.0	.6	.6	1.2	[illegible]
022		.0	.1	.0	.0	.0	.6	.2	.1	(3)	(3)	(3)	(3)	.1
023		.4	.9	3.3	.9	.4	.3	.6	7.	.1	.4	.4	.4	[illegible]
024		.0	.1	.4	.0	.0	.0	.0	.0	.1	.7	.5	.1	[illegible]
020	Particulate ($\mu g/m^3$)	5.0	2.8	3.2	3.2	7.5	14.7	10.5	12.5	17.9	10 6	8.9	10.1	[illegible]
021		4.3	3.7	3.4	2.7	6.9	14.0	10.3	13.3	16.7	11.2	9.2	10 2	6.6
022		4.6	2.8	3.3	2.6	5.7	12.5	7.6	10.3	11.6	8.4	7.9	6 2	[illegible]
023		3.9	2.3	2.8	2.4	5.5	9.8	8.9	11.4	13.4	9 3	7.2	7.1	[illegible]
024		5.3	2.5	3.3	3.0	7.5	14.4	10.2	14.2	16.4	[illegible]	8.1	6 7	[illegible]

[1] Reported data are incorrect because of contaminated manifold.

[2] Reported data may be incorrect because of possible malfunctioning instrument.

[3] Few or no data collected because of instrument malfunction.

[4] Side-by-side monitoring of H_2S in trailer 023 and of SO_2 in trailer 021 was initiated as a data reliability check for three months beginning January 1, 1976. Therefore, no SO_2 analyzer at 021 are reported in the row for 023 for January, February, and March. Data from the second H_2S analyzer at 023 are reported in the row for 021 for January, February, and March.

*50% or less data.

**.0 indicates below limits of detectability of the instruments.

Source: Ashland Oil, Inc., and Occidental Oil Shale, Inc., *Oil Shale Tract C-b–Environmental Baseline Program Final Report (November 1974 through October 1976)*, vol. 3, 1976.

location and those which describe dispersion. Statistics which identify a point of cluster are arithmetic mean, median, and geometric mean. To indicate the extent of data variance about the mean, the standard deviation and standard geometric deviation are used.

d. Problems How to distinguish the pollution sources, which may be a point source (stack), line source (a line of traffic), and an area source (*distant*-source pollution from area cities or local source from shale storage piles). How to determine, from the air quality data, whether baseline and monitoring levels differ. How to assess natural changes when comparing baseline data to development data.

e. Possible solutions Locating air-quality stations in areas of (1) most likely pollution and (2) least likely pollution. Emission data should also be taken at the stacks, and all of the data should be compared with the baseline data. Specific procedures must be established for handling and analyses of data to provide information in required format and at the appropriate time. These procedures should include all aspects of data recordings, validating, storage and retrieval, presentation, and statistical methods of analysis. A possible null hypothesis that could be tested would be H_0: there is no significant difference between the concentrations of gaseous constituents during the baseline period and during the development period. If the first null hypothesis is proven false, a second null hypothesis that could be tested would be H_0: all of the increase in gaseous constituents during commercial development is due to stack emissions.

2. Particulates

a. Methodology A high-volume sampler at each of three air-quality-monitoring sites collects samples 20 ft above ground elevation near planned site development activities.

b. Baseline data and analysis Table 10 gives the two-year monitoring data.

c. Statistical tests None given.

d. Problems Will enough information be taken to (1) assure compliance with regulator standards; (2) evaluate the effectiveness of fugitive dust control measures; (3) determine traffic patterns to minimize airborne particulates; and (4) aid in evaluating the effect of dust deposition on tract vegetation and on water surfaces?

e. Possible solutions Identify development-related perturbation from natural occurrences by positioning high-volume samplers at points of maximum concentrations; other samples upwind of development could provide control. Possible null hypotheses that could be tested are like those for the gaseous constituents.

B. Hydrology

1. Surface water

a. Methodology Standard techniques, including concrete controls at most stations to ensure a stable rating curve. Seeps and springs are also measured, because hydrologic studies and analysis of the two-year baseline data indicate that some of the springs and seeps are hydrologically connected to the upper oil-shale aquifer. Perturbations to springs and seeps will be detected by analysis of flow and water-quality data. Some water-quality information is collected at springs and seeps quarterly or semi-annually. Continuous data are collected at most of the stream-gaging stations on several parameters, including flow, temperature, conductivity, sediment, dissolved oxygen, and pH.

b. Baseline data and analysis Tables 11 and 12 give typical data collected during the two-year baseline period.

c. Statistical tests The nature of the data will determine statistical methods to be used for data analyses. The following hypothesis will be tested on the chemical and physical parameters: H_0: there is no significant difference in

7. Ralph I. Larson, *A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards*, U.S. Govt. Printing Office, Washington, D.C., 1971.

Table 11. Summary of the mineralogy of seven samples from the streambed at the Evacuation Creek gaging sites

Mineral	Percent composition			
	Mean	Standard deviation	Maximum	Minimum
Quartz	33	8	44	21
Potassium feldspar	6	1	8	4
Plagioclase feldspar	7	2	8	4
Calcite	14	3	18	11
Dolomite	17	5	27	12
Clay minerals	14	2	15	10
Analcime	2	1	3	1

Source: Phillips Petroleum Co., Sunoco Energy Development Co., and Sohio Petroleum Co., *First Year Environmental Baseline Report for Tracts U-a and U-b, Utah-White River Shale Project*, VTN Colorado, Inc., vol. 1, May 1976.

Table 12. Results of regression analyses from Evacuation Creek water-quality data collected during the two-year baseline period

Variable[a]		Regression line		Accuracy		
Independent	Dependent	Intercept	Slope	No. of pairs	Correlation coefficient	Standard error of estimate
Specific conductance	Dissolved solids	98	0.83	74	0.85	371.0
Dissolved solids	Calcium	48	0.03	73	0.68	21.0
Dissolved solids	Magnesium	6	0.04	75	0.91	13.0
Dissolved solids	Sodium	-33	0.20	75	0.97	36.0
Dissolved solids	Chloride	-2	0.013	73	0.70	8.1
Dissolved solids	Sulfate	1964	0.030	73	0.05	563.0
Dissolved solids	Copper	29	−0.0001	44	0.01	93.0
Dissolved solids	Iron	118	−0.021	44	0.19	42.0
Dissolved solids	Molybdenum	9	0.007	45	0.25	12.0
Dissolved solids	Selenium	4	0.0003	43	0.008	3.7

[a]Specific conductance in pmhos/cm; 1 pmho/cm = 2.54×10^{-12} mhos/in. All other variables in mg/liter; 1 mg/liter = 8.343×10^{-6} lb/gal.

Source: Phillips Petroleum Co., Sunoco Energy Development Co., and Sohio Petroleum Co., *First Year Environmental Baseline Report for Tracts U-a and U-b, Utah – White River Shale Project*, VTN Colorado, Inc., vol. 1, May 1976.

the chemical composition or physical parameters of the surface waters studied during baseline and development periods at each monitoring station.

d. Problems—(1) How to obtain, from the observed two-year baseline hydrologic data and sparsely available long-term data, a true picture of the hydrologic regime for determining effects during oil-shale development; (2) how many data-collection points must be operated during development to obtain a statistically valid picture of impacts on surface waters; (3) how to accurately gage and evaluate the required surface-flow augmentation to protect existing water rights; (4) how to evaluate the efficacy of mitigating measures, if required, to protect the environment; and (5) how to determine when modified in-situ retorts have reached a state of stability so as to release the lessee from further environmental liability.

e. Possible solutions A sound statistically based network design should be established to show how the surface-water flow and quality will be affected by the modified in-situ development. Monitoring surface-water flows down established channels will be relatively easy; complications arise because the ground-water surface-water systems are intimately connected.

2. Ground water
   a. Methodology – Ground water is monitored by observation wells in the alluvium and in oil-shale aquifers. Water levels are measured continuously in several holes and monthly or quarterly in others. Samples are collected semiannually, quarterly, or more frequently, and analyzed for alkalinity, pH, silica, fluoride, conductivity, temperature, and the major ions. Analyses for trace elements and organics are made quarterly or semiannually. Recording flow meters will be used in the dewatering wells to determine dewatering rates. Water-quality monitoring will concentrate on dewatering well discharges.
   b. Baseline data and analysis Table 13 gives typical ground-water quality data collected during the two-year baseline testing. Modeling techniques also simulated ground-water flow.
   c. Statistical tests - The major ions found during baseline studies were plotted on trilinear diagrams. Water will be categorized by the relative concentration of calcium, magnesium, sulfate, chloride, potassium, bicarbonate, carbonate, silicon dioxide, and fluoride. The following hypothesis will be tested: H_0: there is no significant difference in chemical composition or physical parameters of the ground water between baseline and development periods.
   d. Problems - The problems in testing ground-water data are similar to those for testing the surface-water data.
   e. Possible solutions - The approach will be similar to those used for surface

Table 13. Summary of ground-water quality data collected from the Birds Nest aquifer, Parachute Creek member of the Green River Formation, tract C-a, from November 19, 1974, through November 14, 1975

Description	No. of sample	Mean	Standard deviation	Maximum	Minimum
Conductance (μmhos)[a]	34	4459	1208	6070	1130
Bicarbonate (mg/liter HCO_3)[b]	31	643	319	2010	26
Carbonate (mg/liter CO_3)[b]	30	4.2	21	117	0
Nitrite + Nitrate (mg/liter N)[b,c]	32	1.8	6.1	30	0.00
Hardness (mg/liter Ca, Mg)[b]	31	957	381	1400	22
Calcium (mg/liter)[b,c]	31	131	64	210	5.7
Magnesium (mg/liter)[b,c]	31	153	64	240	0
Sodium (mg/liter)[b,c]	31	807	293	1500	72
Potassium (mg/liter)[b,c]	31	6.8	2.8	13	2.3
Chloride (mg/liter)[b,c]	31	70	22	140	36
Sulfate (mg/liter)[b,c]	31	2009	843	3000	180

[a]1 μmho = 2.54×10^{-6} mhos/in.
[b]1 mg/liter = 8.343×10^{-6} lb/gal.
[c]Elements analyzed only for dissolved fraction.
Source: Phillips Petroleum Co., Sunoco Energy Development Co., and Sohio Petroleum Co., *First year Environmental Baseline Report for Tracts U-a and U-b, Utah – White River Shale Project,* VTN Colorado, Inc., vol. 1, May 1976.

water, with modeling used more frequently.

III. Interrelationships

The lease requires the lessee to study and report to the Mining Supervisor (Area Oil Shale Supervisor) on ecological interrelationships including migratory patterns of birds, mammals, and fish, and plant-animal relationships. This very general requirement, which allows wide latitude in interpretation, could mean as much as a complete ecosystem study, or as little as a qualitative description of some two-factor comparisons.

Not surprisingly, the lessees have monitored interrelationships differently. The C-b lessees are using an International Biological Program (IBP) systems approach. This program, much too complex to present here, is described in the C-b Environmental Baseline Program Final Report.[8] Essentially, the program proposes an ecosystem model where driving variables (precipitation, wind direction and velocity, sulfur-compound emission, ozone, trace metals, fugitive dust, noise and activity disturbance, etc.) are monitored for their effect on five ecosystem response units. These units were formed from 13 plant community/habitat types based on similar vegetation and topographic characteristics. For instance, "General Upland" consists of chained pinyon-juniper rangeland, bunchgrass, sagebrush, and mountain shrub. A wide variety of "State" variables are measured on the response units such as animal numbers and weights, plant standing crop, and litter. Time-series graphs show functional equations over time and space. Impact-response matrices are developed to select the more important cause-and-effect relationships for monitoring and also to provide a mathematical model.

In the more simplistic C-a method, a series of matrices provides the basis for the interrelationship monitoring program. The first matrix compares individual engineering actions against specific biotic and abiotic parameters for (1) magnitude, direction and duration of impact, and (2) importance and measurability of the impact. Another matrix is prepared using the medium- to high-ranked parameters from the previous matrix to form a "mirror image" where biotic and abiotic parameters are compared for relationships among each other. This intra- and interrelationship matrix assures that interdisciplinary studies are coordinated in time and space so correlation-regression and other analyses can be performed. This matrix also provides the basis for cause-and-effect computer models. Important sources of mining impact are expected to be underground retort constructions, cumulative construction activities, dewatering, surface discharge, reinjection, atmospheric venting, storage of raw shale and topsoil, habitat modification, revegetation, and increased human activity.

The major interrelations on tract C-a are abiotic/abiotic, abiotic/biotic, biotic/abiotic, and biotic/biotic. An example of high-level relationships expected among precipitation, vegetation, and mule deer on tract C-a is depicted in the Fig. 7. A diagrammatic presentation of major soil-vegetation topography interactions is shown in Fig. 8.

8. Ashland Oil, Inc., and Occidental Oil Shale, Inc., *Oil Shale Tract C-b* Environmental Baseline Program Final Report (November 1974 through October 1976), vol. 5, 1976.

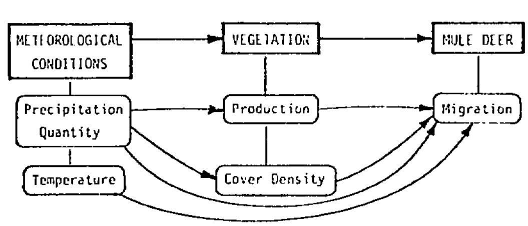


Fig. 7. Diagram of high-level relationships on tract C-a.

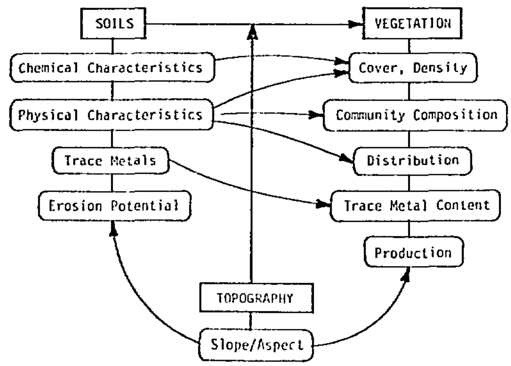


Fig. 8. Diagram of major soil-vegetation topography interactions.

The principal influences of abiotic parameters on other abiotic parameters that are rated as high-level relationships are

1. precipitation (quantity) on ground water (quality, quantity, flow/movement, level, recharge/discharge, availability);
2. precipitation (quantity) on surface water (quality, quantity, flow/velocity, drainage basin, sediment load, stream bed, springs, and seeps);
3. ground-water quantity on surface-water quantity;
4. ground-water quality on surface-water quality;
5. ground-water quantity on quantity of water in springs and seeps;
6. ground-water quality on quality of water in springs and seeps;
7. surface-water flow on surface-water sediment load;
8. soil erosion on surface-water sediment load; and
9. soil chemistry on sediment chemistry.

Nine abiotic-biotic relationship influences of the tract C-a area were ranked as high level:

1. precipitation quantity on vegetation production;
2. precipitation quantity on vegetation cover and density;
3. precipitation quantity on mule deer migrations;
4. ambient air temperature on large mammal migrations;
5. soil chemical characteristics on plant community distribution;
6. soil trace metals on trace metals in plants;
7. soil chemical characteristics on plant community composition;
8. soil physical characteristics (e.g., depth) on plant cover; and
9. slope/aspect on vegetation cover or composition.

Of the number of influences on the abiotic system by biotic components, only one such relationship was considered by the lessee to be highly important—the effect of plant cover and density on soil erosion potential. Moderate-level relationships identified include the influence of

1. vegetation (distribution and cover) on atmospheric particulate levels;
2. plant community composition on soil chemical characteristics;
3. relative abundance of invertebrates on soils (erosion potential, physical and chemical characteristics); and
4. vegetation cover on soil erosion potential.

Relationships among biotic parameters that were ranked as high level include

1. vegetation community composition and cover on small mammal abundance;
2. relative abundance of small mammals on the relative abundance of predatory mammals;
3. vegetation production on mule deer migrations;
4. vegetation cover on mule deer migrations; and
5. periphyton abundance on periphyton productivity and on benthos relative abundance.

A workable mathematical model must be developed to track only the more important parameter interrelationships, utilize the most appropriate statistical methodology, and adapt to changing objectives and methodology.

In conclusion, several statistical problems become evident after a parameter has been selected for monitoring:

1. After a null hypothesis is established, what level of resolution (probability or alpha level) is reasonable for rejection or failure to reject?
2. Is one standard criterion acceptable, such as means will be within one sigma 80% of the time, or should the alpha level vary according to the variability of the parameter?
3. How much latitude can be accepted in departure from the usual statistical assumptions (normal distributions, equal variance of populations, independence of mean deviations)?
4. Because two years does not constitute a basis for premining calibration of dynamic parameters, how can this limited information be used to adjust differences among control and treatment sites?
5. For the most part, interrelationships among variables could not be determined from the baseline data, because of temporal and spatial differences (parameters were measured independent of each other). What procedure should be used to select the most important correlations and regression analyses during production monitoring?
6. The baseline data collection program for air and water parameters addressed mainly the quality control and a comparison with parameter limits established by State and Federal law. Moderate attempts have been made to establish statistical procedures for detecting significant differences among stations, years, seasons, and daily periods. How can statistical tests be used to detect differences due to development as compared with natural effects?
7. There is a need to predict when a level of pollutant or disturbance will result in a significant perturbation in one or more parameters and or their interrelationships. How can statistics be used to predict these disturbances?
8. Environmental damages must be mitigated. What statistical procedures can best be used to test effectiveness of mitigative efforts? What would be a reasonable level of statistical probability?
9. The assessment of interrelationships probably require some type of modeling efforts. What type of cause-and-effect modeling can be used to track the major components of the tract ecosystem during production: How can this model accommodate statistics and be a decision-making tool?

SUMMARY AND CONCLUSIONS

The problems of monitoring are reduced to the need to correlate environmental parameters with forthcoming conceptual and detailed engineering plans and material balances. Emphasis will be placed on physical and chemical properties of gases, liquids, raw oil shale, and various waste products that will emanate from shale-oil production. Transport mechanisms are generally known, but detailed pathways of pollutant transport are little known. Estimation of these factors now will aid in determining anticipated pollutant levels that can be translated into likely stress conditions on the ecosystem.

A statistical design for the monitoring program must include features that will effectively analyze the development, biotic, and abiotic parameters and all important interrelationships.

Problem Discussion 1, Part 1: Assessment of Oil Shale Development—a Problem in Statistical Design

Donald R. Dietz, U.S. Fish and Wildlife Service
Eric Hoffman and Lawrence Barker, U.S. Geological Survey

Donald Dietz: Yesterday, we got cut off a little before we gave our complete conclusions or were able to present a statistical problem that we are concerned with and want help with. I'd like to finish this summary now. After the null hypothesis is established, what level of resolution is reasonable for rejection or failure to reject? Also, we had some comments yesterday about a better procedure than the use of null hypotheses. We would like to pursue this further and encourage your comments. A lot of the lessees or environmental contractors have been using one standard for their criterion for acceptance or rejection such as, the mean should read within one sigma 80% of the time. Should the alpha level vary according to the variability of the parameter, or is one standard enough? How much latitude can be accepted in departure from the user's statistical assumptions, such as normal distributions, etc.? Most of the biological samples with which we work aren't really neat, agronomic parameters. The areas are highly heterogeneous; populations are very dynamic. It's difficult to find areas to replicate that are similar. Because two years does not constitute a basis for premining calibration of dynamic parameters, how can this limited data base of only two years be used to adjust differences among control and treatment sites in trying to determine the impacts of oil shale development? For the most part, interrelationships among variables could not be determined from the baseline data simply because of temporal and spatial differences. What procedures should be insisted upon when we go into production monitoring so that the important correlations and regression analyses can be made between the disciplinary arts? The baseline data collection program for air and water address quality control and how well these parameters meet established state and federal standards. No attempt was made to employ statistical design, null hypotheses, or any other statistical testing with most of the air and water parameter data.

How can statistical tests be used to detect differences in air and water qualities due to project development as compared with the natural effects? There is a need to predict when a level of pollutant or disturbance will result in a significant perturbation in one or more parameters and their interrelationships. How can we predict these disturbances? Environmental damages must be mitigated. What procedures can best be used to test the effectiveness of the mitigation effort? What is a reasonable level of probability to strive for? The assessment of interrelationships requires some type of modeling effort. Will some form or type of cause-and-effect modeling enable us to track the major components of the ecosystems and their interrelationships? How can statistics be built into these models? Can a model be constructed that would be flexible enough to be a decision-making tool for a mining supervisor? These and many other statistical problems confront us, and being nonstatisticians, we hope you will bear with us in our attempt to express what we feel are statistical problems with this huge, essential environmental impact. So, we solicit comments from the audience.

Gary Tietjen, Los Alamos: I would like to make a few rash comments, a few of which I may later regret having made. I think the decision as to whether the oil shale facility will be constructed will not depend upon your study but will be dictated by the demand for energy. It would take a decade or more of studying the environment to encounter all the sources of variation that one might encounter in the few years after the study was completed. You can't take that length of time because the environment will change before you can finish the study; the process of studying it will change it! The amount of rainfall and snowfall will be major perturbations in the environment.

There is absolutely no question in my mind that construction of this facility will alter the environment; thus, one doesn't really need to ask that question. Rather, you should ask, whether the wildlife will *adjust* to the altered environment. The answer is very probably *yes!* The coyotes certainly will adjust; you can't keep them from adapting unless you hire an armed guard.

The whole community of Los Alamos has not disturbed the deer or fox populations seriously; there are enough of them around to disturb the gardeners sufficiently anyway. Bears may be disturbed somewhat, but no one there is trying to encourage a larger bear population. And the skunks will love the place! If the lessee were to irrigate a small field there, he would attract large numbers of deer at night. With the construction of a little cover and some feed, he could attract more birds. Thus, the question should be, "Is the lessee reasonably committed to conservation?" and if so, then he can probably operate, I think, with little concern.

Water supply needs a little different treatment. If the water quality is being degraded, how will you know if it's a serious degradation or whether it may be coming from somewhere upstream of the plant? One way to know positively is to run the plant water through a tank in which you keep trout. If the fish are okay in that tank, I think it is safe to dump the plant water into the stream. What I'm suggesting is that you put less effort into this base study and that you put more emphasis on continuing concern during and after the construction.

Lee Eberhardt, Battelle: I agree wholeheartedly with Gary in the sense that what he has said is partly summarizable as simply, "Let's use common sense." But I'm not sure if the audience as a whole is aware of the amount of effort that's gone into this kind of survey around the country with nuclear power plants and the outer continental shelf studies. I guess I shouldn't try to speak accurately about the requirements of NEPA as to what we should be studying, but it's pretty clear that a lot of the money is being spent on baseline studies and on after-the-fact construction studies. Also a lot of time and effort is going into some sort of experimental design; my reaction is that we often have a single experimental site—a nonrandomly selected, treated site, for which you can pick your controls at random out of the surrounding area if you like. But I am not sure what you do with that in the experimental design. The keynote speaker yesterday talked about a time series problem in which he felt that 100 observations was really not quite enough to do the kind of things he'd like to do. We're being asked here to have three preoperational and perhaps three postoperational samples with a series of six and to do much the same sort of thing. I would appreciate hearing some philosophy from some of the statistical people as to what constitutes a reasonable sort of statistical treatment here. I'm not sure whether all the statisticians are aware that we are talking about single years as single data points. Most of the more important ecological species or situations almost have to be dealt with annually. For the more complex species, the mule deer for example, we have pretty strong evidence of autocorrelation. For small critters like plankton, probably autocorrelation between years is only that that's forced on them by the physical system. So we have a strongly correlated series, a very short one, and I simply do not know what to do with it. I can only state that my philosophy about the use of statistics is correlated to understand what is going on in particular segments of the problem. I guess I return to Gary's comment, "Let's use common sense," but I would argue with each item he suggested, if I had to do it actually in the field.

Larry Barker: The lessees are required to take this data whether it is important or not. Because they must take it, we want something that we can use later. The problem is how to take the data, and what to do with the data after it's been collected. I'm not sure that we made that clear, but I want to make that emphasis right now.

Corwin L. Atwood, EG&G Idaho: I was one of the people who was bothered yesterday about framing everything in terms of testing hypotheses. We know that there is going to be an effect; so why ask the null hypothesis: Is there an effect? If you took enough data you'd find the answer. I was then wondering what to do about this, not in the simple case where you've just got one quantity of interest. Then you can get a confidence interval and visualize what's happening. But what about the harder to visualize case, when you have to keep track of several things at once? I can only draw three dimensions on a two-dimensional piece of paper (Fig. 1). These three things that you are considering might be deviations from an overall mean and an analysis of varying situations. I would like to hear what other people

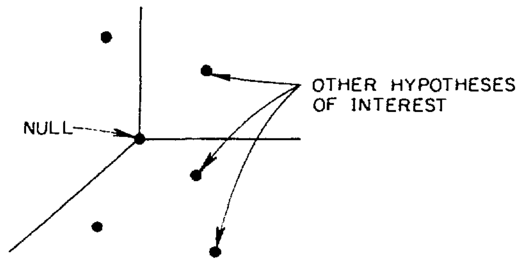


Figure 1

think of one thing that I do: I consider the following: what do the data say about the null hypothesis? what alpha level is wanted? what is the smallest alpha level which could be used and still be able to reject the null hypothesis? But to identify other hypotheses of interest may be one of the various ways that you could get away from the null hypothesis and be just barely unacceptable. There might be a number of these hypotheses, and I would check the data as if I were testing each of these as a null hypothesis. That would give me a ranking among the different possibilities of interest if I could single out a fairly small number of them. Maybe if I don't have much data, my data would accept all of these hypotheses at reasonable alpha levels, and then I'd realize I can not distinguish between them. Maybe the data would reject some of these worrisome ones and accept the null hypothesis, and that would tell me something. There are lots of things that could happen, but this ranking might give me a feeling for what is going on. It isn't in textbooks. I don't know if other people have ideas on this.

Dave Gosslee, Union Carbide Corporation, Nuclear Division: I would like to make two very general points. One is partly a question. To what extent do you intend to involve a statistician with the project, or is a statistician already involved? Second, I have the feeling you can't look at things like species individually. I think you need to look at things like species from a multivariate point of view to get more power into your testing estimation.

Don Dietz: We have a statistician available to us out of CSU, but so far we haven't been able to use him much. We also have the expertise from the Fish and Wildlife Service, from their Biological Services Division; they have done a lot of work with small mammals. We are hoping to get a statistician on board with us. Although we have a systems analyst and a computer specialist now, we still do not have a consulting statistician, and we feel that is our major need right now. We will be working hopefully with some multivariate analyses as we get into more species work. The lease that the oil companies agreed to requires certain things. But once we get the data into a data base, then we will be able to select our own statistical packages and run any analysis that seems appropriate and worthwhile. Right now this is a very simplified approach, and the program is quite dynamic; we still have a few years to consider anything we can get a lead on. One thing I might say about Dr. Tietjen's comments, "As a biologist, he makes a good statistician."

Peter Bloomfield, Princeton: It seems to me in this context it's a little dangerous to talk about our limits of things, for example, one and two sigma limits. I think these are numbers that can be determined from the baseline studies and then sort of written into law and used in a rigid way. As Gary pointed out, it is most unlikely that the baseline study will really come up with enough data to determine those quantities in any accuracy. The baseline study period is to determine good places to use as control sites, and the ongoing monitoring should really be a continuing process for carrying out information from the control site to decide, well, a basis to use for comparison in the development sites.

Frosty Miller, Union Carbide Corporation, Nuclear Division: To try and sample the whole environment at once and determine, shall we say, tolerance levels for change due to an "innovation" in environment is a very difficult problem. There are some ideas to be picked up from other technologies. Chang at Berkeley has a health index which he has proposed. Periodically, he samples people and asks them about their health; what he gets is a high variance on an individual observation that's sort of a control chart idea of the general health of a community over a long period of time. It seems to me that ideas like this are what you need. I didn't hear any discussion yesterday concerning where you were going to place your transects, vis-a-vis the expected changes or degradations in the environment that would inevitably ensue from changing the water flow. That will certainly affect the flora, and I would think that the fauna would go where the water is, at least to a limited extent. Thus, if you plan your transects so that you can discriminate between various hypotheses about what will happen to the water or what will happen to the dust burden from the site in the uncovered areas, you may have a better chance of detecting things than if you merely sample random transects across various altitude zones.

Frank Anscombe, Yale University: I have the notion that whenever one is planning an observational study of any sort, it's a good idea to begin with an exercise, which may seem a little silly, but I believe is really rather sensible: to consider what you would ask for if you were able to ask for absolutely anything whatever, without cost. In this matter of surveying the environment, suppose you could have an army of invisible demons observing absolutely anything that you would like them to observe at whatever frequency you like and suppose that you could have them doing this not merely for two years but for twenty years or two hundred years. What in that case would you like them to observe? But suppose that you are exposed to the risk that you might be expected to process that data when you get it; therefore, it would be foolish to ask for complete information about everything—one should never ask for that! Thus, what would one rather like to see and know about? The difficulty when one is thinking about any actual observational study is that one is too much constrained by what one thinks is practical—of course, one has to be constrained by what is practical—but sometimes considering what could be had in some magic way for the asking will sometimes make very clear that only some features are really of interest, and *that* might help somewhat redirect what one does with the resources that are actually available. Undoubtedly, if one could observe this area in great detail over a long period of time, one would see all sorts of changes going on, and the baseline would be a very wobbly baseline, I think. It would have slopes to it, all sorts of things like that, and to try to think out what sort of summarization would make a tremendous amount of data available will help a good deal in pinpointing the things that would be worthwhile trying to observe and practice.

Lee Eberhardt, Battelle: I think that's a particularly important comment by Dr. Anscombe. The trouble I have personally is that I see reasons to believe that most details of ecosystems are very rigidly controlled, but the systems themselves can be likened to a living animal. Things go precisely in a particular form, which I think one appreciates more or less on a philosophical basis and somewhat on a theoretical basis and more strongly on some understanding of evolution. In practice in the field, though, as a sort of a working rule, we use the coefficients of variation of about 100% on things we measure. I don't believe that's purely a stochastic process I'm looking at, but it's very difficult for me to get that variability for things that are measured much below that. Another thing that I think needs to be said here is that we are talking about something that turns out to be a sort of an adversary process in the end and perhaps in the beginning. We are told in designing studies for environmental impact evaluation that we can expect, sometimes literally, to be put in the witness chair in a court, and the situation then becomes one of legal manuevering. The most obvious thing for the prosecution to ask is, "Well, would you look at this and this and this and this?" To be protected against this sort of thing, the people who have to design the general survey insist that we look at every little item in sight, and in a sense, that's great, but in practice, it compounds the effort hopelessly.

I'll slip in one more comment for Gary, if I may. Many pieces, such as Los Alamos or Hanford, constitute areas that have been a tremendous bonus to the wildlife—to the ecology, if I may use that word in a way I shouldn't. Therefore, Los Alamos knows that not much damage has been done to the environment there, that they have really improved it considerably as far as deer, coyotes, skunks, and all those things are concerned; however, these other situations we're talking about here are going to do a lot of damage. Strip mining will do a lot of damage; the oil spill off the coast clearly will do a lot of damage. There's no doubt we're going to see an impact. The question is how do we deal with the adversary situation and how do we use what tools we have effectively.

Eric Hoffman: I'd like to thank you for those remarks. It's kind of a happy position we suddenly find ourselves in. In fact we have a mass of information on each of the oil shale tracts. Here is just the final baseline data report for tract C-b. It encompasses five volumes and a package of matrix tables which are kind of mind boggling. There are similar reports to the other tracts. Behind that stands about ten bookcases full of detailed statistical and probed-type data that has to be weeded through, and somehow a system has to be designed to compare that mass of information with the data that the lessee is required to obtain during development. Our problem at hand is how can we read through all of that in a statistically sound manner and decide what really would constitute a realistic environmental monitoring program that is statistically defensible. Any ideas or thoughts along that line would certainly be appreciated.

Tony Olsen, Battelle: I'd like to add one specific comment to what Dr. Anscombe said. In addition to deciding what you would like to take, I think you also need to emphasize what size of impact is really important. And then if you are really interested in a hypothesis-testing situation, what you should do is make all of the nice assumptions that one generally likes to make, look at the size of the impact that you want to detect, and find out whether you can detect it at all. I am familiar with an example in the weather modification field: if you make all those nice assumptions and want to detect whether you have an increase in rainfall over an area, you find out that you have to do approximately ten years of experimentation to detect a 50% increase in rainfall when theoretically the meteorologist may not be detecting effective rainfall more than 10% at most. Maybe one would very quickly get away from the idea of testing hypotheses.

John Thomas, Battelle: If you are really going to move away from the idea of testing hypotheses, and we're really going to admit that there is going to be an impact whether we can detect it or not, and we are really committed to doing more work past this baseline thing and we really are going to finally involve a statistician, and we really do have a lot of research questions, it seems to me you have a golden opportunity to do some real research in conjunction with the monitoring.

Lincoln Moses, Stanford University: It seems to me that if you have several hundred pounds of facts, your primary job is to view the problem as a descriptive one. What is it that needs to be described? Several hundred pounds of paper defies description. How would you describe the condition of the environment now; how would you describe changes in it? Hypothesis testing is almost extraneous, and sometimes biological insight would seem to be of primary importance in order to get a dozen, five dozen, I don't know, some number of descriptions that we can monitor there and in nearby areas.

Statistical Aspects of Nuclear Safeguards

Gary L. Tietjen

Los Alamos Scientific Laboratory
Los Alamos, New Mexico

ABSTRACT

A nuclear fuel reprocessing cycle is used to illustrate problems encountered by a statistician when trying to reconcile total amounts of an element at different stages in the recovery cycle. Calculation of errors are discussed along with problems of biases, holdup, and simulation.

INTRODUCTION

Each of the DOE laboratories and contractors is already, or soon will be, deeply immersed in nuclear safeguards and accountability. As I hear the problem discussed from a political viewpoint, there are frequent official references to a "malevolent act," but the term seems to refer more to blackmail threats to a civilian population than to the use of weapons in war, though the latter possibility is always present. The questions are: How can we keep unauthorized persons from getting nuclear material? and how do we tell whether some of it is missing? A *uniform* system of keeping track of our inventory will be necessary because some international control seems imminent and perhaps desirable.

The task is of enormous proportions. Some of the reactors going on-line will process or reprocess 50 kg of plutonium per *day*. Every item, every drop of solution, every piece of scrap metal, and every whiff of powder will have to be accounted for. Moreover, the transactions from one place to another or from one form to another will take place rapidly, so that the accountability will have to be automated on the computer. There will not be time to mull over decisions on a case-by-case basis as we have hitherto done.

When one mentions the word *safeguards*, he may be completely misunderstood. There are many who think of safeguards wholly in terms of *physical* security. At the new plutonium facility at Los Alamos, there probably will be a computer check of your badge, your signature, your fingerprints, and perhaps of your voice before entering the facility. At the same time, you would be monitored for radioactivity, of course. The chemists think the problem of safeguards solved if they have devised *accurate* and *precise* methods of analysis for minute quantities of material. The physicists think of safeguards problems in terms of very *rapid* nondestructive methods of analysis not requiring lengthy sample preparation. (In all of their methods they either count the radioactivity in the sample directly or irradiate the sample first and then count it.) The computer people believe *they* have solved the problem of safeguards if they are able to get the numbers quickly onto a data base with rapid retrieval capability. It is left for the statisticians to try to make some sense of the thousands of numbers that will be generated.

I have chosen one small segment of an actual reprocessing cycle at Los Alamos to illustrate the problems faced by the statistician. I shall try to neither exaggerate nor minimize the difficulties. Of course, material *is* lost during processing. The public and the press don't seem to understand this and have not been sympathetic. The losses are not so large as at first they seem. Losses of uranium at Los Alamos over the last 25 years, if put in metallic form, could be placed in a lady's purse. Of course, even a weight lifter would have difficulty walking out with it.

The situation is somewhat like making cookies. Suppose that you were given a certain amount of flour, sugar, etc., for this purpose and that the ingredients were weighed out to you. After the cookies are baked, they are weighed. You are allowed a certain loss for evaporation, but still there is material *missing*. Where is it? On the beaters, the spoon, in the bowl, and on the dishrag that wiped up the spillage. Taking all this into account, one still has to decide whether the kids running through the kitchen have licked the spoon or made off with a cookie.

Let me get into the example (Fig. 1). We start out with a uranium metal alloy. The concentration of uranium in the metal is determined chemically, and the metal is weighed. A part is then machined from the alloy, and this part is weighed. The difference in the two weighings is the weight of the scrap that is gathered up and put into cans. The scrap itself cannot be weighed because it is oily; it can neither be dissolved safely nor stored safely because it is pyrophoric (i.e., it will catch on fire spontaneously). Consequently, the scrap is burned to an impure oxide and stored in cans in a vault until such time as there is enough of it accumulated for a batch to be reprocessed and until the facilities are ready. Then the ash is dissolved in an acid. The volume of the solution is measured and the concentration determined by an NDA (nondestructive analysis) device called the USAS (uranium solution assay system). At this point we make our first check: concentration × volume = total uranium. The total uranium in solution should be equal to the weight of the scrap metal times the concentration of the metal.

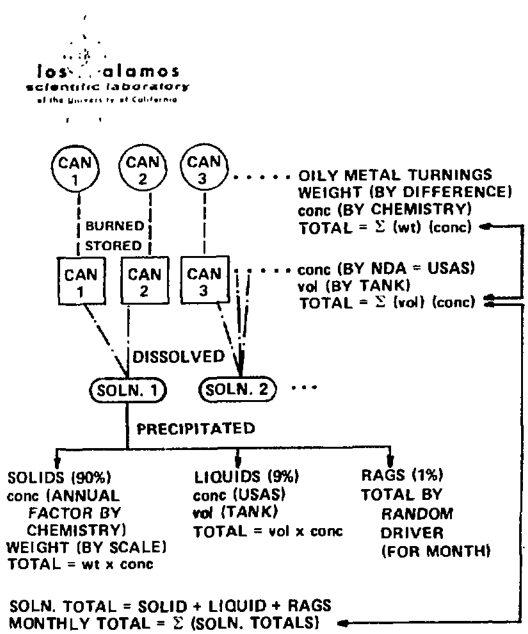


Fig. 1. Metal scrap reprocessing.

Next, something is done to the solution to precipitate the uranium oxide. About 90% of the uranium is precipitated, 9% remains in the filtrate solution, and 1% is on the rags used for cleanup. The precipitate is weighed, and its concentration determined chemically. (For obscure reasons, the concentration of the batch is not used directly. Not every batch is assayed. Instead, the annual average concentration is used. Because of the chemistry involved, this should be quite close to the analysis for any one batch. It is the 90% figure that will vary considerably.) The filtrate solution has its volume measured and its concentration determined by the USAS device. Finally, the collection of rags for an entire month (rather than a batch) is burned, and the amount of uranium in them is determined by an instrument called the Random Driver. Unfortunately the Random Driver has a much larger error than the USAS, but fortunately the amount of material involved is small.

We then add up the total uranium in the precipitate, the filtrate, and the rags, and it should check with the amount found in the solution before precipitation. The differences in the consecutive totals are called MUFs (material unaccounted for) or BPIDs (book physical inventory difference). Each point at which the total uranium may be checked is called an "account." There may be 75–100 such check points at an R&D facility such as Los Alamos.

Although we close the books on the scrap metal at the end of each month, it may be some time before we have all the figures with which to reconcile the totals. What error shall we associate with the three totals we now have? Certainly the totals have different variances. We can enter the figures as shown here (Fig. 2) in a system of multiple entry bookkeeping, but we must allow an extra column for the error or variance of each figure. With the aid of the error column, it is our job to decide whether the books balance. If they do not, there has been an arithmetical error or a diversion of material, and an investigation ensues. This system we might refer to as *statistical* bookkeeping with the statistician acting as the *auditor*.

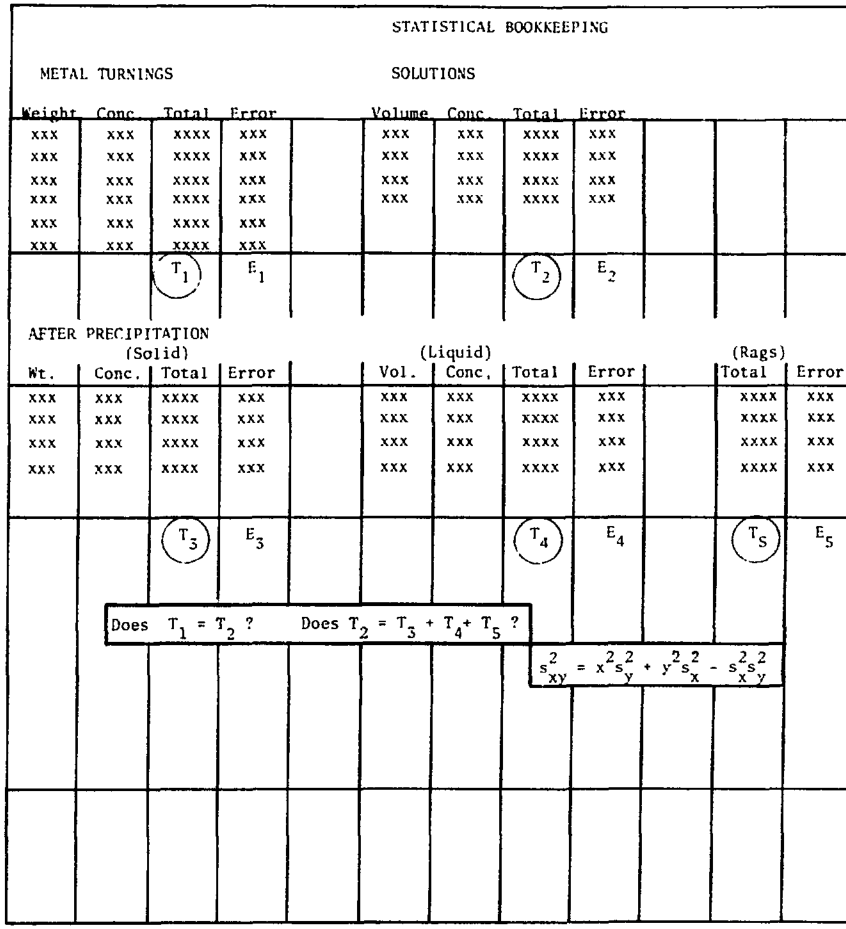


Fig. 2. A statistical bookkeeping system.

How do we calculate these errors? Each entry is a product: concentration times volume. The variance of the product cv is

$$\mu_c^2\sigma_v^2 + \mu_v^2\sigma_c^2 + \sigma_c^2\sigma_v^2 \ .$$

If we replace variances with sample variances and means with sample means, we have one estimator of the sample variance of cv, but it is biased. An unbiased estimate is

$$\bar{c}^2 s_v^2 + \bar{v}^2 s_c^2 - s_c^2 s_v^2(1 - 1/m \ - 1/n) \ .$$

The propagation of error estimate is

$$\bar{c}^2 s_v^2 + \bar{v}^2 s_c^2 \ ,$$

and it, too, is biased. Which do we use? There is still some argument among statisticians. No minimum mean square estimator seems available, the problem being seemingly intractable.

These estimates, however, do not take into account the error in fitting the calibration lines, which can be considerable. To be more explicit, there is a linear calibration line set up for the USAS device (Fig. 3), and the equation of the regression line is $y = a + bx$ where the x's are regarded as fixed (they are known standards). We use this regression line in reverse; that is, we observe y and solve for the corresponding x value: $x = (y - a)/b$. This gives us the concentration. We multiply this by the volume v of the solution and sum over the several solutions processed during the month to obtain the total uranium $\Sigma\, v_i(y_i - a)/b$. For

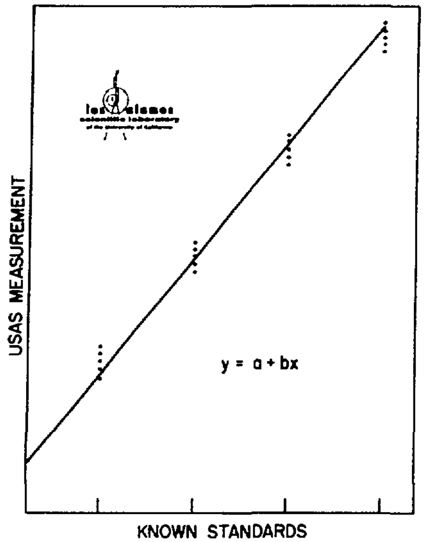


Fig. 3. Calibration line for the USAS device.

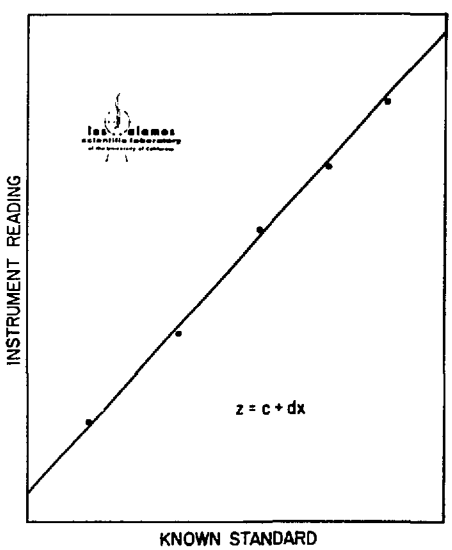


Fig. 4. Calibration line for the chemical results.

the precipitates, we have another calibration line for the chemical results (Fig. 4) with equation $z = c + dx$. Again this is used in reverse: $x = (z - c)/d$. In this case, though, we observe a large number of z's and get an annual average, $\bar{x}$, of the corresponding results as the concentration factor. Multiplying this average concentration by the weight w of a particular precipitate, we obtain $\sum w_i(\bar{z} - c)/d$. For the filtrate, a different calibration curve is used with the USAS device (a different set of standards) and we obtain $\sum q_i(y_i - e)/f$ where the q_i are volumes and the regression line is $y = e + fx$. Finally, for the rags, we use still another line, $y = g + hx$, and we use the single figure $x = (y - g)/h$. The difference between the two sums that should balance is then:

$$Q = \sum v_i(y_i - a)/b - \sum w_i(\bar{z} - c)/d - \sum q_i(y_i - e)/f - (y - g)/h \;;$$

that is,

$$\text{MUF} = \textit{solution} - \text{precipitate} - \text{filtrate} - \text{rags}\;.$$

We could, by propagation of error, find an approximate variance s_Q^2. If Q is unbiased, we would like to test whether it is zero, and if we had enough faith, we might assume asymptotic normality and look at the ratio Q/s_Q. If Q has estimated bias B, we might form the ratio $(Q - B)/(s_Q^2 + s_B^2)^{1/2}$ and compare it to a normal distribution. Is propagation of error the proper tool here?

Some of the sample variances needed for s_Q^2 may be difficult to obtain. The statistician will have to obtain the calibration results and obtain variances for each piece of equipment used. He will need to familiarize himself thoroughly with each step in the process, which will be time-consuming. The variances for volumes can be a real headache. The volume of a tank is calibrated by making marks on the side of the tank to correspond with given volumes. I had always thought that a tank volume, once calibrated, would stay calibrated, but that is not the case here. The calibration is constantly drifting. This tank has to be filled with hollow boron glass cylinders that act as moderators to keep a solution from going critical. The acid solutions eat the glass away, causing the volume to continually increase until recalibrated. Thus we get a curve somewhat like this (Fig. 5). It is not trivial to recalibrate some of these tanks. Even if you fill a tank with a measured container of water, how much air is in that water? What is the density of the water? Some large tanks have to be shielded and are sometimes calibrated as follows: Pour a known volume with a known concentration of strontium into the filled tank. Observe the concentration of the dilute solution. The ratio of the two concentrations is proportional to that of the two volumes, and you can solve for the tank volume. Not even the *weight* of an object will stay fixed in this business. We had some

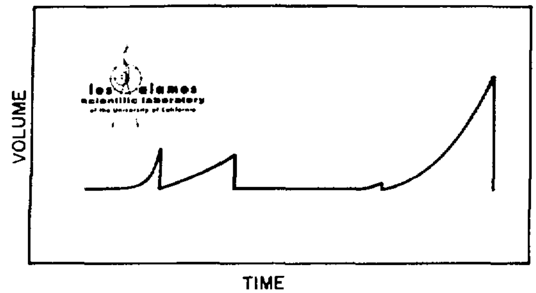


Fig. 5. Curve showing volume of a cylinder before and after recalibrations.

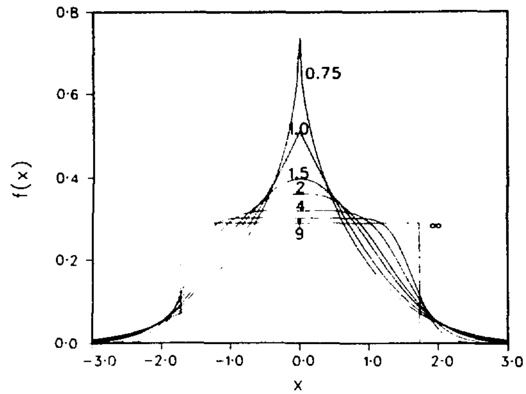


Fig. 6. Densities for selected α values.

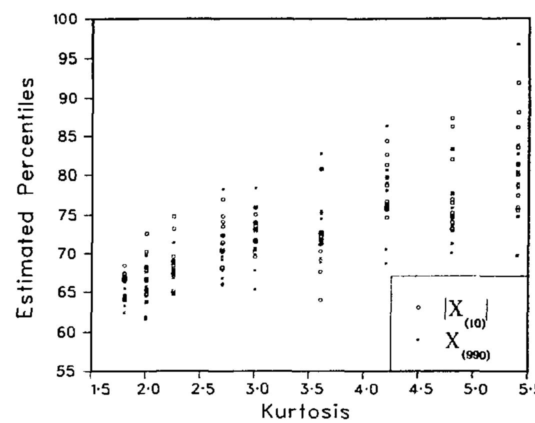


Fig. 7. Safeguards application.

uranium foil in storage and each time it was weighed, the weight was greater. It was assumed that the concentration was unchanged, and the resulting apparent change in total uranium gave the auditors fits until it was realized that the foil was oxidizing in the air!

You can get *real* increases with "holdup" in the tanks. Depending on the acidity, some of the uranium may adhere to the glass cylinders. When a more acid solution is used, you flush this off and get more uranium than you started with. A common case of holdup occurs in glove boxes. A little uranium oxide may be spilled during weighing and left in the glove box. Eventually, perhaps months later, the glove box is thoroughly cleaned, and this buildup added to the account. The result can be observed by watching the account as a function of time. Nearly every loss or low value is followed by a high value in the succeeding month. How do we model this holdup? How do we take it into account?

Another approach we have tried is simulation. We need a confidence interval for the MUF. We do not wish to rely either upon normality of Q nor upon the propagation of error approximation for the variance of Q. To do simulation, however, we shall have to assume certain distributions and parameter values for the random variables involved in Q. Mark and Myrle Johnson at Los Alamos have done some simulation work on this problem. To keep the results from being overly dependent upon a given distribution, one needs a *family* of *reasonable* distributions for the random variables. They have come up with a family, each member of which has mean zero, unit variance, and zero skewness (i.e., they are all symmetric). There is a parameter α that governs the kurtosis (Fig. 6). The family includes the uniform distribution at one extreme ($\beta_2 = 1.8$), the normal distribution ($\beta_2 = 3$), and a very peaked distribution with $\beta_2 = 5.4$. A single algorithm permits easy generation of the variables for any member of the family. They first decided to call these new distributions the NEW DIST family until someone pronounced the acronym too rapidly. By varying the kurtosis, we can study the length of the resulting confidence interval on Q (Fig. 7). This is done by generating a large number of MUFs from a given distribution, sorting them, and picking off the percentiles. We can then choose the longest interval for which we think the kurtosis is reasonable. Of course, one could study a family of asymmetric distributions by exponentiating the random variable we generate.

The simulation approach requires the same amount of work in gathering parameters and variances but has seemed a bit more reasonable and flexible to us than the straight propagation of error. We are looking for further suggestions along these lines.

The picture may be still further complicated by frequent (say, weekly) calibration, which will be insisted upon at the new plutonium facility at LASL. Then we will have to add a few more but similar terms to our expression for Q.

A more disturbing problem is *bias*. What is bias? Some of you were raised on the concept that the bias of an estimator $\hat{\theta}$ is $E(\hat{\theta}) - \theta$. That, by definition, seems to make the bias a *constant*. In a series of influential papers, Churchill Eisenhart at NBS gave a very similar definition, but he has replaced $E(\hat{\theta})$ with the *limiting mean* μ of a set of measurements (under identical circumstances) on a quantity. He then says "the systematic error or bias . . . of a measurement process will ordinarily have *both* constant and variable components." That makes bias a random variable. He illustrates by considering a distance measured with a steel tape. The temperature on the day on which the measurements are made adds a random variable into the limiting mean, hence into the bias. The term "limiting mean" is not so well defined, and thus the concept has expanded into "long-term" and "short-term" systematic errors, which may be either constants or random variables. Not understanding each other, there have been vociferous arguments among statisticians within our DOE community about bias and systematic error and how to correct for them and when not to correct for them, etc., with everyone using his own definition of bias. May I give an example of what confuses us as statisticians, and even more confuses the experimenter? It is to get a form like some we see from the EPA and NBS asking for a series of measurements to be used, say, in standardizing a new method. Here are the questions the experimenter is required to supply under the heading of *Calibration Results*: (1) What is the overall uncertainty on the value of the activity? (2) What is the standard error? (3) Give a 99% confidence limit. (4) The total estimated systematic error is __, comprised of __% due to __ and __% due to __, etc. (5) How are the systematic errors combined? (6) How are the random and systematic errors combined? To fill out such a form requires *agreement* on what the terms mean, and I don't think we have yet reached that agreement among ourselves. We need to do some housecleaning. Indeed, we may be a little disturbed about filling out the form because we think *they* might misinterpret or misuse what we say. I am trying to say that this chemistry business is swarming with biases and systematic errors. The Random Driver, for example, has large errors for small amounts of uranium, moderate errors for moderate amounts of uranium, and large biases for large amounts of uranium. I would like to get a colloquium started on that issue.

Problem Discussion 1, Part 2: Statistical Aspects of Nuclear Safeguards

Gary Tietjen, Los Alamos

Gary Tietjen: I described yesterday a system of multiple entry bookkeeping which I called a statistical bookkeeping for accountability with an added error column. One of the questions was how do you calculate that added error column and do you use it to reconcile the total. I suggested several ways of calculating a variance there and asked which one of those we should use. I gave a mathematical expression for the MUF in one particular case, MUF being the material unaccounted for, and asked if we should try to propagate the error on this MUF to test a hypothesis set at zero. Shall we use simulation to accomplish that purpose or is there something better. How shall we handle changing calibration on volume; how shall we model holdup; and finally, how shall we handle questions of bias in deciding whether this MUF is zero. Let me just make one remark about the holdup. The situation frequently appears something like this (Fig. 1): as we observe the MUF, it will first go below the line and then above the line; the next month there will be a compensating factor, and one will get a curve something like this. There seems to be a type of regularity about this discrepancy as you get one high value, then followed by a low value, and so forth. Perhaps one could do something with that by plotting the values that fell below some line (perhaps zero) against those that fell above that line (Fig. 2) to see if there was perhaps some relationship between those points that fell below and those that fell above the line (with a lag of one month say or several months depending on the process). If there was, then maybe one should infer something about modeling the holdup. Now with relation to the bias, I was just going to say this: Eisenhart gave the example of a steel tape that was calibrated, and to calibrate a steel tape you need to take the temperature of the day into account. He talked about using a steel tape to measure a distance. On a particular day, however, if repeated measurements were made at a temperature below the temperature at which this thing was calibrated, then you would get a kind of bias on that particular day, and that would be a random variable in the sense that it would vary from day to day. So there he's talking about bias that's a random variable, and we as statisticians usually talk about a bias as being something that is constant. Jim Lechner said that his boss at the National Bureau of Standards would like to reserve the term "bias" for that which was characteristic of the measurement process and could not change from day to day. It was a long-term kind of thing, and I would like to talk about perhaps claiming some other terminology or something on

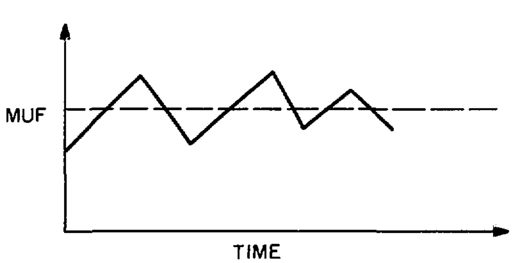


Figure 1

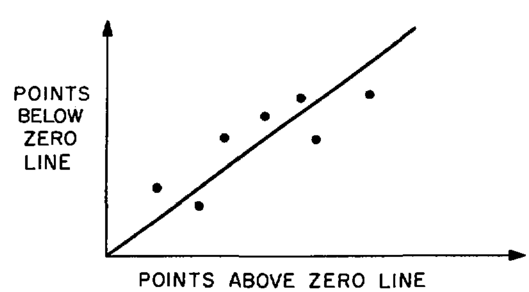


Figure 2

which we could agree and which would help us to describe those sorts of situations.

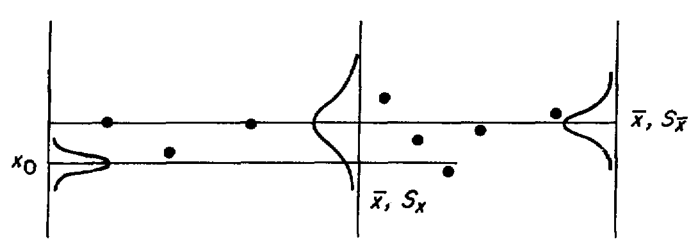
Sylvester Suda, Brookhaven: I have two examples of the variable components of systematic error that I can identify. One of them involves any calibration line. Since Gary has already brought up the problems of volume calibration, I'll start with that as an example.

The typical way to calibrate a tank is to start with two known quantities—a tank and a response system. You don't calibrate the tank; you calibrate the system. Typically in a larger facility, your tank doesn't seem to change. What you really want to worry about is the sensing instrument changing. The typical way is to advance, collect data, and have one calibration line; and you do it again. This is exaggerated, but you do have, if you repeat the calibration so that you get enough data, to assess the uncertainty associated with the calibration process here. There are techniques that are used in the analysis of covariance, which are in Brownlee and have been described in the first paper that John Shepherd and I put together, of looking at this data and deciding whether these calibration lines, in fact, are similar. If they are similar, you can then combine the data and get one average calibration line. Although this calibration line has some uncertainty, you can set confidence intervals on this thing, and we know that the confidence interval tells us that the true line lies somewhere within the band. The calibration line will have some bias associated with it, as, for example, it will be 2 liters too high or too low, and every time we use that value, we won't know. So here we have this variability, this uncertainty associated with the calibration line where, in fact, it is a systematic error that has to be included in the uncertainty associated with that volume determination. Then, of course, you do have operator error, and a lot of things could happen at the time of measurement; there are ways of determining that random error.

You have a similar situation with a weighing system if you have some nominal value, x_0, that you're using as a standard; it is a very good standard, and you get it from the Bureau of Standards. It will have some limit of uncertainty associated with it, however. If you run all month and record observations associated with this thing so that you're measuring a complete program, at the end of the month, first of all, you can look at the distribution of the x's—$\bar{x}$ and some s_x (Fig. 3). You have the distribution of $\bar{x}$ and $s_{\bar{x}}$, and we can define, in this case, this distance here, as an estimate of the bias, which is equal to $\bar{x} - x_0$. I can well attest or determine whether, in fact, that bias is significant or not. If it is, I want to make a correction for it. Let's assume I do make this correction, I have some uncertainty in the way that this $\hat{\beta}$ was determined. I can take the variance of both sides here, and if I assume they are independent, I now have something that I define as the constant complement of systematic error and the variable complement of systematic error. I'd like to hear people's reaction to this.

Jim Lechner, National Bureau of Standards: Having had my name taken more or less in vain by Gary a few minutes ago, I think I need to come up and make a correction here. Apparently, I didn't make myself clear, Gary, on the bus the other day. I would not like to have you quoting me misquoting my boss. That's not healthy.

The term bias (as learned in elementary statistics courses) is the difference between the expectation (that's the statistical term) of an estimator and the quantity that you would like to be estimating.



$$\longrightarrow \hat{\beta} = \bar{x} - x_0$$
$$\longrightarrow \mathrm{Var}(\hat{\beta}) = \mathrm{Var}(\bar{x}) + \mathrm{Var}(x_0)$$

Figure 3

Therefore in order to define a bias, you not only need an estimator of statistics (which was defined), which might be the sample mean, the sample standard deviation, the median, or who knows what else, but also you have to know what quantities you're estimating or wishing you were estimating. Then you can define the bias: it is not so much that my boss would like to have the bias be a permanent characteristic of the measurement process as that it be this mathematically defined quantity depending on both what you are calculating and what you hope you are estimating. This will change each day if the quantity you're calculating is calculated differently for example, by a different calibration line or for day-to-day effects (operator, instrument, or other) that creep into your process. So indeed, the bias does change day to day if there are these effects, as there almost always are. On the other hand, the term "systematic error" has been bandied around in many different directions by many different people John Jaech referred to this last night. The feeling I get from talking to a lot of people (I must admit that most of them are at NBS) is that the term "systematic error" is one they would like to see reserved for a permanent kind of unusable, unchangeable characteristic of a measurement process. Now, anytime you get the measurement, there are components of the error in that measurement, some of which are fairly random in the sense that you can repeat the measurement six times and have six truly independent observations on that component of error. Other components are random, but they only vary every month or with a change of operator or something of that sort. These latter kinds of random components are often included in systematic error, and I'm "guilty" of having done this oftentimes myself. I talk about systematic errors that vary over long periods of time, and then I think I'm not being really careful with my language. I don't know what term to use for these components, other than just calling them components of the error variance. Now that's a precise term that has a well-defined meaning; people know what it says for the most part, I believe, and you don't get into the problem that a lot of us have with the term "systematic error variance." If systematic error is a life-long constant, then it can't have a variance as do the components of error which are like systematic errors for an entire run (maybe all week long or whatever); therefore, we don't like to call them systematic errors. Last night John Jaech said that we should educate people to expect statisticians to disagree. The interpretation of data is something where there's much room for disagreement, more so than we like to see many times, but on terminology we ought to be able to come to an agreement. If I say red, I hope other people know what I mean; and if I say bias, I would like them to know what I mean; and if they say bias or systematic error, I'd sure like to know what they mean. I really think we ought to have some terminology we can all agree on, even if we do disagree on how to interpret data.

Gary Tietjen: I would like to thank Jim for that, and I did misquote him I'm glad he got up and corrected me on that. I wonder if he could make an additional comment or two about combining these types of quantities and what they, at NBS, feel should be done.

Jim Lechner, National Bureau of Standards: Most of my practicing career, which has been longer than I care to admit, has been spent as a probabilist, and not too much as a statistician. I don't speak from lots and lots of experience computing data; however, I've had some. As a general rule, NBS strongly pushes the idea of stating your systematic errors separately from your random errors, not trying to combine the two, because the way they should be combined depends on the use to which you are going to put the results. There is one particular case I will mention. A scientist at NBS came up with a value for, not a basic physical constant, but a definite something that has a true value however you want to define true value of two orders of magnitude better than had been known before. I got his paper, which had a discussion of the systematic and the random errors, for review, and we had to knock some heads together before we could come up with something we both could agree on. In there he had random errors of which he could estimate the variance by his internal repetitions. He also had random errors which were constant throughout his entire experiment. Nevertheless, they are random errors, and he knew the variance in this case because he had worked with the same apparatus so many times and in so many different ways. He knew there were components of error in there, but he knew how they varied from experiment to experiment, so he could say, "I have in here one component of error; I have only one; if it's in there, it's in the whole set of experiments, so I can't have any hint internally on what it is. However, there's a random draw from the distribution which is essentially normal between zero and variance, so he should include that in his random error. You can have random errors which look like systematic errors for your entire set of runs, but you've got to know what you're doing. Basically, your answer for that is don't try to combine the systematic and the random errors.

Let the reader combine them the way he knows is best for his use.

Keith Ziegler, Los Alamos: I want to get away from the subject of bias and systematic errors, because I don't understand them. I would like to discuss the calculation of the uncertainty associated with the MUF. One of the things Gary talked about yesterday was the calibration curve. In the ordinary sense, the calibration curve is written as $y = a + b(x - \bar{x})$, because that is the way I like to write it. But you never use the calibration curve in that sense; you always fall for x as equal to something in terms of y. In a rather controversial paper that was published in *Technometrics* not too long ago by Krutchoff, he advocated that you can get a smaller mean square error if you will not do this type of calibration but actually set $x = c + d(y - \bar{y})$; proceed as though y were the independent variable and x were the dependent variable, and complete the calculation. I would like to suggest that people look at this very strongly when they are calculating their MUF, because it is much easier to calculate the variance on this quantity, and it does seem to give a smaller mean square error. As I have said, this particular paper created quite a controversy when it came out. There was lots of rebuttal and argument as to why this wasn't a really good method. However, when the simulation is studied, it *is* a pretty good method. Because of all the other approximations that go into calculating the MUF, I think that this would be a very minor perturbation on the total calculation.

Now we will not lead the audience to believe that I do not talk to Gary. I know that Gary is fully aware of my views on this, and so he has indicated these ideas at Los Alamos Scientific Laboratory.

Lincoln Moses, Stanford University: In her Ph.D. dissertation, Katherine Lamborn considered two problems. One of them was the Krutchoff paper. She had theoretical arguments indicating that the results he had arranged to be published were quite true, but if you reached outside that range (her analytical results confirmed her simulation was for the range she chose), everything went to hell in a handbasket.

Sylvester Suda, Brookhaven: I'd like to comment on Keith's statement regarding the smaller mean square error. In this propagation of error and LEMUF analysis you've got to be careful about actually seeking the smallest mean square error, because these numbers will be used against you if you claim your measurements are too good, and the inspection facility shows up and samples one of the kinds of things that you measured. We are going to have IAU inspectors in U.S. facilities shortly; the determination of whether you have made an inaccurate statement of your inventory will be based on how many of these differences they discover. Therefore, it is not the small or mean square error that is something you look for. What you need is a realistic estimate of how well you are measuring, so that you are not being cooked in your own juices because you've claimed a smaller limit of error than you really have.

Keith Ziegler, Los Alamos: I guess that I am a little bit in disagreement with the other statistician. I would like to come up with the best estimate but the smallest mean square error. Part of the purpose of the whole safeguard program is to attempt to detect diversion if there is diversion. So you would like to have the smallest reasonable mean square error that you could get, if you're really going to be looking at diversion. It's not the statistician's role, as I see it, to come up with a wide error just so you can explain the MUF.

Carl Bennett, Battelle: I just want to say to Keith that my problem with MUF's over the years has not been with the lack of significance. My problem has been to explain the fact that almost 50% of them usually turn out to be significant. In other words, I guess what happens in this business of using MUF as an index of diversion is trying to eliminate the things from that -- particularly biases, and frequently other sources of error and variability—that tend to make the rather synthetic variances we create so the use of things like LEMUF and so forth are considerable underestimates (and this may be what Syl was talking about) of what the true variability in a measurement is. You can go through life as a statistician in a chem-plant trying to explain why we had significant MUF and why we had MUF that was not consistent with the errors that should have been assigned. There are many notable examples of fairly significant chemical advances and process advances that have come out of the analyses of these significant MUF, particularly MUF which persisted in being significant for let's say 12 months or 24 months in succession. It is this kind of information from the data that says, "Yes, you do need a consistent error of estimate; you need a consistent error, or you need a good knowledge of what your measurement error is and how well you can trust these things." From then on you have to explain all of the other sources of both variation and bias which enter into that index.

Generation of a Typical Meteorological Year

I. J. Hall and R. R. Prairie

Sandia Laboratories
Albuquerque, New Mexico

ABSTRACT

Technology in solar energy is moving very rapidly. For any given solar energy system an important consideration is its performance. There are many methods available for assessing performance, all of which require meteorological input. Thus, a question that is asked is how well will a given solar energy system perform over a typical year for a given location. Our problem is to develop a typical year. A typical year will not be an average but will be a profile of typical fluctuations that occur in weather over a period of time.

There are data available from twenty-six different locations. Each location has hourly data on six different meteorological variables over a number of years. Our intent is to develop a model that represents the joint behavior of the six variables and that can be used for generating a typical year. A set of criteria will be developed based upon the data, and the generated typical year will be checked according to the criteria.

SANDIA ENERGY ACTIVITY

The management structure at Sandia has seven vice-presidents. Under the Vice-President of Research there is a Directorate of Energy and Systems Analysis which has most of the nonnuclear energy work. This Directorate is made up of three departments (Fig. 1): the Solar Energy Department, the Geoenergy Department, and a Systems Analysis Department. The projects in these departments include a total energy program generating both electricity and heat; a solar thermal test tower; wind turbines; and research in the areas of photovoltaic technology, geotechnology, oil shale and drilling technologies, and technical management of DOE's solar irrigation efforts. About 350 people are working in these areas.

DIRECTOR OF ENERGY PROJECTS AND SYSTEM ANALYSIS

- SOLAR ENERGY DEPARTMENT
- GEOENERGY DEPARTMENT
- SYSTEMS ANALYSIS DEPARTMENT

Fig. 1. Sandia energy directorate.

SANDIA PARTICIPATION IN DEVELOPING A TYPICAL METEOROLOGICAL YEAR

A few years ago, a colleague in the Solar Energy Department discussed with members of the Statistics Division some of the statistical problems that he saw in the solar energy area. One problem was the lack of direct normal (DN) radiation data. Direct normal radiation is the energy that is received directly from the sun and does not include the diffuse radiation—radiation received from the clouds, for example. Twenty-six weather stations had collected total horizontal radiation data. Total horizontal (TH) radiation (DN plus diffuse) is the energy received on a flat plate horizontal to the earth's surface. The majority had not collected DN; however, a few weather stations (Albuquerque, New Mexico; Blue Hill, Massachusetts; and Omaha, Nebraska) had both DN and TH radiation data. The availability of these data allowed us to establish an empirical relationship between the TH and DN values for these three locations. This empirical relationship made possible the construction of maps like the one shown in Fig. 2*a*. Figure 2*b* shows the isopleths of TH radiation for January. For some time, TH maps of this nature have been available. Examination of the TH map shows that the isopleths have a tendency to be in an east-west direction across the United States. As shown in Fig. 2*a*, some of the DN isopleths have a tendency to lie in a north-south direction. The map indicates that the northern latitudes receive more DN radiation than one might expect by looking

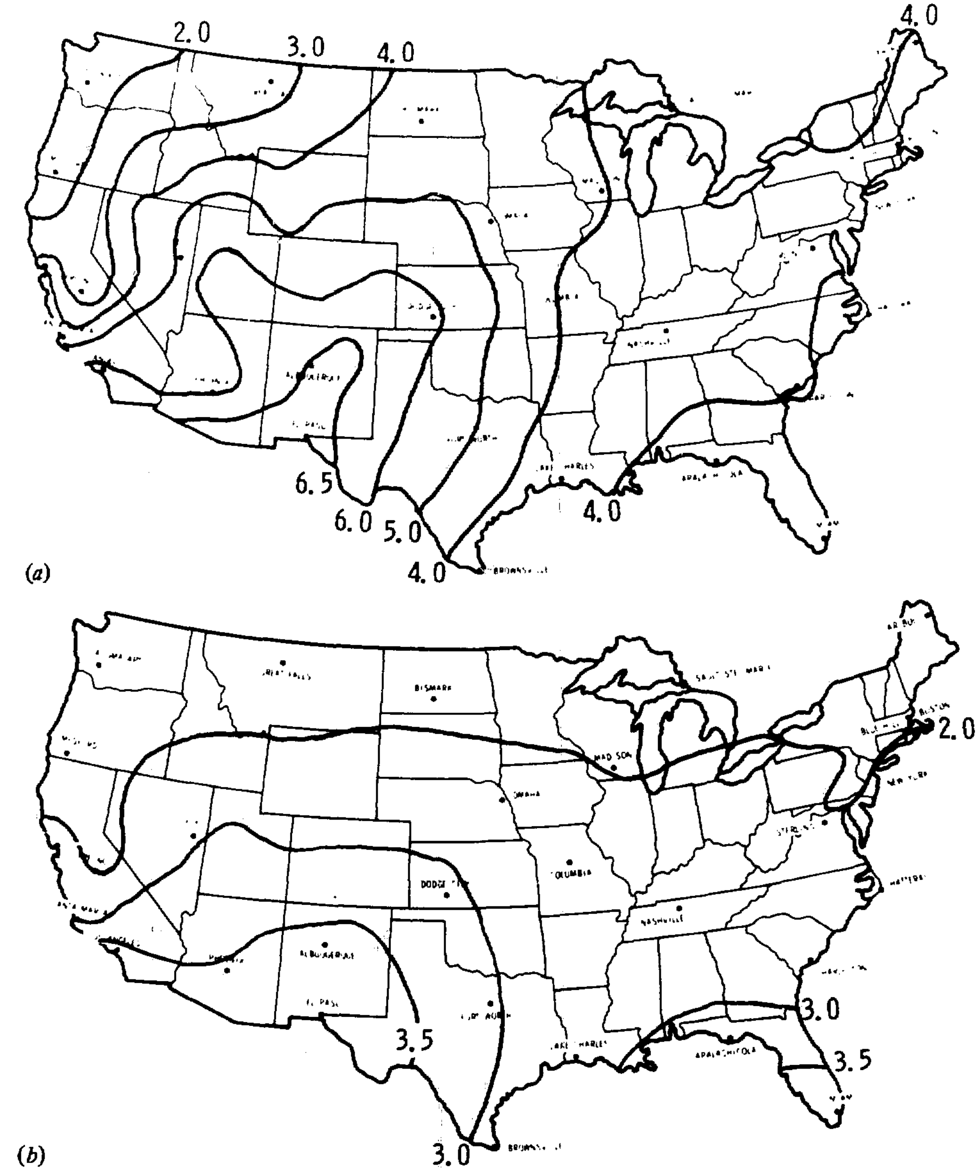


Fig. 2. Solar radiation for January (kilowatt-hours per square meter). (*a*) Mean daily direct normal radiation; (*b*) Total horizontal radiation.

only at the bottom map. The large amount of northern DN radiation was somewhat of a surprising result to us.

Contact with personnel in the solar energy group continued after the above-mentioned project was completed. Earlier this year our colleague in this group suggested that we submit a proposal to the DOE to generate a typical meteorological year (TMY). The proposal was accepted, and we are now in the process of selecting a method to generate a TMY. The motivation for a TMY comes from the need for a common weather data base for each of the 26 stations so that energy systems can be sized and compared.

DATA BASE

The existing data base, to be used for the generation of the TMY, consists of hourly meteorological measures on five variables—dry-bulb temperature, dew point, TH radiation, wind speed, and wind direction. There are 12 years of hourly data available plus a few years of 3-hr data. The data are available at 26 weather stations. The map in Fig. 3 shows the locations of the weather stations. A TMY is to be generated for each of the stations.

The National Climatic Center in Asheville, North Carolina, has had the responsibility of "rehabilitating" the data from each of the locations so we do not have the job of "cleaning it up."

CRITERIA FOR SELECTING A TMY

One of the problems to be faced in developing a TMY is to decide on what indexes might characterize a typical year. As a first step, some summary statistics were obtained. For the dry-bulb temperature, dew point, and wind velocity variables, the mean and the variance were calculated. Also calculated were daily minimums, daily maximums, and daily ranges. Distributions of these variables were determined, an example of which is given in Fig. 4. This figure shows the distribution of the dry-bulb daily maximum temperature for January 1953 at Lake Charles, Louisiana. The data consist of 372 readings—the daily maximum temperature for 12 years. The percentages are also given and are shown in Fig. 4. Distributions for the other variables and other months are being generated for all the locations.

For the solar radiation variable, daily total radiation is calculated and its distribution and summary statistics are determined.

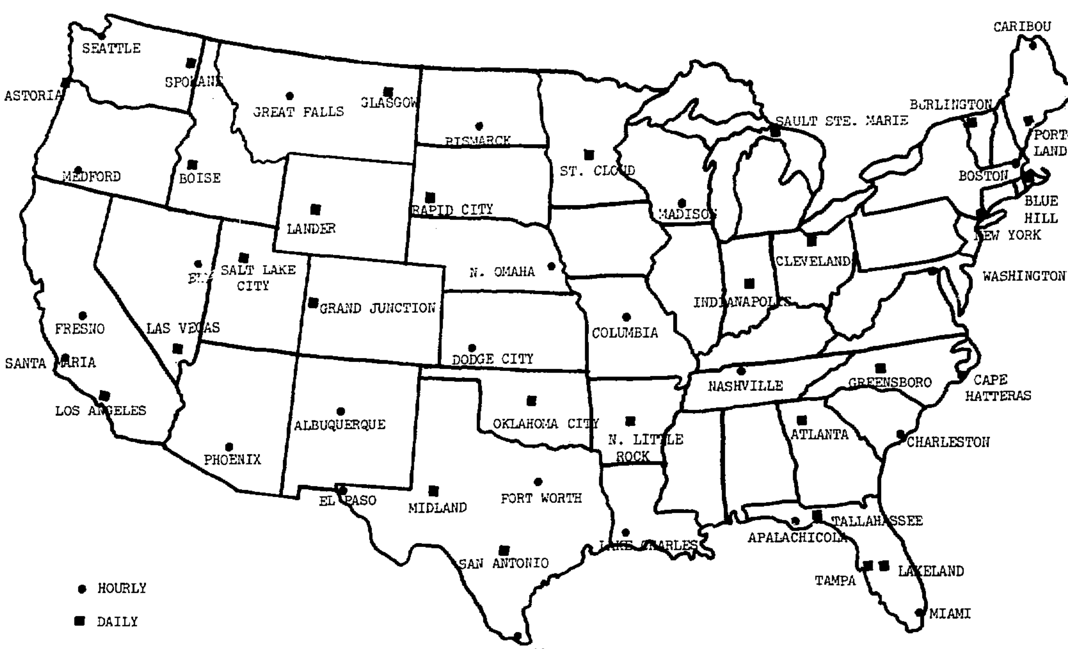


Fig. 3. Solar radiation rehabilitation stations.

```
STATION  3037    JAN    DRY BULB     (MAX.)

N                               .37200E+03
MEAN                            .152[illegible]E+02
STD. DEV.                       .57217E+01
VARIANCE                        .32730E+02
MAX. X                          .25000E+02
MIN. X                         0.
RANGE                           .25000E+02

   MIDPOINT            FREQ

  -.500000E+00          0
   .500000E+00          6     XXXXXX
   .150000E+01          1     X
   .250000E+01          3     XXX
   .350000E+01          7     XXXXXXX
   .450000E+01          2     XX
   .550000E+01          5     XXXXX
   .650000E+01          9     XXXXXXXXX
   .750000E+01         12     XXXXXXXXXXXX
   .850000E+01         11     XXXXXXXXXXX
   .950000E+01          7     XXXXXXX
   .105000E+02         17     XXXXXXXXXXXXXXXXX
   .115000E+02         24     XXXXXXXXXXXXXXXXXXXXXXXX
   .125000E+02         14     XXXXXXXXXXXXXX
   .135000E+02         24     XXXXXXXXXXXXXXXXXXXXXXXX
   .145000E+02         15     XXXXXXXXXXXXXXX
   .155000E+02         31     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
   .165000E+02         25     XXXXXXXXXXXXXXXXXXXXXXXXX
   .175000E+02         20     XXXXXXXXXXXXXXXXXXXX
   .185000E+02         32     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
   .195000E+02          8     XXXXXXXX
   .205000E+02         31     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
   .215000E+02         22     XXXXXXXXXXXXXXXXXXXXXX
   .225000E+02         25     XXXXXXXXXXXXXXXXXXXXXXXXX
   .235000E+02         12     XXXXXXXXXXXX
   .245000E+02          5     XXXXX
   .255000E+02          4     XXXX

                      372

X(01) =     0.00
X(05) =     4.40
X(10) =     7.20
X(25) =    11.70
X(50) =    15.60
X(75) =    20.00
X(90) =    22.20
X(95) =    23.30
X(99) =    24.40
```

Fig. 4. Distribution statistics of January daily dry-bulb temperature maximums for Lake Charles, Louisiana, for 1953–1964.

In addition to these statistics, persistence was also determined. Persistence refers to consecutive days possessing a given characteristic, such as five consecutive days of cold weather. A solar engineer has to worry about such strings in designing a solar system. How much storage he needs to build into his system is affected by such strings. Persistence has been measured by runs of days possessing a certain characteristic. Figure 5 gives the number of runs of "cold days" for the years 1953–1964 for Lake Charles, Louisiana, in January. Here, a cold day means that the daily minimum temperature is less than the 25th percentile over all 12 years. Figure 5 shows, for example, that twice in January 1961 the minimum temperature was less than the 25th percentile for two days in a row and once it was less for three days in a row. From the information in the figure, we can calculate the average number of runs per year and the average run length.

Figure 6 gives similar information on the number of runs in which the maximum wind velocity exceeded the 75th percentile of the maximum wind velocity distribution.

In addition to single-variable persistence, two-variable persistence is important. Figure 7 shows a two-way table for temperature and radiation. The blocks with X's are the pairs which are of primary interest. The lower left X indicates days in which both the temperature and radiation is low—a cold and cloudy day.

Figure 8 gives the number of "cold and cloudy" days. Here cold means the minimum daily tempera-

STATION 3937 JAN
DRY BULB (MIN) <=25 PERCENTILE

YEAR

R.L.	53	54	55	56	57	58	59	60	61	62	63	64	
1	1	1	2	1	0	0	1	2	0	1	1	1	11
2	0	1	1	1	0	1	1	0	2	2	1	2	12
3	0	0	0	1	0	1	3	0	1	0	0	0	6
4	0	1	0	0	1	0	0	0	0	0	0	1	3
5	0	0	0	0	0	0	0	0	0	0	1	0	1
6	0	0	0	0	0	0	0	0	0	1	1	0	2
7	0	0	0	0	0	0	0	0	1	0	0	0	1
8	0	0	0	0	0	0	0	1	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	3	3	3	1	2	5	3	4	4	4	4	37

AVE. NO. OF RUNS/YEAR 3.1

AVE. RUN LENGTH 2.6

Fig. 5. Run length frequencies of January daily dry-bulb temperature minimums less than the 25th percentile for Lake Charles, Louisiana.

STATION 3937 JAN
WIND VEL. (MAX) >=75 PERCENTILE

YEAR

R.L.	53	54	55	56	57	58	59	60	61	62	63	64	
1	0	2	5	0	2	5	2	4	2	1	3	3	29
2	0	2	0	0	2	0	1	1	0	0	3	2	11
3	1	0	0	1	0	0	2	0	0	2	0	1	7
4	1	1	0	0	0	0	0	0	0	1	0	0	3
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	2	0	0	2
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	0	0	0	1
10	1	0	0	0	0	0	0	0	0	0	0	0	1
	4	5	5	1	4	5	5	5	2	6	6	6	54

AVE. NO. OF RUNS/YEAR 4.5

AVE. RUN LENGTH 2.2

Fig. 6. Run length frequencies of January daily dry-bulb temperature maximums greater than the 75th percentile for Lake Charles, Louisiana.

ture was less than the 25th percentile, and cloudy means the daily radiation was less than the 25th percentile. The figure shows that there was one two-day period in which the weather was cold and cloudy. Figure 9 gives the number of runs of sunny (daily radiation greater than the 75th percentile) and cold (daily minimum temperature less than the 25th percentile) days.

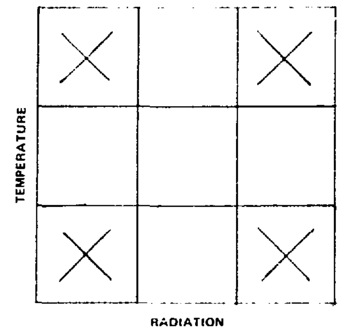


Fig. 7. Joint persistence of dry-bulb temperature and solar radiation.

STATION 3937 JAN
JOINT SOLAR RAD. AND DRY BULB (MIN) <=25% <=25%

YEAR

R.L.	53	54	55	56	57	58	59	60	61	62	63	64	
1	0	2	0	0	1	0	1	2	1	1	2	0	10
2	0	0	0	0	0	0	0	0	0	0	1	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	1	0	0	0	1
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	2	0	0	1	0	1	2	2	1	3	0	12

AVE. NO. OF RUNS/YEAR 1.0

AVE. RUN LENGTH 1.3

Fig. 8. Run length frequencies of January joint minimums of solar radiation and dry-bulb temperature less than the 25th percentile for Lake Charles, Louisiana.

METHODS FOR GENERATING A TMY

To understand what is to be done, it is useful to look at some data in terms of a time series. In Figs. 10 and 11 two such plots are shown.

One approach for constructing a TMY is purely empirical. "Empirical" means that existing segments of data are selected according to some criteria and then pieced together to form a year. For example, the month of January 1963 may be typical for all the Januarys for which we have data, which is mated with February of 1955, etc., for a given location. This approach has the advantage of using actual data and of avoiding the mathematical difficulty with the correlation structure.

STATION 3937 JAN
JOINT SOLAR RAD. AND DRY BULB (MIN) >=75% <=25%

R.L.	53	54	55	56	57	58	59	60	61	62	63	64	
1	1	2	1	1	0	1	1	3	3	1	3	1	18
2	0	0	1	1	1	1	2	1	1	2	1	0	11
3	0	0	0	1	0	0	0	0	0	0	0	1	2
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	2	2	3	1	2	3	4	4	3	4	2	31

AVE. NO. OF RUNS/YEAR 2.6

AVE. RUN LENGTH 1.5

Fig. 9. Run length frequencies of January joint minimums of solar radiation and temperature greater than the 75th percentile for radiation and less than the 25th percentile for temperature for Lake Charles, Louisiana.

A second approach is to fit the data with autoregressive-type models and then generate weather data based on the model. Figure 12 shows two models which have been tried for the temperature data. The H_i ($i = 1, \ldots, 24$) in the model 1 is an hour effect, and the rest of the model consists of autoregressive terms which attempt to relate the present temperature to the temperature of (a) 1, 2, and 3 hr ago, (b) 23, 24, and 25 hr ago, and (c) 47, 48, and 49 hr ago. Model 2 in the figure contains trig functions to account for the daily cycles. Both models have been used to fit the Lake Charles temperature data for all years. The fits have been very good, $R^2 \approx 0.98$, $\sigma = 0.8$. Model 2 contains fewer parameters than the other model. Figures 13 and 14 give examples of simulated results from both models. The models were based on 1957 Lake Charles temperature data. The simulated results do not appear to be unreasonable.

PROBLEMS

1. What criteria should be used to determine when a reasonable model has been determined? In Fig. 15 one method is shown.
2. If a model is selected, how should a typical year be selected? In Figs. 16 and 17 some ideas are given for comparing generated radiation and temperature data with actual data.
3. Wind direction: How do we handle the variable in which a wind direction of 2° and 359° are almost the same but yet are numerically quite different? Figure 18 shows a summary of wind direction for Lake Charles in July. This type of summary may or may not be useful.
4. Multivariate time series: We have five simultaneous weather measurements or a multivariate time series. How do we correctly model these data?
5. How do we know when we have a typical year? We need to know what parameters will characterize a TMY.
6. How do we combine data over years? Do we take averages or should we do something else?

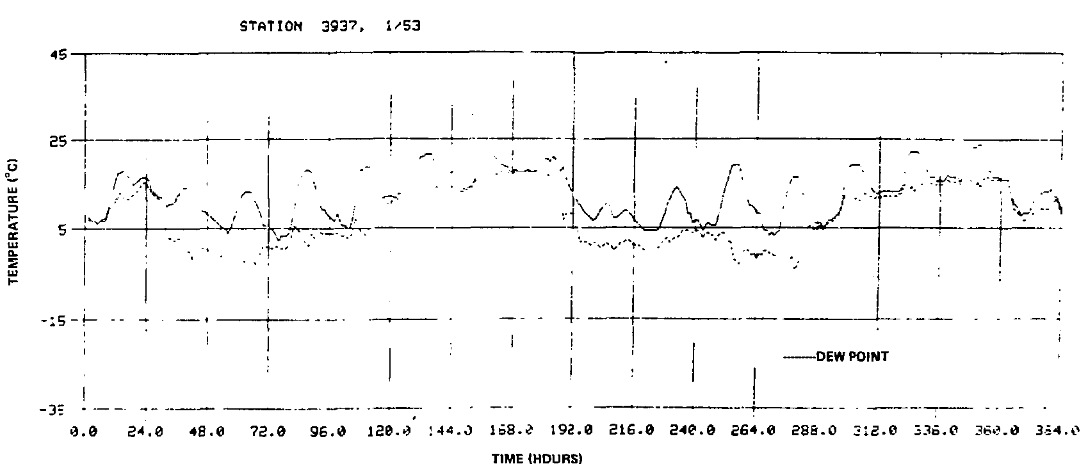


Fig. 10. Plot of dry-bulb and dew point temperature for January 1953 at Lake Charles, Louisiana.

7. How do we use 3-hr data? Our TMY must be on an hourly basis. One possibility is to use the 12 years with hourly data and forget about the 3-hr data.

8. How do we adjust for long-term cycles? Perhaps 12-15 years of data are not enough to detect any long-term cycles.

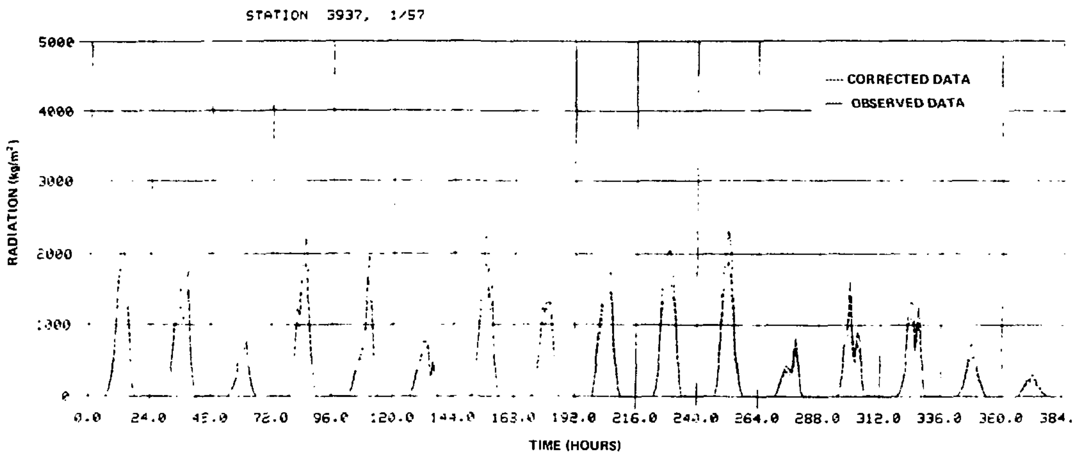


Fig. 11. Plot of solar radiation for January 1957 at Lake Charles, Louisiana.

$$\textit{Model 1: } X_t = \alpha + H_i + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-23} + \beta_5 X_{t-24} + \beta_6 X_{t-25} + \beta_7 X_{t-47} + \beta_8 X_{t-48} + \beta_9 X_{t-49} + \epsilon_t \text{, for } i = t \pmod{24}.$$

$$\textit{Model 2: } X_t = \alpha + \sum_j (a_j \cos w_j t + b_j \sin w_j t) + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-23} + \beta_5 X_{t-24} + \beta_6 X_{t-25} + \epsilon_t \text{, for } w_j = \frac{2\pi}{24}, \frac{2\pi}{12}, \frac{2\pi}{6}, \frac{2\pi}{3}$$

Fig. 12. Autoregressive models used to fit temperature data.

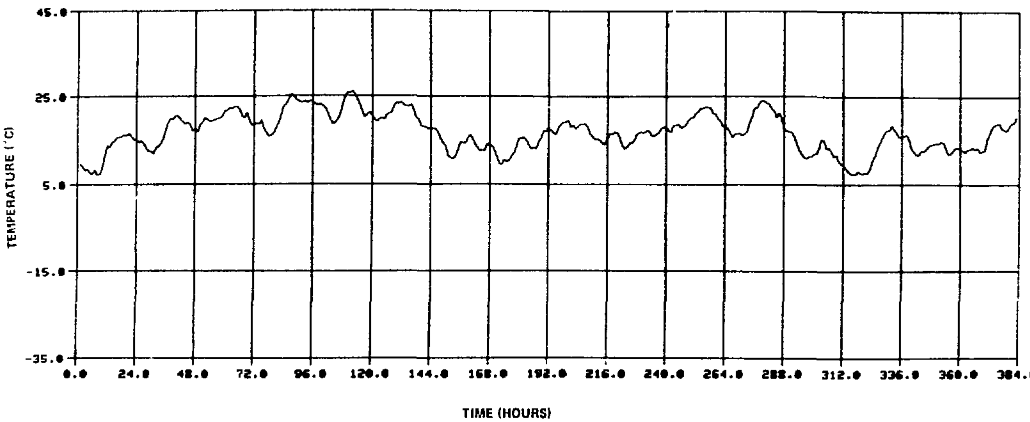


Fig. 13. Simulation results from model 1 based on Lake Charles 1957 temperature data.

(Model 1: $X_t = \alpha + H_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-23} + \beta_5 X_{t-24} + \beta_6 X_{t-25} + \beta_7 X_{t-47} + \beta_8 X_{t-48} + \beta_9 X_{t-49} + \epsilon_t$.)

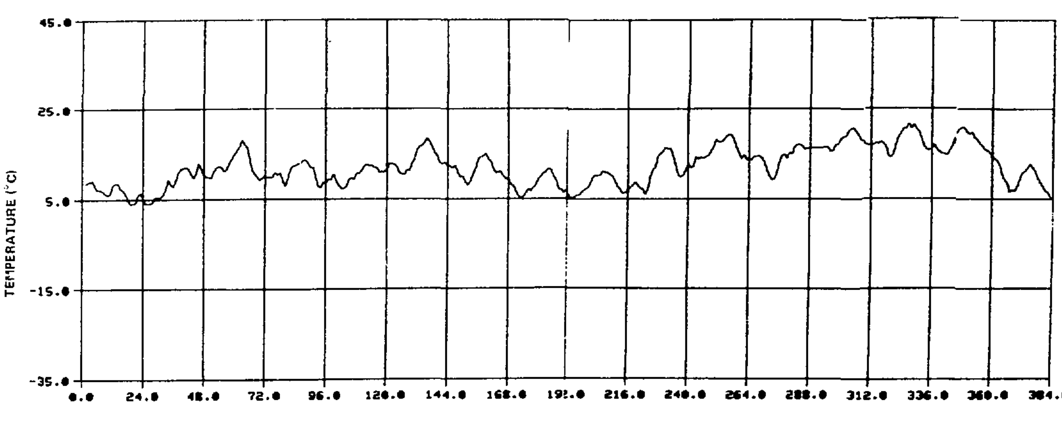


Fig. 14. Simulation results from model 2 based on Lake Charles 1957 temperature data.

(Model 2: $X_t = \alpha + \sum_{i=1}^{3} (a_i \cos w_i t + b_i \sin w_i t) + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-23} + \beta_5 X_{t-24} + \beta_6 X_{t-26} + \epsilon_t$,

where $w_1 = 2\pi/24$, $w_2 = 2\pi/12$, and $w_3 = 2\pi/6$.)

- GENERATE 12 YEARS OF DATA
- CALCULATE SUMMARY STATISTICS AUTO-COVARIANCE AND PERIODOGRAM
- COMPARE WITH 12-YEAR DATA SUMMARY STATISTICS AUTOCOVARIANCE AND PERIODOGRAM

Fig. 15. Criteria to check adequacy of model.

Y = TOTAL DAILY RADIATION
FOR OUR TMY CALCULATE

$\bar{Y}, S_Y^2$, HISTOGRAM OF Y

AUTOCOVARIANCE, PERIODOGRAM

NUMBER OF RUNS OF Y ABOVE $\xi_{0.75}$
AVERAGE LENGTH OF THESE RUNS

NUMBER OF RUNS OF Y BELOW $\xi_{0.25}$
AVERAGE LENGTH OF THESE RUNS

COMPARE WITH DATA VALUES

Fig. 16. Criteria for selecting simulation results from solar radiation.

$\bar{Y}_D$ = DAILY MEAN; $Y_{(N)}$ = DAILY MAX
$Y_{(1)}$ = DAILY MIN; $R = Y_{(N)} - Y_{(1)}$

FOR OUR TMY CALCULATE

$\bar{Y}_{(N)}, S^2_{Y_{(N)}}$, HISTOGRAM OF $Y_{(N)}$

AUTOCOVARIANCE AND PERIODOGRAM OF DRY-BULB TEMPERATURE

NUMBER OF RUNS OF $Y_{(N)}$ ABOVE $\xi_{0.75}$
AVERAGE LENGTH OF THESE RUNS

DO SIMILAR CALCULATIONS FOR OTHER VARIABLES
COMPARE WITH DATA VALUES

Fig. 17. Criteria for selecting simulation results from dry-bulb temperature.

STATION 3937 JUL
WIND DIRECTION

	N	NE	E	SE	S	SW	W	NW
1953	26.7*	9.1	10.2	15.7	14.2	18.0	3.0	3.0
1954	12.1	9.1	7.5	11.7	12.8	25.8	11.2	9.8
1955	19.6	7.4	12.8	22.8	17.1	11.3	4.3	4.7
1956	16.5	.8	2.4	9.3	13.4	29.0	19.8	8.7
1957	34.3	8.2	5.2	5.9	5.2	14.4	16.1	10.6
1958	10.5	2.3	7.8	22.6	26.6	21.4	6.5	2.4
1959	15.3	14.1	19.6	14.1	8.1	14.7	7.5	6.6
1960	47.3	6.5	3.8	1.3	5.2	16.4	12.2	7.3
1961	43.4	2.2	5.4	6.0	22.7	13.3	5.6	1.3
1962	20.8	1.9	4.7	3.5	3.9	34.9	20.0	10.2
1963	13.4	3.6	12.2	11.7	17.6	27.7	6.7	7.0
1964	16.4	5.6	11.8	15.7	18.3	16.0	10.6	5.5
[illegible]	23.0	5.9	8.6	11.7	13.8	20.2	10.3	6.4

* During July the wind blew at Lake Charles, LA, from the north 26.7% of time.

Fig. 18. Wind direction frequencies for July in Lake Charles, Louisiana.

Problem Discussion 2, Part 1: Generation of a Typical Meteorological Year

Irving Hall and Richard Prairie, Sandia Laboratories

Irving Hall: I don't have very much to say that I didn't say yesterday. There is one thing that I want to emphasize: I mentioned that our data base is from 26 sites. Now we visualize making a standard meteorological year for each of these sites; we aren't going to make one for an entire year for the whole country or something like that. There will be one data-base year, typical year, or whatever you want to call it, for each of the 26 sites. If you have your handout, I had a viewgraph of that,* but I didn't have one of a specific problem that we are interested in. The first one is wind directions. Somebody did mention a possible way of handling this problem yesterday. We thought that we understood it at the time, but then we got to looking it over afterwards; either we misunderstood, or I'm not sure it is going to work! But if somebody has some apt comments on any of these things—this multivariate thing—I don't know exactly how they are going to handle it, but maybe there is a straightforward way. Just a little aside here: I mentioned yesterday that I talked to a meteorologist who has done some work for the State of California. One thing he told me was that a statistician couldn't do this job.

Dave Gosslee, Union Carbide Corporation, Nuclear Division: In handling meteorological data or other data that are in a similar form, we want to look at two-dimensional space. There has been quite a bit done with the circular normal distribution, and I think you came up with uniform circular distributions too. I haven't done this since I left some work I did with some climatologists quite a few years ago, so I don't have anything up to date on that. But there certainly are things that can be done, and there is considerable literature about circular normals; whether these would be normally distributed circular or not, I would have no idea.

Francis J. Anscombe, Yale: In the presentation yesterday, I think there was not really any discussion of the purpose of simulating a standard meteorological year. I suspect that if a standard year has been chosen, it will be used for a variety of purposes, and I suppose that these purposes would indicate rather different criteria for what would constitute a suitable standard. I could imagine that one use of the standard meteorological year would be for testing various sorts of theories. For that sort of purpose, I think one of the requirements should be that some few important variables should average out right for that site. I could imagine that it would be desirable to average the total amount of solar energy received, which should come up about correct, and the average amount of rainfall in that year should come out to be about the average rainfall for the region. However, if one were to define a standard meteorological year in which every possible variable was averaged, let's say for the date October 28, you have the average rainfall, average amount of sunlight, average speed and direction of the wind, and so on like that, over many years. In that way, you will altogether compile a year which is fantastically untypical! As I said there will be no storms and none of the variability which is ordinarily perceived. If the purpose of a standard year is meant to be typical weather, it certainly can't be the mean weather. More than a hundred years ago, there were very intense discussions about what is the average man—the mean man, and it was pointed out that if you average all the physical dimensions of a man you may get a description which doesn't correspond to any real man at all. There is obviously a conflict between means, or averages, on the one hand and modes on the other hand. I would suppose that

*I. J. Hall and R. R. Prairie, these Proceedings, Fig. 2.

the objective should be to have a year that in some ways is more or less the mean for certain important variables, but otherwise it is much more the mode than the mean. The mode for the joint distribution for all these variables may easily be quite a long way from the mean.

Corwin L. Atwood, EG&G Idaho: The fact that this year will be used for different purposes will be another reason for trying to persuade the funding agency to accept a collection of several typical years and a couple of extreme years cold ones or wet ones or something.

Ram Uppuluri, Union Carbide Corporation, Nuclear Division: Many questions went through my mind when I heard the talk yesterday, and I was thinking about this important problem of standard meteorological data and about defining the concepts like typical meteorological year. The first question that went through my mind was "Is it as hard to define a typical meteorological season as a meteorological year?" Maybe that is a helpful question for the people who are involved in policy decisions and might lead to the concept of a typical meteorological year. When we start thinking like this, I would like to break it down further: "At a particular site, is there a typical meteorological day?" Maybe we should try to build it up from information which we have. If you persist and try to use the p.d.f. runs, you may run into the problem of independence and dependence. People have to handle these problems using finite Markov chains to look for the transitions or for *peaks* of a different nature when you are looking at problems about persistence.

More than anything else, the concept of a typical manner of year needs to be defined. It is not clear to me how one defines a typical meteorological timespan (whatever the unit of time may be). Does this mean that we are thinking of particular plus or minus sigma limits? or does it mean, as Professor Anscombe pointed out, the modal frequency of a particular variable? Perhaps a greater effort should be made to define the terms; depending on the definition, we have the kind of tools we're looking for. If you want to use tools like quality control, maybe you can define the mean—or perhaps make a lot of pictures around these 26 sites. I imagine we can define ±5–10% of the Δ involved. I suppose the definition has everything to do with the kind of tools we are willing to get into.

A typical meteorological season is more appropriate because we will be eliminating, at least, the problem of seasonal variation, as well as other types of complications. More often than not, I hear people trying to look at the whole problem at one time and trying to get a spectral analysis and getting lost a little more.

Dick Prairie: We are really having great difficulties with the criteria for defining a typical year. We thought about means and modes and runs and ups and downs, and we have charts all over the office trying to look at all these things. We talked to the energy systems people, and we talked to meteorologists hoping we could get some sort of help in defining what the criteria are. Regarding the typical year, we're really thinking now in terms of a typical month possibly going to a typical season, which I would call the weather season (e.g., summer: June, July, and August). One of the problems we are having with this whole thing obviously is what's written. In one sense we've got too much data, and in another we don't have enough. We've got twelve years of hourly data. In reality, we look at a simple thing like temperature, and it varies all over the place from year to year for a given month and a given location. On the other hand, there are not really enough years to look at seasonal variation. In terms of what we're going to use it for, as Irving pointed out yesterday, people are sizing various energy systems all of the time, and what they would like to do is to feed this "typical year" in the little black box and compare various systems that different groups have proposed. So that is really the primary purpose of it. Also this business about upper and lower bounds of some sort is useful, but I don't really know what to do there. Suppose we put in a lower bound of a bad year, but "bad" has to have some specific context, that is, it must be bad with regard to temperature or radiation, for example. Thus, we have to come up with some sort of combination of bounds on badness.

The measures we are using include one measure on solar radiation, two on temperature, and two on wind - velocity and direction. I would say, however, that the biggest problem we have to grapple with right now is actually to set up these criteria (which will in fact satisfy some people) for what a really typical year is. Obviously it is a problem we will never win because, no matter what we come up with, somebody is going to say that it is not typical. One other thing—on the models that we are trying, the fits that we're getting are looking pretty good. The standard deviation that we are coming up with about the model is like three-quarters of a degree right now, and we've had some models that we then simulate from and get what looks like reasonable results to us. Of course we've had some where we've found out that Louisiana gets down to −250°C which we feel is probably slightly atypical at this time.

Dave Gosslee, Union Carbide Corporation, Nuclear Division: I just want to comment on some of the problems that I was involved in some time ago. With respect to the work of your people in Asheville now be sure that you don't read that as Nashville; Nashville, Tennessee, has a lot of records too he's talking about Asheville, North Carolina, where the weather bureau has stored a lot of data. H. C. S. Thom (Herb Thom) is a statistician-climatologist, and he laid down a lot of methodology for some regional, chronological studies. His typical weeks went from March 1 to March 1 to avoid the leap year and the short week; thus, the first week of the year was March 1 through March 6. There is certainly nothing wrong with thinking in terms of other than a chronological year. We were talking about the rain year yesterday, which I think they said went from October 1 to October 1. We would try to establish distributions for these observed variables, whether they be heating-degree days, cooling days, maximum temperatures, and so forth, and this will give you the transformation. We need to try to put out totals of means, medians, and various percentiles. Now I don't know whether this kind of thing really is what we're looking for or not, but there certainly is a lot of work that can be leaned on there and used as a starting point. Dr. Uppuluri brought up a point of persistence. Does anyone have a comment on how to handle that?

Lincoln Moses, Stanford: This idea is not really fully baked; but suppose that one chose an interval like a month or like ten days (one must make a choice, and I can't say anything about what would be smart) and then chose from the 12-year battery a random choice for January and a random choice for February and so forth. Now this will not be a typical year, but it can't help but be more or less representative of the last 12 years' weather. It will contain most of the persistence; you get breaks at the month intervals; and if it were easy to do (as it might be), all kinds of questions of multivariate things will have been solved simply by the way the weather gets made in that area. You can do lots of these "representative" years and get an idea of what the variation was with very little theorizing. However, you would be bound by your 12-year period, and in the case of a drought interval or something, it would not show up, but it would be there.

David Rubinstein, Nuclear Regulatory Commission: Essentially one can establish distance functions between years or months, whatever is regarded as suitable months appeal to me more than years. The distance function would combine, let's say, the difference of the mean temperature or difference of the number of runs, whatever value we are concerned with. Once the distance function is established, we can compare each year in terms of the distance function with every other year and pick the year that was the smallest sum of the distances or something of that sort.

Ronald Thisted, University of Chicago: One of the nice things about these discussion sessions is that we can make suggestions and then leave. I'm very sympathic to the remarks of Professor Anscombe, Dr. Atwood, and Professor Moses. Already over-month boundaries persistences are ignored. I think that is right. Look at January; if there is a cold spell that runs from January 28 to February 3, that's not picked up. So, there's no real loss in picking random months as far as persistence, because that's already considered minor. It seems to me that one of the problems you're facing is that you have a program which simulates years fairly well. It's consistent with the data you have, and you want to know which of these to pick out as a typical year. Germane remarks have been made to the fact that typical years are not things that we should fix on. If we buy a house for the energy demands of a typical year, we may face the same problems that would occur if we built houses for typical families which have 2.3 children. It is very hard to find housing that fits a seven-member family. We may be in the same position of having houses that might withstand extreme weather or be overbuilt for nonextreme weather, which aren't really atypical—just not near the middle. First of all, I think the statistician has an obligation to inform the contractor that what he is asking for may not really be what he wants. Secondly, you have very good programs, apparently, that simulate years; why don't you provide the contractor with a program which would generate typical years, that is, generate a year's worth of data which is consistent with the last 12 years—that's very much the spirit, I think, of Professor Moses's remarks—where your data actually follows from the 12 years but is not tied to any one. People could use this information. Because everybody wouldn't be looking at the same year, every so often, somebody will get something that's just atrocious, like last winter perhaps. I think that's good.

William Conover, Texas Tech: My suggestion is just a political one, because everyone has agreed that there is no such thing as a typical meteorological year. I think it might be a good idea to avoid those words. No matter what you come up with, people are going to attack you and say, "That's not a typical meteorological year." There is no such thing; we've never had a typical day in weather anywhere that I've been. You might want to decide upon another name and that may solve a lot of your problems. You may want to call it a random meteorological year, a modal meteorological year, a mean meteorological year, or a computer-generated meteorological year, and each of these would have different purposes. If you are going to decide whether to fund a solar evergy project, nobody will make that decision on a computer-generated meteorological year or a random meteorological year. Funding agencies want you to use something more stable, because someone will say that it wasn't fair to use that year to compare their site with somebody else's site. That year had a bad storm in it, and it's not something you can expect each year. Your purpose will determine the name you ought to use; then decide what parameters you want to work with, and define these parameters. Do you want first-order serial correlation to be one of your parameters. Take an example from the Rand Corporation random normal deviates. They don't make any claim that there's no serial correlation, and to people who come up and say, "Oh, there's a serial correlation in these normal deviates," they say, "That's all right; we didn't say there wasn't. We just said that if you plot them out, they plot like a normal distribution." So make your foundation fairly firm, and then you'll withstand any criticisms that may come around.

Wes Nicholson, Battelle: I'd like to reemphasize what Conover just said and possibly relate it a little bit to what John Jaech said the other night in his elegant comments. This reminds me a little bit of the problems that you're faced with when you consult with metallurgists, and they can only take a few samples, and they want you to find them a representative sample which has all the different kinds of properties in it that they may face when they look out in the real world. They would like to test a few representative samples and find out the strength of the materials. It's the same problem that is faced by the biologist when it is very expensive to prepare sections; therefore, the biologist wants to find that sacred section which will allow him to scan across it with his microscope and find all the different kinds of anomalies that there could be in the material. This is not a new problem. It's just phrased in a new area, and I'm wondering if maybe this isn't one of those points where we have to bite the bullet, stand up, and say "Well, folks, if what you really want to do is simulate what's going on, you can't do it with something on the average, and you've got to look at the problem as being a distributional problem first of all." We know how to give you good estimates of distributions, and we can tell you how to play games with these distributions, depending upon the questions that you want to answer, but you can't really answer very many real-world questions by looking for things that are typical.

Use of Regression Analysis to Evaluate Environmental Effects: Exploring Methods of Analysis*

J. J. Beauchamp

Computer Sciences Division
Union Carbide Corporation, Nuclear Division
Oak Ridge, Tennessee

C. W. Gehrs

Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

ABSTRACT

The statistical analysis of environmental data presents many interesting problems when the data are taken under field conditions where it is difficult to control factors that may have an effect upon a particular response variable. This paper presents different approaches to the analysis of data from a field experiment that examined the influence of adult density on production of a calanoid copepod zooplankter. Problems are presented involving the analysis of these data using regression and principal components regression.

INTRODUCTION AND SUMMARY

There are many difficulties associated with the use of regression techniques to analyze data from undesigned experimental contexts. For example, the parameter estimates may be unsatisfactory, or meaningful inferences may be difficult to make. The problem to be considered in this paper resulted from a collaborative effort of the authors to analyze data from a field experiment that examined the influence of adult density on reproduction of a calanoid copepod population. The data consisted of observations on clutch size, female size, adult density, and water temperature from samples collected in a pond, representing a closed ecosystem during a full reproductive cycle. The next section of this paper contains a detailed description of the experimental background and a statement of the objectives or goals of the original experiment.

Preliminary data analysis made use of regression techniques to study the population dynamics of the zooplankter. Starting with a complete second-order polynomial regression model in terms of factors related to female size, adult density, and water temperature, the variation in the observed clutch size (i.e., number of eggs per brood) was reasonably explained by a linear model containing ten terms. In addition, heterogeneity of variance and lack-of-fit tests were performed to conclude that it was reasonable to assume the observed number of eggs per brood was a Poisson random variable. The regression analysis was then done on the second-order model

*Research sponsored by the U.S. Department of Energy under contract W-7405-eng-26 with the Union Carbide Corporation. ESD Publication No. 1108, Environmental Sciences Division, Oak Ridge National Laboratory.

using weights appropriate for Poisson-distributed data. This type of analysis made it possible to test for the adequacy of the polynomial approximation and the assumed Poisson distribution. The details of this analytical approach are presented later.

As an alternative to the regression analysis described above, principal components regression was used for analyzing the relation between clutch size and the factors related to female size, adult density, and water temperature. This procedure involved obtaining the principal components of the set of standardized explanatory variables and then calculating their regression upon the clutch size variable. Correlations of these principal components with the original explanatory variables assisted in yielding physical interpretations. A total of nine explanatory variables were used to obtain a new set of nine principal components. Various selection procedures were used to reduce the number of components in the final regression model and also to reduce the number of original explanatory variables without sacrificing the ability of the regression model to explain the observed variation in clutch size. The details of this procedure are presented later.

The results of the regression analyses are presented and compared, and in the last section, conclusions from the analyses are summarized. In addition, problems resulting from the analysis are stated, and alternative methods of analysis are solicited.

BACKGROUND AND OBJECTIVES

Zooplankton are an important component of lake and pond ecosystems, serving as an intermediary in energy flow between algae (e.g., phytoplankton) and fish. One group of zooplankton, calanoid copepods, is extremely important in large bodies of water, with the genus, *Diaptomus*, often being the most abundant of the microcrustacea.[1] These organisms (diaptomids) are primarily herbivores (i.e., eat plants—in this case algae) and detrivors (i.e., eat decomposing organisms), and in turn, they are consumed by larval and juvenile fish. Changes in diaptomid populations, consequently, could quite possibly affect the type and quantity of fish found in a lake.

Their vital role notwithstanding, knowledge concerning the general biology, environmental requirements, and interrelations of this group within the ecosystem is woefully inadequate. It has not been possible to state what role past changes in water quality have had on these organisms, and hence, no prediction can be made concerning what might occur as a result of future changes in water quality.

A substantial research effort was conducted from 1969 through 1972 to gain insight into the ecology of one diaptomid, *Diaptomus clavipes*. This study made use of laboratory and field investigations and was designed to determine what role various factors such as water temperature, food, pH, alkalinity, etc., played in regulating populations of these organisms. Additionally, such organismic and population variables as age class distribution, reproductive rate, female size, and adult density were analyzed for their roles in regulating the populations.[2]

The field study was conducted over an entire reproductive year (February through October) with random samples collected on each sampling date. During the colder months of the year the population is composed only of adults with no immature stages present. Consequently, it was relatively easy to determine how many young were being produced and how they were developing. During the course of the year five distinct generations were produced.[3] The major concern in this study was with those animals that reached adulthood during this one reproductive year. From previous knowledge,[3] it was realistic to assume that the first of these animals reached adult form in early April. Consequently, while data from February and March were available they have been excluded from this study.

Preliminary evaluation of data showed that clutch size was the most important determinant regulating population size[2] and that water temperature, adult density, and female size were the three primary variables affecting clutch size. The purpose of this study was to determine to what level these three variables explain changes in clutch size and what is the relative importance of each of these variables as a determinant. With the goal and purpose of the experiment formulated, the objective of this paper is to describe the ways used in attempting to elucidate the role of these variables or factors in regulating clutch size in *Diaptomus clavipes*.

In the statistical context, this problem involves exploratory analysis dealing with interrelations between variables. More specifically, the purpose is

---

1. Andres Robertson, Carl W. Gehrs, Bryan D. Hardin, and Gary W. Hunt, *Culturing and Ecology of* Diaptomus clavipes *and* Cyclops vernalis, Report EPA-660/3-74-006, U.S. Environmental Protection Agency, 1974.

2. C. W. Gehrs, *Aspects of the Population Dynamics of the Calanoid Copepod,* Diaptomus clavipes *Schacht*. Ph.D. thesis, University of Oklahoma, Norman, 1972.

3. Carl W. Gehrs and Andrew Robertson, "Use of Life Tables in Analyzing the Dynamics of Copepod Populations," *Ecology* 56(3): 665-72 (1972).

to explore ideas for analyzing the relation between a response variable (clutch size) and a set of explanatory variables (female size, adult density, water temperature).

ANALYTICAL APPROACHES

Description of Data

At each sampling date the following variables were recorded:

- Water temperature. In addition to the actual water temperature at the sampling time, estimates of two- and four-week delayed water temperatures were recorded. These delayed water temperatures might affect development (ultimate animal size) during the early formative stages. An average water temperature for the four-week period prior to the time of sampling was calculated by fitting a quadratic function to the sampling, two-week delayed and four-week delayed water temperatures and then calculating the average value of this function over the four-week period.

- Adult density. The adult density was the average number of adults per liter recorded from the samples at each sampling date, and the log of the density was used in all subsequent analyses.

- Female size. The average length of the sampled females was used as the measure of female size.

In addition, for each of the sampled females with clutches, the number of eggs in each clutch was recorded. Table 1 contains a complete listing of the explanatory factors and response variable. Because the main objective of this study is concerned with the dependence of the average number of eggs per clutch (response variable) on the explanatory factors, a regression equation relating the response variable to the explanatory factors appeared to be a logical starting point in analyzing these data.

However, before the regression analysis was performed, testing determined if it was reasonable to assume that the observed number of eggs per clutch at each sampling date followed a Poisson distribution. The first test was a heterogeneity of variance test and used the following statistic to detect extra-Poisson variation:

$$(n-1)s^2/\bar{y}, \qquad (1)$$

where

n = number of clutches observed,

s^2 = sample variance of observed clutch sizes,

$\bar{y}$ = sample average of number of eggs per clutch.

Under the null hypothesis assuming Poisson variability, Eq. (1) should have a chi-square distribution with $(n-1)$ d.f. The results of this test are summarized in Table 2. Although the results of this test were not significant ($P > 0.05$) in 21 of the 25 sampling dates, the overall pooled test indicated the presence of extra-Poisson variation that should be considered in subsequent analysis. In fact, when some "suspect" observations in Table 1 are omitted, the overall chi-square value reduces to 156.06 (d.f. = 149) which is not significant ($P > 0.05$). The second test was a goodness-of-fit test assuming an underlying Poisson distribution and was applied only to those samples having at least ten observations. The results of this test, and the pooled chi-square value, are summarized in the last column of Table 2. None of these tests was significant ($P > 0.05$). Therefore, the Poisson assumption will be used in subsequent analysis.

Regression Equation

Since no model previously had been derived to simulate the change in clutch size as a function of the explanatory variables, a second-order polynomial function of the explanatory variables was the first approximation. That is, the expected value or average value of the response variable was approximated by

$$\begin{aligned} E(y_i) = \beta_0 &+ \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{11} x_{1i}^2 \\ &+ \beta_{22} x_{2i}^2 + \beta_{33} x_{3i}^2 + \beta_{12} x_{1i} x_{2i} \\ &+ \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i}, \end{aligned} \qquad (2)$$

where

y_i = average number of eggs per clutch observed on the ith sampling date,

x_{1i} = average water temperature over the four-week period prior to the ith sampling date,

x_{2i} = logarithm of the adult density on the ith sampling date,

x_{3i} = female size observed on the ith sampling date.

Table 1. Experimental data

Sampling date	Water temperature, x_1 (°C)	Adult density,[a] x_2	Female size,[b] x_3	Number of clutches observed, n	Number of eggs per clutch	Average number of eggs per clutch, y
1	7.26	1.54	102.53	6	38,36,32,31,37,34	34.67
2	7.60	0.81	101.20	2	30,38	34.00
3	7.63	1.44	101.42	7	52,34,40,30,23,34,38	35.86
4	7.10	2.03	100.75	3	35,33,43	37.00
5	9.69	2.69	106.00	4	43,15,38,26	30.50
6	11.30	2.41	102.67	3	20,18,22	20.00
7	11.38	3.72	102.10	9	20,20,18,15,19,18, 15,19,23	18.56
8	11.53	3.53	101.00	3	11,18,19	16.00
9	11.87	1.76	102.00	4	15,28,26,14	20.75
10	10.22	4.87	100.40	7	23,25,24,25,19,7,21	20.57
11	13.55	4.19	99.56	3	25,27,26	26.00
12	17.46	8.11	97.80	4	15,18,20,22	18.75
13	17.78	2.32	100.00	2	11,17	14.00
14	19.98	1.19	95.73	16	15,13,13,15,17,15,11, 17,16,16,19,15,16,17, 14,17	15.38
15	21.95	2.72	99.86	9	28,28,23,34,25,26,26, 29,17	26.22
16	24.44	6.12	99.79	5	19,18,22,12,20	18.20
17	26.10	6.65	96.17	12	12,11,20,11,26,12,14, 16,13,13,18,14	15.00
18	25.18	17.09	94.50	9	13,17,10,25,30,18,10, 9,12	16.00
19	26.84	11.37	93.45	8	12,13,14,11,17,13,17, 11	13.50
20	28.32	11.00	92.53	27	9,12,9,14,12,7,31,11, 8,8,11,8,8,17,8,8,5, 12,8,10,8,8,10,12,12, 6,6	10.30
21	26.09	7.57	89.43	16	16,12,20,12,4,10,11,14, 15,10,15,16,15,6,15,7	12.38
22	23.95	7.20	88.63	6	12,13,14,13,9,17	13.00
23	21.47	2.09	88.33	6	8,14,10,14,14,12	12.00
24	20.59	1.38	94.00	3	21,17,10	16.00
25	17.77	0.39	99.57	2	30,30	30.00

[a] x_2 = ln(adult density) in text; however, values in the table are the average number of adults per liter.
[b] Units are such that if x_3 = 100, the female size is 2.4 mm.

Table 2. Analysis of Poisson Variation[a]

Sampling date	n	s^2	y	$(n-1)s^2/y$	Goodness-of-fit χ^2 (d.f.)
1	6	7.87	34.67	1.13	
2	2	32.00	34.00	0.94	
3	7	81.48	35.86	13.63[b]	
4	3	28.00	37.00	1.51	
5	4	157.67	30.50	15.51[c]	
6	3	4.00	20.00	0.40	
7	9	6.28	18.56	2.71	
8	3	19.00	16.00	2.38	
9	4	52.92	20.75	7.65	
10	7	40.62	20.57	11.85	
11	3	1.00	26.00	0.08	
12	4	8.92	18.75	1.43	
13	2	18.00	14.00	1.29	
14	16	3.85	15.38	3.75	13.88 (6)
15	9	21.44	26.22	6.54	
16	5	14.20	18.20	3.12	
17	12	19.64	15.00	14.40	1.46 (4)
18	9	53.50	16.00	26.75[c]	
19	8	5.71	13.50	2.96	
20	27	24.22	10.30	61.13[c]	5.57 (4)
21	16	17.85	12.38	21.63	8.83 (4)
22	6	6.80	13.00	2.62	
23	6	6.40	12.00	2.67	
24	3	31.00	16.00	3.88	
25	2	0.00	30.00	0.00	
Pooled χ^2				209.96[b]	29.74
d.f.				151	18

[a] n = number of clutches observed; s^2 = sample variance of clutch sizes; and y = sample average of number of eggs per clutch.
[b] $0.01 < P < 0.05$.
[c] $P < 0.01$.

The first step in applying Eq. (2) to describe the variation in the observed clutch size involved obtaining the iterative weighted least-squares estimates of the ten β's in Eq. (2) using the following weights:

$$W_i = n_i/[E(y_i)] , \tag{3}$$

for the ith observation where n_i is the number of clutches observed on the ith sampling date and $i = 1, 2, \ldots, N$. The details of the iterative procedure are presented in the paper by Frome, Kutner, and Beauchamp.[4] Various variable selection procedures were also applied to regression Eq. (2) in an attempt to reduce the dimensionality of the estimation procedure. After obtaining the weighted least-squares estimates of the β's, the following statistic was partitioned and used to test for the presence of extra-Poisson variation and the adequacy of the regression model:

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{m} \hat{y}_i^{-1}(y_{ij} - \hat{y}_i)^2 , \tag{4}$$

where

y_{ij} = the observed number of eggs in the jth clutch on the ith sampling date,

$\hat{y}_i$ = predicted average number of eggs per clutch on the ith sampling date.

The details of this test are also shown in the article by Frome, Kutner, and Beauchamp.[4]

4. E. L. Frome, M. H. Kutner, and J. J. Beauchamp, "Regression Analysis of Poisson-Distributed Data," *J. Am. Stat. Assoc.* 68: 935-40 (1973).

Principal Components Regression

An alternative exploratory approach was applied by obtaining the principal components of the set of standardized explanatory variables in Eq. (2) and then calculating their regression upon the clutch size. The objective for using the principal components is to find a linear transformation of the explanatory variables into a new set, which has desirable properties. Some of the rationale for using the principal components are (1) the principal components are uncorrelated with each other, and (2) each principal component, progressing from the one associated with the largest eigenvalue of the correlation matrix of the explanatory variables to the smallest, accounts for as much of the combined variance of the explanatory variables as possible, consistent with being orthogonal to other principal components. Massy[5] gives a discussion and development of the necessary statistical methods. In this analysis the principal components were calculated from the explanatory variables standardized to have mean zero and unit variance.

The regression of the clutch size, y, on the scaled principal components of the explanatory variables in Eq. (2) is denoted by

$$E(y_i) = \gamma_0 + \gamma_1 P_{1i} + \gamma_2 P_{2i} + \gamma_3 P_{3i} + \gamma_4 P_{4i} + \gamma_5 P_{5i} + \gamma_6 P_{6i} + \gamma_7 P_{7i} + \gamma_8 P_{8i} + \gamma_9 P_{9i}\ , \qquad (5)$$

where

y_i = observed average number of eggs per clutch on the ith sampling date,

P_{pi} = pth principal component, calculated from the correlation matrix of the explanatory variables, on the ith sampling date, for $P = 1, 2, \ldots, 9$ and $i = 1, 2, \ldots, N$.

To examine the possibility of reducing the dimensionality of the problem, the following approaches were considered: (1) the principal components having the smallest eigenvalues were dropped, since these would be relatively unimportant as predictors of the explanatory variables; (2) the components having the smallest correlation with the observed clutch size were dropped; and (3) the principal components resulting from the reduced set of explanatory variables in Eq. (2) were used to explain the variation in the observed clutch size. Weighted least-squares estimates of the γ's in Eq. (5) or its reduced form were obtained using the weights of Eq. (3) and the same iterative procedure. The heterogeneity-of-variance and lack-of-fit tests were also examined.

RESULTS

Regression Analysis

In Fig. 1 plots of the explanatory variables, x_1, x_2, and x_3, against the observed mean clutch size are shown. Table 3 contains the calculated correlation coefficients among the explanatory variables and response variable (clutch size) included in regression Eq. (2). From this figure and table it is clear that many of the explanatory variables are highly correlated with the observed clutch size. However, it is difficult to interpret the effect of the explanatory variables on the response variable because there is also a high correlation present among many of the explanatory variables of Eq. (2).

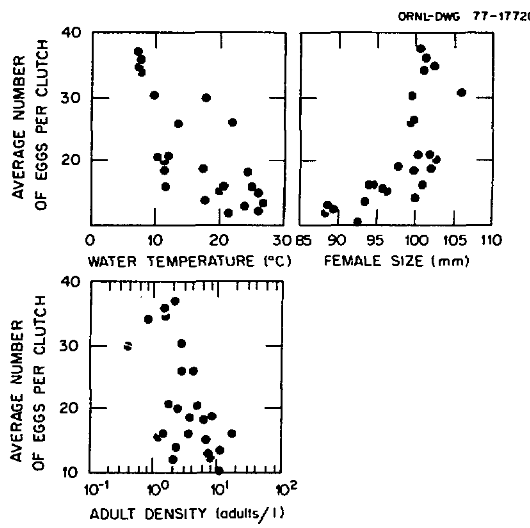


Fig. 1. Plots of clutch size against explanatory variables.

Table 4 summarizes the results of the regression analyses using Eq. (2) and reduced forms of this regression equation. The reduced forms of Eq. (2) were determined by starting with a single-variable model, which contained the single explanatory variable having the largest R^2-value. This single-variable model was then expanded to a two-variable model. At the next stage all two-variable models are considered, and the two-variable model is chosen that gives the maximum R^2-value for all two-variable

5. W. F. Massy, "Principal Components Regression in Exploratory Statistical Research," *J. Am. Stat. Assoc.* **60**: 234 56 (1965).

Table 3. Correlation coefficients among explanatory variables and response variable

Variable	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
x_1	1								
x_2	0.566	1							
x_3	−0.778	−0.415	1						
x_1^2	0.989	0.603	−0.778	1					
x_2^2	0.671	0.887	−0.491	0.710	1				
x_3^2	−0.783	−0.416	0.999	−0.783	−0.497	1			
x_1x_2	0.769	0.934	−0.587	0.816	0.932	−0.590	1		
x_1x_3	0.995	0.558	−0.711	0.979	0.664	−0.718	0.757	1	
x_2x_3	0.527	0.997	−0.356	0.562	0.868	−0.358	0.912	0.524	1
y	−0.758	−0.583	0.681	−0.717	−0.528	0.684	−0.633	−0.741	−0.561

Table 4. Summary of regression analysis

Stage	Variables omitted	Number of variables	R^2	Tests[a] Heterogeneity of variance (χ^2)	Lack-of-fit (F)	Overall (χ^2)
1	Complete	9	0.83	212.03 (151)	2.93 (15,151)	273.65 (166)
2	x_1x_2	8	0.82	211.96 (151)	2.74 (16,151)	273.54 (167)
3	x_2 and x_1x_2	7	0.81	208.23 (151)	4.03 (17,151)	302.64 (168)
4	x_2^2, x_3^2, x_1x_2	6	0.78	212.24 (151)	3.68 (18,151)	305.45 (169)
5	x_2, x_3, x_3^2, x_1x_2	5	0.78	197.47 (151)	5.68 (19,151)	338.56 (170)
6	$x_2, x_3, x_2^2, x_3^2, x_1x_2$	4	0.75	199.78 (151)	5.81 (20,151)	353.47 (171)
7	$x_3, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$	3	0.71	199.82 (151)	6.98 (21,151)	393.70 (172)
8	$x_2, x_3, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$	2	0.63	192.83 (151)	6.99 (22,151)	389.14 (173)
9	$x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$	1	0.58	212.30 (151)	5.23 (23.151)	381.56 (174)

[a]Value in parenthesis is degrees of freedom. See ref. 4 for details.

models. This model can be thought of as the "best" two-variable model in the sense of maximizing the R^2-value. This procedure is repeated in subsequent stages to give the "best" three-variable, four-variable, etc., models. Since some variables may be included at one stage and then omitted at some subsequent stage, there is difficulty in determining the significance, or lack of significance, of terms in the regression equation.

A review of the results in Table 4 shows that the complete nine-variable model (stage 1) does explain a reasonably large amount of the variation (83%) in the observed clutch size. However, the heterogeneity of variance and lack-of-fit tests from the weighted regression analysis were both significant ($P < 0.01$), indicating extra-Poisson variation as well as a significant amount of unexplained variation by the model. Another conclusion from Table 4 and Fig. 2 can be made by observing the increase in R^2 from the "best" single-variable model to the complete nine-variable model. The first variable, x_1, explains more than 69% of the total variation explained by the complete nine-variable model, and the subsequent variables added to the model all increased subsequent R^2 values by less than 10%. However, a comparison of Figs. 3*a* and 3*b*, which plot the observed and calculated clutch sizes from the fitted complete and "best" five-variable model, respectively, indicate a definite improvement

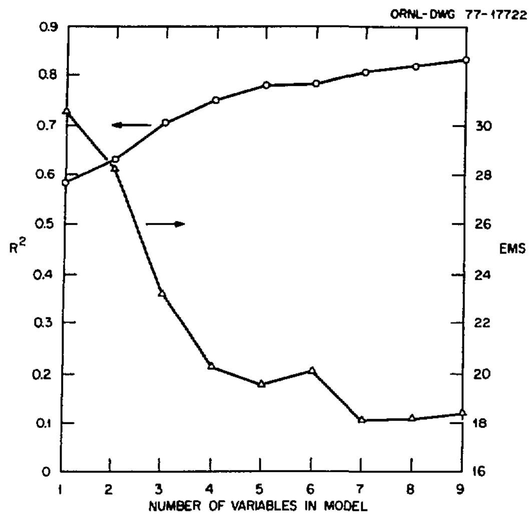


Fig. 2. R^2 and error mean square values (EMS) for different models.

in the explanation of the observed clutch sizes with the complete model. This improvement is especially seen for the larger clutch sizes.

Figure 4 shows plots of the residuals from the fitted complete second-order polynomial against each of the explanatory variables and the predicted clutch size. The most obvious conclusion from an examination of this plot is that there are "suspect" outlier observations falling outside the 2σ and 3σ limits. The investigator would now need to examine these observations closely to determine their influence upon any future conclusions.

An additional difficulty in this particular regression analysis approach arises when a partitioning of the regression sum of squares is done to determine the amount of variation attributable to the different explanatory variables in Eq. (2). Table 5 shows two different partitionings of the regression sum of squares for the "best" eight-variable model. From an examination of this table, the significance or lack of significance of the different explanatory variables would be difficult to determine.

Principal Components Regression

The principal components regression began with the calculation of the principal components of the standardized explanatory variables in Eq. (2) and their correlation with the original explanatory variables. Table 6 displays the eigenvalues and orthonormal eigenvectors from the correlation matrix of the explanatory variables of Eq. (2). These eigenvectors are used to obtain the P's of Eq. (5). From the first part of Table 6 it is clear that the first three principal components are accounting for a majority (>98%) of the variation in the explanatory variables. The physical interpretation of the variation of the new principal components is difficult merely from the eigenvectors that are also shown in Table 6. Therefore, the correlations of each explanatory variable with each principal component, as well as the correla-

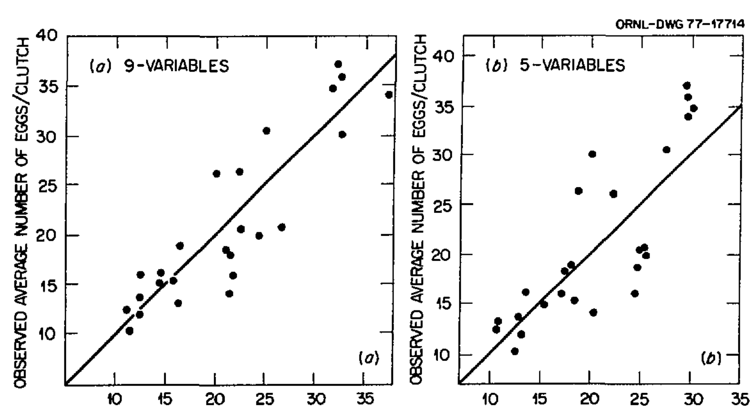


Fig. 3. Comparison of observed and calculated clutch size from nine- and five-variables polynomials. (*a*) Nine variables, (*b*) five variables.

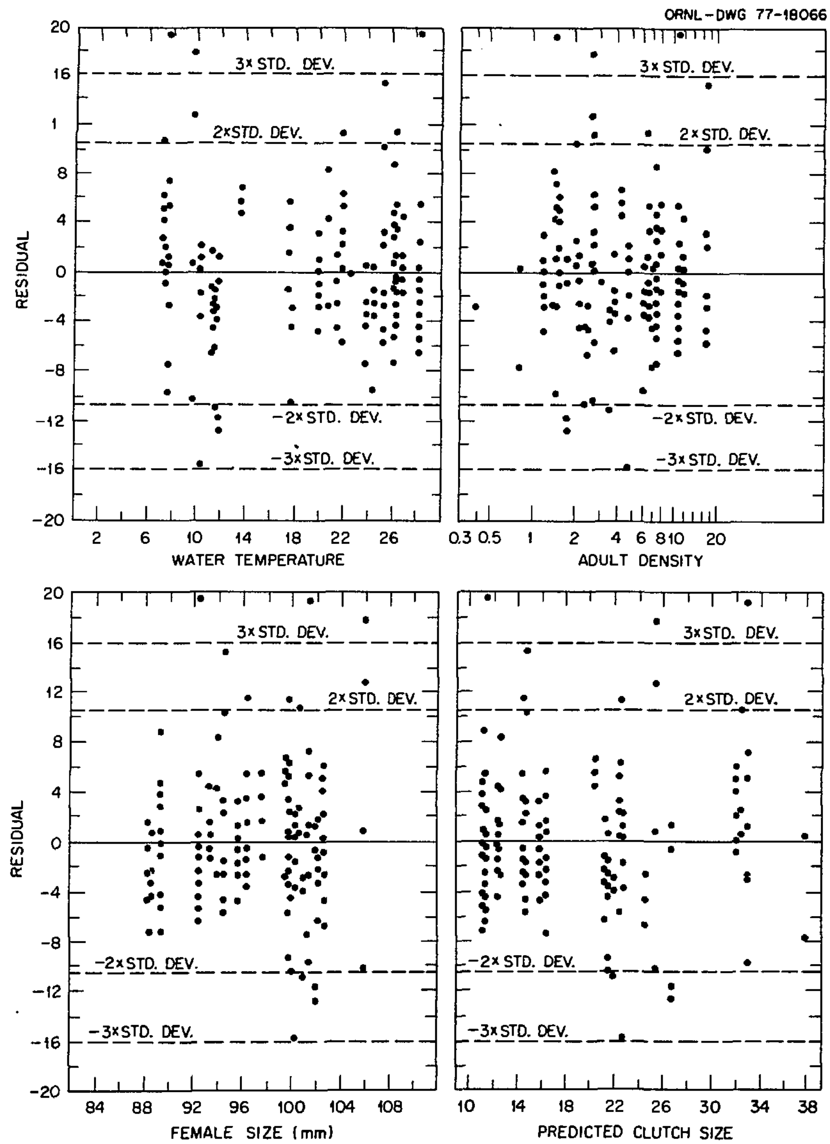


Fig. 4. Residual plots from complete second-order regression equation.

tions of the response variable with each principal component, were calculated. These correlations are summarized in Table 7. As expected, the highest correlations among the explanatory variables and principal components exist in the first three or four principal components. Therefore, the first thought would be to omit the last five principal components in the regression to decrease the dimensionality of the problem. However, the correlations of the response variable do not follow this ordering. In fact, the correlation between the response variable and P_9 was the fourth largest in absolute value. This would imply that P_9 would be relatively important as a predictor of the response variable. These considerations lead to

Table 5. Regression analysis for eight-variable model

Source	d.f.	Sum of squares[a] (1)	Sum of squares[a] (2)
x_1	1	$R(x_1 \mid \text{mean}) = 953.34$	$R(x_1 \mid \text{all other variables}) = 115.79$
x_2	1	$R(x_2 \mid x_1, \text{mean}) = 58.077$	$R(x_2 \mid \text{all other variables}) = 17.782$
x_3	1	$R(x_3 \mid x_1, x_2, \text{mean}) = 39.911$	$R(x_3 \mid \text{all other variables}) = 67.787$
x_1^2	1	$R(x_1^2 \mid x_1, x_2, x_3, \text{mean}) = 180.41$	$R(x_1^2 \mid \text{all other variables}) = 181.492$
x_2^2	1	$R(x_2^2 \mid x_1, x_2, x_3, x_1^2, \text{mean}) = 37.197$	$R(x_2^2 \mid \text{all other variables}) = 20.055$
x_3^2	1	$R(x_3^2 \mid x_1, x_2, x_3, x_1^2, x_2^2, \text{mean}) = 0.00985$	$R(x_3^2 \mid \text{all other variables}) = 64.583$
x_1x_3	1	$R(x_1x_3 \mid x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, \text{mean}) = 75.064$	$R(x_1x_3 \mid \text{all other variables}) = 97.657$
x_2x_3	1	$R(x_2x_3 \mid x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_3, \text{mean}) = 22.869$	$R(x_2x_3 \mid \text{all other variables}) = 22.869$
Error	16	291.031 (Mean square = 18.189)	
Total	24	1657.91	

[a] $R(x \mid y)$ = sum of squares explained by x given y is included in model.

two criteria for deleting components from Eq. (5): (1) delete the components having the smallest eigenvalues, and (2) delete the components having the smallest correlation between the components and the response variable. Both of these criteria were used, and the results of the principal components regression are summarized in Table 8. In addition, the results of the heterogeneity-of-variance and lack-of-fit tests from Eq. (4) are also summarized in this table.

The physical interpretation of the principal components can sometimes be made by an examination of the correlations between the principal components and the explanatory variables as well as the response variable (Table 7) along with plots of one principal component against another (e.g., see Fig. 5). An examination of this figure reveals a clustering of the high- and low-density values into two disjoint groups separated by the dashed line. Thus P_1 and P_2 should give valuable information related to the influence of adult density on the response variable of clutch size.

Additional attempts have been made to reduce the dimensionality of the problem by calculating the principal components from the standardized variables of the "best" reduced variable models summarized in Table 4. The results of these calculations are given in Table 9 along with the heterogeneity-of-variance and lack-of-fit tests.

As expected, the results summarized in Tables 4 and 9 are similar, because the same number of variables were used in each stage with the principal components being only linear combinations of the explanatory variables. Therefore, the only advantages to the use of principal components regression has been the orthogonality of the P's in Eq. (5) and the increasing amount of combined variability of the explanatory variables explained by the P's as one progresses from P_1 to P_2, etc. Both of these advantages should be of some assistance in the physical interpretation of the analytical results.

An additional interesting result is seen from an examination of Fig. 6, in which the observed and calculated clutch sizes are compared for three different forms of Eq. (5). The most obvious conclusion is that using only the principal components with the largest eigenvalues may not do as well in predicting the response variable as another set of principal components that account for a smaller percentage of the combined variance of the explanatory variables.

A graphical aid to the physical interpretation of the principal components is shown in Fig. 7 where the components of the eigenvectors, called loadings, are plotted for each principal component. This figure shows the loadings using only eight of the original explanatory variables. From an examination of this figure, it is possible to easily determine (1) those explanatory variables of major importance for each principal component, that is, those explanatory variables with the largest absolute value loadings; (2) the relative ordering of the explanatory variables for their contribution to each principal component; and (3) the particular explanatory variables being

Table 6. Eigenvalues and eigenvectors of standardized explanatory variables

Nine-variable model

Eigenvalues

Value	6.70	1.61	5.18×10^{-1}	1.34×10^{-1}	3.64×10^{-2}	3.05×10^{-3}	5.19×10^{-4}	1.28×10^{-4}	2.62×10^{-5}
Proportion	0.744	0.179	0.058	0.015	0.004	0.000	0.000	0.000	0.000
Cumulative proportion	0.744	0.923	0.981	0.996	1.000	1.000	1.000	1.000	1.000

Eigenvectors

Vector → Variable ↓	1	2	3	4	5	6	7	8	9
x_1	−0.355	0.232	−0.356	0.101	0.217	0.159	−0.521	−0.121	−0.570
x_2	−0.317	−0.429	0.176	0.319	0.131	−0.103	−0.326	0.653	0.150
x_3	0.303	−0.397	−0.501	−0.007	−0.031	0.000	0.392	0.337	−0.481
x_1^2	−0.362	0.196	−0.317	0.010	−0.430	−0.731	0.081	0.032	0.057
x_2^2	−0.335	−0.302	0.017	−0.850	0.256	−0.085	−0.002	−0.017	0.003
x_3^2	0.305	−0.397	−0.488	0.007	−0.058	−0.045	−0.477	−0.316	0.423
x_1x_2	−0.364	−0.239	0.006	−0.040	−0.719	0.535	0.028	−0.070	−0.016
x_1x_3	−0.347	0.205	−0.481	0.121	0.333	0.324	0.380	0.102	0.470
x_2x_3	−0.303	−0.466	0.145	0.386	0.231	−0.171	0.300	−0.574	−0.135

Eight-variable model

Eigenvalues

Value	5.83	1.50	5.18×10^{-1}	1.34×10^{-1}	1.51×10^{-2}	5.26×10^{-4}	1.70×10^{-4}	2.79×10^{-5}
Proportion	0.729	0.188	0.065	0.017	0.002	0.000	0.000	0.000
Cumulative proportion	0.729	0.917	0.982	0.999	1.00	1.00	1.00	1.00

Eigenvectors

Vector → Variable ↓	1	2	3	4	5	6	7	8
x_1	−0.388	0.190	−0.355	−0.096	0.261	0.517	−0.169	−0.560
x_2	−0.324	−0.488	0.178	−0.309	−0.071	0.348	0.628	0.099
x_3	0.338	−0.369	−0.500	0.006	−0.025	−0.380	0.320	−0.502
x_1^2	−0.393	0.154	−0.317	−0.011	−0.840	−0.114	−0.037	0.042
x_2^2	−0.348	−0.359	0.019	0.861	0.081	−0.005	−0.029	0.002
x_3^2	0.340	−0.368	−0.488	−0.008	−0.069	0.463	−0.311	0.440
x_1x_3	−0.379	0.162	−0.480	−0.115	0.454	−0.356	0.184	0.469
x_2x_3	−0.309	−0.525	0.147	−0.374	0.064	−0.335	−0.585	−0.096

Seven-variable model

Eigenvalues

Value	5.27	1.11	4.89×10^{-1}	1.13×10^{-1}	1.49×10^{-2}	4.43×10^{-4}	3.11×10^{-5}
Proportion	0.753	0.159	0.070	0.016	0.002	0.000	0.000
Cumulative proportion	0.753	0.912	0.982	0.998	1.000	1.000	1.000

Table 6. (continued)

Eigenvectors

Vector → Variable ↓	1	2	3	4	5	6	7
x_1	−0.419	0.090	−0.365	0.076	0.263	−0.561	−0.542
x_3	0.372	−0.386	−0.462	−0.024	−0.022	0.462	−0.534
x_1^2	−0.422	0.050	−0.314	−0.006	−0.843	0.084	0.050
x_2^2	−0.344	−0.516	0.173	−0.761	0.082	−0.009	0.006
x_3^2	0.374	−0.384	−0.451	−0.008	−0.067	−0.533	0.470
x_1x_3	−0.408	0.049	−0.487	0.089	0.456	0.424	0.445
x_2x_3	−0.289	−0.652	0.291	0.638	−0.007	0.008	−0.003

Six-variable model

Eigenvalues

	1	2	3	4	5	6
Value	4.49	1.15	3.42×10^{-1}	1.60×10^{-2}	4.23×10^{-4}	1.58×10^{-4}
Proportion	0.748	0.192	0.057	0.003	0.000	0.000
Cumulative proportion	0.748	0.940	0.997	1.000	1.000	1.000

Eigenvectors

Vector → Variable ↓	1	2	3	4	5	6
x_1	−0.450	−0.246	0.236	−0.275	−0.375	−0.682
x_2	−0.367	0.584	−0.110	0.057	−0.638	0.321
x_3	0.371	0.339	0.853	0.096	−0.076	−0.068
x_1^2	−0.454	−0.205	0.200	0.832	0.132	0.036
x_1x_3	−0.443	−0.231	0.402	−0.455	0.223	0.576
x_2x_3	−0.350	0.623	−0.053	−0.109	0.616	−0.307

Five-variable model

Eigenvalues

	1	2	3	4	5
Value	4.02	8.45×10^{-1}	1.13×10^{-1}	1.90×10^{-2}	3.25×10^{-3}
Proportion	0.804	0.169	0.023	0.004	0.001
Cumulative proportion	0.804	0.973	0.996	1.000	1.00

Eigenvectors

Vector → Variable ↓	1	2	3	4	5
x_1	−0.473	0.341	0.080	0.103	−0.802
x_1^2	−0.478	0.284	−0.007	−0.774	0.303
x_2^2	−0.429	−0.479	−0.760	0.090	−0.015
x_1x_3	−0.471	0.345	0.112	0.618	0.515
x_2x_3	−0.377	−0.675	0.635	−0.022	−0.005

Table 7. Correlations between explanatory variables, response variable, and principal components

Variable	Principal component								
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
				Nine-variable model					
x_1	-0.919	0.295	-0.256	0.037	0.041	0.009	-0.012	-0.001	-0.003
x_2	-0.820	-0.545	0.126	0.117	0.025	-0.006	-0.007	0.007	0.001
x_3	0.785	-0.504	-0.360	-0.003	-0.006	0.000	0.009	0.004	-0.002
x_1^2	-0.937	0.248	-0.228	0.004	-0.082	-0.040	0.002	0.000	0.000
x_2^2	-0.868	-0.384	0.012	-0.312	0.049	-0.005	0.000	0.000	0.000
x_3^2	0.789	-0.504	-0.352	0.002	-0.011	-0.002	-0.011	-0.004	0.002
x_1x_2	-0.942	-0.304	0.004	-0.015	-0.137	0.030	0.001	-0.001	0.000
x_1x_3	-0.898	0.260	-0.346	0.044	0.064	0.018	0.009	0.001	0.002
x_2x_3	-0.785	-0.592	0.104	0.141	0.044	-0.009	0.007	-0.006	-0.001
y	0.759	-0.148	-0.025	-0.278	-0.264	-0.110	0.080	0.052	0.258
				Eight-variable model					
x_1	-0.937	0.232	-0.256	-0.035	0.032	0.012	-0.002	-0.003	
x_2	-0.783	-0.598	0.128	-0.113	-0.009	0.008	0.008	0.001	
x_3	0.816	-0.452	-0.360	0.002	-0.003	-0.009	0.004	-0.003	
x_1^2	-0.950	0.189	-0.228	-0.004	-0.103	-0.003	-0.005	0.000	
x_2^2	-0.841	-0.440	0.014	0.316	0.010	-0.000	-0.000	0.000	
x_3^2	0.820	-0.452	-0.352	-0.003	-0.009	0.011	-0.004	0.002	
x_1x_3	-0.914	0.198	-0.346	-0.042	0.056	-0.008	0.002	0.002	
x_2x_3	-0.746	-0.643	0.106	-0.137	0.008	-0.008	-0.008	-0.001	
y	0.773	-0.089	-0.026	0.272	-0.279	-0.081	0.050	0.241	
				Seven-variable model					
x_1	-0.961	0.095	-0.255	0.026	0.032	-0.012	-0.003		
x_3	0.854	-0.407	-0.323	-0.008	-0.003	0.010	-0.003		
x_1^2	-0.969	0.053	-0.220	-0.002	-0.103	0.002	0.000		
x_2^2	-0.791	-0.543	0.121	-0.255	0.010	-0.000	0.000		
x_3^2	0.859	-0.404	-0.315	-0.003	0.008	-0.011	0.003		
x_1x_3	-0.937	0.052	-0.341	0.030	0.056	0.009	0.002		
x_2x_3	-0.664	-0.686	0.204	0.214	-0.001	0.000	-0.000		
y	0.773	-0.027	-0.021	-0.289	-0.274	0.097	0.217		
				Six-variable model					
x_1	-0.954	-0.265	0.138	-0.035	-0.008	-0.009			
x_2	-0.777	0.627	-0.064	0.007	-0.013	0.004			
x_3	0.786	0.364	0.499	0.012	-0.002	-0.001			
x_1^2	-0.963	-0.220	0.117	0.105	0.003	0.000			
x_1x_3	-0.938	-0.248	0.235	-0.058	0.005	0.007			
x_2x_3	-0.742	0.669	-0.031	-0.014	0.013	-0.004			
y	0.783	0.043	0.093	0.336	-0.017	0.214			
				Five-variable model					
x_1	-0.473	0.341	0.080	0.103	-0.802				
x_1^2	-0.478	0.284	-0.007	0.774	0.303				
x_2^2	-0.429	-0.479	-0.760	0.090	-0.015				
x_1x_3	-0.471	0.345	0.112	0.618	0.515				
x_2x_3	-0.377	-0.675	0.635	-0.022	-0.005				
y	0.742	-0.095	-0.278	-0.124	0.351				

Table 8. Summary of principal components regression using all explanatory variables

Principal components used (P_p)	Number of P_p's	R^2	Tests[a]		
			Heterogeneity of variance (χ^2)	Lack-of-fit (F)	Overall (χ^2)
$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9$	9	0.83	211.94 (151)	2.93 (15,151)	273.55 (166)
$P_1, P_2, P_4, P_5, P_6, P_7, P_8, P_9$	8	0.83	212.58 (151)	2.84 (16,151)	276.65 (167)
$P_1, P_2, P_4, P_5, P_6, P_7, P_9$	7	0.83	210.71 (151)	3.03 (17,151)	281.11 (168)
$P_1, P_2, P_4, P_5, P_6, P_9$	6	0.82	210.25 (151)	2.84 (18,151)	281.51 (169)
P_1, P_2, P_4, P_5, P_9	5	0.81	209.84 (151)	2.74 (19,151)	282.25 (170)
P_1, P_4, P_5, P_9	4	0.79	211.65 (151)	2.84 (20,151)	291.34 (171)
P_1, P_4, P_5	3	0.72	211.47 (151)	4.75 (21,151)	351.30 (172)
P_1, P_4	2	0.65	214.03 (151)	4.93 (22,151)	367.83 (173)
P_1, P_2	2	0.60	212.42 (151)	4.49 (22,151)	351.51 (173)
P_1	1	0.58	220.04 (151)	5.26 (23,151)	396.30 (174)

[a]Values in parentheses are degrees of freedom. See ref. 4 for details.

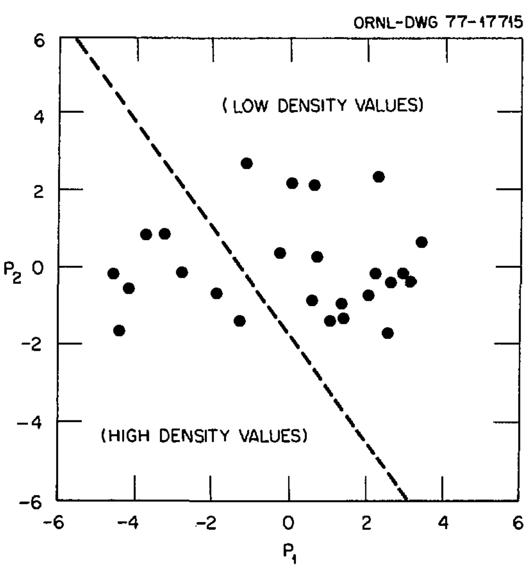


Fig. 5. Plot of P_1 and P_2 from full nine-variable model.

compared for each principal component, that is, the explanatory variables with comparable positive and negative loadings.

Because the correlation between the principal components and the explanatory variables, as well as the response variable, is of major importance in any principal components regression, Fig. 8 is a graphical representation of these correlations to aid in the choice of principal components. In this figure the correlation between the explanatory variables and the principal components is shown for each principal component. In addition, the correlation between each principal component and the response variable is also exhibited. This figure is presented for the eight-variable model and vividly displays the following: (1) the decreasing correlation between the explanatory variables and the principal components as the eigenvalues decrease, that is, as one considers P_1 relative to P_2, P_2 relative to P_3, etc., (2) the explanatory variables highly correlated with the principal components; and (3) the principal components highly correlated with the response variable. A figure of this type is a graphical aid in demonstrating the alternative choices one can make in reducing the dimensionality of the problem, that is, whether the choice of appropriate principal components is based on correlation with the explanatory variables or correlation with the response variable.

CONCLUSIONS AND PROBLEMS

The regression analysis of the field data summarized in this paper has demonstrated the need for the continued interaction between the statistician and investigator during the data analysis process. The original polynomial approximation in Eq. (2) has done a reasonably acceptable job in explaining the variation ($R^2 = 0.83$) in the observed clutch size. The regression analysis and the principal components regression have also provided some plausible choices of important variables and their relationship from which the biologists can choose the biologically most acceptable alternative. From these alternatives, it is possible to suggest some additional laboratory studies to quantify the effects implied from the

Table 9. Principal components regression on reduced models

Stage	Number of variables retained	Variables retained	R^2	Tests[a]		
				Heterogeneity of variance (χ^2)	Lack-of-fit (F)	Overall (χ^2)
1	9	$x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$	0.83	211.94 (151)	2.93 (15,151)	273.55 (166)
2	8	$x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_3, x_2x_3$	0.82	211.99 (151)	2.74 (16,151)	273.59 (167)
3	7	$x_1, x_3, x_1^2, x_2^2, x_3^2, x_1x_3, x_2x_3$	0.81	208.26 (151)	4.03 (17,151)	302.69 (168)
4	6	$x_1, x_2, x_3, x_1^2, x_1x_3, x_2x_3$	0.78	212.24 (151)	3.68 (18,151)	305.46 (169)
5	5	$x_1, x_1^2, x_2^2, x_1x_3, x_2x_3$	0.78	209.23 (151)	4.57 (19,151)	329.58 (170)

[a]Values in parentheses are degrees of freedom. See ref. 4 for details.

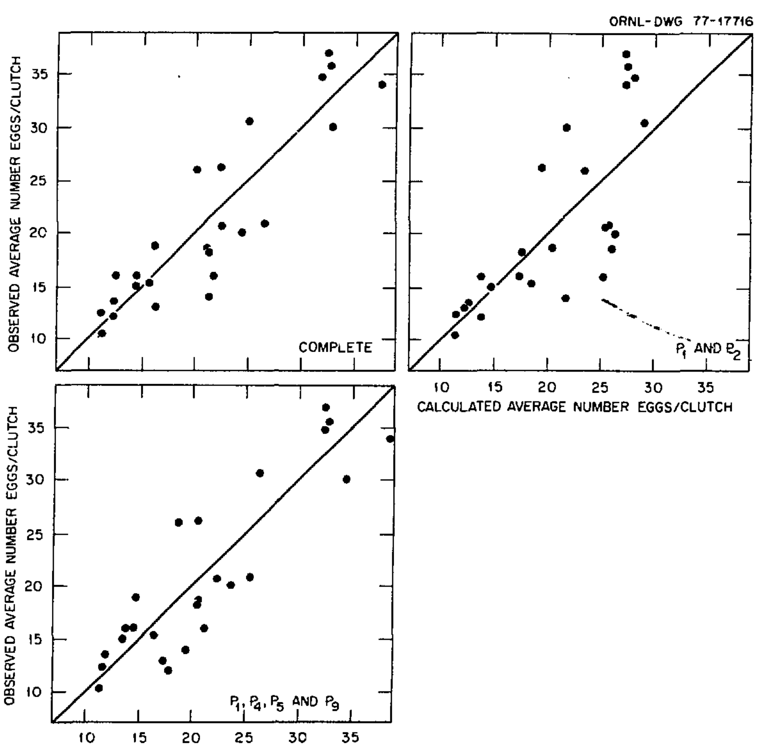


Fig. 6. Comparison of observed and calculated clutch size from various principal components regression models.

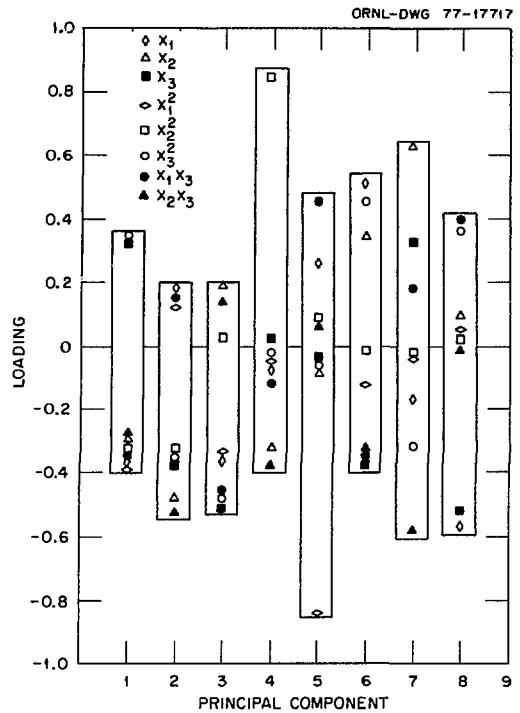


Fig. 7. Loading for eight-variable principal model.

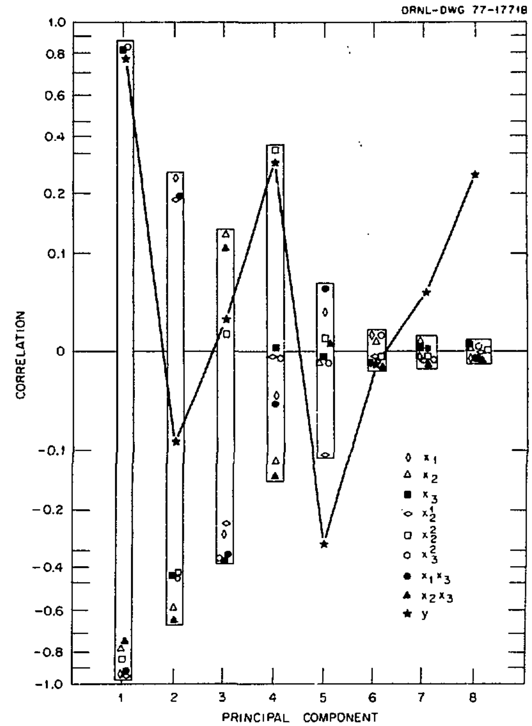


Fig. 8. Correlations between explanatory variables, response variable, and principal components from eight-variable model.

explanatory variables. The results of these laboratory experiments can then be used to refine future field experiments. This iterative process could be continued through many phases.

There have been some specific suggestions for designed laboratory experiments resulting from this analysis. For example, laboratory studies have been initiated to examine the effect of adult density and female size on clutch size when the ranges and values of the two explanatory variables are investigated under known and controlled environmental conditions.

The regression analysis did still leave some problems concerning the adequacy of the model. However, collaboration with the biologist has resulted in some plausible explanations for the apparent unexplained variation by the model that may be incorporated with the laboratory studies to modify future field studies. Some of the explanations are (1) water temperature was the only water quality variable recorded and there may be a need to measure other water quality variables, (2) the use of average water temperature may need to be refined using additional information about the water temperature history, and (3) the inherent variability of any uncontrolled biological system is always difficult to explain.

Any suggestions for alternate analytical approaches or refinements and modifications to the approach as presented are solicted. These approaches should recognize the need to assist in providing plausible explanations and relationships so the biologist and statistician can interact in solving the problem.

Problem Discussion 2, Part 2: Use of Regression Analysis to Evaluate Environmental Effects: Exploring Methods of Analysis

John Beauchamp, Union Carbide Corporation, Nuclear Division

John Beauchamp: For review, my objectives for this problem were to determine if the explanatory variables related to water temperature, adult density, and female size influence clutch size and if they can be used to explain the observed variation in clutch size. In this particular situation, I (and I hope the biologist) am thinking in terms of possibly interpolation and maybe some prediction within the range of the variables of interest. The approaches that I have used there did appear to be some things that we had gained from it and also some problems that were still present both with respect to the polynomial approximation that was used and also to the principle components regression approach that was used. These are some of the things that we talked about yesterday, and I would welcome some comments. I must admit that for the units on the female size, there is a factor missing. This has been brought to my attention, and it will be corrected in the final version. Of course, this factor is not as big as the units given in the paper. They are only fractions of inches rather than 4 in. or so as indicated in the table of data given. I would welcome any comments, or suggestions to alternative approaches that you would have.

Donald Gaver, Naval Postgraduate School: I'm wondering if an alternative analysis transforming the clutch size by taking logarithms could not have been an effective approach, and whether this approach was tried. It seems to me that that might, first of all, remove the necessity for weighting in response to different variances and, second, possibly separate the interaction terms that appeared in the cross product of terms that is part of the original model. It might make those other representations unnecessary.

Ronald Thisted, University of Chicago: Do you have a viewgraph on which you plot the mean against the variance in the different samples? Did you say that you found extra Poisson variation—more variation than you'd expect?

John Beauchamp: Yes, this is the plot I believe you're referring to (Fig. 1). Yes, there were samples where we did find the presence of extra Poisson variation.

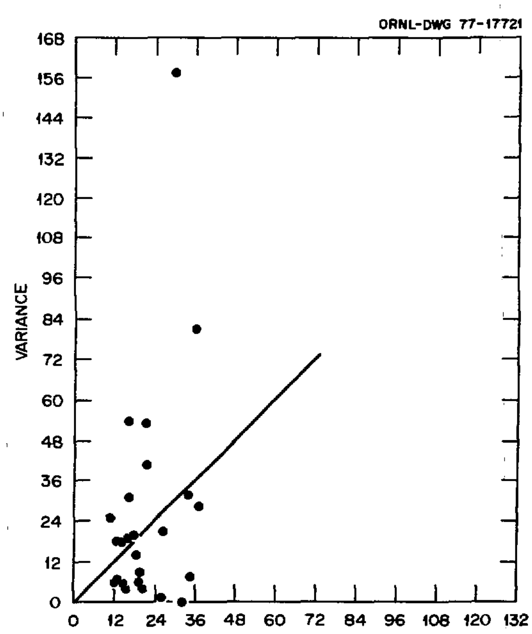


Figure 1

Ronald Thisted, University of Chicago: One thing that comes to mind: the Poisson distribution is appropriate for counting data; however, these really aren't counts, but are tallies of counts. For each female you have a count; one way that you could make the problem very much harder and maybe only a little bit more realistic would be to say that each female has its own Poisson parameter and say that the distributional Poisson parameter is gamma. Then estimate the parameters that introduce extra variation. It's just as if the gamma were a prior distribution except that it really has physical mean, as opposed to a suggested belief, but then you could estimate the parameters of that compound process which could account for the extra variation.

John Beauchamp: Can I ask for a little expansion on this with respect to the presently available data? The individual measurements of length of the females have been lost in the reduction of the data; couldn't this loss cause problems in the approach that you are thinking about? Or am I misunderstanding the approach that you are suggesting?

Ronald Thisted, University of Chicago: Well my approach is a partly baked idea. I think the loss of these measurements would cause problems because you use that as a covariant essentially on the mean, right? I just saw all the individual egg counts in your data display, and I guess I just assumed that all the other measurements were there as well. That could be the fatal blow to my suggestion.

William Conover, Texas Tech: I think that it struck many of us that the asumption of the Poisson distribution is going to be difficult to prove or disprove with the data that you have, where the parameter apparently changes from one day to the next and there are only a few observations. For example, on one day you had two observations, both observations equaled thirty, and you accepted the hypothesis that this was a Poisson distribution. It's very difficult to make a decision on data like that. If you insist on using this statistic $(N-1)s^2/\bar{y}$ that you were using, it seems to me that at least it would be better to use the two-tailed rejection region rather than the one-tailed, because too small of a variance would be just as unsatisfactory as too large a variance. Then when you went through and added the statistics together, the large statistics canceled with the small statistics, and the result was an overall statistics which again to me would not have very much power in detecting lack of a Poisson distribution. So I wonder whether it might not be better to abandon the whole idea of trying to stick with the Poisson distribution and follow something like Don Gaver suggested or perhaps use rank transformation. I suggested rank transformation because you do have a problem with outliers. The outliers appear to be legitimate from the biologist's standpoint, and yet they affect the calculations a great deal. By using the rank, an outlier just has rank one, it doesn't matter how far out it is, and it's not going to affect your data very much. We've had a lot of success using rank transformation when there are outliers. You might try to rank the data and use the same techniques that you are using to see if the results are in agreement. You can work backwards to see how well it predicts what you got, as a backup to the methods you are using.

John Beauchamp: In this particular situation, would you also suggest working with the ranks on the variables that I have referred to as explanatory variables—that is, working with ranks on the entire data set instead of the actual observations. This approach would be interesting.

John Jaech, Exxon Nuclear Co.: In the spirit of the comments made by the other two gentlemen, I think that I would agree. I really don't see any basis for even performing a test of the hypothesis that the data are Poisson. I think we can reject that hypothesis on the basis that they are from different females of different sizes. If they are, in fact, different sizes, then there is no reason to believe that anything more than female size has an effect; there's no reason to believe that the data should be Poisson. If memory serves me correctly, the compound distribution that results from the fact that the Poisson parameter has a gamma distribution is a negative binomial distribution which has a known mean and variance and which is very easy to work with. I think that you can get a lot more insight into what's really going on if you assume at the beginning the negative binomial and work backwards and get a feel for how the actual means per cell are varying according to the underlying gamma. I think that is a real good suggestion.

Donald Gaver, Naval Postgraduate School: I didn't mean to imply in any way that we should focus on the distribution of the cluster sizes to the exclusion of all else. It seems to me that actually the negative binomial distribution is certainly suggested by the scatter plots that were shown and possibly by mechanisms. I believe that some form of the logarithmic transformation will indeed prove to remove the dependence of the mean and the variance. I think that you buy a great deal from this kind of transformation, at least potentially, because you also buy the removal of the interaction that is present in the model. I think there's no hope of determining or verifying any particular distributional form for these cluster sizes; that's a false trail to go down.

Lincoln Moses, Stanford: I never understood exactly what the purpose of this study was. Whether it is to understand causal mechanisms or to interpolate and graduate data over other variables. So I don't know whether the following suggestion helps or hurts. I recall that there was a trend (I forget whether it was up or down) with a clutch number or date arranged in order of this trend. At the same time, you told us that the temperature of the water exhibits a trend with that order. It's quite possible that the density of adults exhibits a trend in order. You might find that putting in time, or even a quadratic in time, would capture a great deal of what you are now capturing against three variables. You could see if any of them add beyond that. I do not know whether it would simplify your interpretation or give you better prediction, but time may be what's underlining several of these variables in part.

John Beauchamp: The problem that I have with including time is that, in any biological mechanism, the critter or animal is not particularly concerned with the time mechanism. Time, true, is a conglomerate of many factors, and that was why I was trying to look at these individual factors—to see if possibly one factor or maybe all of them were of interest. I'm not sure whether the biological significance of time would be very meaningful even if it did explain a great deal, which I think it does. Of course the biologists would have to have some input on this. That's the difficulty that I had with the inclusion of time, though I'm not sure that answers the whole question!

Dave Gosslee, Union Carbide Corporation, Nuclear Division: I thought the point was to look at time as a random variable, so to speak, and then see what additional variation could be explained by temperature, etc.

Lincoln Moses, Stanford: Whether times are suitable depends on what the purpose is. I really can't explain or understand; the biological significance is exceedingly obscure. If you have three variables that are just servants for the unfolding development of this pond, then putting in something on time tells you that you have a problem that is very complicated, I guess! It might be a useful thing to do. If you don't understand, then it is no help that I can see.

Chuck Bayne, Union Carbide Corporation, Nuclear Division: John, I noticed you did a lot of principal component analysis, and one of the difficulties I always find with principal components is to try to interpret what they mean to the experimenter. Now I noticed you were able to put them in some chart form and to determine whether the loadings were high or low. Were you able to use these facts to explain what these principal components actually represented to the biologists?

John Beauchamp: At this particular point the principal components approach is still an open problem. I hope that something may be gained from these charts when I sit down with the biologists. If others have had experience using this approach and trying to explain it to the investigator, I would like to hear about your experience.

Ronald Thisted, University of Chicago: I've been thinking a little bit about the negative binomial suggestion. There are no problems really created by the fact that you haven't recorded the length of each female or the fact that that's gone; this new analysis would follow exactly the same form as the old one. One of the problems now is that you have two parameters involved. The two parameters are the gamma distributions that overlie everything as opposed to the one Poisson parameter. I don't know how to do the regression analysis to get those two pieces. I wouldn't know a reasonable model for that. I think that's where the difficulty is, not in the setup of the covariance.

Francis Anscombe, Yale: Don Gaver has already suggested trying to play the transformation game on the variables, and I would just like to reinforce that suggestion. The regression analysis done here is on quadratic expressions in the three experimental variables. Why stop at quadratic? Why not go to cubic? We all know why not go to cubic; that produces a very large number of coefficients, and clearly one isn't going to be able to estimate them all properly. The trouble is, however, that one does not know that the quadratic expression is adequate. We don't know what kind of expression is adequate, and I'm sure it will be very worthwhile to try playing tricks on transforming the dependent variable and the three experimental variables, making transformations to see whether one can get apparently good approximations to simple linear relations between them. Although there is no guarantee that playing the transformations will greatly simplify the expressions that need to be fitted, sometimes one really does come off very well that way. I think there is a little internal evidence in the plots that I was given that even the quadratic model is not fitting very satisfactorily. Note the figures in the written material here which were labeled Fig. 3.* There are two figures there: one for all nine variables fitted and the other for five. Both of those figures, but particularly the five-variable one, have some suggestion, I think, of a quadratic trend: that is, the residuals appear to be positive when the abscissa value, which is the calculated clutch size, is either very low or very high, and the residual seems to be negative when the clutch size is in the middle. Then in Fig. 4,+ the bottom right diagram there shows residuals plotted against predicted clutch size. This is another plot of a similar kind, I think, if I understood that right, and again there's a tendency for the residuals to be positive at the left side and the right side of the diagram and a trend to be negative in the middle. That sort of curvature does suggest some sort of nonrelativity in the model that is actually being fitted and also suggests, perhaps, that some transformation of the dependent variable still would help, so I would certainly like to support Don Gaver's suggestion that making various scale changes should be tried.

*J. J. Beauchamp, these Proceedings, Fig. 3.

+J. J. Beauchamp, these Proceedings, Fig. 4.

Statistical Methodology for Use in Risk Assessment of Radioactive Waste Disposal in Geologic Media*

Ronald L. Iman

Sandia Laboratories
Albuquerque, New Mexico

ABSTRACT

The development of a statistics methodology for use in risk assessment of radioactive waste disposal in geologic media poses a unique and interesting challenge. The models involved in the assessment are sometimes quite long and involved. Some of the models contain many defined but unknown parameters, and in many cases little or no data are available for parameter estimation. Also, many variables are used, and only a few of them may be of value. Once the individual models are analyzed, it is necessary that the various models interface properly and that some meaning is attached to the output, which is a random function of the input and is dependent upon site selection. We hope to develop the statistical methodology for analyzing such a complex system.

INTRODUCTION

My presentation concerns some of the statistical problems that are associated with the disposal of radioactive waste material in geologic media. This research is done under a contract with the Nuclear Regulatory Commission (NRC) which requires that at least one statistician be assigned to the project. Our group, headed by Dick Prairie, is working on a variety of problems; earlier Irv Hall discussed Sandia's approach to energy problems.

PURPOSES AND APPROACH

The purpose of our work on this problem is to provide insight into the risk associated with radioactive waste disposal and to develop methods and models that can be used in the repository licensing program by the NRC staff. Hopefully, this program would include identification of parameters that determine long-term safety.

There are two steps in the approach, one of which would be involved with risk calculations where we would hope to gain an understanding of the risk of disposal in different geologic media. At the current time there will be a *particular* site for study; different types of geologic media will have to be considered later. The second step in the approach is in regard to the sensitivity studies. We hope to determine the important site and waste characteristics which would in turn be used by the NRC staff to assess a particular site for a disposal site license.

An overview of this problem for a hypothetical site is shown in Fig. 1. (Note the vertical exaggeration.) There is a lens of salt in which the repository would be placed. On either side of the salt there is a shale formation, and above and below the shale there is a sandstone aquifer. We do not want any water to get into the repository from the aquifer and thereby carry the radionuclides to the environment. I would like to emphasize that, while Fig. 1 represents a hypothetical site, we still hope to make it as realistic as possible.

---

*In support of a project being performed by the Fuel Cycle Risk Analysis Division of Sandia Laboratories. Work is funded by the Nuclear Regulatory Commission.

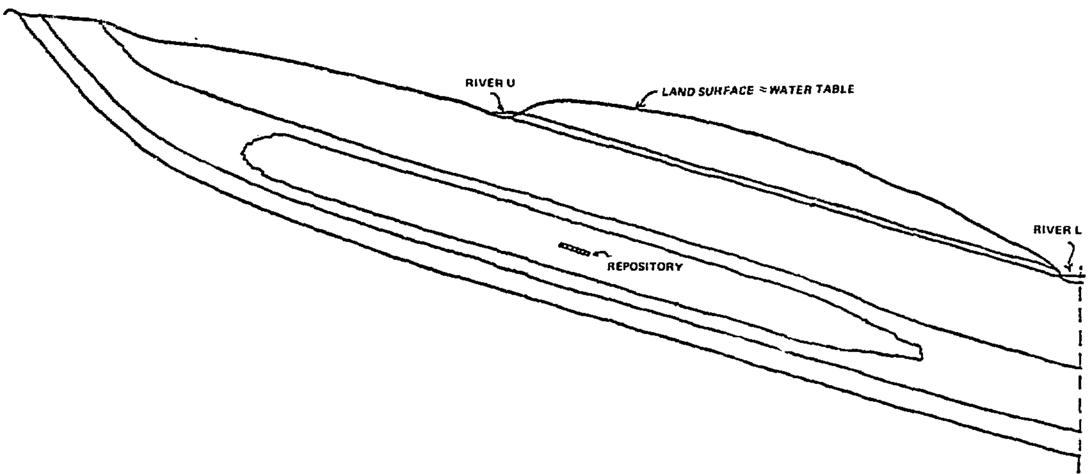


Fig. 1. Hypothetical geologic cross section. (Vertical exaggeration: ×10).

MODEL STRUCTURE AT SANDIA

There are a number of individuals at Sandia involved in this project (Fig. 2). The site description indicated in the upper left-hand corner includes such items as the engineered facilities (shaft and excavation), hydrology, geology, and surface characteristics. The box in the upper right-hand corner (radioactive waste description) will provide information at any one point in time regarding the types and amounts of waste materials. This information includes knowledge of the half-lives of the various nuclides, so in the event of a release at some point in time the chemical composition of the waste is known. The boxes in the middle of Fig. 2 represent models whose development involves at least one person.

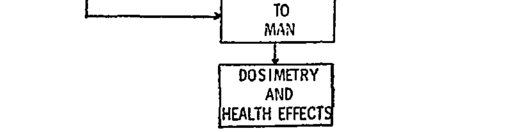


Fig. 2. Problem format.

These models have been developed at Sandia, or on contract by Sandia with Sandia personnel involved. Although these models exist at the present time, not all of them are developed to the point where we are ready to do a sensitivity analysis on them. Once that work is finished, the problem of coupling the models will have to be addressed. The first box in the middle (the release mechanism) will output the probability of release as a function of time as well as the rate of release. Items that this model would consider are those for which a release could take place—for example, an earthquake, undetected faults, a shaft that might fail to seal properly, or something as remote as a meteorite striking the site.

The transport model represents probably the biggest single task because it is a very complicated model with many, many variables. One of the problems associated with this model is that there are little if any data available that pertain to it. Therefore, a lot of the variables are going to require estimates just to determine a physically reasonable range. This type of problem occurs throughout this project. There are so many variables in the transport model that for the last few months two people at Sandia (including a geologist hired in from the outside) have been attempting to reduce this model to a workable size. That work is progressing now, and the sensitivity analysis should soon start on the transport model. When the transport model is coupled into the system, it will provide the rate of discharge of the radionuclides to the environment. Given that there has been a release at some point in time, the nuclides are transported to the underground water system from which they will eventually reach the environment. Once they have reached the environment, the pathways model is used. The pathways model uses the input of the transport model to calculate ingestion and inhalation rates of radionuclides in curies per year for maximum and average diets. This analysis is done with respect to the types of food consumed by a particular individual and with respect to whether or not the individual consumes irrigated food. There are two basic parts to the diet: (1) a water-based diet with foods that come from water sources, and (2) a land-based diet that comes from land-type sources. Also, a combination of these is possible. The dosimetry and health effects model uses the ingestion rates output by the pathways model to calculate the probability of a latent cancer fatality as a function of time.

DEPENDENT VARIABLES FOR THE PATHWAYS MODEL

I have attempted to simplify this problem. For instance, there are many isotopes that could be ingested, but we have looked at only one. Therefore, everything that you see with respect to ingestion rates is only for one nuclide. Also, there is a problem of ingestion rates changing over time. I will show some predictions for a model that uses only one time period, but I do have several plots which indicate that time probably will be a consideration. Next there is a concept of a zone. The pathways model uses the concept of a homogenous zone along some water source with nuclides in it, such as a river. For example, there may be a large area where all the food is irrigated—which would determine one zone. If a dam were put on the river, another zone would be created because of the sediment of the nuclides. There would be yet a different zone below the dam. There could yet be other zones where there is no irrigation. Initially we consider only one zone.

The "average" and "maximum" classifications on Fig. 3 come from the WASH-1400 report descriptions given for an average individual and a maximum individual. The other classification in the two-way table is for individuals having either irrigated or nonirrigated food. I consider that there are basically eight dependent variables with the subscripting notation explained as follows. The subscripts 1, 2, 3, and 4 refer to the individual and irrigation combinations. The second part of the subscript uses W to refer to ingestion from water sources (possibly something as simple as drinking water, but could include sources like invertebrates and fish). The subscript L refers to basically a land-type diet such as plant, milk, or beef. Therefore, in block 1 if there is no irrigation involved with a land-type diet, the ingestion rate is going to be much smaller than it would be where irrigation is involved. The difference in the dependent variables Y_{1L} and Y_{3L} is in the magnitude of the intake; therefore the independent selected variables for these diets should be very similar. Due to the complexity of the pathways model, I attempted to fit a response surface to the output to determine the important independent variables associated with each dependent variable.

INDEPENDENT VARIABLES FOR THE PATHWAYS MODEL

A few comments need to be made with respect to the independent variables (Fig. 4). First, I conveniently refer to these variables as X_1 to X_6 and

Irrigation		Individual: Average	Individual: Maximum
No		Y_{1W} (Water)	Y_{3W}
No		Y_{1L} (Land)	Y_{3L}
Yes		Y_{2W}	Y_{4W}
Yes		Y_{2L}	Y_{4L}

Fig. 3. Dependent variables: ingestion rates of radionuclides in curies per year.

$$10^2 \le X_1 \le 10^5$$

$$2.6 \times 10^5 \le X_2 \le 2.6 \times 10^8$$

$$2.9 \times 10^9 \le X_3 \le 2.9 \times 10^{12}$$

$$6.0 \times 10^5 \le X_4 \le 6.0 \times 10^8$$

$$3.6 \times 10^9 \le X_5 \le 3.6 \times 10^{11}$$

$$1.8 \times 10^9 \le X_6 \le 1.8 \times 10^{13}$$

Fig. 4. Ranges on independent variables.

have not attempted to attach any names to them, as I felt in the format of this meeting it would be just as well to exclude them. However, these variables were carefully selected by the individual who developed the model. He also selected ranges for these variables that he felt were physically reasonable (Fig. 4). A major point with respect to these variables is that there is no probability distribution given with them, which is the case throughout the pathways and transport models. About the best information available are the ranges associated with the different variables. These ranges are relatively broad and are in terms of orders of magnitude.

LATIN HYPERCUBE VARIABLE SELECTION TECHNIQUE

To fit a response surface, it is necessary to run the program for the pathways model several times using various combinations of the independent variables as input. For variable selection I used the Latin hypercube variable selection technique. This technique was presented at the first ERDA symposium in Los Alamos, and Mike McCay, Jay Conover, and Dick Beckman will soon have a paper in *Technometrics* explaining its advantages. Very briefly the Latin hypercube variable selection technique works as follows. Assume there will be N observations used as input into a model. The range of each one of the input variables is divided into exactly N nonoverlapping pieces. The procedure for selecting the pieces may be something as simple as using a uniform distribution which makes all pieces the same width. If the probability distributions associated with these variables is known, the pieces could be selected based on equal probability. Once each range has been divided into N pieces, a value is selected at random in the ith interval for say X_1. Likewise, for X_2 select a value at random in the jth interval. Continue in this manner for each independent variable until each interval has been used exactly once. Next obtain a random mixing of these values as input. A question associated with this technique is what type of distribution should be assumed for these variables—or does it make any difference?

PARTIAL CORRELATION PLOTS

I first assumed a uniform distribution for each of the input variables, as indicated in Fig. 5. For example, on X_1 the range should be divided in exactly N nonoverlapping pieces all of the same width from 10^2 to 10^5. For $N = 50$ this procedure gives one point between 10^2 and 10^3, four points between 10^3 and 10^4, and the remaining 45 points between 10^4 and 10^5. The pathways model was run with these 50 input observations and the partial correlations plotted as a function of time as the pathways model outputs ingestion rates as a function of time. The horizontal axis in Fig. 5 has ten points with the time scale in hundreds of years, so the first point is ingestion rate after 100 years, then 200 years, and so on to 1000 years.

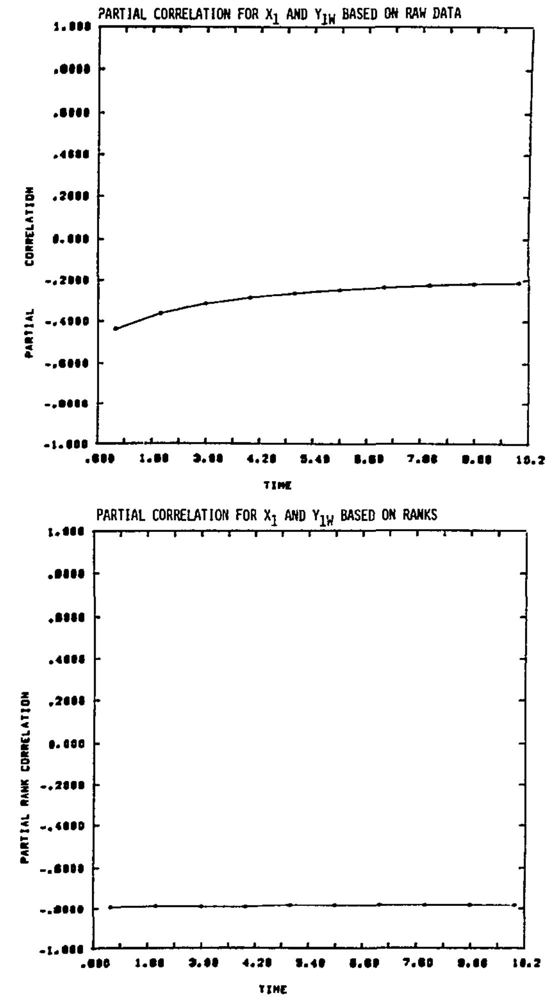


Fig. 5. Partial correlation plots for X_1 and Y_{1W}—uniform distribution.

The top portion of Fig. 5 indicates results when using the raw data. The bottom of Fig. 5 indicates results when using the rank transform on the data (partial rank correlation). Clearly, there is a large disagreement with respect to the importance of X_1. The ranks indicate that X_1 is extremely important as the partial rank coordination remains constant around -0.8. The raw data partial correlation changes over time, starting around -0.45 and getting close to -0.2.

In Fig. 6, a log uniform distribution is used on the independent variables. Taking the $\log_{10}$ on 10^2 and 10^5 will give a range between 2 and 5 for X_1. Divide this range into 50 equal pieces; then one-third of the points will be between 10^2 and 10^3—rather than only one point obtained using the uniform. There will be another one-third between 10^3 and 10^4. The net result of the log uniform is to give more observations toward the lower end of the range on the independent variables. A comparison of the raw data graphs at the top of Figs. 5 and 6 indicates a slightly different story with respect to change over time. Therefore there are going to be some difficulties in determining the important variables with respect to time, depending on the type of distribution that is assumed on the input.

Figure 7 indicates the results of pooling the 50 points from Fig. 5 and the 50 points from Fig. 6. As one might guess, the result is a compromise between Figs. 5 and 6.

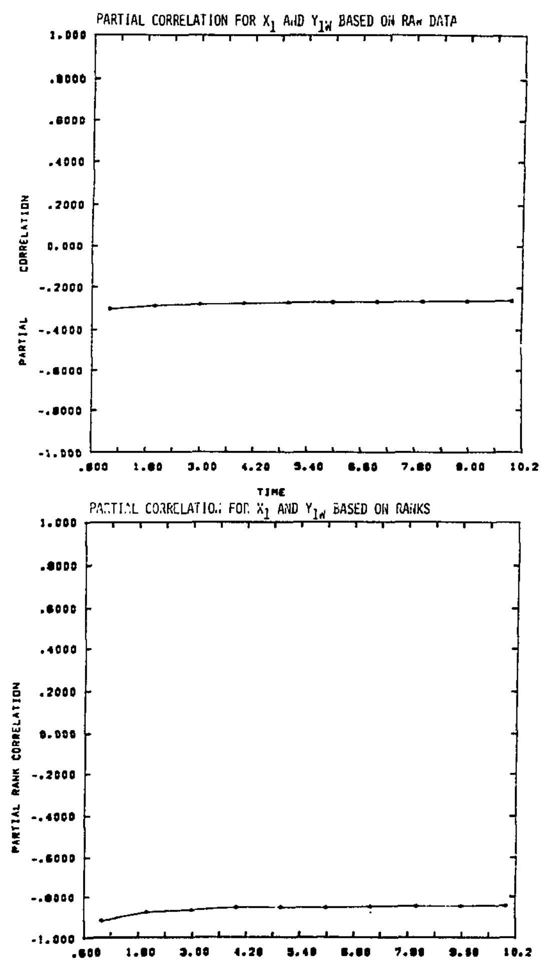


Fig. 6. Partial correlation plots for X_1 and Y_{1W}—log uniform distribution.

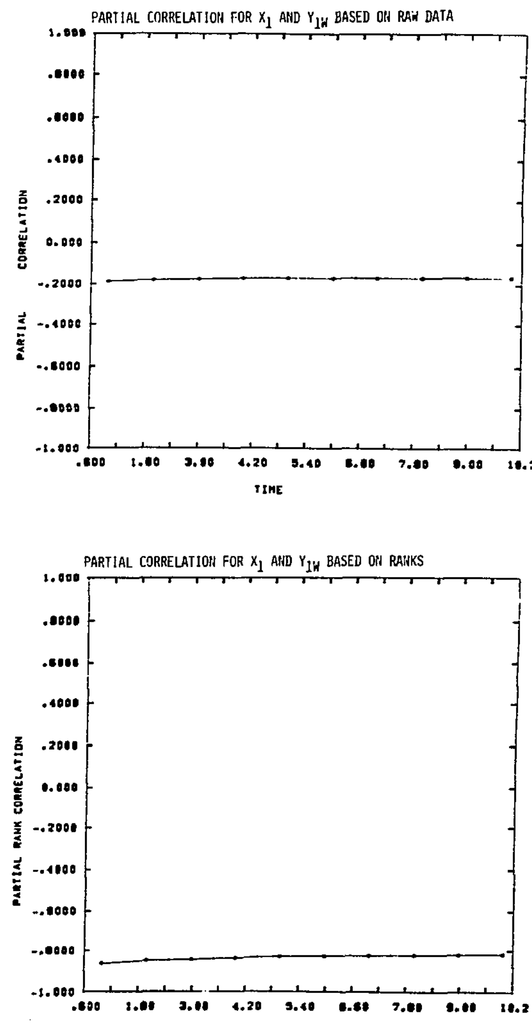


Fig. 7. Partial correlation plots for X_1 and Y_{1W}—mix of uniform and log uniform distributions.

The uniform distribution is shown again in Fig. 8, but now the dependent variable has changed to Y_{1L} (land-based diet, no irrigation). The reason for considering Y_{1L} is that there is no irrigation involved and, since the nuclides are in the groundwater, chances for ingestion of radionuclides are reduced. It would be reasonable to expect different variable selection in fitting a response surface to the land-based diet compared to the water-based diet. One might expect the relation of X_1 with Y_{1L} to be different from what it was before with the Y_{1W}. (Compare Figs. 5 and 8.) The raw data indicate a partial correlation that is essentially zero in Fig. 8, whereas in Fig. 5 the raw data indicate a slightly negative correlation. The analysis on ranks shown on the bottoms of Figs. 5 and 8 changes considerably. Next, I would like to compare Figs. 8 and 9 where time seems to become very important. The raw data indicates the time dependence as the importance of this variable is diminishing across the time axis. For ranks, the correlation starts around −0.6 at time step 1 and is 0.2 at time step 10. This is an interesting point. We are trying to explain the effects of X_1 on Y_{1L}, and at time step 10 the raw data indicates a negative correlation while the ranks indicate a positive correlation.

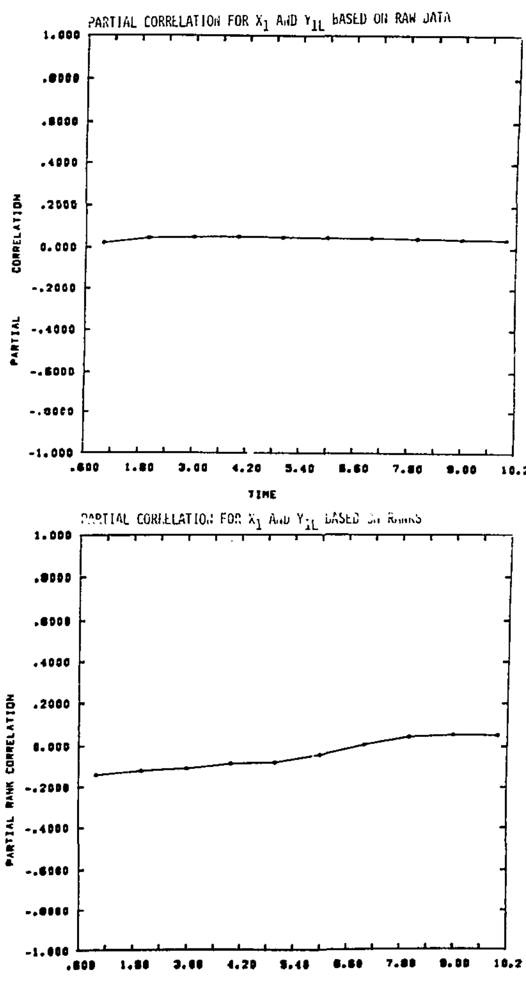


Fig. 8. Partial correlation plots for X_1 and Y_{1L}—uniform distribution.

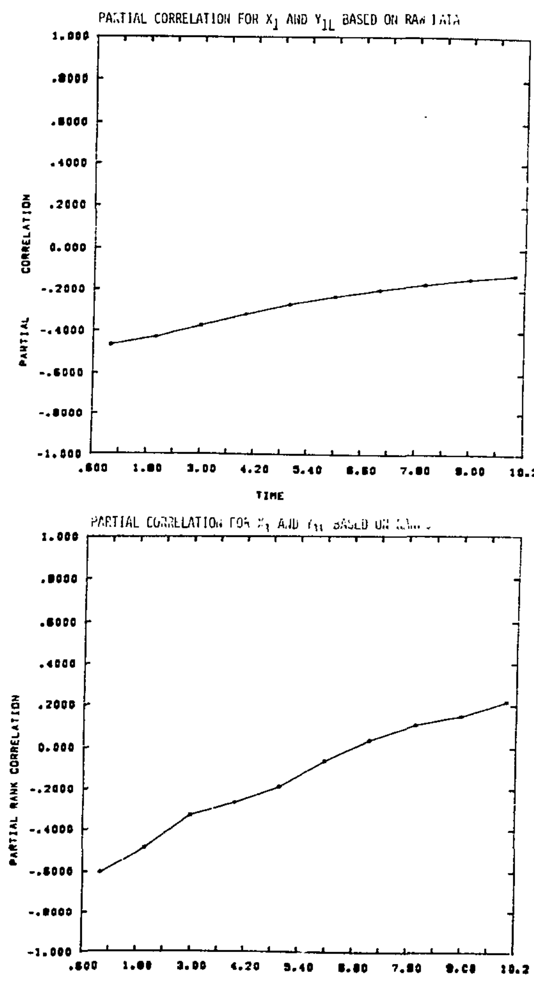


Fig. 9. Partial correlation plots for X_1 and Y_{1L}—log uniform distribution.

Figures 10–13 can be gone through very quickly. The mix on Figs. 10 and 13 again represents the pooling of the 100 points from the previous two figures. In Fig. 11 the independent variable has been changed to X_3 while still using Y_{1L}. The two graphs in Fig. 11 are roughly in agreement indicating that X_3 does not seem to be very important. However, in Fig. 12 where the log uniform was used, the importance of X_3 has changed considerably from Fig. 11.

Problems for consideration would include the following: (1) What type of distribution should one assume on the input variables? Is there a "good" distribution to assume or should one use a sequential procedure such as starting with something like the uniform, finding points where the ingestion rates seem to be quite high, and obtaining more observations in this area by using a distribution something like the log uniform. (2) There is some evidence indicating that different response surfaces are needed. As time changes, how serious is the consideration? (3) Lastly, what type of transformation would be appropriate?

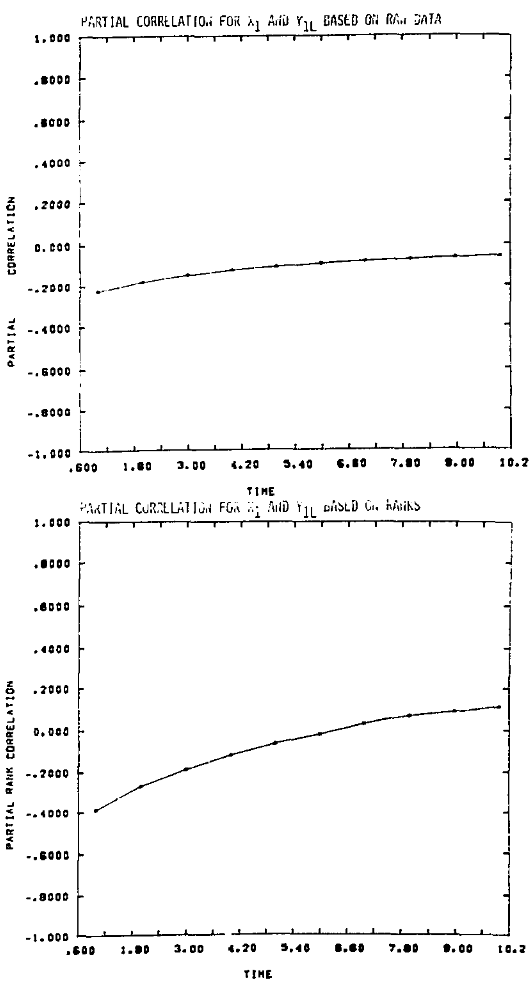


Fig. 10. Partial correlation plots for X_1 and Y_{1L}—mix of uniform and log uniform distributions.

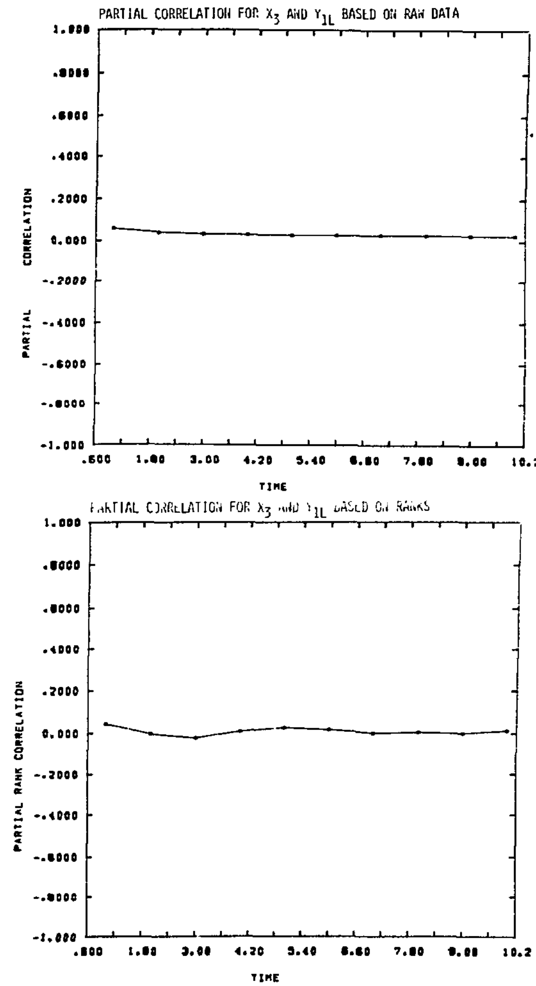


Fig. 11. Partial correlation plots for X_3 and Y_{1L}—uniform distribution.

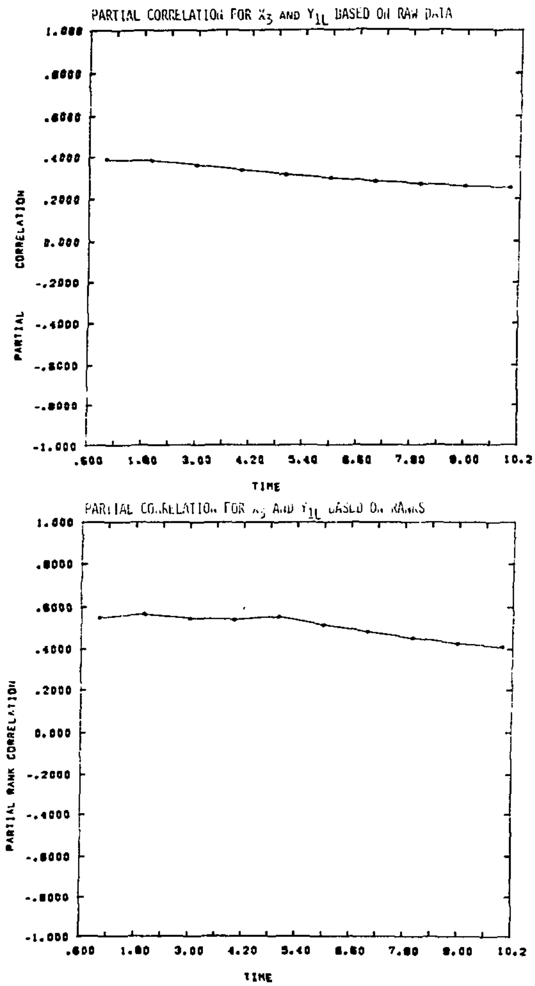


Fig. 12. Partial correlation plots for X_3 and Y_{1L}—log uniform distribution.

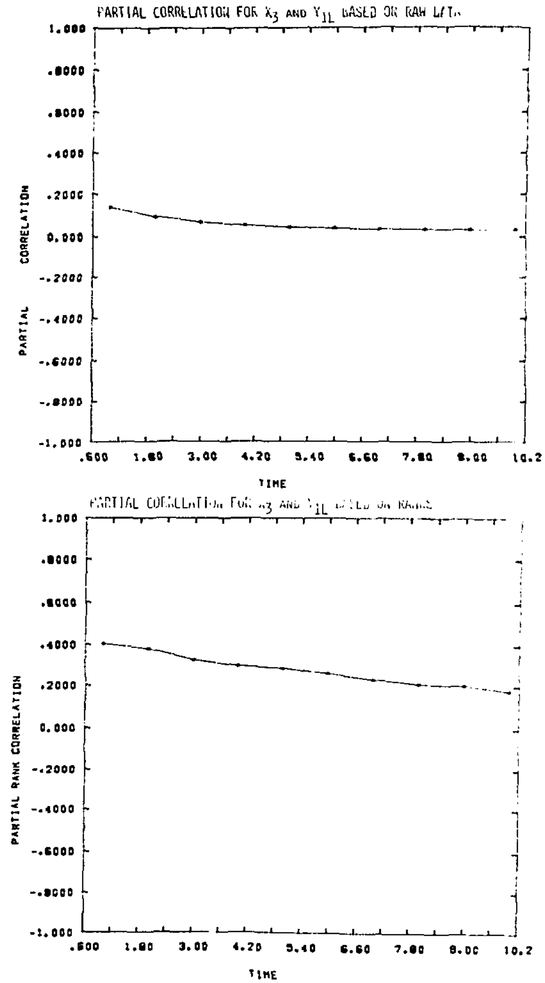


Fig. 13. Partial correlation plots for X_3 and Y_{1L}—mix of uniform and log uniform distributions.

RESPONSE SURFACES

The response surfaces were fit by using the six independent variables indicated in Fig. 4 plus the squares of each and all possible cross products which were put into a forward stepwise regression program. Initially, the raw data was used to see what variables would be selected as important. What was learned is best explained by referring to Fig. 3. Every time the dependent variable changed, a different selection of input variables was noted. That is, there didn't seem to be any consistency of selection, whereas from knowledge of the situation there should be some consistency. In particular, if the water involvement is considered (all dependent variables having subscript W, then the only thing that changes is the magnitude of the ingestion rate. Likewise, the variables Y_{2L} and Y_{4L} will act much like those variables having subscript W, due to the effect of irrigation. In other words the independent variables selected in fitting the response surfaces to these six dependent variables should be similar, and this was not the case when working with raw data. When there is no irrigation, Y_{1L} and Y_{3L}, and only the magnitude of the food intake is involved, the same independent variables should be selected for the response surface fit. Again, this was not the case on raw data. The response surface fits were made for all the dependent variables, and even though they had different variable selections, each fit was used to make predictions for an additional 50 test points. The manner in which the test points were selected is best explained by considering X_1 (Fig. 4), which has a range of 10^2 to 10^5. Sort of a compromise between the uniform and log uniform was used with ten points selected between 10^2 and 10^3, another ten points between 10^3 and 10^4, and the remaining 30 points selected from 10^4 to 10^5. The response surface based on raw data predicted negative ingestion rates for 21 of the 50 points for Y_{1W}, whereas zero is the minimum ingestion rate. The remaining 29 ingestion rates were simply nowhere in the ball park. The same thing happened for the other dependent variables.

The next attempt to fit a response surface used stepwise regression on ranks. Jay Conover and I presented a paper on this technique at the ASA meeting in Chicago in August. It is very simple to use, as the regression program is run on the ranks assigned to the variables and the variable selection noted.

The first thing noted with respect to Fig. 3 was a consistency in the selection of the variables where it should have been consistent, that is, on the four water-type dependent variables plus the two irrigated land situations. The two nonirrigated land situations selected different variables from the other six, but the two were consistent in selecting the same variables. Based on the response surface fit from ranks, predictions were made for the 50 test points. These predictions are given in Fig. 14 for Y_{1W}. $\log_{10}$ of the points was plotted just to make a little nicer graph, and the same scale was used in Figs. 14–16 for ease of comparison. The vertical axis represents the actual ingestion rate, and the horizontal axis represents the predicted ingestion rate. The zeros plotted on these graphs aid in drawing in the line where $\log_{10} Y = \log_{10} \hat{Y}$. The lines on either side represent a one-order-of-magnitude shift. In Fig. 14 the observations vary over the entire range, and the

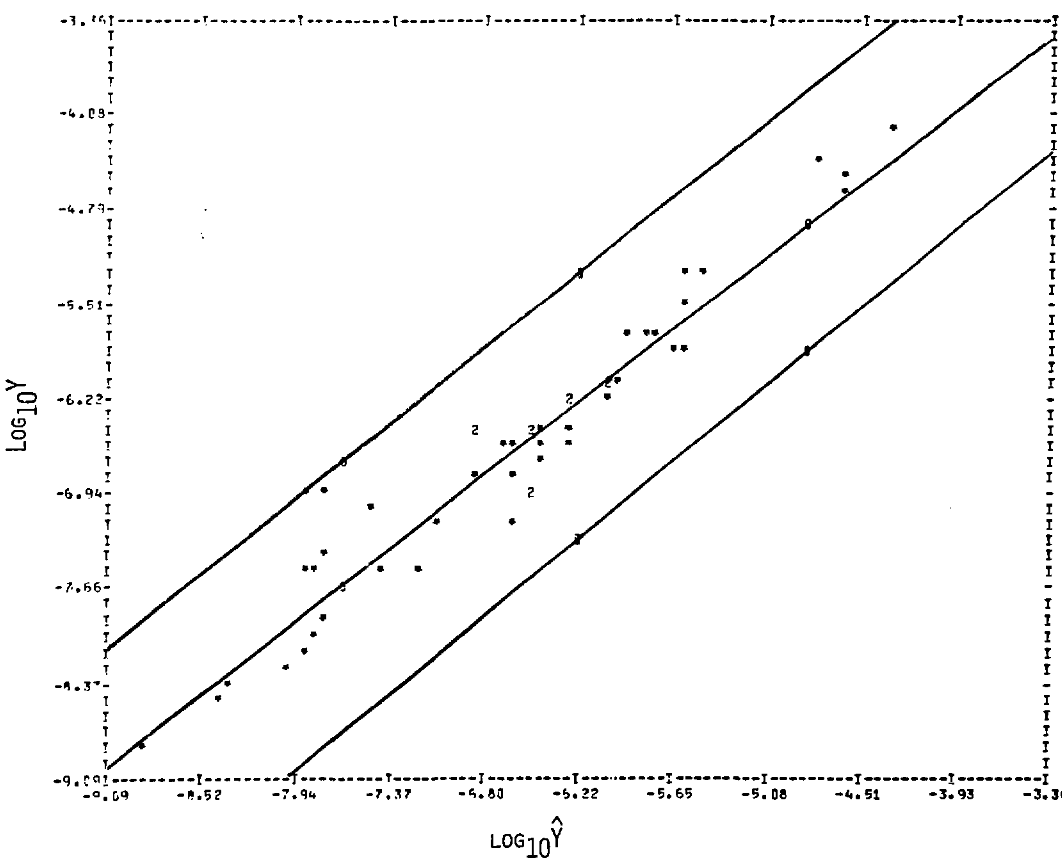


Fig. 14. Log-log plot of actual ingestion rate vs predicted ingestion rate for diet Y_{1W}.

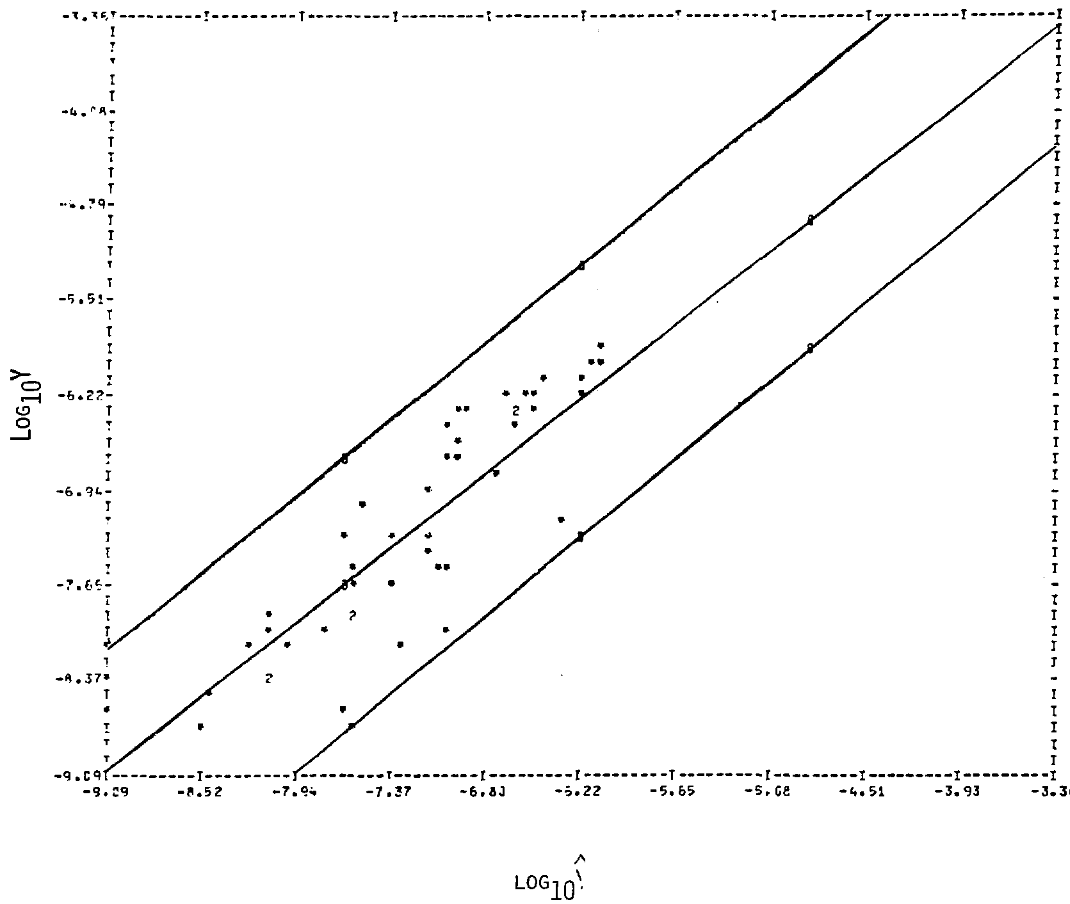


Fig. 15. Log-log plot of actual ingestion rate vs predicted ingestion rate for diet Y_{1L}.

predictions are pretty well in agreement. In Fig. 15 we have Y_{1L} (average individual, no irrigation) with a lower ingestion rate than noted for Y_{1H}, as indicated by the points in the lower left-hand corner. There are no points in the upper right-hand corner, so it seems that the fitted response surface for Y_{1L} worked well for low ingestion rates. On the other hand, when considering Fig. 16 for variable Y_{3L} (maximum individual, irrigated food, land source), the ingestion rates are going to be higher. This fact is indicated by the absence of points in the lower left-hand corner, and it seems the response surface for Y_{3I} is also predicting well for high ingestion rates.

We might consider what happens if a transformation other than the rank transformation is used—for instance, a log transformation. The log transformation was tried, and it seemed to work pretty well with respect to variable selection. It was also considered in terms of the predictions and in particular with respect to the variance associated with the predictions. For comparison purposes some arbitrary parallel lines were added to Fig. 16 on either side of the line $\log_{10} Y = \log_{10} \hat{Y}$ about a quarter inch to three-eighths of an inch away from it. The number of points included in these bounds was counted for the upper half of the graph because these represent high ingestion rates and as such are very critical. I counted 29 of the predicted points between these lines, while the log model had 16 points between the same lines; that is, there was a much greater variability associated with the predictions from the log model, so it was felt that the ranks were doing a little better.

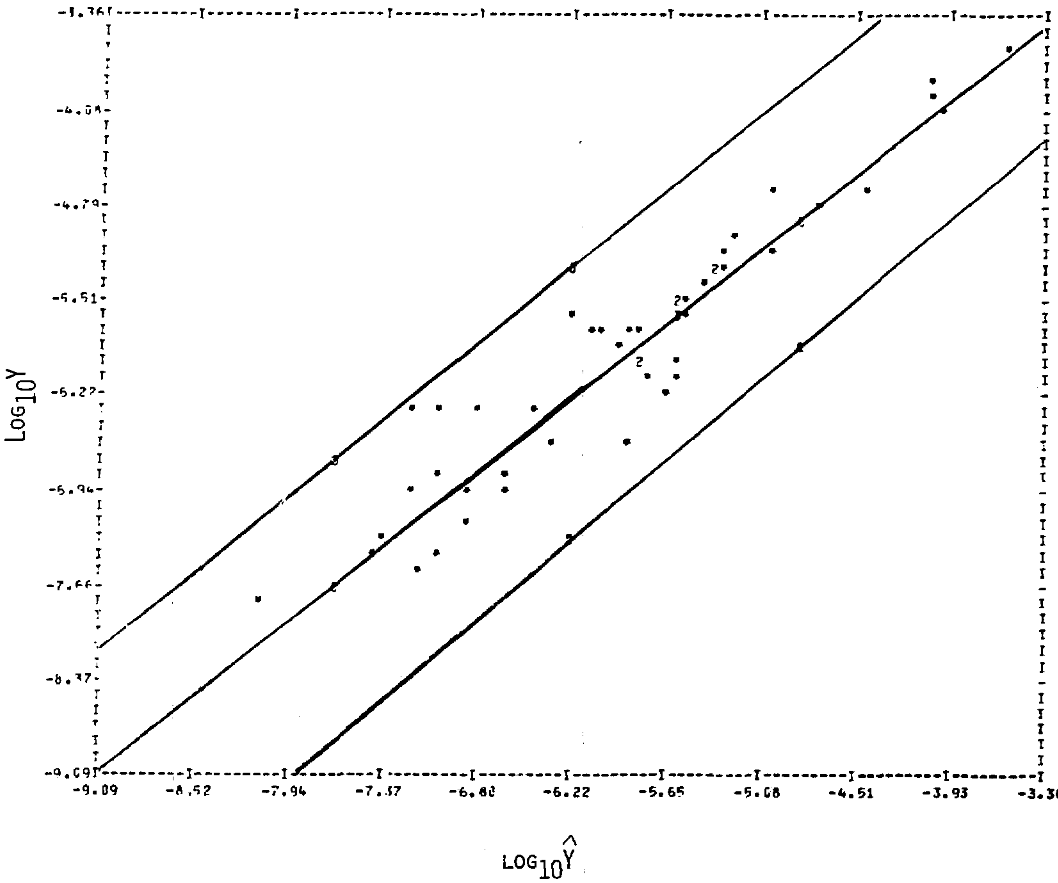


Fig. 16. Log-log plot of actual ingestion rate vs predicted ingestion rate for diet Y_{4L}.

To summarize some of the problems, how does one go about fitting the response surface to these dependent variables—in other words, is there a need for some sort of a transformation? Should an argument be made for some sort of nonlinear model? I believe that examination of the rank transform procedure will show that it is circumventing some of the problems associated with fitting nonlinear models.

Problem Discussion 2, Part 3: Statistical Methodology for Use in Risk Assessment of Radioactive Waste Disposal in Geologic Media

Ronald Iman, Sandia Laboratories

Ronald Iman: The purpose of this problem is to be able to identify the important sites and waste characteristics that would in turn be used by the NRC staff in evaluating the potential applicant for a repository before granting a license to him. The problem over the next few months will become quite large; there are a number of models which have been developed. Presently we are in the process of doing a sensitivity analysis on these models. The pathways model is the only one at this point that is on line ready to go; the others will be ready shortly. We will have some common problems in putting a response surface to the various models to help determine what the important variables or combination of variables might be. Some insight will be needed for proper approach when working with the individual models. Then big problems will come up in tying these various models together to get some sort of a meaningful output for a given scenario. Yesterday, on some of the graphs, I indicated with respect to the partial correlation the types of distributions that are assumed here. I point this out because I know that there are individuals at NRC that seem to feel rather strongly about this. I would like to get your feelings on whether a person should assume some sort of sequential procedure like I've done or whether there is really some distribution that could safely be assumed in all cases.

Gary Tietjen, Los Alamos: Ron, I think you mentioned yesterday that one of your even longer range goals in selecting the site was to decide whether there were any health effects due to injested nuclides. The effect of plutonium on the human body has been studied more extensively than the effects of any other known substance, and the literature is so extensive that people frequently overlook significant previous studies. I read one article which declared that plutonium was the most toxic substance known to man - worse than cobra venom. That phrase caught the public's fancy, and a lot of people believed it. Some 25 years ago, Langham from Los Alamos performed a spectacular experiment with plutonium, which can never be duplicated. You never hear about it except in terms of condemnation on ethical grounds. Yet I regard it as the most important piece of data we have. He took eleven patients (I believe) who had terminal illnesses of various kinds, and he obtained permission from them to inject really large, almost massive, doses of soluble plutonium directly into their blood. Now the effect of this is quite different from inhalation of the insoluble oxide, but the surprising thing was that there were no medically discernible effects. Four of these patients with terminal illnesses lived quite a while. I believe two of them are still alive after 25 years, and there are still no discernible effects. So much for the cobra venom theory! With that in mind, it will be an exercise in futility, I think, to try to pick up health effects caused by small amounts of injested plutonium. One reason for this is that the plutonium goes through the GI tract so rapidly that it has very little time to do much damage.

Michael McKay, Los Alamos: I've worked with these correlation coefficients, and I think it is very helpful to the investigator to see plots of the distribution of the response variable over time. If you find that you have to change your data which have a small range of variation, then what your partial correlation coefficients or partial-rank correlation coefficients seem to tell you can be influenced quite drastically. A correlation very close to 1 with an extremely small range of variation—standard deviation, say—might imply that this variable is extremely important but doesn't do a whole lot. Let me go now to something on the magnitudes of the correlation coefficients. I think Ron is the first person that I've seen get excited about a correlation of 0.2 and whether or not it's different from one that's 0.34. I *have kind of a rule of thumb: I just put my hand over* ±0.5. I don't know what you can say about them. We did a few little studies to get some ideas of the distributions, and it seems that if you don't get around 0.5, you can't indicate what's important. If he had showed us plots of the partial correlation coefficients for all of the X variables on a single graph, I think it could have been very helpful. I wonder (1) if you found anything really outstanding (there was a 0.8 that was displayed for us) and (2) whether you found all of the correlation coefficients kind of huddled around zero. Finally I have a question about the ranked correlation vs the correlation with the unranked data. You have very nicely plotted their display for us, and I have also occasionally managed to get nice smooth plots; however, more often I find that the partial-rank correlation coefficients tend to jump up and down. The response variables, which tend to cross over on the time axes, seem to play havoc with the rank correlations, and I think it's reasonable to expect this. I would also ask whether or not we managed to see anything like that in your data. Again this is tied to how much variations you obtained when you did the study.

Ronald Iman: I didn't realize to what degree I was being carried away here with the magnitude of correlation. It must have been a fit of passion! I agree; 0.2 doesn't really excite me that much either, so maybe we can get that cleared up in the record. I do think that one of things that I was concerned about is determining which of these variables are important and what would be the magnitude of their effect. I feel that the effect of these various variables, as I have indicated here working with the raw data, seems to be disguised; the reason for that could be shown by making some sort of plot of the response variable over time to get some idea of the magnitude of the response variable. What it would indicate is that there are some combinations of independent variables that are giving us very high response rates and this is obviously what's making up the correlation; that is, we need some sort of transformation on it.

I might give you plots of all these independent variables.* In my selection of plots here, I tried to go to all the extreme cases that I could find. For the first case, X_1, Y_1, the raw data were indicating that you have raw correlation to partial correlation in the neighborhood of 0.4 to 0.3 whereas the ranks were indicating this was fairly stable at 0.8. Now this is also true for two more of the six variables, so for a total of three variables, the story is different. Raw *indicated reasonably weak and the other indicated* reasonably strong. Note also[†] the plot of X vs the second dependent variable, which is really Y_2; the correlation seems to change considerably over time. I wondered if there was a time effect. The raw data indicate there is some reasonably mild change, but the rank seems to indicate a rather severe change. In putting the response surface at different points of time, I want to know other different variables at different points of time that I need to consider; that is, the role of a particular variable change over time. So if I'm trying to make these predictions at some point in time, exactly what combination of variables do I need? I think that I'm getting different stories from these, which may have been the cause of the enthusiasm. But ranks tell me "positive correlation" and the raw data tell me "negative," and they do have some bit of difference there concerning the number of observations, although it might be difficult to establish that it is a significant difference. I guess the only other point was with respect to the distribution that is assumed on here. I know that Mike has worked with the Latin hypercube technique as much as anyone, and I would appreciate a comment here with respect to distributions, and again I plead for that.

Chuck Bayne, Union Carbide Corporation, Nuclear Division: I would like to disagree with the author a little bit; he pointed out that the variables he had to work with were given to him by the experimenter. It seems to me that the most important contribution that we could make as statisticians is usually arguing the problem with the experimenter. We should not accept that the variables they give us are what we have to work with; I found in my experience that many times the engineer or chemist gives you variables on a basis that this is what has been done in the past. One of the advantages that we have is that we can take a fresh look at these variables and point out where they may not be applicable to the problem. One other comment: the range of the variables seems quite large and quite variable. For example, there are some ranges to 10^2 and also some to 10^6; it seems to me that if you're using quadratic variables where you are squaring or taking the fourth

---

*Ronald L. Iman, these Proceedings, Fig. 5.

†Ronald L. Iman, these Proceedings, Fig. 9.

power, that you may need to do some type of scaling on these variables.

Judy Mahaffey, Pacific Northwest: One of Gary's comments bothered me. He was talking about one massive exposure. Our experimental work concerns animals, not humans certainly. We've given repeated mid-level and low-level exposures which indicate that we get different results than when we give them one single exposure and never repeat it.

Francis Anscombe, Yale: I would like to ask for, perhaps, an item of clarification. I found myself rather puzzled about just what was being done. My understanding is that regression analysis is being done on something that's called data; however, the data related to a thousand years of experience is not observational data at all, I think, but the output of a mathematical model which is predicting what will happen. Now if that's so, I wonder why regression analysis on presumably some simulated output is done, rather than perhaps a direct study of the consequences of the mathematical model. Why regression rather than a direct study of the model itself. I mean, if the mathematical model is reasonably intelligible-looking, I would have thought it would be possible to infer what properties its solutions would have without really generating some random output and then doing a regression analysis which somehow is probably not particularly related to the form of the mathematical analysis. I haven't seen the model, so I don't know. Maybe the models are too complicated for direct theoretical study.

Ronald Iman: The models that we're working with are extremely complicated differential equations all through the place. It isn't something that simply can be identified immediately, so this is the reason for the choice of trying to fit it to a response surface to help to determine what might be important. The first model, the pathways model, is a little easier to get to as far as the choice of the variables. I worked with the individual who developed the model and relied upon his choice of what he felt would be the important variables to consider.

With respect to Chuck's comment that it is my job to work with this individual and indicate to him that I can help him select the variables. First of all this presupposes that I know a lot about geology - which I do not. We're talking about variables such as suspended solids and things of this nature. After long consideration, we did choose those particular variables. Now some of the other models, for example, the transport model that's coming out, are many times more complicated than this pathways model. In fact the pathways model is going to look simple alongside this. As I indicated yesterday, since June, there's been a group of individuals working on that particular model just to tear it down to a usable size. In other words, they are talking in terms of thousands of variables and are trying to get it down to a handful. There isn't any nice formulation.

Michael McKay, Los Alamos: I'd just like to throw my support behind Ron when he asserted the importance of using the correlation coefficients terms, in fact any sensitivity measure, in identifying important time regimes in the model. I think it can be extremely beneficial to the researcher to be able to tell him, "I think variable X_i is important for $t > 20$," for example, and "There doesn't seem to be much for $0 < t < 20$." We've found in reactor safety studies that it has been of benefit.

Dave Gosslee, Union Carbide Corporation, Nuclear Division: I'm not surprised that when you fit this entire model to eight different dependent variables having correlations among the independent variables that you don't get the same best model in some sense by your stepwise procedure. In doing this correlation, could you substitute one variable for another variable that is highly correlated with it to end up with eight or four or some set of regressions that have some consistency.

Robert Easterling, Sandia: It seems to me that this may be another case of when we need to tell our funders that they are asking us to face the wrong problem. My particular problem is with these known distributions. I've encountered in problems like this that there are multiple sources of error, sources of variations, sources of uncertainty, etc. To pretend you can somehow simulate all these and put them together and come out with *a* distribution, such as log-normal or log-uniform, is terribly naive. I've never known a distribution, and I don't expect to know any, yet we're being asked to work these problems as though we do. I think that this fate will be another situation where we should suggest to the people proposing these problems that maybe the problem should be stated differently.

Dave Gosslee, Union Carbide Corporation, Nuclear Division: Let's go back to the other papers or all three of them. I want to say one thing about the first paper. There has been a lot of looking at the persistence of one meteorological variable. I was quite intrigued by looking at the persistence of combinations of extremes of two variables simultaneously. I think this is a very important approach.

Panel Discussion

Panel Discussion

Richard L. Hooper, Moderator
Pacific Northwest Laboratory
Richland, Washington

Carl A. Bennett
Battelle– Human Affairs Research Center
Seattle, Washington

Fred C. Leone
American Statistical Association
Washington, D.C.

Herbert J. C. Kouts
Brookhaven National Laboratory
Upton, New York

John W. Tukey
Princeton University
Princeton, New Jersey

Richard Hooper: Most of the decisions in formulating a national energy policy have been heavily impacted by consideration in areas such as health effects, nuclear reactor safety, and nuclear materials safeguard. Each of these areas is almost a macrocosm in statistics in that each involves measurement problems, data problems, modeling, and evaluation. I'd like to begin the panel discussion with a question — to obtain from the panelists their views on the role of statisticians in these areas and on the kinds of things that can be done to enhance the statistician's role.

Carl Bennett: We've been presented with a very broad subject, and I think that what happens when presented with a subject like this is already apparent from some of the brief discussions we've had together earlier this morning. Each individual tends to tie this role and his response to it to his own role as a statistician and to his own background and experiences (which I think is probably best under these circumstances). I, in particular, have been involved with the energy problem in a somewhat narrow point of view: first of all from the point of view of nuclear energy and more recently from the point of view of safeguards. On the other hand, we all realize that the adequacy of safeguards is one of the more important elements in gaining acceptance of atomic energy or nuclear energy as the power source. Safeguards in a sense are supposed to establish some kind of control — first of all, over the possibility of misuse of nuclear materials, but in addition to helping establish that control, they also need to help establish the credibility of that control or the credibility of the existence of that control. In the brief period I'm supposed to talk to you, I want to look at two examples which deal with the subject from these two viewpoints in terms of what the statistician's role could be. It is rather interesting to compare the term "safeguards" in the international context with the domestic context. The term is used in this country to indicate the first of the rather expensive procedures that must be carried out either under regulatory standards or under contract to maintain control and accountability for nuclear material. In the international context, the term safeguard is concerned almost entirely with (1) the verification and audit of such control to establish credibility in its existence and (2) the fact that material has not escaped such control. These two completely different aspects of the subject, in my opinion, may call for two somewhat different uses of statistics. The first problem, the problem of control, evidently involves statisticians who become heavily oriented with the production of data, with measurement, with the use of those measurements in a control function for the development of indices (such as the rather famous FTIZ), and with the use of these indices to maintain this control. Here I think our chief problem as statisticians is to be sure that we assist in every way we can to make sure that the data are obtained, that the consequences of that data are understood by people, and that new methods are available to deal with new situations. Yesterday I was saying to some people at the discussion session on safeguards that I'm absolutely convinced that the modern real-time control methods which are being used to obtain timeliness under conditions where timeliness is a very important thing in the control of nuclear materials are going to call for some entirely new statistical methods of dealing with data. For example, it is very difficult to define an appropriate combination of timeliness, and the amount of detection which can be used to parameterize the systems we try to develop. It is one of the kinds of things I think statisticians will have to deal with.

The other side of the picture is the side that I mentioned earlier when I was asked, "What do you think is the most important single problem facing safeguards?" and I responded that the single most important way in which statisticians might help is to be able to state some measure of the effectiveness of safeguards that would be credible to the public: first the public as a whole and second as an international body such as the Sarsi group. Then I could turn to somebody from another country and say, "Yes I do believe that such and such can be accomplished if we put so many inspectors in your plant." In other words, I don't have to be in the position of saying, "Gee, why don't you let me send some inspectors into the plant, and after a while I'll tell you whether we are getting any results and whether they're correct or not." I should be able to convince him that he should accept the intrusiveness of inspectors in his plant, because I can convince him of exactly what the effectiveness is going to be. Now that might not be a statistical problem as such, but it certainly depends heavily on the ability of somebody, systems people or statisticians, to quantify what happens because you have an inspector present, what happens because you have a barrier, what happens because of all the other measures that are usually associated with safeguards. I suppose that what I'm really trying to say is that in this area of establishing credibility in the face of uncertainty (if I can use the term), we really have to think of this first in a positive sense — how to act in the face of uncertainty or how to live with uncertainty (or you can create your own phrase). Finally, because I like it so well, I'm going to borrow from a summarization that John made this morning: I think we also need to have statisticians act in the negative sense: that is, they need to minimize the false attractiveness of highly sophisticated procedures. I will give you a specific example of this. I think we have long overdone the attractiveness to many people of material-balance accounting, limits of error, and so forth, simply because they do produce numbers that are, in a sense, quantitative. It is easy for a statistical procedure to be oversold and over-attractive, simply because it is statistical and because it can be, to some extent, quantified.

Herb Kouts: I guess I have the distinction of being the only member of this panel who is not a statistician. I'm a consumer and not a producer of statistics, and I guess I represent that part of the world. I do have some things that might be useful to say on the subject, however, because I do keep running into statistics in a number of applications and a number of things that I do.

First, let me talk about statistics in a broad sense, that is, the broad utility of statistics and statistical methods toward the energy problem and its solution. When I was getting ready to come here, I went to the library and quite naturally got out the *U.S. Statistical Abstract for 1976* and looked through it to see what kind of data base we're operating on in this country, and I found about ten pages of closely crammed information directly related to energy in a section entitled "Energy of the United States." I also found that almost all of the rest of the *U.S. Statistical Abstract* was full of tables which impacted indirectly on the energy problem. There is certainly a great lesson to be gathered here on the importance of statistics to the solution of the energy problem, because you have a data base which is reliable and comprehensive to start with, and even though it may be a fairly humdrum application of statistics to generate such a data base, it's an extremely important application, it seems to me. One gets involved in this kind of data base in connection with all the energy modeling activities that are popular these days and that are really the analytical input to decision-making in the energy field. If the energy modeling activities start with statistics, a framework is begun within which the statisticians begin to interact with this problem. With respect to the question of the Reactor Safety Study, or the Rasmussen report, since Rich Hooper has brought it up, I might say a few words because it's an example, I think, of the kind of analysis in a specific field which does interact and have importance with decisions in the energy system. The Reactor Safety Study was an analysis of risk in the nuclear field risk associated with the production of power by nuclear reactors. I guess I was associated with it in various depths and various forms from the beginning, when the group under Professor Rasmussen at M.I.T. was first put together, until the time the report finally came out; I'm still associated with it because I'm on a panel that is reviewing the report for the Nuclear Regulatory Commission at the request of Congressman Udall. At any rate, the Rasmussen report is an extremely complicated application of relatively simple statistical methods that is, of propagation of probabilities through model-describing modes of failure, or possible failure, in complex nuclear systems. I know most of the people who worked on it; I was in almost daily contact with the work over a long period of time; and I can assure you that this is a very honest piece of work and about the best level of work that competent people of this character could put together, at this time and with the information they had available. It had several characteristics which every application of statistics to a technical or social problem has. First, it had a data base; second, it had a model and calculations based on that model; and finally, it had an error analysis to be attached to it. These are all important aspects of the application. I think that, in each one of these areas in this particular application, there is room for considerable improvement. The data base for the Rasmussen report needs improvement; everybody says so, especially the people who worked on it. The models for the probabilities of failure are pretty good, but the models for consequences need a lot more work, partly as a result of inadequacy of data, because when the data base gets bigger, then your models can get more complicated, but certainly the modeling needs to be improved. Most important of all is the question of error analysis. This area is one in which I think statisticians really have a great deal to develop and to accomplish, because the error analysis of the results of the Rasmussen report are still rather rudimentary compared to the generation of the values they produced themselves.

I should point out that there is another analysis of a simpler kind, commonly used in reactor safety areas, which is just a calculation of what happens if you have a reactor accident. It's the calculation of the effect of what used to be called the maximum hypothetical accident, now called the designed-base accident—a pipe break, let's say—in a Pressurized Water Reactor. One of the most difficult problems facing people who calculate this kind of thing is an estimate of the error attached to the uncertainty in the numbers in the input to a calculation of this kind. I think a great deal can be done by statisticians to help in this very important field because the analysis of all reactor safety these days for every reactor plant goes through a phase where this is the predominate set of questions to be asked.

Fred Leone: I look at this picture from a somewhat different point of view, partly in terms of the individual, but also in terms of the professional organization. I'm a little bit torn because even within our own board of directors there's a question as to whether the ASA should be a learned society or a professional association. We are a professional association, and somehow we should have some impact both as professionals and as an association on some of the decision-making processes. We have been involved in a few things one is relating to the American National Standards Institute and, at the same time, sending a representative to the International Standards Institute. We have attempted to provide a forum. As most of you know, we had a symposium last year and two years before that on the topic of statistics and the environment. This next year the symposium will be extended, and we're broadening it to include energy, toxicology and environment. We're not speaking to statisticians alone, which I think is important. Unfortunately, most of my own contact is with statisticians, but I would like to see us go beyond this limitation. We're getting into legal aspects; one-third of the program committee is made up of members of the bar association whose particular area is along the environmental and energy lines. We have an ad hoc committee that is working with the American Bar Association and with the committee on environmental guidelines. Somehow there's a lot of frustration because we know that, when the final decisions are made, there's a great deal of political input sometimes evasive on what the actual decision should be. As was mentioned earlier by Carl, we've got to develop a stronger and stronger credibility with reference to the data that are collected, with reference to the data base. I think the time is here for us. I think the public is waiting even though they wonder about the credibility, especially when they see two different results from public opinion polling, and in general, they don't realize that these polls possibly start from different bases. We have tried and will continue to try to have an impact on the legislative process, but as we all know it is extremely difficult to influence legislation, especially since the number of bills that are pouring through Congress has gone up by at least a factor of 10, 20, or more as compared to a few years ago. In some instances, we have taken action, though not in the line of energy; hopefully, we will get more and more into this area. One job I feel I have, as I visit a number of chapters, is beating the drum of the individual statistician, getting involved in the different decision making, first at local levels and then beyond that.

A number of years ago we were asked to produce a report on the statistics system of the Bureau of Mines. That report was completed about 1950 or 1951, and I think that was the last heard of it. Two weeks ago I was asked by the assistant director for metal, minerals, etc., if the ASA was interested in reviewing their statistical systems. He had only been aboard for six and one-half years; he didn't know about the other report. I think a lot has happened since 1950, and perhaps the bureau is hearing some of the sounds about the kind of statistical data that's coming out. Next week I will be sitting down with the acting director of the Bureau of Mines, the individual who contacted me, and two or three statisticians. Let me state that I think there is a role that an association can play as a professional, as a group of professionals. I think the voice has to come from individual statisticians demanding that we take greater steps along this area, but I feel that a national office as such cannot do it without the individuals who are the experts, not in statistics alone but also in the subject matter area, to come forth and propose that they themselves will be part of this role, will be part of this impact.

John Tukey: I suspect most of what I'm likely to say really attaches to things that have been said before, but maybe I can sharpen some of the issues a little. At the 50th (or was it the 75th?) anniversary of the Mathematical Association of America, I was asked to speak about what mathematics can do for the government. I refused until there was a small change in title; I said I was willing to talk about what mathematicians could do for the government. Now if people are going to contribute increasingly to decision activities, they have to plan to be increasingly uncomfortable. Usually, you are not going to contribute to decision-making in major and high-level ways (there may be a few exceptions) by doing statistics as defined in the textbooks, and this means that in a certain sense (it's never bothered *me* very much), you won't sleep so easily at night.

I know of one statistician whose name is known to all of you, and I will be careful not to reveal it either directly or indirectly, who some years ago gave up consulting, because he was deeply concerned, honestly concerned, that he may not have given all his clients the best possible advice every time. I think everybody in the room is aware that if you are going to adhere to this standard, you do have to give up consulting.

Well, the requirement for giving up decision-making or contributing to it is much more stringent than that. I'm not convinced that I understand when the technical facts should really determine the decision. It might have been possible to interpret one of the earliest speaker's words to mean that technical facts ought to settle the matter. Clearly, I don't feel this is always true. If I felt it was always true, I wouldn't have any difficulty in making up my mind when it was true. I think there's a very real responsibility on the profession, or on any profession, to see that the technical facts are, at the very least, reasonably available and in some cases, perhaps, vigorously so. Whether those facts are going to determine the difficult social decision is one matter, and whether they ought to is another matter; so one has to be prepared for oneself and the rest of the technical community to be a voice crying in the wilderness, maybe for a year, or ten years, or indefinitely, and we can't afford to let that get us down. I'm saying that just being a statistician, I would recognize some very strong obligations to go well beyond what is taught in statistics today. If you are concerned with an area, you have an obligation to understand about measurement in that area. You may be lucky enough to find some other people who understand and you may not. It is a narrow responsibility of the statistician, I think, to see that the measurement facts get realized, which isn't easy, and not all the measurement facts are going to be statistical.

Let me follow Carl's suggestion and draw on my recent experience in a couple of directions having to do with the chlorofluoro methane in the ozone layer. Generally speaking, the Academy report got pretty good press. However, I saw a copy of an editorial in one paper, and if I remember right, it wasn't too many miles from here, probably considerably less than 2000, which said either the freons are damaging the ozone layer or they're not, and the Academy of Sciences should have told us which. It's a good joke, gets the reaction of people, including more people in our technical organization than is comfortable for any of us to think about. The issue of gradually educating the public, that they are living with uncertainty whether they think so or not, is very important.

Lay Vaughn Newell, who used to work for the EPA, at one stage had an administrator to replace. Vaughn is a medical type but not a statistician, and he expressed his problem very much in medical terms: it's easily demonstrable medically that it's dangerous to get up out of bed in the morning; you get up, and you may have an accident. Also it's easily demonstrable medically that it's dangerous to stay in bed and not take the exercise. If people could understand that this is the way the world is, that they're living their life under these conditions all the time, and that they need to learn to manage the situation, then I think some of our problems with uncertainty would be less serious. In the report on chlorofreons, or chlorofluoro methanes in the ozone layer, we quoted some uncertainties. Important in those uncertainties were certain chemical rate reactions. Estimated uncertainties for those reactions were used, and these estimates were not the internal uncertainties you would discuss if you analyzed the last set of determinations of that rate constant, if there was one. Based pretty much on professional judgment among chemists skilled in kenetics, the uncertainties would likely be between what you had now and where the thing might settle down. I think statisticians have to look forward to doing more of this.

The reason things are rather different than they were a year or so ago the reason that the committee on impacts of stratospheric changes to some degree must go back to its drawing board is that one of the significant reaction rates has changed by a factor of 38; I believe, though I am not sure, that this is a factor of 38 between the best judgment that could be made before any measurement and the first measurement. One is, I think, still somewhat uncertain as to whether there are more important reactions in the mill, and I am sure that there are other reaction rates which are, and not unreasonably, still in that state. This is the best judgment by analogy with the things that you have measured, because reactions, for example, are between two free radicals at stratospheric pressures and temperature in concentrations that may be a few parts per billion. It is not a thing that one feels nice about or a thing that the next new graduate student would go into the laboratory and measure.

There was a conference in Boulder two months ago on the detectability, by direct ground-base measurements, of trends in the total ozone overhead. When the conference was all over, it looked like some unspecified units, which I will just call units because nobody usually does, existed. Out of four possible units, one and one-half came out of the statistical evaluation of the available data, and you had to allow about one and one-half for that. One-half unit was ample to allow for the stations not being uniformly distributed over the face of the globe. Two units came from the instruments being known to have certain kinds of drift in terms of what you know about how they've been calibrated over the years, and this is a trend that you would expect to see over probably one or two decades. You have to leave two units in there for the measurement, and also you have to be very careful to say you know the ozone goes up and down anyway. You have detected a trend, but it would be unwise to say that we are sure that this is the result of what the human race has been doing. I don't think that's an unfair example of where the quantitative technical issues often come out, in a decision base situation, and we have not yet begun to inquire which uses of the chlorofluoro methanes are important to society. If you are going to contribute the most either narrowly or boardly, you've got to spread your individual responsibility well beyond the courses in the university. The further up the line that you're going to have some influence, the further you're going to have to spread that responsibility and the more things you're going to have to say that you may feel uncomfortable about after you've gone to bed at night.

Richard Hooper: Regarding the second question, I need to summarize a couple of things that were said. Carl made several statements about the need to see that credible data is obtained and the consequences evaluated. He talked about the development of evaluation schools that provide credible information to the public; Dr. Kouts made reference to a very complicated model for which the error analysis that went along with it wasn't nearly as sophisticated as the model; Dr. Leone made reference to a stronger credibility, generally in the data and the data base which is used eventually to support decisions. We in the profession should recognize that most of what we end up doing is the result of our own selling efforts—either selling ourselves and our services to investigators in other areas or selling an idea that we think some sponsoring agency would buy. It seems that our recognition of the necessity of salesmanship in itself might be a mechanism to begin turning some of these problems around, and I'm interested in your views on this.

Carl Bennett: I think the real problem comes in salesmanship of certain types, and I don't mean salesmanship in the sense of going out, if you know what I mean, and advertising or anything like that, but the salesmanship that's talked about in, for example, the book by Ralph Cordner. The description is given in Cordner's book, of the guy who ended up developing for, I believe, Sylvania, one of the fundamental items to patent for television. Between 1927 and 1937, he had his research funds cut off perhaps ten times. But this turned out to be one of the single most important patents that Sylvania ever bought. I think what we're getting back to is the kind of persistence that enables you to look at a data base you have investigated, look at a process which you are familiar with, have a certain amount of confidence in your own feelings that you understand what's going on and that you do have some kind of understanding, which it is important for somebody else to have, and then be able to go through the uncomfortable process of educating somebody else to this point of understanding. In a sense, you have to first analyze the data so you're comfortable with it, but then you have to go through the process of making that understanding credible to someone else, and it's establishing that credibility to someone else, in my opinion, that's going to make you a part of a decision process. It is not enough to understand the process yourself, you see. Unless you can use the results of your statistics to influence others, your work might end up like the report Fred was talking about on the shelf, you see. Now, this can be very uncomfortable; you can get laughed at.

I can remember the expressions on the faces of the people in the room when I first suggested to a group of people that they could learn an awful lot if they'd simply run about 50 fuel elements to rupture. My heavens, there wasn't a single person in that room that wasn't going to have the whole problem solved before they had ten ruptures. I can remember sitting in a room with a group of five production people that just laughed at the idea of deliberately loading and running till rupture until we can get some idea of distribution theory and so forth.

That's the sense of uncomfortableness, I think, that we're talking about. It's the sense of being out there on a limb with, first of all, something you believe in because you've done your best to look at all the data you have, and I can't go through this stale process until I really believe in that fact. That belief has to be established first, but then you have the responsibility for selling it and to establish that you have credibility in someone else. There are two different levels of understanding, and I think that again I am indebted to Morton Schubert for this insight. About 15 years ago, he said that there are two completely different levels of understanding of a problem: one is when you suddenly get the insight and understand it yourself, and the other is the appreciably later time when you have it formalized to the point that you can explain it to someone else. These are two significantly different levels of understanding of the problem. I am admittedly a little turned on because I was triggered by, like I usually am, John's remarks with respect to this business of living dangerously, and the fact that one must live dangerously because it so clearly expresses just what a lot of us are not willing to do.

It's hard to describe to a group like this what the situation was like, let me say about 30 years ago, when you could be in the relatively lonely and uncomfortable position of dealing with data out in the plant. You were sort of looked down upon by the group back at the university as though you were prostituting yourself, and you were looked upon as a way-out longhair by the group you were working with. It is very difficult to understand the sheer uncomfortableness of going out and dealing with data in a time when it was far more enticing to retreat to the nice comfort of good, theoretical development or to go over to the comfort of a handbook which told you what do do. It's that kind of thing that you should challenge, if you're going to become part of the decision-making structure.

John Tukey: Of course, I agree with everything that Carl said. I would expand it just a little bit more. Carl was emphasizing "understanding the data" and those who know me, know I think this is extremely important. In most of these situations, it's also important to understand the problem. In another direction is another level of understanding when you not only understand the data, but you understand how people think about the problem, why they think about it that way, and how the data fits into that. It's a different second level, and if you can combine the two separate levels, so that you not only understand the data but you understand the problem and can transfer the conviction of this understanding, then you've gone a long, long way.

Fred Leone: I'd like to emphasize another point in the matter of salesmanship. Too often I have seen cases, especially in analysis of quality control groups, where the individuals are highly regarded as perhaps being good statisticians but not good scientists. In other words, these individuals may be respected for what they can do in their own small circle, but they somehow do not get it across. Now surely we must be willing to be uncomfortable, uncomfortable to the point where we're willing to be patient to the people who do not understand us at first; then we try to reach them. I'm not talking about those we know we cannot reach and the barriers there, for there's no way to gain their respect, at least initially. We need to understand why the data are such, how they are collected, what the implications are, and how to get that understanding across. I think that too many good statisticians are not ready to take the time to sell good, sound, statistical bases for a decision. I don't think we can underestimate this process at all; I think it is very, very important.

John Tukey: I'd like to come back for a moment to the very important process on the other side of the fence. Jack Youden and Frank Wilcoxon were dedicated chemists for 25 years in the same institution, and my estimate is they argued statistics an hour a day for four days a week on the average over that time, but that made two very good statisticians. It's a question of whether five years should have been enough. The related idea (and I think all of you must be familiar with this) is that it's sometimes often accidental that you start establishing a working relationship in some area and that your best relations have often come by the continued growth of something that seemed to start purely accidentally. You don't get to the places where your expertise is most needed at once; you are lucky if you get to most of them in the long run; and how you get there is sometimes very interesting and round about. In dealing with the subject matter areas, I suspect the most important thing to do is to understand the phenomenology. I've always admired Joe Hurstfellow's title "At the Bikini Tests." He wore the title of chief phenomenologist. Too often the technical people you deal with from other disciplines will not have given the phenomenology enough attention. Where they haven't, if the statistician can learn it and use it in the right way, not too obstreperously, maybe that's a way to take a good step forward.

Herb Kouts: Well, there's a lot of talk about pitfalls in this application to problems of the day, and I'd just like to put some of these in a few different words. Both Carl Bennett and John Tukey have talked about the need to define the problem. I think it is necessary to define the problem in very simple ways. If you can write the problem out in one sentence you stand a chance of solving it. If it takes a page to write the problem out, you might as well give it up. In addition to writing the problem out in one sentence, you ought to write it out in such a way that your wife will understand it. If she can understand it, then you have a good chance of solving it. I'm not joking about this!

In the application to social problems of this kind, it is important to recognize that the social problems are the kinds that affect everybody and the effect on other people is expressed in terms of an inability of other people to understand. Of course there are exceptions to all these things; I doubt that some of the things that took place at Bikini could have been put in a form understandable to one's wife. Other applications of bikinis, yes. There's another pitfall that one runs into—deviating from the initial goal. Frequently sicentists start out knowing what they are going to do, and then, by a process which is like drowning in motion, collision with small problems of the day divert them in completely different directions, and they end up trying to solve other problems. This is often called solving the problem you know rather than the problem that needs to be solved. It happens all the time.

Let me just say something else; if you are going to try to solve social problems, you are going to have to do some selling. You are going to have to convince somebody to let you go where you can do some of this. You are going to have to communicate in understandable terms, which means stating what you intend to do in terms of what's to be had. It's a standard problem in salesmanship.

Every salesman has to make it clear that he has exactly what is wanted. In dealing with the government, the same sort of "rules" apply that are appropriate for a butcher dealing with a housewife. The man in government faces problems that he's really able to condense into a little package usually. These are usually the problems I have to solve. If you go to him, telling him you want to do something for him which doesn't fit into what he has to solve, he's not even going to listen to you, he'll throw you out. This disaster often happens even when there is a wonderful match to be made. I was talking yesterday to a fellow at Brookhaven who has some very nice ideas about stress analysis. He has some beautiful capabilities for doing analysis, for instance, of things that might happen to pipes under very difficult circumstances, but what he wanted to do was to develop codes. Well, I'm sure there's no one in government interested in developing the code that he wants to develop, but there are lots of people in government who are interested in doing calculations or seeing the results of calculations that might be done with codes that he might develop. It's up to him to approach people in a manner to convince them that his product might be beneficial to them.

Richard Hooper: Time is moving on. The panel has indicated its willingness to accept questions from the floor.

David Rubinstein, NRC: We've heard several explanations which go sort of in different directions, and I wonder if the panel can give us guidance in relating them. One is that statisticians should remain "sellative." Not that I wish to state the problem simply in one line and not that I suggest we get out of the decision-making process, beyond what might be formal statistical input, but I think once we step beyond that we are likely to lose our credibility.

John Tukey: If I might respond to that last comment, some will and some won't, and we hope that will determine whether they keep on doing it.

Herb Kouts: I'd like to say it is impossible to get into problems that have social impact without losing some of your credibility.

John Tukey: Going back to the first side of that question where it was suggested that the statisticians would retain their credibility, maybe the message from up here was that they should gain their credibility.

Thomas Woteki, Princeton University: I would like to say I found the comments really encouraging. I think if a statistician is going to play a role in the energy problem, then he has to actively and aggressively claim that role for the following reason: whereas the economist is automatically identified with the economy of energy and the biologist is automatically identified with the biological aspects of energy consumption, etc., the statistician is not automatically identified with any particular aspect of the energy problem. In fact, however, he has a professional role in *every* aspect because perhaps every one of those situations involve collecting data and making measurements and decisions. I don't think we're being automatically identified with any aspect, and we have to go out and aggressively claim our role.

I think it is inherently harder for a statistician to claim his role in some of these tasks than it is for a person in another discipline. If a statistician wants to work on the economic aspects of the energy problem, he has to demonstrate first an interest in some knowledge of economics and only then can he bring to bear the fuel that he had. An economist, however, need not demonstrate any knowledge of statistics or energy or anything else; he has an automatic stake in it for some reason. Nobody really knows much about some of these problems and that's very uncomfortable, but on the other hand, that points to the role of a statistician. Since there's a lot of ignorance about these problems, then the statistician is no more ignorant than anyone else and in fact has a responsibility to contribute to the formulation of the problem. Actually I find that much less uncomfortable than being in a situation where someone comes to me and says, "This is the problem, do it!" or even worse than that, "Here is the data; go analyze it!" Though it may be uncomfortable in some cases, I still think you have to go out and do it.

Fred Leone: I'd like to make a point here. I think part of the problem is with the statistician, himself or herself. You were saying the statistician has to learn the economics and then get into the statistics. Unfortunately, when the statistician learns the economics of the problem, then he has to be called an economist. Often that person is called an engineer. I remember a television interview of one of our most famous statisticians in cancer research, and he did not identify himself as a statistician, but as a cancer expert. So the problem is that too many of us are not willing to say "We are statisticians." I would guess also that most of us who are associated with, for example, the American Statistical Association are also part of at least one other field which would be a subject matter field. Nonetheless, we have to stand up and be counted.

Carl Bennett: I must take the opportunity to say that one of the high points of my career was when the Hanford Laboratory was formed; some important person was visiting, and I was being introduced by the then head of the laboratory from each group, and he said, "I would like for you to meet Carl Bennett, our statistician; he's also a pretty fair scientist."

John Tukey: The message from that is if you want to be important, you must arrange to also be a pretty fair scientist.

Gary Tietjen, Los Alamos: Increasingly I see statisticians becoming involved in legal testimony and coming away from these experiences with unpleasant feelings of inadequacy and frustration. What does the panel see as the role of the statistician in those legal situations.

John Tukey: Well I'll answer that, if you are willing, and preface the answer with an anecdote. I was on the U.S. delegation, the technical working group, to the test-ban negotiation; the first day I was a little perplexed about the atmosphere. But before we got to the second meeting I knew what was familiar: it was exactly the atmosphere of a rate case before a public utilities commission, except there wasn't a strong independent person to chair. I think if our society is going to function well, then interest groups are just as entitled to have quantitative council, as they are to have legal council. We cannot expect that the quantitative council for opposing points of view are going to make entirely compatible statements, but we can expect each of them will keep the other quantitative council honest. Many of you may not know of the case that I mentioned at breakfast, in which Chester Blaizerfield appeared for one party, and Judge Netterfield appeared for the other. One did the analysis model 1 and got one answer, and the other did the analysis of variance model 2 and got the other answer. Now I think it was for the good that this possibility of disagreement and adequate quantitative representation took place. If you are going to appear as quantitative council, that's not the same thing as sitting in the neutral lab; there's no use trying to make it look as if it is.

Carl Bennett: Having spent the last six years over on the other side of the mountain with this Human Affairs Research Centers group (which consists primarily of psychologists, sociologists, lawyers, economists, and a certain number of public affairs type people), I've greatly enjoyed discussing with some of my "legal friends" the difference between the adversary type of proceedings that constitute typical legal procedure and the kinds of procedures we are more accustomed to, where you supposedly analyze the data and come out with the answer. Really in a sense there is a great deal more in common here than you might think. My lawyer friends claim that the adversary proceedings really only start after admission into evidence of a certain set of factual data which both lawyers essentially agree exist. That is, there is a process in law of agreeing on a brief or a body of evidence or data which is taken as fact, and the adversary argument starts from there, in the sense of being interpretation of this evidence, or arguments of how this evidence is interpreted. In a sense we have almost the same procedure here. You can start with a given body of data, and I think it perfectly possible to have two different people agree or disagree on the interpretation of that set of data and what is meaningful to conclude them. To me I think it is worth thinking about.

John Tukey: It seems to me we go a lot further than saying that the statistician professionally internalize this. The motion of the confidence interval is perfectly equivalent to saying that (1) a representative is arguing for each value of the parameter (a different representative for each value) now which ones of those can you not rule out? and (2) the adversary is in a sense supplied, but the successful adversaries are represented by the upper and lower circumferences. Let me also not leave the impression that I think an adversary's position is always best. I've always thought that one of the better compliments I ever had, which I eventually heard secondhand, was in an international discussion where somebody on the other side said, "Well I never understood why you had Tukey along, but now I realize that he's being absolutely as objective as one can get." So, there are times and places; you need to know which is your role and try to fill it.

Richard Hooper: We started out to talk about the roles of the statistician and decisions needed and formulated in national energy policy; we talked at length about the need to understand the problems to the point that we can articulate them in a meaningful way to consumers; we talked at some length about the need to make the application of complicated procedures more credible through the kinds of error analyses and the statements that are associated with them. It seems to me that, time and time again, the conversations turn back to Professor Tukey's statement, which will really have an impact on decisions and enhance and increase our credibility as a profession: "We have to be willing to be uncomfortable."

LIST OF ATTENDEES

Anscombe, Francis J.
Yale University
New Haven, Conn.

Ard, Everett E.
Sandia Laboratories
Albuquerque, N.Mex.

Arnett, L. M.
E. I. Du Pont de Nemours & Co.
Savannah River Laboratory
Aiken, S.C.

Atwood, Corwin L.
EG&G Idaho
Idaho Falls, Idaho

Axelrod, Michael
Lawrence Livermore Laboratory
Livermore, Calif.

Barker, Lawrence K.
U.S.G.S. Conservation Division
Grand Junction, Colo.

Barnes, Madaline
Desert Research Institute
Las Vegas, Nev.

Bayne, Charles K.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Beauchamp, John J.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Bement, Thomas R.
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

Bennett, Carl A.
Battelle - Human Affairs Research Center
Seattle, Wash.

Bloomfield, Peter
Princeton University
Princeton, N.J.

Bowman, K. O.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Bracey, Jere T.
New Brunswick Laboratory
Argonne, Ill.

Buschbom, Ray L.
Pacific Northwest Laboratories
Richland, Wash.

Carr, Daniel B.
Pacific Northwest Laboratories
Richland, Wash.

Choi, John C.
New Brunswick Laboratory
Argonne, Ill.

Conover, William
Texas Tech University
Lubbock, Tex.

Delfiner, Pierre
School of Mines
Center for Geostatistics
Fontainebleau, France

DeVary, J. L.
Westinghouse Hanford
Richland, Wash.

Dietz, Donald R.
U.S. Fish and Wildlife
Grand Junction, Colo.

Doctor, Pamela G.
Pacific Northwest Laboratories
Richland, Wash.

Dutt, Dale
Westinghouse Hanford
Richland, Wash.

Easterling, Robert G.
Sandia Laboratories
Albuquerque, N.Mex.

Friedman, Jerome H.
Stanford Linear Accelerator Center
Stanford, Calif.

Furner, George
Rockwell International — Hanford
Richland, Wash.

Gardiner, Donald A.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Gaver, Donald P.
Naval Postgraduate School
Monterey, Calif.

Gilbert, Ethel S.
Pacific Northwest Laboratories
Richland, Wash.

Gilbert, Richard O.
Pacific Northwest Laboratories
Richland, Wash.

Godbold, James H., Jr.
Oak Ridge Associated Universities
Oak Ridge, Tenn.

Goldstein, Rubin
Combustion Engineering
Windsor, Conn.

Gosslee, David G.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Griffith, William G.
Inhalation Toxicology Research Inst.
Albuquerque, N.Mex.

Gluckman, Perry
Lawrence Livermore Laboratory
Livermore, Calif.

Hall, Irving J.
Sandia Laboratories
Albuquerque, N.Mex.

Hanlen, R. C.
United Nuclear Industries Inc.
Richland, Wash.

Halverson, Galen D.
Allied Chemical Corp.
Idaho Falls, Idaho

Hebble, Thomas
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Hoffman, Eric G.
U.S.G.S. Conservation Division
Grand Junction, Colo.

Hooper, Richard L.
Pacific Northwest Laboratories
Richland, Wash.

Hume, Merril W.
Rockwell International — Rocky Flats
Golden, Colo.

Iman, Ronald L.
Sandia Laboratories
Albuquerque, N.Mex.

Jaech, John L.
Exxon Nuclear Co., Inc.
Richland, Wash.

Johnston, James W.
Pacific Northwest Laboratories
Richland, Wash.

Johnson, Myrle M.
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

Jost, James W.
Westinghouse Hanford
Richland, Wash.

Kao, Samuel C.
Brookhaven National Laboratory
Upton, N.Y.

Kleinknecht, Richard E.
Pacific Northwest Laboratories
Richland, Wash.

Kraft, Tony
Exxon Nuclear Co., Inc.
Richland, Wash.

Kouts, Herb
Brookhaven National Laboratory
Upton, N.Y.

Lechner, James A.
National Bureau of Standards
Washington, D.C.

Leone, Fred C.
American Statistical Association
Washington, D.C.

Lever, William E.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Lohrding, Ronald K.
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

Lookabaugh, Janice
Rockwell International Hanford
Richland, Wash.

Mahaffey, Judy A.
Pacific Northwest Laboratories
Richland, Wash.

Marcus, Allan H.
Washington State University
Pullman, Wash.

Marks, Sidney
Pacific Northwest Laboratories
Richland, Wash.

Mayer, Lawrence S.
Princeton University
Princeton, N.J.

McCammon, Richard
U.S. Geological Survey
Reston, Va.

McKay, Michael
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

McRae, Larry
Rockwell International – Hanford
Richland, Wash.

Mensing, Richard W.
Lawrence Livermore Laboratory
Livermore, Calif.

Meteer, Jim
Allied Chemical Corp.
Idaho Falls, Idaho

Merrill, James A.
Pacific Northwest Laboratories
Richland, Wash.

Miller, Forest L., Jr.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Moses, Lincoln
Stanford University
Stanford, Calif.

Murphy, Chester
Rockwell International Hanford
Richland, Wash.

Nelson, David L.
Boeing Computer Services Inc.
Seattle, Wash.

Nicholson, Wesley L.
Pacific Northwest Laboratories
Richland, Wash.

Olsen, Anthony R.
Pacific Northwest Laboratories
Richland, Wash.

Owen, Pat
Westinghouse Hanford
Richland, Wash.

Popp, Sharon M.
Pacific Northwest Laboratories
Richland, Wash.

Prairie, R. R.
Sandia Laboratories
Albuquerque, N.Mex.

Robertson, Bill
Westinghouse Hanford
Richland, Wash.

Robinson, Peter M.
Harvard University
Cambridge, Mass.

Rubinstein, David
Nuclear Regulatory Commission
Washington, D.C.

Rubinstein, Sol
Rockwell International Hanford
Richland, Wash.

Shepard, Donald F.
Rockwell International Hanford
Richland, Wash.

Shimamoto, Tetsuo
Bettis Atomic Power Laboratory
West Mifflin, Pa.

Smiriga, Nora G.
Lawrence Livermore Laboratory
Livermore, Calif.

Stevens, Donald L., Jr.
Pacific Northwest Laboratories
Richland, Wash.

Suda, Sylvester
Brookhaven National Laboratory
Upton, N.Y.

Tiahrt, Ken
Montana State University
Bozeman, Mont.

Thisted, Ronald A.
University of Chicago
Chicago, Ill.

Thomas, John M.
Pacific Northwest Laboratories
Richland, Wash.

Tietjen, Gary L.
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

Truett, Tykey
Oak Ridge National Laboratory
Oak Ridge, Tenn.

Tukey, John W.
Princeton University
Princeton, N.J.

Tyler, Sylvanus A.
Argonne National Laboratory
Argonne, Ill.

Uppuluri, V. R. R.
Union Carbide Corp., Nuclear Division
Oak Ridge, Tenn.

Waller, Ray A.
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

Waterman, Michael S.
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

Wickramaratne, Priya J.
Raytheon Company
Portsmouth, R.I.

Wincek, Michael A.
Pacific Northwest Laboratories
Richland, Wash.

Woteki, Thomas
Princeton University
Princeton, N.J.

Ziegler, R. Keith
Los Alamos Scientific Laboratory
Los Alamos, N.Mex.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1978-748-189/469