NUREG/CP-0053
LA-10127-C
Conference



# *Proceedings of the Ninth Annual Statistics Symposium on National Energy Issues, October 19-21, 1983*

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

These papers are reproduced as submitted by the authors.

NUREG/CP-0053
LA-10127-C
Conference

AN

# Proceedings of the Ninth Annual Statistics Symposium on National Energy Issues, October 19-21, 1983

Compiled by
Maurice C. Bryson

Manuscript submitted: May 1984
Date published: June 1984

Prepared for
Division of Risk Analysis
Office of Nuclear Regulatory Research
US Nuclear Regulatory Commission
Washington, DC 20555

NRC FIN No. A7225

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

CONTENTS

# PROCEEDINGS OF THE NINTH ANNUAL STATISTICS SYMPOSIUM ON NATIONAL ENERGY ISSUES, OCTOBER 19-21, 1983

Compiled by

Maurice C. Bryson

ABSTRACT

The Ninth Annual Statistics Symposium on National Energy Issues was held in Rockville, Maryland, at the Holiday Inn Crowne Plaza, October 19-21, under the joint sponsorship of Los Alamos National Laboratory and the Nuclear Regulatory Commission. Sessions included two contributed-paper sessions, two tutorial sessions, and one discussion group. Included in these proceedings are those papers for which final copy was provided by the authors, together with a list of papers presented and a list of attendees.

# INTRODUCTION

The Ninth Annual Statistics Symposium on National Energy Issues (formerly, DOE Statistics Symposium) was held in Rockville, MD at the Holiday Inn Crowne Plaza on October 19-21, 1983. Co-hosts for the symposium were the Los Alamos National Laboratory and the Nuclear Regulatory Commission. The symposium has been held annually since 1975 under the auspices of a Steering Committee with membership drawn from the several national laboratories of the Department of Energy. Members of the 1983 Steering Committee were:

Maurice Bryson, Los Alamos National Laboratory - chairman
David Gosslee, Oak Ridge National Laboratory
Ronald Iman, Sandia National Laboratories
Samuel Kao, Brookhaven National Laboratory
David Margolies, Lawrence Livermore National Laboratory
Donald Stevens, Pacific Northwest Laboratory.

The 1983 symposium included both contributed-paper sessions and tutorial sessions with invited participants, as well as an invited-participant panel discussion. Because of the informal nature of some sessions, several participants did not feel that it was appropriate to submit papers or remarks for these proceedings. A comprehensive list of papers is included here, identifying those that are included either in full or in abstract form. Contributed papers were selected from abstracts submitted to a Program Committee, whose membership included the following:

Lawrence Bruckner, Los Alamos National Laboratory - chairman
Cory Atwood, EG&G - Idaho
Daniel Carr, Pacific Northwest Laboratory
Ronald Glaser, Lawrence Livermore National Laboratory
Irving Hall, Sandia National Laboratories
Max Morris, Oak Ridge National Laboratory.

Local arrangements assistance was provided by Dale Rasmuson, Nuclear Regulatory Commission.

Fifty-one persons attended the symposium. A complete list of attendees is included at the end of the Proceedings.

## SUMMARY OF PAPERS PRESENTED

(*) - Denotes paper published in proceedings.
(A) - Denotes abstract only published in proceedings.

KEYNOTE ADDRESS

Victor Stello, Deputy Executive Director for Regulatory Operations, Nuclear Regulatory Commission.

CONTRIBUTED PAPERS (SESSION 1)

(*) "The Estimation of the Hazard Function from Randomly Censored Data by the Kernel Method," by M. A. Tanner and W. H. Wong.

(*) "Combining the Econometric and Time Series Forecasts of Total Product Supplied," by C. Baghelai.

(*) "Statistical Prediction of Instantaneous Residential Electric Demand," by D. J. Anderson, G. J. Collaros, and E. W. Enlow.

GUEST ADDDRESS, LUNCHEON

Roger H. Moore, Senior Mathematical Advisor, U. S. Bureau of the Census.

TUTORIAL SESSION 1: UNCERTAINTY PROPAGATION IN RISK ASSESSMENT

(*) "PRA Uncertainties and the Roles of Sensitivity and Uncertainty Analyses," by W. E. Vesely and D. M. Rasmuson.

"Role of Bayesian Statistical Methods in PRA Uncertainty Analysis," by J. W. Johnson and N. D. Singpurwalla.

(*) "Uncertainty Analysis: Good News and Bad News," by R. D. Beckman and D. E. Whiteman.

(A) "Recent Developments in Sensitivity Analysis," by R. D. Iman.

CONTRIBUTED PAPERS (SESSION 2)

(A) "Toxicokinetics and Risk Assessment," by A. Silvers, P. Ricci, and R. Wyzga.

(A) "Statistical Methods Used in Developing an Optimum Glass for Vitrification and Storage of Nuclear Wastes," by W. M. Bowen and L. A. Chick.

(*) "Probabilistic Fracture Mechanics: Statistical, Mathematical, and Modeling Issues," by D. C. Cox and R. E. Kurth.

(*) "Trend and Pattern Analysis of Operating Reactor Data from Licensee Event Reports," by R. L. Dennig, O. V. Hester, and C. D. Gentillon.

PANEL DISCUSSION: GOVERNMENT AND STATISTICS: THE NRC/ASA EXPERIENCE

C. A. Bennett, Moderator
R. M. Bernero, Panelist
S. H. Bush, Panelist
(*) B. Harris, Panelist
J. R. Rosenblatt, Panelist

TUTORIAL SESSION 2: ACCEPTABLE RISK AND SAFETY GOALS

R. W. Mensing, Moderator
(*) M. L. Ernst, Speaker
V. T. Covello, Speaker
D. MacLean, Speaker
E. Anderson, Speaker

# THE ESTIMATION OF THE HAZARD FUNCTION FROM RANDOMLY CENSORED DATA BY THE KERNEL METHOD

by

Martin A. Tanner and Wing Hung Wong
University of Wisconsin
1210 W. Dayton Street
Madison, WI 53706

(Contributed Paper Session 1)

# THE ESTIMATION OF THE HAZARD FUNCTION FROM RANDOMLY CENSORED DATA BY THE KERNEL METHOD

by

Martin A. Tanner and Wing Hung Wong

SUMMARY

We obtain a direct kernel estimate of the hazard function from censored data by convolution smoothing of the empirical hazard. By exploiting a conditional independence property (the lemma in Section 2) of the random censorship model, it is possible to obtain expressions for bias and variance in small samples. Under regularity conditions, the bias is asymptotically equal to the error resulting from approximating the death hazard function by its convolution with the kernel. The estimator is shown to be mean square consistent as the window size shrinks to zero. Asymptotic normality is also established, the proof of which is based on Hajek's projection approach. The understanding of these properties is prerequisite to the understanding of more complicated procedures based on kernels. The results also give insights to bias correction procedures as well as to the difference between hazard estimation and density estimation.

## 1. Introduction

### 1.1 The problem

In lifetesting, medical follow up, and other studies, the observation of the occurrence of the event of interest (called a failure, or a death) may be prevented for some of the items of the sample by the previous occurrence of some other event (called a loss, or a censoring event). Thus, if $T_1,\ldots,T_n$ are life times (time to failure) for the n items under study and $C_1,\ldots,C_n$ the corresponding censoring times, then it is not possible to observe both $T_i$ and $C_i$. Instead, we can only observe $X_i$ and $\delta_i$ where $X_i = \min(T_i,C_i)$ and $\delta_i = I_{[T_i \leq C_i]}$.

In this paper, $T_1,\ldots,T_n$ are assumed to be i.i.d. from a life distribution $F_T$. We are interested in estimating

(a) the survival function: $S_T(t) = 1 - F_T(t)$
(b) the density function: $f_T(t) = dF_T(t)/dt = -dS_T(t)/dt$
(c) the failure rate (or hazard) function:

$$\lambda_T(t) = f_T(t)/S_T(t) = -d \log S_T(t)/dt.$$

AMS Subject Classification: Primary 62G05, 62P10; Secondary 62E20, 65D10

Key words and phrases: censored data, hazard, survival, kernel method, Hajek's projection method

Theoretically, knowing any one of these functions, we can easily obtain the other two. In practice, when these functions have to be estimated, it is not always possible to directly convert the estimate of one function to estimates of the other two. In this paper, we will focus on the estimation of the failure rate function.

1.2 Kaplan-Meier estimate of survival function

Let $X_{(1)},\ldots X_{(n)}$ be the order statistics of the X's, and $\delta_{(1)},\ldots,\delta_{(n)}$ the corresponding $\delta$'s. The Kaplan-Meier estimator (Kaplan and Meier, 1958) for $S_T$ is

$$\hat{S}^*(x) = \begin{cases} \prod_{j=1}^{i} \frac{n-j}{n-j+1}^{\delta_{(j)}} & \text{if } x < X_{(1)} \\ & \text{if } X_{(i)} \leq x < X_{(i+1)} \\ 0 & \text{if } x \geq X_{(n)} \cdot \end{cases}$$

This estimator is applicable if the lifetime $T_i$ can be assumed to be independent of the (potential) censoring time $C_i$.

1.3 The hazard estimates

$-\log \hat{S}^*(x)$ is a step function with

$$\text{jump at } X_{(i)} = \begin{cases} -\delta_{(i)} \log\left(\frac{n-i}{n-i+1}\right) & i = 1,\ldots,n-1 \\ \infty & i = n \quad . \end{cases}$$

An ad hoc way to remove the awkward infinity is to always treat the last observation as if it were censored, i.e., the survival estimate has no jump at $X_{(n)}$. If this ad hoc convention is adopted, then formal differention of $-\log S^*$ produces the (formal hazard estimate)

$$\sum_{i=1}^{n} -\delta_{(i)} \log\left(1 - \frac{1}{n-i+1}\right)\delta(x - X_{(i)})$$

where $\delta(x - X_{(i)})$ refers to the Dirac Delta function. The infinite spikes at the data points are obviously undesirable. One way to smooth them out is to convolute with a peaking kernel $K_h(x) = \frac{1}{h} K(\frac{x}{h})$, yielding the hazard estimate

$$\hat{\lambda}(x) = \sum_{i=1}^{n} - \delta_{(i)} \log\left(1 - \frac{1}{n-i+1}\right) \frac{1}{h} K\left(\frac{x - X_{(i)}}{h}\right) \quad .$$

Another natural way to obtain hazard estimate is to consider the estimate of the cumulative hazard:

$$\hat{H}(x) = \sum_{X(i) \leq x} a_i \text{ where}$$

$a_i$ = contribution of $X_{(i)}$ to the cumulative hazard

$$= \delta_{(i)} \text{ (1/No. of items at risk)} = \delta_{(i)}(1/n-i+1) \quad .$$

The corresponding hazard function is

$$\sum_{i=1}^{n} \frac{1}{n-i+1} \delta_{(i)} \quad \delta(x - X_{(i)}) \quad .$$

Smoothing this, we obtain the hazard estimate

$$\hat{\lambda}(x) = \sum_{i=1}^{n} \frac{1}{n-i+1} \delta_{(i)} \frac{1}{h} K\left(\frac{x - X_{(i)}}{h}\right) \quad .$$

If there is no censoring, this reduces to the second estimator in Watson and Leadbetter (1964a). It is not hard to see that if $n \to \infty$ and $h \to 0$, then the estimates $\lambda(x)$ and $\hat{\lambda}(x)$ will be equivalent in the limit. Thus the asymptotic properties of the two estimators are similar. In small samples, $\lambda$ is analytically more tractable than $\hat{\lambda}$, and in this paper we will only investigate properties of the former. Rice and Rosenblatt (1976) give asymptotic bounds for the difference between $\lambda$ and $\hat{\lambda}$ in the uncensored case. Figure 1 presents $\lambda$ and $\hat{\lambda}$ obtained from 200 pseudo-random variates in the case of exponential death times and exponential censoring. The two estimates are seen to be quite similar. In fact, Monte Carlo results indicate that they are quite similar for a sample size as small as 20.

1.4 The random censorship assumption

The theoretical properties of the hazard estimate $\hat{\lambda}$ will be developed in this paper under the assumption of random censorship which stipulates that the censoring times $C_1,\ldots,C_n$ also form a random sample, independent of the life times. $F_C$ will denote the common distribution function of $C_1,\ldots,C_n$, and $S_C$, $f_C$, $\lambda_C$, the corresponding survival, density, and hazard functions. In Sections 2 and 4, $\lambda$ is shown to be consistent and asymptotically normal under the random censorship assumption. Although we have not obtained concrete theoretical results outside the random censorship model, we think that $\lambda$ (and its modifications) is also relevant for other types of censoring schemes.

## 1.5 Relation to other work

The Kaplan-Meier estimate for survival function is first discussed in Kaplan and Meier (1958). Meier (1967) emphasizes that this function plays a role in the censored situation similar to the empirical function in the uncensored case. Recently Foldes, Rejto, and Winter (1981) propose estimating the density from censored data by $\Sigma w_i K_h(x-X_i)$, where the $w_i$'s are the jumps in the Kaplan-Meier curve. This is a direct generalization of the usual kernel density estimate in the uncensored case when $w_i = \frac{1}{n}$ is the jump of the empirical cdf. A smooth survival estimate can be obtained by integration and the hazard estimate is obtained by forming the ratio of density and survival estimates. Foldes et al. established strong consistency under certain assumptions. No finite sample results are presented. McNichols and Padgett (1981) examine the same kernel density estimate under a proportional hazard assumption and give expressions for expected value and variance in small samples.

Alternatively, one can use the kernel method directly on the hazard scale, using the weights provided by the jumps of the empirical estimate of the cumulative hazard function given in Nelson (1972) and Aalen (1978). This is the approach taken in this paper.

In the uncensored case, the estimation of the hazard function is considered in detail in Watson and Leadbetter (1964a,b). They examine both approaches, that is, estimation of the hazard function via density estimation, or directly by smoothing empirical hazards. As is clear in their paper, the second approach is analytically more tractible, and they are able to give finite sample bias and variance expressions and conditions for consistency. Our results in Section 2 and Section 3 are generalizations of their results to the censored case. The calculations in Section 2 are conditional and they depend on the crucial observation that $E(\delta_{(j)}|X_{(j)})$ is independent of j. As for the consistency, the direct generalization of Watson and Leadbetter's argument requires conditions which are dependent on the censoring distribution and this dependency may be undesirable. We have therefore presented alternative conditions which avoid this difficulty. More detailed asymptotic results on hazard estimation in the uncensored case are given in Rice and Rosenblatt (1976).

The results in Section 4 concerning asymptotic normality have no analog in the literature. Even in the uncensored case, Watson and Leadbetter cannot obtain asymptotic normality for the direct kernel hazard estimate (the second approach). Our result in Section 4 is a novel application of the projection method (Hajek, 1968) which has traditionally been applied to the asymptotic theory of U-statistics, linear combination of order statistics, and linear rank statistics. It is hoped that our calculations may be useful for the asymptotic theory of sums where each term involves both the rank and magnitude of an individual sample point.

In practice, different degrees of smoothing are needed in different regions. Moreover, it seems natural to use the observations to determine the degree of smoothing. Section 5 contains further comments on these methodological developments, numerical examples, as well as remarks on the significance of the present study.

Finally, we cite works of general interest for the nonparametric estimation of density or hazard functions from censored data. Kimura (1972), Tarter (1979) investigated the Fourier series expansion approach. Bartoszynski, Brown, McBride, and Thompson (1981) apply penalized maximum likelihood methods. Lo (1980) proposes Bayesian nonparametric kernel methods. Finally, the work on piecewise smooth survival function estimation, Friedman (1978), can be regarded as giving piecewise smooth density estimates.

## 2. Calculation of mean and variance

In this and the next two sections we assume the random censorship model:

$$T_1, T_2, \ldots, T_n \sim \text{iid } T_T, \text{ independently of}$$

$$C_1, C_2, \ldots, C_n \sim \text{iid } F_C \ .$$

We observe $X_i = \min(T_i, C_i)$, $\delta_i = I_{[T_i \leq C_i]}$. We will investigate the properties of

$$\hat{\lambda}(x) = \sum_{j=1}^{n} \frac{1}{n-j+1} \delta_{(j)} K_h(x - X_{(j)})$$

as an estimate of $\lambda_T(x)$, where K is a symmetric kernel, $K_h(y) = \frac{1}{h} K(\frac{y}{h})$. The point of interest x is fixed throughout the study, and assumed to satisfy the condition $S_T(x)S_C(x) > 0$. (Without this condition the problem is either trivial or impossible.) We also assume that $f_T$ is a continuous and bounded function.

### 2.1 The mean

The calculation of the mean of $\hat{\lambda}(x)$ requires a conditional argument which is made possible by the following lemma. The lemma says that given the value of $X_{(j)}$, the probability that this corresponds to a censored observation is independent of the ordering j. This result may be surprising at first sight because the unconditional probability of the above event certainly depends on j.

Lemma: $$E(\delta_{(j)} | X_{(j)} = y) = \frac{f_T(y)S_C(y)}{f_X(y)} \overset{\text{def}}{=\!=} m(y)$$

independent of the value of j.

Proof: We need to show, for any Borel set A, that

$$\int_A \frac{f_T(y)S_C(y)}{f_X(y)} \frac{n!}{(j-1)!(n-j)!} F_X^{j-1}(y)(1-F_X(y))^{n-j} f_X(y)dy = \int_{X_{(j)} \in A} \delta_{(j)} dP \ .$$

Now, the right hand side

$$= \sum_{i=1}^{n} P(T_i < C_i,\ \text{rank}(X_i) = j,\ X_i \in A)$$

$$= n\, P(T_1 < C_1,\ \text{rank}(X_1) = j,\ X_1 \in A)$$

$$= n \int_A P(T_1 < C_1,\ \text{rank}(X_1) = j,\ X_1 = y)dy$$

$$= n \int_A P(T_1 < C_1,\ X_1 = y,\ j-1 \text{ out of } X_2,\ldots,X_n < y)dy$$

$$= n \int_A P(T_1 < C_1,\ X_1 = y)P(j-1 \text{ out of } X_2,\ldots,X_n < y)dy$$

$$= n \int_A f_T(y)S_C(y) \binom{n-1}{j-1} F_X^{j-1}(y)(1 - F_X(y))^{n-j}dy$$

as required.

With the help of this lemma, the calculation of the mean is straightforward.

$$(2.1.1)\quad \hat{E}\lambda(x) = \int \sum_{j=1}^{n} E(\delta_{(j)} | X_{(j)} = y) \frac{1}{n-j+1} f_{X_{(j)}}(y)K_h(x - y)dy$$

$$= \int \Big[\sum_{j=1}^{n} \frac{1}{n-1+1} \frac{n!}{(j-1)!(n-j)!} F_X^{j-1}(y)(1-F_X(y)^{n-j})\Big] f_X(y)m(y)K_h(x-y)dy$$

$$= \int (1 - F_X^n(y)) \frac{f_X(y)}{1-F_X(y)} m(y)K_h(x - y)dy$$

$$= \int (1 - F_X^n(y))\lambda_T(y)K_h(x - y)dy$$

Hence

$$\text{Bias} = [\int \lambda_T(y)K_h(x - y)dy - \lambda_T(x)] + [-\int F_X^n(y)\lambda_T(y)K_h(x - y)dy] \ .$$

The first component of the bias is the error of approximating $\lambda_T(x)$ by the convolution $\lambda_T * K_h(x)$, this can be made small by decreasing h. The second term will be seen to go to zero as n increases. This bias expression is important for the purpose of bias correction, see Section 5.2.

2.2 The variance

$$E(\hat{\lambda}(x))^2 = E\Big(\sum_{j=1}^{n} \frac{1}{(n-j+1)^2}\,\delta^2_{(j)} K_h^2(x - X_{(j)})\Big)$$

$$+ 2E\Big(\sum_{r}\sum_{s>r} \frac{1}{n-r+1}\,\frac{1}{n-s+1}\,\delta_{(r)}\delta_{(s)} K_h(x - X_{(r)}) K_h(x - X_{(s)})\Big)$$

$$= (I) + (II) \quad .$$

$$(I) = \int\Big[\sum_{j=1}^{n} \frac{1}{(n-j+1)^2}\, f_{X_{(j)}}(y) E(\delta_{(j)} | X_{(j)} = y)\Big]\, K_h^2(x - y)dy$$

$$= \int\Big[\sum_{j=1}^{n} \frac{1}{(n-j+1)^2}\,\frac{n!}{(j-1)!(n-j)!}\, F_X(y)^{j-1}(1 - F_X(y))^{n-j}\Big] f_X(y) m(y) K_h^2(x - y)dy$$

$$= \int\Big[\sum_{k=0}^{n-1} \frac{1}{(n-k)}\binom{n}{k} F_X(y)^k (1 - F_X(y))^{n-k}\Big]\,\frac{f_X(y)}{1-F_X(y)}\,\frac{f_T(y)S_C(y)}{f_X(y)}\, K_h^2(x - y)dy$$

$$= \int I_n(F_X(y))\lambda_T(y) K_h^2(x - y)dy$$

where

$$I_n(F) \overset{\text{def}}{\equiv\equiv} \sum_{k=0}^{n-1} \frac{1}{n-k}\binom{n}{k} F^k(1 - F)^{n-k} = \int_0^{1-F} \frac{(F+x)^n - F^n}{x}\, dx,$$

a notation introduced by Watson and Leadbetter (1964b). To calculate (II), first note that for $r < s$, $y < z$,

$$E\left(\delta_{(r)}\delta_{(s)} | X_{(r)} = y,\ X_{(s)} = z\right) = P\left(\delta_{(r)} = 1,\ \delta_{(s)} = 1 | X_{(r)} = y,\ X_{(s)} = z\right)$$

$$= P(\delta_1 = 1,\ \delta_2 = 1 | X_1 = y,\ X_2 = z)$$

$$= P(\delta_1 = 1 | X_1 = y) P(\delta_2 = 1 | X_2 = z) = m(y)m(z),$$

by the same kind of argument used in establishing that $E(\delta_{(j)} | X_{(j)} = y)$ is independent of j. Thus, letting

$$t(y,z) = \sum_{r}\sum_{s>r} \frac{1}{n-r+1}\,\frac{1}{n-s+1}\; f_{X_{(r)},X_{(s)}}(y,z)$$

for $y < z$, we have

$$(II) = 2 \int_{y<z} t(y,z)m(y)m(z)K_h(x - y)K_h(x - z)dydz$$

$$t(y,z) = \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \frac{1}{n-r+1} \frac{1}{n-s+1} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F_X^{r-1}(y)f_X(y)[F_X(z)-F_X(y)]^{s-r+1}$$

$$\cdot f_X(z)(1 - F_X(z))^{n-s}$$

$$= \frac{f_X(z)}{1-F_X(z)} \cdot \frac{f_X(y)}{1-F_X(y)} \{ 1 - F_X^n(y) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} [F_X^n(z) - F_X^n(y)] \} .$$

Thus

$$(II) = 2 \int_{y<z} \{1 - F_X^n(y) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} [F_X^n(z) - F_X^n(y)] \} \lambda_T(y)\lambda_T(z)$$

$$\cdot K_h(x - y)K_h(x - z)dydz$$

$$= [ \int \lambda_T(y)K_h(x - y)dy]^2 - 2 \int_{y<z} \{F_X^n(y) + \frac{1-F_X(y)}{F_X(z)-F_X(y)} [F_X^n(z) - F_X^n(y)]\}$$

$$\cdot \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)dydz .$$

Now

$$(E\hat{\lambda}(x))^2 = [\int \lambda_T(y)K_h(x - y)dy)^2$$

$$- 2 (\int \lambda_T(y)K_h(x - y)dy)(\int F_X^n(y)\lambda_T(y)K_h(x - y)dy)$$

$$+ [\int F_X^n(y)\lambda_T(y)K_h(x - y)dy]^2$$

$$[\int \lambda_T(y)K_h(x - y)dy]^2$$

$$- 2 \int_{y,z} F_X^n(y)\lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)dydz$$

$$+ \int_{y,z} F_X^n(y)F_X^n(z) \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)dydz .$$

Hence finally

$$\mathrm{Var}(\hat{\lambda}(x)) = (I) + (II) - (E\hat{\lambda}(x))^2$$

$$= \int I_n(F_X(y))\lambda_T(y)K_h^2(x - y)dy$$

$$+ 2 \int_{y<z} \{F_X^n(z) - F_X^n(y)F_X^n(z) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} [F_X^n(z) - F_X^n(y)]\}$$

$$\cdot \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)dydz \ .$$

## 3. Consistency in mean square

### 3.1 Asymptotic unbiasedness

First consider the convoluation error $\int\lambda(y)K_h(x - y)dy - \lambda(x)$, where $\lambda$ is any hazard function. If $\lambda$ is continuous and bounded, or integrable, then this error vanishes as $h \to 0$. In practice, however, hazard functions might not satisfy these conditions (e.g., Weibull with $\gamma > 1$ : $\lambda(t) = \gamma t^{\gamma-1}$). The crucial factor is, of course, how fast K falls off at infinity. The following condition, given by Watson and Leadbetter (1964b), seems close to the minimal condition needed for the convolution error to vanish in the limit. We say that K is compatible with F if the following condition (A) is satisfied:

(A) : For any fixed $M > 0$, there exists h small enough such that $\frac{1}{h} K(\frac{y-x}{h})/(1 - F(y))$ is uniformly bounded for $|y - x| > M$. (Let us denote this bound by $G_M$.)

If K decreases exponentially, or varies regularly with exponent $< -1$, then a sufficient condition for (A) is that $K(x)/(1 - F(x)) \to 0$ as $x \to \infty$. In particular, a kernel with compact support always satisfies (A).

Theorem 1: Let K be compatible with $F_T$, and $n \to \infty$, $h \to 0$, $nh \to \infty$, then $E\hat{\lambda}(x) \to \lambda_T(x)$.

Proof:

$$|E\hat{\lambda}(x) - \lambda_T(x)| \leq \left| \int\lambda_T(y)K_h(x - y)dy - \lambda_T(x) \right| + \int F_X^n(y)\lambda_T(y)K_h(x - y)dy$$

$$\text{First term} \leq \int_{|y-x|<M} K_h(y - x)|\lambda_T(y) - \lambda_T(x)|dy$$

$$+ \int_{|y-x|>M} K_h(y - x)\lambda_T(y)dy + \int_{|y-x|>M} \lambda_T(x)K_h(y - x)dy$$

$$= (I) + (II) + (III),$$

this is true for any M.

(III) $\leq \lambda_T(x) \int_{|y-x|>M} K_h(y - x)dy = \lambda_T(x) \int_{|t|>\frac{M}{h}} K(t)dt \to 0$

(II): the integrand is dominated by $G_M f_T(y)$ which is integrable; further, the integrand $\to 0$ for each y as $h \to 0$. Thus (II) $\to 0$ as $h \to 0$ by the dominated convergence theorem.

(I) can be made arbitrarily small by choosing M small, since $\lambda_T$ is continuous at x.

Second term $\leq \left(\int_{|y-x|\leq M} + \int_{|y-x|>M}\right) F_X^n(y)\lambda_T(y)K_h(x - y)dy = IV + V,$

IV $\to 0$ exponentially if $F_X(x + M) < 1$, V $\to 0$ again by the dominated convergence theorem.

3.2 Asymptotic expression for the variance

It turns out that the conditions in Theorem 1 are not sufficient to obtain a useful limiting expression for Var $\lambda(x)$. More conditions need to be imposed, and this can be done in two different ways, as the following theorem explains. The proof of this theorem can be skipped without loss of continuity.

Theorem 2: If the conditions of Theorem 1 hold and

either (i) K is also compatible with $F_C$

or (ii) $\int_{|t|>M} K(t)dt = o(M^{-\beta})$ as $M \to \infty$ and

$$h = o(n^{\frac{2}{\beta+1}}),\ \beta > 1$$

then

$$\mathrm{Var}(\hat{\lambda}(x)) = \frac{M_x}{nh} + o\left(\frac{1}{nh}\right)$$

where

$$M_x = \frac{\lambda_T(x)}{S_T(x)S_C(x)} \left(\int K^2(t)dt\right) .$$

Proof: In the case (i), K is compatible with both $F_C$ and $F_T$, the arguments used by Watson and Leadbetter (1964b) to derive their asymptotic formula (theorem 2 of that paper) also apply here, with obvious modifications, to produce the above result. Thus we will only prove the above theorem for the case (ii).

$$\mathrm{Var}(\hat{\lambda}(x)) = \int I_n(F_X(y))\lambda_T(y)K_h^2(x - y)\,dy$$

$$+ 2 \int_{y<z} \left(F_X^n(z) - F_X^n(y)F_X^n(z) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} \left[F_X^n(z) - F_X^n(y)\right]\right)$$

$$\cdot \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)dydz$$

where

$$I_n(F) = \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n}{k} F^k(1 - F)^{n-k} \quad .$$

The theorem is proved if we can show that

$$\text{(a):}\quad \frac{n}{\alpha_h} \int I_n(F_X(y))\lambda_T(y)K_h^2(x - y)dy \to \frac{\lambda_T(x)}{1-F_X(x)} \text{ , and}$$

$$\text{(b):}\quad \frac{n}{\alpha_h} \int_{y<z} \left\{F_X^n(z) - F_X^n(y)F_X^n(z) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} \left[F_X^n(z) - F_X^n(y)\right]\right\}$$

$$\cdot \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - z)dydz \to 0$$

where

$$\alpha_h = \int K_h^2(t)dt = \frac{1}{h} \left(\int K^2(t)dt\right) \quad .$$

Proof of (a): rewrite the left hand side as

$$\frac{n}{\alpha_h} \left(\int_{|y-x|<M} + \int_{|y-x|>M}\right) I_n(F_X(y))\lambda_T(y)K_h^2(x - y)\,dy$$

For the first integral, choose M s.t. $F_X(x+M) < 1$; then $n\, I_n(F_X(y)) \to \frac{1}{1-F_X(y)}$ uniformly in y for $|y - x| \le M$ (Lemma 6 of Watson and Leadbetter, 1964b). Thus the first integral is asymptotically equivalent to

$$\int_{|y-x|<M} \frac{\lambda_T(y)}{1-F_X(y)} \cdot \frac{K_h^2(x-y)}{\alpha_h}\,dy \quad .$$

This converges to $\frac{\lambda_T(x)}{1-F_X(x)}$ since if $K_h$ is a peaking kernel, then so does $K_h^2/\alpha_h$.

For the second integral, it is bounded by

$$\frac{n}{\alpha_h} \int_{|y-x|>M} \left( \sum_{r=0}^{n-1} \frac{1}{n-r} \right) \lambda_T(y) K_h^2(x - y)dy$$

$$= \frac{n}{\alpha_h} \int_{|y-x|>M} (\ell n\, n + O(1))\lambda_T(y)K_h^2(x - y)dy \quad .$$

Thus it suffices to observe that

$$\frac{n\, \ell n\, n}{\alpha_h} \int_{|y-x|>M} \lambda_T(y)K_h^2(x - y)dy \leq \frac{n\, \ell n\, n}{\alpha_h} G_M \cdot (\sup_y f_T(y)) \int_{|y-x|>M} K_h(x-y)dy$$

$$= \text{constant} \cdot (n\, \ell n\, n) \cdot h \cdot \int_{|t|>\frac{M}{h}} K(t)dt$$

$$= o\left(n\, \ell n\, n \cdot h^{\beta+1}\right) \quad .$$

Proof of (b): Let us decompose the integral according to the partition: $\{y < z\}$ = region 1 $\cup$ region 2 $\cup$ region 3 with

region 1: $y < z$, $|y - x| \leq M$, $z > z_0$

region 2: $y < z$, $|y - x| \leq M$, $z < z_0$

region 3: $y < z$, $|y - x| > M$

where m, $z_0$ is chosen such that

$$\begin{cases} F_X(x+M) = \alpha < 1, \; z_0 > x+M, \; F_X(z_0) < 1 \\ F_X(z_0) - F_X(x+M) = \delta > 0 \quad . \end{cases}$$

In region 1:

$$\left| F_X^n(z) - F_X^n(y)F_X^n(z) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} \left(F_X^n(z) - F_X^n(y)\right) \right|$$

$$= \left| \left( \frac{1-F_X(y)}{F_X(z)-F_X(y)} - F_X^n(z) \right) F_X^n(y) + \frac{F_X^n(z)(F_X(z)-F_X(y))-(1-F_X(y))F_X^n(z)}{F_X(z)-F_X(y)} \right|$$

$$\leq \frac{1-\alpha+\delta}{\delta} \alpha^n + \frac{1-F_X(z)}{\delta} F_X^n(z) \quad .$$

Thus the absolute value of the integral in (b) over region 1 is $\leq$

$$\frac{n}{\delta\alpha_h} \int_{\text{region 1}} \{(1 - \alpha + \delta)^n + (1 - F_X(z))F_X^n(z)\}\lambda_T(z)\lambda_T(y)K_h(x - y)K_h(x - z)dydz \ .$$

The first term clearly $\to 0$,

$$2^{\text{nd}} \text{ term} = \frac{1}{\delta}\frac{n}{\alpha_n} \int_{\text{region 1}} (1 - F_C(z))F_X^n(z)\lambda_T(y)f_T(z)K_h(x - z)K_h(x - y)dydz$$

$$\leq \frac{(\sup f_T(z))}{\delta}\frac{n}{\alpha_h}\left(\int_{|y-x|\leq M} \lambda_T(y)K_h(x - y)dy\right)\left(\int_{z>z_o} K_h(x - z)dy\right)$$

$$\leq \frac{(\sup f_T(z))}{\delta}\frac{n}{\alpha_h}\left(\int_{|y-x|\leq M} \lambda_T(y)K_h(x - y)dy\right)\left(\int_{t> \frac{M}{h}} K(t)dt\right)$$

$$= o(n h^{\beta+1}) \ .$$

<u>In region 2</u>: $F_X(y) < F_X(z) < F_X(a_o) = a < 1$ .

$$\left| F_X^n(z) - F_X^n(y)F_X^n(z) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} [F_X^n(z) - F_X^n(y)] \right|$$

$$\leq F_X^n(z)[1 - F_X^n(y)] + (1 - F_X(y))\frac{F_X^n(z)-F_X^n(y)}{F_X(z)-F_X(y)}$$

$$\leq a^n + n a^{n-1} \ .$$

Hence the absolute value of the integral in (b) over region 2 is

$$\leq \frac{n}{\alpha_h}(a^n + n a^{n-1}) \int_{\text{region 2}} \lambda_T(y)\lambda_T(z)K_h(x - y)K_h(x - y)dydz$$

$$\leq \frac{n}{\alpha_h}(a^n + n a^{n-1}) \left[\int \lambda_T(y)K_h(x - y)dy\right]^2 \to 0 \ .$$

<u>In region 3</u>:

$$\left| F_X^n(z) - F_X^n(z)F_X^n(y) - \frac{1-F_X(y)}{F_X(z)-F_X(y)} (F_X^n(z) - F_X^n(y)) \right|$$

$$\leq \left| (1-F_X(y)) [F_X^n(z) (1+F_X(y)+\ldots+F_X^{n-1}(y)) - (F_X^{n-1}(z)+F_X^{n-2}(z)F_X(y)+\ldots+F_X^{n-1}(y))] \right|$$

$$\leq 2n(1 - F_X(y)) \ .$$

Hence the absolute value of the integral in (b) over region 3 is

$$\leq \frac{2n^2}{\alpha_h}\left(\int \lambda_T(z)K_h(x - z)dz\right)\left(\int_{|y-x|>M} f_T(y)K_h(x - y)dy\right)$$

$$= o\left(n^2h^{\beta+1}\right) \quad .$$

This completes the proof of Theorem 2.

3.3 Remarks

First we make the obvious remark that Theorems 1 and 2 together give the conditions for mean square consistency of $\lambda(x)$.

Next we comment on conditions (i) and (ii) of Theorem 2. Condition (i) has the advantage that it imposes no further constraint on h besides those stated in Theorem 1. On the other hand, since $F_X(x) < 1$ and $F_C(x) < 1$, we will eventually observe a number of deaths around the time point x, from which we should be able to accurately estimate $\lambda_T(x)$. Thus it seems undesirable that the tail behavior of $F_C$ should impose conditions on K for the estimation of $\lambda_T(x)$. In this respect, condition (ii), which imposes further conditions on K and h, but removes the dependence of K on $F_C$, is a useful alternative to (i).

Finally, the approximate variance given by Theorem 2 agrees with our intuition that, as the point of interest x goes to infinity, the variability of the estimate increases without limit.

4. Asymptotic normality

4.1 The projection method

$\hat{\lambda}(x)$ is obtained from the data in a complicated way, involving functions of $X_i$'s and their ranks, as well as $\delta$'s. None of the central limit theorems, theorems about combinations of order statistics nor theorems about linear rank statistics can be applied directly in this situation. But as will be shown in the following section, the calculation of the expectation of $\lambda(x)$ conditional on $(X_i,\delta_i)$ is nevertheless possible. It is exactly in this type of situation that the projection method (Hajek, 1968) is most effective in investigating asymptotic normality. Let us first summarize the basics of this method.

Suppose $Y_1,\ldots,Y_n$ are i.i.d., W a statistic based on $\underset{\sim}{Y}$. . The key idea of Hajek's method is that, even though the central limit theorem is concerned with sums of independent random variables, its scope may be extended to statistics asymptotically equivalent to such sums. Thus we can try to approximate W by its projection $\hat{W}$ on the subspace of all such sums of independent terms. Hajek gave the following formulae;

$$\hat{W} = \sum_{i=1}^{n} E(W|Y_i) - (n-1)EW$$

$$E\hat{W} = EW$$

$$E(W - \hat{W})^2 = Var(W) - Var(\hat{W}) \quad .$$

The projection method consists of finding conditions such that (a) the standardized forms of W and $\hat{W}$ have the same asymptotic distribution. (b) the central limit theorem can be applied to establish asymptotic normality for W.

4.2 Calculation of $\hat{W}$

Let $W = \hat{\lambda}(x)$, $Y_i = (X_i, \delta_i)$ $i = 1,\ldots,n$. The projection $\hat{W}$ will now be calculated according to Hájek's formula. For this we need EW and $E(W|Y_i)$. The former is already given in Section 2.1. The calculation of $E(W|Y_i)$ is simplified by the fact that W can be expressed as a sum of identically distributed terms:

$$W = \sum_{j=1}^{n} W_j, \qquad W_j = \frac{1}{n-R_j+1} \delta_j K_h(x - X_j),$$

where $R_j$ is the rank of $X_j$ in the sample $\underline{X}$. Thus

$$E(W|Y_i) = \sum_{j=1}^{n} E(W|Y_i) = E(W_i|Y_i) + (n-1)E(W_j|Y_i), \; j \neq i,$$

where $E(W_i|Y_i)$ and $E(W_j|Y_i)$ is given in (4.2.1) and (4.2.2) below.

(i) $$E(W_i|Y_i) = \delta_i K_h(x - X_i) E\left(\frac{1}{n-R_i+1} \,\middle|\, X_i, \delta_i\right)$$

But given $(X_i, \delta_i)$, $R_i \sim 1 + \text{Binomial}(n-1; F_X(X_i))$

Hence

$$E\left(\frac{1}{n-R_i+1} \,\middle|\, X_i, \delta_i\right) = \sum_{k=0}^{n-1} \frac{1}{n-k} \binom{n-1}{k} F_X(X_i)^k (1 - F_X(X_i))^{n-k-1}$$

$$= \frac{1}{n(1-F_X(X_i))} \left(1 - F_X(X_i)^n\right) \quad .$$

(4.2.1) $$E(W_i|Y_i) = \frac{1-F_X(X_i)^n}{n(1-F_X(X_i))}\ \delta_i\ K_h(x - X_i)$$

(ii) For $j \neq i$,

$$E(W_j|Y_i) = E\left(\frac{\delta_j K_h(x-X_j)}{n-R_j+1} \mid X_i,\delta_i\right)$$

$$= E(\delta_j\ K_h(x - X_j)E\left(\frac{1}{n-R_j+1} \;\middle|\; X_i,\delta_i,X_j,\delta_j\right) \;\middle|\; X_i,\delta_i) \quad .$$

Now given $X_i$, $X_j$, $\delta_i$, $\delta_j$,

$$R_j \sim \begin{cases} 1 + \text{Binomial}(n - 2, F_X(X_j)) & \text{if } X_j < X_i \\ 2 + \text{Binomial}(n - 2, F_X(X_j)) & \text{if } X_j < X_i \end{cases} \quad .$$

Thus, for $X_j < X_i$,

$$E\left(\frac{1}{n-R_j+1} \;\middle|\; X_i,\delta_i,X_j,\delta_j\right) = \sum_{k=0}^{n-2} \frac{1}{(n-k)} \frac{(n-2)!}{k!(n-k-2)!} F_X(X_j)^k(1 - F_X(X_j))^{n-k-2} \ ;$$

for $X_j > X_i$,

$$E\left(\frac{1}{n-R_j+1} \;\middle|\; X_i,\delta_i,X_j,\delta_j\right) = \sum_{k=0}^{n-2} \frac{1}{(n-k-1)} \frac{(n-2)!}{k!(n-k-2)!} F_X(X_j)^k(1 - F_X(X_j))^{n-k-2}$$

$$= \frac{1}{(n-1)(1-F_X(X_j))}[1 - F_X(X_j)^{n-1}] \quad .$$

The above expressions can be combined into

$$E\left(\frac{1}{n-R_j+1} \;\middle|\; X_i,\delta_i,X_j,\delta_j\right) = \frac{1}{(n-1)(1-F_X(X_j))}[1 - F_X(X_j)^{n-1}]$$

$$- \frac{1}{n(n-1)(1-F_X(X_j))^2}[1 - F_X(X_j)^n - nF_X(X_j)^{n-1}(1-F_X(X_j))]$$

$$\cdot\ I_{[X_j<X_i]}$$

Hence

$$(4.2.2)\quad E(W_j|Y_i) = \frac{1}{n-1}\int \frac{1-F_X(y)^{n-1}}{1-F_X(y)}\, m(y)K_h(x-y)f_X(y)dy$$

$$- \frac{1}{n(n-1)}\int \frac{[1-F_X(y)^n - nF_X(y)^{n-1}(1-F_X(y))]}{(1-F_X(y))^2}\, I_{[y<X_i]}m(y)$$

$$\cdot K_h(x-y)f_X(y)dy\ .$$

Finally, the expression for $\hat{W}$ is obtained by putting (2.1.1), (4.2.1), (4.2.2) into

$$\hat{W} = \sum_{i=1}^{n} [E(W_i|Y_i) + (n-1)E(W_j|Y_i)] - (n-1)EW$$

where $j \neq i$ in $E(W_j|Y_i)$ .

Thus, we can write

$$\hat{W} - E\hat{W} = \sum_{i=1}^{n} \{E(W_i|Y_i) + (n-1)E(W_j|Y_i) - EW\}$$

$$= \sum_{i=1}^{n} \{\frac{1}{n} V_n(Y_i) + \frac{1}{n} U_n(Y_i) + \Delta_n\}$$

where $$\Delta_n = -\int F_X^{n-1}(y)m(y)K_h(x-y)f_X(y)dy$$

$$V_n(Y_i) = \frac{1-F_X^n(X_i)}{1-F_X(X_i)}\ \delta_i\ K_h(x-X_i)$$

$$U_n(Y_i) = -\int \frac{[1-F_X(y)^n - nF_X(y)^{n-1}(1-F_X(y))]}{(1-F_X(y))^2}\, I_{[y<X_i]}m(y)f_X(y)K_h(x-y)dy\ .$$

4.3 Asymptotic normality

Theorem 3: If K is compatible with both $F_T$ and $F_C$, then the standardized form of $W = \lambda(x)$ has an asymptotic normal distribution, as $n \to \infty$, $h \to 0$, $h = n^{-\gamma}$, $0 < \gamma < 1$.

Proof: The main steps of the proof are as follows: (i) obtain bounds for $|U_n|$ and $|\Delta_n|$, (ii) obtain approximate moments for $V_n$, (iii) check that $Var(W)/VAR(W) \to 1$, this will guarantee that W and W have the same type of distribution asymptotically, (iv) check Lyapounov's condition for W, thus proving asymptotic normality.

(i) $|U_n| = O(\ell n\ n)$, $|\Delta_n| = O\left(\frac{1}{n(n+1)}\right)$ .

Proof of (i): choose M such that $F_X(x+M) < 1$,

$$0 < -U_n(Y_i) = \left(\int_{|y-x|\le M} + \int_{|y-x|>M}\right) \frac{[1-F_X(y)^n - nF_X(y)^{n-1}(1-F_X(y))]}{1-F_X(y)} m(y) f_X(y) \cdot \frac{K_h(x-y)}{1-F_X(y)}\, dy$$

$$< C + G_M \int_{|y-x|>M} (1 + F_X(y) + \ldots + F_X(y)^{n-1} - nF_X(y)^{n-1})\, dF_X(y)$$

$$< C + G_M \int_0^1 (1 + F + F^2 + \ldots + F^{n-1} - nF^{n-1})\, dF$$

$$= O(\ell n\ n) \quad .$$

$$0 < -\Delta_n < O\left(F_X^{n-1}(x+M)\right) + G_M \int F_X^{n-1}(1 - F_X(y))dF_X(y) = O\left(\frac{1}{n(n+1)}\right) \quad .$$

(ii) $E|V_n(Y_i)|^r = \alpha_{r,h} \left(\frac{1}{1-F_X(x)}\right)^r m(x)f_X(x) + o(\alpha_{r,h})$

where $\alpha_{r,h} = \int K_h^r(y)dy = \frac{1}{h^{r-1}} \int K^r(t)dt$; $r = 1,2,3,\ldots$

Proof of (ii):

$$E|V_n(Y_i)|^r = \int \frac{(1-F_X^n(y))^r}{(1-F_X(y))^r} m(y)K_h^r(x - y)f_X(y)dy$$

$$< \int_{|y-x|<M} \frac{(1-F_X^n(y))^r}{(1-F_X(y))^r} m(y)K_h^r(x - y)f_X(y)dy$$

$$+ G_M^r \int_{|y-x|>M} (1 - F_X^n(y))^r\, m(y)f_X(y)dy$$

$$= \alpha_{r,h} \left(\frac{1}{1-F_X(h)}\right)^r m(x)f_X(x) + o(\alpha_{r,h}) \quad .$$

Here we have used the fact that $K_h^r/\alpha_{r,h}$ is also a peaking kernel and M is chosen such that $F_X(x+M) < 1$.

(iii) $\mathrm{Var}(\hat{W})/\mathrm{Var}(W) \to 1$.

Proof of (iii): making use of (i) and (ii) above, we have

$$\mathrm{Var}(U_n(Y_i)) = O((\ell n\ n)^2),$$

$$\mathrm{Var}(V_n(Y_i)) = \alpha_{2,h} \left(\frac{1}{1-F_X(x)}\right)^2 m(x) f_X(x) + o(\alpha_{2,h})$$

$$= \frac{(\int K^2(t)dt)}{h} \frac{\lambda_T(x)}{1-F_X(x)} + o\left(\frac{1}{h}\right) .$$

Hence,

$$\mathrm{Var}(\hat{W}) = n\ \mathrm{Var}\left(\frac{1}{n} V_n(Y_i) + \frac{1}{n} U_n(Y_i) + \Delta_n\right)$$

$$= \frac{1}{n}\left[\mathrm{Var}(V_n(Y_i)) + 2\ \mathrm{cov}(V_n(Y_i), U_n(Y_i)) + \mathrm{Var}(U_n(Y_i))\right]$$

$$= \frac{(\int K^2(t)dt)}{nh} \frac{\lambda_T(x)}{1-F_X(x)} + o\left(\frac{1}{nh}\right) + o\left(\frac{\ell n\ n}{nh^{1/2}}\right) + O\left(\frac{(\ell n\ n)^2}{n}\right)$$

$$= \frac{(\int K^2(t)dt)}{nh} \frac{\lambda_T(x)}{1-F_X(x)} + o\left(\frac{1}{nh}\right) .$$

Comparing with expression for Var(W) in Theorem 2, we have (iii).

The fact that $\frac{\mathrm{Var}(\hat{W})}{\mathrm{Var}(W)} \to 1$ guarantees that

$$E\left(\frac{\hat{W}-E\hat{W}}{\sqrt{\mathrm{Var}(\hat{W})}} - \frac{E-E\hat{W}}{\sqrt{\mathrm{Var}(\hat{W})}}\right)^2 = \frac{1}{\mathrm{Var}(\hat{W})} E(\hat{W} - W)^2$$

$$= \frac{1}{\mathrm{Var}(\hat{W})} [\mathrm{Var}(\hat{W}) - \mathrm{Var}(W)] \quad \text{(by Hajek's formula)}$$

$$\to 0 .$$

But $\frac{W-EW}{\sqrt{\mathrm{Var}(W)}}$ has the same asymptotic distribution as $\frac{W-E\hat{W}}{\sqrt{\mathrm{Var}(\hat{W})}}$, hence it has the same asymptotic distribution as $\frac{\hat{W}-E\hat{W}}{\sqrt{\mathrm{Var}(\hat{W})}}$.

(iv) Finally it remains to show that $\frac{\hat{W}-E\hat{W}}{\sqrt{Var(\hat{W})}} \to N(0,1)$. Since $\Delta_n$ is negligible, by Lyapounov's theorem, a sufficient condition will be:

$$\frac{1}{Var(\hat{W})^{3/2}} \sum_{i=1}^{n} E\left| \frac{1}{n} V_n(Y_i) + \frac{1}{n} U_n(Y_i) \right|^3 \to 0 \quad .$$

To check this, rewrite the left hand side as

$$LHS = \frac{1}{Var(\hat{W})^{3/2}} \frac{1}{n^2} E|V_n(Y_i) + U_n(Y_i)|^3$$

$$= \frac{1}{Var(\hat{W})^{3/2}} \frac{1}{n^2} [E|V|^3 + 3 (E|V_n|^4 E|U_n|^2)^{1/2} + 3 (E|V_n|^2 E|U_n|^4)^{1/2} + E|U_n|^3] \quad .$$

Now

$$E|V_n|^3 = \frac{1}{h^2} \left( \int K^3(t)dt \right) \frac{m(x) f_X(x)}{(1-F_X(x))^3} + o\left(\frac{1}{h^2}\right)$$

$$E|V_n|^r = O\left(\frac{1}{h^{r-1}}\right)$$

$$E|U_n|^r = O((\ell n\ n)^r)$$

$$Var(\hat{W}) = \frac{\left(\int K^2(t)dt\right)}{nh} \frac{\lambda_T(x)}{1-F_X(x)} + o\left(\frac{1}{nh}\right) \quad .$$

Thus

$$LHS \leq O\left(\frac{1}{\left(\frac{1}{nh}\right)^{3/2}} \cdot \frac{1}{(nh)^2}\right) = O\left(\frac{1}{\sqrt{nh}}\right) ,$$

verifying the condition.

## 5. Discussion

### 5.1 Bias correction near the origin

In calculating the estimate $\hat{\lambda}$ in Figure 1, a bias correction procedure was used for points near the origin. Figure 2 presents $\lambda$ without bias correction. The effect is seen to be quite large for points close to the origin. The reason for the bias is apparent in the dominant terms of the expression for the expected value:

$$E\hat{\lambda}(x) \approx \int_0^\infty \lambda(y)K_h(x - y)dy$$

$$= \int_{\frac{x}{h}}^\infty \lambda(x + ht)K(t)dt$$

$$= \lambda(x) \int_{\frac{x}{h}}^\infty K(t)dt + R \quad .$$

The remainder term R is negligible either when h is small or when $\lambda'(t)$ is small around x. For x large compared to h, $\int_{x/h}^\infty K(t)dt \approx 1$, and even in this case the convolution approximation will be good only if the remainder R is small. Thus, in doing bias correction, we will only correct for the leading term $\lambda(x)\int_{x/h}^\infty K(t)dt$. For x small compared to h this leading term is significantly less than $\lambda(x)$. The bias correction procedure is to divide the estimate $\lambda(x)$ by the factor $\int_{x/h}^\infty K(t)dt$.

5.2 Significance of the present study

Our ultimate goal is to establish the theoretical properties of the fully data adaptive procedures. However, this is a difficult problem. We regard the present paper as solving a significant component problem. One must understand how these estimators behave when the parameters are chosen deterministically as a prerequisite to the analysis of the behavior of the data adaptive procedure. Another component of this problem, dealing with the nearest neighbor hazard estimator, is solved by Tanner (1982).

The present study provides several insights to the relation between density and hazard estimation. The leading term of the bias expression presented in Section 2 is the convolution approximation error $\lambda_T * K_h(x) - \lambda_T(x)$. In the uncensored situation the bias of the kernel density estimate is also a convolution approximation error $f_T * K_h(x) - f_T(x)$. Because $f_T$ is $L^1$, the convolution error is guaranteed to vanish in the limit. However, $\lambda_T$ need not be $L^1$ and hence kernel hazard estimation is at a disadvantage on this point. Convolution is an approximation of the value of a function by a weighted average (according to the kernel) of values at other points. We must therefore ensure that function values at points far away must not be so large that the down-weighting by the kernel is insufficient. This is the essence of the compatibility condition between the kernel and $F_T$ in Sections 3 and 4. We point out that if a kernel with compact support is used then the above difficulty cannot arise. In this case, the asymptotic mean square error behavior of the kernel is exactly the same as that of the kernel density estimate, and all the familiar convergence rate results (in MSE) for kernel density estimates carry over to kernel hazard estimates.

It must now be emphasized that, with censored data, it is no longer possible to express the bias in the kernel density estimates as a convolution approximation error (see next section for a description of kernel density estimates from censored data and relevant references). It appears that in the case of censoring, the hazard function is a more natural entity to analyze than the density. One directly observes $X = \min(T,C)$, and $\delta = I_{[T<C]}$. While the hazard functions are related in a very simple way $\lambda_X = \lambda_T + \lambda_C$ (under independence of T and C), the densities are related in a more complicated way $f_X = f_T S_C + f_C S_T$. Thus, it is not surprising that in this context the estimation of the hazard $\lambda_T$ is a more tractible problem than the estimation of the density $f_T$.

The consideration of the convolution error in the bias also gives us a guide as to when hazard estimation is to be preferred over density estimation. Namely, if over the region of interest we expect stability on the hazard scale, then hazard estimation is preferred since the convolution error will be very small. Conversely, if the density scale is more stable, then density estimation is to be preferred.

In the area of reliability and survival analysis, it is usually the hazard scale which is more stable. For example, the exponential hazard is a constant, and it is the exponential life time which is fundamental to much of reliability and survival analysis. In practice, an exponential model might be too restrictive, but there are often situation where we do not expect the hazard to vary drastically over the region of interest. The kernel hazard estimate should perform well in these situations, since in these cases the window width h need not be very small in order to achieve a good convolution approximation, hence the variance, which is of order $(nh)^{-1}$, can be made much smaller than otherwise possible.

ACKNOWLEDGEMENT

We are grateful to Professors R. R. Bahadur and Stephen M. Stigler for bringing Hajek's projection method to our attention. This research is supported by National Science Foundation Grant No. MCS-8101836.

REFERENCES


1. Aalen, O. (1978). "Nonparametric estimation of partial transition probabilities in multiple decrement models." Annals of Statistics, 6, 534-545.

2. Bartoszyński, R., Brown, B. W., McBride, C. M., and Thompson, J. R. (1981). "Some nonparametric techniques for estimating the intensity function of a cancer related non-stationary Poisson process." Annals of Statistics,

3. Friedman, M. (1978). "Analysis of survival data with covariates and censoring using a piecewise exponential model." Ph.D. dissertation, Department of Statistics, University of Chicago.

4. Földes, A., Rejtö, L., and Winter, B. B. (1981). "Strong consistency properties of nonparametric estimators for randomly censored data, II: estimation of density and failure rate." Periodica Mathematica Hungarica, 12, 15-29.

5. Hajek, J. (1968). "Asymptotic normality of simple linear rank statistics under alternatives." Annals of Mathematical Statistics, 39, 325-346.

6. Kaplan, E. L. and Meier, P. (1958). "Nonparametric estimation from incomplete observations." J.A.S.A., 53, 457-481.

7. Kimura, D. K. (1972). "Fourier series methods for censored data." Ph.D. dissertation, Department of Biostatistics, University of Washington.

8. Lo, A. Y. (1978). "On a class of Bayesian nonparametric estimates; I. Density estimates, II: Rate function estimates." Department of Mathematics and Statistics Technical Report, University of Pittsburgh. (Part I will appear in Annals of Statistics.)

9. Meier, P. (1975). "Estimation of a distribution function from incomplete observations." Perspectives in Probability and Statistics. Edited by J. Gani, pp. 67-87.

10. McNichols, D. T. and Padgett, W. J. (1981). "Kernel density estimation under random censorship." University of South Carolina, Department of Statistics, Technical Report No. 74.

11. Nelson, W. (1972). "Theory and application of hazard plotting for censored failure data." Technometrics, 14, 945-966.

12. Rice, J. and Rosenblatt, M. (1976). "Estimation of the log survivor function and hazard function." Sankhya, A, 38, 60-78.

13. Tanner, M. (1982). "Nearest neighbor and variable kernel estimators of the hazard function from censored data." Technical Report No. 135, Department of Statistics, University of Chicago.

14. Tarter, M. (1979). "Trigonometric maximum likelihood estimation and application to the analysis of incomplete survival information." J.A.S.A., 74, 132-139.

15. Watson, G. S. and Leadbetter, M. R. (1964a). "Hazard analysis. I." Biometrika, 51, 175-184.

16. Watson, G. S. and Leadbetter, M. R. (1964b). "Hazard analysis. II." Sankhya, A, 26, 101-116.

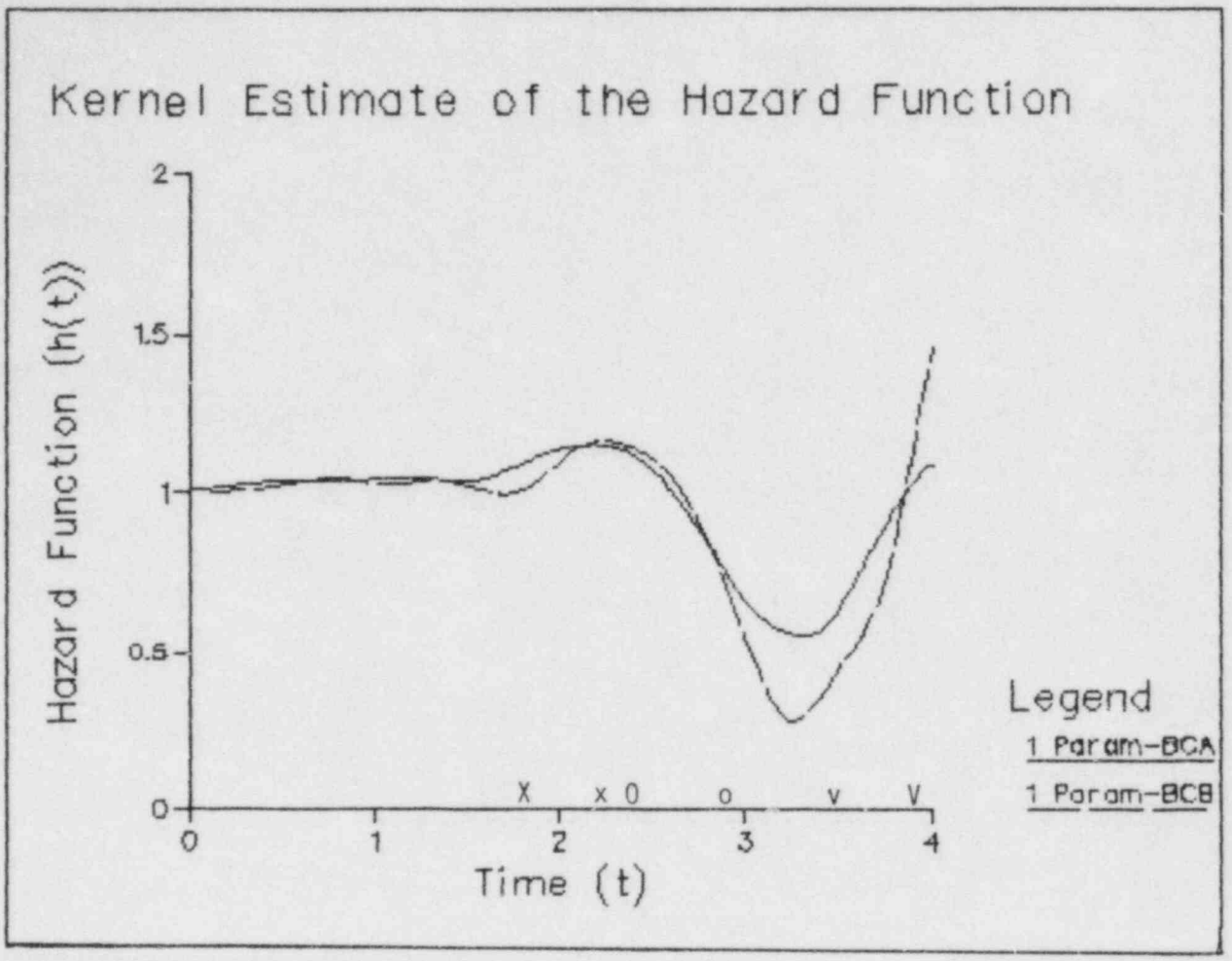



Figure 1

Kernel hazard estimate from 200 observations, with bias correction. Unit exponential lifetimes and exponential censoring times. The symbols X, O, V are the 90th, 95th, and 97th sample percentiles, the small case letters indicate the corresponding population percentiles. The solid and broken lines correspond to the fractional and log-weight versions of the estimator.

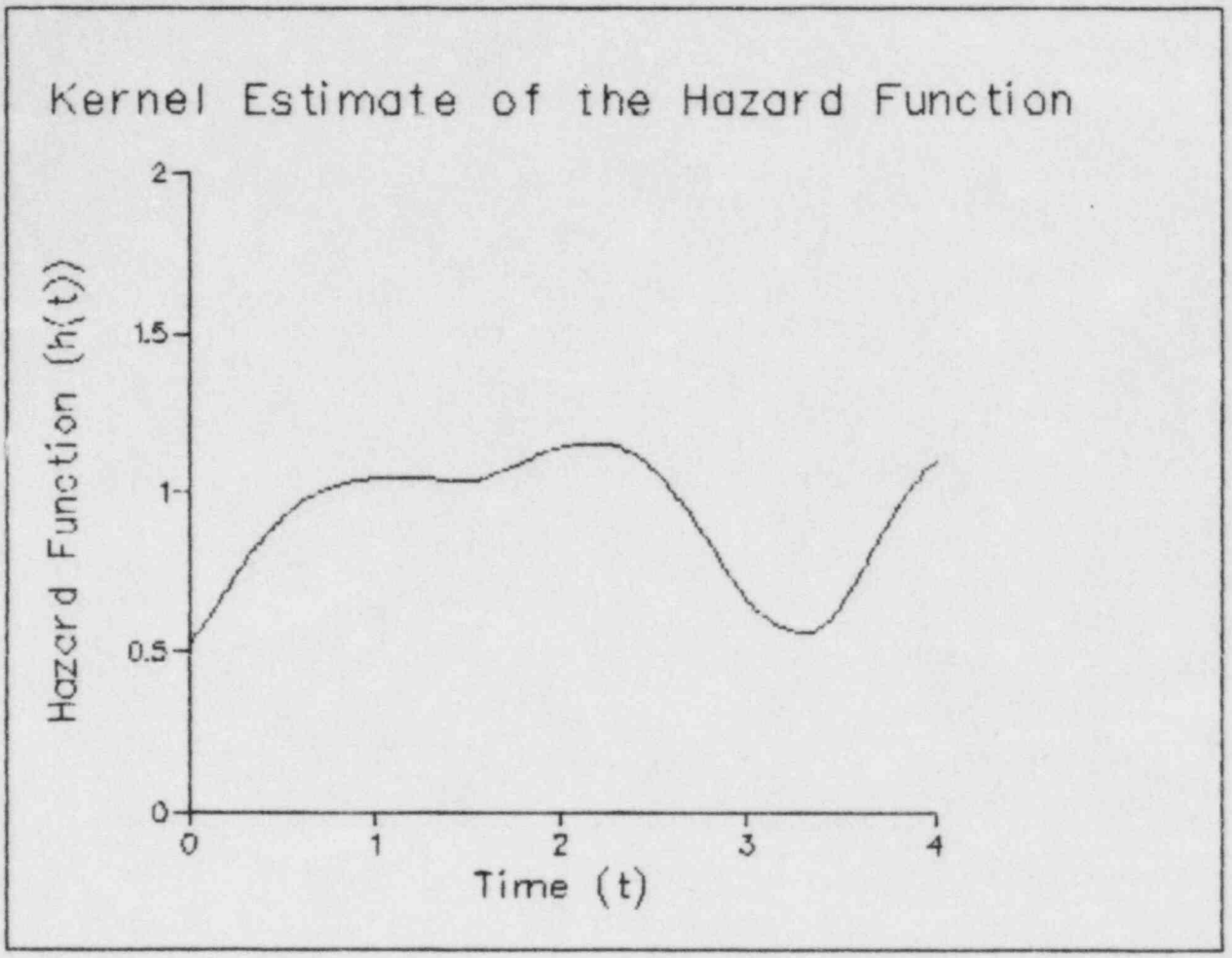



Figure 2

Kernel estimate without the bias correction, from the same data as Figure 1.

# COMBINING ECONOMETRIC AND TIME-SERIES FORECASTS OF DEMAND FOR RESIDUAL FUEL OIL

by

Cyrus Baghelai
Applied Management Sciences, Inc.
962 Wayne Avenue, Suite 701
Silver Spring, Maryland 20910

(Contributed Paper Session 1)

# COMBINING ECONOMETRIC AND TIME-SERIES FORECASTS OF DEMAND FOR RESIDUAL FUEL OIL

CYRUS BAGHELAI

Applied Management Sciences, Inc.
962 Wayne Avenue, Suite 701
Silver Spring, Maryland 20910

## ABSTRACT

The combination forecast methodology is useful because in selecting the "best" single model the analysts may discard independent information of interest, which exists in the rejected models. The methodology of combining forecasts is founded on the axiom of maximum information usage. This paper considers three alternative approaches--econometric, Box-Jenkins, and Winters--to forecasting short-term demand for non-utility residual fuel oil. In the paper, various combinations of these three approaches are used to yield combined forecasts, using the Bates-Granger Technique, applied to monthly data on non-utility residual fuel oil. It is demonstrated that the optimum method proposed is superior to the existing single equation econometric model.

## 1. INTRODUCTION

The essential objective of most forecasting is to provide decisionmakers with the information necessary to permit confidence in the decision made. Since no one forecast includes all the available information or all possible specifications, multiple forecasts are often prepared for the same variable. These multiple forecasts are then examined within the context of the decision objectives, and one forecast is selected as most pertinent to the question at hand. Selecting a single forecast may not make the best use of available information, however, for several reasons. Although there is likely to be some common information among forecasts, rejected forecasts may contain some information not available in the selected forecast. Rejected forecasts could be based on different assumptions, different variables, or different relationships between variables. Thus, combined forecasts may provide a better forecast than any individual forecast. The methodology of combining forecasts is founded on the axiom of maximum information usage.

Improvement and refinement of the demand forecast using time-series (e.g., box-Jenkins) and combination forecasts (e.g., econometric and Box-Jenkins) would enable analysts to produce more accurate short-term projections of the demand for non-utility residual fuel oil. This paper summarizes the findings for non-utility residual fuel oil. It includes: (1) estimation of an econometric model to project the demand; (2) development of alternative approaches (Box-Jenkins and Winters) to the short-term demand forecasting; and (3) development of approaches to combine the econometric and the alternative approaches to the short-term demand forecasting.

## Structure of the Paper

This paper consists of five sections, including this introduction. Section 2 describes the data used for the analysis. Section 3 presents and discusses specification of the models and empirical techniques used. The findings are discussed in Section 4, and Section 5 is a summary with conclusions of the paper.

## 2. DATA

To develop forecasts for non-utility residual fuel oil demand, it was necessary to identify monthly data sources for this variable and the other variables used in the estimation of an econometric model. These data were supplied by the Analysis Branch of the Short-Term Information Division in the U.S. Department of Energy, Energy Information Administration.

The most common data sources identified and used in this paper were:

- Monthly Energy Review, U.S. Department of Energy, Energy Information Administration
- Industrial Production, Board of Governors, U.S. Federal Reserve System
- Monthly State, Regional, and National Heating Degree-Days Weighted by Population, U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Table 1 presents a summary of all variables, sources of data, and units for the variables used in this paper.

## 3. METHODOLOGY

This section presents the methodological approaches used to forecast and to evaluate the demand forecasts of non-utility residual fuel oil. To provide an understanding of the findings presented later in this paper, the discussion that follows focuses on:

- Empirical specification and techniques
- Evaluation of model performance.

The purpose of this section is to describe each of these phases and to provide a framework for the analysis and discussion of findings.

TABLE 1: VARIABLE DEFINITIONS, SOURCES, AND UNITS

| Variable | Definition | Source | Unit |
|---|---|---|---|
| $XDRS_t$ | Non-utility residual fuel oil demand at time t. | Monthly Energy Review and EIA, FPC Form 423 | Million barrels per day |
| $XDRS_{t-1}$ | Non-utility residual fuel oil demand at time t-1. | Monthly Energy Review and EIA, FPC Form 423 | Million barrels per day |
| $XPRS_t$ | Real wholesale price of residual fuel oil at time t. | U.S. Department of Energy, FEA Form P302-M-1, "Petroleum Industry Monthly Report for Product Prices" | 1967 cents per gallon |
| $INDM_t$ | Industrial production index of all manufactured articles at time t (1967=1.00). | Board of Governors, Federal Reserve System, Industrial Production, Monthly, G-12.3 | Index, 1967=1.0 |
| $EHDD_t$ | Heating degree-days at time t, national 1980 population weighted. | U.S. Department of Commerce, National Oceanic and Atmospheric Administration (NOAA) | Average daily degrees above Fahrenheit 65 |
| $GASPR_t$ | Real price of natural gas at time t. | Monthly Energy Review | 1967 cents per million Btu |
| $DUM7677_t$ | A binary variable for natural gas curtailments during the 1967-1977 winters; coded 1 in November 1976 to March 1977, 0 otherwise. | -- | Dimensionless |
| $DUMDEC_t$ | A binary variable for period following total decontrol on January 1981; coded 1 in February 1981 forward, 0 otherwise. | -- | Dimensionless |
| $DUMWIN_t$ | A binary variable for warm winter, coded 1 in November 1981 to March 1982, 0 otherwise. | -- | Dimensionless |

## Empirical Specification and Techniques

Two alternative models are used in the forecasts of non-utility residual fuel oil demand:

- An econometric model
- Time-series models, specifically
  - -- Box-Jenkins method
  - -- Winters method
  - -- Combination methods.

In this section, the empirical specification of these four methodological approaches to forecasting are briefly described.

## Econometric Model

The econometric demand model for non-utility residual fuel oil is specified as a linear function of eight variables:

- The real price of No. 6 residual fuel oil to retail consumers in time t--$XPRS_t$. This variable is included in the demand model to measure the impact of residual fuel oil price on the demand for residual fuel oil.
- The industrial production index of all manufactured articles in time t--$INDM_t$. This is expected to have an effect on the demand for residual fuel oil because it is closely related to consumption of total non-utility residual fuel oil.
- The real price of natural gas in time t--$GASPR_t$. Since natural gas is a substitute for residual fuel oil, the price of natural gas is also included in the specification.
- The national population-weighted heating degree-days in time t--$EHDD_t$. It is included in the specification to explain seasonal variation.
- $DUM7677_t$. This is a binary variable and included in the specification for natural gas curtailments during the winter of 1976 and 1977.
- $DUMDEC_t$. This is a binary variable and included in the specification for the period following total decontrol on January 28, 1981.
- $DUMWIN_t$. This is a binary variable indicating an unusually warm winter.
- $XDRS_{t-1}$. This is a lagged dependent variable included in the specification to allow the last observed value to influence the forecast more directly.

The econometric demand model for non-utility residual fuel oil is:

$$XDRS_t = \beta_0 + \beta_1 XPRS_t + \beta_2 INDM_t + \beta_3 GASPR_t + \beta_4 EHDD_t + \beta_5 DUN7677_t + \beta_6 DUMDEC_t + \beta_7 DUMWIN_t + \beta_8 XDRS_{t-1} + e_t. \quad (1)$$

Equation (1) is estimated using the Cochrane-Orcutt 1/ procedure with monthly data from July 1975 to December 1981.

Box-Jenkins Method

The Box-Jenkins method is a forecasting technique that seeks to develop, in a systematic manner, the forecasting model that is best suited to each time series under investigation.2/ This method can lead to a forecast that is better than those produced by other smoothing models. At the outset, a statistical analysis on the data series is conducted to find the forecast model that gives the best fit. The forecast model is then selected from a collection of models that represent the Box-Jenkins family of models.

Four basic stages are necessary to develop a Box-Jenkins forecast; the first three stages construct the Box-Jenkins model and the last stage produces forecasts. The four stages are:

1. Identification -- The objective of this stage is to select the forecast model that seems most appropriate to the time series under study. The data are used to generate a series of sample autocorrelation functions and partial autocorrelation. These are then compared to certain theoretical autocorrelation and partial autocorrelation functions from known forecast models to seek the best match. The forecast model is then identified and selected. The principle of parsimony is applied: the model with the smallest number of coefficients suitable for the series is the model that is selected.

2. Estimation -- Upon selecting the model, the second stage is initiated whereby the coefficients are estimated. The estimates are found so they yield the fit of past observations which produces the minimum sum of squared residual errors.

3. Diagnostic Checking -- Using the fitted results, the residual errors are examined to determine the adequacy of the fit. A good fit will yield residual errors that are randomly distributed with mean zero and a constant variance. The check is made by way of the autocorrelation functions of the residual errors.

4. Forecasting -- Once an appropriate model has been identified, it may be used to generate a forecast of future values that are optimal in a minimum mean-squared-error sense.

Should the diagnostic check fail, the first three stages are repeated until a model is found that gives acceptable results. Once the model and the corresponding coefficients are selected, the model is then used to forecast

future observations. The forecasts are updated in each time period as each new observation entry becomes available. In the event that the time series seems to be changing, the coefficients of the model may be re-estimated or an entirely new model may be selected. When the appropriate model and coefficient estimates are found, an equation is developed from the model to forecast future values.

Winters Method

Linear and seasonal exponential smoothing as developed by Winters 3/ is based on the assumption that the time series is adequately represented by the model:

$$X_t = (S_t + T_t \tau)\, I_t + e_t, \tag{2}$$

where:

$S_t$ = Base signal or permanent component

$T_t$ = Linear trend component

$I_t$ = Multiplicative seasonal factor

$e_t$ = Random error component

$\tau$ = Lead time for the forecast.

There are three basic equations involved in the Winters model:

$$S_t = \alpha \frac{X_t}{I_{t-1}} + (1-\alpha)\,(S_{t-1} + T_{t-1}), \qquad 0 < \alpha < 1 \tag{3}$$

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)\, T_{t-1}, \qquad 0 < \beta < 1 \tag{4}$$

$$I_t = \partial \frac{X_t}{S_t} + (1-\partial) I_{t-L}, \qquad 0 < \partial < 1 \tag{5}$$

where:

L = Length of seasonality

$\alpha$, $\beta$, and $\partial$ = Smoothing parameters.

Equation (5) is comparable to a seasonal index. This index is the ratio of the current value of the series, $X_t$, divided by the current single smoothed value for the series, $S_t$. If $X_t$ is larger than $S_t$, the ratio will be greater than 1; if it is smaller than $S_t$, the ratio will be less than 1. Equation (4) is used for smoothing the trend because it weights the incremental trend $(S_t - S_{t-1})$ with $\beta$ and the previous trend value $T_{t-1}$ with $(1-\beta)$. Equation (3) is used to obtain the smoothed value; the first term is divided by the seasonal value $I_{t-1}$ to eliminate seasonal fluctuations from $X_t$.

In addition to these equations, a forecast is obtained by Equation (6):

$$X_{t+\tau} = (S_t + T_t \tau)\, I_{t-L+\tau}. \qquad (6)$$

Combination Methods (One Step-Ahead Forecasts)

Bates and Granger propose a technique of producing a forecast by combining the forecasts from two alternative methods in their 1969 seminal paper. 4/ Let $F_t^1$ and $F_t^2$ be two unbiased one-step-ahead forecasts of $X_t$ with errors $e_t^1 = X_t - F_t^1$ and $e_t^2 = X_t - F_t^2$, respectively. Then the combined forecast is

$$X_t^c = W_t F_t^1 + (1-W_t)F_t^2, \qquad (7)$$

where $W_t$ is a combining parameter and is varied at each time period.

Consequently, the combined forecast error is:

$$e_t^c = X_t - X_t^c = X_t - W_t F_t^1 - (1-W_t)F_t^2 = W_t e_t^1 + (1-W_t)e_t^2. \qquad (8)$$

The error variance is:

$$\sigma_c^2 = W^2\sigma_1^2 + (1-W)^2\sigma_2^2 + 2W(1-W)\rho\sigma_1\sigma_2, \qquad (9)$$

where:

$\rho$ = Simple correlation coefficient between $e_t^1$ and $e_t^2$,

$\sigma_1^2$ and $\sigma_2^2$ = Sample variance of $e_t^1$ and $e_t^2$, respectively.

To find the value W, which minimizes the error variance, Bates and Granger differentiate Equation (9) with respect to W:

$$\frac{\partial}{\partial W}(\sigma_c^2) = 2W\sigma_1^2 - 2(1-W)\sigma_2^2 + 2(1-W)\rho\sigma_1\sigma_2 - 2W\rho\sigma_1\sigma_2. \qquad (10)$$

Equating Equation (10) to zero gives:

$$W = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \qquad (11)$$

Substituting Equation (11) for Equation (9) gives the minimum error variance:

$$\sigma_c^2 = \frac{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}. \tag{12}$$

It can be seen that if W is determined by Equation (11), the value $\sigma_c^2$ is no greater than the smaller of the two individual variances unless either $\rho$ is exactly equal to $\sigma_1/\sigma_2$ or to $\sigma_2/\sigma_1$. If either equality holds, the variance of the combined forecast is equal to the smaller of the two error variances.

Bates and Granger list five additional alternative methods for seeking $W_t$.

$$\text{Method 1:} \quad W_t = \frac{\sum_{t=T-v}^{T-1} e_{t,2}^2}{\sum_{t=T-v}^{T-1} (e_{t,1}^2 + e_{t,2}^2)}, \quad t = T-v, \ldots, T-1, \tag{13}$$

where $e_{t,1}$ and $e_{t,2}$ are errors of forecasts 1 and 2, respectively.

$$\text{Method 2:} \quad W_t = \alpha W_{t-1} + \frac{(1-\alpha) \sum_{t=T-v}^{T-1} e_{t,2}^2}{\sum_{t-T-v}^{T-1} (e_{t,1}^2 + e_{t,2}^2)}, \quad 0 < \alpha < 1, \quad t = T-v, \ldots, T-1 \tag{14}$$

where $\alpha$ is a fixed smoothing parameter.

$$\text{Method 3:} \quad W_t = \frac{\sum_{t=1}^{T-1} \beta^t (e_{t,2}^2 - e_{1,t}\, e_{2,t})}{\sum_{t=1}^{T-1} \beta^t (e_{t,1}^2 + e_{t,2}^2 - 2\, e_{t,1}\, e_{t,2})}, \quad \beta \geq 1, \tag{15}$$

where $\beta$ is a discounting parameter.

$$\text{Method 4:} \quad W_t = \frac{\sum_{t=1}^{T-1} \beta^t (e_{2,t}^2)}{\sum_{t=1}^{T-1} \beta^t (e_{t,1}^2 + e_{t,2}^2)}, \quad \beta \geq 1. \tag{16}$$

$$\text{Method 5:} \quad W_t = \alpha W_{t-1} + (1-\alpha) \frac{|e_{t,2}|}{|e_{t,1}| + |e_{t,2}|}, \quad 0 < \alpha < 1. \tag{17}$$

## Evaluation of the Models' Performances

The basic data set to which the non-utility residual fuel oil demand models have been applied is divided into two periods. The first period data are used to determine model identification and to generate forecasts for the second period. Actual values for the second part can then be compared with each of the forecast values, and the applications of statistical measures will permit evaluation of the accuracy of each forecast.

The measures presented in this section have been used to determine the differences in accuracy among the four procedures used for forecasting: the econometric model, the Box-Jenkins model, the Winters model, and the combined models.

Measure 1: Mean Squared Error -- A particular favorite of forecasters using a least-square criterion, mean squared error (MSE) is an obvious measure of forecast quality and is defined as:

$$MSE = \frac{1}{N} \sum_{t=1}^{N} (X_t - F_t)^2, \tag{18}$$

where:

$X_t$ = Actual observations

$F_t$ = Forecasted values.

Measure 2: Theil Inequality Coefficient -- This measure, proposed by H. Theil, 5/ is denoted by $U^2$ and described by the following formula:

$$U^2 = \frac{\frac{1}{N} \sum_{t=1}^{N} (X_t - F_t)^2}{\frac{1}{N} \sum_{t=1}^{N} X_t^2}. \tag{19}$$

$U^2$ equals zero only if all of the forecasts are perfect. $U^2$ equals one when the forecasting procedures leads to the same root-mean-squared error as produced by using naive, no-change, extrapolation. By using the inequality coefficient, the seriousness of the forecast error is measured by the quadratic loss criterion: the zero corresponds with perfection and a value of one corresponds to the loss associated with no-change extrapolation.

Measure 3: Decomposition of Mean Squared Error -- As proposed by Theil, 6/ MSE can be decomposed into three elements, each referring to a particular kind of forecast error. This decomposition can be written in the following ways:

$$U^M = \frac{(\bar{F} - \bar{X})^2}{\frac{1}{N} \sum_{t=1}^{N} (X_t - F_t)^2}, \tag{20}$$

$$U^R = \frac{(S_F - S_X)^2}{\frac{1}{N}\sum_{t=1}^{N}(X_t - F_t)^2}, \tag{21}$$

$$U^D = \frac{2\,(1-r)\,S_X S_F}{\frac{1}{N}\sum_{t=1}^{N}(X_t - F_t)^2}. \tag{22}$$

where:

$\bar{F}$ and $\bar{X}$ = Sample means of the forecasts and actual observations

$S_F$ and $S_X$ = Sample standard derivations of the forecasts and actual observations

$r$ = Sample correlation between the forecasts and the actual observations.

$U^M$ refers to bias proportion, $U^R$ to the regression proportion, and $U^D$ to the disturbance proportion. Both $U^M$ and $U^R$ tend toward zero for the optimum forecast; therefore, $U^D$ should tend toward unity.

Measure 4: Mean Absolute Error -- This measure is defined as:

$$MAE = \frac{\sum_{t=1}^{N} |F_t - X_t|}{N}. \tag{23}$$

Measure 5: Mean Absolute-Percent Error -- This measure is defined as:

$$MAPE = \frac{\sum_{t=1}^{N} \left|\frac{F_t - X_t}{X_t}\right|}{N} \times 100. \tag{24}$$

The forecasting methods used in this study have been evaluated by applying these statistical measures. The results of these comparisons and evaluations are presented in the next section, which reports our findings.

## 4. FINDINGS

In this section, the results of the estimation of the alternative demand forecasting models for non-utility residual fuel oil are presented and discussed. This discussion assesses the implications of statistical measure values used to produce the best short-term forecasts of demand for this product.

## Econometric Model

The econometric demand for non-utility residual fuel oil has been estimated using data from July 1975 through December 1981. The results of this estimation are shown in Table 2. Like all tables referenced in Section 4, Table 2 appears at the end of this paper. As can be seen, the estimated coefficients of the econometric fitted model are all significant at the 0.10 level or better with the exception of INDM, the measure of industrial production.

Evaluation of the estimated equation reveals that this model has relatively good explanatory power. The $R^2$ is 0.911, indicating that 91 percent of the variation in the demand for non-utility residual fuel oil is explained by this model.

The estimated model becomes the basis for forecasting demand for non-utility residual fuel oil for the 9-month period of January 1982 through September 1982. These forecasts will provide a basis for comparison with the alternative forecasts methods described in the sections that follow.

## Box-Jenkins Model

The Box-Jenkins model has been estimated using data for July 1975 through December 1981 on demand for non-utility residual fuel oil. The identification and estimation of the values of the parameter coefficients found in the Box-Jenkins model required that sample autocorrelation and partial autocorrelation functions of various differences in the series be calculated and examined. This examination indicated that, with no differencing, the residual demand series is not stationary, i.e., differencing is needed to achieve stationarity in the series. In addition, it was found that a seasonal pattern exists in the series, with a cycle length of 12 months. To account for this finding, the 12th difference of the first difference ($\nabla\nabla_{12} X_t$) of the residual demand series was calculated. This procedure was undertaken to remove the linear trend and the seasonal pattern.

The sample autocorrelation functions for the series $\nabla\nabla_{12} X_t$ were then examined, and the following tentative model was suggested for further investigation:

$$(1-B)\,(1-B^{12})\,(1-\phi_1 B)\,X_t = (1-\theta_1 B)\,(1-\theta_{12} B^{12})\,a_t. \qquad (25)$$

Using the SAS computer program, the coefficients of the model (25) parametrs were estimated. Results of this estimation are presented in Table 3.

All estimated parameters of the Box-Jenkins model are significant at the 0.10 level or better. The standard error and variance estimates are low, indicating that the model explains relatively well the variation in demand for monthly residual fuel. The application of a Chi-square test to the autocorrelation of the residuals reveals that the residuals are randomly distributed. As a result of the evaluation of the Box-Jenkins model, it is concluded that this model is adequate for forecasting.

Winters Model

The Winters method has been applied to non-utility residual fuel oil demand to produce one-step-ahead forecasts. The first 78 data observations (i.e., July 1975 through December 1981) were used to designate and calculate the smoothing parameters (i.e., Alpha, Beta, and Gamma):

$\alpha = 0.75$
$\beta = 0.10$
$\partial = 0.10.$

These values of the parameters were then used to produce the forecasts for non-utility residual fuel oil, which are the basis for this analysis.

Combined Models

The Bates and Granger methods of combining forecasts discussed in Section 3 have been applied to non-utility residual fuel oil demand to produce one-step-ahead forecasts. The combined models are:

- Econometric model and Box-Jenkins model
- Econometric model and Winters model
- Box-Jenkins model and Winters model

Combined models are developed five different ways, based on the five combining methods discussed in Section 3.

The smoothing parameters used in this paper are:

- $\alpha$ = 0.50, 0.70, and 0.90
- $\beta$ = 1.00, 1.50, 2.0, and 2.50.

The performance of these models is discussed in the next section.

Comparison of Forecasts

Table 4 summarizes the comparisons of alternative forecast methods, presenting several statistical measures of the forecast accuracy which may be used to evaluate the forecasts. With respect to the three initial methods used to forecast demand for non-utility residual fuel oil (i.e., econometric model, Box-Jenkins model, and Winters model), the following conclusions may be drawn:

- Mean Squared Error -- This statistical measure indicates that the Box-Jenkins forecast and the Winters model are preferred to the econometric model. Choosing either of these methods will result in a lower mean squared error than observed for the econometric forecasts.

- Mean Absolute Error and Mean Absolute-Percent Error -- These measures, when applied to the three forecasts, reveals that both a Box-Jenkins model and the Winters model produce forecasts that improve the econometric forecast. Further examination indicates that, on the basis of these measures, the Box-Jenkins model is preferred to the Winters model.

- Thiel's Inequality Coefficient -- This measure indicates that both the Box-Jenkins and the Winters models yield better forecasts than does the econometric model. No differences are observed between these two methods with respect to forecast accuracy.

- Decomposition of the Mean-Squared Error -- The Box-Jenkins method produces the best forecasts, while the Winters model produces forecasts that improve the econometric forecasts.

In the Box-Jenkins model and the Winters model, the five measures of forecast performance considered above indicate that the econometric model does not perform well in comparison with the Box-Jenkins and the Winters models. Overall, the Box-Jenkins model appears to produce the most accurate one-step-ahead forecasts of non-utility residual fuel oil demand.

The data in Table 4 are examined further to evaluate the extent to which the development of combination forecasts, using the alternative forecasts and the alternative weighting methods discussed in Section 3, can improve the forecasts for non-utility residual fuel oil demand. The five statistical measures discussed above have been applied to each of combination forecasts considered. Results of this evaluaton of forecast accuracy indicate that:

1. Combination forecasts can yield more accurate forecasts than the single-method models used for this analysis.

2. The specific combination methods that yield the most accurate forecasts are the Box-Jenkins and Winters model combination, Method 3, $\beta = 2.0$ and $\beta = 2.5$.

A final evaluation of alternative single techniques and combination technique forecast methods is presented in Table 5. This table presents the values of the following ratio:

$$\frac{\text{Mean Squared Error } i}{\text{Mean Squared Error, Econometric}}$$

where i presents single or combination model forecasts. Interpreting this ratio, the degree of improvement in the forecast accuracy possible by choosing method i rather than the econometric model is (1-ratio). Examination of the data in Table 5 again supports the conclusion that the Box-Jenkins and Winters combination, Method 3 yields the greatest opportunity for improvement in forecast accuracy because the ratio is smallest for these alternatives.

## 5. SUMMARY AND CONCLUSIONS

The purpose of this paper has been to examine the potential for improving demand forecasts in the econometric model by developing alternative single and combination forecasting models and by conducting an evaluation of the extent to which these methods do yield more accurate forecasts. The alternative single models examined were the Box-Jenkins model and the Winters model. The forecasts of these models and the econometric model were combined, using five alternative weighting methods, to develop new forecasts that may be able to take advantage of the strengths of the separate forecasts. The statistical measures that had been applied to evaluate the forecasts included:

- Mean squared error
- Mean-absolute error
- Mean-absolute-percent error
- Thiel's inequality coefficient
- Decomposition of mean-squared error.

Examination of the results of the alternative forecasts and application of the statistical measures to the estimates has yielded the following concusions.

1. Of the three single models used to produce short-term forecasts of non-utiity residual fuel oil, the Box-Jenkins model produces the best forecasts. The Winters model also performs better than does the econometric model.

2. The use of combination models can produce better forecasts than the single models examined here. For non-utility residual fuel oil demand, the evaluation of forecast accuracy conducted here indicates that the combination of the Box-Jenkins and Winters models, using Method 3 weighting, yields the most accurate forecast of non-utility residual fuel oil demand.

It is recognized that the usefulness of the one-step-ahead forecast is limited. A methodology that permits the generation of forecast for several steps ahead is preferred. This presents problems for the use of combined forecasts if it cannot be assumed that the single forecasts are stable over time, with respect to accuracy and information of value. Therefore, it is necessary to develop a methodology that permits the weights for each forecast to be time-varying. This is the subject of further research and investigation.

## ACKNOWLEDGEMENT

This paper is based on a recent study by Applied Management Sciences for the Energy Information Administration, Short-Term Information Division, Analysis Branch under Contract No. DE-AC01-80EI-10231. We would like to thank the Analysis Branch, Short-Term Information Division of the Energy Information Administration for its assistance in the re-estimation of the econometric non-utility residual fuel oil demand model; in particular, we thank Mark E. Rodekohr, Chief of the Analysis Branch; Derriel Cato, Project Officer; and Neil Gamson, staff member.

## REFERENCES

1. D. Cochrane, and G.N. Orcutt, "Application of Least Squares to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, 44, 1949.

2. G.E.P. Box and G.M. Jenkins, Time-Series Analysis Forecasting and Control. (San Francisco: Holden-Day, Inc., 1976).

3. D.R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," Management Sciences, April 1960.

4. J.M. Bates, and C.W.J. Granger, "The Combination of Forecasts," Operational Research Quarterly, 20, 1969.

5. H. Theil, Applied Economic Forecasting (Amsterdam: North Holland Publishing Company, 1966).

6. H. Theil, Economic Forecasts and Policy (Amsterdam: North Holland Publishing Company, 1958).

TABLE 2: RESULTS OF ECONOMETRIC FITTED MODEL FOR NON-UTILITY RESIDUAL FUEL OIL DEMAND, JULY 1975 TO DECEMBER 1981

| Variable | Coefficient | t-Ratio |
|---|---|---|
| CONSTANT | 0.713 | 4.721 |
| XDRS (Lag) | 0.170 | 3.583 |
| XPRS | -0.015 | 4.247 |
| INDM | 0.161 | 1.034* |
| GASPR | 0.005 | 2.725 |
| EHDD | 0.001 | 14.975 |
| DUM7677 | 0.149 | 3.932 |
| DUMDEC | -0.261 | 6.744 |
| DUMWIN | -0.242 | 3.432 |

*Not significant at the 0.05 level.

SUMMARY STATISTICS

R-Squared (corrected) = 0.911

Mean-Squared Residual = 0.007

Standard Error of Regression = 0.088

Durbin-Watson Statistics = 2.094

Number of Observations = 78 (July 1975-December 1981)

Rho = 0.308

TABLE 3: RESULTS OF BOX-JENKINS FITTED MODEL FOR NON-UTILITY RESIDUAL FUEL OIL DEMAND, JULY 1975 TO DECEMBER 1981

| Parameter | Coefficient | Standard Error | T-Ratio |
|---|---|---|---|
| AR1($\phi_1$) | 0.439 | 0.290 | 1.51* |
| MA1($\theta_1$) | 0.707 | 0.240 | 2.95** |
| SMA1($\theta_{12}$) | 0.852 | 0.347 | 2.46** |

*Significant at the the 0.10 level.
**Significant at the 0.05 level.

Note: This model is represented by: $(1-B)(1-B^{12})(1-\phi_1 B)\, X_t =$

$(1-\theta_1 B)(1-\theta_{12} B^{12})\, a_t$.

Variance Estimate = 0.014
Standard Error Estimate = 0.120

Autocorrelations of Residual

| Lags | | | | | | |
|---|---|---|---|---|---|---|
| 6 | -0.045 | -0.114 | -0.078 | -0.144 | 0.038 | -0.022 |
| 12 | -0.018 | 0.035 | -0.232 | 0.123 | 0.061 | -0.130 |
| 18 | 0.173 | -0.061 | 0.141 | 0.051 | -0.122 | -0.014 |
| 24 | -0.065 | -0.016 | 0.207 | -0.209 | -0.088 | 0.017 |

Chi-Square = 19.34
Degrees of Freedom = 21

TABLE 4: COMPARISON OF ALTERNATIVE FORECASTING FOR NON-UTILITY RESIDUAL FUEL OIL USING COMBINED METHODS, JANUARY 1982 TO SEPTEMBER 1982

| Statistical Measure | Econ. | Box-Jenkins | Winters | Combined Method 1: Econ. and Box-Jenkins | Combined Method 1: Econ. and Winters | Combined Method 1: Box-Jenkins and Winters | Combined Method 2 ($\alpha = 0.50$): Econ. and Box-Jenkins | Combined Method 2 ($\alpha = 0.50$): Econ. and Winters | Combined Method 2 ($\alpha = 0.50$): Box-Jenkins and Winters |
|---|---|---|---|---|---|---|---|---|---|
| Mean of Actual | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 |
| Mean of Forecast | 1.222 | 1.081 | 1.081 | 1.165 | 1.161 | 1.082 | 1.161 | 1.160 | 1.081 |
| Mean of Error | -0.110 | 0.030 | 0.030 | -0.054 | -0.050 | 0.029 | -0.021 | -0.049 | 0.030 |
| Variance of Actual | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 |
| Variance of Forecast | 0.013 | 0.033 | 0.023 | 0.019 | 0.015 | 0.028 | 0.018 | 0.014 | 0.028 |
| Variance of Error | 0.019 | 0.027 | 0.027 | 0.015 | 0.017 | 0.026 | 0.021 | 0.019 | 0.026 |
| Standard Deviation of Actual | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 |
| Standard Deviation of Forecast | 0.114 | 0.183 | 0.151 | 0.138 | 0.123 | 0.166 | 0.133 | 0.118 | 0.166 |
| Standard Deviation of Error | 0.138 | 0.165 | 0.164 | 0.123 | 0.130 | 0.161 | 0.144 | 0.137 | 0.163 |
| Mean-Absolute Error | 0.153 | 0.118 | 0.120 | 0.118 | 0.119 | 0.118 | 0.127 | 0.126 | 0.119 |
| Mean-Absolute-Percent Error | 14.946 | 10.013 | 10.327 | 10.893 | 11.081 | 10.061 | 11.738 | 11.645 | 10.116 |
| Mean-Squared Error | 0.029 | 0.025 | 0.025 | 0.016 | 0.017 | 0.024 | 0.019 | 0.019 | 0.024 |
| Root-Mean-Squared Error | 0.171 | 0.159 | 0.159 | 0.127 | 0.130 | 0.155 | 0.137 | 0.137 | 0.156 |
| Thiel's Inequality Coefficient | 0.023 | 0.020 | 0.020 | 0.013 | 0.013 | 0.019 | 0.015 | 0.015 | 0.019 |
| Correlation Coefficient | 0.84 | 0.67 | 0.66 | 0.86 | 0.86 | 0.68 | 0.81 | 0.83 | 0.67 |
| Decomposition of MSE | | | | | | | | | |
| $U^M$ (Bias Proportion) | 0.42 | 0.04 | 0.04 | 0.18 | 0.15 | 0.04 | 0.13 | 0.13 | 0.04 |
| $U^R$ (Regression Proportion) | 0.38 | 0.05 | 0.19 | 0.41 | 0.54 | 0.12 | 0.39 | 0.54 | 0.12 |
| $U^D$ (Disturbance Proportion) | 0.20 | 0.91 | 0.77 | 0.41 | 0.45 | 0.84 | 0.48 | 0.33 | 0.84 |

TABLE 4: COMPARISON OF ALTERNATIVE FORECASTING FOR NON-UTILITY RESIDUAL FUEL OIL USING COMBINED METHODS, JANUARY 1982 TO SEPTEMBER 1982 (Continued)

| Statistical Measure | Combined Method 2 ($\alpha = 0.70$) | | | Combined Method 2 ($\alpha = 0.90$) | | | Combined Method 3 ($\beta = 1.0$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters |
| Mean of Actual | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 |
| Mean of Forecast | 1.158 | 1.159 | 1.081 | 1.157 | 1.159 | 1.081 | 1.121 | 1.127 | 1.101 |
| Mean of Error | -0.048 | -0.048 | 0.030 | -0.046 | -0.048 | 0.031 | -0.010 | -0.016 | 0.010 |
| Variance of Actual | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 |
| Variance of Forecast | 0.018 | 0.014 | 0.027 | 0.017 | 0.013 | 0.027 | 0.034 | 0.029 | 0.030 |
| Variance of Error | 0.020 | 0.020 | 0.027 | 0.022 | 0.020 | 0.027 | 0.013 | 0.016 | 0.018 |
| Standard Deviation of Actual | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 |
| Standard Deviation of Forecast | 0.132 | 0.116 | 0.166 | 0.132 | 0.115 | 0.166 | 0.185 | 0.169 | 0.174 |
| Standard Deviation of Error | 0.141 | 0.140 | 0.163 | 0.148 | 0.143 | 0.164 | 0.115 | 0.127 | 0.135 |
| Mean-Absolute Error | 0.131 | 0.128 | 0.119 | 0.135 | 0.131 | 0.119 | 0.099 | 0.105 | 0.104 |
| Mean-Absolute-Percent Error | 12.082 | 11.884 | 10.161 | 12.408 | 12.089 | 10.175 | 8.901 | 9.384 | 9.178 |
| Mean-Squared Error | 0.020 | 0.020 | 0.025 | 0.022 | 0.020 | 0.025 | 0.012 | 0.015 | 0.016 |
| Root-Mean-Squared Error | 0.142 | 0.140 | 0.157 | 0.147 | 0.143 | 0.157 | 0.110 | 0.122 | 0.126 |
| Thiel's Inequality Coefficient | 0.016 | 0.016 | 0.020 | 0.017 | 0.016 | 0.020 | 0.009 | 0.012 | 0.013 |
| Correlation Coefficient | 0.79 | 0.82 | 0.067 | 0.75 | 0.81 | 0.67 | 0.85 | 0.81 | 0.79 |
| Decomposition of MSE | | | | | | | | | |
| $U^M$ (Bias Proportion) | 0.11 | 0.12 | 0.04 | 0.10 | 0.12 | 0.04 | 0.01 | 0.02 | 0.01 |
| $U^R$ (Regression Proportion) | 0.38 | 0.53 | 0.11 | 0.34 | 0.54 | 0.11 | 0.10 | 0.17 | 0.13 |
| $U^D$ (Disturbance Proportion) | 0.51 | 0.35 | 0.85 | 0.56 | 0.34 | 0.85 | 0.89 | 0.81 | 0.86 |

TABLE 4: COMPARISON OF ALTERNATIVE FORECASTING FOR NON-UTILITY RESIDUAL FUEL OIL USING COMBINED METHODS, JANUARY 1982 TO SEPTEMBER 1982 (Continued)

| Statistical Measure | Combined Method 3 ($\beta$ = 1.5) | | | Combined Method 3 ($\beta$ = 2.0) | | | Combined Method 3 ($\beta$ = 2.50) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters |
| Mean of Actual | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 |
| Mean of forecast | 1.092 | 1.098 | 1.119 | 1.074 | 1.084 | 1.127 | 1.065 | 1.080 | 1.131 |
| Mean of Error | 0.019 | 0.013 | -0.008 | 0.037 | 0.027 | -0.016 | 0.046 | 0.031 | -0.020 |
| Variance of Actual | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 |
| Variance of Forecast | 0.055 | 0.054 | 0.037 | 0.068 | 0.070 | 0.048 | 0.074 | 0.078 | 0.061 |
| Variance of Error | 0.015 | 0.018 | 0.010 | 0.018 | 0.020 | 0.006 | 0.020 | 0.022 | 0.006 |
| Standard Deviation of Actual | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 |
| Standard Deviation of Forecast | 0.234 | 0.233 | 0.191 | 0.260 | 0.264 | 0.220 | 0.272 | 0.279 | 0.246 |
| Standard Deviation of Error | 0.124 | 0.133 | 0.100 | 0.136 | 0.143 | 0.078 | 0.141 | 0.148 | 0.079 |
| Mean-Absolute Error | 0.104 | 0.106 | 0.081 | 0.112 | 0.119 | 0.056 | 0.116 | 0.123 | 0.059 |
| Mean-Absolute-Percent Error | 9.532 | 9.754 | 7.329 | 10.594 | 11.474 | 5.207 | 11.101 | 12.115 | 5.496 |
| Mean-Squared Error | 0.014 | 0.016 | 0.009 | 0.018 | 0.019 | 0.006 | 0.020 | 0.020 | 0.006 |
| Root-Mean-Squared Error | 0.118 | 0.126 | 0.095 | 0.134 | 0.138 | 0.077 | 0.141 | 0.141 | 0.077 |
| Thiel's Inequality Coefficient | 0.011 | 0.013 | 0.007 | 0.014 | 0.015 | 0.005 | 0.016 | 0.016 | 0.005 |
| Correlation Coefficient | 0.85 | 0.83 | 0.89 | 0.85 | 0.84 | 0.94 | 0.86 | 0.85 | 0.95 |
| Decomposition of MSE | | | | | | | | | |
| $U^M$ (Bias Proportion) | 0.03 | 0.01 | 0.01 | 0.08 | 0.04 | 0.04 | 0.11 | 0.05 | 0.07 |
| $U^R$ (Regression Proportion) | 0.02 | 0.01 | 0.09 | 0.09 | 0.11 | 0.00 | 0.14 | 0.18 | 0.12 |
| $U^D$ (Disturbance Proportion) | 0.95 | 0.98 | 0.90 | 0.83 | 0.85 | 0.96 | 0.75 | 0.77 | 0.81 |

TABLE 4: COMPARISON OF ALTERNATIVE FORECASTING FOR NON-UTILITY RESIDUAL FUEL OIL USING COMBINED METHODS, JANUARY 1982 TO SEPTEMBER 1982 (Continued)

| Statistical Measure | Combined Method 4 ($\beta$ = 1.0) | | | Combined Method 4 ($\beta$ = 1.5) | | | Combined Method 4 ($\beta$ = 2.0) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters |
| Mean of Actual | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 |
| Mean of Forecast | 1.165 | 1.161 | 1.082 | 1.158 | 1.153 | 1.082 | 1.153 | 1.149 | 1.082 |
| Mean of Error | -0.054 | -0.050 | 0.029 | -0.047 | -0.042 | 0.029 | -0.042 | -0.038 | -0.029 |
| Variance of Actual | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 |
| Variance of Forecast | 0.019 | 0.019 | 0.028 | 0.025 | 0.021 | 0.028 | 0.029 | 0.024 | 0.028 |
| Variance of Error | 0.0[illegible] | 0.017 | 0.026 | 0.013 | 0.014 | 0.026 | 0.012 | 0.014 | 0.026 |
| Standard Deviation of Actual | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 |
| Standard Deviation of Forecast | 0.138 | 0.123 | 0.166 | 0.158 | 0.144 | 0.166 | 0.171 | 0.156 | 0.167 |
| Standard Deviation of Error | 0.123 | 0.130 | 0.161 | 0.112 | 0.120 | 0.161 | 0.109 | 0.116 | 0.160 |
| Mean-Absolute Error | 0.118 | 0.119 | 0.118 | 0.103 | 0.104 | 0.117 | 0.099 | 0.100 | 0.117 |
| Mean-Absolute-Percent Error | 10.900 | 11.066 | 10.054 | 9.382 | 9.497 | 10.021 | 8.973 | 9.132 | 9.977 |
| Mean-Squared Error | 0.016 | 0.017 | 0.024 | 0.013 | 0.015 | 0.024 | 0.012 | 0.013 | 0.024 |
| Root-Mean-Squared Error | 0.127 | 0.130 | 0.155 | 0.114 | 0.122 | 0.155 | 0.110 | 0.114 | 0.155 |
| Theil's Inequality Coefficient | 0.013 | 0.013 | 0.019 | 0.010 | 0.012 | 0.019 | 0.009 | 0.010 | 0.019 |
| Correlation Coefficient | 0.86 | 0.86 | 0.68 | 0.87 | 0.86 | 0.68 | 0.87 | 0.86 | 0.69 |
| Decomposition of MSE | | | | | | | | | |
| $U^M$ (Bias Proportion) | 0.18 | 0.15 | 0.04 | 0.17 | 0.12 | 0.04 | 0.15 | 0.11 | 0.04 |
| $U^R$ (Regression Proportion) | 0.41 | 0.54 | 0.12 | 0.29 | 0.38 | 0.12 | 0.19 | 0.31 | 0.11 |
| $U^D$ (Disturbance Proportion) | 0.41 | 0.31 | 0.84 | 0.54 | 0.51 | 0.84 | 0.66 | 0.58 | 0.85 |

TABLE 4: COMPARISON OF ALTERNATIVE FORECASTING FOR NON-UTILITY RESIDUAL FUEL OIL USING COMBINED METHODS, JANUARY 1982 TO SEPTEMBER 1982 (Continued)

| Statistical Measure | Combined Method 4 ($\beta$ = 2.5) | | | Combined Method 5 ($\alpha$ = 0.50) | | | Combined Method 5 ($\alpha$ = 0.70) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters | Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters |
| Mean of Actual | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 | 1.111 |
| Mean of Forecast | 1.151 | 1.147 | 1.081 | 1.141 | 1.146 | 1.078 | 1.138 | 1.144 | 1.078 |
| Mean of Error | -0.040 | -0.036 | 0.030 | -0.030 | -0.035 | 0.033 | -0.027 | -0.033 | 0.033 |
| Variance of Actual | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 |
| Variance of Forecast | 0.031 | 0.026 | 0.628 | 0.026 | 0.021 | 0.028 | 0.026 | 0.020 | 0.028 |
| Variance of Error | 0.012 | 0.013 | 0.025 | 0.014 | 0.015 | 0.026 | 0.016 | 0.017 | 0.026 |
| Standard Deviation of Actual | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 | 0.219 |
| Standard Deviation of Forecast | 0.177 | 0.161 | 0.167 | 0.163 | 0.145 | 0.166 | 0.161 | 0.142 | 0.166 |
| Standard Deviation of Error | 0.108 | 0.115 | 0.159 | 0.117 | 0.124 | 0.160 | 0.127 | 0.131 | 0.160 |
| Mean-Absolute Error | 0.096 | 0.098 | 0.116 | 0.105 | 0.108 | 0.116 | 0.115 | 0.116 | 0.116 |
| Mean-Absolute-Percent Error | 8.700 | 8.908 | 9.936 | 9.368 | 9.850 | 9.867 | 10.269 | 10.563 | 0.099 |
| Mean-Squared Error | 0.012 | 0.013 | 0.023 | 0.013 | 0.015 | 0.024 | 0.015 | 0.016 | 0.024 |
| Root-Mean-Squared Error | 0.110 | 0.114 | 0.152 | 0.114 | 0.122 | 0.155 | 0.122 | 0.126 | 0.155 |
| Theil's Inequality Coefficient | 0.009 | 0.010 | 0.018 | 0.010 | 0.012 | 0.019 | 0.012 | 0.013 | 0.019 |
| Correlation Coefficient | 0.87 | 0.86 | 0.69 | 0.85 | 0.84 | 0.68 | 0.82 | 0.82 | 0.68 |
| Decomposition of MSE | | | | | | | | | |
| $U^M$ (Bias Proportion) | 0.13 | 0.10 | 0.04 | 0.07 | 0.00 | 0.05 | 0.05 | 0.07 | 0.05 |
| $U^R$ (Regression Proportion) | 0.15 | 0.26 | 0.12 | 0.24 | 0.37 | 0.12 | 0.22 | 0.37 | 0.12 |
| $U^D$ (Disturbance Proportion) | 0.72 | 0.64 | 0.84 | 0.69 | 0.63 | 0.83 | 0.73 | 0.56 | 0.83 |

TABLE 4: COMPARISON OF ALTERNATIVE FORECASTING FOR NON-UTILITY RESIDUAL FUEL OIL USING COMBINED METHODS, JANUARY 1982 TO SEPTEMBER 1982 (Continued)

| Statistical Measure | Combined Method 5 ($\alpha = 0.90$) Econ. and Box-Jenkins | Econ. and Winters | Box-Jenkins and Winters |
|---|---|---|---|
| Mean of Actual | 1.111 | 1.111 | 1.111 |
| Mean of Forecast | 1.135 | 1.142 | 1.077 |
| Mean of Error | -0.024 | -0.031 | 0.033 |
| Variance of Actual | 0.048 | 0.048 | 0.048 |
| Variance of Forecast | 0.026 | 0.020 | 0.028 |
| Variance of Error | 0.019 | 0.019 | 0.026 |
| Standard Deviation of Actual | 0.219 | 0.219 | 0.219 |
| Standard Deviation of Forecast | 0.161 | 0.140 | 0.166 |
| Standard of Deviation of Error | 0.139 | 0.139 | 0.161 |
| Mean-Absolute Error | 0.125 | 0.124 | 0.117 |
| Mean-Absolute-Percent Error | 11.170 | 11.275 | 0.155 |
| Mean-Squared Error | 0.018 | 0.018 | 0.024 |
| Root-Mean-Squared Error | 0.134 | 0.134 | 10.175 |
| Thiel's Inequality Coefficient | 0.014 | 0.014 | 0.019 |
| Correlation Coefficient | 0.77 | 0.79 | 0.68 |
| Decomposition of MSE | | | |
| $U^M$ (Bias Proportion) | 0.03 | 0.05 | 0.05 |
| $U^R$ (Regression Proportion) | 0.19 | 0.35 | 0.12 |
| $U^D$ (Distribution Proportion | 0.78 | 0.60 | 0.83 |

TABLE 5: MEAN-SQUARED ERROR RATIOS FOR ALTERNATIVE NON-UTILITY RESIDUAL FUEL OIL MODELS AND ECONOMETRIC MODEL, JANUARY 1982 TO SEPTEMBER 1982

| | Ratio | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Alternative Models | No $\alpha$ or $\beta$ | $\alpha$=0.50 | $\alpha$=0.70 | $\alpha$=0.90 | $\beta$=1 | $\beta$=1.5 | $\beta$=2.0 | $\beta$=2.5 |
| Box-Jenkins | 0.86 | -- | -- | -- | -- | -- | -- | -- |
| Winters | 0.86 | -- | -- | -- | -- | -- | -- | -- |
| Combined | | | | | | | | |
| Method 1 | | | | | | | | |
| Econ. & Box-Jenkins | 0.55 | -- | -- | -- | -- | -- | -- | -- |
| Econ. & Winters | 0.59 | -- | -- | -- | -- | -- | -- | -- |
| Box-Jenkins & Winters | 0.83 | -- | -- | -- | -- | -- | -- | -- |
| Method 2 | | | | | | | | |
| Econ. & Box-Jenkins | -- | 0.66 | 0.69 | 0.76 | -- | -- | -- | -- |
| Econ. & Winters | -- | 0.66 | 0.69 | 0.69 | -- | -- | -- | -- |
| Box-Jenkins & Winters | -- | 0.83 | 0.86 | 0.86 | -- | -- | -- | -- |
| Method 3 | | | | | | | | |
| Econ. & Box-Jenkins | -- | -- | -- | -- | 0.41 | 0.48 | 0.62 | 0.69 |
| Econ. & Winters | -- | -- | -- | -- | 0.52 | 0.55 | 0.66 | 0.69 |
| Box-Jenkins & Winters | -- | -- | -- | -- | 0.55 | 0.31 | 0.21* | 0.21* |
| Method 4 | | | | | | | | |
| Econ. & Box-Jenkins | -- | -- | -- | -- | 0.55 | 0.45 | 0.41 | 0.41 |
| Econ. & Winters | -- | -- | -- | -- | 0.59 | 0.52 | 0.45 | 0.45 |
| Box-Jenkins & Winters | -- | -- | -- | -- | 0.83 | 0.83 | [illegible]3 | 0.79 |
| Method 5 | | | | | | | | |
| Econ. & Box-Jenkins | -- | 0.45 | 0.52 | 0.62 | -- | -- | -- | -- |
| Econ. & Winters | -- | 0.52 | 0.55 | 0.62 | -- | -- | -- | -- |
| Box-Jenkins & Winters | -- | 0.83 | 0.83 | 0.83 | -- | -- | -- | -- |

*Lowest ratio.

# STATISTICAL PREDICTION OF INSTANTANEOUS RESIDENTIAL ELECTRIC DEMAND

by

D. J. Anderson, G. J. Collaros, and E. W. Enlow
The BDM Corporation
1801 Randolph Road S.E.
Albuquerque, NM 87106

(Contributed Paper Session 1)

# STATISTICAL PREDICTION OF INSTANTANEOUS RESIDENTIAL ELECTRIC DEMAND*

D. J. Anderson, G. J. Collaros, E. W. Enlow
The BDM Corporation, 1801 Randolph Road S.E., Albuquerque, New Mexico 87106
(505)848-5000

## ABSTRACT

Use of time average residential electric load data has been assumed to significantly under- or over-estimate photovoltaic system performance measures. Based on a large sample of instantaneous electric load data, a statistical model was developed which converts real time-average load data into estimated instantaneous data. Periodic regression analysis, cluster analysis, and classification analysis were performed to characterize the instantaneous load data. The model was structured into an algorithm which was incorporated into a photovoltaic systems analysis simulation code. This simulation model was exercised to investigate the effects of instantaneous versus time-average load data, with varying parameters, on residential photovoltaic system performance.

## INTRODUCTION

A residential load profile is the pattern of electric use in a residence over a period of time. Utilities routinely monitor residences and use these profiles to analyze the demand for electricity by a particular class of user. When a utility constructs a profile, the time interval used is usually between 15 minutes and 1 hour. While this interval is adequate for utility requirements, it is too large to reveal the instantaneous behavior of the load. It is generally assumed that use of time-average data in residential photovoltaic (PV) system simulations significantly underestimates the interaction between a PV system and the utility. The designer of a PV system must be able to predict the total energy that will be produced by an array and the fractions of energy that will be used directly by the load or purchased from (or sold to) the utility. These performance predictions are required to properly size a system to meet demand and to determine the system's economic worth. Prior performance predictions of residential PV power systems had to be conducted using models that could accept only hourly time-averaged load data, which does not reveal the instantaneous load behavior caused by residential appliance usage.

Issues of using instantaneous load data versus time-averaged data to predict energy fractions can be simply illustrated by figure 1. This figure shows a hypothetical PV system power output, assumed to be constant, and a load at three different time intervals: instantaneous, 15 minute, and 1 hour. As the time interval becomes larger, the amount of variability in the load declines until, at the longer time interval, little interaction takes place between the PV array, the residential load, and the utility.

*Work performed under contract 62-3977 to Sandia National Laboratories, Albuquerque, New Mexico.

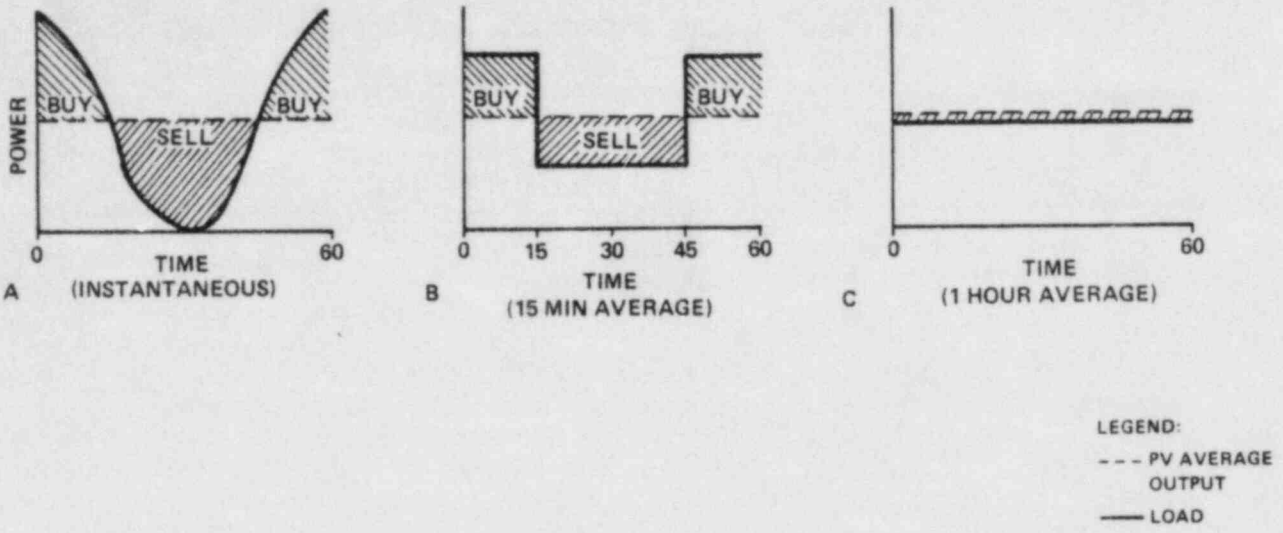



Figure 1. Time-Averaged Versus Instantaneous Load Profile

Recent studies of the effects of instantaneous loads have used statistical approaches involving the complete characterization of the residential load based on the probabi ties of various household appliance usage for average durations and individual appliance load requirements (references 1 and 2). These statistical studies showed instantaneous loads to have an effect on annual residential PV system performance. However, the numerous appliance on/off probability estimates required for these studies add uncertainty to the results. Adequate samples of instantaneous residential loads did not exist to verify these results.

Objectives of the Detailed Residential Electric Load Determination Program were to measure real, instantaneous load data, and to develop a statistical model based on these measured data. Since utilities throughout the nation currently collect residential load data at 15-minute intervals, the model was designed to predict instantaneous residential electric load from 15-minute time-averaged data. With this large data base, the statistical model can be used to predict instantaneous load demand for a variety of PV applications.

Four residences in three regions of the country were instrumented and the load measured at 5-second intervals. Each residence was continuously monitored for 2 weeks during the summer and winter to obtain seasonal load variations. The load data were statistically analyzed and incorporated into a probabilistic model. Annual performance predictions were then conducted to determine the amount of energy supplied by a PV array to the load. To accomplish the performance predictions, the statistical model was incorporated into SOLCEL, a Sandia National Laboratories (SNLA) computer simulation code designed specifically for the analysis of PV systems.

## DATA COLLECTION

Data acquisition hardware for direct measurement of instantaneous load consisted of measurement equipment, located at each of four monitored residences, and a computer controller. The complete measurement system consisted of current loop transformers connected to a Scientific Columbus Joule Electronic Meter (JEM), which measured load data at 5-second intervals in units of kVAh. After

digitization within the JEM, the accumulated data were collected at regular intervals by the computer controller via telephone lines. Data were stored on floppy disks and later transferred to the SNLA mainframe computer system.

The Public Service Company of New Mexico, Georgia Power of Atlanta, Georgia, and the Public Service Electric and Gas Company of Newark, New Jersey, were selected for participation in the program. Residences were selected by the utilities to obtain a typical residence from their current residential monitoring program. Selection was made on the basis of electrical consumption that approximated the sample mean for residences from the utility load data bases.

Residential electric load data were directly measured at 5-second intervals on the four residences selected in the northeast, southeast, and southwest regions of the country. Sample instantaneous load profiles, along with the 15-minute time averages, are illustrated in figure 2. A load data base consisting of 5,080 complete, continuous, 15-minute intervals was established (1,270 hours or 53 days of data). Load data were collected for 2-week periods during the summer and again for 2-week periods in the winter at the selected residences. To date, this data base is the largest source of instantaneous load data available. Data collected for each residence is summarized in table 1.

TABLE 1. DATA COLLECTION BY RESIDENCE

| | Hours of Load Data Collection | | | |
|---|---|---|---|---|
| | ALB-1 | ALB-2 | Atlanta | Newark |
| Summer | 215.5 | 224.5 | 149.5 | 190.5 |
| Winter | 146.5 | 104.0 | 155.5 | 84.0 |
| Total | 362.0 | 328.5 | 305.0 | 274.5 |

## LOAD MODEL DEVELOPMENT

Initially, the instantaneous load model was planned to involve multiple linear regression prediction of peaks (load spikes), peak durations, and frequencies per 15-minute interval, based on the time-averaged load. However, analysis indicated that the instantaneous profiles, or patterns, are driven by major appliances only, and that smaller appliances simply added to the base load. The peak loads, durations, and frequencies were thus based on particular appliance load profiles.

A new modeling approach was undertaken which involved three basic stages plus the final prediction stage:

(1) Instantaneous load profile characterization for each 15-minute interval.

(2) Determination of the most prevalent profile groups based on analysis of the characterization measures.

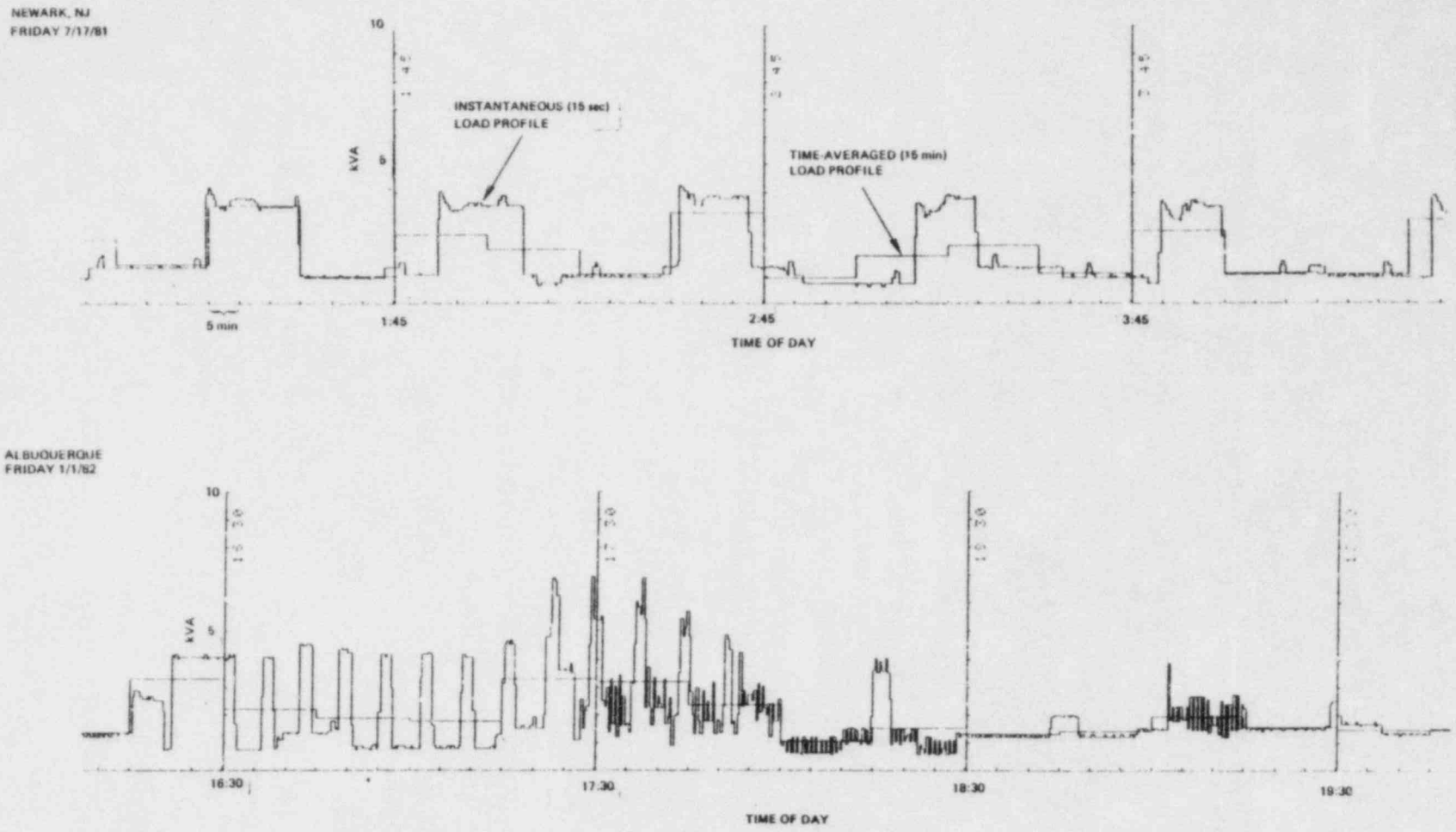



Figure 2. Sample Instantaneous Load Profiles

(3) Representation of the prevalent profile groups.

(4) Prediction of which profile group representation to use based on the time average.

Characterization of the 15-minute interval load profiles was accomplished by periodic regression analysis, a technique applicable to periodic data and similar to polynomial regression and Fourier analysis. This analysis uses functions of sine and cosine curves as independent variables in a multiple regression equation, the coefficients of which characterize the instantaneous profile.

Periodic regression analysis was first performed on 4-second interval data collected during the initial stages of the direct measurement program. An early conclusion was that data could be aggregated to 16-second intervals for modeling. Doing so reduced the roundoff effects in the data, adequately maintained the instantaneous profile information, and advantageously reduced the amount of data to be analyzed by periodic regression for each 15-minute interval.

An example of the regression analysis results, from program BMDP1R (reference 3), and the coefficient estimates for a selected 15-minute interval is shown in figure 3. Three pairs of sine and cosine functions of the timeline variable, X, were included in the periodic regression example to form three harmonics. The high squared multiple correlation coefficient of 0.936 indicates the periodic regression model fits the data quite well, with only three harmonics. Coefficients for the sine and cosine pairs define the presence and

MULTIPLE R 0.9676
MULTIPLE R-SQUARE 0.9362
STD. ERROR OF EST. 0.1754

ANALYSIS OF VARIANCE

| | SUM OF SQUARES | DF | MEAN SQUARE | F RATIO |
|---|---|---|---|---|
| REGRESSION | 22.149336 | 6 | 3.691556 | 119.923 |
| RESIDUAL | 1.5083482 | 49 | 0.3078262E-01 | |

VARIABLES IN EQUATION

| VARIABLE | COEFFICIENT | STD. ERROR OF COEFF | STD. REG COEFF | F TO REMOVE |
|---|---|---|---|---|
| Y-INTERCEPT | 1.232 | | | |
| COSX | 0.256 | 0.033 | 0.279 | 59.626 |
| SINX | 0.791 | 0.033 | 0.860 | 569.068 |
| COS2X | -0.140 | 0.033 | -0.152 | 17.820 |
| SIN2X | 0.204 | 0.033 | 0.222 | 37.826 |
| COS3X | 0.028 | 0.033 | 0.030 | 0.695 |
| SIN3X | 0.195 | 0.033 | 0.212 | 34.506 |

Figure 3. Periodic Regression Example with Three Harmonics

magnitude of the corresponding harmonic in the equation. The Y-intercept value is the 15-minute interval average load. Due to the orthogonality of the sine and cosine functions, individual coefficient estimates remain the same, no matter how many other harmonics are in the equation. The orthogonality also results in identical standard errors of the coefficients in the equation. Coefficients for each sine and cosine pair, say a and b respectively, can be combined to provide the amplitude, A, and phase angle, $\theta$, of the sine curve harmonic by the following relations:

$$A = \sqrt{a^2 + b^2}$$

$$\tan \theta = a/b$$

The 15-minute interval contains 56 intervals 16 seconds long, with the last data point excluded. This corresponds to the degrees of freedom shown in the figure. A plot of the observed data (O) for the 15-minute interval in the example and the predicted values (P) from the periodic regression equation in the previous figure is presented in figure 4. The plot displays the adequacy of the periodic regression characterization of the instantaneous profile.

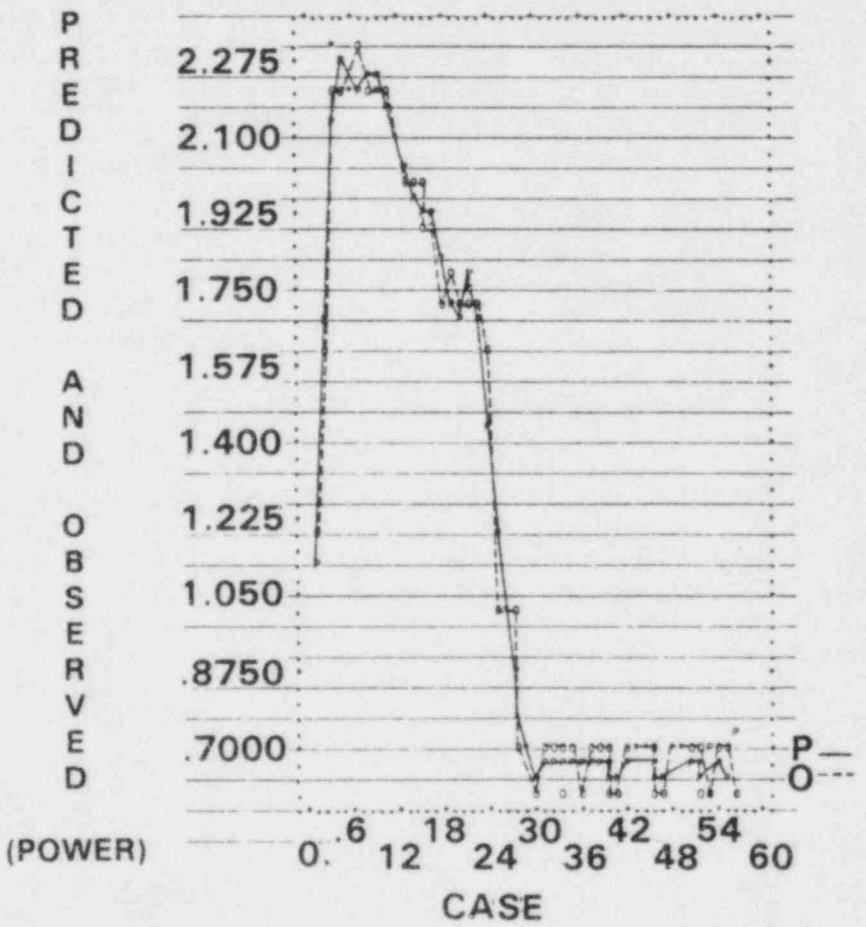



Figure 4. Plot of Predicted and Observed Load Versus Time

Equipment calibration and check-out were accomplished during the initial direct measurement phase. Prior to collection of the actual load data base, changes made to the monitoring instrumentation to reduce rounding errors present in low readings necessitated lengthening the sampling interval to 5 seconds.

Prior to performing periodic regression on each of the intervals in the load data base, the 5-second interval data were aggregated into 15-second interval load values, for the reasons previously stated, producing 60 measurements per 15-minute interval. Since no prior information existed on how many harmonics were necessary to adequately model the highly variable load profiles, it was decided to use eight harmonics in each periodic regression analysis. The use of eight harmonics was considered adequate for most intervals and was also an upper bound in consideration of the large volume of information which would require file storage.

Periodic regression analysis was performed on each of the 5,080 15-minute intervals in the load data base using IMSL (reference 4) subroutines. Coefficients from each of the periodic regression equations and descriptive information relating the house, season, day, date, time, minimum, maximum, energy above average, squared multiple correlation, standard error of estimate, and standard error of coefficient were concurrently stored on file for further analysis. With 72.5 percent of the equations having squared multiple correlations ($R^2$) greater than 70.0 percent, the periodic regression stage was considered to have completely accomplished the goal of adequate profile characterization.

Three criteria were then used to determine whether profiles were sufficiently active to require further modeling: time of day, energy above average, and the squared multiple correlation. Very little load activity was observed between midnight and 5:30 a.m. The energy above average measures the energy (area) below the instantaneous profile and above the 15-minute average, and thus relates the activity of the profile. Active profiles would necessarily have high energy above average values. Squared multiple correlation measures the amount of variability accounted for by the regression equation in contrast to a horizontal line. Thus, a low correlation indicates that the time average value (Y-intercept or constant term in the equation) is adequate for modeling and that the profile is inactive.

Plots of energy above average versus quartile (15-minute time interval of the day) and squared multiple correlation for the 5,080 intervals are presented in figures 5 and 6 respectively. Also indicated on the plots are the cut-off levels for the three criteria used to define active profiles:

(1) Quartiles greater than 22 (5:30 - 24:00) (figure 5)

(2) Energy above average greater than .06 kVAh (figures 5 and 6)

(3) Squared multiple correlation greater than or equal to 70 percent (figure 6).

The breakdown of the 5,080 intervals according to the cut-off levels of the criteria is shown in table 2. These cut-off levels were determined by observation of the figures and evaluation of the possible outlier intervals. (A more

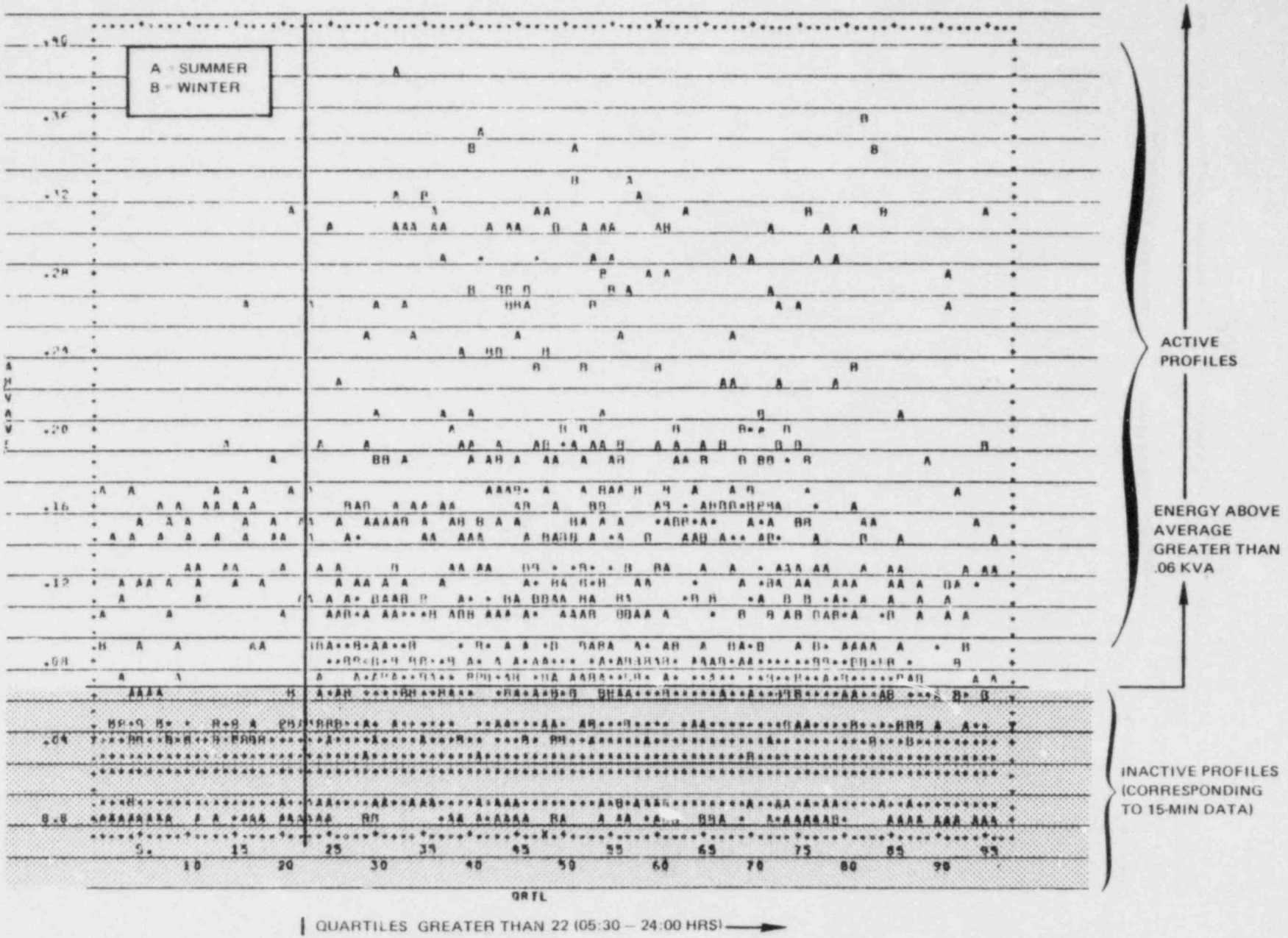



Figure 5. Plot of Energy Above Average (ABAVE) Versus Quartile (QTRL)

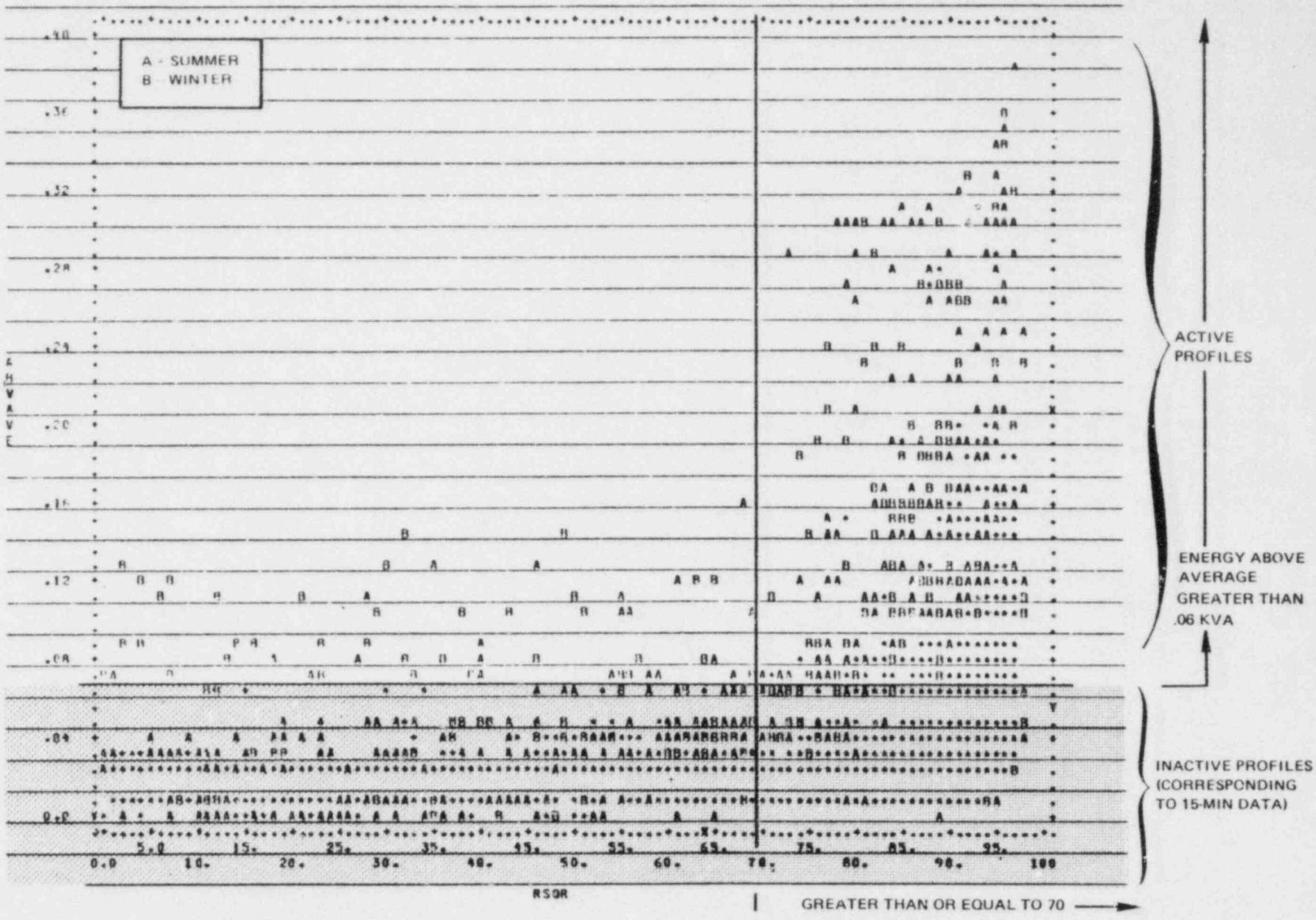



Figure 6. Plot of Energy Above Average (ABAVE) Versus R-Squared (RSQR)

detailed discussion is presented in reference 5.) The three criteria resulted in classifying 734 out of the 5,080 15-minute interval profiles as being active.

TABLE 2. BREAKDOWN OF INTERVALS BY CRITERIA FOR ACTIVE PROFILES

| ENERGY ABOVE AVERAGE | QUARTILE | CORRELATION | |
|---|---|---|---|
| | | LT 70 | GE 70 |
| LE .06 | LE 22 | 313 | 801 |
| | GT 22 | 1045 | 2069 |
| GT .06 | LE 22 | 0 | 60 |
| | GT 22 | 58 | 734 |

NOTE:

GE - GREATER THAN OR EQUAL TO

GT - GREATER THAN

LE - LESS THAN OR EQUAL TO

LT - LESS THAN

Determination of the most prevalent types of active load profiles was the next analysis goal. Based on the premise that active profiles are mostly due to a few major appliances, load profiles were grouped according to the coefficients of the characteristic periodic regression equations. Even though base loads may be different, resulting in different 15-minute average load values, load profiles due to a particular appliance should be similar. In fact, each distinct profile group developed could probably be associated with an individual appliance load or combination of appliance loads.

Cluster analysis and contingency table layouts were used to examine the groupings resulting from various criteria. In cluster analysis, values of the coefficients are recoded as belonging to specified intervals, and cases (profiles) having common codes for the coefficients are grouped together. The specification of the intervals for each coefficient and the number of coefficients included greatly affected the groupings, resulting in a large number of trial and error analyses. Output from the BMDP3M cluster analysis program is quite lengthy and is not included here.

Initially, the coefficient value for the sine and cosine functions were used to group the intervals. Several problems were encountered involving the common scale of the variables, symmetry of the variables about zero, and the large number of variables (up to 17) and cases (734). It was decided to transform the coefficients into the harmonic amplitude and phase angle values and to reduce the number of harmonics used to cluster the profiles. Intervals for the amplitudes were reduced to four: low, medium, high, and very high. Phase angles were not included individually, but combined to form phase angle differences between pairs of harmonics when both were in at least the medium category. The phase angle defines where along the time axis the sine curve lies. This location was not considered important since it is dependent upon when exactly the data collection occurred and when appliances were turned on. However, when two or more harmonics were at the medium or higher categories, the phase angle difference between the pairs of harmonics did aid in grouping the profiles. Phase angle differences were placed in three categories: in phase, between phase, and out of phase.

Several analyses using this structure led to a good grouping of the profiles. One conclusion was that only the first two harmonics were needed to group the profiles. The final grouping structure is summarized in a contingency table layout shown in table 3. Displayed in the table are the number of 15-minute interval profiles in each group and the group classification number. Thirteen different profiles groups were identified, with groups 10 and 12 being a combination of two somewhat similar groups and group 13 being somewhat of a "grab bag" group. All profiles not falling into a distinct group were placed in group 13. With the distinct profile groups defined, remaining modeling steps were to form representations of the groups and to develop a prediction or selection strategy.

To construct the statistical model predicting instantaneous load profiles, precise definition of the predicted profiles (and therefore the profile groups) was necessary. Mathematical formulation of the profiles using the sine curve (harmonic) amplitude and phase angle differences was considered, but was found to smooth out the profiles. In view of how the SOLCEL PV analysis code operates, and with the desire to maintain the instantaneous nature of the profiles, it was decided to use the interval of real instantaneous data which was closest to the average or centroid of the profile group. That is, real data profiles were used to represent the distinct profile groups.

Mathematical averages of the first four harmonic amplitudes and the difference between the first two harmonic phase angles were calculated for each profile group. From the interval profiles in a particular group, the interval with amplitudes and the phase angle difference closest to the averages for that group was determined. Closest was defined by several criteria: minimum sum of squared deviations, minimum sum of absolute deviations, and two different weighting schemes used to weight the deviations from the amplitudes more than the deviation from the phase angle difference. The representative intervals chosen matched on at least three of the above four criteria for all the groups.

The final load modeling step was to develop a rule for determining the presence of an active profile, and selecting which representative profile to use. Possible relationships to be included in the prediction of load profile were time of day, day of week, season, region (NE, SE, SW), and the 15-minute average load. These possible relationships were investigated, through analysis of contingency tables and plots, to determine whether particular profiles occurred only at

TABLE 3. CLASSIFICATION TABLE DISPLAYING DISTINCT PROFILES

| 1ST HARMONIC AMPLITUDE | 2ND HARMONIC AMPLITUDE: LOW - LE .05 | | MED - GT .05, LE .10 | | HIGH - GT .10 | | |
|---|---|---|---|---|---|---|---|
| | NUMBER OF PROFILES | PROFILE GROUP | NUMBER OF PROFILES | PROFILE GROUP | NUMBER OF PROFILES | PROFILE GROUP | |
| LE .05 LOW | 52 | 1 | 55 | 2 | 45 | 3 | |
| GT .05, LE .10 MED | 136 | 4 | 66 | 5 | 9 | 13 | IN |
| | | | 49 | 6 | 8 | 13 | BTWN |
| | | | 13 | 13 | 1 | 13 | OUT |
| GT .10, LE .15 HIGH | 80 | 7 | 40 | 8 | 14 | 10 | IN |
| | | | 43 | 9 | 7 | 13 | BTWN |
| | | | 2 | 13 | 1 | 13 | OUT |
| GT .15 VERY HIGH | 38 | 11 | 9 | 13 | 14 | 10 | IN |
| | | | 17 | 12 | 27 | 12 | BTWN |
| | | | 2 | 13 | 6 | 13 | OUT |

IN = PHASE ANGLES IN PHASE
BTWN = PHASE ANGLES BETWEEN PHASE
OUT = PHASE ANGLES OUT OF PHASE
LE = LESS THAN OR EQUAL TO
GT = GREATER THAN

specific times of day, on weekdays or weekends, during a particular season, only in a particular region, or only for selective time-average loads. However, since the loads are appliance-driven, an argument could be made that the differences were due to use of different kinds of appliances, each with different efficiencies, load requirements, and usage practices. With a sample of only four households of appliances, determining precise relationships was not feasible. In addition, no clear relationships were determined.

A basic approach involving frequencies of occurrence was decided upon for the profile prediction. The frequency of occurrence of active profiles out of the total possible intervals when an active profile could occur was used to develop the baseline probability of an active profile. Referring to the

breakdown of the sampled intervals presented in table 2, the first and third rows were excluded as being possible active profiles. The resultant probability of an active profile was calculated to be

$$(58 + 734)/(1045 + 2069 + 734 + 58) = 0.203$$

Finally, given the presence of an active profile, the selection of one of the 13 representative profiles was based on the frequencies of occurrence of the profile groups out of the 734 total active profiles. Even though profiles are predicted by this relatively simple probabilistic prediction scheme, annual simulation results are considered to be realistic. However, comparing a predicted instantaneous load profile to an actual instantaneous load for a selected 15-minute interval could show large discrepancies.

An algorithm incorporating the statistical load model prediction scheme and the representative profiles was implemented into SOLCEL (reference 6). Using annual load data and solar/weather data, the program simulates the annual performance of specified PV and battery storage systems; computes the load requirements met by the PV system, battery storage, and the utility; and performs an economic analysis. The representative load profiles were normalized to the 15-minute time average load value and included as deviations from the 15-minute time averaged load. SOLCEL steps through the simulation at 15-minute intervals, but if the 15-minute time average value is greater than 0.5 kW (less than 0.5 kW is considered unconsequential) and an active profile is predicted, all computations are performed at 15-second interval steps. For active profiles, the algorithm uses the chosen profile deviations and the current time-averaged load to calculate the 15-second load values for that 15-minute interval. The probability of an active profile, to be input by the user, can range from 0.0 (no use of instantaneous load, 15-minute time-average only) to 1.0 (complete use of 15-second interval profiles of the statistical model between 5:30 a.m. to midnight).

## PERFORMANCE COMPARISONS

The SOLCEL simulation code with the statistical load model algorithm was exercised using various PV array sizes, probabilities of an active profile, and time intervals. Typical Meteorological Year solar/weather data for Albuquerque and an annual set of 15-minute load data for an Albuquerque residence, similar to those sampled in this project, were used as inputs to define PV output and average load. Performance of residential PV arrays is usually measured by the percentage load supplied by the PV output. Using the baseline estimate of probability of an active profile, .203, and a 4 kW array size (considered appropriate for the residences in this study), a baseline test case resulted in the PV array supplying 33.3 percent of the annual energy demand of the residence. This estimate represents the most accurate prediction of the actual PV fraction.

This result was then compared with results obtained using 15-minute and hourly time-average load data and using the statistical model with twice the occurrence of active profiles. Results for these four initial simulations are displayed in table 4 in the form of annual percentages of load met by the 4 kW PV system. As expected, with smaller time intervals and the use of the 15-second load profiles from the load model, the percentage of load satisfied by the PV system declines. However, the hourly result overestimates the baseline model

result by only 1.3 percentage points, or a 3.9 percent error. Even operating the model with twice the probability of occurrence of an active profile produces little difference in the amount of energy supplied by the array to the load.

TABLE 4. RESULTS FROM SOLCEL SIMULATIONS

| | MODEL WITH BASELINE PROBABILITY | 15 MIN | HOURLY | MODEL WITH TWICE PROBABILITY |
|---|---|---|---|---|
| PV DIRECT FRACTION | .333 | .339 | .346 | .327 |

When array size is considered, and the probability of an active load is increased, instantaneous load data produce a significant decline in PV fractions only for very small array sizes and unrealistically active loads. This effect is displayed in figure 7. The probability of an active load was varied from 0.0 (use of 15-minute load data with 15-minute simulation time-step) to 1.0 (a continuously active load between 5:30 a.m. and midnight with 15-second simulation time-step). A probability of an active profile of 0.4 was considered an extreme active case for load variability. This probability was extended to 1.0 to investigate sensitivities. Array size was varied from 1 to 9 kW. For more detailed discussion of these performance comparisons, see reference 5.

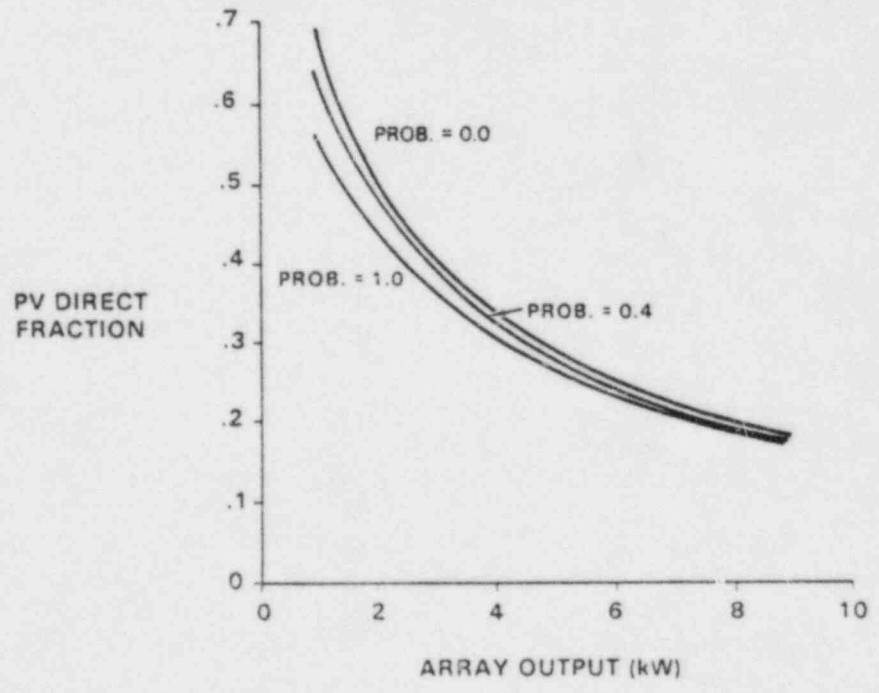



Figure 7. PV Performance Versus Array Size for Various Probabilities of an Active Profile

## CONCLUSIONS

A major finding of this project was the limited occurrence of active instantaneous load profiles. Results from the simulations for typically sized PV arrays and observed load activity indicate that errors in annual PV system performance measures arising from use of time-average load data versus instantaneous load data are not significant. Only extremely small PV arrays, where the average load demand would approximately equal the PV output, lead to significant over-estimation of the annual PV direct fraction by use of time-averaged load data. These results are contrary to results from previous probabilistic studies, which were not based on measured data. With the abundance of 15-minute time-averaged residential load data that currently exists with utilities, 15-minute load data is considered adequate for simulation and analysis of annual performance measures of residential PV systems.

## REFERENCES


1. Sanford Sillman, "On-Site Use and Sellback in Residential Photovoltaic Systems Based on Monitored Load Data," SERI/TP-254-1739, Golden, Colorado, October 1982.

2. Jay W. Lathrop, "Investigation of Load Management Strategies for Residential Photovoltaic Systems," Third Photovoltaic Systems Definition and Projects Integration Meeting, Proceedings, Sandia, 1982.

3. BMDP (1981), "Biomedical Computer Programs," University of California, Los Angeles, California.

4. IMSL (Edition 8), "International Mathematical and Statistical Libraries," Houston, Texas, International Mathematical and Statistical Libraries, Inc.

5. The BDM Corporation, "Detailed Residential Electric Load Determination Final Report," Albuquerque, New Mexico, SAND82-7114, BDM/A-82-299-TR-R1, 1983.

6. Sandia National Laboratories, "SOLCEL 2, A User's Manual," Albuquerque, New Mexico, SAND79-1785, 1979.

# PRA UNCERTAINTIES AND THE ROLES OF SENSITIVITY AND UNCERTAINTY ANALYSES

by

W. E. Vesely
Battelle Columbus Laboratory
505 King Avenue
Columbus, OH 43201

D. M. Rasmuson
Nuclear Regulatory Commission
Division of Risk Analysis Dept.
Washington, DC 20555

(Tutorial Session 1)

# PRA UNCERTAINTIES AND THE ROLES OF SENSITIVITY AND UNCERTAINTY ANALYSES

W. E. Vesely, Battelle Columbus Laboratories
D. M. Rasmuson, Division of Risk Analysis, NRC

## ABSTRACT

Seven categories of uncertainties are identified in a PRA and each category is divided into subcategories to better differentiate the implications of the different types of uncertainties. In most PRA's, the majority of uncertainties are not adequately treated or are not treated at all. By elevating sensitivity analyses to a major role in a PRA, the PRA can become a more meaningful and credible information source for decision making. Uncertainty analyses can also be made more useful information sources than they are now. The fact that PRAs allow uncertainties to be explicitly evaluated is a unique strength of PRAs which needs to be explioted much more than it is now.

## 1.0 INTRODUCTION

Probabilistic Risk Analysis, or PRA for short, has as its general objective the quantification of risks from man-made and natural activities. The quantification of risk is achieved by calculating frequencies and consequences of various accidents which can occur. PRA really started with WASH-1400 (1), which quantified the risks from nuclear power plant accidents, and today most effort in PRA is devoted to evaluating risks from nuclear power plants. In referring to PRA's we shall thus mean nuclear power plant PRA's; however, the discussions will be generally applicable to any PRA.

The usual application of a nuclear power plant PRA first involves constructing accident scenarios (accident sequences) which define the specific accidents to be considered. The accident scenarios consist of accident initiating events and safety system failures which must occur to produce significant consequences. The accident scenarios are defined by logic event trees.

The frequencies of occurrence of the accidents and the consequences from the accidents are quantified using reliability approaches, physical analyses, and statistical analyses. In quantifying the accident frequencies, a variety of accident contributors are considered such as component failures, human errors, and environmental stresses. Various consequences are calculated, including fatalities, radiation exposure, and property damage. The results from a PRA include probability versus consequence distribution curves, from which various characteristic risk values are derived such as the expected consequences from an accident. Figure 1 illustrates the steps involved in a PRA. References (2,3) discuss in more detail the techniques and approaches which are used in a PRA.

Because of the lack of experimental data, the assumptions, models, and data in a PRA involve a great deal of engineering assessments and subjective judgements. Different analysts and experts sometimes have different opinions on how likely accidents are and what the consequences will be. A PRA thus has considerable uncertainties associated with it, and these uncertainties have sometimes caused controversy over the meaningfulness and interpretations of a PRA. Because there are various uncertainties associated with a PRA, confusion often occurs when PRA uncertainties are discussed. The goal of this paper is to differentiate the uncertainties which are associated with a PRA.

According to Webster's New Collegiate Dictionary, "uncertainty" is the quality or state of being uncertain. "Uncertainty" in turn means indefinite, indeterminate, not certain to occur, not reliable, not known beyond doubt, not clearly identified, or not constant. "Uncertainty" is thus a general term which can mean the quality or state of being random, or of being unreliable, doubtful, vague, or changeable.

Probabilistic risk analysis (PRA) has uncertainties with regard to all the above meanings, and these different meanings are what often cause confusion when PRA uncertainties are discussed. In this paper we define the different types of uncertainty which arise in a PRA. The different types of uncertainties are quite different in nature and have different ramifications on the results produced by a PRA.

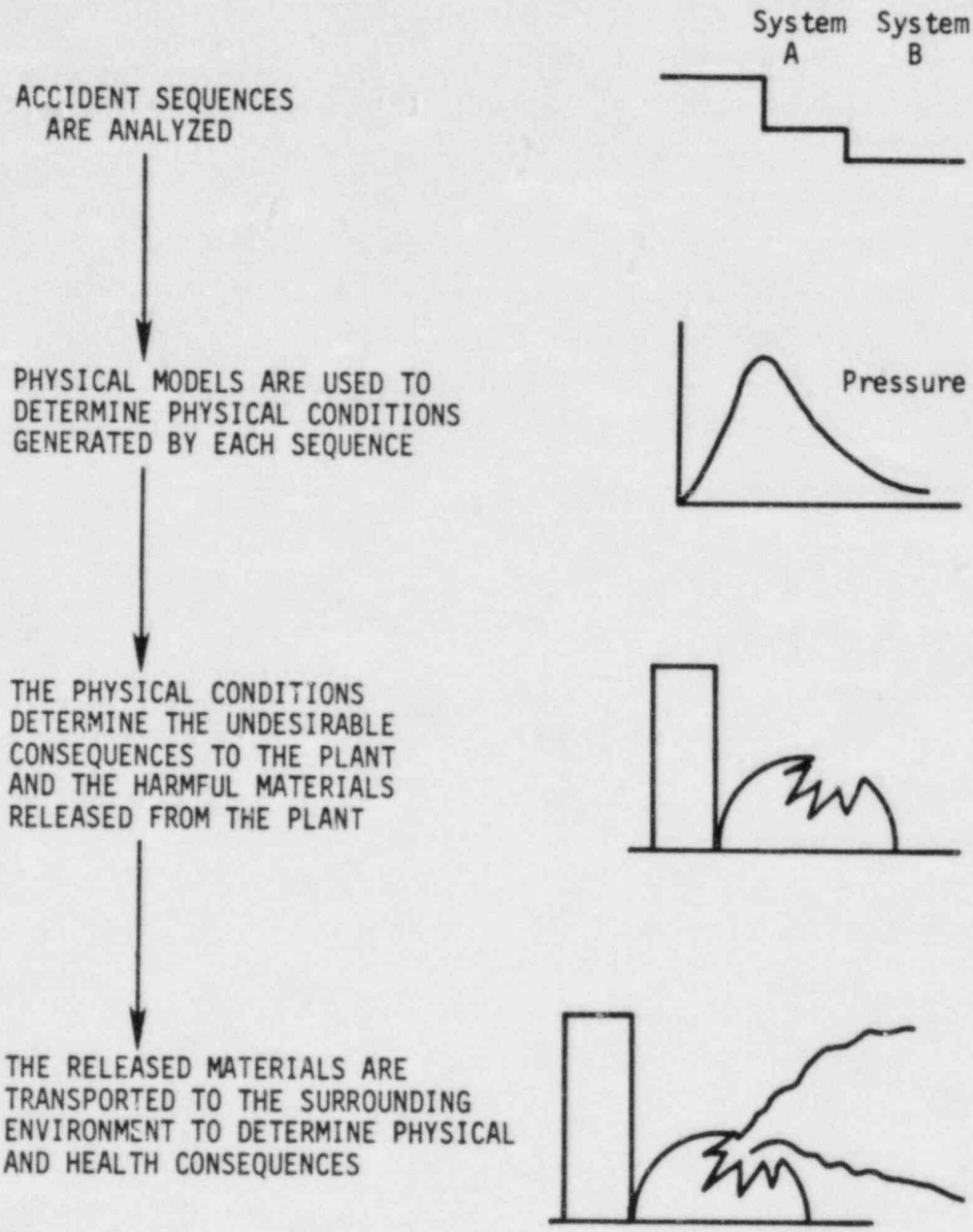



FIGURE 1. THE BASIC STEPS IN A PRA

After categorizing the different types of uncertainties, we discuss the roles of uncertainty and sensitivity analyses in a PRA. Many PRA uncertainties are not addressed by the usual uncertainty or sensitivity analyses performed in a PRA. We suggest possible ways in which uncertainty analyses can be made more complete and more useful.

The paper is organized as follows. In the next section, Section 2.0, we define the major types of uncertainty which exist in a PRA. In Section 3.0 we further categorize different types of uncertainty according to their specific properties and impacts. In Section 4.0, we discuss uncertainty analyses and sensitivity analyses and ways they can better contribute to the usefulness of a PRA. Finally, in the section on conclusions and observations, we comment on the present state of the art of uncertainty and sensitivity analyses and needs which exist as we see them.

## 2.0 THE MAJOR TYPES OF UNCERTAINTY

With regard to their different ramifications, especially for PRA's, we need to first of all distinguish between two major types of uncertainty:

1. Uncertainty due to physical variability

and

2. Uncertainty due to lack of knowledge.

Uncertainty due to physical variability is actual, random behavior in some physically measurable quantity. We can imagine an experiment being conducted which produces a specific value of the quantity; when the experiment is repeated, different values will be produced because of some underlying physical variability. "Experiment" is used in the general context and can refer to an actual physical experiment, a specific scenario producing a specific value of some variable, or a given sample measurement having a specific value. Examples of uncertainty due to physical variability are variations in weather, variations in stock market prices from one day to another, variations in component failure times from one observation to another, and variations in consequences from one accident to another. "Uncertainty due to experimental variability" would thus be another label for this type of uncertainty. For ease of reference, we shall call uncertainty due to physical variability "experimental uncertainty".

The second type of uncertainty, uncertainty due to lack of knowledge, is quite different from experimental uncertainty. Uncertainty due to lack of knowledge is vagueness, indefiniteness, or imprecision in an analysis, a stated conclusion, or stated value. The uncertainty exists because of a lack of knowledge; if we had more information and more knowledge, the uncertainty would decrease or would not exist. Examples of uncertainty due to lack of knowledge are uncertainties in the appropriateness of an economic model, uncertainties in a conclusion concerning the acceptability of a risk value, and uncertainties associated with an estimated value of a parameter. (A parameter is defined here to be an unknown constant.) We shall call uncertainties due to lack of knowledge, "knowledge uncertainty".

As indicated above, what differentiates experimental uncertainty from knowledge uncertainty is the impact that additional knowledge has. As

we gain more knowledge, knowledge uncertainty will decrease. However, as we gain more knowledge, experimental uncertainty will not decrease; we will know the variable behavior better and be able to quantify it more precisely; however, the variability itself will not diminish.

Variation in weather is a common example of experimental variability (the "experiment" here is an observation on a given day, for example). As we collect additional meteorological readings at a site, day to day temperatures and day to day rainfall amounts will not vary less. However, as we collect more meteorological data, we may be able to more precisely estimate a fixed parameter value in some weather model reducing our knowledge uncertainty about that parameter value.

There is the argument that if we had perfect, complete knowledge, there would be no experimental uncertainty as well as no knowledge uncertainty. We would know all the causes of physical and random variations and would be able to precisely predict the value that will occur in the next experiment. However, this would not remove the variation from experiment to experiment, even though under specific circumstances we might be able to control the variation. Furthermore, perfect knowledge is an ideal state which is really never achieved. We therefore believe the differentiation between experiment uncertainty and knowledge uncertainty to be meaningful and important.

Both Bayesian and classical statistics textbooks generally agree on the interpretation and analysis of experimental uncertainty. Experimental uncertainty is associated with a random variable which can vary from experiment to experiment; the probability of specific values of the random variable is characterized by a sampling distribution. The variation in the values of the random variable are viewed as being frequency based; the empirical probability of an event for a finite number of experiments is defined as the measured number of occurrences of the event divided by the number of experiments conducted. As the number of trials increase, the empirical probability approaches the true probability of the event. The set of repeated experiments constitutes the "collective" in the terminology of von Mises, one of the founders of the frequency based approach of statistics (4).

Knowledge uncertainty is more complex and more nebulous than experimental uncertainty. Knowledge uncertainty consists not only of imprecision in parameter estimates, but also incompleteness in modeling and analysis, vagueness in appropriate data values and data ranges, indefinitieness in the applicability of the model, and doubtfulness and vagueness in the interpretability of results produced by a model.

Statistics, both Bayesian or classical, generally deal with only one specific type of knowledge uncertainty, the imprecision in an estimated parameter value. The imprecision is described by a confidence or probability interval for the parameter which reflects uncertainty due to insufficient numbers of measurements, and which assumes the data are appropriate and all models are applicable and exactly known. The usual Bayesian and classical statistical uncertainty analyses performed in PRA's treat only data imprecisions and their impacts on the PRA results. Consequently, a large portion of the uncertainties which exist in a PRA are not considered by these uncertainty analyses.

Some nonconventional statistic approaches have been developed for handling additional types of knowledge uncertainty (5), however, these approaches generally have been limited in their application and are not used in PRA's. In addition to the nonconventional statistical approaches, a theory termed fuzzy set theory has undergone rapid development in the past several years. Fuzzy set theory (6,7) attempts to address the various types of knowledge uncertainty which are not addressed by conventional statistical approaches. In the following section, we will utilize fuzzy set theory concepts, as well as statistical concepts when appropriate, to identify the different types of uncertainty which exist in a PRA.

## 3.0 THE DIFFERENT TYPES OF UNCERTAINTIES IN A PRA

The IEEE/ANS PRA Procedures Guide (2) identified three general types of uncertainties which exist in a PRA: completeness uncertainty, modeling uncertainty, and data uncertainty. Completeness uncertainty generally refers to uncertainty in identifying all the elements and contributors to risk. Modeling uncertainty generally refers to uncertainty in the appropriateness of the models used in a PRA. Data uncertainty refers to uncertainty in the parameter values input to a PRA.

These three types of uncertainties generally refer to different types of knowledge uncertainties; however, experimental uncertainties can also be included in some cases. It is useful to expand these three uncertainty types into finer categories to better characterize the uncertainties which exist in a PRA. The expanded categories will allow more specific investigation of the ramifications of the different uncertainties. Table 1 gives the expanded categories of uncertainties which we have identified.

In Table 1, seven major categories of uncertainties are identified, with each category divided into subcategories where necessary for further specification. The seven categories are: (1) data uncertainties, (2) analyst uncertainties, (3) modeling uncertainties, (4) completeness uncertainties, (5) frequency uncertainites, (6) consequence uncertainties, and (7) interpretation uncertainties. These seven categories, going from Category 1 to Category 7, represent a progression from PRA input uncertainties to higher level uncertainties associated with PRA results. Even though particularly applied to a PRA, we believe these categories are generally applicable to any modeling exercise. We now briefly discuss the different categories and subcategories of uncertainties.

### Data Uncertainties

We begin with data uncertainties since they are generally the uncertainties which are treated in a PRA. Data uncertainties refer to uncertainties in the parameters which are input to a PRA. The parameters include those constants which are used in estimating the frequency of different accidents

TABLE 1. CLASSIFICATION OF PRA UNCERTAINTIES

| | Category | Subcategory | General Type |
|---|---|---|---|
| 1. | Data Uncertainties | Variation in parameter values from one population to another | Experimental Uncertainty |
| | | Imprecision in estimated parameter values | Knowledge Uncertainty |
| | | Vagueness in parameter values or parameter ranges | Knowledge Uncertainty |
| | | Indefiniteness in applicability of data | Knowledge Uncertainty |
| 2. | Analyst Uncertainties | Variation in results from analyst to analyst | Knowledge Uncertainty |
| 3. | Modeling Uncertainties | Indefiniteness in the comprehensiveness of the model | Knowledge Uncertainty |
| | | Indefiniteness in the characterizations used in the model | Knowledge Uncertainty |
| 4. | Completeness Uncertainties | Indefiniteness as to whether all significant contributors are included | Knowledge Uncertainty |
| | | Indefiniteness as to whether the contributors are included in the proper context and in the correct relative manner | Knowledge Uncertainty |
| 5. | Frequency Uncertainties | Variation in the occurrence frequency from one accident to another | Experimental Uncertainty |
| | | Uncertainty in the occurrence frequency of of a given accident resulting from data, analyst modeling, and completeness uncertainties | Knowledge Uncertainty |
| 6. | Consequence Uncertainties | Variation in consequences from one accident to another | Experimental Uncertainty |
| | | Uncertainty in the consequences of a given accident resulting from data, analyst modeling, and completeness uncertainties | Knowledge Uncertainty |
| 7. | Interpretation Uncertainties | Doubtfulness or vagueness in the interpretability of the results produced by the analysis | Knowledge Uncertainty |

and those constants which are used in estimating the consequences of the accidents. Table 2 lists various parameters which are input to the likelihood and consequence evaluations in a PRA.

The data uncertainty category consists of four subcategories (1) population variations, (2) imprecisions in values, (3) vagueness in values, and (4) indefiniteness in applicability. Population variation occurs when the parameters vary from scenario to scenario within the analyses; the variation can generally be related to physical causes. The parameters may vary over time, or over different spatial regions, or over different individuals. The parameters become random variables which assume different values under these different situations. When we assume the parameters are constants, we ignore these variations. For example, in PRA's, component failure rates are generally treated as being constant over time and over similar components, however, more comprehensive modeling would allow the failure rates to randomly vary with time, with the system in which the component is located, and with the unique environment experienced by the component. This type of uncertainty is not adequately treated by PRA's, or at most is treated in a very limited manner.

The second type of data uncertainty, parameter imprecision, occurs when limited measurements are only available to estimate the parameter values. This is the type of data uncertainty which is standardly treated by statistical analysis, either Bayesian or classical. Assuming the measurements are completely applicable, and assuming an exactly known given probability distribution (likelihood) for the occurrence of the measurements, a confidence interval or probability distribution (a sampling distribution or posterior distribution) is then derived to describe the imprecision in the parameter value inferred from the measurements. The confidence intervals or distributions are then propagated through the PRA to obtain the corresponding intervals or distributions on the PRA results. Data imprecisions are generally the uncertainties, and the only uncertainties, which are quantified and propagated in PRA uncertainty analyses.

The third type of data uncertainty, parameter vagueness, is different from parameter imprecision and is the type of uncertainty which is addressed by fuzzy set theory. Parameter vagueness refers to the situation where definitive values or definitive intervals are not able to be assigned to

TABLE 2. PARAMETERS UTILIZED IN A PRA EVALUATION

| |
|---|
| **Frequency Parameters** |
| Frequencies of Accident Initiating Events |
| Component Failure Rates |
| Human Error Rates |
| Dependent Failure Probabilities (e.g., beta factors) |
| Test intervals and Durations |
| Maintenance Intervals and Durations |
| **In-Plant Consequences Parameters** |
| Number of compartments for the containment; initial temperatures, pressures, humidities, and dimensions of the compartments |
| Densities, heat capacities, and thermal conductivities of the heat sinks |
| Thickness, density, thermal conductivity, and initial temperature of containment floor |
| Flow rates, shutoff pressures, and failure temperatures of coolant pumps |
| Flow rate, water temperature, and spray drop diameter of spray system |
| Active fuel height, mass of $VO_2$, fuel-pellet diameter, hydraulic diameter, cladding thickness |
| Heat capacity, temperatures, and heat transfer area of vessel structures. |
| **Ex-Plant Consequence Parameters** |
| Inventories of all radionuclides at the time of the accident |
| Time and duration of radioactive releases, warning time for evacuation |
| Elevation buoyancy and duration of releases |
| Particle sizes and chemical properties of the released radionuclides |
| Meteorological data; hourly wind speed, duration, stability category, precipitation index |
| Population density, evaluation speed, dry-deposition velocity |
| Value of land, relocation costs, decontamination costs |

a parameter; in fuzzy set theory this is often described as not being able to assign "crisp" values and intervals (6). No one "best" estimate is able to be assigned to a parameter, or if an interval is ascribed for a parameter, one is not able to assign 75 percent, 90 percent, 99 percent, or some other definite confidence value or probability value to the interval. Parameter vagueness may arise from a variety of causes. The conditions under which the data were collected may not be exactly known. Prior knowledge (such as the prior distribution) about the parameter may be fuzzy and only partially quantifiable; for example, a fuzzy range for the parameter may only be specifiable. Parameter vagueness is a common situation in PRA's and what one generally does is ignore the vagueness and to assume some well defined best estimate, probability interval, and associated probability distribution. Parameter vagueness is thus not treated in PRAs and this vagueness can significantly increase the data uncertainties now assessed in current PRA's.

The final type of data uncertainty, parameter applicability, refers to extrapolation of the parameter values to the specific situations in the analyses. The available data bases or measurements may pertain to different situations than those being analyzed, or general (generic) parameter values may only exist which need to be specialized to the given situation. An example where applicability uncertainty arises is extrapolating component failure data for normal environments to accident environments. Another example is extrapolating human behavior on plant simulators to behavior in actual accident situations. Available data may be precise but not very applicable.

When applicable, plant specific data does not exist for a PRA then generic data is often used. In current PRA's, "generic data" is a nebulous term which means data representative of some class. The constitutuency of the class is not well defined or is rather arbitrarily described by some assumed probability distribution. The relevance of the class-averaged data or class description to the particular case being evaluated is not generally addressed. The individual case, such as an individual component, is assumed to be the same as the average or median of the ill-defined class. Alternatively the individual case is assumed to be a random selection from the class with some rather arbitrarily assumed distribution. Uncertainties due to applicability are thus not treated in PRAs or are glossed over.

## Analyst Uncertainty

Analyst uncertainty is the second major category of uncertainty which we have identified, and refers to the variation which exists in PRA modeling and quantification due to individual analyst interpretations. Analyst uncertainty is sometimes lumped with data uncertainty, however, it is a separate uncertainty contribution. Given the same problem and same basic information, there will be variation in the results among different analysts (different PRA teams) because of their different analysis framework. There is a potential for large variations because of the latitude afforded by sparse data and sparse experience on accident occurrences. We categorize this type of uncertainty as a knowledge experimental uncertainty. This uncertainty is not addressed in current PRA's; it can be a significant uncertainty contribution in comparing PRA's performed by different analysts even with supposedly the same ground rules on the PRAs.

## Modeling Uncertainties

The third major category of uncertainties, modeling uncertainties, refers to uncertainties in the applicability and precision of the models which are used in a PRA. Numerous models are utilized in a PRA to model initiating event occurrences, safety system and component failures, human errors, physical phenomena which are associated with an accident, containment and mitigating system behavior, radiological transport, and health effects and other consequences which result from the accidents. Table 3 lists models which are utilized in a PRA.

We have divided modeling uncertainty into two subcategories: (1) indefiniteness in the model's comprehensiveness, and (2) indefiniteness in the model's characterizations.

Indefiniteness in model comprehensiveness refers to the uncertainty as to whether the model accounts for all the variables which can significantly affect the results. For example, in modeling operator response in an accident, the question arises as to whether all the performance shaping factors are considered which can significantly affect the operator's behavior. In modeling the stresses which are exerted on the reactor containment, the

TABLE 3. MODELS UTILIZED IN A PRA

| |
|---|
| **Frequency Models** |
| Accident Occurrence Models |
| Component Failure Models |
| System Failure Models |
| Human Response and Failure Models |
| Testing and Maintenance Models |
| Fire Propagation Models |
| Seismic Response and Fragility Models |
| Flood and High Wind Models |
| **In-Plant Consequence Models** |
| Meltdown Thermal Hydraulics Models |
| Radionuclide Release Models |
| Steam Explosion Models |
| Hydrogen Combustion Models |
| Containment Response Models |
| Radionuclide Transport Models |
| Containment Failure Mode Models |
| **Ex-Plant Models** |
| Atmospheric Dispersion Models |
| Cloud Depletion Models |
| Ground Contamination Models |
| Dosimetry Models |
| Evaluation Models |
| Health Effects Models |
| Property Damage Models |

question arises as to whether all the pertinent phenomena are considered wich can result in high containment pressure.

Indefiniteness in model characterization refers to the uncertainties in the relations and descriptions used in the model. Even if the pertinent variables are included in the model, appropriate relationships among the variables may not be described. Are the phenomena treated to sufficient detail to allow meaningful results to be obtained? Are the phenomena meaningfully characterized in the model? Should the phenomena be treated in a probabilistic or deterministic fashion? How are failures defined? Should the variables be treated as random variables or as parameters? These are all questions which affect the uncertainty in the model's characterizations. Modeling uncertainties in general are not well addressed in a PRA and it is our experience that they are as important or more important than data uncertainties.

Completeness Uncertainties

The fourth major category of uncertainties is completeness uncertainties. Completeness uncertainties are the uncertainties as to whether all the significant phenomena and all the significant relationships have been considered in the PRA. Completeness uncertainties are similar in nature to modeling uncertainties but occur at the initial, identification stage in the PRA. To perform a PRA, we must first ask what risks are to be considered, what types of accidents are to be considered, and what types of accident contributors are to be considered. These questions relate to completeness uncertainties. The selection of models to then produce the accident probabilities and accident consequences relates to modeling uncertainties.

There are two subcategories of completeness uncertainties, (1) contributor uncertainties, and (2) relationship uncertainties. Contributor uncertainties refer to the uncertainty as to whether all the pertinent risks and all the important accidents have been included. Relationship uncertainties refer to the uncertainty as to whether all the significant relationships are identified which exist among the contributors and variables. Table 4 gives a more detailed breakdown of items which need to be considered under completeness-contributor uncertainties and those which need to be considered under completeness-relationship uncertainties.

TABLE 4. ELEMENTS IMPACTING COMPLETENESS UNCERTAINTY

| |
|---|
| **Elements Impacting Contributor Uncertainty** |
| The completeness of the accident initiating events |
| The completeness of system and component failure states |
| The completeness of human responses |
| The completeness of the causes of system and component failures |
| The completeness of physical processes involved in core-meltdown |
| The completeness of hydrogen burning phenomena |
| The completeness of phenomena impacting radionuclide transport |
| The completeness of consequences and health treated effects. |
| **Elements Impacting Relationship Uncertainty** |
| The completeness of interactions defined between the initiating event and system failures |
| The completeness of interactions among system failures, component failures, and human errors |
| The completeness of interactions between system failures and physical processes |
| The completeness of interactions among in-vessel and out-of vessel phenomena |
| The completeness of interactions among radionuclide transport processes |
| The completeness of interactions among consequences and health effects. |

Completeness uncertainty acts as a constraint and limitation on a PRA. The PRA evaluates the risk from only those accident scenarios which are identified. These will never be exhaustive nor will the analyses ever be completely comprehensive. Ways thus need to be devised to utilize PRAs which are consonant with their strengths and weaknesses and which account for their possible incompletenesses. If high risk contributors are not found then it may be due to the incompleteness of the PRA. Completeness uncertainties are thus as important as modeling and data uncertainties and especially so when low risk numbers are calculated.

Frequency Uncertainties

The last three categories of uncertainties--(5) frequency uncertainties, (6) consequence uncertainties, and (7) interpretation uncertainties deal with uncertainties in PRA outputs and results. Frequency uncertainties are uncertainties associated with the accident probabilities and accident frequencies produced by a PRA. The frequency uncertainties result from the data, analyst, modeling, and completeness uncertainties which were previously discussed and which propagate through the PRA to the calculated accident probabilities and frequencies.

Frequency uncertainties are divided into two subcategories, accident to accident variations which comprise experimental uncertainties, and knowledge uncertainties in the estimated frequency of any given accident. Accident to accident variations are the differences in accident frequencies which are due to the different events involved in different accidents. In most PRA's, a finite number of accidents are considered and the variation in accident frequency is represented as a histogram or smoothed curve of frequency versus accident description. Variations in parameter values which are physically caused (i.e., experimental, data uncertainties), contribute to the accident to accident variation. By treating certain random variables in the accidents as fixed constants, the accident variation is artifically truncated; however, compared to other uncertainties, PRA's handle this type of uncertainty reasonably well for those accidents identified.

Frequency uncertainties due to lack of knowledge arise from knowledge uncertainties in data, modeling, and completeness; analyst to analyst variations also contribute. In PRA's, imprecisions in data (parameters) are generally the only uncertainties which are propagated through the analysis to obtain frequency uncertainties due to lack of knowledge. The other data uncertainties, analyst uncertainties, modeling uncertainties, and completeness uncertainties which are generally not addressed, all contribute as much or more uncertainty in most problems.

## Consequence Uncertainties

Consequence uncertainties are similar in nature to frequency uncertainties and consist of accident to accident variations and lack of knowledge uncertainties. The accident to accident variations are physical variations in consequences due to the different events and physical realizations of variables which occur. The knowledge uncertainties are uncertainties in the consequences of any given accident due to data, analyst, modeling, and completeness uncertainties. Like the frequency uncertainties, most PRAs only estimate the uncertainties which are due to parameter imprecisions, which is in many, if not most, PRAs not the dominant contributor. Figure 2 illustrates the frequency and conseqence uncertainties as applied to a PRA-calculated "risk curve", i.e., a complete cumulative distribution of frequency versus consequences.

## Interpretation Uncertainties

Interpretation or implementation uncertainty is the last uncertainty category. Interpretation uncertainty is the uncertainty the decision maker, manager, or the public has in understanding and utilizing PRA results. This uncertainty is as real as the other uncertainties. The user of PRA results, or the intended audience toward which a PRA is directed, is not in general the same as the individuals who were involved in performing the PRA. The transfer of knowledge from the doer to the user involves a loss of information and introduces additional uncertainties.

The communication of information in a PRA is not given near the attention it should. Obtuse tables, abstract distribution curves, and uncoded computer printouts are produced in abundance in a typical PRA report. Page after page of the report is filled with technical details which serve as an obstacle to understanding. What makes matters worse is that a full scale PRA often results in more than a 1000 pages of report. The format of a typical PRA is a maze; we know, since we've had to review numerous PRAs and have tried to retrieve information from them.

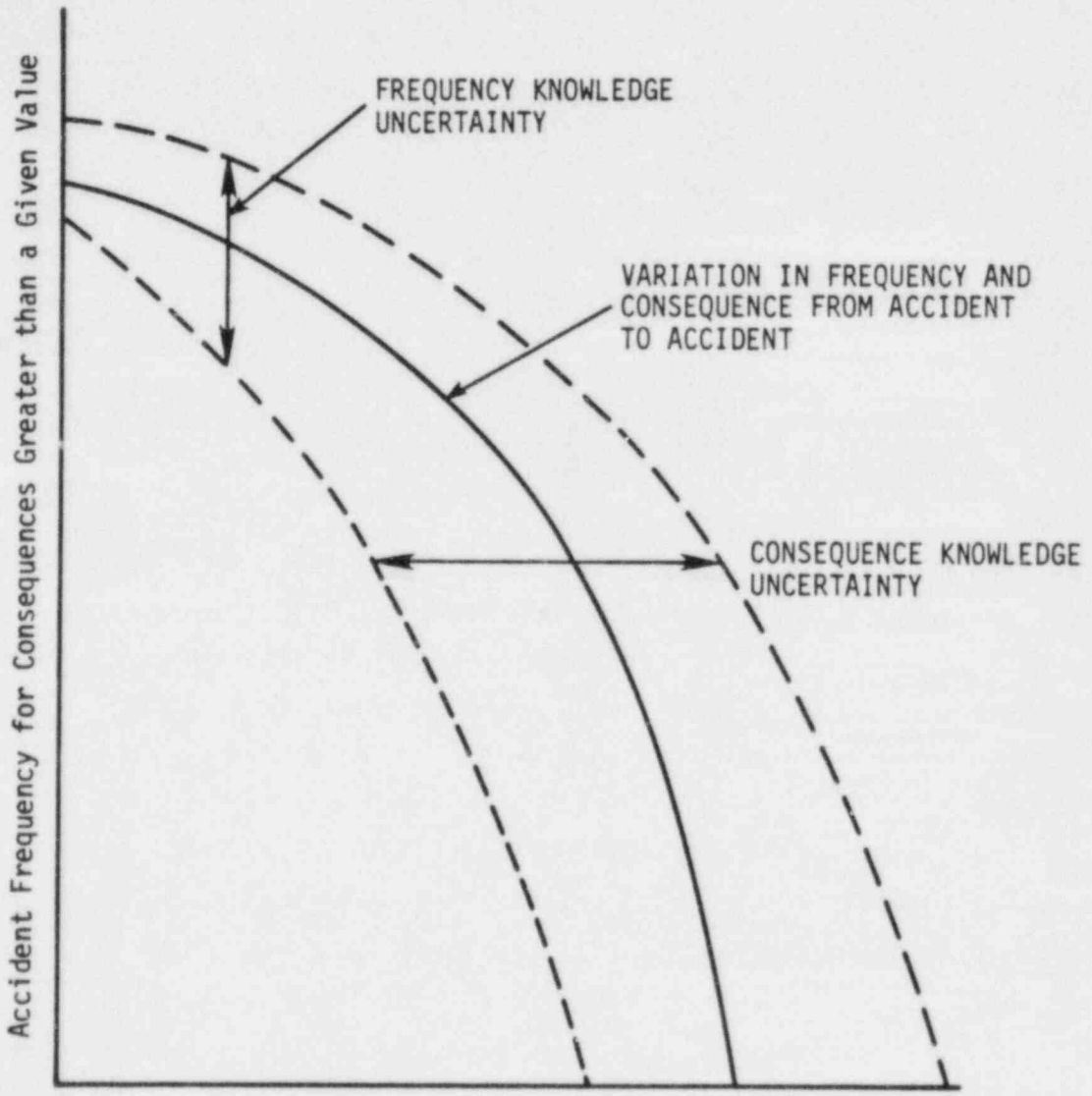



FIGURE 2. FREQUENCY AND CONSEQUENCE UNCERTAINTIES

# 4.0 SENSITIVITY AND UNCERTAINTY ANALYSES

Sensitivity and uncertainty analyses are analyses performed to investigate the impacts of uncertainties in PRA assumptions, models, and data. Sensitivity and uncertainty analyses are similar in that they both have as a general objective the evaluation of variations in the results (output) which can occur because of variations in the assumptions, models, and data (input). Even though they have a similar, general objective, the two analyses are different in the approaches they use and the information they supply. Sensitivity and uncertainty analyses are not exploited enough in a typical PRA; when fully utilized, sensitivity and uncertainty analyses can provide a rich variety of information which can significantly increase the usefulness and credibility of a PRA.

## Sensitivity Analyses

Sensitivity analysis is the most straightforward of the two types of analyses and involves changing one or more of the inputs and determining the resulting changes in the PRA output. Sensitivity analyses can involve changing the inputs one at a time, two at a time, up to all at a time. The inputs changed are those deemed most subject to possible variations and uncertainties. Multiple inputs are simultaneously changed if they are thought to be related, for example due to dependencies or if large interactions due to the simultaneous changes are suspected. A set of inputs could also be simultaneously changed if the set of modified values represented a conservative or optimistic bound on the PRA evaluation.

The size of the input change depends upon the focus of the sensitivity study. Larger changes in the inputs are generally used to represent bounding cases or alternative hypotheses; conservative or optimistic results are often obtained. Small changes are used in local sensitivity studies. For these local sensitivity studies, partial derivatives of the output with regard to one or more of the inputs are often used to characterize the output sensitivities. Taylor series or response surfaces can also used to represent the output variations in the vicinity of some nominal input value set.

We regard importance evaluations as a part of sensitivity analyses. Importance evaluations identify the "importance" or contribution of inputs to the output results. The importances of inputs are generally obtained by determining the changes in outputs which result from prescribed changes in inputs. In this regard, they are thus types of sensitivity analyses. Importance evaluations in PRAs are discussed in more detail in References 8 and 9.

Figure 3 illustrates formats of systematic sensitivity analyses that can be instituted as part of a PRA. In each sensitivity analyses table, a specific assumption, model, piece of data, or combination of inputs, is modified in the left column and the impact on the PRA calculated likelihood and/or consequence is described in the table. Core melt frequency impacts, radioactive release impacts, early fatality impacts, latent fatality impacts, and property damage impacts can be specifically identified and can form the table headings across the top of the table.

The size of impacts in the tables can be determined from the sensitivity calculations which are performed. The impacts can also be coded; for example, high, medium, and low with associated ranges of values could be assigned if impacts are subjectively estimated. The direction of the impact, as to whether the result is increased or decreased, can also be identified (sometimes called the downside and upside impacts). Table 5 gives examples of the specific sensitivity analyses which can be performed in a PRA and the results shown in a codified table such as in Figure 3. These sensitivity analyses would substantially increase the information provided by the PRA. This list is certainly not complete and can be supplemented by additional studies for particular implementations.

## Uncertainty Analyses

Uncertainty analysis is different from sensitivity analysis in that uncertainty analyses attempts to describe the likelihood for different size variations while sensitivity analyses does not. The description of the likelihood of different variation can be carried out formally using classical statistics, Bayesian statistics, or fuzzy set theory, or can be carried out informally using qualitative descriptors such as describing "likely" ranges for values.

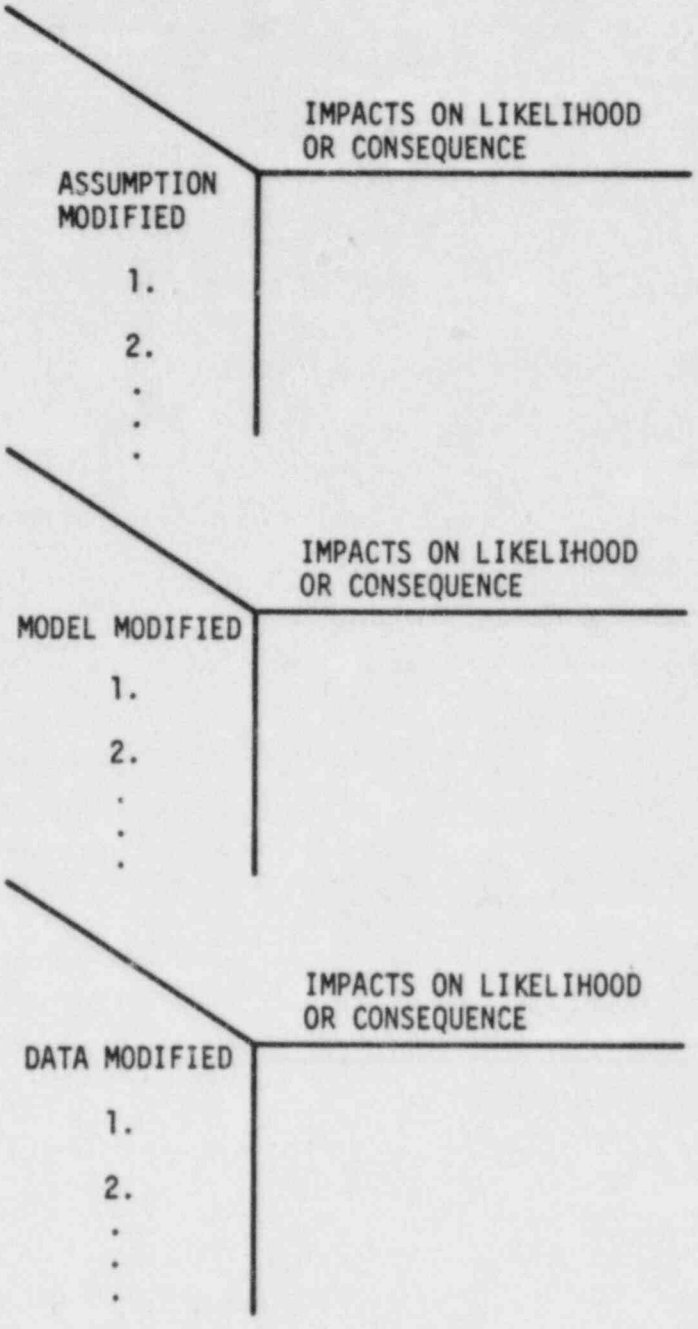



FIGURE 3. SENSITIVITY STUDY TABLES

TABLE 5. EXAMPLES OF SENSITIVITY ANALYSES

1. Change system failure definitions where relevant according to regulatory definitions and according to best estimate definitions.

2. Assume the probability of the second and subsequent human errors to be unity for moderate or highly coupled human actions according to NUREG/CR-1278 (10). This would show the impact of multiple human errors committed due to common procedural errors or "mindsets".

3. Assume the probability of the second and subsequent component failures to be unity for those components of the same generic type which are under a common maintenance or testing program. This would show the potential impacts of systematic maintenance or testing deficiencies.

4. Change all priors used in Bayesian analyses to likelihood-dominated priors (essentially flat priors on a linear or log scale) to show the impacts of a priori assumptions on data.

5. Calculate quantitative importances of individual components, human actions, test and maintenance activities, and systems.

6. Increase all human error rates associated with maintenance by a factor of 2 to 3 to show the impact of increased maintenance error frequency.

7. Change occurrence frequencies, fragilities, and responses in seismic risk analyses to show the impacts of design, operation, and modeling effects; calculate quantitative importances for all the various contributors.

The formal uncertainty approaches are more attractive in that they explicitly quantify the uncertainties and the contribution to the uncertainty. However, these formal approaches require distributions (functions) to be assigned to quantify the likelihood of individual values being realized. For classical statistic approaches, sampling distributions are required.* For Bayesian statistics, prior distributions are required. For fuzzy set theory, membership functions are required. In performing the formal uncertainty analyses, the uncertainty distributions are assigned to input variables and are propagated through the PRA analyses to obtain uncertainty distributions on the output results.

The assignment of distributions in formal uncertainty analyses is a problem since it can involve as much uncertainty as that to be quantified. The assigned uncertainty distributions, particularly for Bayesian and fuzzy set analyses, have little empirical basis and indicate more the subjective viewpoints of the PRA analyst.

The PRA analyst generally selects specific distributions (e.g., discrete log normal) for formal uncertainty analyses based on his own personal rationale and thus wrongly places himself in the position of the decision maker. We support the philosoply of Leamer (11), which interpreted in a PRA context says that a PRA uncertainty analyses should provide a mapping to show how different decision-maker viewpoints are transformed by the PRA to updated or posterior assessments. At a minimum this means using different prior distributions in sensitivity studies to show their impacts on the PRA results derived.

With the present state of the art only data uncertainties, and specifically data imprecisions, are treatable by formal classical and Bayesian uncertainty analyses approaches. Data uncertainty is however only one of the many uncertainties which exist in a PRA--an important uncertainty but only one uncertainty.

---

*Classical confidence intervals can also be utilized, however, they require the same type of information.

Fuzzy set theory has potential but has not been applied to PRAs. Informal uncertainty analyses approaches can be useful, but sufficient work has not been performed to understand and codify their applications. Formal uncertainty analyses approaches also have potential applicability to the other categories of uncertainties which exist in a PRA, but at the present time this is a potential and not a realization.

Consequently, the majority of uncertainties in a PRA can at the present time be treated only by sensitivity analyses which necessarily must form a critical supplement to uncertainty analyses. Table 6 lists the uncertainties which we have identified in a PRA and the type of analyses, uncertainty or sensitivity analyses, which presently is available to address the uncertainty. In Table 6, uncertainty analyses is restricted to either classical or Bayesian statistical analyses since fuzzy set theory has not yet been applied to PRAs. Wherever uncertainty analyses appears by itself in Table 6 it can of course be replaced or be supplemented by sensitivity analyses. As observed in the table sensitivity analyses clearly predominates as the most relevant technique.

TABLE 6. APPROACHES AVAILABLE TO ADDRESS PRA UNCERTAINTIES

| Category | Subcategory | Available Approach |
|---|---|---|
| 1. Data Uncertainties | Variation in parameter values from one population to another | Uncertainty Analyses |
| | Imprecision in estimated parameter values | Uncertainty Analyses |
| | Vagueness in parameter values or parameter ranges | Sensitivity Analyses |
| | Indefiniteness in applicability of data | Sensitivity Analyses |
| 2. Analyst Uncertainties | Variation in results from analyst to analyst | Sensitivity Analyses |
| 3. Modeling Uncertainties | Indefiniteness in the comprehensiveness of the model | Sensitivity Analyses |
| | Indefiniteness in the characterizations used in the model | Sensitivity Analyses |
| 4. Completeness Uncertainties | Indefiniteness as to whether all significant contributors are included | Sensitivity Analyses |
| | Indefiniteness as to whether the contributors are included in the proper context and in the correct relative manner | Sensitivity Analyses |
| 5. Frequency Uncertainties | Variation in the occurrence frequency from one accident to another | Uncertainty Analyses |
| | Uncertainty in the occurrence frequency of of a given accident resulting from data, analyst modeling, and completeness uncertainties | Sensitivity Analyses |
| 6. Consequence Uncertainties | Variation in consequences from one accident to another | Uncertainty Analyses |
| | Uncertainty in the consequences of a given accident resulting from data, analyst modeling, and completeness uncertainties | Sensitivity Analyses |
| 7. Interpretation Uncertainties | Doubtfulness or vagueness in the interpretability of the results produced by the analysis | Sensitivity Analyses |

## 5.0 SUMMARIZATIONS AND CONCLUSIONS

Seven categories of uncertainties have been identified in a PRA: (1) data uncertainties, (2) analyst uncertainties, (3) modeling uncertainties, (4) completeness uncertainties, (5) frequency uncertainties, (6) consequence uncertainties, and (7) interpretation uncertainties. The first four categories represent input uncertainties and the last three represent output and implementation uncertainties. These different uncertainties have different ramifications in a PRA and they all are important.

In most PRAs, the majority of uncertainties which we have identified are not adequately treated, and in fact, are not likely to be treated at all. Inadequate sensitivity analyses are performed in a PRA and if they are performed they are rather unorganized and subjugated to a minor role. When uncertainty analyses are performed in a PRA they are narrowly focused with assumed distributions and with the analyst playing the role of decision-maker.

We believe sensitivity analyses need to be elevated to a major role in a PRA. The information provided by sensitivity analyses can be as useful or even more useful than the bottom-line probabilities and consequences which are calculated in a PRA. We believe a PRA uncertainty analysis needs to provide a mapping to show how different a priori decision-maker assessments are transformed by the PRAs; at minimum this means using different priors for Bayesian analyses.

Because of the different uncertainties which exist in a PRA, sensitivity analyses need to importantly supplement formal uncertainty analyses which are performed in a PRA. We think the systemization and codification of sensitivity and uncertainty analyses on a PRA needs significant attention. As importantly, the results of these studies need to be presented in a format understandable to the decision-maker.

With regard to research, we think additional uncertainty analyses approaches, both formal and informal, need to be developed to address the range of PRA uncertainties which exist in a PRA. As an important part of this research, practical ways of utilizing uncertainty information in decision-making needs to be identified. Through these efforts, we believe PRAs will become more meaningful and credible information sources for decision-making.

The uncertainties which occur in a PRA are not different from those which occur in other types of safety analyses or other analyses in general. PRA uncertainties have been the focus of much attention and much criticism because they tend to be more apparent than in other approaches which tend to obscure the involved assumptions and uncertainties. The fact that PRAs allow uncertainties and their impacts to be made evident is a unique strength of PRAs, not a weakness. This strength needs to be exploited and we feel this can be done by pursuing the directions discussed in this paper.

# UNCERTAINTY ANALYSIS: GOOD NEWS AND BAD NEWS

by

R. J. Beckman and D. E. Whiteman
Los Alamos National Laboratory
P. O. Box 1663
Los Alamos, NM 87545

(Tutorial Session 1)

# UNCERTAINTY ANALYSIS: GOOD NEWS AND BAD NEWS

R. J. Beckman and D. W. Whiteman
Los Alamos National Laboratory
P. O. Box 1663
Los Alamos, New Mexico 87545

Abstract

The good and bad characteristics of three methods of uncertainty analysis, propagation of errors, techniques for changing the input distribution, and Latin Hypercube Sampling, are investigated. For fault tree analysis, Monte Carlo is shown to be the best technique. However, nagging questions about the appropiateness of placing distributions on the input variables may render the technique useless. Two methods are given for changing the input distributions to a large computer code, but in some cases these methods are shown to be inefficient. Unbiasedness in Latin Hypercube Sampling is shown to be induced by random selection of the design matrix, and the most efficient designs are not related to the monotonicity of the function.

## I. Introduction

Propagation of errors through fault trees, methods for investigating the sensitivity of the output variables to changes in the input variables and Latin Hypercube Sampling are methods commonly employed in uncertainty analysis. The good characteristics of these methods have for the most part been well documented by the originators and in applications. On the other hand, apparent severe drawbacks with each have been generally ignored or dismissed. The purpose of this manuscript is to critique these three techniques of uncertainty analysis; showing both the good and bad characteristics of each.

Uncertainty analysis is viewed in two ways. The first is error analysis in which the variances or "errors" in the inputs to a computer code are propagated by various methods to obtain the errors in the output variables. The second is sensitivity analysis where the investigator deduces the most important input variables in terms of increasing the "errors" in the output variables. While these two analyses are related their objectives are completely different, and as such require different techniques, and while some techniques are appropriate for one type of analysis they may not be appropriate for the other. For each method investigated the type of analysis; either error or sensitivity, for which the method was developed will be stated.

In section 2, methods are investigated for the propagation of errors through given fault trees. Methods for the propagation of errors in the distribution functions of the input variables is given in section 3, while Latin Hypercube sampling is investigated in section 4.

## II. Propagation of Errors in Fault Trees.

In a study by Martz, Beckman, Campbell, Whiteman and Booker,[1] various methods were investigated for the propagation of errors in coupled nuclear power plant safety system fault tree models for the Arkansas Nuclear One Unit 1

(ANO-1) power plant. Fault trees models were constructed for the front-line safety system and were analyzed by Sandia Laboratories using the SETS code[2] to produce a Boolean expression in terms of the minimal cut sets. An example, of a fault tree, the two out of three monitoring system found in Henley and Kumamoto[3], is given in Figure 1. The Boolean expression for the system unavailability for this model is

$$P_s = P_1P_2 + P_2P_3 + P_1P_3 - 2\,P_1P_2P_3 \quad ,$$

where $P_1$ is the component unavailability. Martz et. al.[1] assumed that the uncertainty in the basic event probabilities, $P_i$, could be represented by a probability distribution, and that these uncertainties could be propagated to obtain the uncertainty in the overall system unavailability probability $P_s$.

The methods investigated error for propagation were: (1) the method of moments where the distribution for the top event was either the same as the basic events (MM), normal (MM-N), or lognormal (MM-LN), (2) the method of moments using Tchebyshev's inequality (MM-TI), (3) propagation by discrete probability distributions (DPP) of Kaplin[4], and (4) Monte Carlo, (MC). Comparisons across the four schemes are facilitated by box plots which are defined in Figure 2 and illustrated for a typical fault tree in Figures 3 and 4. The "truth" was determined by a large scale Monte Carlo simulation. The figures show the only reliable method of propagation to be Monte Carlo. The "good news" from this study is: given the "Boolean" expression for the fault trees, Monte Carlo produced the best results and these simulations can be accomplished in a finite amount of time using a $3000 home computer. The "bad new" from the study concerns the assumptions. If only the first two moments are assumed known and no other assumptions are made about the form of the distribution, the best that can be done in estimating the unavailability percentiles is given by the Tchebyshev inequality. These bounds could be very conservative for the upper percentiles.

## III. Changing the Input Distribution

Many time investigators desire to change the distribution of the inputs of large scale Monte Carlo computer codes. Due to cost constraints however, these computer runs are rarely made, and thus one is unsure of the sensitivity of the output to the distribution of the input.

The good news is that two techniques which permit the investigator to change the input distributions without rerunning the code have been developed. The first method (I) is a weighting method, like importance sampling, which gives unbiased estimates of functions of the output variables. The second technique (II) involves a random selection of the existing input data which changes the distribution of the output.

Let the input variables $X_i$ = 1,2,...,n have density $f(x_i)$ and suppose that the output variable $Y_i = h(X_i)$. Also suppose that $\theta_f = E_f(g^*(Y)) = F_f\{1/\Sigma g(Y_i)\}$ is the parameter to be estimated. Then under density f

$$E_f\{g^*(Y)\} = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} g(y_i) f(x_i)\, dx \quad .$$

An unbiased estimator of the parameter $\theta_{\tilde{f}}$ when the $X_i$ come from density $\tilde{f}(x)$, can be obtained by weighting $g(Y_i)$ by $w_i = \tilde{f}(x_i)/f(x_i)$. That is

$$g_w(Y) = \frac{1}{n} \Sigma\, w_i g(Y_i).$$

Then, since

$$E_f(w_i g(Y_i)) = \int_{-\infty}^{\infty} w_i g(y_i) f(x)dx = \int_{-\infty}^{\infty} g(y_i)\tilde{f}(x)dx = E_{\tilde{f}}\,(g(Y_i)) \quad ,$$

$g_w(Y)$ is an unbiased estimator of $E_{\tilde{f}}(g^*(Y))$.

A second method for changing the input distribution involves discarding pairs of $(x_i, y_i)$ leaving $x_i$ with the "pseudodensity" $\tilde{f}(x)$. Let M be a uniform bound (if one exists) for the ratio $\tilde{f}(x)/f(x) \leq M$. Also let the random variable V given $X = x$ have a uniform distribution between 0 and $(Mf(x))$. Then $x_i$ and the corresponding value of $y_i$ are retained in the sample if the realization v of $V/X$ is less than $\tilde{f}(x_i)/Mf(x_i)$. Letting $X^* = x_i$ if $x_i$ is selected then

$$\Pr\{X^* < x\} = \Pr\{X^* < x \text{ and } X^* = x_i\}/\Pr\{X^* = x_i\}$$

$$= \int_{-\infty}^{x} f(x) \cdot \frac{\tilde{f}(x)}{Mf(x)}\, dx \;/\; \int_{-\infty}^{\infty} \frac{f(x)\,\tilde{f}(x)}{Mf(x)}\, dx$$

$$= \int_{-\infty}^{x} \tilde{f}(x)\, dx = \tilde{F}(x).$$

Hence, the selected $X_i$ have $\tilde{f}$ as a density.

At this time little is known of the properties of these two schemes. However, some unfavorable characteristics of them are evident. First, for Method I, the resulting estimators are not bounded. This may lead to some very poor estimators. For example, in the special case of estimating the distribution function of the output variables by the estimator

$$g^*(y,Y) = \frac{1}{n} \Sigma I(y - Y_i) \quad ,$$

where $I(z) = 1$ if $z \geq 0$, and $I(z) = 0$ otherwise, the estimator of the distribution function under $\tilde{f}$ becomes

$$g_w(y,Y) = \frac{1}{n} \Sigma\, (\tilde{f}(x_i)/f(x_i))\; I(y-Y_i) \quad .$$

Since no restriction placed on the ratio is $\tilde{f}(x_i)/f(x_i)$, $g_w(y,Y)$ can be greater than 1. If these results are scaled to 1, then unbiasedness is lost. Therefore, with this method there may not be a good estimator of the distribution in function of h(x) under $\tilde{f}$. In addition, the variance of the estimator may be very large or not exist. For example let $f(x)$, $\tilde{f}(x)$ be normal with variance 1 and means $\mu_1$ and $\mu_2$, and $Y = h(X) = X$. Then, $g^*_w = 1/n\, \Sigma\, X_i \exp$

$[(\mu_2 - \mu_1)X_i + .5(\mu_1^2 - \mu_2^2)]$ and the variance of $g^*_w$ is

$$\mathrm{Var}(g^*_w) = [(2\mu_2 - \mu_1)^2 + 1] \exp[(\mu_2 - \mu_1)^2] - \mu_2^2$$

which is inceasing in $|\mu_2 - \mu_1|$.

Since a uniform bound $M \geq \tilde{f}(x)/f(x)$ may not exist, it is not always possible to implement Method II. This happens in a surprisingly large number of cases. A simple example of this is the preceeding one of a shift of the mean in normal samples with common variances. In this case the ratio $\tilde{f}(x)/f(x)$ is given by $\exp[(\mu_1 - \mu_2)x + 1/2(\mu_2^2 - \mu_1^2)]$, which is a monotone unbounded function of x on the support $(-\infty, \infty)$. The lack of a uniform bound can also occur for densities with finite support. For example if $f(x) \propto x^{\alpha_1}(1 - X)^{\beta_1}$ and $\tilde{f}(x) \propto x^{\alpha_2}(1-X)^{\beta_2}$ over the support [0,1], the ratio $\tilde{f}(x)/f(x) = CX^{\alpha_2-\alpha_1}(1 - X)^{\beta_2-\beta_1}$ is finite over [0,1] only if $\alpha_2 > \alpha_1$ and $\beta_2 > \beta_1$.

## IV. Latin Hypercube Sampling

McKay, Conover and Beckman[5] developed a technique for sensitivity analysis called Latin Hypercube Sampling (LHS). In LHS the experimenter partitions the sample ranges of each of the input variables; he then chooses at random one of the designs, where each input variable occurs once and only once in each partition. For example, suppose there are two input variables, $X_1$ and $X_2$, over the range [0,1]. Also suppose that only three samples are permitted. Then, the ranges of the two input variables are partitioned into three intervals, $I_1 = [0,1/3]$, $I_2 = [1/3, 2/3]$ and $I_3 = [2/3, 1]$, and pairs of intervals are drawn at random without replacement for both of the input variables. This results in six possible combinations or designs. Let $(I_i, I_j)$ represent the fact that $X_1$ is in interval $I_i$, while $X_2$ is in interval $I_j$. Then the six possible designs for three samples are:

$$D_1 = [(I_1, I_1), (I_2, I_2), (I_3, I_3)]$$

$$D_2 = \{(I_1, I_1), (I_2, I_3), (I_3, I_2)\}$$

$$D_3 = \{(I_1, I_2), (I_2, I_1), (I_3, I_3)\}$$

$$D_4 = \{(I_1, I_3), (I_2, I_1), (I_3, I_2)\}$$

$$D_5 = \{(I_1, I_2), (I_2, I_3), (I_3, I_1)\}$$

$$D_6 = \{(I_1, I_3), (I_2, I_2), (I_3, I_1)\}$$

The Latin Hypercube Procedure for this example requires that one of these six designs be drawn at random and $X_i$'s drawn according to their distribution over the selected intervals. It should be noted that the LHS procedure forces the values of the $X_i$ to be spread across their entire range. Therefore, in terms of sensitivity analysis, this procedure is not as likely as random sampling to miss those portions of the range of the input variables which could greatly influence the output variable. In addition, McKay, Conover and Beckman[5] gave unbiased estimators of parameters of the output variables, which for many fuctions have a smaller variance than unbiased estimators obtained using random samples.

It would seem then that LHS sampling has the desired properties of being both useful in sensitivity analysis and producing reduced variance unbiased estimators for error analysis. This is <u>not</u> true! The unbiasedness of the LHS estimators is an artifact of the random selection of the design. Once the

design has been chosen, the resulting estimator may not be unbiased. The following examples will illustrate this point.

Consider the following four functions of the input variables $X_1$ and $X_2$:

$$f_1(X_1,X_2) = X_1 + X_2,$$

$$f_2(X_1,X_2) = X_1X_2,$$

$$f_3(X_1,X_2) = (X_2 - \mathrm{Sin}(\pi X_1))^2,$$

and $f_4(X_1,X_2)$, a continuous monotone function which is closely approximated by the bivariate step function which take on the values.

| | | | | |
|---|---|---|---|---|
| | $I_3$ | 18 | 18 | 0 |
| $X_2$ | $I_2$ | 18 | 9 | 0 |
| | $I_1$ | 9 | 0 | 0 |
| | | $I_1$ | $I_2$ | $I_3$ |
| | | | $X_1$ | |

$X_1$ and $X_2$ are assumed to have a uniform distribution and $E(Y) = E[f_1(X_1,X_2)]$ is to be estimated. The expected value of the functions and their bias, given any of the six designs, are given in Table 1. Table 2 contains the variance of the average of three realizations of the function and the efficiencies of the six designs, where the efficiencies are measured by the ratio of the mean square error of the estimator for each design to the variance of the function obtained using random sampling.

The LHS estimator of the function $f_1(X_1,X_2) = X_1 + X_2$ is unbiased given any of the six designs, and is 27 times more efficient than random sampling independent of the design. LHS estimators of the means of linear functions, such as $f_1$, are always unbiased reguardless of the design chosen, and they are more efficient than those obtained by random sampling.

Nonlinear functions such as, $f_2(X_1,X_2) = X_1 X_2$, are usually biased given a LHS design. From Table 1, the LHS estimator is most biased for the two extreme (in the sense of the correlation between $X_1$ and $X_2$) designs $D_1$ and $D_6$. In addition, the efficiencies for either of these two designs is a factor of two lower than the efficiency of any of the other four designs.

One might expect that with more samples estimators of the mean of $f_2$ using designs $D_1$ or $D_6$ would become unbiased. Actually, the opposite is true, for as n, the number of cells in the design approaches $\infty$, the expected value of $f_2$ using $D_1$ approaches .33 and the bias goes to .08. The bias in $f_2$ under design $D_1$ is smallest when there is one observed value of $(X_1,X_2)$, and it increases monotonically with n.

example, consider the function $f_3(X_1,X_2) = [X_1 - \sin(\pi X_2)]^2$. This function is both unbiased and has highest efficiency for designs $D_1$ and $D_6$, while the mean square error using designs $D_3$ and $D_4$ is actually larger than that for random sampling.

That function $f_3$ is better for designs $D_1$ and $D_6$ is not necessarily due to the nonmonotinicity of $f_3$. Consider, for example, function $f_4$. For the sake of easily obtaining the variance of the estimator for the various designs we assume that $f_4$ is a step function taking on the values in the function definition. We can see from Tables 1 and 2 that the "linear" design $D_6$ is the best design while the intermediate design $D_5$ is the worst. Therefore, the most efficient designs can not be predicted even if the underlying function is monotone in its variables.

## V. Conclusions

It has been shown that three methods which either have been or will be used in uncertainty analysis have both their good and bad characteristics. We have shown that "error" propagation or analysis for fault trees can be carried out by Monte Carlo at a reasonable cost, but that we may be "wrong" in placing distributions on the input variables. In addition we showed two methods for changing the distribution of input variables, but the methods may lead to highly variable estimators or the variance of the estimators may not exist. Finally we have shown that in most applications the use of Latin Hypercube Sampling leads to biased estimators, but in terms of mean square error, these estimators are usually better than those based on random sampling.

## References


1. H. F. Martz, R. J. Beckman, K. Campbell, D. E. Whiteman, J. M. Booker, "A Comparison of Methods for Uncertainty Analysis of Nuclear Power Plant Safety System Fault Tree Models", Los Alamos National Laboratory report, NUREG/CR-3263, LA-9729-MS (1983).

2. R. B. Worrell and D. W. Stack, "A SETS Users' Manual for the Fault Tree Analyst," Sandia National Laboratories report SAND77-2051 (1978).

3. E. Henley and H. Kumamoto, Reliability engineering and Risk Assessment (Prentice-Hall, New Jersey, 1981).

4. S. Kaplan, "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations-Application to Sersmic Risk Assessment," Risk Anal. 1, 189-146 (1981).

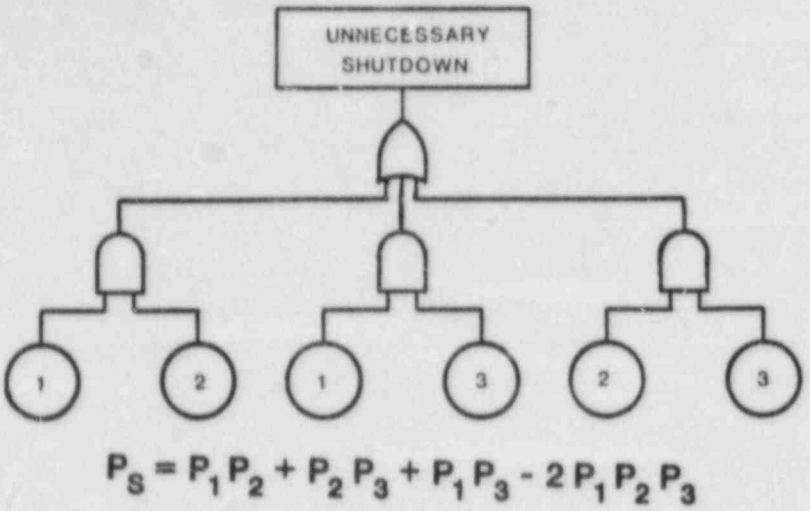
5. M. D. McKay, W. J. Conover, R. J. Beckman, "A Comparison of Three Methods for Selecting Values of Input Variables on the Analysis of Output from a Computer Code," Technometrics, 21, 239-245 (1979).

Table 1. Expected Values and Bias for Four Functions and Six Design Matrices

| Function | Expected Value | $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_5$ | $D_6$ |
|---|---|---|---|---|---|---|---|
| $f_1 = X_1 + X_2$ | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $f_2 = X_1 X_2$ | .25 | -.07 | .04 | .04 | -.04 | -.04 | .07 |
| $f_3 = [X_1 - \sin(\pi X_2)]^2$ | .20 | .00 | -.11 | .11 | .11 | .11 | .00 |
| $f_4$ | 8.0 | -2.00 | 1.00 | -2.00 | 02.00 | 4.00 | 1.00 |

Table 2. Variance and Efficiencies for Four Functions and Six Design Matrices

| Function | Variance | $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_5$ | $D_6$ |
|---|---|---|---|---|---|---|---|
| $f_1 = X_1 + X_2$ | .056 | 27.0 | 27.0 | 27.0 | 27.0 | 27.0 | 27.0 |
| $f_2 = X_1 X_2$ | .016 | 2.3 | 4.4 | 4.4 | 4.4 | 4.4 | 2.3 |
| $f_3 = [X_1 - \sin(\pi X_2)]^2$ | .017 | 1.7 | 1.0 | .8 | .8 | 1.0 | 1.7 |
| $\sim f_4$ | 20.67 | 5.2 | 20.6 | 10.3 | 10.3 | 1.3 | 20.6 |




$$P_S = P_1P_2 + P_2P_3 + P_1P_3 - 2P_1P_2P_3$$

Figure 1.

# NOTATION FOR BOXPLOTS

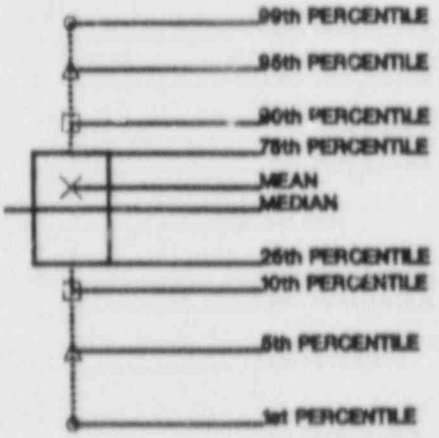



Figure 2.

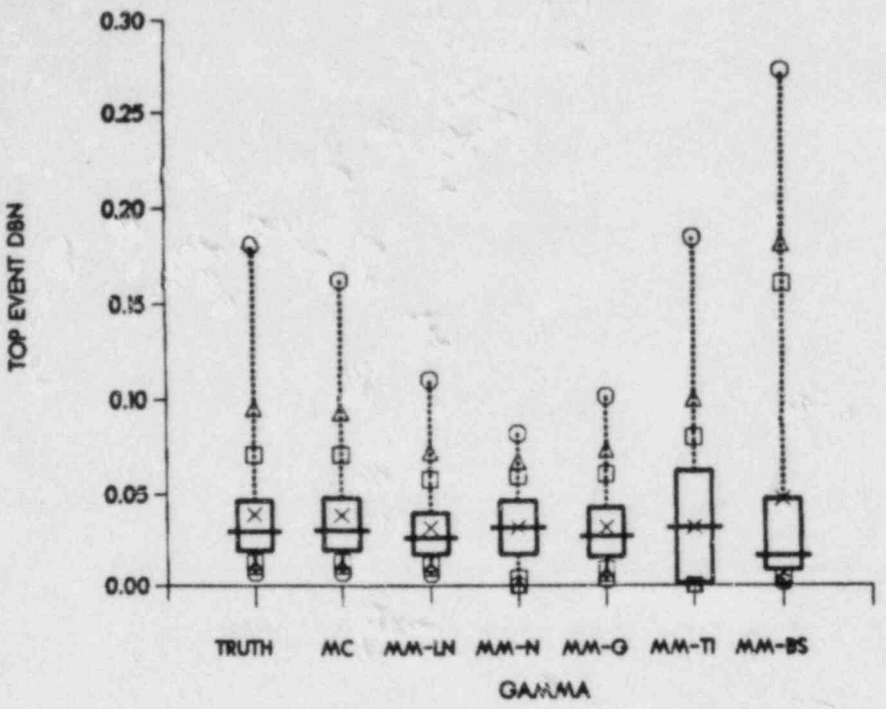



Figure 3.

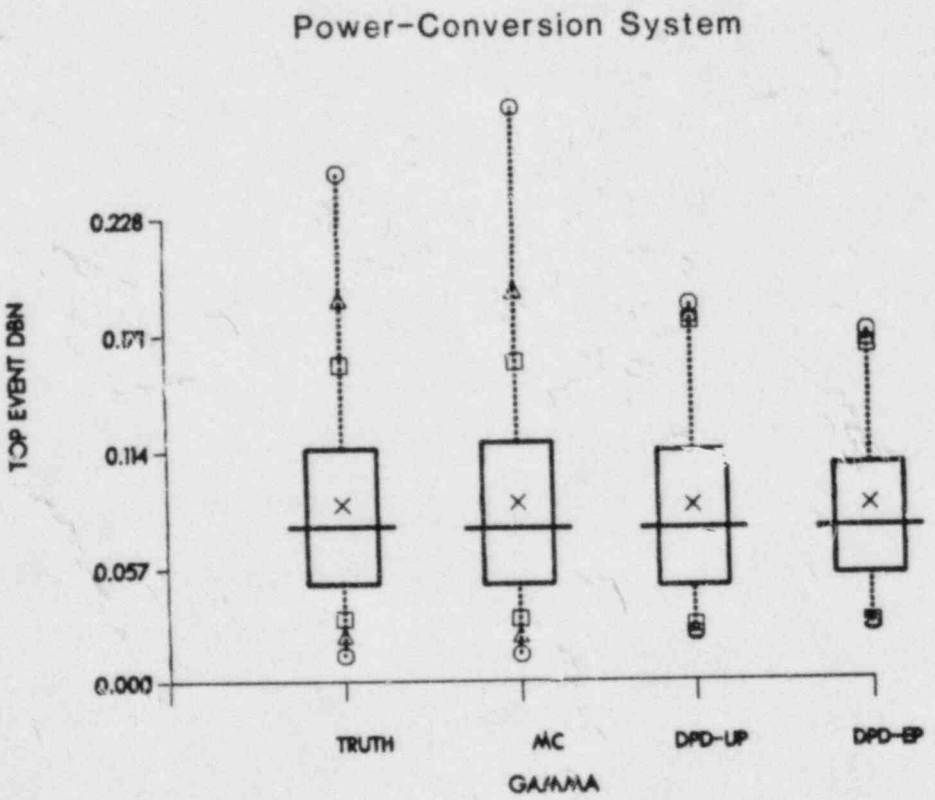



Figure 4.

# RECENT DEVELOPMENTS IN SENSITIVITY ANALYSIS

by

Ronald L. Iman
Sandia National Laboratories
Division 1223
Albuquerque, NM 87185

(Abstract)

(Tutorial Session 1)

# RECENT DEVELOPMENTS IN SENSITIVITY ANALYSIS

Ronald L. Iman

## ABSTRACT

This presentation was a preview of a much larger effort at Sandia National Laboratories involving the comparison of techniques for uncertainty analysis and sensitivity analysis for use with computer models in risk assessment applications. That effort compares the techniques of (1) response surface fitting using fractional factorial designs, (2) Latin hypercube sampling, and (3) differential analysis. The comparison utilizes three real computer models used in risk assessment applications associated with severe accidents at nuclear reactors and with geologic isolation of radioactive waste. Since the paper with the results of these comparisons is quite lengthy, no attempt is made to summarize the results in these proceedings; rather it is suggested that interested readers contact the author directly to receive a copy.

# TOXICOKINETICS AND RISK ASSESSMENT

by

Abe Silvers, Paolo Ricci, and Ron Wyzga
Electrical Power Research Institute
74 Roosevelt Circle
Palo Alto, CA 94306

(Abstract)
Contributed Paper Session 2

# TOXICOKINETICS AND RISK ASSESSMENT

Abe Silvers, Paolo Ricci, and Ron Wyzga

## ABSTRACT

Health risk assessment, the estimation of the probability of incurring an adverse health effect, given exposure to a toxicant, relies heavily on the use of dose-response functions and the extrapolation of animal data to man. Recent work (1) with the chemical vinyl chloride has illustrated the possible usefulness of incorporating chemical kinetic data into a risk assessment. Other authors have recognized the significance of these approaches (2). To further study the use of toxicokinetics in risk assessment, an Electric Power Research Institute workshop was held on the subject. The objectives of the workshop were to explore various questions such as the following:

1. What are the strengths and weaknesses of toxicokinetics in the extrapolation of animal data to man?

2. Can the toxicokinetic study concept, as now practiced, be modified to enhance its application in extrapolation problems?

3. Are there generic structures which can readily be used as models for toxicokinetic studies?

4. Are there generic structures (such as metals) which cannot be successfully studied by accepted toxicokinetic techniques? Are there modifications of approach which could be considered?

5. What areas of risk assessment modeling, incorporating kinetic parameters, should be pursued?

6. What toxicokinetic research problems, associated with risk assessment, should EPRI support?

A fundamental research problem considered at the workshop is that presented by the difference between the administered dose, and the dose delivered to the biological target actually causing the adverse response. In a study, the chemical can be administered intravenously, orally through inhalation, or intramuscularly. These routes may differ from inhalation, ingestion and dermal absorption, normally the ways through which man first comes in contact with a toxicant. A fraction of the original amount administered to a test animal may be delivered to the target organ or cell. Following absorption (for example, diffusion or osmosis), a chemical may undergo such chemical processes as hydroxilation and rendered "safe" for blood transport and excretion. The concentration of the remaining unbound substance may be a fraction of the absorbed concentration. The free chemical may be further converted to a metabolite, the active toxic substance. For instance, the

metabolic process may involve epoxidation, as is the case for vinyl chloride, trichlorethylene and other organic compounds. Thus, for example, although the compound to which man is initially exposed is benzo(a)phyrene, the active carcinogen is an epoxide. The establishment of the target tissue dose (effective dose) is basic to this research.

For most cancers, the actual site for carcinogenic attack is DNA. However, even though data on DNA transformation is not normally available, it was recognized that information collected from metabolic studies, with data indicating length of study, survivorship, and other experimental factors could reduce the uncertainties in the estimates of parameters describing adverse health effects.

Various research directions were suggested. Among them were projects to study the following:

(a) The use of short-term bioassays with pharmacokinetics to estimate risk.

(b) The use of DNA adducts to estimate risk.

(c) The correlation of toxicity to effective dose.

(d) The correlation of various types of effective doses (free or bound concentrations, metabolite concentration, or uptake rate) to outcomes such as tumor incidence or DNA adduct concentration.

A volume of the complete discussions will be published shortly.

References

1. Gehring, P. J., Watanable, P. G., and Park, C. N. "Resolution of Dose-Response Toxicity Date for Chemicals Requiring Metabolic Activation: Example--Vinyl Chloride," Toxicology and Applied Pharmacology, 44, pp. 581-591 (1978).

2. Hoel, D. G., Kaplan, N. L., and Anderson, M. W., "Implications of Nonlinear Kinetics on Risk Assessment in Carcinogenesis," Science, Vol. 215, pp. 1032-1036, March 1983.

# STATISTICAL METHODS USED IN DEVELOPING AN OPTIMUM GLASS FOR VITRIFICATION AND STORAGE OF NUCLEAR WASTES

by

W. M. Bowen and L. A. Chick
Pacific Northwest Laboratory
P.O. Box 999, Richland, WA 99352

(Abstract)
Contributed Paper Session 2

# STATISTICAL METHODS USED IN DEVELOPING AN OPTIMUM GLASS FOR VITRIFICATION AND STORAGE OF NUCLEAR WASTES*

W. M. Bowen and L. A. Chick
Pacific Northwest Laboratory
P.O. Box 999, Richland, WA 99352
(509) 375-2979

A proposed method of storing certain nuclear wastes, which has been under study for some time, involves adding glass-forming chemicals to the waste stream so that the mixture can be melted to form a glass. The resulting melt is then poured into a metal canister for cooling and storage in a geologic repository. In developing a waste glass composition for this purpose there are two composition dependent properties which must be minimized. These are: 1) leaching (i.e., dissolution by contact with ground water) and 2) crystal formation as the glass cools from 1000° to 500°C. In minimizing these two properties, the only feasible compositions are those which are compatible with the liquid-fed ceramic melter (LFCM). There are three major LFCM process influence properties which must be observed. First, viscosity of the glass must be approximately 100 poise at the operating temperature 1150°C to facilitate homogeneity and pouring of the glass. Second, there must be minimal crystallization between 1000° and 1150°C to avoid crystal and sludge buildup at "cool spots" in the melter. Third, electrical conductivity of the composition must be between 0.15 and 0.50 $(ohm\text{-}cm)^{-1}$ in order to melt the glass under current power supply restrictions.

This paper presents the statistical methods used in a current study to develop a waste glass composition with minimum leaching and crystallization, subject to constraints on the three LFCM process influence properties. The computer-aided design of a D-optimal seven component mixture experiment is outlined with a discussion of anomalies, modifications, and deviations from classical design of mixture experiments. Methods for analyzing, displaying and interpreting the component effects are presented. Scheffe polynomial models, fitted to the experimental data, express the five properties of interest as separate functions of composition. Model selection and validation are described. The fitted models are then utilized with nonlinear optimization techniques to locate the optimum waste glass composition, subject to constraints on the seven components and on the three LFCM process influence properties.

A formal PNL technical report which gives a detailed account of this project will be available on request from the authors after March 1984. We apologize for not being able to meet the deadline for inclusion in the Proceedings.

---

Prepared for the U.S. Department of Energy under Contract DE-AC06-76RLO 1930.

# PROBABILISTIC FRACTURE MECHANICS: STATISTICAL, MATHEMATICAL, AND MODELING ISSUES

by

David C. Cox and Robert E. Kurth
Battelle Columbus Laboratory
505 King Avenue
Columbus, OH 43201

(Abstract)
Contributed Paper Session 2

# PROBABILISTIC FRACTURE MECHANICS: STATISTICAL, MATHEMATICAL, AND MODELING ISSUES

by

David C. Cox and Robert E. Kurth

Battelle's Columbus Laboratories

Presented at

1983 Statistics Symposium on National Energy Issues

October 20, 1983
Rockville, Maryland

## ABSTRACT

Over the last several years, Battelle has been developing a probabilistic fracture mechanics (PFM) capability. The purpose of the program is two fold: first to examine, modify, or develop probabilistic methods for use in the analysis of structural reliability and, secondly, to demonstrate the use of such methods through their application in simplified structural analysis.

This paper presents the results of the probabilistic model development effort for three probabilistic techniques: Monte Carlo, Markov chain, and discrete probability distributions. In addition, three case studies are presented in which each of these methods was compared and recommendations for their suitability in structural analysis were made. The structures included a bridge component, a pipe-to-vessel weld in a nuclear piping system and a steel plate. The results of these three studies indicate that Monte Carlo analysis is the preferred technique for use in structural analysis with the Markov chain model being used for sensitivity analysis. The discrete probability distribution method is found to be accurate if enough data points are used to represent the probability density function of the random variable. However, since classic confidence bounds cannot be estimated for the discrete probability method, several runs are usually necessary to have confidence in the results. KEY WORDS: STRUCTURAL RELIABILITY, MONTE CARLO, MARKOV CHAIN, DISCRETE PROBABILITY DISTRIBUTIONS, PROBABILISTIC ANALYSIS

## INTRODUCTION

The purpose of this paper is to discuss some of the statistical and modeling problems encountered during a progam to develop probabilistic fracture mechanics (PFM) models and present the methods used to address these problems. Included in the discussion are problems, together with the associated technique used to solve them, encountered during both the development phase and the subsequent application.

All of the discussion in this paper is in the context of stable crack growth which is governed by Linear Elastic Fracture Mechanics (LEFM). It is important to note that PFM models are not so narrow in scope. In fact, we have developed such models for fast fracture and crack initiation processes which are not governed by LEFM. However, since the primary focus of this paper is on statistical and probabilistic modeling methods, only the LEFM models are presented to minimize the amount of fracture mechanics theory which needs to be presented.

The paper is divided into three sections. The first defines PFM models and how they can be used in structural reliability applications. The second presents the actual PFM models and how they are used during structural analysis. Finally, the analysis performed during three case studies is given during which a discussion of the modeling problems and solutions is presented.

## Probabilistic Fracture Mechanics Models - A Defintion

The ultimate goal of any fracture mechanics analysis is to assess the effect of defect growth on structural integrity. For example, any welded structure, such as bridges, nuclear piping systems, offshore oil platforms, etc., will have defects in them due to imperfect welding processes. These types of structures are the primary focus of this paper; they can be analyzed at least conservatively, by Linear Elastic Fracture Mechanics (LEFM). According to the LEFM theory, the growth of cracks is governed primarily by the current crack size, normally denoted a, and the applied stress, denoted $\sigma$. Thus, if the crack[1] is small enough or the stress level low then no, or very little, crack growth will occur. (The empirical relationship governing crack growth is discussed later). For laboratory conditions in which the material laws are well-defined, the initial crack size is precisely known and the applied stresses tightly controlled very good predictions of crack growth are obtained. Unfortunately during service, there is uncertainty and random fluctuations in each of these processes. The purpose of developing PFM models is to account for the uncertainty and random variation in parameters and models so as to include the stochastic nature of the fracture process during the structural

---

[1] In this paper the terms crack and defect are used interchangably. This is for the convenience of the authors. Many fracture analysis reserve the term defect for the material state prior to crack initiation.

evaluation. In developing PFM models and their associated data bases, no distinction is made between uncertainty (due to a lack of knowledge) and random fluctuations (due to true stochastic variations).

A PFM model can be constructed using many different probabilistic techniques. These techniques can be broadly classified into two categories: random-parameter and shock models. Random-parameter models are based on an underlying mechanistic (deterministic) theory in which the inputs are distributions rather than point values. The selected probabilistic methods, e.g., Monte Carlo, is then used to combine these distributions according to the underlying mechanistic principle to obtain an output distribution for the quantity or quantities of interest. For PFM analysis, this is normally the time dependent distribution of crack sizes from which the expected time to failure, probability of failure, and so on, can be determined. On the other hand, shock models require no underlying mechanistic principle but rather the philosophy is that experimental results will dictate the correct choice of model parameters which represent the mechanistic process without actually having derived this relationship. The term shock is used because, for PFM models, two assumptions are necessary for this type of development. First, it is assumed that the applied load is viewed as a "shock" at the crack tip. If this shock level exceeds a critical value, then the crack advances; if not, there is no growth. The second assumption is that the process is Markovian, that is, the advance of the crack is only dependent on the current crack size and the shock level.

In summary, a PFM model is a technique for describing the growth of cracks in a structure using either random-parameter or shock model methods for including the uncertainty and stochastic variation of material properties, initial defect sizes, environments, and loads in a structural assessment.

## PFM Model Descriptions

During the course of PFM model development at Battelle, several different types of probabilistic techniques have been examined for use in the model construction. Three of the techniques are described and their method of application explained in this paper. These are: Monte Carlo, Discrete Probability Distributions (DPD's) and Markov chain. Before describing these methods, the basic principles of LEFM are presented.

### LEFM Theory

LEFM determines the amount of crack growth from a correlation relating crack growth to a measure of the stress field strength at the crack tip, called the intensity factor, denoted K. For simple loadings and geometries, K can be calculated analytically. For more complex problems, numerical solutions are required. Since the primary focus of this study is on probabilistic methods, relatively simple structural configurations have been studied. While a variety of crack growth laws have been proposed, the earliest, and simplest,

of these forms is used throughout the three case studies. This form is known as the Paris' Law equation[1] and is given by

$$da/dN = C'(\Delta K)^n \tag{1}$$

where

$\Delta K = \sigma f(a,w) \sqrt{\pi a}$
f(a,n): function of geometry which is set equal to one for these studies
c',n: empirically determined constants
N : number of stress cycles
a : crack length

For constant applied stress ranges[2] and a known initial crack size equation (1) can be integrated to yield

$$a_f = (a_{om} + C\pi^{n/2}\sigma^n N_f)^{1/m} \tag{2}$$

where

$m = 2 - n/2$
$C = n/2\ C'(2/(2 - n))$
$N_F$ = final cycle
$a_f$ = final crack size

Because the applied stress, material properties, and initial crack size are not known precisely this deterministic form can only provide a rough estimate of the final crack size. Because of the variable loadings, we use a Taylor series approximation to equation (2) in which the interval of integration, denoted $\Delta N$, is assumed to be small enough that the stress range can be assumed to be constant. The series is terminated after the linear terms to yield

$$a_{t+1} = a_t^m + C\ \sigma^n \Delta N \tag{3}$$

This is the basic underlying deterministic model for the random--parameter models.

---

[2] Stress ranges are used since fatigue crack growth is being considered. Creep crack growth is not discussed.

## Random-Parameter Models

During the three case studies, two random-parameter models were used: Monte Carlo and DPD's. Each of these is presented below.

Monte Carlo. The Monte Carlo technique is a simple method for adding a probabilistic structure to a deterministic model. Suppose the output, $L_c(t)$, is related to the individual inputs, $L_i(t)$, by a function

$$L_c(t) = f\,(L_1(t), L_2(t), \ldots, L_n(t))$$

where the function f may not even be analytic, e.g., a computer program. If each of the inputs has been characterized by a probability density function (PDF) during data analysis, then the following procedure is used during a Monte Carlo simulation. The cumulative distribution function (CDF) of each input is generated by integrating the PDF. A random number between 0 and 1 is generated, call it $r_1$. The CDF is inverted and the value for the load, $L_j(1)$ is determined. This method of choosing the value is repeated for each individual load. A value of $L_c$ is then calculated as

$$L_c(1) = f\,(L_1(1), L_2(1), \ldots, L_n(1)) \quad .$$

The entire process is repeated a large number of times, say M. What results is a M-dimensional vector of the output: $(L_c(1), L_c(2), \ldots, L_c(M))$. This vector is used to construct a histogram which can be analyzed statistically to obtain estimates of the mean, kurtosis, probability of failure and so on. Obviously, in the limit as M tends to infinity, the continuous distribution will be asymtoptically approached. Equally obvious the computer time will also increase. As an alternate method of sampling, importance sampling schemes[2] can be employed to reduce computational time.

Discrete Probability Distributions (DPD's). The description of DPD's follows the conventions set forth by Kaplan (3). In this method, the initial input distributions are discretized into m values. Each value of each variable is then assigned a probability of occurrence. Additionally, the various forms of any probabilistic function are assigned a probability of being correct. If these discrete values are paired with their probabilities, the following vectors of ordered pairs result for two loads X and Y.

$$X = [(x_1, p_1), (x_2, p_2), \ldots, (xm, pm)]$$
$$Y = [(y_1, q_1), (y_1, q_2), \ldots, (ym, qm)] \quad .$$

The number of discrete points in each of these vectors has been chosen to be the same although it is not necessary to do so. The addition of two discrete vectors is defined by

$$Z = Y + X$$
$$Z = (Y_i, p_j) + (X_j, q_j) \quad , \text{ and}$$
$$Z = (X_j + Y_i, P_i * q_j) \text{ for all } i \text{ and } j \quad .$$

Therefore, the addition of two vectors containing m ordered pairs each results in a vector which has $m^2$ ordered pairs. The multiplication of DPD's is similarly defined.

Since, even for relatively small values of m and K (on the order of 10), the computer storage capability will quickly be exceeded, it is necessary to examine some procedure for reducing this vector's size. This leads to an examination of the condensation procedure discussed below.

In order to illustrate the condensation procedure, assume that the initial DPD for two inputs contain 20 ordered pairs, respectively. The output distribution will then be a vector of 400 ordered pairs after each of individual DPD's have been combined. However, it has been assumed that 20 orderd pairs adequately describe the distribution. Suppose the range of possible values is divided up into equal intervals. Further, for the sake of example, assume that the new values between 144 and 188 in this vector, denoted L', fall in the 6th interval. Then

$$P_6 = \sum_{i=144}^{188} p_i \quad , \text{ and}$$

$$Z_6 = \frac{1}{P6} \sum_{i=144}^{188} p_i \, Z_i$$

where $L' = [(Z_i, p_i)]$, $i = 1, 2, \ldots, 400$. This procedure can be written in general as

$$\hat{P}_i = \sum_{S_i} p_j$$

$$Z_i = \frac{1}{p_i} \sum_{S_i} p_j \, Z_j$$

where

$$S_i = \{j \mid d_i < Z_j < d_{i+1}\}$$
$$\Delta = (a_{max} - a_{min})/20$$
$$d_i = a_{min}, \; d_i + 1 = d_1 + \Delta$$

For many DPD applications, it is not possible to use equally spaced intervals. In fact, logarithmically spaced or other time independent unequal interval spacing schemes may lead to the same problem. A method for calculating time dependent bin sizes has been devised by Battelle and is used to condense the DPD at each time step. In this method, after each time step, the largest and smallest values of the vector L' are determined. The intervals are then determined from

$$b_1 = a_{min}$$
$$b_1 = b_{i-1} + 2(i - 1)\ (a_{max} - a_{min})/N/(N - 1),\ i = 2, \ldots, N$$

where

$a_{min}$ = minimum value
$a_{max}$ = maximum value
$N$ = number of discrete intervals
$b_i$ = interval endpoints for condensation.

Markov Chain Models. In discussing the Markov model, some of the conventions put forth by Bogdanoff[4] are used. In the Markov mode the crack is defined by discrete states with time being measured by duty cycles. In addition, the Markov assumption is made, i.e., the probability that the crack which is currently in state i will be in state j during the next duty cycle is only dependent on its present state and not on the previous load history. Consider

$$t = 0, 1, 2, 3, \ldots$$

which need not be of equal duration. The crack is defined by variable states

$$a = 0, 1, 2, \ldots, n$$

where a = 0 implies no crack and a = n may be defined as the limit state. The initial distribution of cracks is defined by

$$I = (i_1, i_2, \ldots, i_n)$$

by design codes or other data. The transition probabilities, $P_{ij}$, are defined as the probability that given the crack is currently in state i it will next be in state j. The evolution of the crack growth process is given by

$$P_t = I\ M^t$$

where

M = matrix of transition probabilities $P_{ij}$
$P_t$ = ($P_t(1)$, $P_t(2)$, ..., $P_t(n)$, PDF of loads at time t.

Once the transition probabilities, $P_{ij}$, are defined, the Markov model provides an inexpensive, simple method for calculating the distributions at any time. However, several drawbacks to the Markov model exist. First, recent crack growth rate data which show very limited scatter in growth rate under steady state (K = constant) testing conditions[5] suggest that the assumption of a shock model for the crack growth may not be correct. Secondly, it is difficult, if not impossible, to find the appropriate crack growth data in sufficient quantity to calculate the transition probabilities. To generate such data experimentally would be expensive, although it would only be necessary to do so once for a specified load distribution. If crack growth data cannot be found, the transition probabilities can be determined from material life data for a given load distribution. In this case, however, the Markov model provides a curve fitting procedure and it is difficult to attach physical significance to the Markov model parameters.

The evaluation of each of these probabilistic methods has been performed during the course of three case studies. Each of these studies are summarized below.

## Case Studies for the Evaluation of Probabilistic Fracture Mechanics Models

Three case studies have been performed to evaluate the effect of including probabilistic techniques in the structural integrity models. The first case study examines the effect of using random load sequencing in place of a deterministically fixed sequence of loads which represents the expected frequency of the individual load transients. The second case study compared Monte Carlo and Markov Chain models. Finally, a comparison of Monte Carlo and DPD models was made. The details and results of each study are summarized below.

### Case Study I: Probabilistic Modeling of Loads in Nuclear Piping Systems

The piping System in a nuclear power plant must be designed so that the coolant pressure boundary is maintained throughout the plant life. To insure that the piping system performs this function, many analyses must be performed, including determination of the piping system fatigue crack growth rates. Because of the variability in the initial defect sizes and loads, a deterministic analysis of the fatigue crack growth process must necessarily make worst case assumptions about the initial defect size and coupling of the pipe stresses due to the steady state plant operation and transient events. In this type of analysis, severe transient events, e.g., design basis

earthquakes, which have a relatively low probability of occurrence, are applied at the start of the plant life in order to be conservative during the analysis. The purpose of this study is to compare and contrast the use of such a deterministic analysis with the results of a probabilistic fracture mechanics (PFM) analysis of a selected piping system in a nuclear power plant.

Certain transient load events, while unlikely, are significant for the analysis since the induced stresses in the pipe are large, causing significant crack growth and reducing the overall system life and threatening its integrity. Current analyses tend to hypothesize that major transient events occur early in the plant life. This assumption may causes the crack to reach a size greater than the threshold size associated with typical service stresses very early in its life. The hypothesis that statistically rare transient events, which induce large stresses in the piping system, occur early in the plant life is referred to in this study as the worst case coupling of loads.

An alternative to this type of analysis is probabilistic ordering of the loading coupled with probabilistic fracture mechanics (PFM) methods. The use of probabilistic models allows the analyst to avoid the overly conservative bias in the assumptions of a deterministic analysis with probabilistic information to estimate the behavior of crack growth in a manner that reflects the random nature of the materials, the loadings, and the cracking process.

The purpose of this study is to compare piping system integrity, calculated using each type of analysis, to estimate the relative level of conservatism. The cold leg piping system of a selected nuclear power plant has chosen as a vehicle for this comparison. The deterministic analysis will use many of the techniques and methods described in the "Cold Leg Integrity Evaluation" by M. E. Mayfield, et al.[6]. The probabilistic calculations will be made using a PFM model which employs a Monte Carlo simulation.

The initial crack depth, aspect ratio, yield stress, critical stress intensity factor, and failure crack size for this analysis are presented in Table 1. The chosen initial crack depth was so large so as to almost insure that the threshold would not be a significant factor. For this reason, threshold effects were excluded from this analysis.

TABLE 1. FRACTURE MECHANICS ANALYSIS PARAMETERS

| | |
|---|---|
| Initial Defect Size (1) | 1.43 cm |
| Aspect Ratio | 0.1 |
| Crack Depth at Failure (1) | 3.81 cm |
| Crack Orientation | Axial |

The remaining parameters needed for the deterministic analysis are the loads and load sequence encountered in one lifetime. The number of times each transient occurs has been taken from the vendors design analysis as was done in the Cold Leg Integrity Study. The eleven transient events shown in Table 2 are included in the analysis. It is important to note that during the deterministic analysis, immediately after the plant start-up, Transient 1, an earthquake, is postulated to occur, certainly a good example of worst case loading.

TABLE 2. TRANSIENT EVENTS INCLUDED IN THE ANALYSIS OF PIPING SYSTEM

| Transient Number | Transient Description | Design Cycles |
|---|---|---|
| 1 | Plant Start-Up or Shutdown | 240 |
| 2 | Power Loading or Unloading | 48,000 |
| 3 | Step Increase/Decrease in Power | 8,000 |
| 4 | Reactor Trip | 470 |
| 5 | Turbine Trip | 390 |
| 6 | Rapid Depressurization | 80 |
| 7 | Steam Line Failure | 1 |
| 8 | Hydrotest of 3,125 psi at 400 F | 20 |
| 9 | Earthquake (OBE) | 650 |
| 10 | Earthquake (SSE) | 1 |
| 11 | Vibration | $2.1 \times 10^{10}$ |

For the probabilistic analysis, two key parameters change. First, the initial crack size is a random variable, not a constant. Based on the results of reference(7), it is assumed that the initial crack depth distribution is a Rayleigh distribution. This distribution is skewed right and has a relatively long tail, although not as dramatic as for the lognormal distribution. The parameters for the initial defect distribution are given in Table 3. The second random variable is the frequency of the transient events. For the present analysis, the frequency of these events is chosen to match the frequency of these events over an assumed 40 year plant life used to the Cold Leg Integrity Study. For example, a plant start-up or shutdown (transient 1) will, on the average, occur 240 times during the plant life as given in Table 2. The frequency of all 11 transients has been detailed in Table 2.

Several simplifications were made for the deterministic model to reduce the cost of the analysis to facilitate meeting the objectives of this study. These include: (1) no crack interaction modeling, (2) no threshold effects, (3) one-dimensional crack growth, (4) no elastic-plastic modeling, and (5) trapezoidel rule for integration. These assumptions are different

TABLE 3. PARAMETERS FOR INITIAL CRACK SIZE PROBABILITY DENSITY FUNCTION (PBF)

| PDF Form | Rayleigh |
|---|---|
| Minimum Value | 0.635 cm |
| Modal Value | 1.270 cm |
| Maximum Value | 1.905 cm |

enough from those of the Cold Leg Study to inhibit direct comparison of the deterministic results of this study with the Cold Leg Study. Therefore the deterministic calculation has been made independently of the Cold Leg Study analysis.

The result of the deterministic calculation, gives a failure time of approximately $3.7 \times 10^9$ cycles. Since the loading spectrum adopted in reference[6] was used in this study, the relationship of these cycles to time may be made and is found to be approximately 8 years. Given the assumption that the plant is designed for a 40 year life, this corresponds to a factor of safety, denoted $\beta$, of approximately 0.2.

The results of the probabilistic calculations are presented in Figures 1 and 2. Because the initial crack size is a random variable in the probabilistic model and because of the final crack size's sensitivity to this value, a skewed distribution for the initial size was chosen. This is crucial so that the effect of the worst case loading scheme on the crack growth may be examined without the effect of the initial defect size obscuring the results.

The empirically constructed cumulative distribution function (CDF) of the initial defect size is given in Figure 1. The modal and mean values of this distribution are 1.270 cm and 1.383 cm, respectively. Since the deterministic calculation started at a value of 1.43 cm, this implies that approximately 54 percent of all probabilistic calculations began with a crack size larger than the deterministic value.

The results of the probabilistic fracture mechanics analysis are shown in Figure 2 in terms of the cumulative distribution function of the safety factor, $\beta$. The parameters of the distribution of $\beta$ are

$$\mu = 2.64 \qquad \rho = 1.50$$

where $\mu$ is the mean value and $\rho$ is the standard deviation. The median value from Figure 3 is given as 1.77. Therefore, the best estimate value for the remaining life of the piping system is approximately 70 years.

The immediate conclusion one reaches in comparing these results is that the worst case loading of a nuclear piping system predicts that the

expected life of that system is shortened by a factor of nine. Even if a conservative PFM analysis is being performed, the deterministic worst case sequencing of the loading predicts about a factor of 5.0 reduction in the plant life as compared to the PFM analysis when the 10 percentile value of $\beta$ is used.

As noted earlier, this study made a number of simplifications in the deterministic analysis as compared to reference (1). That deterministic analysis for the same point in the piping system indicated a mean safety factor of 0.77. This is significantly larger than the value of 0.2 calculated during this study, a difference that is not unexpected given the simplifying and conservative assumptions made. However, the PFM analysis, which was performed using the same simplifying assumptions employed during this study's deterministic analysis, predicts that 97 percent of the time the safety factor will exceed that found during the Cold Leg Study. This indicates that the assumption of worst case load sequencing often used in deterministic analysis performed (for example, the Cold Leg Study) introduces a degree of overconservatism sufficient to make the additional complexity, and cost, of the more detailed analysis questionable.

Observe that the present study indicates a factor of nine difference in the safety factor determined from probabilistic and deterministic analyses. While this difference is quite significant, it is biased towards a lower bound in that the present PFM formulation is significantly biased towards lower safety factors for several reasons, as follows. First, in the deterministic load sequence between plant start-up and shutdown, a sampling without replacement scheme is used. In the PFM analysis, a sampling with replacement scheme has been adopted. For calculations in which the safety factor is significantly greater than one, this should not result in any significant difference between the two analyses since the frequency of the various transients will approach the same value, on the average. However, in the present analysis, the component's expected life is less than the 40 year period and it is expected that the sampling with replacement will lead to slightly more conservative results than sampling without replacement. Secondly, the selected density function for the initial crack size leads to a mean value that is higher than the initial crack size used in the Cold Leg Study. Finally, and most importantly, no threshold for the stress intensity factor was used during this study. Because of the prohibitive cost of making Monte Carlo calculations for this case, a statisticaly significant set of crack curves were not generated. However, several sample paths for the crack growth were calculated and the indication is that deleting the threshold stress intensity factor from the PFM analysis reduces the safety factor by an order of magnitude.

## Case Study II: Comparison of Monte Carlo and Markov Chain Methods

The determination of the reliability of structures, such as joints in bridges, requires that the uncertainty in material properties, flaw sizes and location, environments, and loadings be considered in the analysis. Two

popular probabilistic models used to incorporate these uncertainties in crack growth analysis are: (1) Monte Carlo, and (2) Markov chain. This study compares results developed by applying both methods to a problem in which cracks are considered to be growing in bridge-type materials. The effects of variable crack sizes and variables loadings on the growth of a crack are investigated using both methods. The results of these investigations are used to judge each method's applicability to such problems. Before one can proceed with the application of Monte Carlo and Markov chain models (previously described), a basic load cycle together with the load distribution must be defined and sources of variability charcaterized. The situation of a bridge is used as a vehicle for this discussion.

The normal practice for defining a load cycle for a bridge is to examine the stress at a selected point on the bridge during the time it takes a vehicle to cross the bridge. In this case, the load cycle is equal to the time it takes the vehicle to cross the bridge. Since millions of vehicles will cross a bridge during its life, somewhat longer cycles are defined, such as the hourly or daily number of trucks crossing multiplied by this single cycle load. However, for many bridges several vehicles can be on the bridge simultaneously. Then, since the stress at a point on the bridge is a combination of the stresses induced by each vehicle, the probabilistic methods cannot sample from the weight distribution for the vehicles independently.

To formulate stress distributions for some location, the first step taken was to redefine the basic stress cycle as the time between the entry of a vehicle onto the bridge and the first subsequent time at which no vehicles are on the bridge. A calculation of the duration of the load cycle defined in this way was made and compared to the duration under the above noted more common definition. For purposes of this comparison, the speed of the vehicles was assumed to be uniformly distributed between 30 and 40 mph and the interarrival time between vehicles was assumed to be exponentially distributed with a mean of 20 seconds. For this case, the mean duration of the load cycle was determined to be approximately 14 seconds, a result which compares favorably with the 10.4 second duration of the load cycle developed under the more common defintion. Thus, the new definition does not appear to significantly affect the load duration. In addition, this new definition of the load cycle is numerically convenient because it allows a stress distribution to be calculated which is statistically independent between successive cycles. Consequently, this indpendence property simplifies the Monte Carlo model and reduces, to some extent, the computational time of the program. Since the stress distribution due to multiple vehicle crossings cannot be determined from independent sampling of the individual vehicle weight with the former definition of the load cycle, the new definition was adopted for use in this study.

Data on vehicle (truck) weights were obtained from several reports[8-10]. These data were synthesized and input to a stress analysis code using multiple span beam theory to calculate a stress histogram for multiple truck crossings. Thereafter, a Rayleigh distribution was fit to these results. This distribution is used in the subsequent probabilistic calculations. Stresses so determined were assumed to exist in the vicinity of a through edge crack in the flange of a main grider.

Having defined the probabilistic techniques to be used and the method for calculating the statistical distribution for the stress, the first step in applying the PFM methods is to define and calculate the model parameters. These parameters are obtained from the literature for all cases, except for the Markov transition probabilities which are derived as detailed later.

Parameters needed for the Paris law equation are available for typical bridge steels from several sources. In this study the values found in reference[11] were used. These values, which represent A514 steel at finite growth rates, are:

$$c' = 2.4 \text{ E-10 (inch/cycle)/ksi}$$

$$m = 3.0.$$

Data in References[8-10] were used to synthesize the distribution of truck weights. As previously discussed, the stress frequency of occurrence distribution is represented by the Rayleigh distribution, with the truck interarrival time being obtained from Reference[8]. The only remaining parameters for the Monte Carlo analysis are those which define the distribution of the initial crack sizes. Data from Reference[11] suggest the initial crack size as a Rayleigh distribution with the following parameters:

$$a_{modal} = 0.01 \text{ inch}$$

$$a_{min} = 0.005 \text{ inch} \quad .$$

At this point, the Monte Carlo analysis can be performed. It remains to determine the transition probabilities for the Markov analysis.

In order to determine the transition probabilities for the Markov model, it is necessary to have a statistically significant number of crack growth curves for the given applied stress history. Data for crack growth under actual loading conditions are relatively limited. This condition, at first consideration, would suggest that the use of the Markov model is impossible. However, while actual crack data may not exist in sufficient quantity to calculate the transition probabilities directly, these probabilities can be estimated from data giving the time to failure. In this case, there is no way to attach any physical significance to the damage state, since, during the process of estimating the transition probabilities, the number of states is being adjusted so that the Markov calculation will reproduce the cumulative distribution function of the time to failure. It is not appealing to use the Markov model as a curve fitting procedure. Furthermore, the comparison of the two models would be inappropriate if the Markov model were used in this way (since the parameters could be adjusted to provide as accurate a fit to the data as desired). For these reasons, the results of the Monte Carlo calculation are used to simulate a set of crack growth curves.

It is then assumed that these curves represent real data so that the transition probabilities are calculated from this data set.

The data set generated includes 1,000 crack growth curves. The crack growth calculation is stopped when the crack size reaches 0.5 inch. For purposes of Markov analysis, it is assumed that the damage is discretized into equal intervals of 0.025 inch. To calculate the transition probabilities, each of the 1,000 curves is examined to determine how many cracks which are in state i at time t remain in state i, how many move to state i+1, and so on. Knowing the state the crack started in, and examining the next time interval for the state that crack is predicted to be in, the transition probablities are easily calculated.

Consider first the results of the Monte Carlo method shown in Figures 3 and 4. The probability of failure, shown in Figure 3, is indicated to be near zero up to approximately 600,000 cycles, at which time it increases very rapidly. The average crack growth behavior, presented in Figure 4, starts off very slowly but it too increases rapidly at 600,000 cycles.

Figures 5 and 6 also show the result of the Markov chain application to cracks growing in bridges for failure probability and average crack size, respectively. For the transition probability calculation (shown in Table 4), it was initially assumed that there are 20 damage states. What is immediately obvious from these figures is that the Markov model is overpredicting the spread in the crack growth process. In fact, using the following formulas for a unit step model:

$$E(t_f) = \sum_{i=1}^{19} (1 + P_i/(1 - P_i))$$

$$(t_f) = \sum_{i=1}^{19} (P_i/(1 - P_i)^2)^{1/2}$$

where

$E(t_f)$ = expected value of the time to failure
$(t_f)$ = standard deviation of the time to failure

we find

$E(t_f) = 8 \times 10^5$ cycles
$(t_f) = 1.4 \times 10^4$ cycles.

(These are only approximate values since the model is not strictly a unit step model.) The corresponding results for the Monte Carlo analyses are

TABLE 4. TRANSITION PROBABILITY MATRIX FOR THE MARKOV MODEL

| Damage State I | Transition Probability of Crack Moving from Damage State I to J* | | |
|---|---|---|---|
| | J = I | J = I + 1 | J = I + 2 |
| 1 | .9486 | .0514 | 0 |
| 2 | .9812 | .0188 | 0 |
| 3 | .9580 | .0420 | 0 |
| 4 | .9297 | .0703 | 0 |
| 5 | .8966 | .1391 | 0 |
| 6 | .8609 | .1391 | 0 |
| 7 | .8196 | .1804 | 0 |
| 8 | .7788 | .2212 | 0 |
| 9 | .7294 | .2706 | 0 |
| 10 | .6807 | .3193 | 0 |
| 11 | .6375 | .3675 | 0 |
| 12 | .5737 | .4263 | 0 |
| 13 | .5220 | .4780 | 0 |
| 14 | .4681 | .5319 | 0 |
| 15 | .3980 | .6020 | 0 |
| 16 | .3391 | .6609 | 0 |
| 17 | .2811 | .7189 | 0 |
| 18 | .2005 | .7923 | .0072 |
| 19 | .1532 | .8408 | 0 |
| 20 | 1.0 | 0 | 0 |

* All other entries (not shown) are equal to zero.

$$E(t_f) = 8 \times 10^5 \text{ cycles}$$
$$t_f = 3.2 \times 10^3 \text{ cycles.}$$

Thus, the standard deviation of the time to failure for the Markov chain calculation is a factor of 4.5 larger than the Monte Carlo results predict. Since we are free to choose how the damage and time are discretized in the Markov model, the transition probabilities were regenerated for a 40 state case. This result is also shown in Figure 8. Observe for this case that the variance of the time to failure is less than for the 20 state run. Clearly then the number of damage states exerts a strong influence on the variance. While it is true that the number of states could be increased to more closely approximate the standard deviation of the Monte Carlo results used to estimate the transition probabilities, the added computational cost does not justify this procedure.

The comparison of the Monte Carlo and Markov chain models provided one immediate major result. Regardless of how much data is available to calculate transition probabilities, parameters in the Markov model must still be adjusted to reproduce the data. Thus, the Markov model is not unique nor are exact material property data and precise stress distributions sufficient to allow the Markov model to accurately predict the crack growth characteristics.

If confidence can be placed in the Markov results, the advantage of the method is its extremely low cost. In contrast, while the Monte Carlo analysis seems to reasonably replicate crack growth trends observed in structures, it does so at a high computational cost. (The Markov analyses performed required two orders of magnitude less computational time than the Monte Carlo analysis). Since probabilistic analysis of crack growth is usually performed for risk assessment or in sensitivity studies of crack growth models, it will usually be the case that many runs of the code are required. During this study, one run of the Monte Carlo analysis cost approximately $500 and the number of runs performed was minimized to make the results statistically significant but at a loss of statistical accuracy. Obviously then, while it is an appropriate model for PFM analysis, its cost may preclude its general use.

It appears from the results generated that probabilistic analysis (of a bridge or other structure) may be best accomplished by a combination of PFM formulations. The present results suggest that it is best to perform a limited number of Monte Carlo analyses for several of the potential stress distributions which the structure will see. Then, these results can be used to accurately determine the Markov transition probabilities for the range of stress distributions of interest. Thereafter, the Markov model can be used with confidence in repeated applications in order to minimize the cost of analysis for probabilistic studies.

## Case Study III: Comparison of Monte Carlo and Discrete Probability Distribution Methods

Recently, Discrete Probability Distributions (DPD's) have been suggested for use in risk analysis calculations to simplify the numerical computations which must be performed to determine failure probabilities. Specifically, DPD's have been developed to investigate probabilistic functions, that is, functions whose exact form is uncertain. The analysis of defect growth in materials by Probabilistic Fracture Mechanics (PFM) models provides an example in which the probabilistic function plays an important role. This study compares and contrasts Monte Carlo simulation and DPD's as tools for calculating material failure due to fatigue crack growth.

The remainder of this study examines the following topics. First, the necessary mathematics for each of the two methods is developed. Next, the application of these methods to cracks growing in steels is made. Finally, a comparison of the two methods is provided and conclusions about the use of each method are reached.

### CRACK GROWTH RATE LAW

Over the last few years, several investigators have examined the growth of cracks in materials, resulting in a relationship between the rate of crack growth with respect to time, da/dN, and a measure of the stress near the crack tip, denoted the stress intensity factor. The relationship used during these studies is:

$$da/dN = C\Delta K^m \quad , \tag{1}$$

where

$\Delta K$ = stress intensity factor, and
C,m = empirically determined constants.

The simplest formula for calculating $\Delta K$ is given by

$$\Delta K = \sigma \sqrt{\pi a} \quad .$$

As a vehicle for discussion, a problem will be examined in which the material is a carbon steel used in nuclear piping systems. Following the data derived by Harris[2], we obtain from Equation (1)

$$da/dN = C\,\sigma^4\,\pi^2\,a^2 \quad , \tag{2}$$

where

$m = 4$
$C$ = lognormally distributed with mean $9.14 \times 10^{-12}$ and standard deviation $2.20 \times 10^{-11}$.

Integrating Equation (2) yields

$$a_{n+1} = a_n + C\,\sigma^4\,\pi^2\,\Delta N \quad , \tag{3}$$

where it is assumed that $\Delta N$ is small enough so that $\sigma$ can be assumed constant, and n represents discrete time in units of $\Delta N$.

Equation (3) contains three random variables: the present defect distribution, $a_n$; the stress, $\sigma$; and the empirical constant, C. The initial defect distribution $a_o$ is assumed to be described by a Rayleigh distribution:

$$R(a) \sim a' \exp(-(a')^2/2) \quad ,$$

where $a' = (a - a_{min})/(a_{modal} - a_{min})$.

The stress distribution is also given by a Rayleigh distribution, denoted by $P(\gamma)$. Table 5 gives the parameter values for each of these distributions.

TABLE 5. INITIAL DEFECT AND STRESS DISTRIBUTION PARAMETER VALUES

| Parameter | Initial Defect Distribution | Stress Distribution |
|---|---|---|
| Minimum Value, $a_{min}$ | 0.005 inch | 10 Ksi |
| Modal Value, $a_{modal}$ | 0.01 inch | 35 Ksi |

We are interested in the DPD of the initial crack size after $N_F$ cycles, then

$$A^{(N_F)} = (\hat{Z}_i, \hat{p}_i) \text{ for } 1 \leq i \leq 20 \quad .$$

However, since C was fixed throughout this calculation, what has actually been obtained is

$$A^{(N_F)} = (\hat{Z}_i, \hat{p}_i\ (\hat{Z}_i|C1)) \equiv A^{(N_F)}\ (C_1) \quad .$$

That is, the probability being calculated is conditional on $C_1$. If the calculation is repeated for all of the discrete values of C, then

$$A^{(N_F)}(C_j) = Z_i, P_i(Z_i, P_i(Z_i|C_j)) \quad .$$

By combining $A^{(N_F)}(C_j)$ with the DPD representing the distribution for C and condensing the resulting DPD, the final crack size distribution is obtained.

The results of the crack growth calculations are shown in Table 6. For the Monte Carlo results, 2000 crack sample paths were generated. The DPD calculation requires a selection for the number of discrete points to be made. Four cases are shown in Table 6; for example, "DPD 10" means that each distribution was represented by 10 discrete points.

TABLE 6. RESULTS OF CRACK GROWTH CALCULATIONS USING DPD AND MONTE CARLO METHODS AFTER 100,000 STRESS CYCLES

| Method | Mean Crack Size (in.) | Standard Deviation (in.) | Failure Probability (%) | Computer Time (sec) |
|---|---|---|---|---|
| DPD 10 | .08361 | .1460 | 11.3 | 5 |
| DPD15 | .07778 | .1368 | 9.0 | 16 |
| DPD 20 | .07753 | .1361 | 8.5 | 40 |
| DPD 25 | .07698 | .1363 | 8.4 | 87 |
| Monte Carlo | .07546 | .1350 | 7.9 | 45 |
| 95% confidence intervals for Monte Carlo results | .06954-.08138 | .1307-.1391 | 6.7-9.1 | |

Table 6 shows that, for this problem, the DPD method is faster than Monte Carlo for comparable accuracy. The DPD-15 estimates of mean crack size, standard deviation, and failure probability each fall within the 95 percent confidence bounds for the corresponding parameter, based on the Monte Carlo results. The DPD-15 calculation takes 16 seconds as compared to 45 for Monte Carlo. In practice, however, it may be difficult to realize the maximum possible improvement in computation time using the DPD method. This is because the computational costs of the DPD method are very sensitive to the number of discrete points used to represent the distributions. For example, increasing the number of points from 15 to 20 more than doubles the computation time. In general, the computation time will increase as:

$$C_2 = \left(\frac{R_2}{N_1}\right)^M C_1$$

where

$C_1$ = computation time for first DPD calculation with $N_1$ discrete points
$R_1$ = number of discrete points for first DPD calculation
$N_2$ = number of discrete points for the DPD calculation being estimated
$M$ = number of random variables included in the DPD calculation
$C_2$ = estimate of DPD calculation time for $N_2$ discrete points.

Therefore, the inevitable trial-and-error needed to produce stable DPD results may be computationally expensive, unless runs are carefully planned and judicious extrapolation methods are employed.

The computational advantages of the DPD method may be greater in practical risk assessment problems where very low failure probabilities are encountered. For example, for a failure probability of $10^{-4}$, 1.7 million Monte Carlo runs would be required to obtain the accuracy exhibited by the Monte Carlo estimate of the failure probability in Table 6 (95 percent confidence interval of $\pm$ 15 percent). Improvements in the crude Monte Carlo method, such as importance sampling[2] tend to be ad hoc, and rather problematical in practice.

To obtain estimates of the probability of failure with the DPD method, no alterations to the algorithm need to be made, yet low probabilities can be calculated by placing more of the discrete points in the tails of the distributions for the random input variables. By maintaining the same total number of discrete points, the cost will remain the same, yet the low failure probability can be calculated. Because of the relatively low cost of the DPD calculation for such cases, several runs can be made to determine the failure probability's sensitivity to the discretization chosen for the random variable distributions. To date, however, the performance of the DPD method in such low-probability evaluations has not been precisely characterized.

Finally, it should be noted that the DPD method is not merely a convenient device for approximating inherently continuous calculations. It can be argued that many actual risk assessment calculations should properly be treated as discrete. For example, the fitting of continuous distributions to sparse data inevitably involves further assumptions not directly supported by that data. Thus, the DPD calculation, making direct use of discrete data, may actually be more natural than others based on continuous distributions.

## ACKNOWLEDGEMENT

This work was supported by Battelle Memorial Institute Corporate Technical Development under Contract No. 587-K-7346/B-1333-1430.

## REFERENCES


(1) Paris, P. C., and Erdogan, F., "A Critical Analysis of Crack Propagation Laws", J. Basic Engng. Ser. D, 85, 4, 1963.

(2) Halton, J. H., "A Retrospective and Prospective Survey of the Monte Carlo Method", SIAM Review, 12, 1, 1961.

(3) Kaplan, S., "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations--Application to Seismic Risk Assessment", Risk Analysis, 1, No. 3, 1981.

(4) Bogdanoff, J. L., and Krieger, W., "A New Cumulative Damage Mode, Part 2", J. Appl. Mech., 45, No. 2, 1978.

(5) Broek, D., and Leis, B. N., "Similitude and Anomalies in Crack Growth Rates", Materials, Experimentation, and Design in Fatigue, IPC Press, 1981, pp 129-148.

(6) Mayfield, M. E., et al, "Cold Leg Integrity Evaluation", NUREG/CR-1319, February 1980.

(7) Kurth, R. E., Leis, B. N., Cox, D. C., "Probabilistic Methods for Structures Under Random Loads", ASCE Fall Convention, New Orleans, October 1982.

(8) Moses, F., and Pavia, A., "Probability Theory for Highway Bridge Fatigue Stresses Part II", Ohio DOT-02-76, 1976.

(9) Yamada, K., and Albrecht, P., "Fatigue Design of Welded Bridge Details for Surface Stresses", Transportation Research Record 607, 1981.

(10) Goble, G. G., Moses, F., and Pavia, A. A., "Final Measurements and Laboratory Testing of Bridge Components", Final Report to Ohio Department of Transportation, 1974.

(11) Harris, D. O., Lim, E. Y., and Dedhia, D. D., "Probability of Pipe Fracture in the Primary Coolant Loop of a PWR Plant, Volume 5: Probabilistic Fracture Mechanics Analysis", NUREG/CR-2189, Vol. 5, 1981.

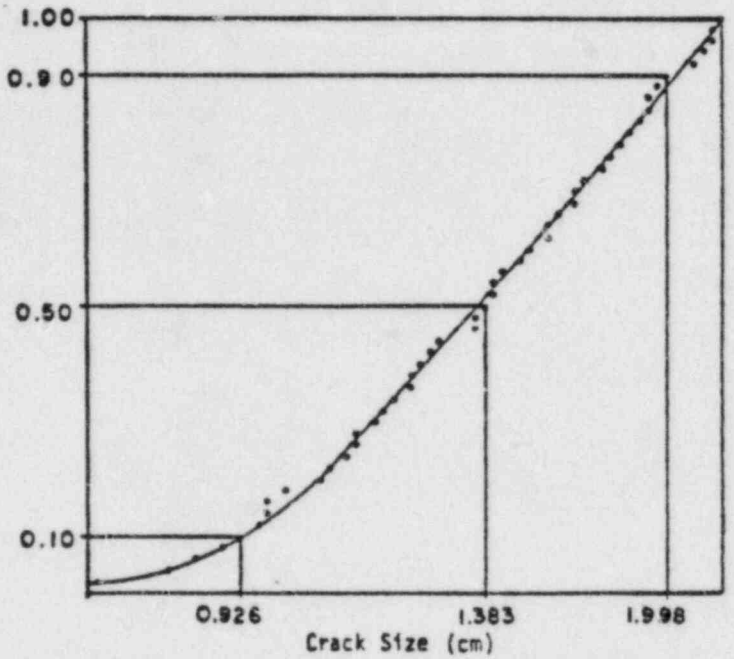



FIGURE 1. EMPIRICALLY DETERMINED CUMULATIVE DISTRIBUTION FUNCTION (CDF) FOR THE INITIAL DEFECT SIZE

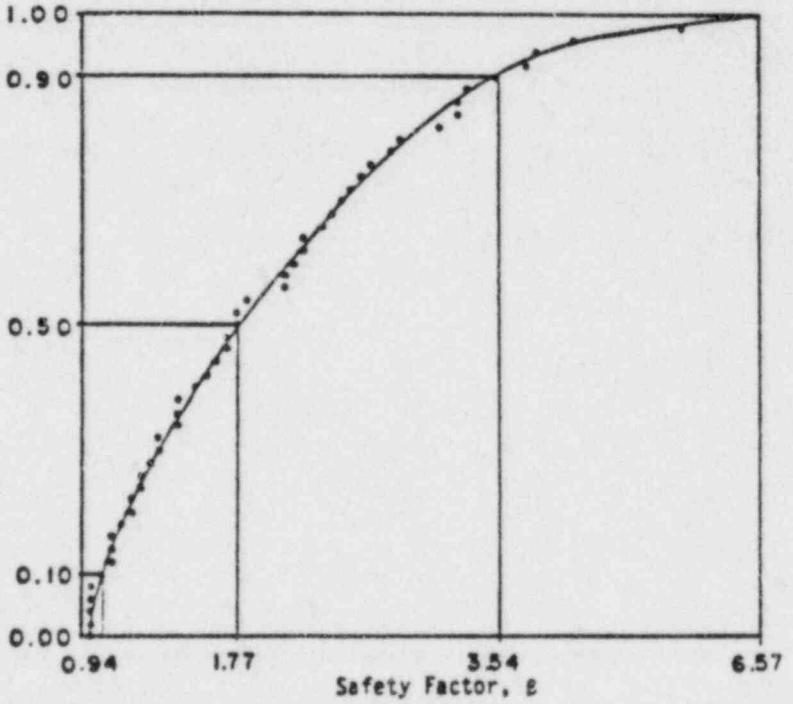



FIGURE 2. EMPIRICALLY DETERMINED CUMULATIVE DISTRIBUTION FUNCTION (CDF) FOR SAFETY FACTOR BASED ON 40 YEAR PLANT LIFE

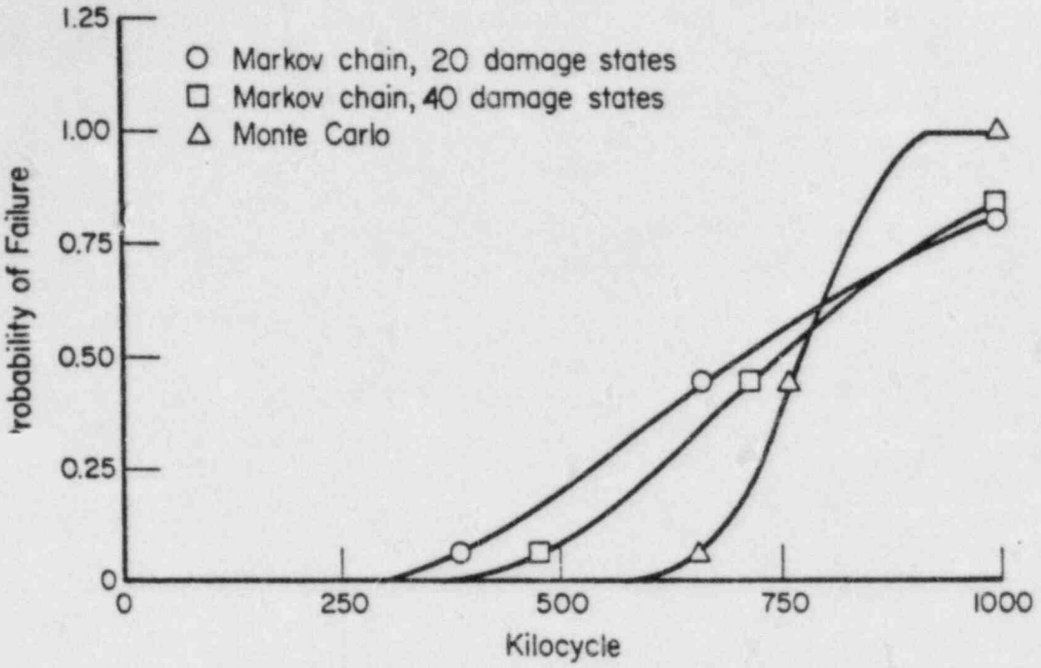



FIGURE 3. COMPARISON OF THE PROBABILITY OF FAILURE CALCULATED BY MONTE CARLO AND MARKOV METHODS

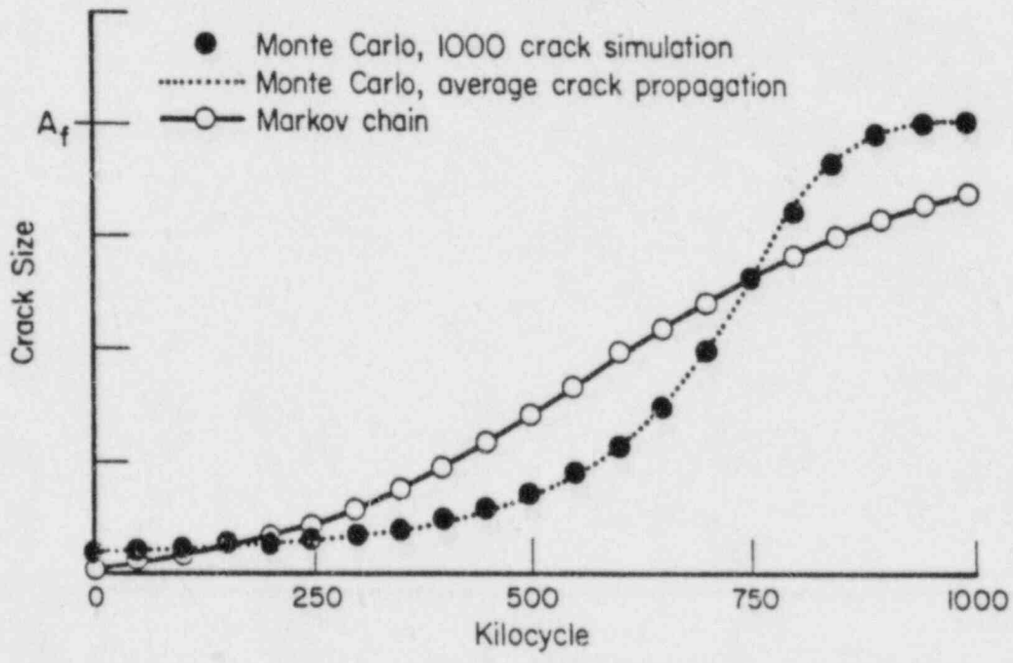



FIGURE 4. COMPARISON OF THE AVERAGE CRACK SIZE CALCULATED BY MONTE CARLO AND MARKOV METHODS

# TREND AND PATTERN ANALYSIS OF OPERATING REACTOR DATA FROM LICENSEE EVENT REPORTS

by

Robert L. Dennig
Nuclear Regulatory Commission
AEOD-263A-EWS
Washington, DC 20555

O. V. Hester and C. D. Gentillon
EG&G Idaho-Inc.
P. O. Box 1625
Idaho Falls, ID 83415

Contributed Paper Session 2

# TREND AND PATTERN ANALYSIS OF OPERATING REACTOR DATA FROM LICENSEE EVENT REPORTS

Robert L. Dennig
U.S. Nuclear Regulatory Commission

O. V. Hester and C. D. Gentillon
EG&G Idaho, Inc.

## ABSTRACT

"Trend and pattern analysis" are oft-cited goals in the review of operational data reported by nuclear utilities to the Nuclear Regulatory Commission (NRC). The NRC Office for Analysis and Evaluation of Operational Data in conjunction with EG&G Idaho, Inc., has developed software which uses contingency table techniques to perform such analysis. The objective of the analysis is to identify outliers and anomolous behavior within the data which would be good candidates for detailed engineering follow-up.

This paper briefly discusses the Sequence Coding and Search System (SCSS) Licensee Event Report (LER) data base which contains data in a form amenable to cross-classification, the data retrieval and statistical software employed in conjunction with the SCSS, and the results of a trial application to the 1981 LER data.

## BACKGROUND

The pre-eminent source of operating incident data for U.S. nuclear power plants is the Licensee Event Report (LER). Reactor licensees submit an LER to the Nuclear Regulatory Commission (NRC) when an incident at a plant meets one or more of the reporting criteria incorporated in their operating license technical specifications. Table 1 shows the volume of LERs received over the last 5 complete years, and a received-to-date figure for 1983.

TABLE 1. NUMBER OF LERS BY CALENDAR YEAR

| Year | 1978 | 1979 | 1980 | 1981 | 1982 | 1983* |
|---|---|---|---|---|---|---|
| Number of LERs | 3168 | 3164 | 3850 | 4016 | 4399 | 2571 |

*Partial year; reports through July 1983.

Each individual LER receives wide distribution both within and outside of the NRC, and is scrutinized in some detail by engineering personnel. This case-by-case review is sometimes supplemented by reference to previous reports, but in general the recognition of incident recurrence, increasing rate of occurrence, or a pattern of occurrence is delimited by the perception and memory of the individual engaged in the review. While confident that event-by-event review identifies events of immediate safety significance, we have been concerned that safety-significant situations of high or wide-spread incidence of lower level go unrecognized in this approach. To address this concern we have developed computer software which allows rapid and flexible statistical analysis of the LER incident data.

## SOFTWARE DESCRIPTION

The Sequence Coding and Search System (SCSS) LER data base was developed by the NRC Office for Analysis and Evaluation of Operational Data (AEOD) with assistance from Oak Ridge National Laboratory (ORNL) in order to support both ad hoc data retrieval (i.e., identification of all LERS which describe a given problem) and broader form statistical analysis. The production version is implemented in System 1022 on a PDP-10 at ORNL;[1] the version which works in concert with the trends and patterns software is implemented in the Control Data Corporation (CDC) "DMS-170" data management system at the Idaho National Engineering Laboratory (INEL).

The most fundamental data structure in SCSS is the step record. Each step record contains information about a single reported incidence of hardware fault or human error, and step records are strung together to model a sequence of events described in an LER. SCSS step records are conceptually equivalent to the "one-line" component fault records cataloged in the previously published LER data summaries.[2] A single LER generates multiple step records in the SCSS data base.

By manipulation of the step records and related data stored in other segments of the SCSS we can cross classify individual incidents to build multiway contingency tables. The dimensions of these tables are selected principally from those listed in Table 2.

The software which accesses the SCSS database and constructs the desired tables is the CONTING program written by E. Henry and L. R. Fitch of INEL.[3] CONTING is a user-prompting interactive program which translates the analyst's table specifications into the required searches and sorts of the SCSS and if desired will automatically prepare an input file for execution by the P4F program of the "BMDP" statistical software package.[4]

TABLE 2. MAJOR INCIDENT ATTRIBUTES FOR CROSS-CLASSIFICATION

| Abbreviation | Attribute | Explanation |
|---|---|---|
| FID | Facility Identification | A four digit alpha numeric code unique to each licensed power reactor |
| CAUSE | Cause | The proximate cause of the incident |
| COMP | Component | An item of hardware, a person, or a designation for a train or all trains of a system |
| PSYSTEM | Principal System | The system in which a component is installed or, in the case of personnel, the activity engaged in when the incident occurred |
| ISYSTEM | Interfacing System | Additional system information for components at system boundaries |
| VENDOR | Vendor | The component manufacturer |
| EFFECT | Effect | The observed component state or behavior |
| EVDATE | Time | The time interval containing the incident's event date |

CONTING has the capability to perform selection and tabulation over the entire SCSS database, or over any pre-selected subset. Pre-selection may be performed by CONTING, or outside of CONTING using the CDC program Query/Update (QU).[5] This capability allows the user greater flexibility in the specification of the relevant population of records than is afforded by the selection of levels for a given table dimension or factor within CONTING. It is most useful for exclusion of a single level in a dimension we don't wish to see in the final table itself.

The CONTING user builds a table by selecting one dimension at a time from attributes such as those listed in Table 2. Once a dimension is selected the user has the capability to:

- List all the level values permitted in the SCSS data base, along with the counts of records containing each value.

- Select all permitted values, delete selectively from this list, or build up the values list one at a time.

- Use a file of previously selected values.

The construction of time cells by CONTING for event dates is particularly flexible and convenient. The user can select from input options such as the following:

- Start date, end date, number of cells

- Start date, months per cell, number of cells

- End date, months per cell, number of cells

- Last N months (prior to ending date for SCSS data)

- User-specified time cells.

When specification of the levels or cells for the selected dimension is completed, CONTING will summarize the status of the table: each dimension and its corresponding number of levels, and the total table size (total cells). The user then may either select an additional dimension or proceed to execute the data categorization already specified.

CONTING can determine counts for up to 50,000 cells as defined by the table dimension/level input. Tables with up to 200,000 cells can be built through batch execution; the input needed for batch execution is automatically prepared. Outputs from CONTING are specified by the user and can include:

- A history of the CONTING interactive session

- A computer file of the table input which can be saved and re-used at a future time

- Cell counts for the table in a list format with cell identifying labels

- A computer file of the control language and data needed for multiway table analysis by BMDP-P4F. This data can include exposure time in each cell in reactor calendar hours or reactor critical hours, as well as cell counts.

The ability to calculate cell exposure time is another important and convenient feature of CONTING. This feature uses a special auxiliary file of the INEL SCSS data base which contains Gray Book data[6] for calculation of reactor critical hours.

BMDP-P4F provides numerous options for table display and log-linear modeling. The standard P4F input file prepared by CONTING may contain requests for table-formatted cell counts and exposure times for all cells as well as cell counts for all one- and two-way marginals. Percent-of-total count figures are also displayed for each cell. No requests for modeling are included in the standard file. The user adds the necessary P4F modeling commands on a case-by-case basis by editing the input file before submitting it for execution.

Following the notation used in the BMDP documentation,[4] the log-linear model in three dimensions is written as:

$$\ln F_{123} = \theta + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{123}, \qquad (1)$$

where

$F_{123}$ = the expected cell counts in each cell of a three-dimensional array indexed by i = 1, I; j = 1, J; k = 1, K; for dimensions (attributes) 1, 2 and 3, respectively (I is the number of categories or levels of the first attribute; J, the second; and K, the third),

$\theta$ = the grand mean effect,

$\lambda_1$, $\lambda_2$, $\lambda_3$ = the main effects for the three dimensions,

$\lambda_{12}$, $\lambda_{13}$, $\lambda_{23}$ = second order effects, and

$\lambda_{123}$ = the third order effect.

In this notation, one may use further subscripts to denote effects for individual levels of the attributes; e.g., $\lambda_{12(ij)}$ is the second-order effect describing the interaction of the $i^{th}$ level of the first table dimension attribute with the $j^{th}$ level of the second attribute. When the additional subscripts are suppressed, the notation refers to the set of all such effects, over all the levels of the variables or attributes indexed by the subscripts. Standard log-linear modeling includes constraints which make the effects unique. They are a measure of the magnitude each term contributes to the expected cell frequency.

The model of Equation (1) is "saturated"; estimates for the $\lambda$'s can be found so that the expected cell counts match the observed cell counts exactly. One may test the adequacy of log-linear models with selected sets of the effects assumed to be zero and thus excluded. For example, one may test the hypothesis that the $\lambda_{123}$ are all zero. P4F finds maximum likelihood estimates of expected cell counts for hierarchical models; i.e., models for which the inclusion of an effect of a specified order implies the inclusion of all lower-order effect sets involving the same dimensions. The expected cell counts under assumed Poisson sampling are estimated using iterative proportional fitting (IPF).[7]

After estimating expected cell counts, P4F uses one of two algorithms (described in Appendix A.7 of Reference 4 and in the July 1983 BMDP Newsletter) for calculating parameter estimates (the $\lambda$'s in Equation 1) and their standard errors. Pertinent to later discussions is the method used when the model is indirect (i.e., the iteration is required to calculate estimates of expected cell counts) and the cell estimates include one or more zeros. In this case, P4F obtains parameter estimates by finding the weighted least squares solution given by

$$\hat{\underset{\sim}{\lambda}} = (X' \, W \, X)^{-1} \, X' \, W \, \underset{\sim}{Y} \qquad (2)$$

where

$\underset{\sim}{Y}$ = a column vector of the logs of cell estimates,

X = an (nxp) design matrix for the model (where n = number of non-zero cell estimates and p = number of parameters to be estimated), and

W = an (nxn) diagonal matrix with the reciprocal of the cell estimates as the diagonal entries; i.e., W is the asymptotic variance-covariance matrix of Y.

As mentioned earlier, CONTING has the ability to calculate the exposure time in plant calendar hours or plant critical hours for each cell in a table, and to pass the values to BMPD-P4F. The IPF algorithm involves an initial fit matrix which generally consists of all ones. P4F allows the user to override this default; one may for example declare certain cells to be structural zeros (empty cells; combinations of attributes which do not ever occur) by inputting zeros in the initial matrix for those cells. Specifying exposure times for each cell in a table as an initial fit matrix allows the user to take advantage of an even stronger feature, namely, the log-linear hazard modeling methodology demonstrated by Laird and Olivier.[8] Laird and Olivier showed that if the exposure times are used

as initial values and piecewise exponential sampling is assumed, then IPF yields cell estimates which, when divided by their respective exposure times, in turn yield a set of maximum likelihood estimates for occurrence or hazard rates for a given log-linear model. Further, likelihood ratio tests based on the fitted counts are valid in making inferences about the hazard rates. With an initial fit matrix of exposure times, these rates are produced by P4F as expected values from the specified model.

## TRIAL APPLICATION

The log-linear modeling approach to investigating the pattern or trend which may exist in a data set was applied to a contingency table for personnel steps in the SCSS data base.

Personnel errors were used in this initial application of log-linear modeling techniques for a host of reasons. Among these reasons are the fact that personnel steps represent a unique subset of the SCSS database, the way personnel steps easily lend themselves to a configuration consisting of well defined categories with a managable size table, and the ability to construct a table without an overriding number of zeros. The last reason stated is a significant factor which will be discussed later.

The original contingency table built through CONTING consists of three SCSS fields, namely, FID, PSYSTEM, and EVDATE. For FID six plants were used. These six plants represented the three pressurized water reactor (PWR) and boiling water reactor (BWR) plants with the highest amount of LER reporting for 1981. The plants are listed below.

| BWRs | | | PWRs | | |
|---|---|---|---|---|---|
| BEP2 | - | Brunswick Unit 2 | MGS1 | - | McGuire Unit 1 |
| EIH1 | - | Hatch Unit 1 | SGS2 | - | Salem Unit 2 |
| EIH2 | - | Hatch Unit 2 | SNP1 | - | Sequoyah Unit 1 |

The SCSS PSYSTEM field for personnel steps reflects the activity engaged in when an incident occurs. Five activities were chosen for the modeling. The selection was based on their being deemed "interesting" from an engineering perspective and there being an appreciable number of steps for each of these activities in the data base. The following activities were used:

| | |
|---|---|
| PD | Design Activity |
| PM | Maintenance/Repair Activity |
| PO | Operation Activity |
| PT | Test/Calibration Activity |
| PZ | Unknown Activity |

Finally, only 1981 data were used in developing a model. The year was partitioned into four quarters. This partitioning was motivated by the interest in assessing whether or not the events being studied could be characterized by any underlying temporal trend or pattern. The four quarters of 1981 are labelled T1, T2, T3, and T4, respectively.

One other SCSS field was involved in this analysis. The SCSS EFFECT code when used in conjunction with personnel describes whether an omission or commission was involved. Some personnel steps have the effect code UF ("Desired Commission"). These steps provide further information about reported events but are excluded from this study because they do not describe personnel errors (faults).

With the basic structure of the table defined, the CONTING software package was then used. An initial run sorting on the eleven personnel effect codes other than UF was used to form a set of step records describing personnel faults. A subsequent CONTING run using this set generated counts for a 6 x 5 x 4 table incorporating the plant (FID), activity (PSYSTEM) and EVDATE variables defined above. This run invoked the option for building a BMDP input deck for program P4F. The BMDP deck for this run, with the count data excluded, appears in Figure 1. The deck was executed and an evaluation was made of the resulting contingency table. Figure 2 shows the output, with percent of total and margin tables attached. The table has 120 cells and 295 observations, but 29 cells are zero. If the effect dimension were included for study, the same 295 observations would be spread among 1320 cells and the table would be extremely sparse.

Zero entries always pose some problem for log-linear modeling, since expected cell counts that follow a log model cannot be zero. The zeros in Figure 2 are sampling zeros rather than structural zeros; given enough time for observation, eventually every cell would be nonzero. In the meantime, the data may not provide enough information to estimate all the parameters of specified log-linear models.

The implications in having a multidimensional table dominated by zeros vary with the particular log-linear analysis code being used. Problems occur with the BMDP code when all the cells associated with an effect being estimated for a specified model are zero counts; when this is the case, the effect is unestimable and the algorithm loses track of which effects were estimable and which estimates are associated with which effects. However, this is not a problem for some log-linear codes. In these codes for each effect whose cell counts are all zero the effects itself is assumed to be zero and no attempt is made to estimate it. In essence the cells are treated as structural zeros and all the remaining effects with nonzero cell counts are estimated. The BMDP program is to be modified in the future so that it handles this problem in the manner discussed above.

A number of additional alternatives for handling tables with sparse entries exist. Included are deletion of the rows and/or columns in the table whose sparsity inhibits the modeling, adding a small number (delta) to each cell in the table, and collapsing of categories into a lesser number of groups. The use of some of these alternatives is discussed below; References 7, 9 and 10 provide further insights.

```
OVHBM,T37,P1,STANY.
ACCNT,ID=OVH,ORG=3540,BIN=TM3.
ATTACH,BMDP4F,ID=BMDP.
BMDP4F,W=50000.
*EOR
/ PROBLEM    TITLE IS
 'PERSONNEL MODELING 1981
/ INPUT      VARIABLES ARE 3.
             TABLE IS   5,  6,  4.
             FORMAT IS FREE.
/ VARIABLE   NAMES ARE PSYSTEM,FID    ,EVDATE .
/ TABLE      INDICES ARE PSYSTEM, FID    , EVDATE .
             SYMBOLS ARE A, B, C.
/ COMMENT '
          THE FOLLOWING TABLE SHOWS THE TIME CELLS
          AND CORRESPONDING EVENT DATE BOUNDARIES

        ------------------------------------------------------------
          CELL         BOUNDARY         CELL         BOUNDARY
        ------------------------------------------------------------
          T1      19810101-19810401    T3      19810701-19811001
          T2      19810401-19810701    T4      19811001-19820101
        ------------------------------------------------------------
            '.
/ CATEGORY
 CODES(3) ARE 1 TO   4
 NAMES(3) ARE
 'T1  ', 'T2  ', 'T3  ', 'T4  '.
 CODES(2) ARE 1 TO   6
 NAMES(2) ARE
 'BEP2', 'EIH1', 'EIH2', 'MGS1', 'SGS2', 'SNP1'.
 CODES(1) ARE 1 TO   5
 NAMES(1) ARE
 'PD  ', 'PM  ', 'PO  ', 'PT  ', 'PZ  '.
/ PRINT      OBS.PERC=TOT.MARGINALS=2.
/ COMMENT '
          SORT WAS PERFORMED ON SET  1 OF OLDSETS

          RECORD HITS WERE USED FOR FREQUENCY ACCUMULATION

          RECORDS PERTAINING BOTH FAILURES AND COMMAND
          FAULTS WERE SORTED

            '.
/ END
```

Figure 1. Initial BMDP-P4F deck.

***** OBSERVED FREQUENCY TABLE 1

| EVDATE | FID | PSYSTEM PD | PM | PO | PT | PZ | TOTAL |
|---|---|---|---|---|---|---|---|
| T1 | BEP2 | 5 | 0 | 0 | 3 | 0 | 8 |
| | EIH1 | 2 | 0 | 0 | 5 | 1 | 9 |
| | EIH2 | 2 | 0 | 1 | 5 | 1 | 9 |
| | MGS1 | 6 | 4 | 3 | 1 | 1 | 15 |
| | SGS2 | 0 | 0 | 1 | 0 | 0 | 1 |
| | SNP1 | 5 | 2 | 4 | 6 | 2 | 19 |
| | TOTAL | 20 | 6 | 9 | 15 | 5 | 55 |
| T2 | BEP2 | 1 | 1 | 2 | 2 | 0 | 6 |
| | EIH1 | 7 | 1 | 0 | 2 | 2 | 12 |
| | EIH2 | 1 | 4 | 5 | 2 | 1 | 13 |
| | MGS1 | 6 | 15 | 15 | 5 | 2 | 43 |
| | SGS2 | 3 | 3 | 6 | 4 | 0 | 16 |
| | SNP1 | 6 | 2 | 1 | 1 | 1 | 11 |
| | TOTAL | 24 | 26 | 29 | 16 | 6 | 101 |
| T3 | BEP2 | 3 | 2 | 0 | 3 | 2 | 10 |
| | EIH1 | 1 | 5 | 2 | 1 | 1 | 10 |
| | EIH2 | 1 | 1 | 0 | 0 | 1 | 2 |
| | MGS1 | 11 | 14 | 13 | 2 | 1 | 41 |
| | SGS2 | 1 | 0 | 3 | 1 | 1 | 6 |
| | SNP1 | 4 | 2 | 0 | 6 | 3 | 15 |
| | TOTAL | 21 | 24 | 18 | 13 | 8 | 84 |
| T4 | BEP2 | 6 | 6 | 0 | 3 | 0 | 15 |
| | EIH1 | 2 | 0 | 3 | 1 | 0 | 6 |
| | EIH2 | 0 | 1 | 1 | 1 | 0 | 3 |
| | MGS1 | 4 | 2 | 3 | 6 | 2 | 17 |
| | SGS2 | 0 | 2 | 3 | 0 | 1 | 6 |
| | SNP1 | 2 | 2 | 3 | 0 | 1 | 8 |
| | TOTAL | 14 | 13 | 13 | 11 | 4 | 55 |

TOTAL OF THE OBSERVED FREQUENCY TABLE IS 295

Figure 2. Output from initial BMDP-P4F run.

***** PERCENTS OF THE TABLE TOTAL -- TABLE 1

| EVDATE | FID | PSYSTEM PD | PM | PO | PT | PZ | TOTAL |
|---|---|---|---|---|---|---|---|
| T1 | BEP2 | 1.7 | 0.0 | 0.0 | 1.0 | 0.0 | 2.7 |
| | EIH1 | .7 | 0.0 | 0.0 | 0.0 | .3 | 1.0 |
| | EIH2 | .7 | 0.0 | .3 | 1.7 | .3 | 3.1 |
| | MGS1 | 2.0 | 1.4 | 1.0 | .3 | .3 | 5.1 |
| | SGS2 | 0.0 | 0.0 | .3 | 0.0 | 0.0 | .3 |
| | SNP1 | 1.7 | .7 | 1.4 | 2.0 | .7 | 6.4 |
| | TOTAL | 6.8 | 2.0 | 3.1 | 5.1 | 1.7 | 18.6 |
| T2 | BEP2 | .3 | .3 | .7 | .7 | 0.0 | 2.0 |
| | EIH1 | 2.4 | .3 | 0.0 | .7 | .7 | 4.1 |
| | EIH2 | .3 | 1.4 | 1.7 | .7 | .3 | 4.4 |
| | MGS1 | 2.0 | 5.1 | 5.1 | 1.7 | .7 | 14.6 |
| | SGS2 | 1.0 | 1.0 | 2.0 | 1.4 | 0.0 | 5.4 |
| | SNP1 | 2.0 | .7 | .3 | .3 | .3 | 3.7 |
| | TOTAL | 8.1 | 8.8 | 9.8 | 5.4 | 2.0 | 34.2 |
| T3 | BEP2 | 1.0 | .7 | 0.0 | 1.0 | .7 | 3.4 |
| | EIH1 | .3 | 1.7 | .7 | .3 | .3 | 3.4 |
| | EIH2 | .3 | .3 | 0.0 | 0.0 | 0.0 | .7 |
| | MGS1 | 3.7 | 4.7 | 4.4 | .7 | .3 | 13.9 |
| | SGS2 | .3 | 0.0 | 1.0 | .3 | .3 | 2.0 |
| | SNP1 | 1.4 | .7 | 0.0 | 2.0 | 1.0 | 5.1 |
| | TOTAL | 7.1 | 8.1 | 6.1 | 4.4 | 2.7 | 28.5 |
| T4 | BEP2 | 2.0 | 2.0 | 0.0 | 1.0 | 0.0 | 5.1 |
| | EIH1 | .7 | 0.0 | 1.0 | .3 | 0.0 | 2.0 |
| | EIH2 | 0.0 | .3 | .3 | .3 | 0.0 | 1.0 |
| | MGS1 | 1.4 | .7 | 1.0 | 2.0 | .7 | 5.8 |
| | SGS2 | 0.0 | .7 | 1.0 | 0.0 | .3 | 2.0 |
| | SNP1 | .7 | .7 | 1.0 | 0.0 | .3 | 2.7 |
| | TOTAL | 4.7 | 4.4 | 4.4 | 3.7 | 1.4 | 18.6 |

Figure 2. (continued).

***** MARGINAL SUBTABLE -- TABLE 1

PSYSTEM

| PD | PM | PO | PT | PZ | TOTAL |
|---|---|---|---|---|---|
| 79 | 69 | 69 | 55 | 23 | 295 |

***** MARGINAL SUBTABLE -- TABLE 1

FID

| BEP2 | EIH1 | EIH2 | MGS1 | SGS2 | SNP1 | TOTAL |
|---|---|---|---|---|---|---|
| 39 | 31 | 27 | 116 | 29 | 53 | 295 |

***** MARGINAL SUBTABLE -- TABLE 1

EVDATE

| T1 | T2 | T3 | T4 | TOTAL |
|---|---|---|---|---|
| 55 | 101 | 84 | 55 | 295 |

***** MARGINAL SUBTABLE -- TABLE 1

| FID \ PSYSTEM | PD | PM | PO | PT | PZ | TOTAL |
|---|---|---|---|---|---|---|
| BEP2 | 15 | 9 | 2 | 11 | 2 | 39 |
| EIH1 | 12 | 6 | 5 | 4 | 4 | 31 |
| EIH2 | 4 | 6 | 7 | 8 | 2 | 27 |
| MGS1 | 27 | 35 | 34 | 14 | 6 | 116 |
| SGS2 | 4 | 5 | 13 | 5 | 2 | 29 |
| SNP1 | 17 | 8 | 8 | 13 | 7 | 53 |
| TOTAL | 79 | 69 | 69 | 55 | 23 | 295 |

Figure 2. (continued).

***** MARGINAL SUBTABLE -- TABLE 1

| EVDATE | PSYSTEM | | | | | |
|---|---|---|---|---|---|---|
| | PD | PM | PO | PT | PZ | TOTAL |
| T1 | 20 | 6 | 9 | 15 | 5 | 55 |
| T2 | 24 | 26 | 29 | 16 | 6 | 101 |
| T3 | 21 | 24 | 18 | 13 | 8 | 84 |
| T4 | 14 | 13 | 13 | 11 | 4 | 55 |
| TOTAL | 79 | 69 | 69 | 55 | 23 | 295 |

***** MARGINAL SUBTABLE -- TABLE 1

| EVDATE | FID | | | | | |
|---|---|---|---|---|---|---|
| | BEP2 | EIH1 | EIH2 | MGS1 | SGS2 | SNP1 | TOTAL |
| T1 | 8 | 3 | 9 | 15 | 1 | 19 | 55 |
| T2 | 6 | 12 | 13 | 43 | 16 | 11 | 101 |
| T3 | 10 | 10 | 2 | 41 | 6 | 15 | 84 |
| T4 | 15 | 6 | 3 | 17 | 6 | 8 | 55 |
| TOTAL | 39 | 31 | 27 | 116 | 29 | 53 | 295 |

Figure 2. (continued).

Returning to our particular model, after a number of iterations the decision was made to consolidate plants into the two groups, BWRs and PWRs. This change was made through a simple modification of the original BMDP deck produced by CONTING. Figure 3 shows the modified BMDP deck with additional options specified for modeling. Commands in the deck which are new or modified are highlighted. The contingency table for this new configuration appears in Figure 4.

In Figure 3, we see that the three fields PSYSTEM, FID and EVDATE are referenced, respectively, by the symbols P, F, and T (rather than the dimension numbers 1, 2, and 3 of Equation 1). Also, notice that a number of models for the data have been specified. Finding a "good" model is an exploratory process. The models specified in the figure represent only a small subset of those tested on the data. The ensuing discussion pertains only to the model ultimately chosen to characterize the data. However, there is no one model which can objectively be classified as the "best" model. Factors such as past knowledge of relationships between variables, physical constraints and cost must be taken into consideration in development of any model.

The model selected to characterize the data is PF,T. That is, non-zero effects for the PSYSTEM-facility interaction and for event date will be included in the model. Since this is a hierarchial model, P and F are also terms whose main effects are included. The effect sets $\lambda_{PT}$, $\lambda_{FT}$, and $\lambda_{PFT}$ are assumed to be zero for this model.

Sections of the P4F output pertinent to this model are contained in Figure 5. Figure 5(b)[a] shows the expected cell frequencies generated.

Estimates of the effects for the model are given in Figure 5(c). Equation (1) together with the model implies that the (i,j,k) cell frequency can be estimated as

$$\exp(\theta) \cdot \exp(\lambda_{P(i)}) \cdot \exp(\lambda_{F(j)}) \cdot \exp(\lambda_{T(k)}) \cdot \exp(\lambda_{PF(ij)}), \qquad (3)$$

and the "multiplicative parameters" in the listing give these values. For example, the multiplicative "PWR" effect (see Figure 5(d)) of 1.43 shows that counts for the PWR plant cells in the table are on the average roughly 43% higher than the overall average. Detailed interpretation of these effects and the model itself in general requires insight and knowledge on the part of the individual doing the analysis, regarding the "real world" meaning of the factors being modeled in addition to an understanding of the mathematics involved in the modeling procedure itself.

---

a. Figure 5 is marked with circled letters. These correspond to the letters in parentheses in the text figure references.

```
OVHBM,T37,P1,STANY.
ACCNT,ID=OVH,ORG=3540,BIN=TM3.
ATTACH,BMDP4F,ID=BMDP.
BMDP4F,W=50000.
*EOR
/ PROBLEM    TITLE IS
 'PERSONNEL MODELING 1981
/ INPUT      VARIABLES ARE 3.
             TABLE IS   5,  6,  4.
             FORMAT IS FREE.
/ VARIABLE   NAMES ARE PSYSTEM,FID    ,EVDATE .
/ TABLE      INDICES ARE PSYSTEM, FID    , EVDATE .
  SYMBOLS ARE P,F,T.
/FIT MODEL IS PF,FT,PT.
 CELL=STAN.STEP=8.PROB=.25.
   ADD IS MULTIPLE.DELETE IS SIMPLE.STRATA IS PSYSTEM.
/FIT MODEL IS FP,T.
 CELL=STAN.STEP=8.PROB=.25.
  ADD IS MULTIPLE.DELETE IS SIMPLE.STRATA IS PSYSTEM.
/FIT MODEL IS F,PT.
 CELL=STAN.STEP=8.PROB=.25.

  ADD IS MULTIPLE .DELETE IS SIMPLE.STRATA IS PSYSTEM.
/FIT MODEL IS FT,P.
 CELL=STAN.STEP=8.PROB=.25.
  ADD IS MULTIPLE .DELETE IS SIMPLE .STRATA IS PSYSTEM.
/FIT MODEL IS  P,F,T.
CELL=STAN.STEP=8.PROB=.25.
  ADD IS MULTIPLE.DELETE IS SIMPLE.STRATA IS PSYSTEM.
/PRINT  OBS.EXP.LAMBDA.BETA.PERC=TOT.
```

Figure 3. Modified BMDP-P4F deck.

```
/ COMMENT '
          THE FOLLOWING TABLE SHOWS THE TIME CELLS
          AND CORRESPONDING EVENT DATE BOUNDARIES

        ------------------------------------------------------------
          CELL          BOUNDARY          CELL          BOUNDARY
        ------------------------------------------------------------
          T1       19810101-19810401      T3       19810701-19811001
          T2       19810401-19810701      T4       19811001-19820101
        ------------------------------------------------------------
          '.
/ CATEGORY
 CODES(3) ARE 1 TO   4
 NAMES(3) ARE
 'T1  ', 'T2  ', 'T3  ', 'T4  '.
 CODES(2) ARE 1 TO   6
 NAMES(2) ARE
+'BWR', 'BWR', 'BWR', 'PWR', 'PWR', 'PWR'.
 CODES(1) ARE 1 TO   5
 NAMES(1) ARE
 'PD  ', 'PM  ', 'PO  ', 'PT  ', 'PZ  '.
/ PRINT      OBS.PERC=TOT.MARGINALS=2.
/ COMMENT '
          SORT WAS PERFORMED ON SET  1 OF OLDSETS

          RECORD HITS WERE USED FOR FREQUENCY ACCUMULATION

          RECORDS PERTAINING BOTH FAILURES AND COMMAND

          FAULTS WERE SORTED

            '.
/ END
```

Figure 3. (continued).

***** OBSERVED FREQUENCY TABLE 1

| EVDATE | FID | PSYSTEM PD | PM | PO | PT | PZ | TOTAL |
|---|---|---|---|---|---|---|---|
| T1 | BWR | 9 | 0 | 1 | 8 | 2 | 20 |
| | PWR | 11 | 6 | 8 | 7 | 3 | 35 |
| | TOTAL | 20 | 6 | 9 | 15 | 5 | 55 |
| T2 | BWR | 9 | 6 | 7 | 6 | 3 | 31 |
| | PWR | 15 | 20 | 22 | 10 | 3 | 70 |
| | TOTAL | 24 | 26 | 29 | 16 | 6 | 101 |
| T3 | BWR | 5 | 8 | 2 | 4 | 3 | 22 |
| | PWR | 16 | 16 | 16 | 9 | 5 | 62 |
| | TOTAL | 21 | 24 | 18 | 13 | 8 | 84 |
| T4 | BWR | 8 | 7 | 4 | 5 | 0 | 24 |
| | PWR | 6 | 6 | 9 | 6 | 4 | 31 |
| | TOTAL | 14 | 13 | 13 | 11 | 4 | 55 |

TOTAL OF THE OBSERVED FREQUENCY TABLE IS 295

Figure 4. Observed frequency table from modified BMDP-P4F run.

| MODEL | D.F. | LIKELIHOOD-RATIO CHI-SQUARE | PROB | PEARSON CHI-SQUARE | PROB | ITER. |
|---|---|---|---|---|---|---|
| PF,T. | 27 | 35.31 | .1313 | 30.28 | .3015 | 2 |

(a)

***** EXPECTED VALUES USING ABOVE MODEL

(b)

| EVDATE | FID | PSYSTEM PD | PM | PD | PT | PZ | TOTAL |
|---|---|---|---|---|---|---|---|
| T1 | BWR | 5.8 | 3.9 | 2.6 | 4.3 | 1.5 | 18.1 |
| | PWR | 8.9 | 8.9 | 10.3 | 6.0 | 2.8 | 36.9 |
| | TOTAL | 14.7 | 12.9 | 12.9 | 10.3 | 4.3 | 55.0 |
| T2 | BWR | 10.6 | 7.2 | 4.8 | 7.9 | 2.7 | 33.2 |
| | PWR | 16.4 | 16.4 | 18.8 | 11.0 | 5.1 | 67.8 |
| | TOTAL | 27.0 | 23.6 | 23.6 | 18.8 | 7.9 | 101.0 |
| T3 | BWR | 8.8 | 6.9 | 4.0 | 6.5 | 2.3 | 27.6 |
| | PWR | 13.7 | 13.7 | 15.7 | 9.1 | 4.3 | 56.4 |
| | TOTAL | 22.5 | 19.6 | 19.6 | 15.7 | 6.5 | 84.0 |
| T4 | BWR | 5.8 | 3.9 | 2.6 | 4.3 | 1.5 | 18.1 |
| | PWR | 8.9 | 8.9 | 10.3 | 6.0 | 2.8 | 36.9 |
| | TOTAL | 14.7 | 12.9 | 12.9 | 10.3 | 4.3 | 55.0 |

Figure 5. Modified BMDP-P4F run output for selected model.

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE
THETA(MEAN)  1.8038

ESTIMATES OF THE MULTIPLICATIVE PARAMETERS (BETA = EXP(LAMBDA)
EXP(THETA)  6.0725

***** ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

| | PSYSTEM | | | |
|---|---|---|---|---|
| PD | PM | PO | PT | PZ |
| .427 | .232 | .098 | .075 | -.832 |

***** RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES (LAMBDA) TO ITS STANDARD ERROR

| | PSYSTEM | | | |
|---|---|---|---|---|
| PD | PM | PO | PT | PZ |
| 3.783 | 1.894 | .723 | .593 | -4.543 |

***** ESTIMATES OF THE MULTIPLICATIVE PARAMETERS (BETA = EXP(LAMBDA)

| | PSYSTEM | | | |
|---|---|---|---|---|
| PD | PM | PO | PT | PZ |
| 1.533 | 1.261 | 1.103 | 1.078 | .435 |

***** ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

| FID | |
|---|---|
| BWR | PWR |
| -.359 | .359 |

***** RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES (LAMBDA) TO ITS STANDARD ERROR

| FID | |
|---|---|
| BWR | PWR |
| -5.196 | 5.196 |

***** ESTIMATES OF THE MULTIPLICATIVE PARAMETERS (BETA = EXP(LAMBDA)

| FID | |
|---|---|
| BWR | PWR |
| .698 | 1.432 |

Figure 5. (continued).

***** ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

| | EVDATE | | | |
|---|---|---|---|---|
| | T1 | T2 | T3 | T4 |
| | -.258 | .350 | .166 | -.258 |

***** RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES (LAMBDA) TO ITS STANDARD ERROR

| | EVDATE | | | |
|---|---|---|---|---|
| | T1 | T2 | T3 | T4 |
| | -2.285 | 3.777 | 1.692 | -2.285 |

***** ESTIMATES OF THE MULTIPLICATIVE PARAMETERS (BETA = EXP(LAMBDA)

| | EVDATE | | | |
|---|---|---|---|---|
| | T1 | T2 | T3 | T4 |
| | .773 | 1.419 | 1.180 | .773 |

***** ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

| FID | PSYSTEM | | | | |
|---|---|---|---|---|---|
| | PD | PM | PO | PT | PZ |
| BWR | .140 | -.054 | -.325 | .194 | .045 |
| PWR | -.140 | .054 | .325 | -.194 | -.045 |

***** RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES (LAMBDA) TO ITS STANDARD ERROR

| FID | PSYSTEM | | | | |
|---|---|---|---|---|---|
| | PD | PM | PO | PT | PZ |
| BWR | 1.245 | -.442 | -2.408 | 1.534 | .245 |
| PWR | -1.245 | .442 | 2.408 | -1.534 | -.245 |

***** ESTIMATES OF THE MULTIPLICATIVE PARAMETERS (BETA = EXP(LAMBDA)

| FID | PSYSTEM | | | | |
|---|---|---|---|---|---|
| | PD | PM | PO | PT | PZ |
| BWR | 1.151 | .947 | .723 | 1.214 | 1.046 |
| PWR | .869 | 1.056 | 1.384 | .824 | .956 |

(e)

Figure 5. (continued).

```
MODELS FORMED BY ADDING TERMS TO MODEL --   PF,T.

                                  MULTIPLE                 LIKELIHOOD-RATIO              PEARSON
MODEL                             EFFECT        D.F.       CHI-SQUARE      PROB          CHI-SQUARE      PROB
-----                             ------        ----       ----------      ----          ----------      ----
PT,PF.                                          15          21.01         .1367           18.08         .2586
          DIFF. DUE TO  ADDING    PT.           12          14.30         .2819           12.21         .4292
FT,PF.                                          24          30.26         .1763    (f)    24.64         .4256
          DIFF. DUE TO  ADDING    FT.            3           5.05         .1684            5.65         .1302
PFT.                                             0           0.00        1.0000            0.00        1.0000
          DIFF. DUE TO  ADDING    PFT.          27          35.31         .1313           30.28         .3015
STEP  1.          BEST MODEL FOUND IS -- PFT.
STEPPING STOPS DUE TO CRITERION PROBABILITY (      .250).

MODELS FORMED BY DELETING TERMS FROM MODEL -- PF,T.

                                  SIMPLE                   LIKELIHOOD-RATIO              PEARSON
MODEL                             EFFECT        D.F.       CHI-SQUARE      PROB          CHI-SQUARE      PROB
-----                             ------        ----       ----------      ----          ----------      ----
PF.                                             30          56.15         .0026           48.47         .0178
          DIFF. DUE TO DELETING   T.             3          20.64         .0001           18.19         .0004
F,P,T.                                          31          44.25         .0581    (g)    39.29         .1458
          DIFF. DUE TO DELETING   PF.            4           8.94         .0625            9.01         .0609
STEP  1.          BEST MODEL FOUND IS -- F,P,T.
STEPPING STOPS DUE TO CRITERION PROBABILITY (      .250).

   *****DELETION OF STRATA

VARIABLE  CATEGORY  CHISQUARE  D.F.    PROB.
PSYSTEM   PD            14.85     9   .09521
PSYSTEM   PM             8.28     9   .50656    (h)
PSYSTEM   PO            14.83     9   .09561
PSYSTEM   PT            14.87     9   .09450
PSYSTEM   PZ            10.00     9   .35076
```

Figure 5. (continued).

A basic way of assessing how good the model fits the data is through evaluating the chi-square ($\chi^2$) statistic computed from the observed and expected data. The $\chi^2$ statistic can be computed either through the likelihood-ratio (LR) method or the Pearson method. In both cases the statistic is a measure of the overall amount of deviation between the expected and observed cell frequency counts. These statistics are presented in Figure 5(a). The magnitude of the $\chi^2$ statistics must be evaluated relative to the specific configuration for a prescribed table and relative to the specified model. For this model the $\chi^2$ statistics and associated probability levels are (35.31, P = .13) and (30.28, P = .30), respectively for the LR and Pearson statistics. The probability level can be interpreted as the probability of getting a "larger" $\chi^2$ value under the hypothesis that the model is "true". Low probability values imply that either the model is not acceptable or the observed data is rare for that model. The observed values for the PF,T model indicate a moderately good fit with this model; that is, the data gives us no reason to suspect that the model is totally inadequate. However, it could be of interest and benefit to pursue trying to enhance the fit of the model to the data. Two common ways this is done are through addition or deletion of terms to the model and exclusion of "other" cells from the data.

In Figure 5(f) we see the effect of including additional terms in the model. Notice a better fit is obtained when the term FT is added. For the resulting model (FT,PF) the LR and Pearson $\chi^2$ statistics are .17 and .42, respectively. The "significance" of this observed change in the adequacy of the fit, again, is a subjective question which can only be answered specific to the particular application.

One might also try to enhance the model by deletion of terms. In particular, we will examine the term PF. The effects associated with the interaction between PSYSTEM and type of facility (PF) are shown in Figure 5(e). For example, we see that, after adjusting for the main (overall) differences between the selected BWR and PWR plants, for design activities (PD) the BWR plant counts on the average remain about 15% higher than the average, while for operation activities (PO) PWR plant counts remain about 38% higher than the average. This indicates the relationship between PSYSTEM and FID may depend on the particular PSYSTEM and FID combination. The need to include this term in the model is reflected in the $\chi^2$ statistics we obtain with the term deleted. Figure 5(g) shows that the LR and Pearson $\chi^2$ statistics for the model (P, F, T) (without the interaction term) are .05 and .14, respectively; these in contrast to .13 and .30 with PF included.

The final section of the BMDP output, see Figure 5(h), provides information about the level of PSYSTEM which has the greatest impact on the fit of the model. Maintenance is the type of activity which is least accomodated by the selected model.

The fact that the PF,T model provides a reasonable fit coupled with the fact that the $x^2$ statistics do not improve through adding PT to the model (Pearson $x^2$ actually decreases) indicates that the distribution of the personnel activities studied doe not significantly change from one quarter to another in 1981 for the high-reporting plants.

## CONCLUSIONS

In this paper we have illustrated a system for automated trend and pattern analysis of operational data from nuclear power plants. The system permits one to analyze broad segments of the reported data and generate tables of counts. These tables are useful for obtaining overviews of the data and qualitatively identifying outliers, even without the log-linear modeling capability. The time-consuming nature of producing such tables in the past has now been overcome.

With log-linear modeling, a capability to statistically investigate the relationships present in the data has been added and demonstrated. The application of this capability to SCSS data is in its preliminary stages; for example, the normalizing capability for studying hazard rates was not demonstrated in the trial application because its use in conjunction with collapsing categories to produce less sparse tables is still under development. However, we are confident that log-linear hazard rate modeling is appropriate for this data and we expect to learn more about operational events through engineering analysis of events flagged by these methods as they evolve further.

## REFERENCES


1. Sequence Coding and Search System: User's Manual for the SCSS Data Base, ORNL/NSIC-190 (Draft), April 5, 1982.

2. M. Trojovsky, Data Summaries of Licensee Event Reports of Pumps at U.S. Commercial Nuclear Power Plants, January 1, 1972 to September 30, 1980, NUREG/CR-1205, Rev. 1, (EGG-EA-5524), January 1982.

3. E. B. Henry, C. D. Gentillon, A User's Guide to CONTING, EGG-EA-6402, September, 1983 (Draft; to be published as a NUREG).

4. W. J. Dixon, (Chief Editor), BMDP Statistical Software 1981, Berkeley, California: University of California Press, 1981.

5. Query/Update Reference Manual, Control Data Corporation Manual No. 60498300G.

6. Operating Units Status Reports, Licensed Operating Reactors, U.S. Nuclear Regulatory Commission, NUREG-0020 (published monthly).

7. Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, Discrete Multivariate Analysis: Theory and Practices, Cambridge, Mass.: MIT Press, 1975.

8. N. Laird and D. Olivier, "Covariance Analysis of Censored Survival Data Using Log-linear Analysis Techniques", Journal of the American Statistical Association (76:374), June 1981, pp. 231-240.

9. S. E. Fienberg, The Analysis of Cross-Classified Categorical Data, Cambridge, Mass.: The MIT Press, 1977.

10. D. Edwards and S. Kreiner, "The Analysis of Contingency Tables by Graphical Models", Biometrika (70:3), December 1983, pp. 553-565.

# GOVERNMENT AND STATISTICS: THE ASA/NRC EXPERIENCE

by

Bernard Harris
Mathematics Research Center
and Statistics Department
University of Wisconsin
Madison, WI 53706

Panel Discussion

# GOVERNMENT AND STATISTICS: THE ASA/NRC EXPERIENCE

Bernard Harris

## Abstract

The following material is part of the panel discussion on the activities of the ASA Ad Hoc Advisory Committee on Nuclear Regulatory Research. The discussion that follows is intended to describe the operations of the committee and its interactions with the Nuclear Regulatory Commission (NRC) from the perspective of a committee member.

## GOVERNMENT AND STATISTICS: THE ASA/NRC EXPERIENCE

In this brief discourse, I wish to describe the interactions of the ASA Ad Hoc Advisory Committee on Nuclear Regulatory Research with the Nuclear Regulatory Research Branch of the Nuclear Regulatory Commission (NRC). The remarks that follow will reflect my personal views of these interactions and their consequences to both organizations. These comments need not necessarily be the views of other committee members.

In order to put these comments into perspective, it seems desirable to first briefly describe how the committee came into existence and the organization of the committee. In response to a perceived need, the American Statistical Association entered into discussions with the Nuclear Regulatory Commission and following some negotiations, the committee came into being in October 1980.

Initially, the committee divided itself into three subcommittees:

1. Models and Analysis
2. Data Problems
3. Communication.

This division is significant to this discussion, since it reflects the committee's view of how it expected to operate, the areas in which it hoped to provide advice and the manner in which it hoped to provide such advice. That is, the committee envisioned three areas in which advice could be useful and divided themselves into these three subcommittees according to their talents and interests. This reflects a perception of the needs of the Nuclear Regulatory Commission in assessing nuclear reactor safety as follows:

### 1. Models and Analysis Subcommittee

This subcommittee would study mathematical and statistical models for reliability in order to ascertain their applicability and utility in problems of concern to the Nuclear Regulatory Commission. They would also study the methodology employed by the NRC to determine if it is adequate or if some modifications or perhaps even substantially different methodology is needed.

The committee planned to study system reliability modelling and the application of some recent proposals for modelling of complex systems, such as the MAXIMUS proposal. The agenda also included the analysis of common mode and common cause failure models, and the study of mathematical models for the frequency of initiating events, such as earthquakes, floods, or fires. The subcommittee would study mathematical models for human errors. In addition, the study of uncertainty and sensitivity analyses was considered to be within their purview. In any of these areas, it was deemed useful to create new models whenever appropriate.

### 2. Data Subcommittee

This subcommittee reviewed the data collection processes employed by the NRC and made appropriate recommendations. They studied and familiarized themselves with LER's and other data sources. In addition, they considered possible changes in processes that would increase accessibility of data by using suitable automated procedures for organizing data.

### 3. Communication Subcommittee

This subcommittee was concerned with advising on methods of communicating results and conclusions concerning nuclear reactor safety to the general public, political bodies and appropriate professional organizations by interpreting the statistical evidence in a manner conducive to facilitating such communication.

The above structure was abandoned after one year and only the first two functions were continued fro the balance of the life of the committee.

The Nuclear Regulatory Commission appeared to have somewhat different ideas concerning the utilization of the committee. Many of the meetings of the committee were devoted to two principal activities.

1. Briefings by the staff of the Nuclear Regulatory Commission. These served the purpose of informing the committee about their current projects and areas of concern. In addition, this also provided information on the methods that were currently being employed by the Nuclear Regulatory Commission in studying these problems.

2. Reviews of NRC research projects completed under external grants and contracts. Naturally, some time was needed for the committee members to become familiar with the terminology used by the NRC staff and with the problems that confronted NRC, in particular, those with which the committee could render some useful assistance. Nevertheless, from our standpoint, a significant fraction of the time allocated to such meetings was quite unproductive. In particular, I had the feeling that reviewing reports prepared under contract should be done by the NRC internally. However, some exceptions are certainly justified. Specifically, some of the reports did provide some insight into NRC concerns.

Despite the above negative comments, there were a number of significant benefits derived from these meetings. In particular, we became acquainted with a number of the phenomena of concern to the NRC. These include:

1. Damage to reactors due to floods, earthquakes, and so on.

2. Models for release and transport of radionuclides, such as Gaussian plumes, top-hat plumes, and others.

3. Maintenance problems, including pipe breakage and failure of diesel engines to start on demand.

4. The tendency of reactor vessels to become brittle.

5. Human error rates.

6. The inadequacies of data on component and system failures due to regulations as well as record keeping problems.

7. The inadequacies of data due to lack of accessibility of records

I feel that such meetings would have been substantially more productive, if, one day prior to the meeting, some time had been allocated to screening possible briefings in order to ascertain whether the committee might benefit from them. A similar procedure could have been undertaken in screening the reports to be reviewed.

Inasmuch as most of my activities on the committee were devoted to the Models and Analysis Subcommittee, a disproportionate amount of the comments that follow will deal with the activities of that subcommittee and in particular with my own activities.

Accordingly, I will now list the specific problem areas to which I have gained insight and to which I hope that I have made a contribution.

1. Probabilistic Risk Assessments. This is the most commonly used method for estimating the rate of nuclear reactor accidents. The Reactor Safety Study (1975) (also known as the Rasmussen Report) is the prototype of such studies.

2. Common Mode and Common Cause Failures. Some of the methods previously employed for the estimation of the probability of such failures have been highly controversial. For example, the procedures used in the Reactor Safety Study were severely criticized in the Lewis Report.

3. Precursor Methods. These are methods for evaluating the frequency of nuclear reactor accidents in which the observed occurrence of those events which can be completed to accidents are combined with estimates of the probabilities of the sequences needed for completion. This enables one to obtain estimates of the accident rates.

The committee did make a number of reasonable accomplishments once there was a clear perception of some areas in which they could be useful. A listing of some of these accomplishments follows.

1. Reports prepared under NRC Research Grants were reviewed and comments evaluating them were released to the NRC.

2. Recommendations on data collection and data processing were made.

3. The PRA (Probabilistic Risk Assessment) technique was reviewed and recommendations for improvements in the PRA process were made.

4. Some particular cases of implementations of PRA studies were reviewed and reports commenting on them were provided to the NRC.

5. A report on common mode and common cause failures has been sent to the NRC.

6. A report discussing precursor methods has been prepared and should be released to the NRC shortly.

I would like to close this brief discussion with some comments and suggestions for future advisory committees and how their interactions with government agencies may be improved.

First, a clearer understanding of potential contributions of the committee are needed. These should be related to the expectations of the government agency in such a way that both bodies will be satisfied with the advisory process.

An additional problem is a consequence of the Federal Advisory Committee Act. This imposes severe restrictions on the appointment and operation of advisory committees. Presumably the purpose of this act is to avoid the undue influence of special interest groups and to prevent conflicts of interest, as well as to insure that public access to such meetings is guaranteed. Personally, I do not regard this as being so serious a problem as to warrant the restrictions imposed by this act. In particular, while a special interest group may make suggestions to a government agency, the agency is capable of exercising discretion and is under no obligation to adopt the suggestions.

I would now like to close this commentary with some remarks concerning the benefits to my own career and activities that accrued as a result of my association with the committee.

A significant fraction of my own research is in the area of reliability theory. My activities with the committee provided me with a source of interesting problems and some insight into an important area of application.

I also became acquainted with the PRA methodology and even made some contributions to the PRA Guide through my membership on William Vesely's committee.

I learned a substantial amount about the present state of research on the propagation of uncertainty and on sensitivity analysis.

I will terminate this discussion with one negative note. I have sent a number of reports to the Office of Nuclear Regulatory Research, but as of this date, I have received no comments or feedback. I hope that some communication will be forthcoming soon.

# PROBABILISTIC RISK ASSESSMENT (PRA)
# AND
# DECISIONMAKING UNDER UNCERTAINTY

by

Malcolm L. Ernst
Division of Risk Analysis
U.S. Nuclear Regulatory Commission
Washington, DC 20555

Tutorial Session 2

# PROBABILISTIC RISK ASSESSMENT (PRA) AND DECISIONMAKING UNDER UNCERTAINTY

Malcolm L. Ernst, Acting Director
Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission

## Abstract

Decisions must be made, in spite of uncertainty. If the uncertainty is too large, whether in a probabilistic or a deterministic sense, then the regulatory decision might be not to proceed with the activity, or to proceed with appropriate conservatism expressed in cautions and limitations--but a decision is made nonetheless. While most regulatory decisions are made in a deterministic sense using appropriate bounding conservatisms, in most cases the basic uncertainties involved are much the same as those inherent in a PRA. The use of PRA adds a new dimension for the decisionmaker--one of a more realistic display of the integrated interactions of all systems, including the human. These insights are important, but the bottom-line risk or core melt numbers must be used with caution because of the inherent uncertainties.

Use of PRA as an influence on decisions suffers from the tendency to go too quickly to the bottom line, which is the weakest part of a PRA. Numerical criteria such as safety goals need to be constructed and implemented in such a way as not to drive one even faster to the bottom line. One needs to stop along the way and pay particular attention to the design and operations insights derived from the analyses. Therefore, the performance of the PRAs and the display of results and uncertainties should be suitably constructed so as to provide convenient and scrutable stopping places for the decisionmaker, to encourage viewing these insights and understanding the underlying assumptions and uncertainties, and to discourage undue fixation on the bottom line. Only then will regulation be able to draw fully upon the potential benefits of PRA as an information source and regulatory tool.

Decisions cannot be made by-the-numbers in a complex technology. Neither PRAs nor safety goals can be used as conveyances for easy decisions. Acceptable risk as portrayed in the Commission's safety goals is a subjective judgment that cannot and should not be interpreted as a clear go no-go numerical criterion. One must remember that PRA results are uncertain, but also their bottom-line results are being compared to objectives that are believed to be a conservative, and perhaps equally uncertain, estimate of society's perceived tolerance of risk.

The general topic for this session of the Statistics Symposium is acceptable risk and safety goals, and I was asked to discuss experiences and strategies at the NRC. After giving this subject considerable thought, I felt I would

be responsive to this charge if I tried to put into perspective the past practices of the NRC with regard to acceptable risk and decisionmaking under uncertainty, the evolving role of PRA as an analytical tool and aid to decisionmaking, and the thinking behind the Commission's safety goal. Indeed, because of the uncertainties involved, there must be enough grist in these topics to keep any statistician's mill grinding.

The focus will be principally on PRA, because it is the only quantitative tool that can be used to estimate risk. The focus will not be on the definition of acceptable risk--the Commission's safety goals will be taken as a given in this area. With regard to acceptable risk, past NRC policies and practice have used a qualitative approach--reasonable assurance of no undue risk. The safety goals policy statement is an attempt to define acceptable risk in a more quantitative manner; to make the NRC's policy more explicit and understandable and to set more definitive boundaries on regulation.

## What is a Decision?

There are very few decisions made on any subject today that do not have some degree of uncertainty. The only sure things in life are death and taxes; but there is precious little that we, as decisionmakers, can do to alter the outcome of these events. Therefore, we are stuck with making the best of uncertainty in the vast majority of our decisions.

While the NRC is specifically charged with protecting the public health and safety in matters involving the peaceful uses of nuclear energy, it is recognized that the backdrop of this charge is the anticipation that, with proper regulation, such uses can occur without an undue threat to public health and safety. Therefore, the NRC is clearly charged to strive toward a reasonable balance in regulation, and cessation of the activity it regulates should only take place if there is not reasonable assurance that the public is being adequately protected.

The interesting thing about decisions is that, if you don't take any action, that is still a decision. In this context, the lack of a decision to issue a nuclear power plant operating license would signify unacceptable risk, and the lack of a decision to shut down an operating plant would signify acceptable risk. Both such inactions clearly are decisions. Thus, every day the NRC, through action or inaction, is called upon to make decisions; and these decisions must be made using the best available knowledge and in spite of uncertainty.

## How Does the NRC Regulate and Make Decisions?

The traditional regulatory process for limiting nuclear power risks is deterministic, based on the concept of defense-in-depth which involves plant design, operations, siting, and emergency planning. The defense-in-depth concept emphasizes good management; quality assurance; conservative design, contruction, and operations; prevention of core damage accidents by requiring appropriate emergency shutdown and cooling systems; mitigation of any accidents that might lead to core damage through the use of systems that reduce the amount of fission products released to the environment; siting in areas that

are not in close proximity to highly populated areas; and good emergency planning. Analyses to demonstrate compliance with NRC's requirements have generally been based on conservative engineering judgment, through the use of design basis accidents (DBAs), with little emphasis on probabilistic assessments as to the likelihood of meeting the engineering intent of the requirements. However, even the NRC's deterministic approach to licensing has been sprinkled with judgments regarding the likelihood of occurrence of certain events. These are apparent, for example, in the establishment of requirements for redundancy and diversity, including use of the single failure criterion.

The DBAs are a set of accidents chosen to envelope credible accident conditions. Design and operations should be such that these accidents will not cause substantial degradation of the core, and conservative estimates of the radiological impact of such accidents have to be limited to prescribed values using engineered safety features and appropriate siting. The success of this approach has been demonstrated in that it is widely recognized that accidents outside of the design basis envelope dominate the estimated low levels of risk associated with the operation of nuclear power plants. The more probable causes of such accidents are believed to originate from multiple failures or human errors outside the domain of the single failure criterion and those common cause failure mechanisms currently addressed in the regulations (seismic qualification, safeguards, fire protection, etc.). While this fact does not negate the effectiveness of the NRC's regulatory practices, it does raise a question as to whether additional protection for accidents beyond the design basis should be provided--thus, the current widespread interest in degraded core accidents and PRA.

The principal shortcoming of the traditional deterministic process is that it does not include an effective means for conducting an integrated and systematic systems analysis of the plants. PRA is designed to fill this gap. It is a powerful tool for organizing information into a logical framework and providing insights into the complex interrelationship among systems and humans in a nuclear power plant.

PRA can provide a comprehensive, balanced, realistic model for predicting reactor risks by performing a systematic review of the design and operation of a nuclear power plant from a risk and reliability perspective. It is not constrained to design basis events, but instead provides an integrated assessment of front-line safety systems, support systems, and plant operation with respect to core melt, containment failure, and radiological consequences.

PRA provides the regulator with additional tools to: (1) conduct a broad, realistic inquiry into plant vulnerability to severe accidents; (2) understand the character, variety, and importance of these vulnerabilities; and (3) estimate the net risk reduction value of alterations in design or operation. It also provides the regulator with a frame of reference to better understand and interpret the lessons from experience with operating reactors.

Through the modeling of plant performance and the interactions between systems and humans, and the subsequent quantification of success/failure paths, a number of potential weaknesses in plant design and operating, test, and maintenance procedures have been identified, even though a plant may meet the NRC's deterministic requirements. Thus, PRA can be effectively used by

the industry and the NRC to supplement conventional engineering evaluation techniques to enhance safety as well as to improve plant availability.

In spite of the potential advantages offered by PRA, however, the uncertainties inherent in the assessments have appropriately limited their usefulness in regulation. These uncertainties are pervasive, since they are a function of the assumptions made by the individual analysts, the state of knowledge of plant response to a given stimulus, and the quality of the data base; and the PRAs are so complex that it is difficult not only to identify the uncertainties, but also to quantify and propagate them so as to measure their importance relative to central estimates of absolute risk. The end result is that, after a PRA is completed, the pervasive uncertainties result in doubt as to the actual importance of the identified design and operational weaknesses and questions as to whether other weaknesses of perhaps even more risk importance have been overlooked. However, this does not negate the qualitative and quantitative insights that are gained.

Uncertainties arise in many areas, as we are all aware. Principal among the uncertainties in PRA are the modeling of human actions, assumptions on success/failure criteria, effects of test and maintenance, completeness of accident initiators (e.g., sabotage), common cause failure mechanisms (including seismic, flood, and fire-induced failures), the phenomena of core melt progression and in-plant and ex-plant fission product transport, threats to containment integrity, partial failures, rectification, design and fabrication errors, equipment performance, treatment of generic and plant-specific data, failure and recovery times, meteorological conditions, radiological health effects, and the effectiveness of offsite emergency actions.

Clearly, the uncertainties are not all in one direction. Many may result in an underestimate of the risk, such as incompleteness in accident initiators and common mode failures. Others may overestimate the risk, such as any use of conservative failure criteria, failure to consider appropriately partial failures or ad hoc operator intervention, and the use of conservative source terms. Therefore, it is difficult to judge whether the bottom line results of any PRA represent an underestimate or an overestimate of risk. However, most experts in the field agree that reasonable bounds on PRA estimates of the risk from a nuclear power plant are about an order of magnitude around the central estimate, and the bounds likely are larger for external phenomena (such as seismic, fire, and flood) acting on the plant systems.

No technical analysis, whether deterministic in nature or probabilistic, is ever formally complete or completely certain. Also, in most instances, the uncertainties identified in PRAs are equally applicable to the more deterministic analyses. Therefore, it is important that the decisionmaker understand all significant uncertainties so as to make more optimum use of all provided analyses, including the information contained in PRAs.

Given these large uncertainties, how can PRA be used in the regulatory process with any degree of confidence? That is the question one so often hears. I will ask a different question. Given the strengths of PRA in providing an integrated look at the performance and interactions of systems, components, and humans, how can PRA not be used in the regulatory process? Our task must be not to decide whether to use PRA, but how to use it most effectively.

It should be emphasized that PRA is not a decisionmaking tool nor is it a magic formula which can be substituted for sound judgment. PRA is an analytic tool which is being and should be used to supplement (not supplant) the deterministic analyses traditionally used by the NRC staff. The process of considering deterministic and probabilistic information, weighing policy alternatives, and selecting the most appropriate regulatory action requires integrating PRA insights with engineering judgment and operating data as well as with social, economic, and political concerns. No one said that decision-making was easy, and certainly PRA should not be expected to make it much easier.

What are Potential Uses of PRA in Decisionmaking?

The current situation regarding PRA applications in regulatory decisionmaking is one of increasingly widespread use as an analytical tool to add an additional perspective to safety analysis, but also with increasing concern over uncertainties and the credibility of bottom-line numbers. With only the rarest exceptions, there have not been direct applications of PRA in plant-specific decisions; but the methods are now almost routinely used in the "high-medium-low" sense for assigning priorities among both generic and plant-specific safety issues and in considering regulatory revisions. One recent application of great importance is the heavy reliance by all parties in the special Indian Point ASLB hearing on the methods and results of PRA. Insights from the Indian Point PRA resulted in a few relatively inexpensive modifications and procedural changes that offered substantial safety benefits at modest cost. Another example is the use of PRA insights in the SEP (Systematic Evaluation Program) review of the ten oldest operating plants, to help in decisionmaking on hardware backfits and procedural changes. Still another is the continuing use of probabilistic perspectives in resolving unresolved safety issues (USIs). Examples include ATWS, station blackout, shutdown heat removal, pressurized thermal shock, and DC power. A plant-specific application of some note was for Big Rock Point: the utility-sponsored PRA was used to demonstrate that many suggested safety-related retrofits would not be cost-beneficial because of the specific design, size, and siting of the Big Rock Point station.

Some areas where PRA might someday contribute importantly are still in an evolving stage of development. These include accidents initiating from fires, where the first PRA applications on a broad systems level have shown the techniques to be useful but in need of further development; and from earthquakes, where there has already been substantial development under NRC and private sponsorship which now allows quite useful insights, even though the quantitative results of the analyses are quite uncertain and presently should not be credibly compared directly to risks from internal accident initiators. Another example is the study of core melt progression and fission product transport, where the incorporation of a better understanding of the physical phenomena and containment performance into probabilistic models is now in a very active stage of development.

More important, several regulatory initiatives now of greatest visibility within the NRC apply probabilistic thinking or analysis as an integral part of the approach: examples include the severe accident arena, the rethinking of the siting and emergency preparedness regulations, the human factors area, the

analysis of operational data and events, and improved approaches to regulating reactor operations. Perhaps the most important area now moving toward regulatory closure is the "severe accident" arena, where possible rulemaking or changes in other aspects of regulation is now under active consideration within NRC, and where probabilistic methods and insights that include numerous PRA studies are anticipated to form an important information base to aid decisionmaking.

Potential uses of PRA in the future fall into three basic categorizations: prioritization of NRC resources; generic regulatory applications; and plant-specific usage. Each of these categories places somewhat different demands on the quantitative credibility of the PRA results.

Prioritization of resources--Even considering the inherent incompleteness of the models and the uncertainties associated with the quantification of models in a probabilistic risk assessment, because of its integrated nature and reliance on realistic information, a probabilistic risk analysis presents our best available information concerning the specific ways in which the critical safety functions at nuclear power plants can fail to be performed, and the importance of such failures. This information can be used to guide and focus a wide spectrum of activities designed to improve the state of knowledge regarding the safety of nuclear power plants. The resources of NRC, as well as those of the industry, are limited; and the application of probabilistic risk analysis techniques or insights from previous studies could permit the decisionmaker to allocate these resources to issues most likely to reduce risk or better define or limit the uncertainties. Examples of areas where prioritization is required include allocating resources to the resolution of generic safety issues, establishing priorities for limited research funds, and allocating staff to those inspection modules and activities of most importance to safety.

All issues in the above areas are not really amenable to reasonable quantification. For example, it would be difficult to quantify the importance of a quality assurance issue. However, most issues can be reasonably quantified, and the nature of the decisions necessary to allocate regulatory resources does not require great precision in PRA results. It is sufficient to prioritize confirmatory research and the efforts used to resolve generic safety issues generically into broad categories (e.g., high, medium, and low). The reasoning is that one would not dismiss a potential safety issue unless it were clearly of low risk. Thus, a few completed PRA studies can be selected as surrogates even though it is apparent they do not fully represent the characteristics of some plants, provided the nature of these differences are reasonably understood and can at least be qualitatively evaluated.

The uncertainties involved in the risk measures used for prioritization are such that only large (at least an order of magnitude) variations in the comparative results should be considered significant. Thus, if severe core damage frequency were one of the measures, it would be improper to conclude that an issue having an estimated damage frequency of $3 \times 10^{-5}$/RY is significantly more important than an issue assessed as $1 \times 10^{-5}$/RY, but it would normally be appropriate to prioritize on the basis that an issue assessed as $10^{-4}$/RY is substantially more important than another issue evaluated as $10^{-6}$/RY.

Generic regulatory applications--The insights gained from the integral view of the probabilistic risk analysis can identify gaps in the present regulatory concept of defense-in-depth or in the detailed application of that concept. Due to the disciplined, integrated nature of the review, virtually every probabilistic risk analysis performed to date has identified some feature of the plant, previously unrecognized, which has had a measurable impact on either the frequency of severe core damage or the risk to the public associated with the facility. Many times the weakness involves system interactions or dependencies. It is possible to examine these gaps and, if necessary, develop deterministic criteria which remove these weaknesses from further regulatory consideration.

However, there is an important question relating to the applicability of the results of existing PRAs to all plants. This is the so-called surrogate problem, which I would prefer to call the problem of generic applicability.

At one time, it was thought that WASH-1400 might adequately display the risk and dominant contributors to risk for all LWRs. Work completed since then, however, has clearly indicated that class-specific and plant-specific differences can substantially affect the estimated core melt frequency and risk of a plant, as well as the dominant accident contributors. This is so because many of the risk significant features of a plant are dependent on balance-of-plant design. Therefore, if one wanted to establish surrogate plant classes based on risk estimates, the number of plant classes would be larger than the number of NSSS designs with containment variations.

While there is substantial ongoing work on the surrogate question, both by the NRC and the industry, it appears that, if reasonable accuracy is desired, there will likely need to be several dozen plant classes to describe the risk; and these classes still could not be used to provide an accurate central estimate of risk of a specific plant due to the inherent uncertainties and the potential for important plant-specific design or operational differences.

How then can one use PRA as an effective tool for generic decisionmaking? Again, I will answer that question by asking how we can afford not to use the insights available from PRA. My answer implies that some information, albeit incomplete, on risk is still better than no information. There is no requirement for perfect knowledge or perfect equity in regulatory decisionmaking; therefore, these standards should not be drawn across the trail of PRA--they are red herrings. Given a recognized safety problem for a given plant, the NRC frequently and in a deterministic manner has expanded the search to all plants and made generic, design-specific (e.g., based on NSSS design) decisions on fixes that clearly could be and have been argued as being uncertain and inequitable on a plant-specific basis. The use of PRA information would at least better define the importance of the uncertainties, and some of the inequity might even be reduced. Does that mean that PRA insights should not be used just because there will still be residual uncertainty and inequity? I think not. But one must be very careful not to place undue weight on the numbers themselves.

The principal benefits of PRAs in generic regulatory application are the ability to identify, for further study, potential plant weaknesses and to assess in a cost-benefit mode the effectiveness and relative merits of

alternative fixes. Even though one may not be able to describe generically the risk of a large number of plants using a single surrogate, one might find that a single fix may have comparable risk reduction potential across a number of different design classes. This could occur because a modification to design or procedures could affect a number of different accident sequences, one or more of which might be dominant in one design, while others might be dominant in other designs. Therefore, while there likely will be a large number of plant classes, there might be a significantly smaller number of "surrogate" (generic) potential risk reduction modifications.

The types of analyses that could support generic regulatory decisions vary in scope and depth. They could range from full-scope PRAs (even including external accident initiators, such as seismic) to limited scope reliability analyses similar to that performed several years ago on auxiliary feedwater systems. Uncertainties in the fuller scope analyses would have to be considered in detail to assess the likely generic implications with regard to the dominance of the sequences, the impact of various alternative fixes, and the applicability to other designs (the equity problem). For the limited scope analyses the equity problem still exists, because the system being reviewed will have more or less risk importance depending on the specifics of the overall plant design. However, as far as the study itself the uncertainties probably would be diminished, since relative insights would likely be sought using a prescribed analytical process which would tend to cancel out some uncertainties. The basic objective of such narrow studies would be to identify and assess risk-important system design and procedure alternatives across a number of different plant designs in the search for correctable weaknesses in system reliability. Of course, even these reliability studies generally still should be performed in an accident sequence context, to maintain some sensitivity to plant-specific design differences of major risk importance.

Given a reasonably thorough understanding of the risks and dominant sequences, the limitations of the insights as they derive from the scope and depth of the studies performed, and an understanding of the uncertainties involved in the analyses, the insights gained from probabilistic risk analyses can be used to identify areas in which regulatory action is necessary either to significantly lower the probability or consequences of certain types of accident sequences, or to relax regulatory requirements when they do not have a significant impact on either the estimated risk to the public or the estimated frequency of core melting.

It is possible that there presently is a sufficient collection of PRAs such that the risks of one or more accident sequences and the benefits of possible fixes can be reasonably understood as a function of plant class, including an understanding of the uncertainties in the analyses. Of course, there could be plant-specific risk outliers that would swamp the relative importance of particular accident sequences or possible fixes at a given plant; but this still would not alter the absolute importance of that sequence or fix. Also, it is recognized that there could be some plant-specific design or operational characteristics that would substantially alter both the absolute and relative importance of particular accident sequences or fixes at a particular plant; but that should not be a strong argument for not taking appropriate action on a significant safety problem.

Plant-specific applications--The third area for the potential use of PRAs in decisionmaking lies in plant-specific applications. As recognized in the Commission's safety goal policy statement, the use of safety goals (and thus the bottom-line numbers of plant-specific PRAs) as licensing criteria that must be met is not appropriate at the present time. Such use focuses on the weakest part of a PRA--the accuracy of the bottom-line results. However, there are other important potential uses of a plant-specific PRA that could be beneficial.

A plant-specific probabilistic risk analysis, performed early in the design process, can yield a large number of insights regarding integral performance of the plant to the designers as they perform their detailed design; and the assumptions and boundary conditions used in the PRA can drive the design process so that they are included in the detailed design. At the same time, it can be used to focus quality assurance activities during the detailed design and construction, as well as during the development of operating, test, and maintenance procedures on those items which have the highest potential for affecting risk. The real significance of such an analysis is not the numerical values calculated, but rather the insights on important features of the design and critical man-machine interfaces which are identified and therefore can be considered in depth.

For operating plants, there are a number of potential uses of a PRA that would not be strongly dependent on the accuracy of the bottom-line numbers. Examples of these uses include:

1. The possibility of assessing plant-specific alternatives to the resolution of generic safety issues.

2. The possibility of plant-specific prioritization of inspection activities which, coupled with operating experience feedback, could more effectively utilize inspection resources in important areas such as quality assurance, maintenance, and testing.

3. Assessment of the risk importance of operating events, including assisting decisionmaking with regard to requests for relief from LCO requirements.

4. Identification of plant-specific design weaknesses such as functional systems interactions and other common cause failure modes, including those initiated by seismic events, fires, and floods.

5. Evaluation of operating and emergency procedures and the assessments of off-normal events that have the potential of progressing to a degraded core accident.

As indicated, the above uses do not draw their basic strength from the bottom-line numbers. The basic strength is the ability to pinpoint important areas that must be properly monitored, to understand accident sequences sufficiently well to be able to estimate margins of safety, and to better understand remedial actions that should be taken during accident situations. While all of these uses are of regulatory importance, they would be even more important to the utility itself, to help it run the plant more safely and reliably.

## What is the Role of Safety Goals in Regulatory Decisions?

On March 14, 1983, a policy statement on safety goals was issued by the Commission for public comment and a two-year evaluation period. This policy statement includes qualitative safety goals, as well as quantitative design objectives which could serve in the future as risk benchmarks for use by the NRC as part of the decisionmaking process on matters relating to nuclear safety. The Commission's policy statement and evaluation plan explicitly exclude the safety goals from use in licensing cases, and as a principal decision criterion in regulation, for the two-year evaluation period.

This statement of NRC safety policy expresses the Commission's views on the acceptable level of risks to public health and safety and on the safety-cost tradeoffs in regulatory decisionmaking. However, as clearly stated in the policy statement the quantitative design objectives are only aiming points, not firm requirements or limits. They are goals which plant designers and operators should meet where feasible, and they are not substitutes for existing regulations. Also, the basic purpose of the evaluation period is to permit a better understanding of the strengths and weaknesses of the techniques (PRA) by which one judges whether these objectives have been met, and to judge the effectiveness of the goals and design objectives.

Therefore, at the present time the safety goals have no real role in decision-making. However, what might their future role be? It would be difficult, if not presumptuous, to speculate; but one can make some judgments as to factors affecting the potential future role.

First, it is clear that in implementing the safety goals great care will have to be exercised to differentiate between risk levels that must be met and goals or aiming points that are only desirable to meet. Second, since there will always be substantial uncertainty surrounding the results of a PRA, careful thought will have to be given to the degree of confidence to be required in estimating whether the goals are reached. For example, if a 90 percent confidence factor is expected, either explicitly or de facto, this would be equivalent to increasing the stringency of the design objectives by as much as a factor of 10 in most cases; or by even more where the PRA results are even more uncertain, such as in estimations of seismic risk.

Finally, in implementing the safety goals care will have to be taken that the principal use and thrust of PRAs does not become the precise matching of bottom-line numbers with the quantitative design objectives. Avoiding such use will be difficult to do, given the structure of the safety goals. However, if this happens, the focus will be on the weakest element of a PRA; and the substantial insights to be drawn with regard to accident sequences, system reliability, and human performance will tend to be downgraded or even lost.

The role to be played by the safety goal is yet to be determined. However, it is clear that, practically speaking, the role will be strongly dependent on the final structure of the implementation plan.

## How Can PRA Best Serve the Decisionmaker?

Regardless of advances in the state of the art of PRA and the experience data base, there will always be substantial uncertainties in the results and opportunities for unintentional or even intentional bias. Because of this fact, the strengths and weaknesses of PRA must be kept firmly in mind when crafting an implementation plan for the Commission's safety goals. PRA is only a tool for providing information, and the safety goals can only be as effective as the tool permits. A mismatch between the strengths of the tool and the de facto implementation of the safety goals would do a disservice to both.

Recognizing the uncertainties involved in PRAs, several things can be done to help the decisionmaker understand the results of a PRA and factor the results appropriately into his or her decision rationale, as discussed below. This is an important subject, whether or not quantitative design criteria are finally adopted by the Commission for widespread use; because PRAs will still exist, and the results of the PRAs will be used by decisionmakers whether or not formal decision standards or criteria are adopted by the Commission.

Prescriptive methodology and assumptions--It is clear that some degree of prescriptiveness must occur, otherwise one would not know whether differences between PRAs were reflective of plant design and operating differences, or merely the result of the ideas of individual analysts. The degree of prescriptiveness of the methodology is an issue of substantial controversiality, since one does not want to destroy innovative thought.

Peer review--Some comfort can be drawn from a substantial peer review of the PRA, both during and after the conduct of the PRA; and the regulatory review process itself certainly will add credibility to the analyses.

Uncertainty analyses--Standard procedures will have to be prescribed for the performance of uncertainty and confidence analyses. This will have to cover analytical assumptions and phenomenology, as well as the stochastic uncertainties of the data base. The existence of an unusual degree of uncertainty must be identified for analysis and understanding.

Sensitivity analyses--Standard procedures will have to be prescribed for the performance of sensitivity and importance analyses, so that the decisionmaker can determine the weight that he or she should give to uncertainties.

Display of results--Because of the almost unlimited number of areas where uncertainty and sensitivity analyses can play an important role, careful thought will have to be given to reasonably standard requirements to enhance the full and scrutable display of results for decisionmakers. Scrutability and completeness will play opposing roles, and the challenge will be to achieve the first without sacrificing the second unduly.

## In Summary

Decisions must be made, in spite of uncertainty. If the uncertainty is too large, whether in a probabilistic or a deterministic sense, then the regulatory decision might be not to proceed with the activity, or to proceed with

appropriate conservatism expressed in cautions and limitations--but a decision is made nonetheless. While most regulatory decisions are made in a deterministic sense using appropriate bounding conservatisms, in most cases the basic uncertainties involved are much the same as those inherent in a PRA. The use of PRA adds a new dimension for the decisionmaker--one of a more realistic display of the integrated interactions of all systems, including the human. These insights are important, but the bottom-line risk or core melt numbers must be used with caution because of the inherent uncertainties.

Use of PRA as an influence on decisions suffers from the tendency to go too quickly to the bottom line, which is the weakest part of a PRA. Numerical criteria such as safety goals need to be constructed and implemented in such a way as not to drive one even faster to the bottom line. One needs to stop along the way and pay particular attention to the design and operations insights derived from the analyses. Therefore, the performance of the PRAs and the display of results and uncertainties should be suitably constructed so as to provide convenient and scrutable stopping places for the decisionmaker, to encourage viewing these insights and understanding the underlying assumptions and uncertainties, and to discourage undue fixation on the bottom line. Only then will regulation be able to draw fully upon the potential benefits of PRA as an information source and regulatory tool.

Decisions cannot be made by-the-numbers in a complex technology. Neither PRAs nor safety goals can be used as conveyances for easy decisions. Acceptable risk as portrayed in the Commission's safety goals is a subjective judgment that cannot and should not be interpreted as a clear go no-go numerical criterion. One must remember that PRA results are uncertain, but also their bottom-line results are being compared to objectives that are believed to be a conservative, and perhaps equally uncertain, estimate of society's perceived tolerance of risk.

## REFERENCES


1. Reactor Safety Study--An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400 (NUREG-75/014), October 1975.

2. PRA Procedures Guide, NUREG/CR-2300, January 1983.

3. N. J. McCormick, Reliability and Risk Analysis, Academic Press, New York, 1981.

4. R. von Mises, "On the Foundation of Probability and Statistics", Annual of Mathematics and Statistics, 12, 1941, pp. 191-205.

5. V. Barnett, Comparative Statistical Influences, Wiley and Sons, New York, 1982.

6. D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.

7. A. Kandel, Fuzzy Techniques in Pattern Recognition, Wiley and Sons, New York, 1982.

8. W. E. Vesely, T. C. Davis, R. S. Denning, and N. Saltos, Measures of Risk Importance and Their Applications, NUREG/CR-3385, July 1983.

9. W. E. Vesely, T. C. Davis, and N. Saltos, Measures of the Risk Impacts of Testing and Maintenance Activities, NUREG/CR-3541, October 1983.

10. A. D. Swain and H. E. Guttmann, Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, NUREG/CR-1278, October 1980.

11. E. E. Leamer, Specification Searches: Ad Hoc Influence with Nonexperimental Data, Wiley and Sons, New York, 1978.

Attendees:

1983 Statistical Symposium on National Energy Issues

Lee Abramson
Nuclear Regulatory Commission
MNBB 7602
Washington, DC 20555

Fred Balkovetz
EG&G Idaho-Inc.
P. O. Box 1625
Idaho Falls, ID 83415

Iris Banz
Westinghouse Electric Corporation
P. O. Box 355, MNC-416
Pittsburgh, PA 15230

Richard Beckman
Los Alamos National Laboratory
Mail Stop F600
Los Alamos, NM 87545

Thomas P. Bement
Los Alamos National Laboratory
Mail Stop F600
Los Alamos, NM 87545

Steve Bengston
EG&G Idaho-Inc.
P. O. Box 1625
Idaho Falls, ID 83415

Carl A. Bennett
Battelle Human Affairs Research Center
4000 NW 41st St.
Seattle, WA 98105

Deborah E. Bennett
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

Jane M. Booker
Los Alamos National Laboratory
Mail Stop F600
Los Alamos, NM 87545

W. M. Bowen
Battelle Northwest
P. O. Box 999, Battelle Blvd.
Richland, WA 99352

Lawrence A. Bruckner
Los Alamos National Laboratory
Mail Stop F600
Los Alamos, NM 87545

Maurice C. Bryson
Los Alamos National Laboratory
Mail Stop F600
Los Alamos, NM 87545

Beverly G. Cassidy
Westinghouse Electric Corporation
P. O. Box 355, MNC-416
Pittsburgh, PA 15230

David C. Cox
Battelle Columbus Laboratory
505 King Avenue
Columbus, OH 43201

Robert L. Dennig
Nuclear Regulatory Commission
AEOD-263A-EWS
Washington, DC 20555

Pamela G. Doctor
Battelle Northwest
P.O. Box 999, Battelle Blvd.
Richland, WA 99352

Mark Durst
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

David E. Fields
Oak Ridge National Laboratory
P. O. Box X, Bldg. 7509
Oak Ridge, TN 37850

Sandra Frattalli
Nuclear Regulatory Commission
MNBB 7602
Washington, DC 20555

Cindy D. Gentillon
EG&G Idaho-Inc.
P. O. Box 1625
Idaho Falls, ID 83415

Ronald E. Glaser
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

Rubin Goldstein
Combustion Engineering, Inc.
Dept. 9485-2408, 1000 Prospect Hill Rd.
Windsor, CT 06095

Peter Groer
Oak Ridge Associated Universities
Box 117
Oak Ridge, TN 37831

Oren V. Hester
EG&G Idaho-Inc.
P. O. Box 1625
Idaho Falls, ID 83415

Ronald L. Iman
Sandia National Laboratories
Division 1223
Albuquerque, NM 87185

Leslie Johnson
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

Samuel C. Kao
Brookhaven National Laboratory
AMD 515, BNL
Upton, NY 11973

Robert E. Kurth
Battelle Memorial Institute
505 W. King NE
Columbus, OH 43201

James A. Lechner
National Bureau of Standards
Administration A337
Washington, DC 20234

Ernest Linder
Penn State University
219 Pond Laboratory
Department of Statistics
University Park, PA 16802

J. H. Lofthouse
EG&G Idaho-Inc.
P. O. Box 1625
Idaho Falls, ID 83415

Dan Lurie
Nuclear Regulatory Commission
MNBB 7602
Washington, DC 20555

David Margolies
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

Michael D. McKay
Los Alamos National Laboratory
Mail Stop F600
Los Alamos, NM 87545

Richard W. Mensing
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

Joann M. Mines
Exxon Nuclear Idaho Co., Inc.
P. O. Box 2800, CPP-668
Idaho Falls, ID 83401

David L. Nelson
Boeing Computer Services
565 Andover Park West, MS 9C-01
Tukwila, WA 98188

L. D. Y. Ong
Nuclear Regulatory Commission
Office of Policy Evaluation
1717 H. St. N.W.
Washington, DC 20555

Jerome Puskin
Nuclear Regulatory Commission
MNBB 7602
Washington, DC 20555

Joan R. Rosenblatt
National Bureau of Standards
A438 Administration
Washington, DC 20234

David Rubinstein
Nuclear Regulatory Commission
MNBB 7602
Washington, DC 20555

John Van Ryzin
Columbia University and
Brookhaven National Laboratory
600 West 168th St.
New York, NY 10032

Rama Sastry
U.S. Department of Energy
EP-32
Washington, DC 20584

F. A. Seiler
Lovelace Research Center
P. O. Box 5890
Albuquerque, NM 87185

Abraham Silvers
Electrical Power Research Institute
74 Roosevelt Circle
Palo Alto, CA 94306

Martin Tanner
University of Wisconsin
1210 W. Dayton
Madison, WI 53706

Jacqueline K. Telford
Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, MD 20707

Terrance N. Troy
National Bureau of Standards
Office of the Comptroller
Washington, DC 20234

V. R. R. Uppuluri
Oak Ridge National Laboratory
Bldg. 9704-1
P. O. Box Y
Oak Ridge, TN 37830

Steve Verrill
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

Richard Weisinger
Lawrence Livermore National Laboratory
P. O. Box 808, L-316
Livermore, CA 94550

DISTRIBUTION

| | Copies |
|---|---|
| Nuclear Regulatory Commission, Bethesda, Maryland | 149 |
| Technical Information Center, Oak Ridge, Tennessee | 2 |
| Los Alamos National Laboratory, Los Alamos, New Mexico | 50 |
| Total: | 201 |

NRC FORM 335 (6-83) U.S. NUCLEAR REGULATORY COMMISSION

# BIBLIOGRAPHIC DATA SHEET

1 REPORT NUMBER (Assigned by TIDC, add Vol. No., if any)
NUREG/CP-0053
LA-10127-C

2 Leave blank

3 TITLE AND SUBTITLE
Proceedings of the Ninth Annual Statistics Symposium on National Energy Issues, October 19-21, 1983

4 RECIPIENT'S ACCESSION NUMBER

5 DATE REPORT COMPLETED
MONTH: May YEAR: 1984

6 AUTHOR(S)
Compiled by Maurice C. Bryson

7 DATE REPORT ISSUED
MONTH: June YEAR: 1984

8 PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Los Alamos National Laboratory
Los Alamos, NM 87545

9 PROJECT/TASK/WORK UNIT NUMBER

10 FIN NUMBER
A7225

11 SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555

12a TYPE OF REPORT
Conference Proceedings

12b PERIOD COVERED (Inclusive dates)

13 SUPPLEMENTARY NOTES

14 ABSTRACT (200 words or less)

The Ninth Annual Statistics Symposium on National Energy Issues was held in Rockville, Maryland, at the Holiday Inn Crowne Plaza, October 19-21, 1983, under the joint sponsorship of Los Alamos National Laboratory and the Nuclear Regulatory Commission. Sessions included two contributed-paper sessions, two tutorial sessions, and one discussion group. Included in these proceedings are those papers for which final copy was provided by the authors, together with a list of papers presented and a list of attendees.

15a KEY WORDS AND DOCUMENT ANALYSIS

15b DESCRIPTORS

16 AVAILABILITY STATEMENT
Unlimited

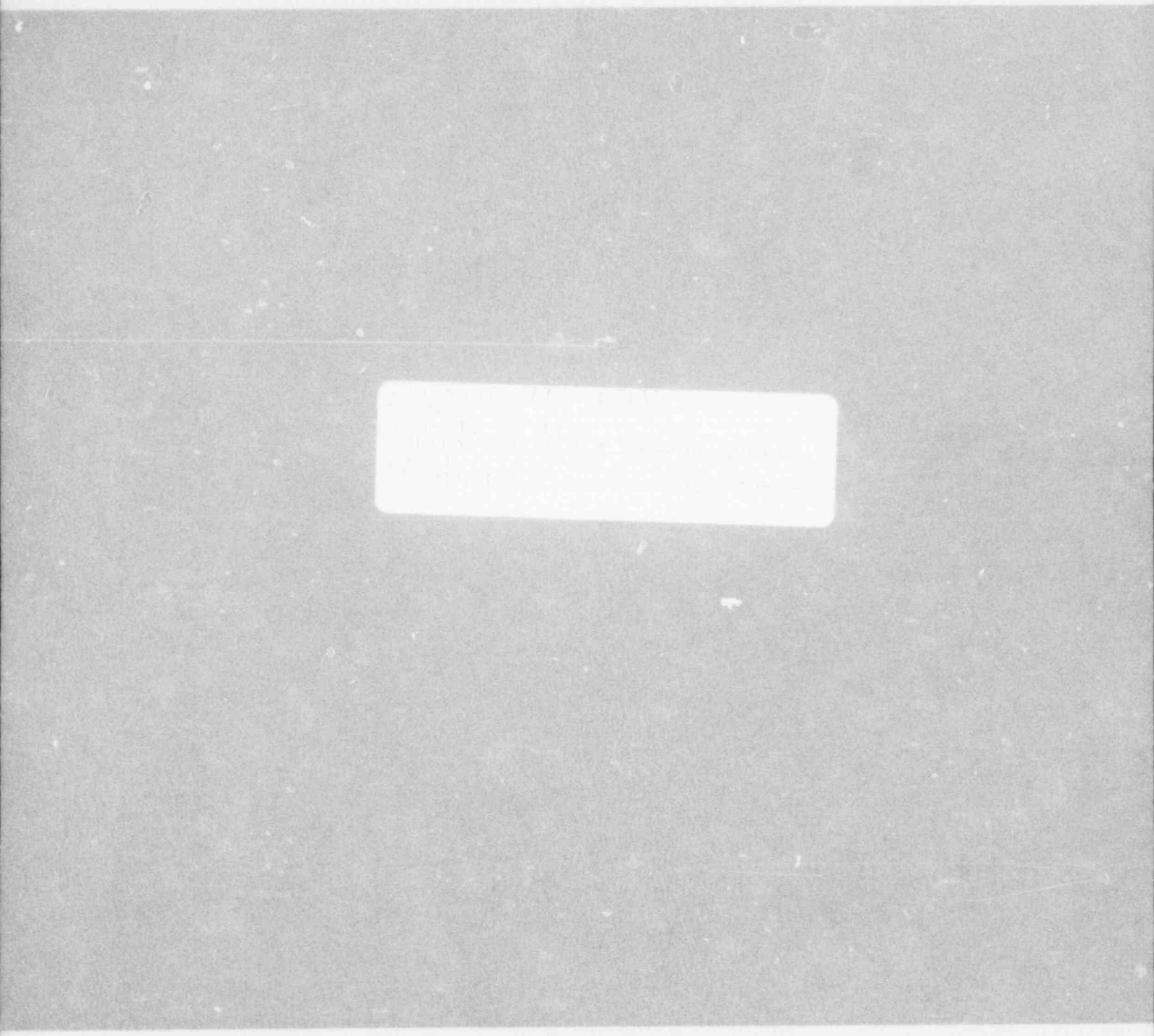
17 SECURITY CLASSIFICATION (This report)
Unclassified

18 NUMBER OF PAGES

19 SECURITY CLASSIFICATION (This page)
Unclassified

20 PRICE
$

☆U.S. GOVERNMENT PRINTING OFFICE 1984—776-026/4102



Los Alamos