LENDING REQUEST

TN: 323601
Borrower: OIP
Lending String: *UAF,WYU,DUQ,OKT,PMC
Patron: Eberle, W David

Journal Title: Risk analysis ; an official publication of the Society for Risk Analysis.

Volume: 14 **Issue:** 4
Month/Year: 8 1994 **Pages:** 395-404

Article Author/Title: Lognormal distributions for fish consumption by the general U.S. population

ILL Number: 107066941

Call #: T174.5.R55 PER
Location: Level 3
Request Date: 7/16/2013 12:13:29 PM

Shipping Address:
Boise State University
Albertsons Library - ILL
P.O. Box 46
Boise, ID 83707-0046

Fax: 208-334-2268
Ariel: 132.178.252.16
Email: libraryill@boisestate.edu

Attention:
ODYSSEY: 132.178.252.8

Initial when Updated to Found:

Risk Analysis, Vol. 14, No. 4, 1994

Lognormal Distributions for Fish Consumption by the General U.S. Population

Betsy Ruffle,[1] David E. Burmaster,[2] Paul D. Anderson,[3] and Henry D. Gordon[2]

Received June 3, 1993; revised January 3, 1994

The rate of fish consumption is a critical variable in the assessment of human health risk from water bodies affected by chemical contamination and in the establishment of federal and state Ambient Water Quality Criteria (AWQC). For 1973 and 1974, the National Marine Fisheries Service (NMFS) analyzed data on the consumption of salt-water finfish, shellfish, and freshwater finfish from all sources in 10 regions of the United States for three age groups in the general population: children (ages 1 through 11 years), teenagers (ages 12 through 18 years), and adults (ages 19 through 98 years). Even though the NMFS data reported in Ref. 14 are 20 years old, they remain the most complete data on the overall consumption of all fish by the general U.S. population and they have been widely used to select point values for consumption. Using three methods, we fit lognormal distributions to the results of the survey as analyzed and published in Ref. 14. Strong lognormal fits were obtained for most of the 90 separate data sets. These results cannot necessarily be used to model the consumption of fish by sport or subsistence anglers from specific sites or from single water bodies.

KEY WORDS: Fish consumption; lognormal distributions; general U.S. population.

1. INTRODUCTION

Consumption of fish containing environmentally persistent chemicals represents a potentially important pathway for human exposure. While this is of particular concern for individuals who consume large amounts of fish, it is also useful to know fish consumption rates for the general U.S. population for use in public health risk assessments and the development of Ambient Water Quality Criteria (AWQC) (see Refs. 4 and 21). The U.S. Environmental Protection Agency (U.S. EPA) and state agencies commonly rely on point estimates of the amount of fish consumed daily by various members of the U.S. population. For example, in developing freshwater Ambient Water Quality Control (AWQC), the U.S. EPA has historically used a per capita consumption rate of 6.5 g per day of freshwater and estuarine fish and shellfish for the general U.S. fish consumer.[20,21] Several states, including New York and Minnesota, use a consumption rate of ≥30 g per day everyday when setting AWQC.

Analyses of the distributions of freshwater fish consumption by the general public and sport anglers by Anderson *et al.*[2] and ChemRisk[6] suggest that many of these point values overestimate the average consumption for both groups. To date, no one has presented parametric distributions of saltwater fish and shellfish consumption by the general public. Such distributions would be useful in Monte Carlo simulations to quantify the *variability* associated with the risk to public health from fish consumption.

In 1980, Rupp *et al.*[14] published an analysis of the U.S. National Marine Fisheries Service's (NMFS) na-

[1] ENSR Consulting and Engineering, 35 Nagog Park, Acton, Massachusetts 01720.
[2] Alceon Corporation, P.O. Box 2669, Cambridge, Massachusetts 02238-2669.
[3] Ogden Environmental and Energy Services, 239 Littleton Road, Suite 7-C, Westford, Massachusetts 01886.

0272-4332/94/0800-0395$07.00/1

Table I. States Comprising the Census Regions of the United States

Region	States (postal abbreviation)
NewEng	ME, NH, VT, MA, RI, CT
MidAtl	NY, NJ, PA
SoAtl	DE, MD, DC, VA, WV, NC, SC, GA, FL
ENoCent	OH, IN, IL, MI, WI
ESoCent	KY, TN, AL, MI
WNoCent	MN, IA, MO, ND, SD, NE, KS
WSoCent	AR, LA, OK, TX
Mtn	MT, ID, WY, CO, NM, AZ, UT, NV
Pac	WA, OR, CA, AK, HI

tionwide survey designed to provide a representative sample of fish consumption patterns among the continental U.S. population. The results of the year-long survey, originally commissioned by the Tuna Research Institute (TRI) and conducted by NPD Research, Inc., in 1973 and 1974, were obtained and reanalyzed by the NMFS, with assistance from the U.S. Department of Agriculture (USDA), the U.S. Food and Drug Administration (FDA), and TRI. One-twelfth of the survey pool received the survey during each of the 12 months and were asked to record the number of meals and serving size for each type of fish eaten. The 12-month duration of the survey was designed to account for seasonal variation in fish consumption.

For the sample pool of 23,213 participants of known age, Rupp *et al.*[14] published selected percentiles (50th or median, 90th, and 99th), averages, maxima, and sample sizes for annual fish consumption for three age groups (children, ages 1 through 11 years; teenagers, ages 12 through 18 years; adults, ages 19 through 98 years), 10 regions of the country (the 9 census regions of the U.S. and the entire country as shown in Table I), and three categories of fish (saltwater finfish, shellfish, and freshwater finfish). The minimum regional sample size was 108 (for teenagers from the East South Central states), and the maximum regional sample size was 3303 (for adults from the Mid-Atlantic states). For each age group and fish category, Rupp *et al.* also analyzed the data from all states as a single group.

We contacted Rupp *et al.*[14] to determine whether additional percentiles are available. The two of the three authors who were reached reported that neither the original data used to develop the percentiles nor additional percentiles are now available. While the original NPD magnetic data tapes are available from the U.S. National Technical Information Service (NTIS), a review of the data indicated that reanalyzing the data on the tapes would be time intensive and redundant of the efforts already performed by Rupp *et al.*

Table I lists the states in each U.S. Census Region at the time of the NMFS survey, and Table II tabulates the 90 data sets for the daily consumption rates (denoted DCR, in units of g/day every day) as converted from Rupp *et al.* For each of the 90 data sets, Table II summarizes the results from the NMFS survey in terms of the 50th, 90th, and 99th percentiles of consumption (DCR50, DCR90, and DCR99, respectively), the average and maximum of consumption (DCR_{avg} and DCR_{max}, respectively), and the regional sample size (Count). We do not analyze the data for children eating freshwater finfish in New England because all the entries for percentiles and average are zero.

We worked with the summary statistics presented by Rupp *et al.*[14] for each of the 89 working data sets. In each data set, the DCR50 and DCR_{avg} values supply good information about the central location of the DCR distribution, and the DCR90, DCR99, DCR_{max}, and Count values supply good information about the upper tail of the distribution. The DCR_{avg} values contain information on the lower tail of the distribution in the sense that DCR_{avg} has contributions from each datum recorded.

In Table II, each of the DCR values for consumption of saltwater finfish is strictly positive. However, many of the DCR50 values and some of the DCR90 values for the consumption of shellfish and freshwater finfish are reported as zero, even though the NMFS survey included groups with 108 to >3000 people who were asked to respond for 1 month. The zero values reported in Table II make the statistical analysis of the results more difficult as explained below. However, it is possible that many of the zero values are, in fact, small nonzero values. The five working data sets with zeros reported for both the DCR50 and the DCR90 values probably reflect too short a period to capture the consumption by those people who eat shellfish or freshwater finfish infrequently. Specifically, because respondents reported on fish consumption for 1 month, the survey had <10% chance of capturing someone who eats only one fish meal per year. Assuming that the average fish portion for a single meal is ~200 g, such a person has a DCR50 of ~0.55 g/day, and not 0 g/day as reported by Rupp *et al.*[14] With a longer measurement period, the probability of measuring infrequent consumption would increase, resulting in small, but nonzero values at various percentiles.

The NMFS survey has other limitations. The participants may not have reported all fish consumed during the month-long survey, which would result in an underestimate of the amount of fish consumed. Participants also may have overestimated or underestimated portion

Table II. Daily Consumption Rates of Saltwater Finfish, Shellfish, and Freshwater Finfish (Data from Ref. 14)

Age group and region	Saltwater finfish						Shellfish						Freshwater finfish					
	DCR50 of data (g/day)	DCR90 of data (g/day)	DCR99 of data (g/day)	DCR_{avg} of data (g/day)	DCR_{max} of data (g/day)	Count (n)	DCR50 of data (g/day)	DCR90 of data (g/day)	DCR99 of data (g/day)	DCR_{avg} of data (g/day)	DCR_{max} of data (g/day)	Count (n)	DCR50 of data (g/day)	DCR90 of data (g/day)	DCR99 of data (g/day)	DCR_{avg} of data (g/day)	DCR_{max} of data (g/day)	Count (n)
Adults																		
All	7.29	23.73	57.59	10.68	179.12	15,576	0.00	10.96	31.56	3.59	120.93	15,576	0.00	5.12	22.99	1.48	158.03	15,576
NewEng	9.48	26.99	55.53	12.47	81.53	885	2.90	17.26	55.26	6.19	96.55	885	0.00	0.00	6.68	0.30	22.49	885
MidAtl	8.49	26.85	64.58	12.27	164.68	3,303	0.00	11.97	34.14	3.92	120.93	3,303	0.00	1.64	19.40	0.96	61.15	3,303
SoAtl	8.25	24.00	52.47	11.10	122.44	1,993	0.00	12.33	34.82	4.85	65.10	1,993	0.00	4.33	18.05	1.10	59.70	1,993
ENoCent	6.60	21.34	54.08	9.81	170.19	2,924	0.00	6.71	20.77	2.14	78.55	2,924	0.00	6.25	25.75	1.95	158.03	2,924
ESoCent	6.52	20.33	52.49	9.21	59.73	744	0.00	8.93	26.03	2.85	84.03	744	0.00	6.68	27.48	2.33	67.51	744
WNoCent	5.78	20.03	49.26	8.74	113.59	1,503	0.00	6.25	19.29	2.19	88.08	1,503	0.00	6.68	25.32	2.25	91.78	1,503
WSoCent	6.41	21.84	60.05	9.78	85.48	1,424	0.00	12.22	40.82	3.84	68.36	1,424	0.00	7.21	25.92	2.27	77.53	1,424
Mtn	6.58	19.59	42.25	8.90	74.41	659	0.00	8.93	23.84	2.85	52.88	659	0.00	5.07	24.63	1.48	61.53	659
Pac	7.84	25.59	56.93	11.37	179.12	2,141	0.00	11.64	29.62	4.05	86.16	2,141	0.00	4.05	17.40	1.07	48.71	2,141
Teenagers																		
All	5.15	15.81	37.62	7.23	100.14	2,274	0.00	3.75	10.71	0.93	38.68	2,274	0.00	2.77	15.48	0.85	31.37	2,274
NewEng	5.97	16.44	36.99	8.77	64.96	134	0.00	3.75	8.85	1.21	20.66	134	0.00	0.00	3.95	0.25	22.49	134
MidAtl	5.97	19.29	41.56	8.55	100.14	430	0.00	2.99	9.34	0.74	24.33	430	0.00	0.00	7.89	0.30	31.37	430
SoAtl	5.70	16.60	31.21	7.42	33.56	305	0.00	4.77	14.99	1.42	20.66	305	0.00	0.00	11.84	0.58	17.75	305
ENoCent	4.82	14.66	33.70	6.66	48.30	475	0.00	2.49	8.49	0.68	23.32	475	0.00	5.15	16.90	1.23	29.26	475
ESoCent	3.67	11.56	20.71	5.64	32.49	108	0.00	2.60	7.89	0.79	15.51	108	0.00	3.95	8.22	1.01	19.45	108
WNoCent	4.00	11.89	25.32	5.53	43.67	225	0.00	2.60	7.51	0.63	10.33	225	0.00	5.15	13.42	1.18	18.41	225
WSoCent	3.92	13.48	42.79	6.58	78.41	194	0.00	5.42	15.51	1.59	38.68	194	0.00	5.15	17.42	1.48	21.64	194
Mtn	5.45	13.34	20.68	6.49	28.93	114	0.00	3.75	8.16	0.85	16.58	114	0.00	3.95	12.22	1.40	25.78	114
Pac	5.59	16.44	39.23	7.86	59.84	289	0.00	3.75	10.19	1.01	21.53	289	0.00	0.88	12.71	0.63	24.85	289
Children																		
All	3.01	9.84	21.81	4.33	90.71	5,363	0.00	5.15	15.48	1.45	35.45	5,363	0.00	1.51	8.88	0.49	40.49	5,363
NewEng	3.78	10.79	19.40	4.99	32.08	311	0.00	7.01	13.42	2.38	14.68	311	0.00	0.00	0.00	0.00	0.00	311
MidAtl	3.37	10.49	22.79	4.77	62.30	1,263	0.00	5.15	12.11	1.56	20.11	1,263	0.00	1.26	6.66	0.33	27.59	1,263
SoAtl	3.26	10.14	19.78	4.38	31.37	549	0.00	8.08	15.78	2.22	16.44	549	0.00	0.00	7.51	0.38	17.23	549
ENoCent	2.82	9.07	19.01	3.97	33.15	1,003	0.00	3.95	10.19	0.90	35.45	1,003	0.00	1.51	8.38	0.55	40.49	1,003
ESoCent	2.74	10.33	18.36	4.44	39.45	221	0.00	2.25	8.58	0.96	19.78	221	0.00	3.95	8.22	1.01	9.67	221
WNoCent	2.19	7.34	18.05	3.26	90.71	577	0.00	3.34	6.25	0.74	10.38	577	0.00	2.85	7.89	0.66	13.42	577
WSoCent	2.60	9.73	23.56	4.16	41.42	510	0.00	5.15	26.93	1.92	31.53	510	0.00	2.68	11.23	0.77	21.64	510
Mtn	2.90	9.67	19.56	4.33	25.53	260	0.00	4.66	9.86	1.32	12.63	260	0.00	2.60	8.88	0.63	11.01	260
Pac	3.26	10.36	23.59	4.77	45.59	669	0.00	5.15	15.48	1.59	18.58	669	0.00	2.00	9.01	0.55	22.58	669

size, the effects of which would tend to balance. The type of fish consumed may also have been mistaken (e.g., tuna instead of whitefish). Because this type of miscategorization is more likely to occur across species within the main categories of saltwater fish, shellfish, and freshwater finfish than across categories, this limitation is not expected to affect overall consumption rates for the three broad categories examined here. Consumption of certain fish types on a long-term basis may also be underestimated. Someone may eat a certain kind of fish, such as rainbow trout, only a few times a year, and not during the month-long survey period. The NMFS survey did not focus on the consumption patterns of sport or subsistence anglers who catch and eat fish from particular water bodies.

Finally, the single greatest limitation of the data is that overall fish consumption has increased since the survey was conducted. To address this concern, we reviewed several sources to estimate the change in fish consumption by the general U.S. population, including two USDA sources[17,18] and one U.S. Department of Commerce (USDC) source.[19] According to the USDA, in the 10 years from 1977 to 1987, per capita fish consumption increased by ~16%. According to the USDC, per capita consumption of fish and shellfish increased ~24% between 1975 and 1985 and ~27% between 1975 and 1990. None of these sources distinguish between the broad categories of fish studied here. (In the Discussion below, we show how to adjust the results for the increased consumption.)

Despite these limitations, many of which are common to nearly all surveys, the NMFS survey included large regional sample sizes and was conducted over 1 year, so that Rupp *et al.*[14] believed that the consumption patterns were representative of year-round consumption across the continental United States in the 1970s.

We also note a number of strengths of the NMFS survey. Given that Rupp *et al.*[14] considered 23,213 individuals in their analyses, their results and our further analyses have few statistical problems related to small sample sizes. Because of the relatively large regional sample sizes and the fact that the original NMFS survey was designed to capture fish consumers throughout the continental United States and over 1 entire year, the consumption rates reported by Rupp *et al.* and the distributions presented here are representative of fish consumption patterns by the general population. Further, to the extent that the data include people reporting consumption >150 g/day, they include a subset of the general population that eats very large amounts of fish.

2. STATISTICAL METHODS AND RESULTS

Using graphical and numerical techniques from exploratory data analysis,[15,16] we found that the 89 working data sets in Table II do not come from truncated normal (or Gaussian) distributions. First, we note that DCR50 < DCR_{avg} for each of the 89 working data sets, implying positively skewed distributions. Second, using normal probability plots[10] written in Mathematica,[22] we fit (truncated) normal distributions to each of the 89 working data sets. We do not report these results because the fits strongly failed visual and quantitative tests for goodness of fit. We next investigated whether the 89 working data sets could be well fit by exponential distributions. We found that the longer-tailed exponential distributions do give better fits, but the results suggested that a family of distributions with even longer tails might fit even better.

We next investigated (two-parameter) lognormal distributions for two reasons. First, lognormal distributions are much easier to manipulate and fit to percentile data than gamma distributions, the next-most reasonable alternatives also with longer tails. Second, other studies of other consumption-related exposure variables have found that lognormal distributions provide good fits to the data sets.[12,13] We chose this form of the lognormal distribution:

$$\ln \mathrm{DCR} \sim N(\mu, \sigma) \iff \mathrm{DCR} \sim \exp[N(\mu, \sigma)]$$

where ln represents the natural logarithm, exp represents the exponential function, $N(\mu, \sigma)$ represents a normal or Gaussian distribution with parameters μ for the mean and σ for the standard deviation, and the double-headed arrow denotes equivalence.

We used three methods to fit lognormal distributions to the 89 working data sets.

2.1. LogNormal Distributions Fit by a NonLinear Optimization Method (NLO Method)

If the lognormal model holds exactly for a particular data set, then these five relationships obtain for appropriate values of μ and σ:[11]

$$\begin{aligned}
\mathrm{DCR50} &= \exp(\mu) \\
\mathrm{DCR90} &= \exp[\mu + z(0.90) \cdot \sigma] \\
\mathrm{DCR99} &= \exp[\mu + z(0.99) \cdot \sigma] \\
\mathrm{DCR}_{avg} &= \exp[\mu + 0.5 \cdot \sigma^2] \\
\mathrm{DCR}_{max} &= \exp[\mu + z(f_{max}) \cdot \sigma]
\end{aligned}$$

Here the function $z(f)$ computes the z score for the percentile located at fractile f, with $0<f<1$.[3]

For each of the 89 working data sets, we used a nonlinear optimization (NLO) method to find the optimal values for μ and σ using an equally weighted least-squares objective function. We used Mathematica[22] to find the values of μ and σ which minimize the sum of the squares of the discrepancies:

$$\text{Select } (\mu, \sigma) \text{ to Minimize } \langle \Delta_{50}^2 + \Delta_{90}^2 + \Delta_{99}^2 + \Delta_{avg}^2 + \Delta_{max}^2 \rangle$$

where, for example, $\Delta_{50} = \text{DCR50} - \exp(\mu)$. A zero minimum value for the objective function shows that the lognormal model fits a particular data set exactly, while a small minimum value of the objective function shows that the lognormal model fits a particular data set reasonably well. This NLO method is a "full information method" in the sense that it uses all six values reported by Rupp *et al.*[14] (DCR50, DCR90, DCR99, DCR_{avg}, DCR_{max}, and Count in Table II). Table III presents the optimal values for μ and σ from this NLO method, along with the minimum value of the objective function.

2.2. LogNormal Distributions Fit by a First-Probability Plot Method (PP1 Method)

If the lognormal model holds exactly for a particular data set, then the data points will plot in a straight line on a lognormal probability plot with the z values on the abscissa and the lnDCR values on the ordinate.[10] If the lognormal model does not hold exactly but does hold reasonably well for a particular data set, then the data points will plot in almost a straight line (with small scatter and little curvature) on the axes just described. On these axes, the linear regression of ln DCR as a function of z has an intercept equal to μ and a slope equal to σ. A high adjusted R^2 value (aR^2 value) for the linear regression supports the conclusion that a lognormal model fits the data well.

We used Mathematica to find the intercept and slope of the linear regression line on the logarithmic probability plot of lnDCR values against z values. Note that this PP1 method is not a full information method. Table III presents the optimal values for μ and σ from this PP1 method, along with the aR^2 value from the linear regression.

[3] The mathematical function for $z(f)$ is $z(f)$ = sqrt[2] · inverseErf [$2 \cdot f-1$], where inverseErf denotes the inverse error function.[1] By convention,[5,8,15] $z(f_{max})$ is computed as z[(Count−0.5)/Count], a value that changes for each data set analyzed.

2.3. LogNormal Distributions Fit by a Second-Probability Plot Method (PP2 Method)

To overcome the limitation that many of the data sets have DCR50 values reported as zeros, we also analyzed the data sets in Table II using a second probability plot (PP2) method. Again, we used Mathematica to estimate values of μ and σ by fitting linear regression lines to the 84 of the 89 working data sets with values of DCR90, DCR99, and DCR_{max} greater than zero. This PP2 method simply discards the DCR50 value (whether zero or positive) and estimates μ and σ by fitting a straight line to only three data points on the upper tail of the distribution. Note that the PP2 method is not a full information method. Again, a high aR^2 value supports the conclusion that a lognormal model fits the upper tail of a particular data set well. Table III presents the optimal values for μ and σ from this PP2 method, along with the aR^2 value from the linear regression.

3. DISCUSSION AND COMPARISON OF THE RESULTS

3.1. Comparison of the Goodness-of-Fit Measures

Using scatterplots, we found that the goodness-of-fit measures from the three methods used to test for lognormality are highly correlated. In other words, low minimum values of the NLO objective function are highly correlated with high aR^2 values for the linear regressions in the PP1 and PP2 methods.

Figure 1 presents a 3×3 array of graphs which compare the goodness-of-fit measures of the three methods for three data sets. In this figure, the graphs down each column pertain to one statistical method, and the graphs across each row illustrate the quality of the fit for one data set. The data set in the top row has excellent goodness-of-fit measures for all three methods, the data set in the middle row has acceptable goodness-of-fit measures for all three methods, and the data set for the bottom row has poor goodness of fit-measures for all three methods.

Table III. Best Fits of Lognormal Distributions Using Three Algorithms

Age group and region	Saltwater finfish NLO (μ)	NLO (σ)	NLO (minSS)	PP1 (μ)	PP1 (σ)	PP1 (aR^2)	PP2 (μ)	PP2 (σ)	PP2 (aR^2)
Adults									
All	2.311	0.720	30.13	2.078	0.800	0.99100	2.264	0.738	0.99244
NewEng	2.570	0.572	65.32	2.341	0.670	0.97422	2.617	0.563	0.96063
MidAtl	2.350	0.763	15.66	2.191	0.822	0.99617	2.319	0.776	0.99661
SoAtl	2.208	0.748	2.18	2.143	0.774	0.99847	2.231	0.741	0.99994
ENoCent	1.878	0.910	0.19	1.890	0.906	0.99996	1.898	0.903	0.99989
ESoCent	2.437	0.541	222.27	2.000	0.724	0.92742	2.391	0.571	0.74622
WNoCent	1.977	0.811	6.79	1.810	0.876	0.99596	1.965	0.818	0.99817
WSoCent	2.490	0.591	235.69	1.993	0.788	0.94698	2.368	0.647	0.85098
Mtn	2.045	0.716	4.64	1.930	0.766	0.99565	2.078	0.707	0.99860
Pac	1.919	0.934	19.41	2.063	0.883	0.99735	2.074	0.880	0.99043
Teenagers									
All	1.691	0.830	0.33	1.657	0.843	0.99964	1.703	0.826	0.99999
NewEng	1.325	1.046	59.08	1.746	0.857	0.98211	1.565	0.937	0.93813
MidAtl	1.374	1.057	46.01	1.773	0.900	0.98972	1.725	0.919	0.95795
SoAtl	2.098	0.508	39.39	1.856	0.623	0.93599	2.286	0.444	0.84634
ENoCent	1.874	0.661	18.73	1.638	0.763	0.98503	1.860	0.673	0.96647
ESoCent	1.286	0.816	18.80	1.334	0.793	0.97780	1.495	0.721	0.87741
WNoCent	1.258	0.879	5.54	1.390	0.823	0.99610	1.405	0.817	0.98057
WSoCent	1.014	1.194	5.60	1.317	1.065	0.99580	1.113	1.152	0.99727
Mtn	1.714	0.610	8.80	1.728	0.605	0.97836	1.882	0.536	0.89029
Pac	1.821	0.780	2.78	1.737	0.817	0.99879	1.796	0.792	0.99659
Children									
All	0.881	0.970	4.31	1.090	0.901	0.99627	1.063	0.911	0.98835
NewEng	1.376	0.706	2.32	1.372	0.710	0.99137	1.529	0.645	0.97929
MidAtl	1.049	0.917	3.84	1.214	0.857	0.99676	1.210	0.859	0.98766
SoAtl	1.427	0.651	2.52	1.264	0.725	0.98483	1.534	0.616	0.99865
ENoCent	1.347	0.658	3.39	1.128	0.750	0.98411	1.399	0.646	0.99309
ESoCent	0.678	1.041	27.09	1.051	0.888	0.96982	1.222	0.815	0.85115
WNoCent	−1.396	1.884	31.08	0.631	1.135	0.94346	0.133	1.334	0.88979
WSoCent	1.251	0.802	2.39	1.026	0.896	0.99196	1.263	0.801	0.99686
Mtn	1.391	0.650	4.20	1.155	0.758	0.97735	1.501	0.612	0.98590
Pac	1.307	0.792	0.38	1.219	0.829	0.99776	1.337	0.782	0.99996

Age group and region	Shellfish NLO (μ)	NLO (σ)	NLO (minSS)	PP1 (μ)	PP1 (σ)	PP1 (aR^2)	PP2 (μ)	PP2 (σ)	PP2 (aR^2)
Adults									
All	1.370	0.858	27.57				1.324	0.877	0.99190
NewEng	2.010	0.794	117.57	1.249	1.093	0.96279	1.804	0.877	0.94492
MidAtl	1.155	1.008	13.18				1.216	0.992	0.99993
SoAtl	1.702	0.718	72.00				1.633	0.753	0.94230
ENoCent	0.494	1.081	3.62				0.537	1.069	0.99994
ESoCent	0.280	1.294	6.43				0.658	1.160	0.98857
WNoCent	−0.217	1.379	3.42				0.174	1.248	0.98833
WSoCent	1.775	0.733	133.53				1.576	0.816	0.89396
Mtn	0.847	0.986	7.40				0.983	0.941	1.00000
Pac	1.177	0.938	13.01				1.294	0.903	0.99994
Teenagers									
All	−0.183	1.092	1.19				−0.033	1.046	0.99913
NewEng	−2.247	1.963	8.18				−0.193	1.132	0.85057
MidAtl	−0.784	1.305	0.66				−0.447	1.182	0.99366
SoAtl	0.539	0.862	10.24				0.454	0.906	0.94601
ENoCent	−0.959	1.336	0.35				−0.701	1.243	0.99691
ESoCent	−2.356	1.948	3.16				−0.713	1.272	0.93250
WNoCent	−0.239	0.919	1.82				−0.171	0.902	0.97631
WSoCent	−1.203	1.733	9.35				0.030	1.248	0.93948
Mtn	−1.537	1.640	7.15				−0.028	1.013	0.84058
Pac	−0.442	1.200	1.17				−0.053	1.052	0.98932
Children									
All	0.854	0.730	16.06				0.753	0.775	0.94476
NewEng	0.986	0.609	16.93				1.400	0.463	0.86194
MidAtl	0.724	0.688	9.37				0.853	0.657	0.95991
SoAtl	1.215	0.545	34.37				1.652	0.400	0.70704
ENoCent	−0.546	1.250	1.80				−0.084	1.089	0.97950
ESoCent	−1.431	1.554	0.43				−0.981	1.379	0.99236
WNoCent	0.029	0.745	2.12				0.418	0.612	0.99963
WSoCent	1.191	0.757	97.56				0.487	1.032	0.75487
Mtn	0.428	0.747	5.03				0.752	0.632	0.97500
Pac	0.916	0.657	27.02				0.870	0.692	0.78725

Age group and region	Freshwater finfish NLO (μ)	NLO (σ)	NLO (minSS)	PP1 (μ)	PP1 (σ)	PP1 (aR^2)	PP2 (μ)	PP2 (σ)	PP2 (aR^2)
Adults									
All	0.334	1.183	6.45				0.101	1.253	0.99264
NewEng	−1.548	1.433	2.44						
MidAtl	0.373	1.037	29.56				−1.151	1.527	0.86000
SoAtl	0.256	1.102	5.86				0.003	1.191	0.98742
ENoCent	−0.111	1.445	1.45				0.012	1.406	0.99932
ESoCent	0.662	1.109	10.90				0.401	1.206	0.98767
WNoCent	0.372	1.219	3.17				0.331	1.235	0.99931
WSoCent	0.678	1.085	7.89				0.565	1.127	0.99511
Mtn	0.379	1.182	10.16				−0.019	1.328	0.98325
Pac	0.443	0.986	10.67				0.068	1.117	0.96516
Teenagers									
All	0.578	0.822	23.51				−0.152	1.075	0.85362
NewEng	−7.698	4.033	3.97						
MidAtl	−2.584	1.981	1.08						
SoAtl	−0.467	1.153	16.23						
ENoCent	0.540	0.930	9.55				0.432	0.979	0.97397
ESoCent	−3.861	2.617	12.58				−0.060	1.068	0.73754
WNoCent	0.421	0.889	6.00				0.598	0.830	0.98350
WSoCent	0.541	0.926	12.93				0.424	0.984	0.94426
Mtn	−2.352	2.129	8.34				−0.358	1.316	0.92449
Pac	−0.741	1.358	6.30				−2.682	2.093	0.93822
Children									
All	−0.559	1.141	2.19				−1.144	1.325	0.96619
NewEng									
MidAtl	−1.415	1.411	0.25				−1.640	1.488	0.99660
SoAtl	−1.133	1.281	4.42						
ENoCent	−1.687	1.637	0.06				−1.684	1.636	1.00000
ESoCent	0.264	0.730	4.25				0.641	0.593	0.95066
WNoCent	−0.034	0.848	2.46				0.008	0.844	0.97792
WSoCent	−0.089	1.027	3.68				−0.441	1.164	0.96930
Mtn	−0.020	0.860	4.95				−0.173	0.932	0.90954
Pac	−0.597	1.171	1.21				−0.905	1.286	0.98745

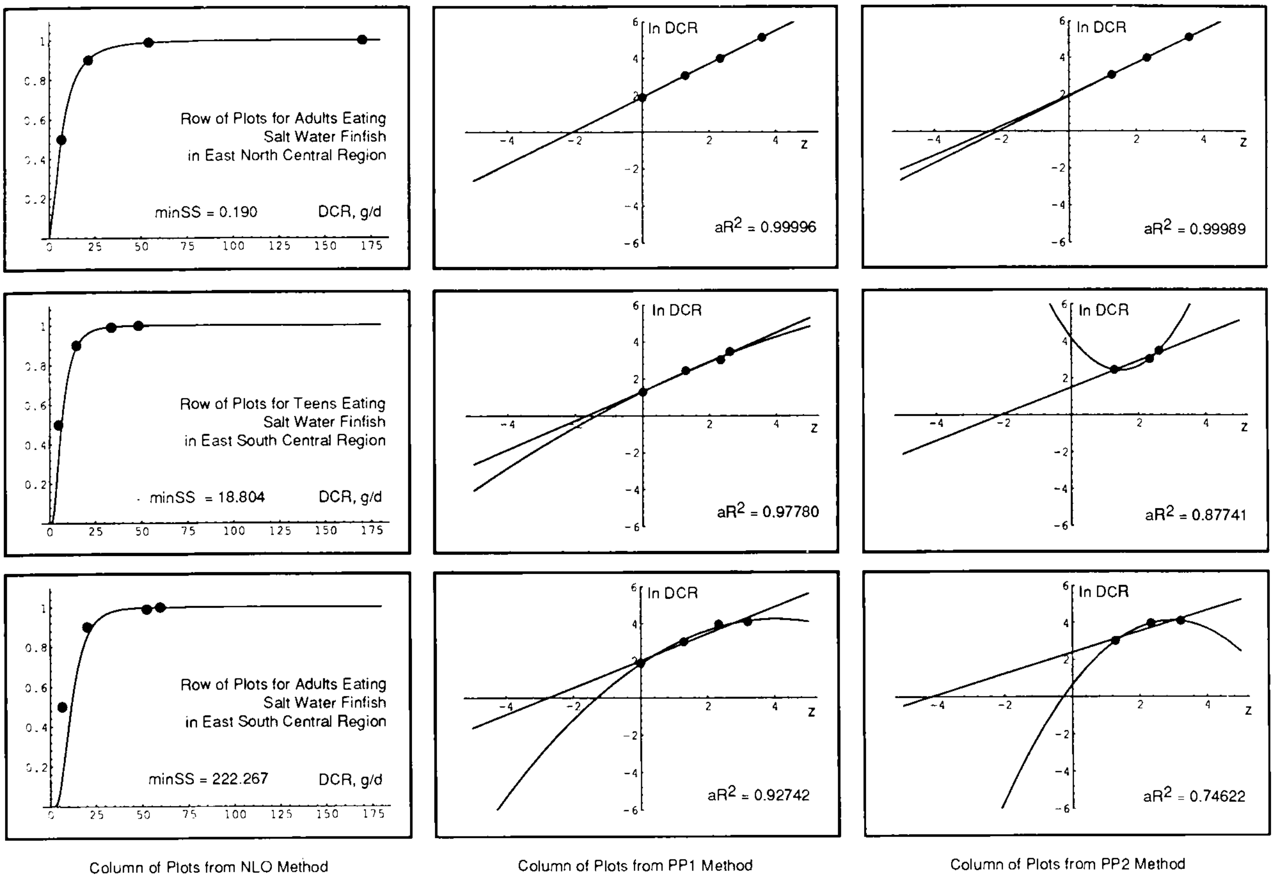

Fig. 1. Visual comparison of fits for three data sets by the NLO, PP1, and PP2 algorithms.

3.2. Comparison of Best-Fit Parameters

Using scatterplots, we found that the best-fit parameters from each method are highly correlated with each other. For each method alone, and for the three methods together, a strong and inverse relationship exists between the optimal value of μ and that of σ.

3.3. Comparison of Reported and Predicted DCR Values

Again, we used scatterplots to compare the reported and predicted DCR values for each of the three methods. Figure 2 visualizes the absolute and relative ability of the lognormal distributions fit by the NLO and PP2 methods to reproduce the DCR data reported by Rupp *et al.*[14] In Fig. 2, the plots in the top and bottom rows, respectively, compare the results from the NLO and the PP2 methods. In each plot, each point refers to one data set, the lighter line shows the lowess regression line fit to the points, and the darker line shows the 45° locus of perfect prediction. By interactively "brushing" data points on the computer screen in the Systat program,[15] we found that most of the more widely scattered points (relative to the locus of perfect prediction) arise from data sets with poor goodness-of-fit measures. After restricting the data sets, the four scatterplots in the top row in Fig. 2 show the points and lowess regression lines for 77 data sets with minSS < 30 as an empirical criterion, while the corresponding plots in the bottom row show the scatterplots and lowess regression lines for the 73 data sets with $aR^2 > 0.90$ as an empirical criterion. The selection criteria reduce the number of data sets by approximately the same number (NLO, $89 - 12 = 77$; PP2, $84 - 11 = 73$). With the restrictions in place, the NLO method predicts with greater accuracy and less bias than does the PP2 method for DCR50, DCR_{avg}, and

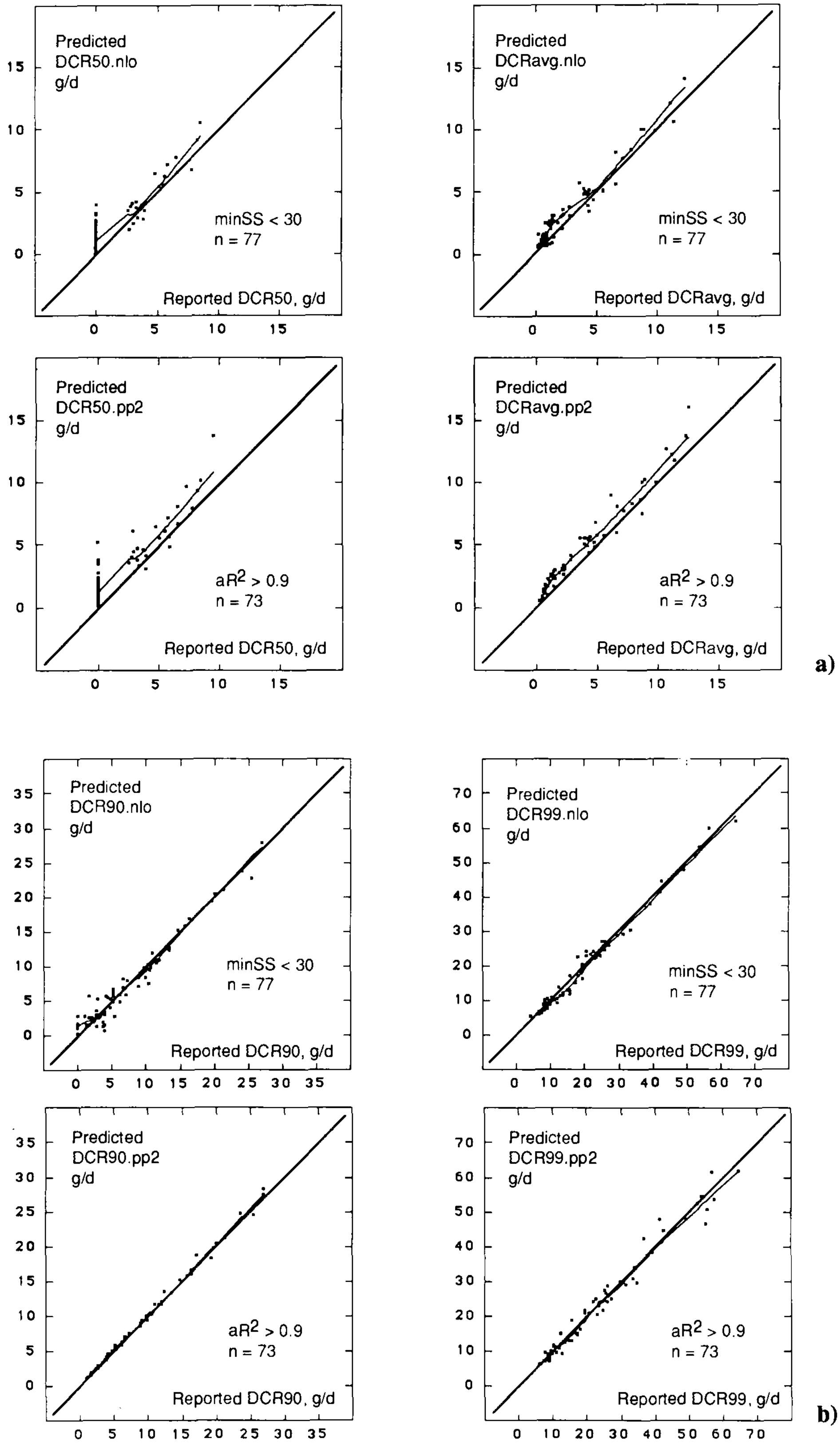

Fig. 2. Comparison of reported and predicted DCR values for selected data sets.

DCR_{max}. When applied to fewer data sets, the PP2 method outperforms the NLO method for predicting DCR90.

We conclude that the results for 77 data sets fit by the "full-information" NLO method with minSS < 30 are well suited for risk assessment (with the adjustment discussed below) that focuses on the diet of people in the general population. However, we also conclude that the results for the remaining 12 data sets fit by the NLO method may also be appropriate for use when exercised with due care and with sensitivity analyses.

4. OVERALL DISCUSSION

The lognormal distributions developed here are useful in the discipline of public health risk assessment both to estimate (i) point values for variables in "deterministic" risk assessments and (ii) distributions for "probabilistic" or Monte Carlo risk assessments involving the consumption of finfish and shellfish. Because they are more informative and inherently more representative, we recommend the use of distributions rather than single point estimates in public health risk assessments. This type of approach can readily be incorporated into the development of AWQCs and enables regulators to characterize the distributions of exposures and decide whether to protect to the 50th, 90th, 95th, or other percentile of the population at a particular allowable risk target. Further, complicated exposure scenarios for combinations of regions or combinations of fish categories can be simulated using several of the distributions in an appropriate model. Thus, the distributions developed here represent a toolkit for the analysis of many novel problems heretofore unaddressable by region, age, or type of fish.

We recognize limitations in this manuscript that arise from limitations in the original publication by Rupp *et al.*[14] First, the data from Rupp *et al.* pertain to the consumption of fish from all sources (i.e., purchased from stores or individuals, consumed in restaurants, received as gifts, and consumed by sport anglers) by all people in the general population; they do not apply (solely) to the consumption of fish caught in a particular stream, river, or estuary. Hence, they may not be appropriate for developing site-specific water quality standards where fishing patterns differ from those of the general U.S. population. For site-specific studies where public health concerns focus on the consumption patterns of sport anglers or ethnic groups with unusual diets, we recommend that the sponsors undertake new, site-specific surveys which distinguish among purchased, restaurant, gift, and self-caught fish. Second, Rupp *et al.* provide no information on correlations, if any, among the consumption rates of the three types of fish. It remains an open question, for example, whether a person who eats a larger or smaller than average amount of saltwater finfish eats a larger or smaller than average amount of freshwater finfish. Third, we draw no inferences on the uncertainty in the model specification or in the best-fit parameters. Fourth, fish consumption has increased since the NMFS survey was conducted in 1973 and 1974 by ~16 to ~27%. An increase in overall fish consumption >27% seems unlikely given the increasing prices in several desirable species due to declining harvests and continuing fish consumption advisories. Whether the increase in consumption applies to all types of fish and from all sources is unknown. To account for this, the location of the distributions for the consumption rates can be increased appropriately if the age of the NMFS survey is a concern to the reader.

Despite the overall increase in fish consumption by the general U.S. population, it is likely that the distributions are still lognormal in shape and that the tails (minima and maxima) are essentially the same. Those people who have always eaten very little fish probably still eat very little, and those who have always eaten large quantities probably still eat fish at the same high rates (as limited by caloric balance). While a small fraction of the U.S. population may consume fish at rates equal to the combined intake of red meat, poultry, and fish, there is a limit on the protein intake which will not have changed in the last 20 years. Therefore, the increase in overall fish consumption is likely to be reflected in the body of the distribution, which includes individuals who eat moderate amounts of fish. For these individuals, the peaks of the distributions are likely to have shifted to the right, resulting in medians ~25% higher than indicated by the distributions presented in this paper. The simple addition of $0.22 = \ln(1.25)$ to each of the μ values fit in this paper increases each percentile (and therefore each average) of each distribution by 25%, a conservative adjustment to account for the general increase in fish consumption in the United States since the time of the survey.

It should be noted that the U.S. EPA's average fish consumption rate of 6.5 g/day is consistent with Rupp *et al.*'s consumption rates for these types of fish by all U.S. consumers. The combined DCR_{avg} of shellfish and finfish of 5.0 g/day for all adults supports the 6.5 g/day rate historically used by the U.S. EPA in health risk assessments and in setting AWQCs, even with the conservative 25% adjustment factor.

While we acknowledge certain limitations, while we agree that site-specific surveys are often useful or necessary for estimating recreational or subsistence fishing rates, and while we agree that a new long-term (>12-month) nationwide survey that disaggregates fish consumption into categories of fish (i.e., salt vs fresh water, and finfish vs shellfish) would provide additional useful information, we believe that the results here are strong and unique in that no other such national survey now exists. Consequently, we believe that these results—with the adjustment just discussed—are useful now in practical risk assessments and in the setting of federal and state AWQCs. Until and unless a new and large national survey that disaggregates saltwater finfish, shellfish, and freshwater finfish is funded and conducted, the data of Rupp *et al.*[14] provide an excellent basis for Monte Carlo analyses.

ACKNOWLEDGMENTS

We thank William Gillespie at the National Council for Air and Stream Improvement for his support of earlier studies of fish consumption rates. We also thank Paul Scott Price of ChemRisk, Inc., and Andrew E. Smith of the Harvard School of Public Health for many helpful comments on an early version of the manuscript. We also thank two anonymous reviewers for suggestions incorporated in the final version. Alceon Corporation and ENSR Consulting and Engineering funded this research.

REFERENCES

1. M. Abramowitz and I. A. Stegun (eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* National Bureau of Standards, Applied Mathematics Series Number 55 (Superintendent of Documents, U.S. Government Printing Office, Washington, DC, June 1964).
2. P. D. Anderson, B. Ruffle, and W. Gillespie. "A Monte Carlo Analysis of Dioxin Risk from Consumption of Fish Caught in Fresh Waters of the US," Presented at TAPPI Environmental Conference, Richmond, VA, Apr. 12 to 15 (1992).
3. D. E. Burmaster and K. E. von Stackelberg. "Using Monte Carlo Simulations in Public Health Risk Assessments: Estimating and Presenting Full Distributions of Risk," *J. Expos. Anal. Environ. Epidemiol.* **1**(4), 491–512 (1991).
4. D. E. Burmaster and K. E. von Stackelberg. "A Discussion of the US EPA Methodology for Determining Water Quality Standards (WQS)," *Qual. Assur. Good Pract. Regulat. Law* **1**(3), 192–206 (1992).
5. J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis* (Wadsworth International Group, Belmont, CA, and Duxbury Press, Boston, MA, 1983).
6. ChemRisk, Inc. *Consumption of Fresh Water Fish by Maine Anglers* (in cooperation with HBRS, Inc., Portland, ME, July 31, 1991).
7. W. S. Cleveland. *Visualizing Data* (Hobart Press, Summit, NJ, 1993).
8. W. S. Cleveland. *The Elements of Graphing Data* (Wadsworth Advanced Books, Monterey, CA, 1985).
9. E. L. Crow and K. Shimizu. *Lognormal Distributions: Theory & Applications* (Marcel Dekker, New York, 1988).
10. R. B. D'Agostino and M. A. Stephens (eds.). *Goodness-of-Fit Techniques* (Marcel Dekker, New York, 1986).
11. M. Evans, N. A. J. Hastings, and J. B. Peacock. *Statistical Distributions: A Handbook for Students and Practitioners,* 2nd ed. (Wiley & Sons, New York, 1993).
12. K. M. Thompson and D. E. Burmaster. "Parametric Distributions for Soil Ingestion by Children," *Risk Anal.* **11,** 339–342.
13. A. M. Roseberry and D. E. Burmaster. "Lognormal Distributions for Water Intake by Children and Adults," *Risk Anal.* **12,** 99–104.
14. E. Rupp, F. Miller, and C. Baes. "Some Results of Recent Surveys of Fish and Shellfish Consumption by Age and Region of Some US Residents," *Health Phys.* **39,** 165–174.
15. Systat, Inc. *Systat, Sygraph, The System for Graphics,* Version 5.2 (Systat, Evanston, IL, 1992).
16. J. W. Tukey. *Exploratory Data Analysis* (Addison–Wesley, Reading, MA, 1977).
17. U.S. Department of Agriculture, Economic Research Service. *Annual Statistics on Per Capita Fish Consumption (Edible Weight) in the US* (World Almanac Book of Facts, 1980–1989).
18. U.S. Department of Agriculture. *Agricultural Statistics—1986* (Economic Research Service, Washington, DC, 1986).
19. U.S. Department of Commerce. *Statistical Abstract of the United States—1992* (Washington, DC, 1992).
20. U.S. Environmental Protection Agency. *Ambient Water Quality Criteria (AWQC) for 2,3,7,8-Tetrachlorodibenzo(p)dioxin,* EPA 440/5-84-007 (Environmental Criteria and Assessment Office, Cincinnati, OH, 1984).
21. U.S. Environmental Protection Agency. Draft Technical Support Document for Water Quality-Based Toxics Control (Office of Water, Washington, DC, April 1990).
22. S. Wolfram. *Mathematica, a system for Doing Mathematics by Computer,* 2nd ed. (Addison-Wesley, Redwood City, CA, 1991).