FC
USGS
OFR
78-1003

GL00165

UNIVERSITY OF UTAH
RESEARCH INSTITUTE
EARTH SCIENCE LAB.

Methodology of determining the uncertainty in the accessible geothermal resource base of identified hydrothermal convection systems

Manuel Nathenson

U. S. Geological Survey

Menlo Park, California 94025

December, 1978

U. S. Geological Survey

OPEN-FILE REPORT 78-1003

This report is preliminary and has not been edited or reviewed for conformity with Geological Survey standards. Any use of trade names and trademarks in this publication is for descriptive purposes only and does not constitute endorsement by the U. S. Geological Survey.

Abstract

In order to quantify the uncertainty of estimates of the geothermal resource base in identified hydrothermal convection systems, a methodology is presented for combining estimates with uncertainties for temperature, area, and thickness of a geothermal reservoir into an estimate of the stored energy with uncertainty. Probability density functions for temperature, area, and thickness are assumed to be triangular in form. In order to calculate the probability distribution function for the stored energy in a single system or in many systems, a computer program for aggregating the input distribution functions using the Monte-Carlo method has been developed.

To calculate the probability distribution of stored energy in a single system, an analytical expression is also obtained that is useful for calibrating the Monte Carlo approximation. For the probability distributions of stored energy in a single and in many systems, the central limit approximation is shown to give results ranging from good to poor.

Introduction

The purpose of this paper is to document a methodology for providing measures of the uncertainty in estimates of the accessible geothermal resource base in identified hydrothermal convection systems. One could simply make a judgmental statement concerning this quantity (i.e. the geothermal resource base is x cal with a standard deviation of $\pm$ y cal). The basis for making such a statement is very weak, because one has very little intuitive feel for the measurement of stored energy. Just as the estimate of identified accessible geothermal resource base is built up by estimating the temperature, area, and thickness of a system, calculating the stored energy, and summing the values for the various systems, one can make estimates of the uncertainty of each of these primitive quantities, and develop a methodology for aggregating these estimates into values for derived quantities. The uncertainties of temperature, areal extent, and thickness are relatively easy to estimate and vary considerably from system to system depending upon the amount of available data.

The starting point of the calculation is to assume that the probability density functions for temperature, area, and thickness for a hydrothermal system are triangular. The parameters of each density have been estimated using the available geologic, geochemical, and geophysical data. These parameters are then used to calculate means and standard deviations for temperature, area, and thickness. Two types of derived quantities are then calculated from the parameters of the density functions. The first type is the mean and standard deviation of the stored

energy in a single system and the mean and standard deviation of a sum of the stored energy of many systems. Formulas for these quantities are easily developed. The second type of derived quantity is the probability density and distribution functions for stored energy in a single system and for a sum of many systems. These distribution functions are necessary to obtain confidence limits. The calculation for a single system can be done analytically but the calculation for many systems is analytically intractable. Instead, the probability distribution for stored energy in many systems is calculated using a Monte-Carlo simulation. Similar calculations for petroleum resources have been done by Miller and others (1975) and White and others (1975).

The type of statistical analysis that is discussed is not the only kind that is applicable to these data. La Mori (1976) analyzed the data in Circular 726 (Renner and others, 1975) for the relationship between number of systems and grade (temperature) in the usual mineral resource context in order to try and relate the sample of identified system to the population of all systems. In unpublished work, he also looked at number versus size relations for the same data set. McNitt (1977) has also made a study of number versus size but for producing fields and fields defined by drilling in different areas of the world to check the log-normal properties of the relationship between number of systems and size. No such analysis will be attempted in this study.

It should be pointed out that a statistician will find nothing new in the results of this study. They are all essentially textbook results put together in a particular combination. The casual reader would do

better to read the summary in U. S. Geological Survey Circular 790 (Brook and others, 1979) for the highlights. The point of this paper is to satisfy the reader who wants to know how a particular calculation is done. The basic notions of probability have been freely borrowed from Papoulis (1965) and Feller (1950) and only particular results or statements that are not easily found have been referenced.

Random variables; distribution functions; density functions; mean; standard deviation

In developing these notions of the behavior of a single random variable, the area of a system is used as an example. The results apply equally to temperature, thickness, stored energy, etc. The first point is the definition of a random variable. A random variable functionally relates an outcome to a numeric value. For example, in dealing with outcomes such as whether the result of a coin toss is heads or tails, the random variable "number of heads" might be assigned the value one when a head appears and zero when a tail appears. For the outcome of measuring the area of a geothermal system, the random variable (denoted A) is simply the value of the area. Random variables will be symbolized by using capital letters. An event $A \leq a$ is the set of values A such that they are all less than or equal to a. The probability of the event $A \leq a$ is denoted by $P(A \leq a)$ and is a number between 0 (the event never occurs) and 1 (it is certain to occur). The distribution function for the random variable A is $F_A(a)$ and gives the values of the probability $P(A \leq a)$, i. e.

$$P(A \leq a) = F_A(a). \qquad (1)$$

The distribution function is usually defined for a between plus and minus infinity. All of our random variables take only positive values, so we will restrict a to be nonnegative. We will usually be interested in the event $A \geq a$, and the probability of this event $P(A \geq a)$ is then $1 - F_A(a)$. The density (function) is defined as

$$f_A(a) = \frac{dF_A(a)}{da} \tag{2}$$

and may be used to calculate the distribution function from

$$F_A(a) = \int_o^a f_A(x)\, dx \tag{3}$$

where x is a dummy variable for the definite integral. The density and distribution are both useful in that the distribution gives actual probabilities while the density is used for most calculations. The quantity $f_A(a)da$ may be interpreted as the probability that A is between the values a and a+da, i. e.

$$f_A(a)da = P\{a \leq A \leq a + da\}. \tag{4}$$

Figure 1 shows the triangular density and its associated distribution for the estimate of the area of the geothermal reservoir at Heber, California (Renner, 1976).

A density may also be described by certain of its moments. The two most commonly used are mean (expected value) and standard deviation. The mean is defined as

$$\langle A \rangle = \int_o^\infty a\, f_A(a)da \tag{5}$$

and the variance as

$$\mathrm{Var}(A) = \langle (A - \langle A \rangle)^2 \rangle = \int_o^\infty (a - \langle A \rangle)^2 f_A(a)da. \tag{6}$$

The standard deviation σ_A is simply the square root of the variance. Expanding the squared term on the right hand side of equation (6), we obtain the useful relation

$$\begin{aligned} \langle (A - \langle A \rangle)^2 \rangle &= \int_o^\infty (a^2 - 2a\langle A \rangle + \langle A \rangle^2) f_A(a)da \\ &= \langle A^2 \rangle - 2\langle A \rangle\langle A \rangle + \langle A \rangle^2 \\ &= \langle A^2 \rangle - \langle A \rangle^2, \end{aligned} \tag{7}$$

i. e., the variance can be calculated from the second moment and the square of the mean. On figure 1 the mean and the mean plus or minus one standard deviation are shown. The mean gives us a measure of where most of the density function is concentrated while the standard deviation gives a measure of the spread of the density. Other measures of where most of the density function is concentrated are also used. The median a_m is the value of a such that $P\{A \leq a_m\} = 1/2$ and $P\{A \geq a_m\} = 1/2$. The mode or most likely value a_{m1} is the value of a such that $f_A(a)$ is at its maximum. For a symmetrical density, the mean, median, and mode are all the same. The density function shown in figure 1 is not symmetrical so the three quantities all have different values.

Example: A triangular density

Since one of our primary densities is the triangular density, it is worthwhile to go through and calculate the various quantities defined above in terms of the parameters of the density function. Triangular densities are a three parameter family. As shown on figure 1, the three parameters that we will use are a_1, a_2, a_3. The quantity a_1 is the minimum value--the value for which the probability that $A \geq a_1$ is unity. The quantity a_2 is the most likely value (mode)--the value for which the probability $P(a_2 \leq A \leq a_2 + da)$ is a maximum. The quantity a_3 is a maximum value; the value for which the probability that $A \geq a_3$ is zero. Introducing the unit step function

$$\alpha(x) = \begin{cases} 0 & x<0 \\ 1 & x \geq 0 \end{cases} \tag{8}$$

we may write out the expression for the triangular density as

$$f_A(a) = 2\left\{\frac{(a - a_1)\alpha(a - a_1)}{(a_3 - a_1)(a_2 - a_1)} + \frac{(a - a_2)\alpha(a - a_2)}{(a_2 - a_3)(a_2 - a_1)} + \frac{(a - a_3)\alpha(a - a_3)}{(a_3 - a_1)(a_3 - a_2)}\right\}. \quad (9)$$

The convenience of this form is that the expression is valid for a between zero and infinity. The associated distribution function is obtained by integration

$$F_A(a) = \frac{(a - a_1)^2\alpha(a - a_1)}{(a_3 - a_1)(a_2 - a_1)} + \frac{(a - a_2)^2\alpha(a - a_2)}{(a_2 - a_3)(a_2 - a_1)} + \frac{(a - a_3)^2\alpha(a - a_3)}{(a_3 - a_1)(a_3 - a_2)}. \quad (10)$$

The modal value of $f_A(a)$ is given by equation (9) with $a = a_2$, and we obtain

$$f_A(a_2) = \frac{2}{a_3 - a_1}. \quad (11)$$

The mean and variance are obtained by appropriate integrals of equation (9). With some algebraic rearrangement, we may write these as

$$\langle A \rangle = (a_1 + a_2 + a_3) / 3 \quad (12a)$$

$$\mathrm{Var}(A) = \langle A^2 \rangle - \langle A \rangle^2$$

$$= [(a_2 - a_1)^2 + (a_3 - a_2)^2 + (a_2 - a_1)(a_3 - a_2)] / 18. \quad (12b)$$

For easy reference, these results and results for the normal and log-normal density functions are given in appendix 1.

Equation (12a) for the mean shows that the minimum and maximum values carry equal weight with the most likely value in determining

the position of the mean. The mean may thus be quite different than the most likely value. Figure 2 shows a highly skewed type of density as a solid line. The minimum, most likely, and maximum values have been taken from Renner (1976). The relationship between the strong peak and the long tail was chosen so as to maximize the graphical disparity between the two approximations to be discussed and does not necessarily reflect the true density function for the estimate of the area of the hydrothermal system at Coso, California. If we try to model this with a triangular density, the choice of whether to use the density shown as a broken line in figures 2a or 2b as the approximation has a dramatic effect on the mean. We are more interested in the mean than in the regions with extremely low values of probability density, and therefore we should use the broken line in figure 2b. Since the long positive tail of the true density is rapidly lost in the aggregation of comparable random variables, this is a satisfactory approximation. It is also important to realize that not all of the density functions have a long positive tail. If all had this tail, there would be a systematic bias in the use of a triangular density and some other density would have to be used.

Products and sums of random variables: mean and variance

There are three functions of random variables for which we will be interested in finding the mean and variance. The first is the volume v for a single system of area a and thickness δ:

$$v = a\,\delta \tag{13}$$

The second is the stored energy q for a single system of area a, thickness δ, and temperature t:

$$q = \rho c(t - t_o) \, a \, \delta \tag{14}$$

where t_o is a reference mean annual surface temperature taken as 15°C and the product ρc is the volumetric specific heat of the rock plus water system (here taken 0.6 cal/cm^{3}C). The third function is the sum of the stored energy sq for n systems:

$$sq = \sum_{i=1}^{n} q_i \tag{15}$$

where a, δ, t, v, and q in equations (13) and (14) should all be subscripted as being for i th system, but for clarity we will delete the subscripts except where necessary.

The mean of v is by definition

$$\langle V \rangle = \int_o^\infty v f_V(v) dv. \tag{16}$$

A fundamental theorem in statistics is that this may be calculated from

$$\langle V \rangle = \int_o^\infty \int_o^\infty a\delta f_V(a, \delta) \, da \, d\delta \tag{17}$$

where $f_V(a, \delta)$ is the joint density of random variables A and Δ (Papoulis, 1965, p. 206). The next paragraph gives the reasons for assuming that T, A, and Δ are statistically independent random variables in order to allow a significant simplification of equation (17).

The process of estimating area, thickness, and temperature generally involves the use of different indicators for each quantity. Temperature is estimated from geothermometry, hot spring temperatures and mixing models, or it is known from well data. Areas are estimated from hot spring occurrances, temperature gradients in shallow wells, surface geophysical measurements of gravity, resistivity patterns, magnetics, topography, etc. Thicknesses are constrained below by a maximum of three kilometers (normal drillable depth), above by temperature gradient

data, and are estimated by geologic analogy. These processes of estimating area, thickness, and temperature for identified geothermal systems are essentially independent. Some common denominators are temperature gradient holes and resistivity soundings for estimates of thickness and area. They are used in fundamentally different ways for estimating the two quantities, however. For areal extent, one is interested in the difference in gradient (or resistivity) from hole to hole (sounding to sounding) to define the anomaly while for thickness one looks at a particular hole (sounding) to estimate at what depth the reservoir temperature is likely to occur. One is thus led to the assumption that area, thickness, and temperature are statistically independent random variables for which that the joint density in equation (17) may be written as

$$f_V(a, \delta) = f_A(a) f_\Delta(\delta). \tag{18}$$

and $\langle V \rangle$ may then be written as

$$\begin{aligned} \langle V \rangle &= \{\int_0^\infty a\, f_A(a)da\}\{\int_0^\infty \delta f_\Delta(\delta)d\delta\} \\ &= \langle A \rangle \langle \Delta \rangle \end{aligned} \tag{19}$$

Similarly, the second moment is

$$\begin{aligned} \langle V^2 \rangle &= \{\int_0^\infty a^2 f_A(a)da\}\{\int_0^\infty \delta^2 f_\Delta(\delta)d\delta\} \\ &= \langle A^2 \rangle \langle \delta^2 \rangle \end{aligned} \tag{20}$$

and we may write the variance as

$$\begin{aligned} \mathrm{var}(V) &= \langle V^2 \rangle - \langle V \rangle^2 \\ &= \langle A^2 \rangle \langle \Delta^2 \rangle - \langle A \rangle^2 \langle \Delta \rangle^2 \\ &= [\mathrm{Var}(A) + \langle A \rangle^2][\mathrm{Var}(\Delta) + \langle \Delta \rangle^2] - \langle A \rangle^2 \langle \Delta \rangle^2 \end{aligned} \tag{21}$$

Similar relations can be developed for stored thermal energy, and they

are given in Appendix 1.

A possible point of confusion concerning the assumption of statistical independence is the relationship of estimating the statistical properties of a single system to the underlying population of geothermal systems with perfectly measured properties. There is a frequency distribution

$$n\,(v_i,\ t_i) \tag{22}$$

for the number of systems having a volume with the value v_i to $v_i + \Delta v_i$ in the temperature range t_i to $t_i + \Delta t_i$. This frequency distribution is most easily conceptualized in terms of discrete random variables. A sample estimate of this frequency distribution could be provided by taking the mean values of the volume V_i and temperature T_i for each system in the 1978 data set (Brook and others, 1979), making a grid of cells of size T_i to $T_i + dT_i$ and V_i to $V_i + dV_i$, and assigning each system to a cell. For this frequency distribution, temperature and size are not statistically independent random variables. In one case we are dealing with uncertainties of a measuring process while the other looks at the characteristics of perfectly measured quantities.

The mean and variance of SQ may be obtained similarly, i. e.

$$\langle SQ\rangle = \left\langle \sum_{i=1}^{n} Q_i \right\rangle$$

$$= \int_0^\infty \cdots \int_0^\infty \left(\sum_{i=1}^{n} q_i\right) f_{SQ}(q_1, \ldots, q_n)\, dq_1 \ldots dq_n$$

$$= \int_0^\infty \cdots \int_0^\infty \left(\sum_{i=1}^{n} q_i\right) f_{Q_i}(q_1) \ldots f_{Q_n}(q_n)\, dq_1 \ldots dq_n$$

$$= \sum_{i=1}^{n} \int_0^\infty \cdots \int_0^\infty q_i f_{Q_1}(q_1) \cdots f_{Q_n}(q_n) dq_1 \cdots dq_n$$

$$= \sum_{i=1}^{n} \int_0^\infty q_i f_{Q_i}(q_i) dq_i$$

$$= \sum_{i=1}^{n} \langle Q_i \rangle \tag{23}$$

$$\langle (SQ - \langle SQ \rangle)^2 \rangle = \sum_{i=1}^{n} (\langle Q_i^2 \rangle - \langle Q_i \rangle^2) \tag{24a}$$

$$\text{var}(SQ) = \sum_{i=1}^{n} \text{var}(Q_i) \tag{24b}$$

where we have again assumed that the Q_i are statistically independent random variables in going from line 2 to line 3 above and in obtaining equations (24).

Products and sums of random variables: distributions and density functions

In the previous section, relations for the mean and variance of products and sums of random variables were obtained without finding the density or distribution of the derived random variables. In this section, we will present a formal solution for the resultant density functions. In special cases, the calculation can actually be done and we will present results for the products of random variables, each having a triangular density.

The random variable $V = A\,\Delta$ has a distribution function $F_V(v) = P\{V \leq v\}$. To determine $F_V(v)$, we must find the probability of the event $\{V \leq v\}$. In the a δ plane, the event $\{V \leq v\}$ corresponds to $a\delta \leq v$

which defines a region in the aδ plane shown cross-hatched in figure 3. We will call this region the domain D_V, then we have

$$P\{V \leq v\} = P\{(A, \Delta] \in D_V\} \tag{25}$$

where the second grouping is read as the probability that A and Δ are in the domain D_V. This later probability is given by

$$F_V(v) = P\{(A, \Delta) \in D_V\} = \int_{D_V} \int f_{A\Delta}(a, \delta)\,da\,d\delta. \tag{26}$$

For statistically independent random variables, we have

$$F_V(v) = \int_{D_V} \int f_A(a) f_\Delta(\delta)\,da\,d\delta \tag{27}$$

and the density can be obtained by differentiating the result. For our particular case, the integral may be written over the domain D_V as

$$F_V(v) = \int_o^\infty \int_o^{v/\delta} f_A(a) f_\Delta(\delta)\,da\,d\delta \tag{28}$$

where the diagram in figure 3 shows that we first integrate a strip of width dδ from the axis a=o to the curve v/δ and then integrate all of these strips from δ=0 to δ=∞ (see Papoulis, 1965, p. 189, for a similar manipulation). Differentiating the integral with respect to v and using Leibnitz Rule, we obtain

$$f_V(v) = \int_o^\infty \frac{1}{\delta} f_A\left(\frac{v}{\delta}\right) f_\Delta(\delta)\,d\delta. \tag{29}$$

By identical reasoning, the density for stored energy in a single system is

$$f_Q(q) = \int_o^\infty \frac{1}{\rho c(t-t_o)}\, f_T(t) f_V\left(\frac{q}{\rho c(t-t_o)}\right) dt. \tag{30}$$

For the sum of random variables, it is easiest to solve for the resultant density in steps. Defining a variable $sq' = q_1 + q_2$, similar reasoning gives the result

$$f_{SQ'}(sq') = \int_0^{sq'} fq_1(sq' - q_2)fq_2(q_2)dq_2. \tag{31}$$

Defining a variable $sq'' = q_1 + q_2 + q_3$, we obtain by similar reasoning

$$f_{SQ''}(sq'') = \int_0^{sq''} \int_0^{sq''-q_3} fq_1(sq''-q_3-q_2)fq_2(q_2)fq_3(q_3)dq_2dq_3 \tag{32}$$

and ad infinitum as the number of variables increases to more than a hundred. The form of equations (31) and (32) is what is known as a convolution integral, and for some simple density functions, the use of transforms would permit the evaluation of the integral in closed form.

Equations (29) and (30) are the formal solution for the density function of stored energy in a single system and equations (31) and (32) are the first two of a series of equations that constitute the formal solution of obtaining the density function of stored energy in many systems from the density functions of stored energy in each of the single systems. Assuming triangular densities for temperature, area, and thickness, it is practical to use these formal solutions only as far as calculating the probability density and distribution of the volume and stored energy in a single system. To obtain the density function of volume, we need the density function of area given in equation (9) and one for thickness that has the same form with just the symbols changed. Substituting into equation (29), the first term is (there are eight more)

$$f_V(v) = \int_{o\delta}^{\infty 4} \frac{(\frac{v}{\delta} - a_1)(\delta - \delta_1)\alpha(\frac{v}{a_1} - \delta)\alpha(\delta - \delta_1)}{(\delta_3-\delta_1)(\delta_2 - \delta_1)(a_3 - a_1)(a_2 - a_1)}d\delta + \ldots \tag{33}$$

The two step functions restrict the effective domain of integration to values of δ that are between δ_1 and v/a_1. Consequently, the first term may be written as

$$f_V(v) = \alpha(v - a_1\delta_1)\int_{\delta_1}^{v/a_1} \frac{4(\frac{v}{\delta} - a_1)(1 - \frac{\delta}{\delta_1})d\delta + \ldots}{(\delta_3 - \delta_1)(\delta_2 - \delta_1)(\delta_3 - \delta_1)(\delta_2 - \delta_1)}$$

$$= \frac{4[2(a_1\delta_1 - v) + (a_1\delta_1 + v)\ln \frac{v}{a_1\delta_1}]\alpha(v-a_1\delta_1)}{(\delta_3 - \delta_1)(\delta_2 - \delta_1)(\delta_3 - a_1)(a_2 - a_1)} + \ldots \qquad (34)$$

and the other eight terms can be written by a rotation of subscripts. The complete result for volume and the result for stored thermal energy are given in Appendix 1. A computer program for evaluating the density and distribution function for stored thermal energy in a single system is given in Appendix 2.

Central limit theorem for products and sums of random variables

As a large number of random variables (satisfying certain technical conditions) are summed to form a resultant random variable, the form of the individual distribution functions tends to become swamped by the convolution process and the resultant distribution tends toward a normal with a mean given by the sum of the means and a variance given by the sum of the variances. The random variables added to obtain the stored energy of the various systems are likely to satisfy these conditions (see Parzen, 1960, pp. 430-432 for a precise statement), though one must

be careful of the attribute that most of the variance is contributed by a relatively small number of large systems. The central limit approximation is useful in either: 1) providing an approximate answer to converting standard deviations to confidence intervals, or 2) providing a semi-quantitative check of Monte Carlo results. The normal density has the unfortunate property of being defined from minus infinity to plus infinity. Thus it will yield a small but finite probability that the aggregate of stored energy will be less than zero. In practice, this turns out to not really be a problem. The probability that the stored energy is greater than zero is usually very close to one (examples given in discussion section).

An easy extension of the central limit theorem may be made for the stored energy in a single system as related to the temperature, area, and thickness by equation (14) by defining

$$\gamma = \ln q \tag{35a}$$

$$\beta_1 = \ln (t - t_o) \tag{35b}$$

$$\beta_2 = \ln (a) \tag{35c}$$

$$\beta_3 = \ln (\delta) \tag{35d}$$

so that we obtain the relation

$$\gamma = \beta_1 + \beta_2 + \beta_3 + \ln (\rho c). \tag{36}$$

We have thus transformed a product to a sum of random variables. Assuming that the central limit theorem holds, the variable γ should be distributed normally. This is exactly the definition of a log-normal distribution for the variable q--that its logarithm is distributed normally. In practice, it turns out that the log-normal approximation is not very good. As will be shown below, a normal distribution frequently

represents the distribution of q for a single system more accurately than a log-normal.

Monte-Carlo methods for aggregating random variables

For any but the simplest density function, the manipulations required to calculate a density function from eqs. (31) and (32) are very difficult and sometimes not analytically possible. One could write computer programs to perform the integrations. However, a minor change in the relationship between random variables might require a whole new program. A technique that has been developed that solves the problem in a fundamentally different manner is the Monte Carlo method (see Mize and Cox, 1968, for an introduction). The basic idea is to generate a random outcome (e.g., toss a coin), use this random outcome as an input to a model, and to record the output from the model. This process is usually repeated many times. In order to do a good simulation, the randomness of the outcomes must be guaranteed (e. g., the coin must be a very fair coin) or the results will be biased.

In order to describe the technique it is easiest to go through the formal solution of an example. Let q_1 and q_2 be the stored energy in two systems with distributions $F_1(q_1)$ and $F_2(q_2)$. We want to find the resultant distribution $F(q)$ where $q = q_1 + q_2$. The first step is to generate a set of random numbers in the interval 0 to 1, say 800 numbers. We choose two random numbers u_1 and u_2 and find the values of q_1 and q_2 such that

$$q_1 = F_1^{-1}(u_1) \tag{37a}$$

$$q_2 = F_2^{-1}(u_2) \tag{37b}$$

where F^{-1} denotes an inverse, i. e., the value of stored heat q_1 that

has a probability u_1 that the random variable $Q_1 \leq q_1$. We then sum q_1 and q_2 to obtain a value q. We have now performed one trial. This process is repeated four hundred times and yields a table of q values that can be sorted into increasing order. Define F*(x)=(no. q values $\leq$ x)/(no. of trials). It is well known that F*(q) is a good estimate of F(q) and that the accuracy of this estimate increases as the sample size increases. F*(q) is known as the sample distribution function. Clearly, the computations involved are quite simple. Different choices of distributions and relations among random variables are easy to incorporate into fairly short computer programs. The choice of distributions that can be inverted analytically results in significant computational efficiencies. For the triangular distribution of equation (10), the inversion is

$$a(F_A) = \begin{cases} a_1 + ((a_3 - a_1)(a_2 - a_1)F_A)^{1/2} & 0 \leq F_A \leq \dfrac{(a_2 - a_1)}{(a_3 - a_1)} \\ a_3 - ((a_3 - a_1)(a_3 - a_2)[1 - F_A])^{1/2} & \dfrac{(a_2 - a_1)}{(a_3 - a_1)} < F_A \leq 1. \end{cases} \tag{38}$$

Harold Javitz of SRI International has written such a program for this study and a listing is given in Appendix 3. Figure 4 shows for 400 trials the degree to which the sample density function and sample distribution function reproduced by this program agree with the exact triangular density and the resulting distribution function. Many more trials are required to reproduce the density function than the distribution function, and we will only use the Monte-Carlo program for producing sample distribution functions. The mean and standard deviation calculated from the Monte-Carlo simulation are 2.003 and 0.408, and

these are within 0.2% of the true values of 2 and 0.408. The analytical solution for a product of three random variables will be used to make further comparisons below.

Discussion and example using 1975 data set

As part of the effort to estimate the resource base in hydrothermal convection systems in Renner and others (1975), a file was set up and estimates made (Renner, 1976) to perform a statistical analysis of the stored energy. Minimum, maximum, and best estimates for temperature, area, and thickness, and other quantities were derived. The lack of a methodology and the press of time prevented this aspect of the study from being carried to completion. We can now use some of this data set to provide examples for the methodology developed here.

There are two inadequacies in the 1975 data set caused by the lack of a methodology that are important to remember in what follows. The first inadequacy is that there was no real assumption as to the form of the density function. Having now assumed a triangular density, the problem mentioned above of dealing with long tails becomes important. The data for Coso, CA in figure 2 exemplify this. The maximum area was estimated from the arcuate faults found by Duffield in preliminary mapping and it was quite speculative that the system might be that large (Duffield, 1978, pers. commun.). Thus the maximum should have been lowered in order to deal with the limitations of a triangular density. This example is an easily documented case of a problem that probably occurs in other places in the data set. The second inadequacy is that

the best values reported in Renner (1976) are not necessarily the most likely values (Renner, pers. commun., 1978). In some cases they are the most likely, but in others they are actually the mean values. This distinction will be ignored, and the best value will be treated as the most likely. For these reasons, we caution the reader that the values given in this report are not to be regarded as a revision to Circular 726.

Figures 5, 6, and 7 show the density functions for temperature, area, thickness, and stored energy for the three systems at Heber and Long Valley, CA and Valles Caldera, N. M. These three examples were chosen because of their very different forms of the density function for area. Heber is relatively symmetric, Long Valley has a long tail toward zero, and Valles Caldera has a long tail towards infinity. In all three cases, the density function for stored energy is smooth and possesses a distinct tail. Taking a product of random variables tends to produce an asymmetric density with a longer tail towards infinity. Thus the rather sharp break for the area of Long Valley becomes less sharp in the stored energy, but the long tail towards infinity in the area of Valles Caldera is preserved. The distribution for stored energy in each of the single systems are also shown on figures 5, 6, and 7 along with Monte-Carlo derived sample distribution functions. The Monte-Carlo simulations are generally very good representations of the analytical solutions, and this good agreement gives confidence in using the Monte-Carlo aggregation for summing the individual systems to produce an aggregate distribution function.

Figures 8, 9, and 10 show the density and distribution functions for stored energy for each system along with the approximations based on using normal and log-normal densities and distribution functions. The normal and log-normal densities have the same mean and standard deviations as the true density. Other matching schemes are possible, but this preserves the important statistical measures that can be manipulated without knowing the form of the density function. For Heber and Long Valley, the normal is a better approximation to the density than the log-normal and is essentially indistinguishable from the true when comparing distribution functions. For Valles Caldera, the log-normal is a very good approximation to the true density. The normal is a reasonable approximation to the distribution function, even though it asserts a probability of 0.03 that the stored energy is less than zero.

Figure 11 shows a comparison between a Monte-Carlo derived sample distribution function of the total stored energy in the three systems and distribution function approximations based on a normal and log-normal density using the true mean and standard deviation. Since the process of summing random variables tends to result in a normal density, one might expect a normal approximation to be a good one. That some of the densities of stored energy for a single system are approximated well by a normal density also tends to help the tendency towards normality for the sum of random variables. Both the normal and the log-normal seem to be reasonable representations of the Monte-Carlo simulation. Because of the large variability of the shape of the input distributions as parameters are varied, it is not possible to make any strong a priori

statement about how well either distribution will fit for the sum of a large number of systems. However, the Monte-Carlo produced sample distribution function will converge to the true distribution function as the number of trials increases.

Application to U. S. Geological Survey Circular 790 (Brook and others, 1979)

This report has discussed a number of aspects of the problem of finding resultant statistics for products and sums of random variables. For the assessment of geothermal resources in hydrothermal convection systems (Brook and others, 1979), only some of this material is important. The path that is taken in that document is to estimate the minimum, most likely, and maximum values for temperature, area, and thickness. Assuming a triangular density function, these values are then used to obtain a mean and standard deviation for each quantity. The relations for products of random variables are then used to convert these values into means and standard deviations for reservoir volume and thermal energy for individual systems. The relations for sums of random variables are then used to obtain means and standard deviations for various groupings of systems, e. g. all hot-water systems >150°C, hot-water systems >150°C plus vapor-dominated systems, etc. In each case, the standard deviation is used to provide a rough measure of the uncertainty in each quantity. The minimum, most likely, and maximum values of temperature, area, and thickness are then used as input to Monte Carlo computer programs to calculate sample distributions for stored energy in various groupings of the systems.

References cited

Brook, C. A., Mariner, R. H., Mabey, D. R., Swanson, James, Guffanti, Marianne, and Muffler, L. J. P., 1979, Hydrothermal convection systems with reservoir temperatures $\geq$ 90°C, in Muffler, L. J. P., ed., Assessment of Geothermal Resources of the United States--1978: U. S. Geological Survey Circular 790, in press.

Feller, William, 1950, An introduction to probability theory and its application, Volume 1 (2d. ed.): New York, John Wiley, 461 p.

La Mori, P. N., 1976, Hydrothermal geothermal resources and growth in utilization, in Hill, R. F., ed., Energy Technology III; Commercialization: Proceedings of the Energy Technology Conference, 3d., Washington, Government Institutes, Inc., p. 103-114.

McNitt, J. R., 1977, The United Nations' approach to geothermal resources assessment, in Larderello Workshop on Geothermal Resource Assessment and Reservoir Engineering, Larderello, Italy, 1977, Proceedings: Studi e Ricerche, Roma, Ente Nasionale per l'Energia Elettrica, p. 327-350.

Miller, B. M., Thomsen, H. L., Dolton, G. L., Coury, A. B., Hendricks, T. A., Lennartz, F. E., Powers, R. B., Sable, E. G., and Varnes, K. L., 1975, Geological estimates of undiscovered recoverable oil and gas resources in the United States: U. S. Geological Survey Circular 725, 78 p.

Mize, J. H., and Cox, J. G., 1968, Essentials of simulation: Englewood Cliffs, New Jersey, Prentice-Hall, 234 p.

Papoulis, Athanasios, 1965, Probability, random variables, and stochastic processes: New York, McGraw-Hill, 583 p.

Parzen, Emanuel, 1960, Modern probability theory and its applications: New York, John Wiley, 464 p.

Renner, J. L., 1976, Selected geothermal resource data: Hydrothermal convection systems in the states of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming: U. S. Geological Survey, USGS-CD-76-001, Springfield, Va. National Technical Information Service, PB-250377/9WE, 358 p.

Renner, J. L., White, D. E., and Williams, D. L., 1975, Hydrothermal convection systems, in White, D. E. and Williams, D. L., eds., Assessment of Geothermal Resources of the United States—1975: U. S. Geological Survey Circular 726, p. 5-57.

White, D. A., Garrett, R. W., Jr., Marsh, G. R., Baker, R. A., and Gehman, H. M., 1975, Assessing regional oil and gas potential, in Haun, J. D., ed., Methods of Estimating the Volume of Undiscovered Oil and Gas Resources: American Association of Petroleum Geologists Studies in Geology No. 1, p. 143-159.

Appendix 1 Useful formulas

For a single system

$$\langle V\rangle = \langle A\rangle\langle \Delta\rangle$$

$$\langle Q\rangle = \rho c(\langle T\rangle - t_o)\langle A\rangle\langle \Delta\rangle$$

$$\mathrm{var}(V) = \left[\mathrm{var}(A) + \langle A\rangle^2\right]\left[\mathrm{var}(\Delta) + \langle \Delta\rangle^2\right] - \langle A\rangle^2\langle \Delta\rangle^2$$

$$\mathrm{var}(Q) = (\rho c)^2\left[\mathrm{var}(T) + (\langle T\rangle - t_o)^2\right]\left[\mathrm{var}(A) + \langle A\rangle^2\right]\left[\mathrm{var}(\Delta) + \langle \Delta\rangle^2\right] - (\rho c)^2(\langle T\rangle - T_o)^2\langle A\rangle^2\langle \Delta\rangle^2$$

For a sum of n systems

$$\langle SQ\rangle = \sum_{i=1}^{n} \langle Q_i\rangle$$

$$\mathrm{var}(SQ) = \sum_{i=1}^{n} \mathrm{var}(Q_i)$$

Properties of a triangular density

$$\alpha(x) = \begin{cases} 0 & x<0 \\ 1 & x\geq 1 \end{cases}$$

$$f_A(a) = 2\left\{\frac{(a - a_1)\alpha(a - a_1)}{(a_3 - a_1)(a_2 - a_1)} + \frac{(a - a_2)\alpha(a - a_2)}{(a_2 - a_3)(a_2 - a_1)} + \frac{(a - a_3)\alpha(a - a_3)}{(a_3 - a_1)(a_3 - a_2)}\right\}$$

$$F_A(a) = \frac{(a - a_1)^2\alpha(a - a_1)}{(a_3 - a_1)(a_2 - a_1)} + \frac{(a - a_2)^2\alpha(a - a_2)}{(a_2 - a_3)(a_2 - a_1)} + \frac{(a - a_3)^2\alpha(a - a_3)}{(a_3 - a_1)(a_3 - a_2)}$$

$$\langle A\rangle = (a_1 + a_2 + a_3)/3$$

$$\mathrm{var}(A) = \left[(a_2 - a_1)^2 + (a_3 - a_2)^2 + (a_2 - a_1)(a_3 - a_2)\right]/18$$

Properties of a normal density

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-\frac{(x - m)^2}{2\sigma^2}\}$$

$$F(x) = \frac{1}{2}\left\{1 + \text{erf}\left[\frac{x - m}{(2\sigma^2)^{1/2}}\right]\right\}$$

$$\langle X \rangle = m$$

$$\text{var}(X) = \sigma^2$$

Properties of a log-normal density

$$f(x) = \frac{1}{x(2\pi b^2)^{1/2}} \exp\left\{-\frac{(\ln x - a)}{2b^2}\right\}$$

$$F(x) = \frac{1}{2}\left\{1 + \text{erf}\,\frac{\ln x - a}{(2b^2)^{1/2}}\right\}$$

$$\langle X \rangle = \exp\{a + \frac{b^2}{2}\}$$

$$\langle X^2 \rangle = \exp\{2a + 2b^2\}$$

$$a = \ln\{\langle X \rangle^2 / (\langle X^2 \rangle)^{1/2}\}$$

$$b^2 = \ln\{\langle X^2 \rangle / \langle X \rangle^2\}$$

Density and distribution for volume and stored energy for a single system

$$\text{del}(a_i, j) \equiv \begin{cases} (a_1 - a_3)(a_1 - a_2) & j=1 \\ (a_2 - a_3)(a_2 - a_1) & j=2 \\ (a_3 - a_1)(a_3 - a_2) & j=3 \end{cases}$$

$$f_V(v) = \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{4\left[2(a_j\delta_k - v) + (a_j\delta_k + v)\ln\frac{v}{a_j\delta_k}\right]\alpha(v - a_j\delta_k)}{\text{del}(a_i, j)\ \text{del}(\delta_i, k)}$$

$$F_V(v) = \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{\left[3(a_j^2\delta_k^2 - v^2) + v(4a_j\delta_k + 2v)\ln\frac{v}{a_j\delta_k}\right]\alpha(v - a_j\delta_k)}{\text{del}(a_i, j)\ \text{del}(\delta_i, k)}$$

$$q_{jkm} = \rho c a_j \delta_k (t_m - t_o)$$

$$f_Q(q) = \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{m=1}^{3} \frac{8q}{(\rho c)^2} \left[\left(1 - \frac{q_{jkm}}{q}\right)\left(6 + \frac{1}{2} \ln\left(\frac{q}{q_{jkm}}\right)^2\right) - 3\left(1 + \frac{q_{jkm}}{q}\right) \ln \frac{q}{q_{jkm}} \right] \frac{\alpha(q - q_{jkm})}{\mathrm{del}(a_i, j)\,\mathrm{del}(\delta_i, k)\,\mathrm{del}(t_i, m)}$$

$$F_Q(q) = \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{m=1}^{3} 8\left[a_j \delta_k (t_m - t_o)\right]^2 \left\{ \frac{31}{8}\left(\frac{q^2}{q_{jkm}^2} - 1\right) - 4\left(\frac{q}{q_{jkm}} - 1\right) + \left(\frac{q^2}{4q_{jkm}^2} - \frac{q}{2q_{jkm}}\right)\left(\ln \frac{q}{q_{jkm}}\right)^2 - \left(\frac{7}{4}\frac{q^2}{q_{jkm}^2} + 2\frac{q}{q_{jkm}}\right) \ln \frac{q}{q_{jkm}} \right\}$$

$$\times \frac{\alpha(q - q_{jkm})}{\mathrm{del}(a_i, j)\ \mathrm{del}(\delta_i, k)\ \mathrm{del}(t_i, m)}$$

Appendix 2 Computer program for calculating the stored energy for a single hydrothermal convection system

The equations for the stored energy in a single system have been used to write a program in PL/I for the Honeywell 68/80 computer. The program is a straightforward calculation routine. The only obscure aspect is that it is written for a time share mode, so that the program is called from the user's working directory, the program then requests the input data from the user, sends the output to various files that the user may then inspect and have printed. For batch use, the put statements would have to be reorganized accordingly.

```
fqi: procedure;
/*-------------------------------------------------------------------
   This program calculates the values of the probability density and
distribution function for stored heat.  Values of the minumum,
most likely, and maximum of temperature, area, and thickness for triangular
density are requested from a terminal after the program has been called.
The values are printed in file "dataout" for printing and in files "f"
and "F" for transmission to the 4051.

                                                    March,1978
-------------------------------------------------------------------*/
  declare (sysin) file stream input;
  declare (dataout) file stream output;
  declare (f)       file stream output;
  declare (F)       file stream output;
  declare (sysprint)file stream output;
  open file (sysin) input stream;
  open file (dataout) output stream print pagesize(58) linesize(100);
  open file (f)       output stream;
  open file (F)       output stream;
  open file (sysprint) output stream;

  dif: procedure(a,i,delt);
       declare (a(3),delt) float(30),i fixed binary (30);
       if i=1 then delt=(a(1)-a(3))*(a(1)-a(2));
       if i=2 then delt=(a(2)-a(1))*(a(2)-a(3));
       if i=3 then delt=(a(3)-a(1))*(a(3)-a(2));
       return ;
       end;

  mv: procedure(min,ml,max,mean,var);
       declare (min,ml,max,mean,var) float (30);
       mean=(min+ml+max)/3e0;
       var=((min-ml)*(min-ml)+(max-ml)*(max-ml)+(ml-min)*(max-ml))/18e0;
       return;
       end;

  declare (t(3),a(3),d(3),qa(3),tmean,tvar,amean,avar,dmean,dvar,qmean,qvar,
           rhoc,q,fq,Fq,qc,den,delta1,delta2,delta3) float (30),
           (i,j,k,l) fixed binary (30),
           name character (25)  varying;
  declare (sd,unit,del) float (30);

  rhoc=0.6e-3;

  put file (sysprint) edit
       ("You need values of the probability density and distribution function",
        "for stored heat. Input the min,ml, and max of",
        "temperature(C), area(km**2), and thickness(km) and the calculation interval and ",
        "the system name in double quotes. Hit return and leave the rest to me.")
        (4(skip,a));
  get file (sysin) list (t,a,d,unit,name);

  call mv(t(1),t(2),t(3),tmean,tvar);
  call mv(a(1),a(2),a(3),amean,avar);
  call mv(d(1),d(2),d(3),dmean,dvar);

  qmean=rhoc*(tmean-15e0)*amean*dmean;
  qvar=rhoc*rhoc*((tvar+tmean*tmean-30e0*tmean+225e0)*(avar+amean*amean)*(dvar+dmean*dmean)
       -(tmean-15e0)*(tmean-15e0)*amean*amean*dmean*dmean);
```

```
    qa=rhoc*(t-15e0)*a*d;
    sd=sqrt(qvar);


put file (dataout) edit (name,"-stored heat","min","ml","max","mean","s.d.",
        "temperature(C)",t,tmean,sqrt(tvar),
        "area(km**2)",a,amean,sqrt(avar),
        "depth(km)",d,dmean,sqrt(dvar),
        "heat(e18 cal)",qa,qmean,sqrt(qvar))
        (skip,a,a,skip,x(20),5(a(8)),skip,a(16),5(f(8,0)),
         skip,a(16),5(f(8,1)),skip,a(16),5(f(8,1)), skip,a(16),5(f(8,3)));
put file (dataout) edit ("q","f(q)","1-F(q)") (skip,x(4),3(a(10)));

do i=1 to 400;
   del=i;
   q=trunc(qa(1))+del*unit;
   fq=0e0; Fq=0e0;
   do j=1 to 3;
   do k=1 to 3;
   do l=1 to 3;

   qc=q/(rhoc*a(j)*d(k)*(t(l)-15e0));
   if qc >= 1e0 then do;
      call dif(a,j,delta1);
      call dif(d,k,delta2);
      call dif(t,l,delta3);
      den=delta1*delta2*delta3;
      fq=fq+((8e0*q)/(rhoc*rhoc))*((1e0-1e0/qc)*(6e0+0.5e0*log(qc)*log(qc))
            -3e0*(1e0+1e0/qc)*log(qc))/den;
      Fq=Fq+8e0*((a(j)*d(k)*(t(l)-15e0))**2)*(31e0*(qc*qc-1e0)/8e0-4e0*(qc-1e0)
            +qc*(qc/4e0-0.5e0)*log(qc)*log(qc)-(1.75e0*qc*qc+2e0*qc)*log(qc))/den;
      end;
   end; end; end;
   put file (dataout) edit (q,fq,1e0-Fq) (skip,f(8,2),x(2),f(10,5),f(10,4));
   put file (f) edit (q,",",fq,",") (f(9,3),a,f(10,4),a);
   put file (F) edit (q,",",1e0-Fq,",") (f(9,3),a,f(10,4),a);
   if lineno (dataout) >57 then
      put file (dataout) edit
      (name,"-stored heat","q","f(q)","1-F(q)") (skip,a,a,skip,x(4),3(a(10)));
   if Fq>0.99995e0 then go to out;
   end;

out:
put file (sysprint) edit ("x,f, and 1-F are in file dataout for printing",
                          "purposes. x and f are in file f and x and 1-F",
                          "are in file F for sending to 4051") (3(skip,a));
close file (sysin);
close file (dataout);
close file (f);
close file (F);
close file (sysprint);
end;
```

Appendix 3 Computer program for performing a Monte Carlo simulation to calculate the distribution of stored energy in one or many systems

Harold Javitz of SRI International wrote a Fortran program to calculate the distribution function for stored energy with triangular input density functions for temperature, area, and thickness. The basic relations used are equation (38), the inversion formula for a distribution function from a triangular density function, and equation (14), the relation between temperature, area, and thickness, and stored energy. The program uses the subroutine ggu4 written by International Mathematical and Statistical Libraries, Inc. For use at the U. S. Geological Survey, the program was rewritten in PL/I for the Honeywell 68/80 with the MULTICS interactive operating system. In order to use the compiled version of the program, the first step is to type in the command

```
asr >iml>imsl -after working_dir
```

to add the imsl library to the search rules, a file called "pinc" must then be created with input data. All data items need only be separated by a space. The first three data items are the number of systems, the number of Monte Carlo trials to be performed, and the seed for the random number generator. A reasonable starting seed according to Mr. Javitz is 123457. Part of the output is the final value of the seed to be used for the next run. The minimum, most likely, and maximum values for temperature (^{o}C), area (km^2), and thickness (km) then follow for each system. Degenerate density functions with minimum=most likely=maximum may be used for any quantity that is known precisely. The output gives the

input seed, the input parameters for the density functions, the mean, variance, and standard deviation for the Monte Carlo simulation of the resulting distribution and the value of the seed for the next run. The Monte Carlo trials are sorted in ascending order and printed with the number of occurrences in all the previous trials as the last number in each line. The estimated probability that the stored energy is greater than any value found in the table is one minus the number of trials up to and including that value divided by the total number of trials. Using 400 trials means each entry increments the estimated probability by 0.0025.

```
PROGRAM     LOCATE     TRACE                          CDC 6700 FTN V3.0-355F OPT=0  78/0
                       PROGRAM LOCATE(INPUT,TAPE1=INPUT,OUTPUT)
                       DIMENSION S(300,3,6)
                       DIMENSION U(3,300),TEMP(300)
                       DIMENSION E(400),X(3)
                       READ(1,4) NLOC,ITER,ISEED
                       PRINT 203,NLOC,ITER,ISEED
                   203 FORMAT (*1*,*NLOC = *,I4,*   ITER = *,I4,*   ISEED = *,I10,//)
                     4 FORMAT(2I3,I10)
                       DO 200 I=1,NLOC
                       READ(1,8)((S(I,J,K),K=1,3),J=1,3)
                     6 FORMAT(3(3F6.1,5X))
                     8 FORMAT(9F6.1)
                       IF(EOF(1).NE.0)GO TO 201
                   100 PRINT 6,((S(I,J,K),K=1,3),J=1,3)
                       DO 200 I=1,NLOC
                       DO 200 J=1,3
                       IF(ABS(S(I,J,3)-S(I,J,1)).LE. .00001)GO TO 110
                       S(I,J,4)=(S(I,J,2)-S(I,J,1))/(S(I,J,3)-S(I,J,1))
                       GO TO 120
                   110 S(I,J,4)=1.0
                   120 CONTINUE
                       S(I,J,5)=(S(I,J,3)-S(I,J,1))*(S(I,J,2)-S(I,J,1))
                       S(I,J,6)=(S(I,J,3)-S(I,J,1))*(S(I,J,3)-S(I,J,2))
                   200 CONTINUE
                       IOPT=1
                       DO 300 N=1,ITER
                       DO 310 K=1,3
                       CALL GGU4(ISEED,NLOC,IOPT,TEMP)
                       DO 320 L=1,NLOC
                   320 U(K,L)=TEMP(L)
                   310 IOPT=0
                       E(N)=0.0
                       DO 400 I=1,NLOC
                       DO 500 J=1,3
                       IF(U(J,I).LT.S(I,J,4))X(J)=S(I,J,1)+SQRT(U(J,I)*S(I,J,5))
                       IF(U(J,I).GE.S(I,J,4))X(J)=S(I,J,3)-SQRT((1.0-U(J,I))*S(I,J,6))
                   500 CONTINUE
                   400 E(N)=E(N)+(X(1)-15)*X(2)*X(3)
                   300 CONTINUE
                       XITER=ITER
                       SUME=0.0
                       SUME2=0.0
                       DO 600 N=1,ITER
                       SUME=SUME+E(N)
                   600 SUME2=SUME2+E(N)**2
                       EMEAN=SUME/XITER
                       EVAR=SUME2/XITER-EMEAN**2
                       EVAR=EVAR*XITER/(XITER-1.0)
                       ESD=SQRT(EVAR)
                       PRINT 7,EMEAN,EVAR,ESD,ISEED
                     7 FORMAT(*0*,*EMEAN= *,E12.5,5X,*EVAR= *,E12.5,5X,
                    **ESTDEV = *,E12.5,5X,*ISEED = *,I10)
                       CALL BSORT(E,ITER)
                       STOP
                   201 PRINT 3

PROGRAM     LOCATE     TRACE                          CDC 6700 FTN V3.0-355F OPT=0  78/05/
                     3 FORMAT(*0*,*ILLEGAL END OF DATA*)
                       END
```

```
      SUBROUTINE  BSORT     TRACE                    CDC 6700 FTN V3.0-355F OPT=0   7

                            SUBROUTINE BSORT(E,ITER)
                            DIMENSION E(ITER)
                            DIMENSION A(10)
                            LOGICAL SWITCH
                5           REAL ITEMP
                          1 SWITCH=.FALSE.
                            ITER1=ITER-1
                            DO 10 I=1,ITER1
                            IF(E(I).LE.E(I+1))GO TO 10
               10           SWITCH=.TRUE.
                            ITEMP=E(I)
                            E(I)=E(I+1)
                            E(I+1)=ITEMP
                         10 CONTINUE
               15           IF(SWITCH)GO TO 1
                            NPRINT=0
                        125 NREM=ITER-NPRINT
                            IF(NREM.EQ.0)GO TO 121
                            NNOW=10
               20           IF(NREM .LT. 10)NNOW=NREM
                            DO 100 J=1,NNOW
                        100 A(J)=E(NPRINT+J)
                            IF(NNOW .LT. 10)GO TO 105
                            NPRINT=NPRINT+10
               25           PRINT 110,(A(J),J=1,10),NPRINT
                        110 FORMAT(*0*,10E12.5,I6)
                            GO TO 125
                        105 PRINT 120,(A(J),J=1,NNOW)
                        120 FORMAT(*0*,10E12.5)
               30       121 CONTINUE
                            RETURN
                            END
```

```
montcar: procedure;
/*   Monte Carlo simulation for calculating probability distribution function
 of stored energy in a single or in many systems.  Some variables are defined as:
      nloc = number of systems
      iter = number of Monte Carlo cycles
      s(i,j,k) = parameters for all systems
             i = the i th system
             j = the j th property
                 1 = temperature, deg C
                 2 = area, km**2
                 3 = thickness, km
             k = the k th parameter
                 1 = mimumum
                 2 = most likely
                 3 = maximum
                 4,5 = quantities derived from 1, 2, and 3

 Written by Harold Javitz, SRI International and translated into PL/I by
                 USGS.
                                                    May, 1978       */
bsort: procedure (e,iter);
 declare (iter,s,i,nprint,nnow,nrem) fixed binary (31);
 declare (e(*),itemp,a(10)) float (16);
 switch: s=1;
 do i=1 to iter-1;
    if e(i)<=e(i+1) then go to again;
    s=0;
    itemp=e(i);
    e(i)=e(i+1);
    e(i+1)=itemp;
 again: end;
 if s=0 then go to switch;
 nprint=0;
 more: nrem=iter-nprint;
 nnow=10;
 if nrem<10 then nnow=nrem;
 do j=1 to nnow; a(j)=e(nprint+j); end;
 if nnow < 10 then go to last;
 nprint=nprint+10;
 put file (mc) edit ((a(j) do j=1 to 10),nprint) (skip,10(f(8,3)),f(6));
 go to more;
 last: if nnow>0 then put file (mc) edit ((a(j) do j=1 to nnow))
                                          (skip,10(f(8,3)));
 return;
end;

declare (pinc) file stream input;
declare (mc)   file stream output;
open file (pinc) input stream;
open file (mc)   output stream print pagesize (58) linesize (136);

declare (s(300,3,6),u(3,300),temp(300),x(3),xiter,
         sume,sume2,emean,evar,esd) float (16);
declare e(iter) float (16) controlled;
declare ggu4 entry options (variable);
declare (nloc,iter,iseed,iopt,i,j,k,l,n) fixed binary (31);
 get file (pinc) list (nloc,iter,iseed);
 put file (mc) edit ("seed in",iseed) (skip,a(10),f(13));
 allocate e;
 do i=1 to nloc;
```

```
 get file (pinc) list (((s(i,j,k) do k=1 to 3) do j=1 to 3));
 put file (mc)   edit (((s(i,j,k) do k=1 to 3) do j=1 to 3)) (skip,9(f(6,1)));
end;
do i=1 to nloc;
 do j=1 to 3;
  if abs(s(i,j,3)-s(i,j,1))<=1e-5 then
     s(i,j,4)=1e0; else
     s(i,j,4)=(s(i,j,2)-s(i,j,1))/(s(i,j,3)-s(i,j,1));
  s(i,j,5)=(s(i,j,3)-s(i,j,1))*(s(i,j,2)-s(i,j,1));
  s(i,j,6)=(s(i,j,3)-s(i,j,1))*(s(i,j,3)-s(i,j,2));
end; end;
iopt=1;
do n=1 to iter;
 do k=1 to 3;
  call ggu4(iseed,nloc,iopt,temp);
  do l=1 to nloc;
   u(k,l)=temp(l); end;
 iopt=0; end;
 e(n)=0e0;
 do i=1 to nloc;
  do j=1 to 3;
   if u(j,i)<s(i,j,4) then
      x(j)=s(i,j,1)+sqrt(u(j,i)*s(i,j,5)); else
      x(j)=s(i,j,3)-sqrt((1e0-u(j,i))*s(i,j,6));
  end;
  e(n)=e(n)+(x(1)-15e0)*x(2)*x(3);
 end;
 e(n)=0.6e-3*e(n);
end;
xiter=iter;
sume=0e0;
sume2=0e0;
do n=1 to iter;
 sume=sume+e(n);
 sume2=sume2+e(n)*e(n);
end;
emean=sume/xiter;
evar=sume2/xiter-emean*emean;
evar=evar*xiter/(xiter-1e0);
esd=sqrt(evar);

put file (mc) edit ("mean","var","s.d.","seed",emean,evar,esd,iseed)
                   (skip,x(3),4(a(8)),skip,3(f(8,3)),f(15));
call bsort(e,iter);
close file (pinc);
close file (mc);
end;
```

seed in 1324647809

150.0 190.0 225.0 10.0 50.0 85.0 1.5 2.0 2.3

mean	var	s.d.	seed							
9.570	10.990	3.315	2070723536							
2.188	2.304	2.319	2.943	3.026	3.179	3.357	3.431	3.507	3.610	10
3.634	3.706	3.719	3.777	3.809	3.902	3.909	4.016	4.079	4.188	20
4.202	4.241	4.373	4.448	4.532	4.689	4.712	4.748	4.785	5.010	30
5.072	5.073	5.127	5.137	5.203	5.203	5.215	5.220	5.289	5.295	40
5.332	5.365	5.482	5.532	5.546	5.601	5.649	5.654	5.662	5.759	50
5.797	5.802	5.813	5.833	5.925	5.968	6.003	6.011	6.017	6.018	60
6.039	6.084	6.089	6.114	6.149	6.183	6.193	6.234	6.239	6.270	70
6.286	6.316	6.354	6.450	6.453	6.476	6.511	6.602	6.613	6.614	80
6.617	6.626	6.651	6.722	6.771	6.786	6.812	6.833	6.944	6.954	90
6.978	6.980	7.034	7.092	7.104	7.119	7.191	7.217	7.237	7.253	100
7.254	7.284	7.332	7.375	7.375	7.387	7.396	7.403	7.481	7.496	110
7.509	7.511	7.534	7.568	7.580	7.593	7.599	7.619	7.622	7.654	120
7.664	7.675	7.690	7.706	7.754	7.761	7.806	7.825	7.833	7.903	130
7.915	7.946	8.009	8.046	8.081	8.089	8.109	8.116	8.128	8.147	140
8.168	8.190	8.210	8.222	8.225	8.240	8.258	8.267	8.269	8.282	150
8.283	8.287	8.315	8.346	8.354	8.396	8.401	8.452	8.473	8.480	160
8.483	8.490	8.560	8.569	8.581	8.597	8.618	8.680	8.681	8.703	170
8.732	8.741	8.746	8.748	8.750	8.773	8.806	8.818	8.866	8.889	180
8.891	8.988	9.038	9.042	9.049	9.116	9.135	9.153	9.165	9.170	190
9.184	9.184	9.191	9.199	9.217	9.240	9.296	9.322	9.336	9.354	200
9.357	9.367	9.438	9.471	9.502	9.513	9.538	9.599	9.618	9.636	210
9.636	9.641	9.652	9.655	9.684	9.771	9.851	9.856	9.883	9.885	220
9.892	9.893	9.963	9.983	10.023	10.045	10.098	10.103	10.127	10.144	230
10.158	10.244	10.285	10.335	10.343	10.354	10.367	10.387	10.394	10.396	240
10.401	10.442	10.448	10.450	10.460	10.533	10.534	10.549	10.589	10.606	250
10.608	10.609	10.638	10.652	10.679	10.680	10.681	10.713	10.804	10.820	260
10.851	10.872	10.890	10.914	10.921	10.949	10.982	10.997	11.003	11.019	270
11.022	11.028	11.030	11.035	11.036	11.062	11.124	11.129	11.162	11.233	280
11.246	11.272	11.279	11.328	11.332	11.340	11.341	11.365	11.393	11.393	290
11.393	11.423	11.449	11.548	11.578	11.647	11.666	11.674	11.708	11.713	300
11.716	11.733	11.755	11.855	11.856	11.908	11.928	11.970	11.981	11.990	310
11.991	12.044	12.071	12.092	12.100	12.189	12.244	12.259	12.303	12.321	320
12.322	12.376	12.390	12.410	12.465	12.491	12.503	12.517	12.522	12.561	330
12.574	12.625	12.651	12.680	12.686	12.728	12.854	12.907	13.154	13.157	340
13.235	13.296	13.323	13.563	13.577	13.579	13.629	13.633	13.661	13.684	350
13.698	13.698	13.707	13.742	13.847	13.859	13.930	13.953	14.034	14.046	360
14.065	14.123	14.127	14.128	14.160	14.277	14.384	14.468	14.762	14.865	370
14.976	15.139	15.171	15.223	15.236	15.242	15.262	15.384	15.400	15.429	380
15.594	15.612	15.655	15.724	15.804	15.923	15.952	15.953	16.020	16.024	390
16.231	16.676	16.713	16.813	16.982	16.995	17.034	17.693	17.897	19.990	400

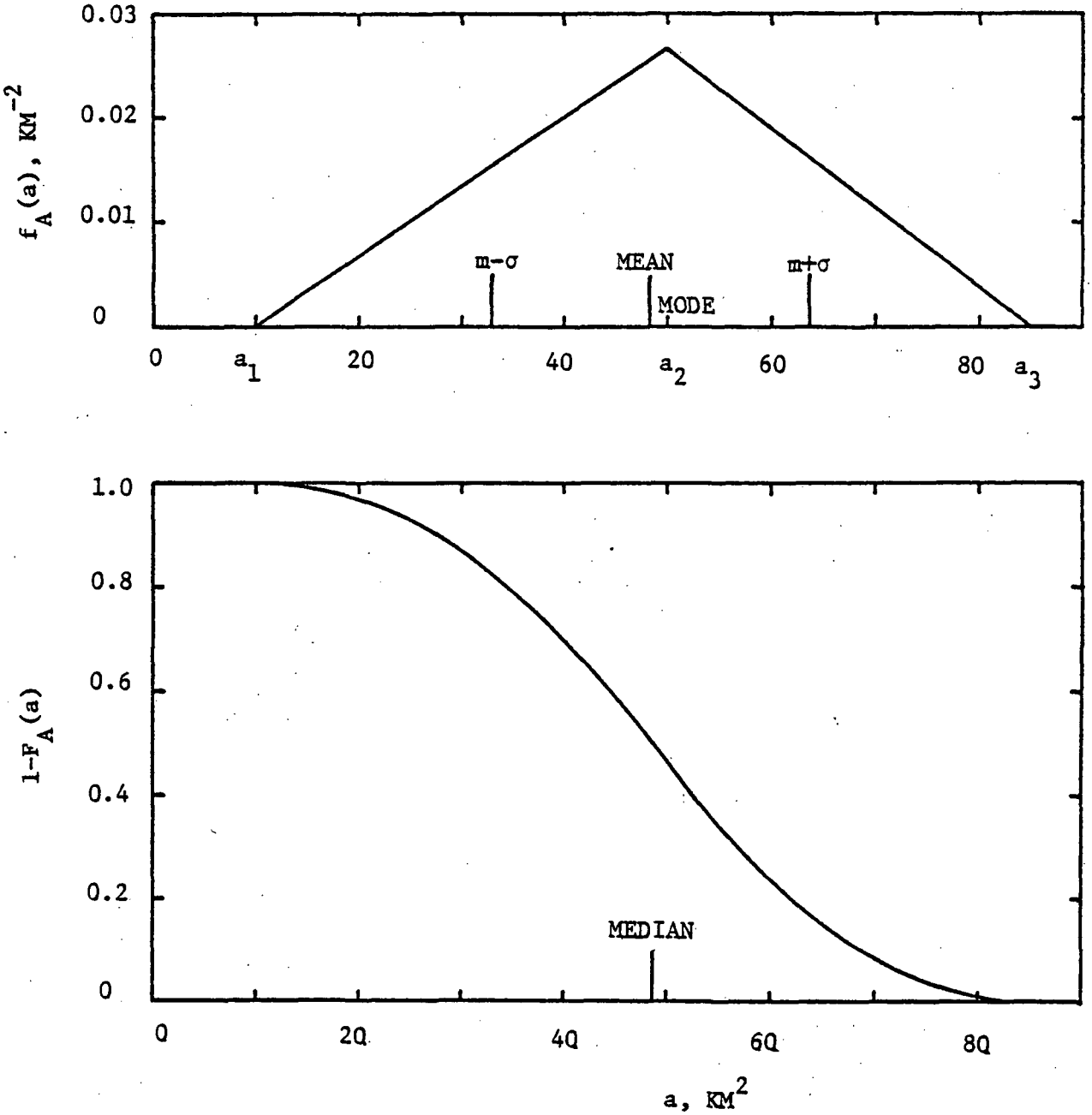

Figure 1.--Density and distribution for the area of the hydrothermal system at Heber, CA--1975 data. Mean, mean plus or minus one standard deviation (σ), mode, and median are shown.

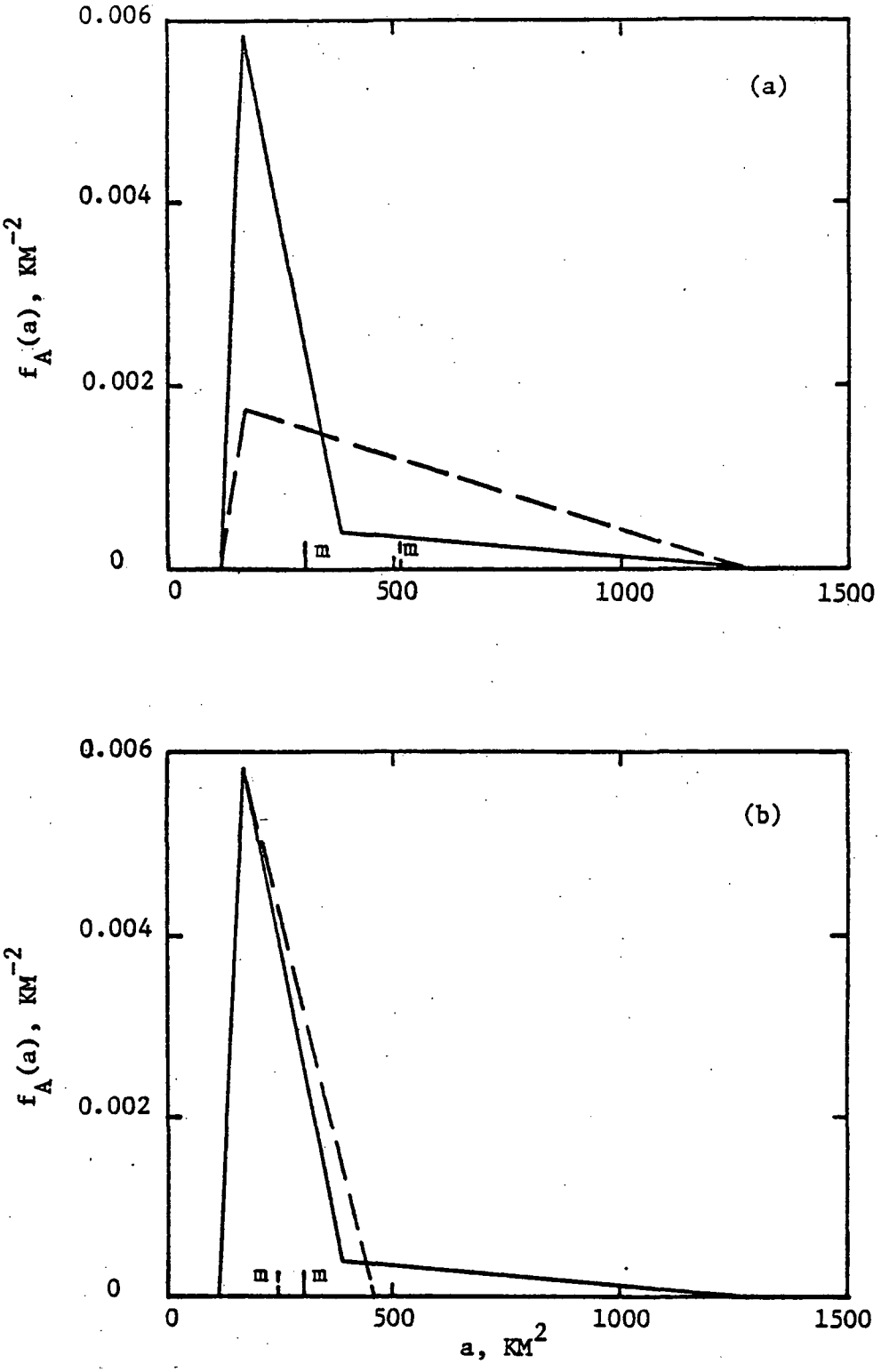

Figure 2.—Several hypothetical density functions based on 1975 data for Coso, CA. Solid line is proposed true density and broken lines are two approximations to this true density. Mean m for each density shown on the area axis.

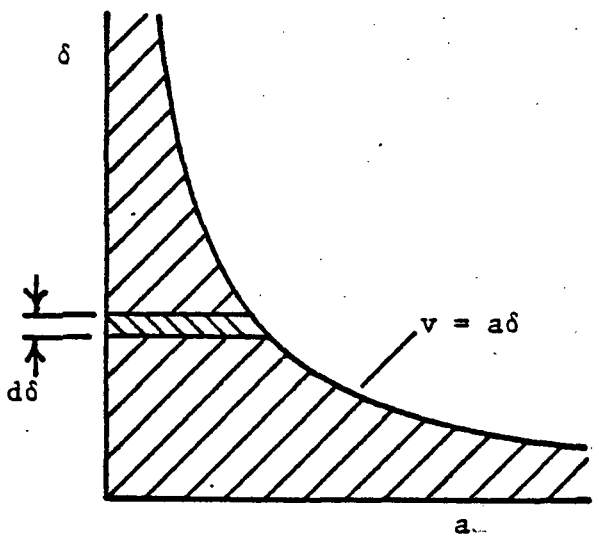

Figure 3.--Region of integration for determining the probability $P\{V \leq v\}$.

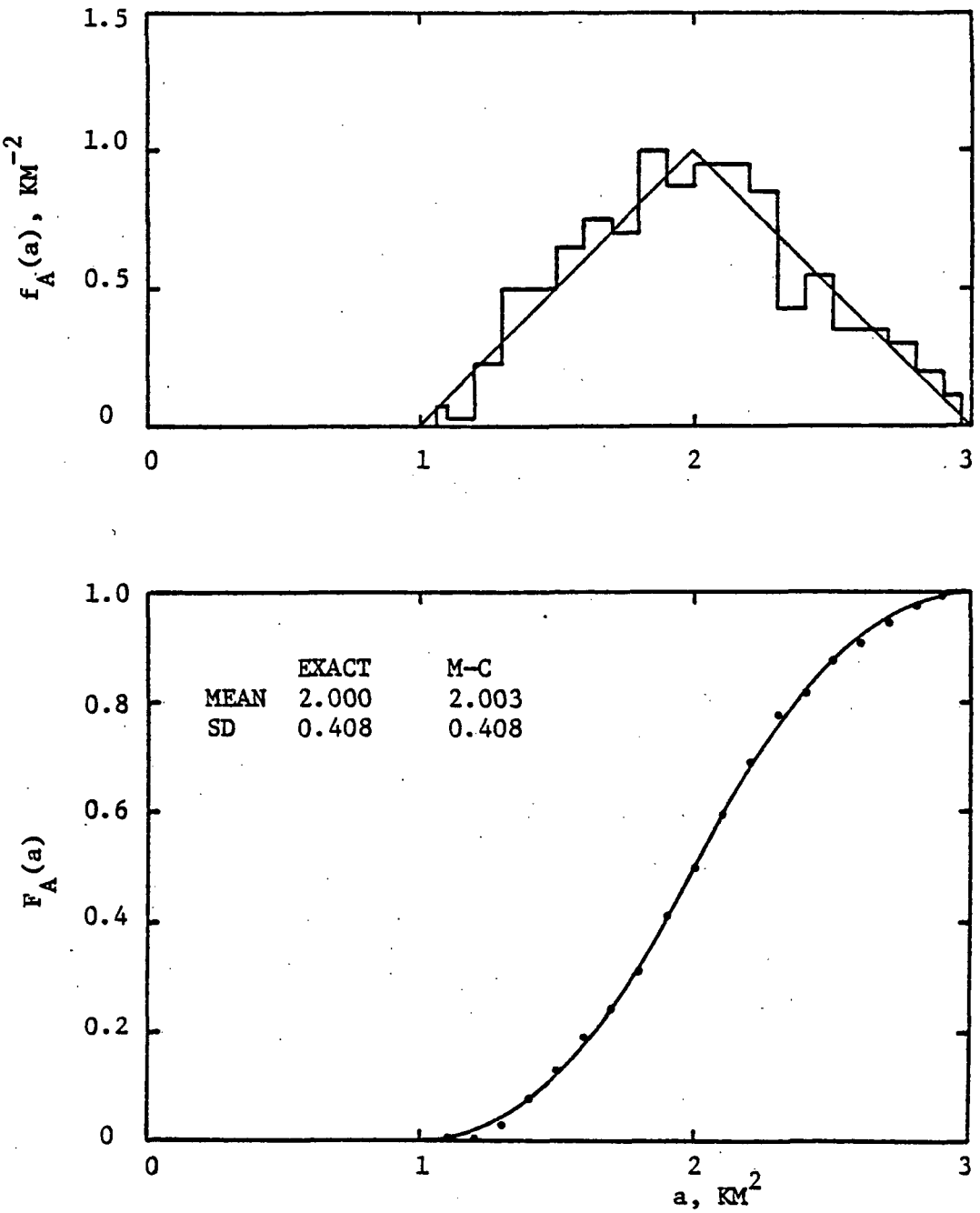

Figure 4.—Monte-Carlo simulation of a triangular density function shown as a histogram (top) and the resulting sample distribution function shown as dots (bottom), compared to exact distribution function.

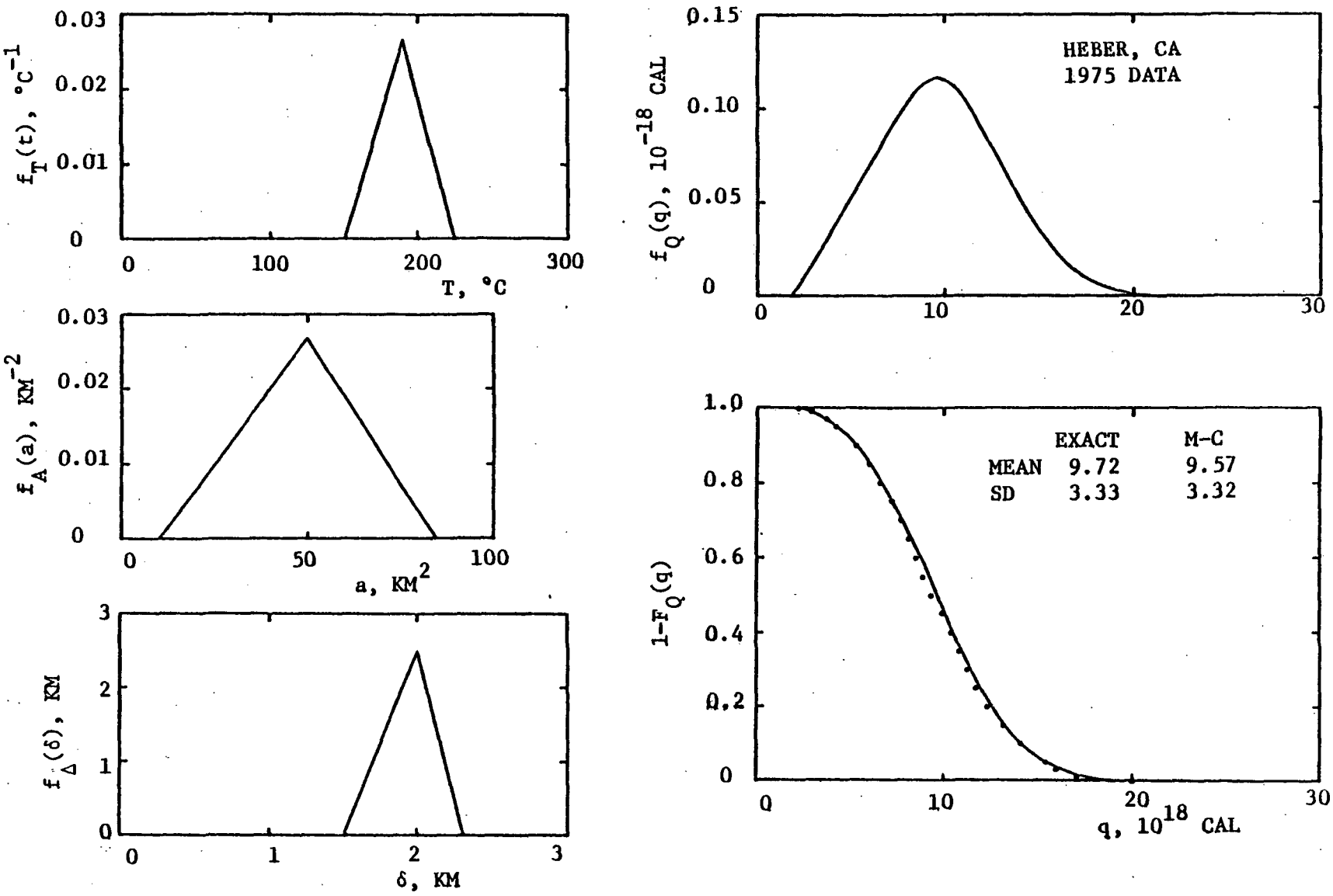

Figure 5.--Probability density functions of temperature, area, thickness, and stored energy and distribution function of stored energy for hydrothermal system at Heber, CA. Monte-Carlo sample distribution function for stored energy shown as dots.

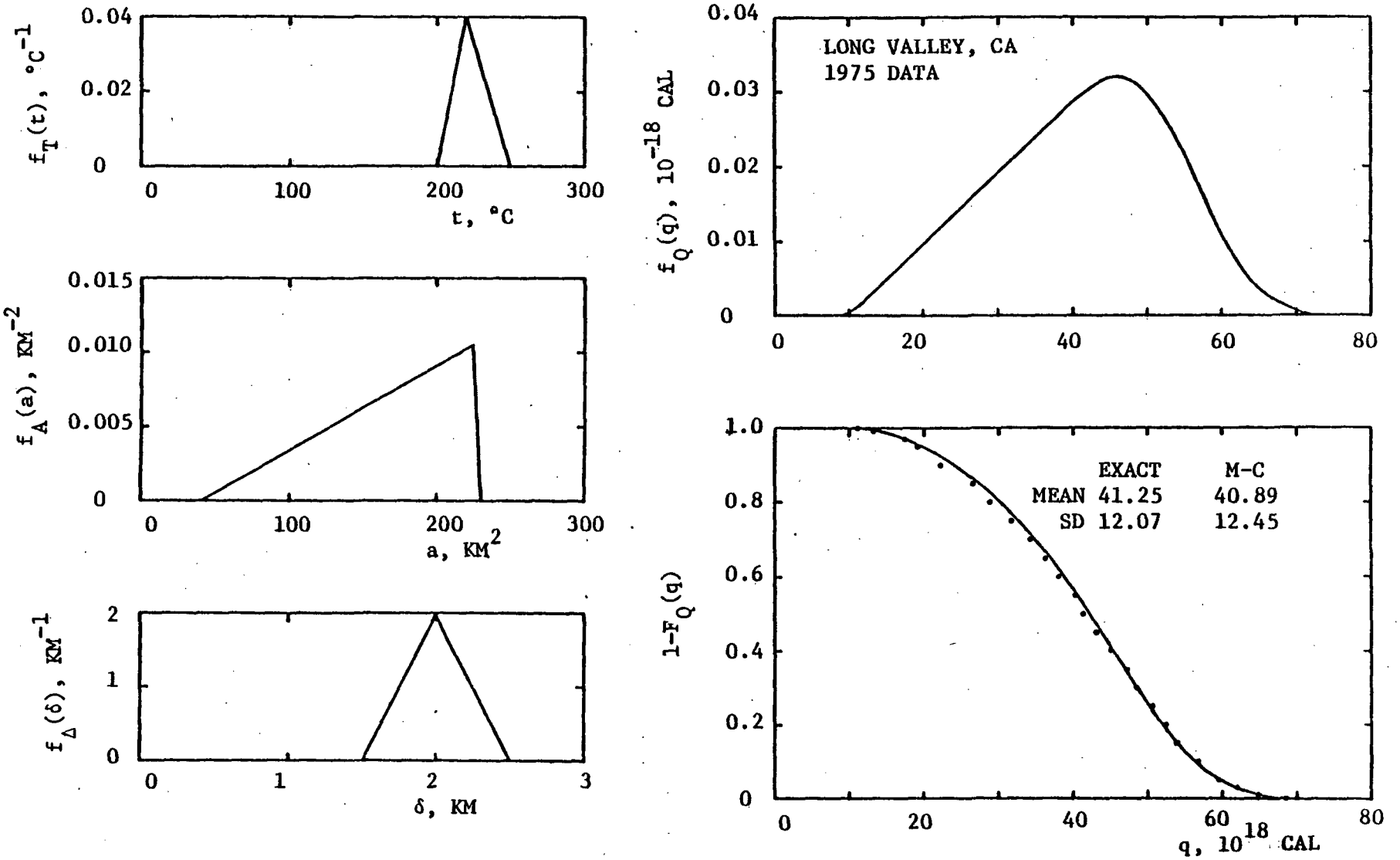

Figure 6.--Probability density functions of temperature, area, thickness, and stored energy for hydrothermal system at Long Valley, CA. Monte-Carlo sample distribution function for stored energy shown as dots.

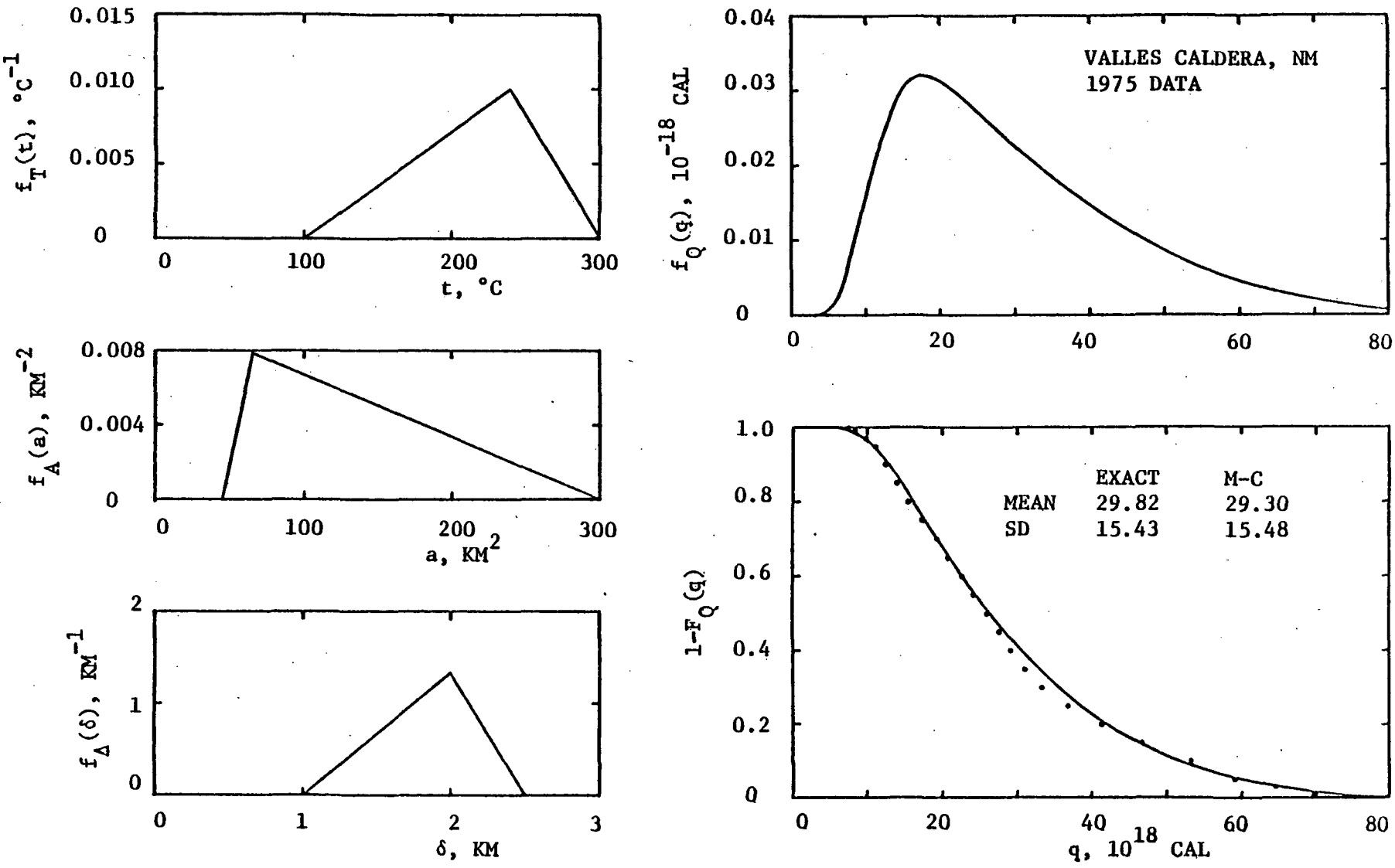

Figure 7.--Probability density functions of temperature, area, thickness, and stored energy and distribution function of stored energy for hydrothermal system at Valles Caldera, N. M. Monte-Carlo sample distribution function for stored energy shown as dots.

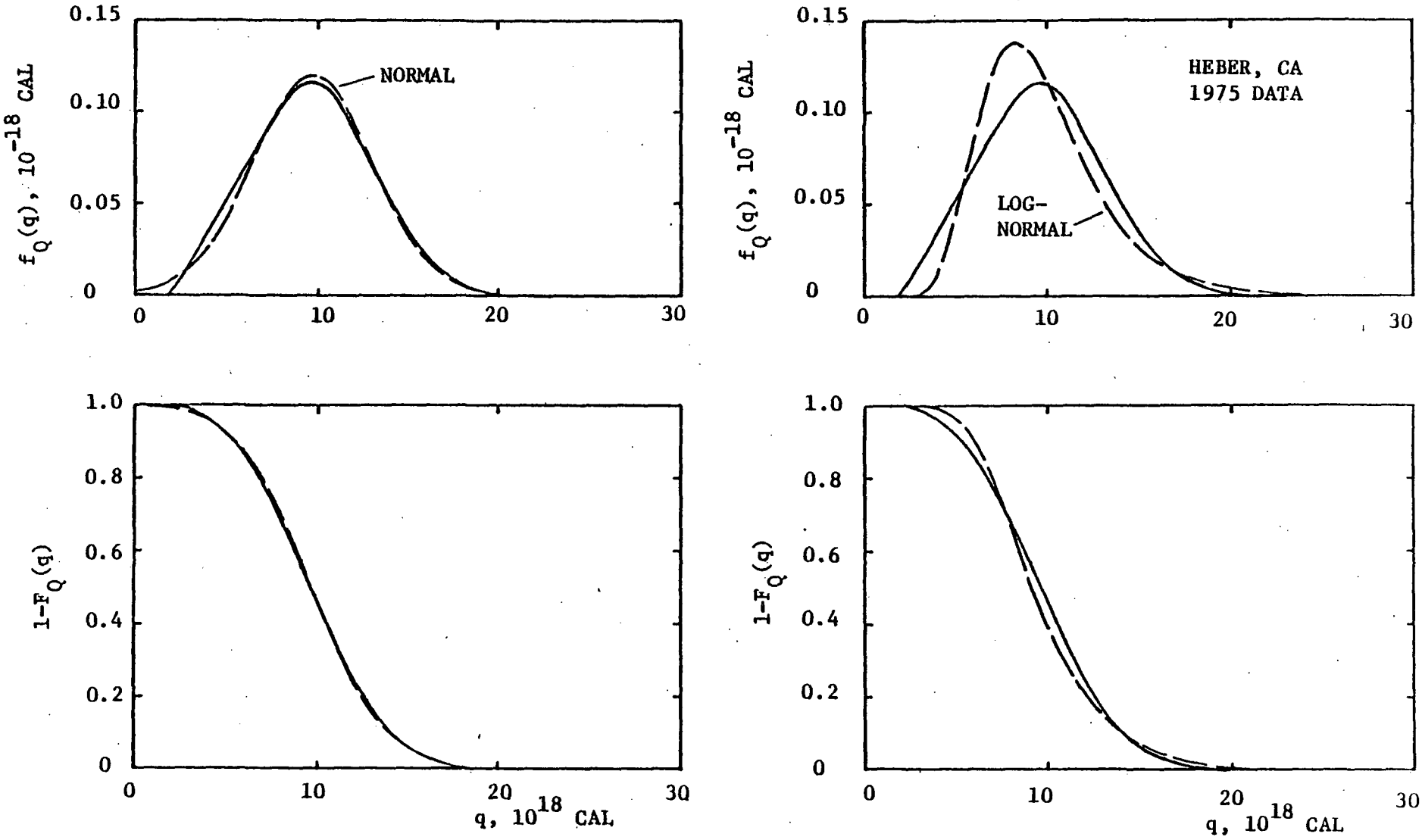

Figure 8.--Probability density and distribution functions of stored energy for hydrothermal system at Heber, CA compared to normal and log-normal functions with same mean and standard deviation.

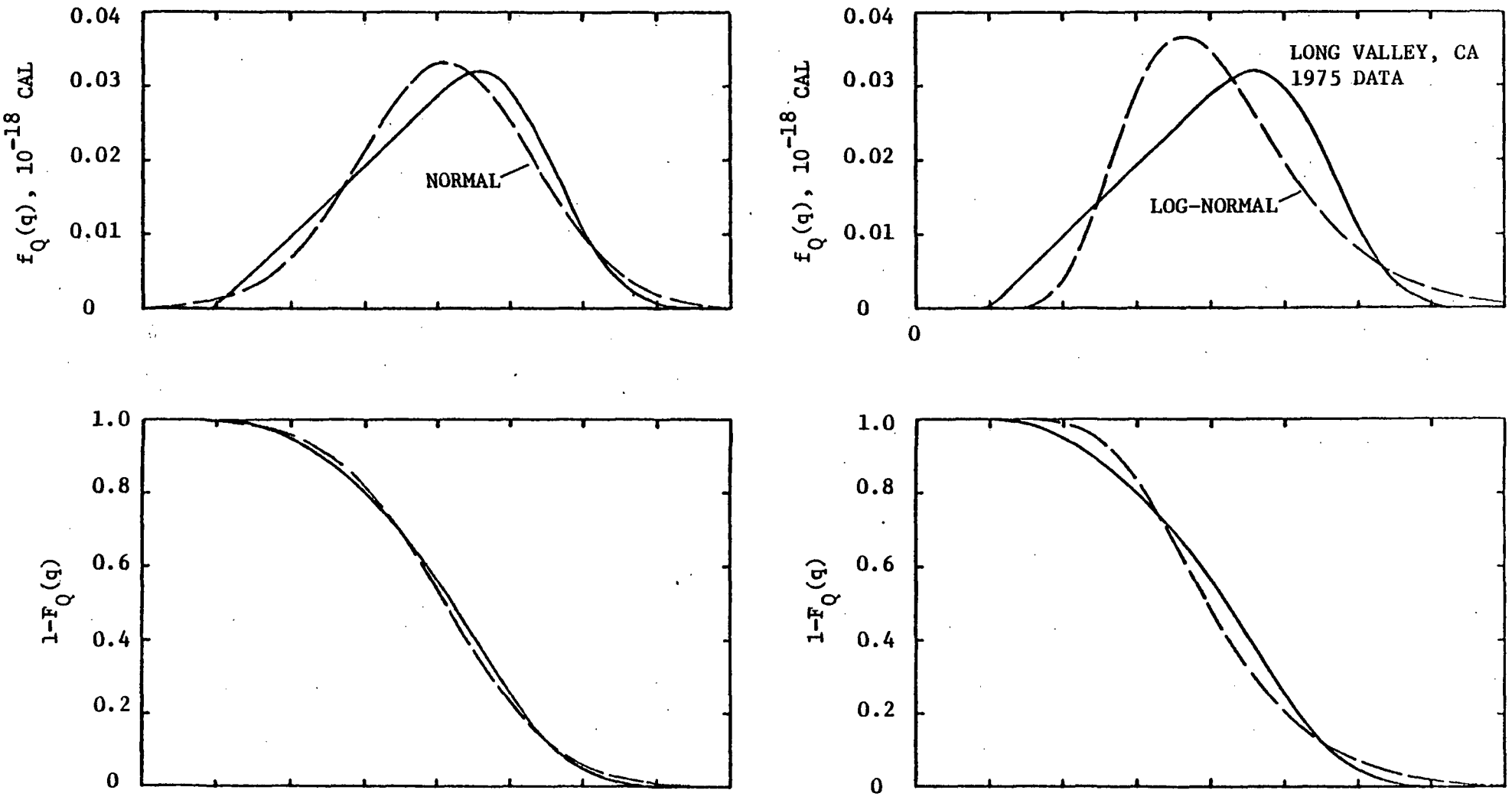

Figure 9.--Probability density and distribution functions of stored energy for hydrothermal system at Long Valley, CA compared to normal and log-normal functions with same mean and standard deviation.

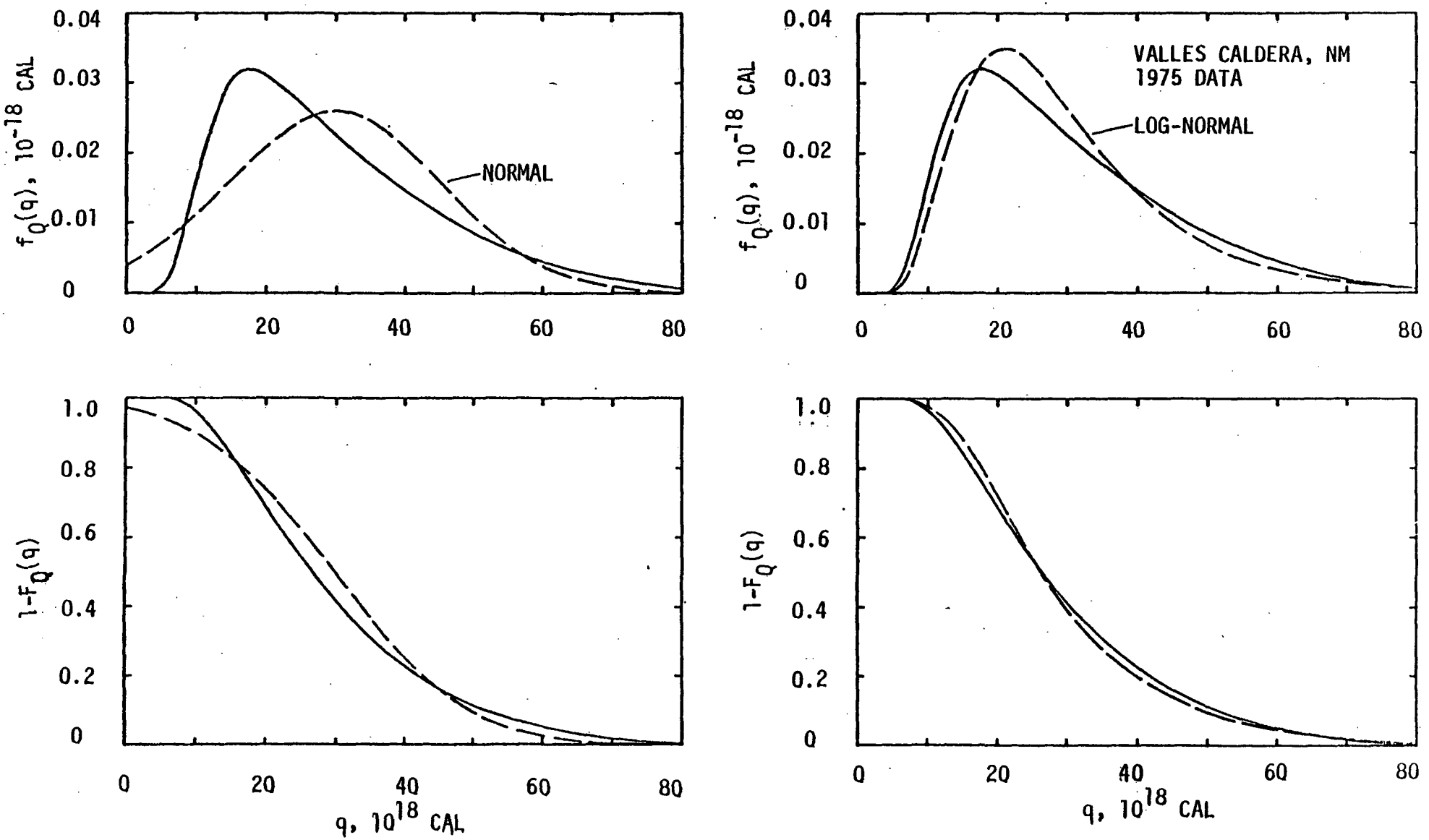

Fig. 10.--Probability density and distribution functions of stored energy for hydrothermal system at Valles Caldera, N. M. compared to normal and log-normal functions with same mean and standard deviation.

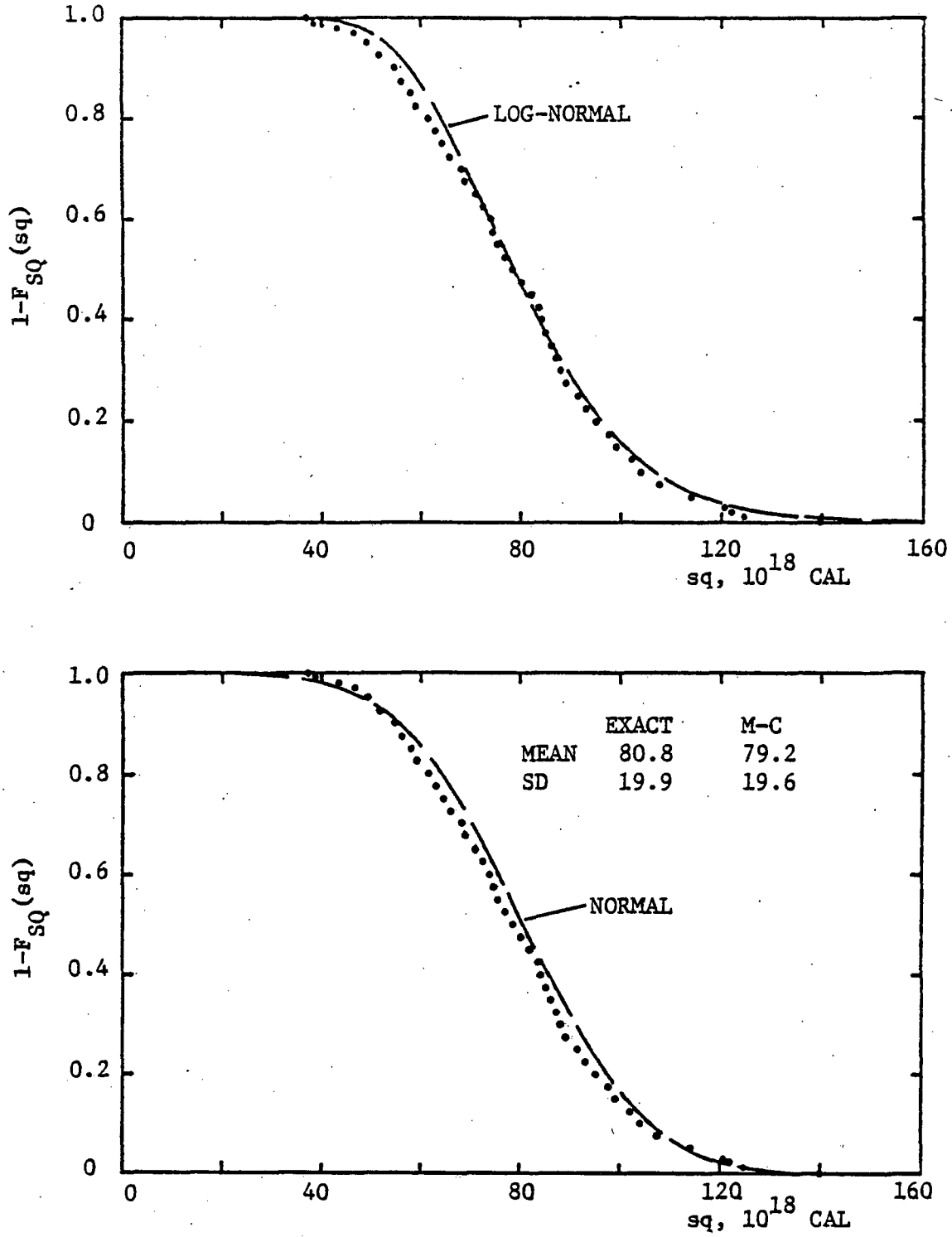

Figure 11. Probability distribution functions for the sum of the stored energy in the hydrothermal systems at Heber, CA, Long Valley, CA, and Valles Caldera, N. M. obtained by Monte-Carlo simulation. Normal and log-normal approximation with same mean and standard deviation.