Chapter 2

METHODOLOGY

2.1 Terminologies used in reserve calculation

Volumetric equations:

$$\text{Oil reserve} = 6.2898 * \frac{A * h * GCF * N/G\_ratio * \phi * (1 - S_W) * RF}{B_{oi}} \quad (1)$$

$$\text{Gas reserve} = 35.3147 * \frac{A * h * GCF * N/G\_ratio * \phi * (1 - S_W) * RF}{B_{gi}} \quad (2)$$

Oil reserve is in stock tank barrels (STB) and gas reserve is in standard cubic feet (scf). Area and thickness are in metric units. The numbers of 6.2898 and 35.3147 are conversion factors to change cubic meters to barrels and cubic feet.

Discovered Hydrocarbons-Initially-In-Place: are those quantities of oil or gas which are estimated, for given data, to be contained in known accumulations, plus those quantities already produced therefrom (after Rose, 2001). They can be OIIP or OOIP or GIIP or OGIP.

Reserves: are those quantities of petroleum which are anticipated to be commercially recovered from known accumulations from given data forward. There are three levels of reserves that are distinguished by levels of uncertainty of geological and engineering data and of economical and political conditions. Proved reserves are those quantities of petroleum which, by analysis of geological and engineering data, can be estimated with reasonable certainty to be commercially recoverable, from a given data forward, from known reservoirs and under current economic conditions, operating methods, and government regulations. If probabilistic methods are used, there should be at least a 90 percent probability that the quantities actually recovered will equal or exceed the estimate. Probable reserves are those unproved reserves, which analysis of geological and engineering data suggests, are more likely than not to be recoverable. In this context, when probabilistic methods are used, there should be at least a 50 percent probability that the quantities actually recovered will equal or exceed the sum of estimated proved plus probable reserves. Possible reserves are those quantities of

petroleum, which analysis of geological and engineering data suggests, are less likely to be recoverable than probable reserves. In this context, when probabilistic methods are used, there should be at least a 10 percent probability that the quantities actually recovered will equal or exceed the sum of estimated proved plus probable plus possible reserves (after Rose, 2001).

Area (A): is the area of the oil-bearing or gas-bearing reservoir bounded by the contour which corresponds to the hydrocarbon-water contact. Area is measured on the structural map of the top reservoir. The structural maps are constructed using seismic and well log data. From one set of data, many possible maps can be created for just one horizon. This is one of the sources that create the random variation of area. In the case that the created map is considered to reasonably show the geological structure, there is still another uncertainty. Figure 2.1 is an example of the uncertainty involved.

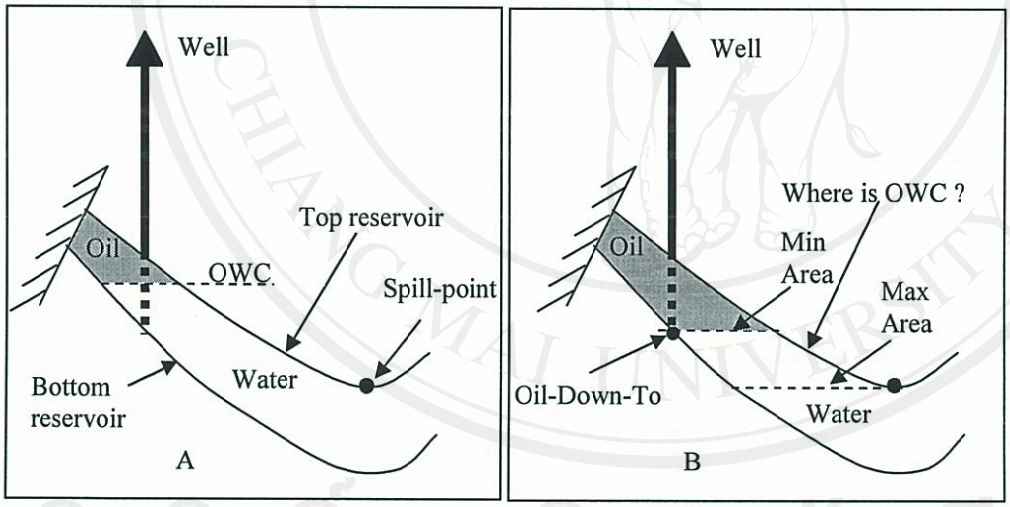

Figure 2.1 Possible extensions of an oil body

If the oil-water contact is penetrated in the well (Figure 2.1A), then the extension of the oil body can be estimated quite correctly. In this case, the value of area is considered to be constant. If the well does not penetrate the oil-water contact, it is difficult to determine the extension of the oil accumulation (Figure 2.1B). In this case, the area inside the contour corresponding to the oil-down-to is normally considered to be the minimum possible area, the area inside the contour corresponding to spill-point is considered to be the maximum possible area. In this

study, there are no structure maps for layer KR2-6 and KR2-7. In order to estimate the areas of these two layers, the structure map of S4 horizon, which is the top of layer KR2-8, was used with the assumption that the top surfaces of KR2-6 and KR2-7 are parallel to the S4 surface. The S4 structure map was shifted upward to KR2-6 and KR2-7 using the average distances from top of those layers to S4 measured in wells. Then the areas on each layer were determined by using the shifted maps (Figure 2.2).

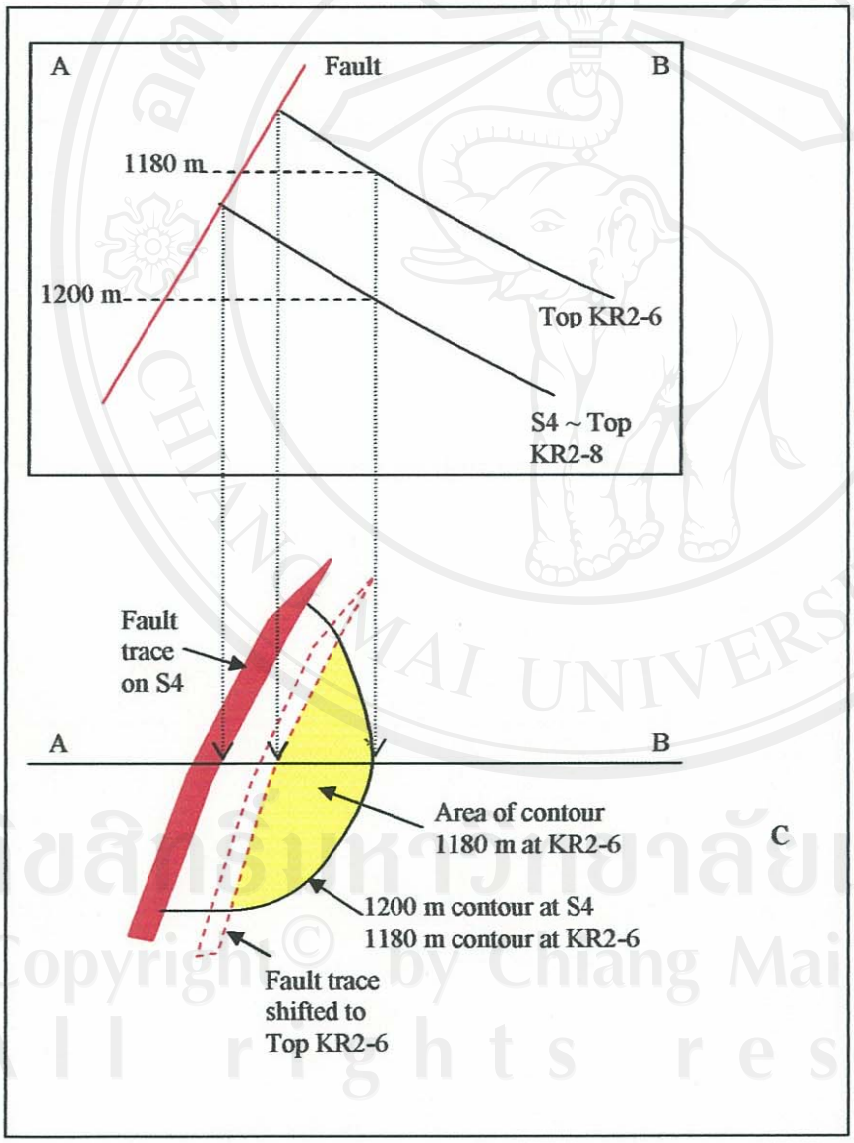

Figure 2.2 Method of shifting faults and contours on a structure map

In case of lack of data, the triangular distribution was normally used to describe the variation of area. In this study, there were only two data points of area, minimum

area and maximum area, which could be determined. The average of these two values, which corresponding to the middle contour between the deepest oil-down-to point and the shallowest water-up-to point or spill-point, is usually used as the limit for the P50 area. The minimum, maximum, and average areas were used to describe the triangular distribution in this report.

Thickness (h): is the true vertical thickness of the reservoir, vertically measured from top to bottom of the reservoir. If the reservoirs are thin, it is not accurate to determine thickness from seismic lines because seismic waves can not define layers thinner than 25-30 m. Thus, the thicknesses of reservoirs are determined from well logs. In this study, after the well correlation was made, the true vertical thicknesses of only sand bodies were selected. The true vertical thickness is calculated by:

$$TVT = MLT\left[\cos(\alpha) \pm \sin(\alpha)\cos(\gamma)\tan(\beta)\right] \quad (3)$$

When the bed is horizontal, equation (3) becomes:

$$TVT = MLT\cos(\alpha) \quad (4)$$

On the structure maps of S3 and S4 the bed dip of the layers is only about five degrees. The correction for this small bed dip is negligible. Thus, bed thickness was calculated by equation (4).

Net to Gross ratio (N/G_ratio): in a heterogeneous reservoir, there may be some tight layers, such as claystone, mudstone, or tight sandstone, that have very low porosity and permeability. Even these layers contain some hydrocarbon, but the interstitial fluids can not flow. It is necessary to eliminate such layers from volumetric calculations. Net to gross ratio is used to do this. This is the ratio between the total thickness of porous layers inside the reservoir and the total vertical thickness of the reservoir. It is expressed in fraction or percentage. Net to gross ratio is normally determined from well logs. In this study, the net to gross ratio was one because only the thicknesses of sand bodies were used. But even within a sandstone layer, the whole layer may not yield hydrocarbons. Some parts of it could be non-reservoirs because of clay contamination. Clay can reduce the pore space that can yield fluids. The remaining parts of the rock unit, which are reservoirs, are the volume of sand that equals one minus the volume of clay:

$$Vss = 1 - Vcl \tag{5}$$

with V_{SS} being the volume of sand in one unit volume of rock and V_{cl} being the volume of clay in one unit volume of rock. So that, the actual reservoir rock volume for sand layers becomes:

$$\text{Reservoir rock volume} = A * h * V_{SS} * GCF \tag{6}$$

In each sand bed, the volume of clay was calculated by using the gamma ray log and the relationship for Tertiary rocks (Rider, 1996):

$$Vcl = 0.083\left(2^{3.7I_{GR}} - 1\right) \tag{7}$$

where: Vcl is volume of clay in one unit of volume of rock

I_{GR} is gamma ray index, calculated by

$$I_{GR} = \frac{GR_{\log} - GR_{\min}}{GR_{\max} - GR_{\min}} \tag{8}$$

Sand volume data were listed in Appendix B.

Geometry Correction Factor (GCF): In the volumetric equation, the product A*h is the volume of a parallelepiped or cylinder geometry. In fact, there are very rare reservoirs with such geometries. Geometric correction factor is used to adjust the calculated volume A*h to the real geometry of the reservoir. This factor is shown in fraction. The product of area, thickness, net to gross ratio, and geometric correction factor is the total rock volume of porous reservoir that bears hydrocarbons. The geometric correction factor is normally derived from a chart based on the shape and the thickness of the reservoir (Figure 2.3). The three layers within the Lower Zone are nearly similar. Thus, the geometric correction factor estimated for one layer can be applied to the others. From the structure maps of the S4 horizon, the ratio of length over width of the structure is about 1, including the missing segment of the structure caused by the faulting (Rose, 2001), and the geometric shape would be type 3. The closure on the S4 structure map is 155 m between the highest contour 1120 m and the closure contour 1275 m. The average thickness of KR2-8 in 10 wells is 7.9 m. The ratio of sand thickness over the height of closure is about 0.05.

From the chart, the geometric correction factor is 0.97 for KR2-8 and also for the KR2-6 and KR2-7.

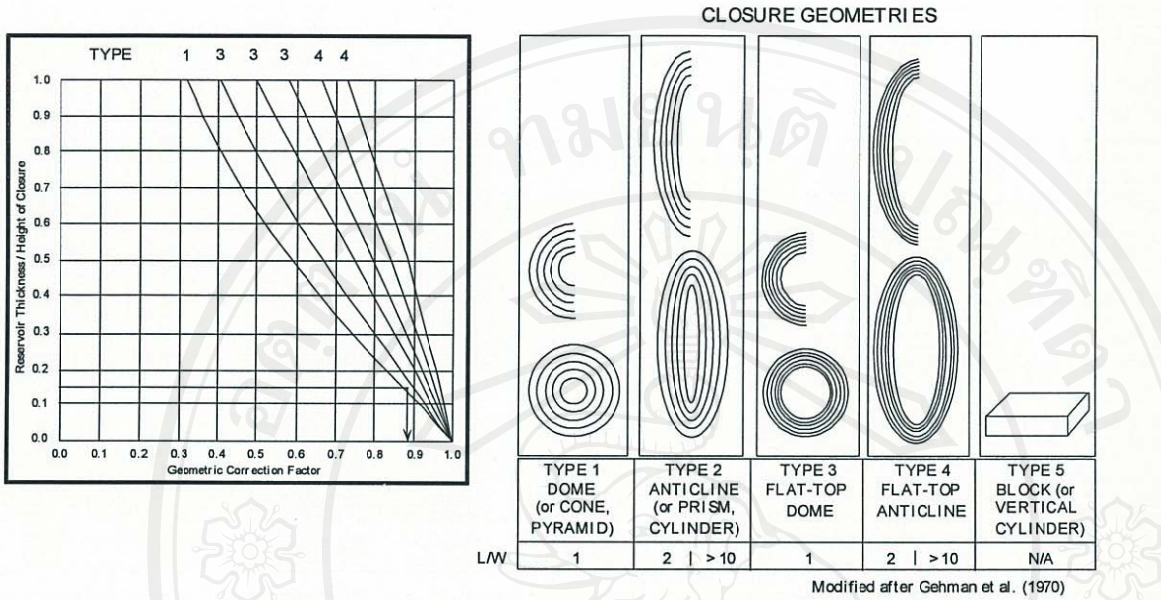

Figure 2.3 Determination of geometric correction factor (after Rose, 2001)

Porosity (Φ): is the ratio between pore volume and the bulk rock volume. It is shown as a fraction or as a percentage. The product of total rock volume and porosity is the total pore volume of the reservoir. Porosity is determined from well logs. Core data, if available, are used to correlate the log-derived values. In this study, due to the lack of core data, the log-derived porosities and water saturations were used directly. Porosity was determined from sonic logs using the Raymer - Hunt equation and from density and neutron logs using Schlumberger's charts after borehole corrections and clay volume correction. For the wells where both density and neutron logs were available, the neutron-density porosity was used; otherwise, porosity from only the sonic, density or neutron log was used.

Sonic porosity was calculated by:

$$\phi_S = \frac{\Delta t - \Delta t_{ma}}{\Delta t_f - \Delta t_{ma}} \qquad (9)$$

Density porosity was calculated by:

$$\phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} \qquad (10)$$

Neutron porosity was read directly on the neutron log and corrected for sandstone matrix using Schlumberger's correction chart.

Neutron-density porosity was calculated by:

$$\phi_{N-D} = \frac{\phi_N + \phi_D}{2} \quad (11)$$

The porosities of each layer in each well are shown in Appendix B.

For each layer, the distributions of porosity used in the simulations were constructed using the fitting tool of the Crystal Ball software based on the log-derived porosity data. Some distribution types, such as lognormal and logistic distributions, vary from minus infinity to plus infinity. If these distributions are used to describe some parameters, such as porosity and saturation, they need to be truncated by the constraints of the parameters which they describe. For example, saturation, net to gross ratio, and V_{ss} always vary from zero to one. Therefore, at least the ranges from minus infinity to zero and from one to plus infinity of the distributions must be truncated because it is impossible for those ranges to occur in nature. Within a reservoir, those parameters normally vary in a certain range but not the whole range from zero to one. In this study, the distribution curves of porosity and water saturation were truncated using the actual range of the data.

Saturation: is the volumetric proportion of one liquid in the total pore volume of the reservoir. In an oil zone or a gas zone, there are both hydrocarbons and water co-existing. Thus, one zone has two saturations: oil saturation (S_o) or gas saturation (S_g) and water saturation (S_w). From well logs, only water saturation can be calculated. In order to calculate oil saturation and gas saturation, the following formulae are used:

$$S_o = 1 - S_w \quad (12)$$

$$S_g = 1 - S_w \quad (13)$$

In a laboratory, oil and water saturations are measured directly from core samples. But these measured values are not correct due to the contamination of drilling mud and the loss of water and oil during sampling. Thus, the log-derived water saturations are normally used. Water saturation is determined from spontaneous potential and resistivity logs using the Archie equation:

$$S_W = \sqrt[n]{\frac{F.R_W}{R_t}} \quad (14)$$

F is calculated by the equation, which is the best average for sandstones (Rider, 1996):

$$F = \frac{0.62}{\phi^{2.15}} \quad (15)$$

A water resistivity, R_W, of 0.8 ohm meter at surface temperature used to calculate the water saturation was measured in the laboratory of PTTEP.

For the wells where deep, shallow, and micro resistivity, density, and neutron logs are available, the dual water model was used to calculate water saturation. In wells UT1-7/D5, UT1-7/D6, and UT1-7/D7, the log data were not enough. That led to unconfident results for water saturation. Then, the water saturations of these wells were not used in the study.

The water saturations of each layer in each well are shown in Appendix B.

For each layer, the distributions of water saturation used in the simulations were constructed by using the fitting tool of the Crystal Ball software based on the log-derived data and then truncated by the actual range of the data.

Initial oil or gas formation volume factor (B_{oi}, B_{gi}): presents the change in volume when one unit volume of oil or gas at reservoir pressure and temperature is brought to the surface. It is the ratio between the volume of oil or gas at reservoir conditions and one unit volume of oil or gas at standard surface conditions. B_{oi} and B_{gi} are measured in PVT laboratories. In thin, small reservoirs, where the reservoir pressure and temperature and the oil compositions do not vary much, B_{oi} can be considered to be constant throughout the reservoir. For the oil from the U-Thong field, the measured B_{oi} was 1.11 bbl/STB at reservoir conditions.

Recovery factor (RF): is the ratio between the ultimate amount of produced hydrocarbons and the hydrocarbons initially in place in the reservoir. This factor represents the percentage of the hydrocarbons initially in place in the reservoir that

can be recovered. It can be expressed as a fraction or as a percentage. The recovery factor for an oil field and a gas field are defined as:

$$RF = \text{Recovered oil volume/ OIIP} \quad (16)$$

$$RF = \text{Recovered gas volume / GIIP} \quad (17)$$

Recovery factor is affected by two main elements: natural factors, such as reservoir heterogeneity, aquifer size, pore geometry, and fluid properties, and artificial factors, such as abandonment pressures, which is the pressure at which the natural energy of the field is considered to be exhausted, abandonment rates, and well spacing. Estimation of recovery factor is the most difficult task. Until now, all the oil fields in Thailand are still producing. Therefore, the ultimate recovery factors of other fields are not available for analogy. In PTTEP's calculation, the reserves of the U-Thong field were calculated with an assigned recovery factor of 0.3 for the lower zone. This value was used in this study to check the accuracy of the Monte Carlo simulation in comparison to the volumetric method.

2. 2 Monte Carlo Simulation

The Monte Carlo method was originally developed by Nicholas Metropolis and Stanislaw Ulam while working on the Manhattan Project in Los Alamos, New Mexico, U.S.A. during the time of World War II as a way to model the random behavior of sub-atomic particles. Since then, the theory has been used by numerous industries and for countless purposes to solve deterministic problems through the use of random numbers. By the early 1970s petroleum engineers began to use the technique to model reserves estimates. The Monte Carlo method has two requirements. The first is a mathematical model, or simulator. The second is the knowledge of the probabilistic cumulative distribution functions (CDFs) of the variables to be fed into the mathematical model. When the CDFs are known, each variable needed in the model is randomly sampled and the model is used to calculate the unknown quantity. This process, known as a trial, is repeated many times until a sufficient number of trials have been made to create a distribution of the unknown quantity. The process of performing an adequate number of trials is called a Monte Carlo simulation.

2. 2.1 The mathematical model

For the recoverable hydrocarbon volume, or reserves, approach using Monte Carlo simulation, the mathematical model used was the volumetric equations (1) and (2) mentioned earlier.

In the case of a multi-layer field, the mathematical model is:

$$\text{Field reserves} = \Sigma\,(\text{Zone reserves})_i \qquad (18)$$

The term $(\text{Zone reserves})_i$ is the reserves of production zone i^{th} calculated by either equation (1) or (2). The distribution of the total field reserves may be calculated by:

$$\text{Distribution of field reserves} = \Sigma\,\text{Distributions of } (\text{Zone reserves})_i \qquad (19)$$

In this study, the Crystal Ball 2000 Professional package, version 5.0, was used to perform the Monte Carlo simulation for hydrocarbon reserves. The Crystal Ball can sample the distributions of the input variables and perform the mathematical calculations as described by the model. To do this, the program generates a random number between 0 and 1. Once the random number has been generated, the program reads the corresponding data value from the CDF plot and inserts it into the model. For each trial, a random number is generated and the corresponding parameter value is determined for each of the seven input variables required in the volumetric model. These seven input variables are area, thickness, net to gross ratio, porosity, water saturation, hydrocarbon formation volume factor, and recovery factor. The process is repeated for each trial. A simulation can be a few hundreds to a few thousands trials. As the number of trials increase, the results represent more reliable probability distributions but it becomes a time consuming procedure. However, the time needed to run the simulation depends largely on the capacity of the computer used.

The CDF of each parameter in the right-hand side of the volumetric equation is fed into the model.

2. 2. 2 Probability Distributions

Some basic terminologies concerning probability distributions are:

- Statistics: is a set of mathematical methods for collecting, organizing, and interpreting data, as well as drawing conclusions and making reasonable decisions based on such analysis.
- Population: is a collection of a finite number of measurements or virtually infinitely large collection of data about something of interest.
- Sample: is a representative subset selected from the population. A good sample must reflect the essential features of the population from which it is drawn.
- Random sample: is a sample in which each member of a population has an equal chance of being included in the sample.
- Random variable: is a form of presenting any unsampled, or unknown, value z, the probability distribution of which models the uncertainty about z. The variable can be continuous or discrete.
- Probability function of a random variable Z: is a mathematical function that assigns a probability to each realization z of the random variable Z: P(Z=z).
- Expected value, EV, or mean: is the probability-weighted sum of all the possible occurrences of the random variable Z.

$$EV(Z) = m = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \tag{20}$$

There are several ways to calculate the mean from a cumulative distribution curve:

- Arithmetic mean = $\frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i}$ (21)
- True statistical mean = $e^{\left(\mu_{\ln z} + 0.5\sigma_{\ln z}^{2}\right)}$ (22)
- Swanson's mean = 0.3 * P90 + 0.4 * P50 + 0.3 * P10 (23)

This mean is good approximation for low-medium variance distributions.

- Variance of the random variable Z: is defined as the expected squared deviation of Z about its mean.

$$Var\{Z\} = EV\{Z-m\}^2 = EV\{Z^2\} - m^2 \geq 0 \quad (24)$$

- Standard deviation: is the square root of the variance.

$$SD = \sqrt{Var\{Z\}} \quad (25)$$

- Median: is the value of variable Z at the cumulative probability of 50 percent.
- Mode: is the value of variable Z which has highest frequency on a probability density curve.
- Cumulative distribution function, CDF: is defined as:

$$F(z) = Prob\{Z \leq z\} \in [0,1] \quad (26)$$

This formula gives the area under the probability density function of the random variable Z, and is the probability that the random variable Z is less than or equal to a threshold value of z. Then the probability of exceeding any of the threshold values of z can be written:

$$Prob\{Z>z\} = 1 - F(z) \quad (27)$$

All CDFs are monotonically increasing and continuous to the right. Monotonically increasing means that as z increases, F(z) must not decrease. Continuous to the right means that small changes in z result in small changes in F(z). Additionally, for all CDFs, as z goes to infinity, F(z) approaches the value of 1 and as z goes to negative infinity, F(z) approaches the value of zero.

When reservoir parameters are analyzed as random variables and the distributions of their values are created, it is possible to make statistical statements about the variables. For example, it might be said of a given reservoir that 90 percent of the porosity values measured will fall below 30 percent. This comes directly from the CDF. In this case, at the point where F(z) = 0.9, the porosity equals 30 percent. Similar readings of the CDF

can be made and the corresponding values of porosity can be determined to have a certain probability. In the example above, a porosity of 30 percent is considered the P90 value. It is common to note the P10, P50, and P90 values for distributions since they give a good representation of the entire distribution with just three numbers. By definition, P50 is equal to the median value of the distribution. Figure 2.1 gives an example of CDF.

- Probability density function, PDF: is the derivative of the CDF and can be written:

$$f(z) = F'(z) = \lim_{dz \to 0} \frac{F(z+dz) - F(z)}{dz} \tag{28}$$

Inversely, CDF can be obtained from integrating the PDF:

$$F(z) = \int_{-\infty}^{z} f(z)dz \tag{29}$$

The PDF displays the same information as the CDF, but in a different format. A familiar type of PDF is the histogram. PDF relates the probability of occurrence of the variable f(z) for a small range of z. The variable z should not be thought of as a fixed number, but as a small range of values because the probability f(z) that z is any one exact number is zero. For all PDFs, $f(z) \geq 0$ for all z. Additionally, the sum of the values of f(z) over the infinite range, also the area under the PDF curve, must be equal to one. Figure 2.4 gives an example of PDF.

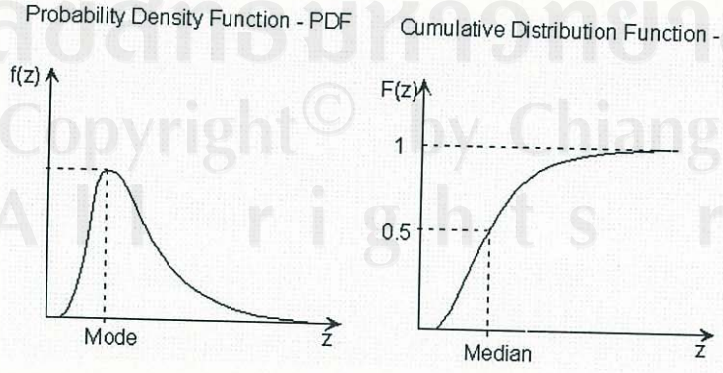

Figure 2.4 Example of PDF and CDF

Depending on the data, there are many different types that the CDF and PDF can be.

Mathematicians have developed models of the CDF and PDF for certain distribution types that frequently occur in applications. The following are some examples of distribution types.

Normal Distribution (or Gaussian distribution)

The normal distribution has a PDF given by the following equation:

$$f(z)=\frac{1}{SD\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{z-m}{SD}\right)^2\right] \tag{30}$$

The standard normal PDF $f_0(z)$ has a mean of zero and a standard deviation of one:

$$f_0(z)=\frac{1}{\sqrt{2\pi}}\exp\left[-\left(\frac{z}{2}\right)^2\right] \tag{31}$$

The CDF of the normal distribution F(z) is written as:

$$F(z)=\int_{-\infty}^{z} f(z)dz = F_0\left(\frac{z-m}{SD}\right) \tag{32}$$

The CDF of the standard normal distribution $F_0(z)$ is written as:

$$F_0(z)=\int_{-\infty}^{z} f_0(z)dz \tag{33}$$

The expected value of the normally distributed random variable is equal to the mean and also to the median. The normally distributed PDF and CDF are symmetrical in shape (Figure 2.5).

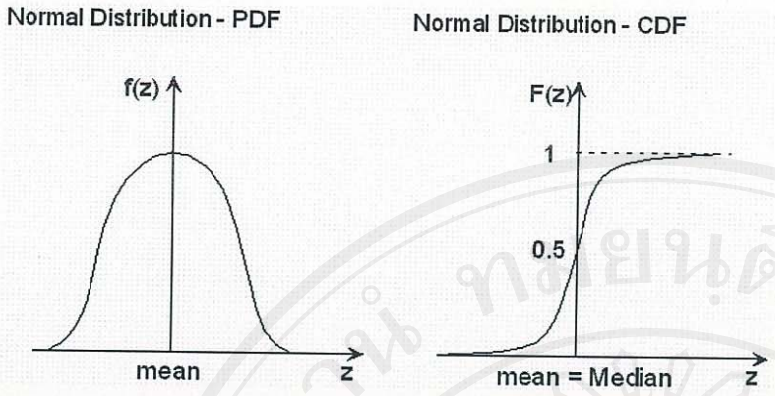

Figure 2.5 Normal PDF and CDF

Lognormal Distribution

The lognormal distribution is another common distribution in which the logarithm of the random variable is normally distributed. The lognormal distribution is frequently associated with processes that divide large quantities into smaller ones. This is exactly the type of process that occurs during sediment deposition. Therefore, lognormal distributions are common in the petroleum industry. The PDF for the lognormal distribution is:

$$f(z) = \frac{1}{z\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z - m_{\ln z}}{\sigma}\right)^2\right] \tag{34}$$

and the CDF is given by:

$$F(z) = \frac{1}{z\sigma\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right] dz \tag{35}$$

When drawn on probability paper in which the x-axis has a logarithmic scale, lognormal distributions plot as straight line.

Figure 2.6 shows the shape of the PDF and CDF curves for a lognormal distribution. The PDF of a lognormal distribution is said to be positively skewed, as is seen by the large value of f(z) for small values of z and a long tail extending to the right.

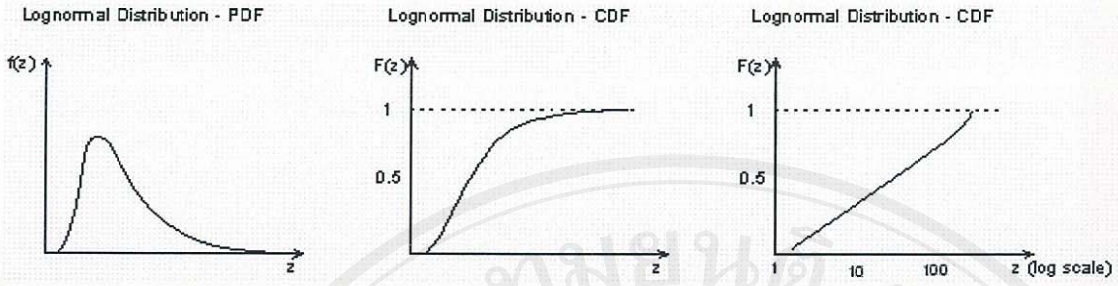

Figure 2.6 Lognormal PDF and CDF and lognormal CDF plotted on probability paper

Note that a lognormal distribution will be a straight line on probability paper in which the y-axis is a linear scale and the x-axis is a logarithmic scale.

Triangular Distribution

The triangular distribution is useful when limited data are available. Data are distributed between a minimum, a maximum, and the likeliest value. For an oil field, the number of wells drilled limits the number of reliable data points. Therefore, the triangular distribution is used to describe some reservoir data such as reservoir thickness. The equations for the PDF are:

$f(z) = 0$ $\quad z < a$ (36)

$$f(z) = 2(z-a)\frac{1}{(c-a)(b-a)} \quad a \le z < b \quad (37)$$

$$f(z) = 2(c-z)\frac{1}{(c-a)(c-b)} \quad b \le z < c \quad (38)$$

$f(z) = 0$ $\quad c \le z$ (39)

The CDF is given by the following:

$F(z) = 0$ $\quad z < a$ (40)

$$F(z) = (z-a)^2\frac{1}{(c-a)(b-a)} \quad a \le z < b \quad (41)$$

$$F(z) = 1-(c-z)^2\frac{1}{(c-a)(c-b)} \quad b \le z < c \quad (42)$$

$F(z) = 1$ $\quad c \le z$ (43)

where a is the minimum, b is the likeliest, and c is the maximum possible value of z. An example of triangular distribution is shown in Figure 2.7.

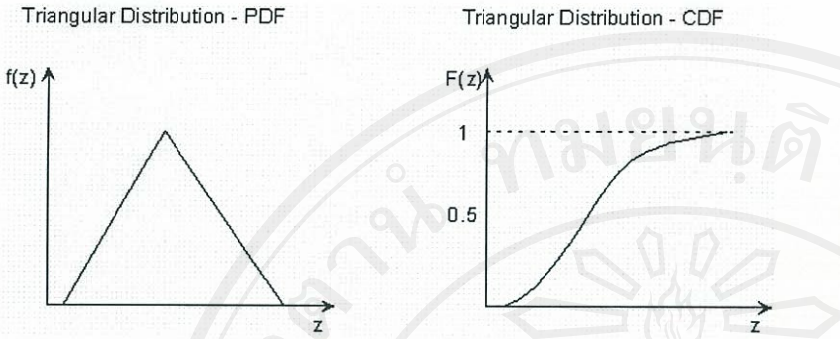

Figure 2.7 Triangular PDF and CDF

Uniform Distribution

The uniform distribution is a two-parameter distribution and is the simplest of all distributions. It is useful when little is known about the distribution of the subject parameter other than a maximum and a minimum. The PDF of this distribution is given by:

$f(z) = 0 \qquad z < a \qquad (44)$

$f(x) = 1/(b-a) \qquad a \leq z < b \qquad (45)$

$f(z) = 0 \qquad b \leq z \qquad (46)$

The CDF is:

$F(z) = 0 \qquad z < a \qquad (47)$

$F(z) = (z-a)/(b-a) \qquad a \leq z < b \qquad (48)$

$F(z) = 1 \qquad b \leq z \qquad (49)$

where a is the minimum possible value and b is the maximum possible value. Figure 2.8 shows the shape of the PDF and CDF for the uniform distribution.

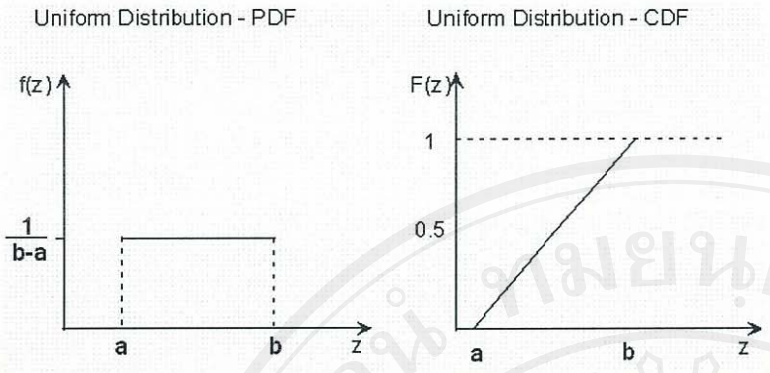

Figure 2.8 Uniform PDF and CDF

Central Limit Theorem

A very important and useful concept in statistics is the central limit theorem. There are three things that are important in understanding about any distribution:

1) The location of its center

2) Its width

3) Its distribution

The central limit theorem helps to approximate all three.

The central limit theorem states that as sample size increases, the sampling distribution of sample means approaches that of a normal distribution with a mean the same as the population and a standard deviation equal to the standard deviation of the population divided by the square root of the sample size.

The central limit theorem is simply that the sum of a great number of independent equally distributed standardized random variables tends to be normally distributed.

The corollary of the central limit theorem states that the product of a great number of independent, identically distributed random variables tends to be lognormally distributed.

The model used for hydrocarbon reserves estimation is the result of multiplying the distributions of area, thickness, porosity, water saturation, and recovery factor together. Thus, this model would be expected to be approximately lognormal according to the corollary.

2.2.3 Creating probability distributions of the parameters, PDFs and CDFs

In Monte Carlo simulation, the program uses PDFs for calculations. The first duty of the estimator is to create CDFs for the data.

To create the CDF of a set of data, the following steps are taken. First, the data are sorted from smallest to largest, Zi=1 to Zi=N. Then, the corresponding value of $F(z)_i$ is calculated by the following formula:

$$F(z)i = (Ni - 0.5) / N \quad (50)$$

where Ni is the rank of the data point and N is the total number of data measurements. A plot of F(z) versus Zi yields the empirical CDF plot. A CDF may be created manually or by computer programs.

To calculate the empirical PDF, representative ranges of values for the data, or bins - Az, are created. The sizes of the bins are equal and are in general determined by the formula:

$$Az = 5(ZN - Z1)/N \quad (51)$$

The data are then examined to determine the number of occurrences of values that fall into each bin. These occurrences are tabulated and then the value of f(z) is calculated by dividing the number of occurrences in each bin by the total number of data points. A plot of f(z) versus Zi yields the empirical PDF plot. A PDF can also be created either manually or by computer programs.

After the CDFs and PDFs are created, the best fit types for these functions to input into the simulator need to be chosen. There are some ways to evaluate the empirical CDFs and PDFs to find out what types of distribution they are. The first is a visual inspection and comparison to the idealized distribution CDF and PDF shapes. The second, and more reliable, is to plot the data on probability paper. For a lognormal distribution, the logarithm of the data forms a straight line on probability paper.