

# Modeling petroleum resources in provinces of the Former Soviet Union

**Roberto F. Aguilera**

*Centre for Research in Energy and Minerals Economics (CREME),  
Curtin University, GPO Box 1987, Perth 6845, Australia  
Email: roberto.aguilera@curtin.edu.au*

(Received 18 February 2011; accepted 6 June 2011)

## Abstract

This paper estimates petroleum endowment volumes for provinces of the Former Soviet Union (FSU) that have not been previously assessed by other organizations. The study uses the United States Geological Survey World Petroleum Assessment (USGS, 2000) as a starting point. It then utilizes nonlinear regression to estimate parameters of a Variable Shape Distribution (VSD) model that calculates the total petroleum endowment throughout the FSU. Earlier size distribution models used to evaluate unassessed petroleum resources relied mainly on the fractal and lognormal distributions. In fact, all the methods used historically have been based on an assumed form of the size distribution of nature's endowment of petroleum resources. The VSD model is different in that it allows the actual petroleum resource data from USGS (2000) to determine the form of the size distribution of petroleum resources. The model is validated by a good fit of actual data, supported by coefficients of determination ( $R^2$ ) equal to 0.98 or greater. It is concluded that there is a large petroleum endowment in the FSU that will last for several decades and can contribute significantly to domestic energy needs as well as export requirements.

**Keywords:** Conventional petroleum, Former Soviet Union, Petroleum assessment

## 1. INTRODUCTION

The USGS (2000) study provides conventional petroleum endowment volumes for provinces of the FSU that are thought likely to contribute significant recoverable volumes over a 30 year time horizon.<sup>1</sup> As defined by the USGS, endowment volumes are equal to known plus undiscovered volumes (Fig. 1), while conventional petroleum is the summation of conventional oil, gas, and natural gas liquids (NGL). The USGS

---

<sup>1</sup>A province is a USGS-defined "area having characteristic dimensions of perhaps hundreds to thousands of squared kilometers encompassing a natural geologic entity (for example, sedimentary basin, thrust belt, delta) or some combination of contiguous geologic entities."

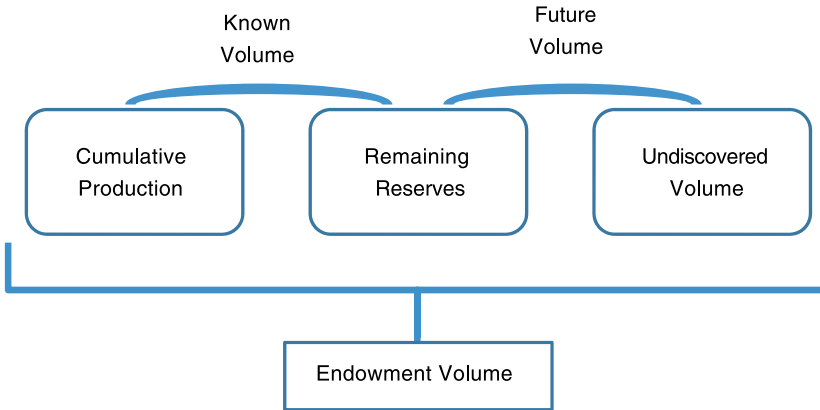


Figure 1. The relationship between cumulative production, remaining reserves and undiscovered volumes, based on definitions from USGS (2000).

(2000) study includes estimates of oil and NGL endowment volumes for 18 provinces, as well as natural gas endowment volumes for 31 provinces. In total, the USGS states that the FSU region can be divided into 141 provinces. Thus, the oil and NGL volumes in 123 provinces, and gas volumes in 110 provinces, are not estimated in the study. This paper provides endowment volumes for the unassessed provinces using a size distribution model called the Variable Shape Distribution (VSD) model. The VSD is a statistical method known as size distribution analysis that has gained success for many decades as a simple and effective complement to more complex and time-consuming geological methods. When various assessment methodologies converge in their results, a higher degree of confidence in those results is provided. Although geological methods are the most dependable means of estimating endowment volumes, size distribution models have proven useful as they are” applicable to different types of data including accumulations, fields, and reservoirs, over varying regions from plays, to petroleum systems to petroleum provinces, to the world” (Barton, 1995).

The focus on FSU is partly due to the large number of unassessed provinces in the region, as well as the large endowment volumes thought to exist in both assessed and unassessed provinces.

The selection by the USGS of 18 oil and NGL provinces, and 31 gas provinces, represents a data generation process and affects the ultimate estimate of the endowment volumes in 141 provinces. As stated by USGS (2000), “the assessed areas were those judged to be significant on a world scale in terms of known petroleum volumes, geologic potential for new petroleum discoveries, and political or societal importance.”

The petroleum endowment volumes estimated by the USGS (2000) study do not include offshore conventional oil deeper than 2,000 meters in some cases and 4,000 meters in others. Nor do the volumes include unconventional oil such as heavy oil, oil sands and shale oil. Similarly, the natural gas endowment does not include unconventional gas, such as tight gas, shale gas, coalbed methane and natural gas hydrates. It also does not include offshore gas deeper than 2,000 meters in some cases and 4,000 meters in others.

**Table 1. Oil, natural gas and NGL endowment in FSU (without including unconventional oil and unconventional gas).**

	<b>Oil Endowment</b> Billion Barrels	<b>Gas Endowment</b> Trillion Cubic Feet	<b>NGL Endowment</b> Billion Barrels
<b>Former Soviet Union</b>			
USGS estimate for assessed provinces	380	3,716	71
VSD estimate for assessed provinces	379	3,714	71
VSD estimate for assessed + unassessed provinces	556	4,815	163
VSD estimate for assessed + unassessed provinces reserve growth	795	7,507	254

Table 1 shows the conventional oil, gas and NGL endowments for FSU without including unconventional oil or gas. The oil endowment published by USGS (2000) in FSU is 380 billion barrels of oil equivalent (bboe). A very similar volume, 379 bboe, is obtained with the use of the VSD model. The natural gas endowment as estimated by the USGS is 3,716 trillion cubic feet (tcf). This compares well with 3,714 tcf calculated with the VSD model. Based on the good comparison, the VSD is used to estimate total endowments, including provinces not assessed by the USGS. This leads to an oil estimate of 556 bboe, a gas estimate of 4,815 tcf, and a NGL estimate of 163 bboe.

The volumes are very large and, with additional research and development, may help for several decades to meet the energy needs of the FSU and surrounding regions. The following sections describe the procedures used to obtain the volumes presented in Table 1.

## 2. EARLIER SIZE DISTRIBUTION METHODS

The lognormal and Pareto distributions are common size distribution models used to estimate petroleum volumes of unassessed areas.<sup>2,3</sup> Some researchers believe that the distribution of nature's petroleum resources follow lognormal distributions (Kaufman, 2005). However, several researchers claim that the lognormal distribution provides overly pessimistic results (Drew, 1997). They observe that, with additional exploration, there is an on-going discovery process that can better be modeled with a Pareto distribution. The difference between the two distributions can be seen in Figure 2, where they are shown as density distributions.

Over the past five decades, "scientists have engaged in honest debate about the form of the distribution. In the USGS, thinking has evolved from regarding the entire natural

<sup>2</sup>The log-normal distribution is a continuous distribution in which the logarithm of a random variable is normally distributed.

<sup>3</sup>The Pareto distribution is also known as the power law, Bradford, hyperbolic, fractal, scaling, Zipf (when the slope is 1.0), log-geometric, and J-shape distributions. The Pareto distribution, named after the Italian economist Vilfredo Pareto, is a power law distribution, where the exponent of the power law is constant.

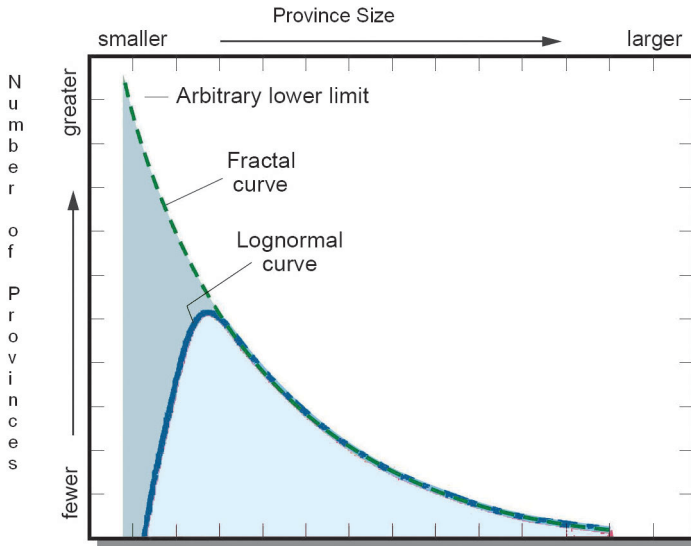


Figure 2. Density distributions of number of provinces versus province size (adapted from Barton, 1995).

population as lognormal to recognizing that a fractal or power-law distribution better predicts the increasing numbers of small accumulations” (Barton, 1995). More recently, it has generally been acknowledged (Aguilera, 2006) that the Pareto distribution tends to overestimate petroleum resources, while the lognormal distribution tends to underestimate them.

These types of statistical distributions are believed to be representative of many natural and social occurrences (e.g. resource distribution in nature; income distribution across population). Given that the subdivision of FSU by the USGS is based on natural geological features, theory would indicate that methods like the VSD model are dependent on this type of natural disaggregation.

## 2.1. The pareto distribution

Xiao *et al.* (2011) also indicates that the Pareto (i.e. fractal) distribution could accurately model the size distribution of petroleum resources.<sup>4</sup> The Pareto distribution specifies that a log-log plot of the cumulative number of discovered petroleum accumulations versus the size of the accumulation could result in an approximate straight line with a constant negative slope. The straight line is generally observed for the larger fields, as shown by the solid circles in Figure 3. The slope of the dashed straight line is known as the shape parameter of the Pareto distribution. The open circles represent discovered (known) volumes of oil and gas below levels that are presently economically viable. The area under the curve is the cumulative volume of discovered oil and gas. The straight line is extrapolated to an arbitrary minimum volume to calculate the undiscovered oil and gas, which is equal to the area between the straight line and the open circles.

<sup>4</sup>The Pareto distribution is the probability distribution characteristic of fractals. A fractal is a structure that contains an organized arrangement of repeating patterns over many ranges of scale. In fractals, the part is reminiscent of the whole. Some fundamental properties of fractals are self-similarity, self-affinity and scale invariance.

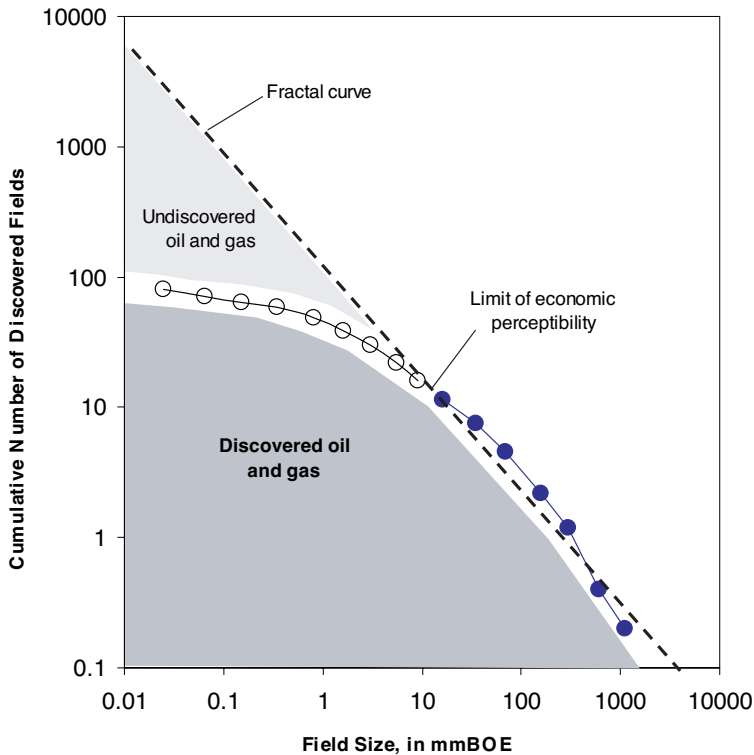


Figure 3. Cumulative number of discovered oil and gas fields versus size of the field (adapted from Barton, 1995).

### 3. VARIABLE SHAPE DISTRIBUTION (VSD) MODEL

Historically, all the methods used to forecast oil and gas volumes have been “based on an assumed form of the size-frequency distribution of the natural population of oil and gas accumulations” (Barton, 1995). The VSD model is different in that we start by observing the curvature (on a log-log plot) given by the size and number of assessed FSU provinces from USGS (2000). We then develop the VSD model which allows the data to determine the relationship between the size and number of provinces. Next, the model is extended out of sample such that (1) the right end point of the curve does not change (i.e. the largest province is fixed and assumed to be found) and (2) the smallest province ( $V_m$ ) is also constant and known. As with all size distribution models, the original sample used to estimate the parameters contains the largest data in terms of endowment volumes. This allows us to estimate the slope and intercepts of the straight line given by the largest provinces.

The VSD model has been used with success, typically with coefficients of determination ( $R^2$ ) equal to or greater than 0.98, for fitting available endowment data from conventional oil, natural gas and NGL in provinces evaluated by the USGS (Aguilera, 2006). The good matches allow for forecasting of the petroleum endowments in provinces not evaluated in the past.

Equation 1 presents the VSD as a nonlinear least squares (NLS) model. In particular, the problem is:

$$\min_{\{V_x, a_p, V_s, \psi, S\}} \sum_{i=1}^n (V_i - \hat{V}_i)^2 \tag{1}$$

Subject to:

$$\hat{V}_i = \frac{\left\{ \left[ \left( \frac{1}{N_t} - \left( \frac{V_m}{V_x} \right)^{\left( \frac{\log N_x - \log N_m}{\log V_x - \log V_m} \right)^{a_p}} + \frac{V_m}{V_x} \right) \cdot V_x \right] \times (\psi) \right\}}{(\psi) + [1 - (\psi)] \cdot \left[ 1 - \exp \left\{ - \left[ \left[ \left( \frac{1}{N_t} - \left( \frac{V_m}{V_x} \right)^{\left( \frac{\log N_x - \log N_m}{\log V_x - \log V_m} \right)^{a_p}} + \frac{V_m}{V_x} \right) \cdot V_x \right] \div V_s \right] \right\}^S} \tag{2}$$

Where:

- $a_p$  - slope of straight line approximated from USGS sample points (same as slope of Pareto distribution).
- $N_m$  - minimum number of USGS provinces (= 1).
- $N_t$  - cumulative number of provinces.
- $N_x$  - maximum number of provinces.
- $S$  - severity exponent that controls the steepness of the slope of the estimated VSD curve where it separates from the Pareto straight line (on the right tail of the distribution, typically near the largest volumes).
- $V_m$  - minimum USGS province volume.
- $V_s$  - approximate volume at which the USGS data begins to deviate from the Pareto straight line (on the right tail of the distribution, near the largest volumes).
- $\hat{V}_i$  - estimated volume of a province.
- $V_x$  - maximum volume (boe) given by the Pareto straight line (at  $N_m = 1$ ). The actual maximum volume could be larger, equal to, or smaller than  $V_x$ .
- $\psi$  - separation ratio that controls the amount of separation between the Pareto straight line and the estimated VSD curve (on the right tail of the distribution, near the largest volumes).

Equation 1 shows five parameters of the VSD model that are estimated with nonlinear regression:  $V_x$ ,  $a_p$ ,  $V_s$ ,  $\psi$ , and  $S$ . These parameters are used to obtain the best possible fit of the sample of FSU provinces for which petroleum endowment data from USGS (2000) exists. The parameters are estimated by examining the coefficients of determination ( $R^2$ ), comparing the USGS and VSD-calculated endowments, and visually inspecting the curves.

The same parameters used for obtaining a good fit of the provinces evaluated by USGS (2000) are used to forecast the oil, gas and NGL endowments of unassessed provinces in the region. Appendix A provides the mathematical development of Equations 1 to 2 and explains the nomenclature in detail.

### 3. FORMER SOVIET UNION

#### 3.1. Oil endowment

Figure 4 is a plot of cumulative number of provinces versus size of provinces. The

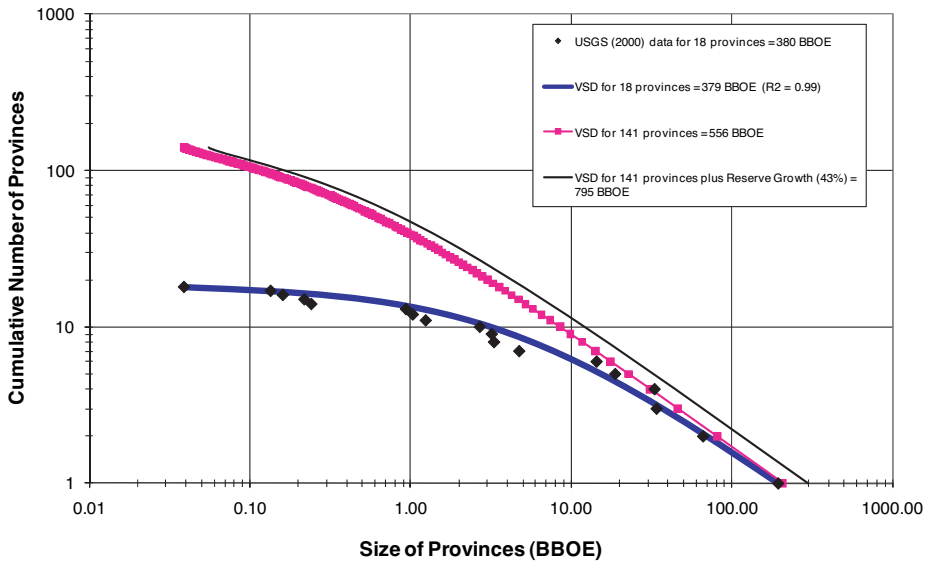


Figure 4. Conventional oil endowment in FSU.

black diamonds represent endowments of conventional oil published by the USGS (2000) for 18 provinces in FSU - equal to 380 bboe. A similar volume, 379 bboe, is calculated by the VSD model using the parameters presented in the second column of Table 2. The VSD data are represented by the thick continuous solid line in Figure 4. Visual inspection shows a good fit supported by an  $R^2$  equal to 0.99. Given the good fit, the same model and parameters are used for estimating the oil endowment in 141 provinces recognized by the USGS to exist in FSU. The calculated data are represented by squares in Figure 4. The only change is that in the second case, 141 rather than 18 provinces are used in the calculations. Other parameters remain identical to the ones used in the first VSD calculation. The endowment in the second case, which includes assessed and unassessed provinces, is calculated to be 556 bboe.

**Table 2. Parameters used for calculating assessed and unassessed oil, natural gas, and NGL endowments in FSU. Parameters are defined under Equation 4.**

Parameter	Oil, FSU	Gas, FSU	NGL, FSU
$a_p$	0.74	0.25	1.06
$N_m$	1	1	1
$N_x$	18	31	18
$S$	9	0.8	21
$V_m$	0.039	0.1	0.254
$V_s$	50,000	11,893,126	5,000
$V_x$	210	44,000,000	25
$\Psi$	0.28	0.000046	50
$R^2$	0.99	0.99	0.98

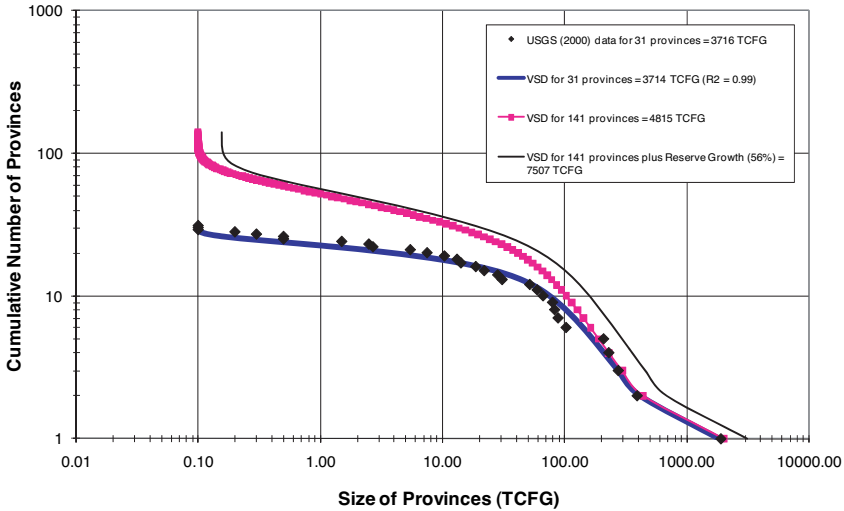


Figure 5. Conventional natural gas endowment in FSU.

### 3.2. Natural gas endowment

The same approach and the same symbols discussed above are used to construct Figure 5. The 31 provinces for which USGS (2000) provides volumes yield a natural gas endowment of 3,716 tcf. Nearly the same volume (3,714 tcf) is calculated by the VSD model using the parameters shown in the third column of Table 2. The coefficient of determination is 0.99. The assessed plus unassessed endowment in this case is 4,815 tcf.

### 3.3. NGL endowment

The endowment of NGL is presented in Figure 6. The volume calculated by the USGS for 18 provinces is 71 bboe. Using the parameters shown in the fourth column of Table 2, the same volume is calculated with the VSD model and results in an  $R^2$  value of 0.98. Extending the evaluation to 141 provinces gives an assessed plus unassessed endowment equal to 163 bboe.

## 4. RESERVE GROWTH

Reserve growth is a factor defined by the USGS as the increase in reserves of a previously discovered field through time. It provides very significant increase to petroleum volumes. As classified by the USGS, reserve growth applies to 'known' volumes (cumulative production plus remaining reserves). In this paper, reserve growth also applies to endowment volumes, under the assumption that it is possible that even the total endowment may experience future growth in the same way known volumes have. It is estimated by calculating a percentage for reserve growth, based on 'known' volumes from USGS (2000), and applying it to estimated petroleum endowment volumes of both assessed and unassessed provinces.

Reserve growth percentages have been calculated using values from Table 3. In the 'world total' section, for instance, the reserve growth percentage of known oil is calculated to be 43%. This comes from dividing reserve growth (688 BBOE) by the summation of cumulative production plus remaining reserves (710 BBOE + 891 BBOE).



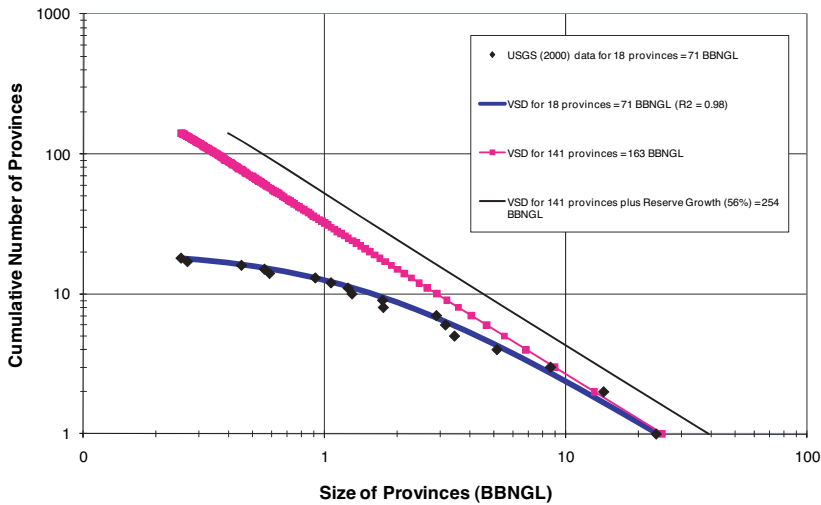


Figure 6. NGL endowment in FSU.

**Table 3. Calculation of Reserve Growth Percentages, based on data from USGS (2000).**

	Mean Oil (Billion barrels)	Mean Gas (BBOE)	Mean NGL (Billion barrels)	Total Mean Petroleum (Billion barrels)
<b>World (excluding USA)</b>				
Undiscovered conventional	649	778	207	1634
Reserve growth (conventional)	612	551	42	1205
Remaining reserves	859	770	68	1697
Cumulative production	539	150	7	696
<b>Total</b>	<b>2659</b>	<b>2249</b>	<b>324</b>	<b>5232</b>
Known volumes	1398	920	75	2393
<b>Reserve growth based on known volumes (%)</b>	<b>43.78</b>	<b>59.89</b>	<b>56.00</b>	<b>50.36</b>
	Mean Oil (Billion barrels)	Mean Gas (BBOE)	Mean NGL (Billion barrels)	Total Mean Petroleum (Billion barrels)
<b>USA</b>				
Undiscovered conventional	83	88	combined with oil	171
Reserve growth (conventional)	76	59	combined with oil	135
Remaining reserves	32	29	combined with oil	61
Cumulative production	171	142	combined with oil	313
<b>Total</b>	<b>362</b>	<b>318</b>	<b>combined with oil</b>	<b>680</b>
Known volumes	203	171	combined with oil	374
<b>Reserve growth based on known volumes (%)</b>	<b>37.44</b>	<b>34.50</b>	<b>combined with oil</b>	<b>36.10</b>

	Mean Oil (Billion barrels)	Mean Gas (BBOE)	Mean NGL (Billion barrels)	Total Mean Petroleum (Billion barrels)
<b>World Total</b>				
Undiscovered conventional	732	866	207	1805
Reserve growth (conventional)	688	610	42	1340
Remaining reserves	891	799	68	1758
Cumulative production	710	292	7	1009
<b>Total</b>	<b>3021</b>	<b>2567</b>	<b>324</b>	<b>5912</b>
Known volumes	1601	1091	75	2767
<b>Reserve growth based on known volumes (%)</b>	<b>42.97</b>	<b>55.91</b>	<b>56.00</b>	<b>48.43</b>

The basic assumptions in the application of reserve growth are that the reserve growth percentages based on known volumes will be the same for endowment volumes. Also, the reserve growth percentage estimated by USGS (2000) for the world will be applicable to the FSU region. Finally, the ranking of endowment volumes plus reserve growth, by size, will be the same as the ranking of the endowment volumes without reserve growth.

The estimated VSD curve of oil endowment plus reserve growth (43%), for 141 provinces, is presented in Figure 4. It lies to the right of the curve showing the VSD estimate of 141 provinces without reserve growth. It shows that at each province, each volume is now 43% greater than the original estimate. The estimated oil endowment volume plus reserve growth, for 141 provinces, is 795 billion BOE. The same reserve growth methodology is applied to natural gas and NGL, estimated at 7,507 tcf and 254 bboe, respectively.

## 5. LIFE EXPECTANCIES

Table 4 indicates that there is plenty of oil, natural gas and NGL to last for several decades in the FSU. The second column presents 'future' volumes, i.e., the endowment minus the cumulative production reported in the USGS (2000) study. The third column gives the average annual production between 2007 and 2009, as reported in the British Petroleum Statistical Review of World Energy (BP, 2010). The fourth column shows life expectancies in years, at various growth rates in production (0%, 2% and 5% per year). The fifth column provides the average annual growth in production over 30 years, between 1979 and 2009 (BP, 2010). For a hypothetical production growth rate of 2% per year, the life expectancy for oil, including unassessed provinces and reserve growth, would be 87 years. In the case of gas, the remaining life would be 92 years. It is important to note that the estimates do not include unconventional oil and gas.

In spite of the considerable petroleum resources in the FSU, temporary setbacks can be anticipated. It is known from past history that supply disruptions can occur due to lack of spare capacity, cartels, political instability, natural disasters, strikes, shortage of qualified workers, shortage of refining capacity, commodity manipulation by speculators, and the power of national oil companies. The important point is that there are vast petroleum volumes in the region that could be recovered over the long term, given sufficient research and development and appropriate public policies.

**Table 4. FSU life expectancies at various growth rates in production (without including unconventional oil or gas).**

1	2	3	4	5
Conventional Petroleum Resource	Future Volumes (Oil and NGL in boe; gas in CF)	2007-2009 <sup>a</sup> Average Annual Production (Oil in boe; gas in CF)	Life Expectancy in Years, at Various <sup>b</sup> Growth Rates in Production	Average Annual <sup>c</sup> Growth in Production, 1979-2009 (%)
FSU oil from USGS (2000)	2.68E+11		0%	
FSU oil including unassessed provinces	4.44E+11	2.94E+09	91	34
FSU oil including unassessed provinces and reserve growth	6.83E+11		151	43
			232	51
FSU gas from USGS (2000)	3.33E+15		124	39
FSU gas including unassessed provinces	4.43E+15	2.69E+13	165	44
FSU gas including unassessed provinces and reserve growth	7.12E+15		265	53
FSU NGL from USGS (2000)	6.92E+10		-	-
FSU NGL including unassessed provinces	1.61E+11	-	-	-
FSU NGL including unassessed provinces and reserve growth	2.52E+11		-	-

Notes:

<sup>a</sup> Average annual production comes from British Petroleum (2010)<sup>b</sup> Life expectancies estimated by this study<sup>c</sup> Average annual growth in production calculated from British Petroleum (2010)

## 6. CONCLUSIONS

The Variable Shape Distribution (VSD), a size distribution model, is unique in that it allows actual petroleum resource data to determine the relation between the size and number of petroleum provinces. In the past, all methods were based on an assumed form of the size distribution of petroleum resources (e.g. fractal or lognormal).

The VSD model is estimated and validated using data from USGS (2000). That study presents FSU oil and NGL volumes for 18 provinces, as well as gas volumes for 31 provinces, out of a total of 141. Therefore, there are 123 provinces left unassessed for oil and NGL, and 110 left unassessed for gas. Given the validations of the VSD demonstrated in this paper, the model can be used to estimate reasonable endowment volumes for all 141 provinces in the region. For the case of oil, the VSD model estimates 556 BBOE in 141 provinces. The VSD gas endowment is calculated at 4,815 TCFG, while NGL amounts to 163 BBNGL.

These endowments are larger than commonly assumed, since there is a tendency to overlook volumes that come from unassessed provinces and reserve growth. An important implication is that FSU petroleum will last for several decades and can contribute significantly to the energy requirements of the region.

## APPENDIX A: DEVELOPMENT OF THE VSD MODEL

### A.1 Fractal and pareto distributions

A fractal distribution is provided by a power law of the form (Barton and Scholz, 1995):

$$N(r) = Cr^{-D} \quad (\text{A.1})$$

where:

$D$  - constant fractal dimension.

$C$  - constant of proportionality.

$r$  - some quantity (e.g. natural gas volume in TCFG).

$N(r)$  - number of objects with a quantity equal to or greater than  $r$ .

For conventional petroleum, volumes ( $V$ ),

$$V = r^3 \quad (\text{A.2})$$

and consequently, Equation A.1 becomes:

$$N(V) = CV^{-D/3} \quad (\text{A.3})$$

where:

$N(V)$  - distribution function of  $V$ .

The density function,  $n(V)$ , is the derivative of the distribution function (Equation A.3), and is given by:

$$n(V) = \frac{dN(V)}{dV} \quad (\text{A.4})$$

Conversely, the distribution function,  $N(V)$ , is the integral of the density function, and is given by:

$$N(V) = \int_{V_{\min}}^{V_{\max}} n(V) dV \quad (\text{A.5})$$

where  $V_{\min}$  and  $V_{\max}$  are the minimum and maximum volumes, respectively.

The distribution function, Equation A.3, can be written as a Pareto power law, or Pareto distribution, as follows:

$$N(V) = CV^{-a_p} \quad (\text{A.6})$$

where  $a_p$  is a shape parameter (also known as shape exponent, Pareto exponent, Pareto constant, or fractal dimension). Note that based on the above development, the shape parameter ( $a_p$ ) is equal to 3 times the fractal dimension (D).

Taking logarithms of both sides of Equation A.6 leads to:

$$\log N(V) = \log(C) - a_p \log V \quad (\text{A.7})$$

Equation A.7 indicates that a log-log plot of  $N(V)$  versus  $V$  should result in a straight line with a slope equal to  $-a_p$  and an intercept, at  $V = 1$ , equal to  $C$ .

## A.2 Description of VSD model

As mentioned earlier, development of the VSD starts by observing the curvature given by the data points from USGS (2000) on the log-log plot. We then create the VSD model which allows the data to determine the specified relationship between the size and number of provinces.

Following the work of Barton and Scholz (1995) and Pareto, the VSD is given by:

$$N_i = \frac{1}{r_i^{a_i}} \quad (\text{A.8})$$

where:

$a_i$  - unknown, variable shape exponent.

$N_i$  - cumulative number of provinces, bound by USGS (2000).

$r_i$  - estimated normalized volume of a province.

Thus, the VSD equation is somewhat similar to Pareto's equation.<sup>5</sup> The significant difference is that the shape exponent in Pareto's equation is a constant, while the shape exponent in the VSD model can vary. When the shape exponent in the VSD model is constant, the VSD equation is a Pareto equation. By using normalized volumes, the constant of proportionality,  $C$ , is equal to 1. Since there are two unknowns in Equation A.8,  $a_i$  and  $r_i$ , an auxiliary equation has been developed for calculating the two unknowns. The equation has the form:

<sup>5</sup>Other extensions of the Pareto distribution include the Parabolic Fractal Distribution (Bettini, 1987) and the Log-Pareto Distribution (Laherrere, 1996). These extensions, like the Pareto and lognormal distributions, are based on assumed forms of the size distribution of oil and gas accumulations.

$$N_t = \frac{1}{r_t^{a_t}} = \frac{1}{r_v^{a_p} + r_m^{a_m}} \quad (\text{A.9})$$

where:

$a_m$  - slope of line between (1) point of maximum number of provinces ( $N_x$ ) and minimum USGS province volume ( $V_m$ ) and (2) point of minimum number of provinces ( $N_m = 1$ ) and maximum volume ( $V_x$ ) given by Pareto straight line at  $N_m = 1$ .<sup>6</sup>

$a_p$  - slope of straight line approximated from USGS data points; same as slope of Pareto distribution.

$r_m$  - normalized, minimum USGS province volume ( $r_m = V_m/V_x$ , where  $V_m$  is the minimum USGS province volume, and  $V_x$  is the maximum volume given by the Pareto straight line at a number of provinces equal to 1).

$r_v$  - estimated normalized volume of a province minus the normalized minimum USGS volume.

The right hand side of Equation A.9 has been empirically estimated and tested against the data published in (USGS, 2000), providing reasonable results. Since the number of provinces ( $N_t$ ) and the normalized, minimum province volume ( $r_m$ ) are known from USGS (2000), and the slopes  $a_p$  and  $a_m$  can be determined from the graphed data, it is possible to re-arrange Equation A.9 to determine  $r_v$ , as follows:

$$r_v = \left( \frac{1}{N_t} - r_m^{a_m} \right)^{\frac{1}{a_p}} \quad (\text{A.10})$$

Equation A.10 gives the estimated normalized volume of a province minus the normalized minimum USGS volume ( $r_m$ ). The minimum volume must be added to  $r_v$  to get the estimated, normalized volume of a province, as follows:

$$r_t = r_v + r_m \quad (\text{A.11})$$

Then, to 'de-normalize'  $r_t$ , and get the estimated volume of a province, the following equation is used:

$$V_t = r_t V_x \quad (\text{A.12})$$

where:

$V_t$  - estimated volume of a province before right-tail adjustment.

$V_x$  - maximum volume given by Pareto straight line at number of provinces = 1.

After calculating  $r_v$  and  $r_t$  using Equations A.10 and A.11, respectively, it is possible to re-arrange Equation A.9 to solve for the variable shape exponent ( $a_t$ ), as follows:

$$a_t = \frac{\log[r_v^{a_p} + r_m^{a_m}]}{\log r_t} \quad (\text{A.13})$$

<sup>6</sup>In equation form, this slope is given by  $a_m = \frac{\log N_x - \log N_m}{\log V_x - \log V_m}$

where:

$a_i$  - variable shape exponent; slope of line between point of any estimated volume and the maximum volume point given by the Pareto straight line at number of provinces equal to 1.

### A.3 Adjusting right tail (largest volumes) of VSD model

Notice that in Figure 5, the data point for the largest province,  $V_L$ , does not lie on the Pareto straight line. Often, more than one of the larger volumes will not lie on the line.<sup>7</sup>

To get a more accurate, estimated curve that fits over the largest volumes, an exponential function has been incorporated empirically into the VSD model. The function has the effect of smoothly bending the estimated curve in the direction of the actual USGS data. The exponential function is as follows:

$$f(V_i, V_s) = (1 - \exp(-V_i/V_s))^S \quad (\text{A.14})$$

where:

$S$  - severity exponent that is determined by trial and error. It controls the extent to which the estimated VSD curve will separate from the Pareto straight line (on the right tail of the distribution, near the largest volumes).

$V_s$  - parameter that corresponds approximately to the point at which the largest USGS provinces begin to deviate from the Pareto straight line, on the right tail of the distribution, near the largest volumes.

$V_i$  - estimated volume of a province before right-tail adjustment.

Another parameter incorporated into the VSD model to get a more accurate fit of the largest USGS volumes is approximated by:

$$\psi = \frac{V_L}{V_X} \quad (\text{A.15})$$

where:

$\psi$  - separation ratio that controls the amount of separation between the Pareto straight line and the estimated VSD curve (on the right tail of the distribution, near the largest volumes).

$V_L$  - volume of the largest USGS province.

$V_X$  - maximum volume given by Pareto straight line.

Three parameters ( $V_s$ ,  $\psi$ , and  $S$ ) only affect the right tail of the VSD curve, near the largest volumes. They are incorporated so that the VSD curve will fit over the largest volumes that do not lie on the Pareto straight line.

The exponential function (see Equation A.14) and the separation ratio ( $\psi$ ) are used to improve the estimate of a province's volume as follows:

$$\hat{V}_i = \frac{V_i \times \psi}{\psi + (1 - \psi) \cdot f(V_i, V_s)} \quad (\text{A.16})$$

<sup>7</sup>This is a common occurrence in Pareto distributions. It has been interpreted by Barton and Scholz (1995) to be the result of sub-populations, each of which have the same Pareto shape exponent (slope), but different largest volumes.

where:

$\hat{V}_i$  - estimated volume of a province.

Expanding the exponential function,  $f(V_t, V_s)$ , in Equation A.16 gives:

$$\hat{V}_i = \frac{V_t \times \psi}{\psi + (1 - \psi) \cdot (1 - \exp(-V_t/V_s))^S} \tag{A.17}$$

where, as shown earlier in Equation A.12:

$$V_t = r_t V_x = \left[ \left( \frac{1}{N_t} - r_m^{a_m} \right)^{\frac{1}{a_p}} + r_m \right] \cdot V_x \tag{A.18}$$

Further expansion of Equation A.18 results in:

$$V_t = \left[ \left( \frac{1}{N_t} - \left( \frac{V_m}{V_x} \right)^{\left( \frac{\log N_x - \log N_m}{\log V_x - \log V_m} \right)^{\frac{1}{a_p}}} \right)^{\frac{1}{a_p}} + \frac{V_m}{V_x} \right] \cdot V_x \tag{A.19}$$

When  $V_t$ , from Equation A.19, is substituted into Equation A.17, all the terms are expanded, giving:

$$\hat{V}_i = \frac{\left\{ \left[ \left( \frac{1}{N_t} - \left( \frac{V_m}{V_x} \right)^{\left( \frac{\log N_x - \log N_m}{\log V_x - \log V_m} \right)^{\frac{1}{a_p}}} \right)^{\frac{1}{a_p}} + \frac{V_m}{V_x} \right] \cdot V_x \right\} \times (\psi)}{(\psi) + [1 - (\psi)] \cdot \left[ 1 - \exp \left( - \left\{ \left[ \left( \frac{1}{N_t} - \left( \frac{V_m}{V_x} \right)^{\left( \frac{\log N_x - \log N_m}{\log V_x - \log V_m} \right)^{\frac{1}{a_p}}} \right)^{\frac{1}{a_p}} + \frac{V_m}{V_x} \right] \cdot V_x \right\} \div V_s \right) \right]^S} \tag{A.20}$$

Equation A.20 is used to generate, via non-linear regression, the estimated VSD curves shown in Figures 4 to 6. The validity of the equation can be confirmed by visual inspection of the Figures. Further confirmation is provided by the coefficients of determination (0.98 or 0.99) and the volumetric estimates presented earlier that compare well with the actual volumes published by the USGS (2000) study.

**REFERENCES**

Aguilera R.F., 2006. Assessing the long run availability of global fossil energy resources. Doctoral Thesis, Colorado School of Mines, Golden, Colorado, United States.

Barton C.C. and Scholz C.H., 1995. The fractal size and spatial distribution of hydrocarbon accumulations. In: Barton C.C. and LaPointe P.R. (Eds.), *Fractals in Petroleum Geology and Earth Processes*. Plenum Press, New York.



- Barton C.C., 1995. A new approach to estimating hydrocarbon resources. United States Geological Survey Fact Sheet. Reston, Virginia, United States.
- Bettini C., 1987. Forecasting populations of undiscovered oil fields with the Log-Pareto distribution. Doctoral Thesis, Stanford University, United States.
- British Petroleum, 2010. Statistical Review of World Energy. London, United Kingdom.
- Drew L.J., 1997. Undiscovered Petroleum and Mineral Resources, Assessment and Controversy. Plenum Press, New York and London.
- Kaufman G.M., 2005. Where have we been? Where are we going? *Natural Resources Research* **14(3)**, 145–152.
- Laherrere J.H., 1996. Parabolic Fractal Distributions in Nature. *CR Academy of Sciences* **322(7)**, 535–541.
- United States Geological Survey, 2000. World Petroleum Assessment. CD-ROM, Virginia.
- Xie H., Guo Q., Li F., Li J., Wu N., Hu S. and Liang K., 2011. Prediction of petroleum exploration risk and subterranean spatial distribution of hydrocarbon accumulations. *Petroleum Science* **8(1)**, 17–23.

