

文章编号: 1001-6112(2010)05-0517-04

基于广义帕莱托分布的关键参数求取方法探讨

徐忠美¹, 金之钧², 孙红军², 徐京新³

(1. 中国地质大学 能源学院, 北京 100083; 2. 中国石油化工股份有限公司
石油勘探开发研究院, 北京 100083; 3. 中国石油大学, 北京 102249)

摘要: 广义帕莱托法是进行资源潜力评价的一种重要方法, 主要用于高勘探程度区。该方法可以预测地质目标中油气藏的个数与规模, 进而对地质目标的资源总量和资源结构开展定量预测。该方法的核心是如何求取 4 个关键参数: q_0 (最小油藏规模)、 q_{\max} (最大油藏规模)、 r (油气藏规模中位数)、 λ (分布特征参数)。针对中低勘探程度地区, 在资源丰度和最大油藏规模不确定的情况下, 原有的方法在求取上述 4 个关键参数中存在一定的困难。该文从广义帕莱托方法原理及分布模型出发, 探索了上述 4 个关键参数求取的一种新思路, 并初步推导了 4 个关键参数的求取公式, 这对广义帕莱托法的进一步推广和应用具有重要意义。

关键词: 广义帕莱托模型; 关键参数; 资源结构; 资源评价

中图分类号: TE155

文献标识码: A

THE DISCUSSION ABOUT HOW TO OBTAIN THE KEY PARAMETERS OF GENERALIZED PARETO DISTRIBUTION

Xu Zhongmei¹, Jin Zhijun², Sun Hongjun², Xu Jingxin³

(1. *China University of Geosciences, Beijing, 100083, China;*
2. *SINOPEC Exploration and Production Research Institute, Beijing 100083, China;*
3. *China University of Petroleum, Beijing 102249, China*)

Abstract: Generalized Pareto is an important method of resource assessment, this method is widely used in forecasting the total quantitative and structure of resource by calculating the number and scale of reservoir. How to obtain four key parameters is important in this method: q_0 (minimum pool size), q_{\max} (maximum pool size), r (median oil and gas pool size), λ (distribution parameter). For low degree of exploration areas, because resource abundance and the maximum reservoir size uncertain, the key parameter is difficult ascertain by the conventional method. According to this, auther discussed how to obtain the four key parameters and deduced the formula, this is important to further application of this method.

Key words: generalized Pareto distribution; key parameters; resource structure; resource evaluation

勘探决策的首要问题就是对地质目标的资源潜力评价, 地质目标资源潜力的大小取决于地质目标中油气藏的个数与规模。为了解决这一问题, 众多地质学家通过研究已发现油气藏的规律特征, 建立起了各种预测模型和相应的评价方法。金之钧 1992 年在对西西伯利亚盆地的 2 600 个油气藏数据分析基础上提出了广义帕莱托分布模型^[1]。在该模型的应用中, 参数的确定是方法应用的关键, 所以对不同勘探程度的地区, 如何求取关键参数就成为研究的重点^[2]。笔者通过调研广义帕莱托方法应用现状, 分析了现有的关键参数确定方法和思路, 并进一步探讨了在低勘探程度地区关键参数的

求取思路。

1 广义帕莱托分布模型

油气勘探实践证明, 在一个含油气盆地中, 储量大的油气藏出现的概率很小, 中小型油气藏出现的概率很大(图 1)。假设它们之间的关系可用概率密度函数来表示, 如果能够找到这一概率密度函数的表达式, 则可以预测该预测单元中的资源量及其结构^[1]。

金之钧统计了西西伯利亚盆地的 2 600 个油气藏的数据, 划分选定了 3 个含油气区, 并以 5 年为一勘探阶段, 划分出 16 个勘探样本, 研究建立了

收稿日期: 2010-06-05; 修订日期: 2010-09-25。

作者简介: 徐忠美(1980—), 男(苗族), 博士生, 从事含油气系统模拟、资源评价、油气成藏机理与分布规律的研究。E-mail: xzm-79@163.com。

基金项目: 国家重大专项“海相碳酸盐岩层系大中型油气田分布规律及勘探评价”(2008ZX05005-001)。

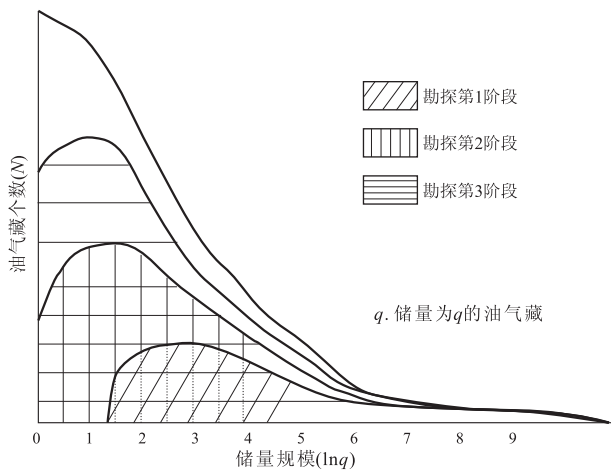


图 1 不同勘探阶段油气藏发现概率及规模趋势
金之钧,1991。

Fig. 1 The discovery probability and discovery trend of reservoirs in different stage of exploration

如下分布函数^[1]：

$$F(q) = 1 - \left(\frac{q_0 + r}{q + r}\right)^\lambda + \left(\frac{q_0 + r}{q_{\max} + r}\right)^\lambda \quad (1)$$

$$f(q) = \frac{\lambda(q_0 + r)^\lambda}{(q + r)^{\lambda+1}} \quad (2)$$

式中： r 为样本中中位数油气藏储量大小； λ 是一特征参数，为油气藏规模分布系数（ $\lambda \geq 0$ ）； q_0, q_{\max} 分别为最小和最大油藏规模。

通过研究，还发现 r, λ 还与资源丰度、资源转化率及最小油藏规模存在如下关系。

$$r = 0.1\rho(1 - \alpha) + 3.0q_0 \quad (3)$$

$$\lambda = 1.22 - 0.001\rho \quad (4)$$

式中： α 为资源转化率； ρ 为资源丰度。

2 资源结构预测及参数确定现状

由上述模型可知，广义帕莱托方法应用涉及到 4 个关键参数： $q_0, q_{\max}, r, \lambda$ 。基于广义帕莱托分布模型，在已知评价区资源丰度或资源总量的情况下，刘金侠等^[2]推导了上述 4 个关键参数的求取方法，并提出了应用广义帕莱托方法进行资源结构预测的公式和流程。具体步骤如下：1) 研究基本石油地质特征，划分成藏体系，确定统计预测单元；2) 建立勘探样本，在所选研究区里按时间序列建立勘探样本序列；3) 根据勘探样本中给出的具体数字建立 r 与 α 的函数关系： $r = k\rho(1 - \alpha) + Bq_0$ ，式中： α 为资源转化率， ρ 为资源丰度。当研究区确定后， ρ 与 q_0 为已知，求 r 与 α 的对应值可以确定回归系数 k, B ；4) 求 λ 值：首先根据研究区实际地质特征（如

资源丰度等）、勘探样本与 λ 值建立关系，求取初始值 λ_0 ，而后用最小二乘法检验来确定 λ 值；5) χ^2 检验：检验所建立的分布是否符合统计要求；6) 预测待发现的资源结构。

应用该方法计算内容包括：总平均油气藏储量 \bar{q} ，油气藏总个数 N ，某一储量级别区间 $[q_1, q_2]$ 的资源量 Q 及油气藏个数 n 。作者也进一步指出了其地质意义在于：通过预测不同级别油气藏的储量和油气藏个数的百分比，比例大的是应优先考虑的勘探领域。

由上所述可知，刘金侠等对广义帕莱托分布模型进行资源结构预测已经有了完整的推导和求取过程。上述的推导中，在已知最大油气藏规模时，还可以进一步计算勘探目标的资源总量，即求取最小油气藏与最大油气藏间的资源量就是该地区的资源总量。在关键参数的求取中，其基本思路为： q_0 为最小经济油气藏规模； r, λ 基于已知资源丰度和 q_0 通过回归来求取； q_{\max} 可以通过地质分析来给定。

3 关键参数确定思路探讨

对于中、低勘探程度地区，最大油气藏规模往往是不可知或者不确定的。笔者在此探讨另外一种计算思路与方法，其特点是不知道或者无法给出确定的资源丰度和最大油藏规模的情况下，根据已有勘探样本来建立勘探目标的分布模型，进一步预测资源规模和资源量。

3.1 参数分析及优化思路确定

基于最初的广义帕莱托分布模型，公式(1)、(2)中涉及到 4 个关键参数： $q_0, q_{\max}, r, \lambda$ 。 q_0 可以根据经济油藏规模下限值来设定，而其它 3 个参数是需要进一步求取的。由于假设资源丰度为不确定，因此就不能根据资源丰度来求取 r, λ 。在此确定了如下研究思路：首先基于广义帕莱托分布模型，来推导 r, q_{\max} 的数学公式；通过生成的随机样本，来拟合 λ 与 r 的关系式；最终建立 r, λ, q_{\max} 的数学公式。针对实际的勘探样本，利用本次推导的关键参数求取公式，就可以基于广义帕莱托模型来预测资源结构及资源量。

3.2 公式分析及推导

3.2.1 油气藏规模中位数的公式分析

大量勘探实践表明，较大规模油气藏(田)在勘探早期发现的概率较大，体现在理论随机抽样模拟中表现为样本中位数随抽样过程的持续进行而变小(图2)。如将勘探发现过程看成是不放回随机

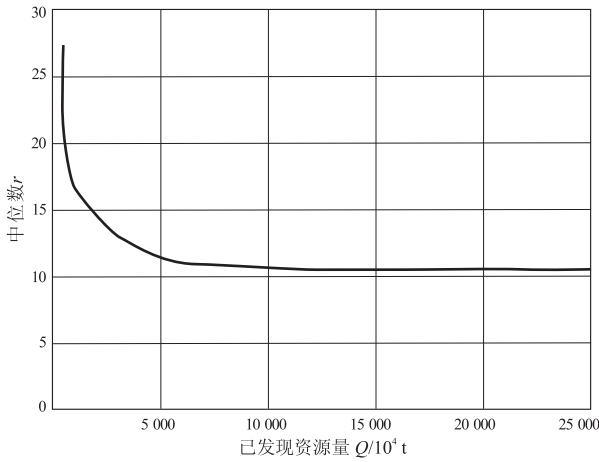


图 2 中位数与已发现资源总量关系

Fig. 2 The relations figure between the median of reservoirs and resources

抽样过程,则各勘探阶段已发现总量与勘探阶段发现油气藏(田)规模中位数应有下列关系:

$$r_t = r \exp(B/Q_t) \tag{5}$$

式中: r_t 为 t 勘探阶段已发现油气藏规模中位数; Q_t 为 t 勘探阶段已发现资源量。

利用上述公式,把已知勘探样本划分为 2 个或 2 个以上的勘探阶段,我们就可以确定总体样本的 r (样本中位数油气藏储量大小)。

3.2.2 最大油气藏规模的公式推导

依据广义帕莱托相关参数的定义与设定方法,当 q_0 取 0 时,广义帕莱托分布函数可简化为:

$$F(q) = \begin{cases} 0 & q=0 \\ 1 + \text{pow}(r/(q_{\max} + r), \lambda) - \text{pow}(r/(q + r), \lambda) & 0 < q < q_{\max} \\ 1 & q \geq q_{\max} \end{cases} \tag{6}$$

则:当 q 取值为油气藏规模中位数时, $F(q)$ 取值为 0.5,则公式(6)可变形为:

$$\begin{aligned} 1 + \text{pow}(r/(q_{\max} + r), \lambda) - \text{pow}(r/(r + r), \lambda) &= 0.5 \\ \text{pow}(r/(q_{\max} + r), \lambda) &= \text{pow}(r/(r + r), \lambda) - 0.5 \\ \lambda \log_{10}(r/(q_{\max} + r)) &= \log_{10}(\text{pow}(0.5, \lambda) - 0.5) \\ \log_{10}(r/(q_{\max} + r)) &= \log_{10}(\text{pow}(0.5, \lambda) - 0.5) / \lambda \\ r/(q_{\max} + r) &= \text{pow}(10, \log_{10}(\text{pow}(0.5, \lambda) - 0.5) / \lambda) \end{aligned}$$

再进一步变形,就可推导出 q_{\max} 与 r, λ 具有如下关系:

$$q_{\max} = r / \text{pow}(10, \log_{10}(\text{pow}(0.5, \lambda) - 0.5) / \lambda) - r \tag{7}$$

由公式(7)可知,对任一总体样本 r 已知,如果知道 λ 就可以确定 q_{\max} 。

3.3 数学抽样模拟及 λ 参数关系式确定

由于 λ 为特征参数,无法通过公式推导得出,

通过分析可知 λ 与同为特征参数的标准差(σ)具有密切关系。通过模拟产生的随机样本,首先根据已推导的 r 公式,来模拟 λ 与样本中标准差(σ)和中位数(r)的关系式。

3.3.1 油气藏规模中位数的公式分析

图 3 是 5 组不放回随机抽样过程分段统计中位数与累积量关系图,从这 5 组数据可以看出,样本中位数随抽样过程的持续进行而变小的规律是存在的。图 4 是同一总体样本的 3 组不放回随机抽样,分段统计其中位数与累积量关系及回归分析图。

从图 4 可以看出,这 3 组数据样本中位数与累积量之间存在下述关系:

$$r = A \exp(B/q)$$

3.3.2 油气藏规模分布系数的公式分析

图 5 为不同中位数在相同分布系数下的样本

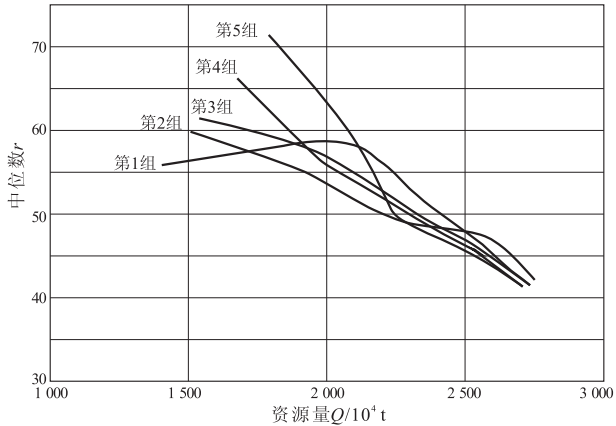


图 3 5 组样本的中位数变化规律

Fig. 3 The change rule of median variation about five groups sample

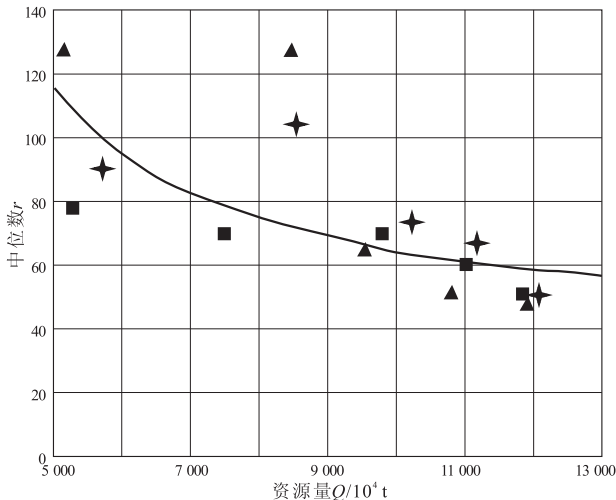


图 4 中位数与累积量回归分析

Fig. 4 The regression analysis figure between the median of reservoirs and cumulative amount

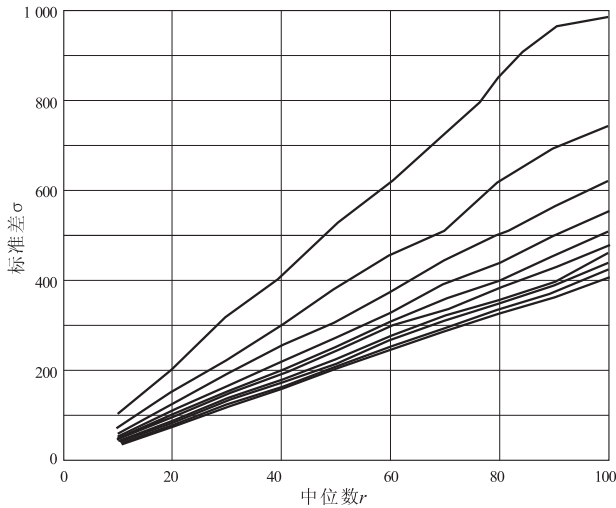


图 5 分布系数相同时中位数与标准差的关系

Fig. 5 The relational figure between the median of reservoirs and standard deviation

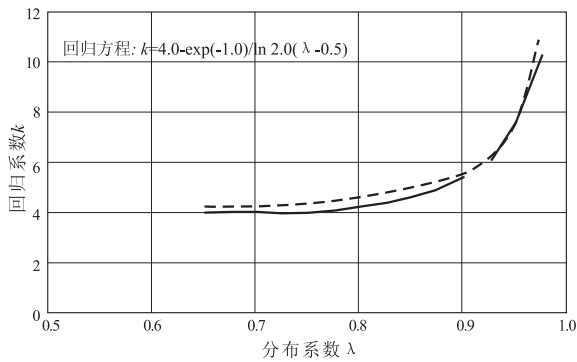


图 6 分布系数与回归系数的关系

Fig. 6 The relational figure between distribution coefficient and regression coefficients

标准差统计分析图,由图 5 可知分布系数(λ)相同时中位数(r)与标准差(σ)存在下列关系: $\sigma=k\lambda r$ 。

图 6 为分布系数与“中位数与标准差关系斜率”关系,结合分布系数相同时中位数与标准差关系可以得出如下关系式:

$$\lambda = \exp(-1 / (\exp(1) \times (\sigma / r - 4))) / 2 + 0.5 \quad (8)$$

至此,基于广义帕莱托法原理与分布模型,推导了 r 和 q_{\max} 的数学公式,并基于随机样本模拟,

确定了 λ 的数学公式,加上 q_0 的较易求取,有关广义帕莱托分布的 4 个关键参数就已经确定了。

4 结论

1) 国内外学者在实际地区的应用中,通过与其它几种具代表性的概率分布模型相比^[3-9],普遍认为广义帕莱托分布更适用于不同勘探程度探区资源结构的定量预测^[5,6]。所以推动该方法在不同勘探程度探区的规范应用意义重大。

2) 前人对广义帕莱托方法在中高勘探区(资源丰度已知)的应用进行了推导和研究,并确立了其研究流程和步骤,极大地推动了广义帕莱托方法的应用,同时也为后面的研究奠定和积累了很好基础。

3) 对于中等偏低以及低勘探程度地区,笔者基于金之钧教授提出的广义帕莱托分布模型,通过分析确定了关键参数的求取思路,推导了 3 个关键参数 r , q_{\max} 和 λ 的数学公式。其意义在于,基于前人对中高勘探程度广义帕莱托方法的应用,进一步拓展了该方法在低勘探程度区的应用。

参考文献:

- [1] 金之钧,施比伊曼,武守诚. 油气资源定量评价系统[J]. 地质论评,1996,42(增刊): 247-258.
- [2] 刘金侠,金之钧. 板桥北大港成藏体系资源结构定量预测[J]. 大庆石油学院学报,2004,28(1): 44-46.
- [3] 周总瑛,白森舒,何宏. 成因法与统计法油气资源评价对比分析[J]. 石油实验地质,2005,27(1): 67-73.
- [4] 金之钧. 五种油气藏规模概率分布模型比较研究[J]. 石油学报,1995,74(3): 12-15.
- [5] 金之钧,张金川. 油气资源评价技术[M]. 北京:石油工业出版社,1999: 1-85.
- [6] 谌卓恒,金之钧. 川东石炭系天然气成藏组合总体特征预测[J]. 石油勘探与开发,1996,23(6): 12-23.
- [7] ARPS J J, ROBERTS TG. Economics of drilling for Cretaceous oil on east flank of Denver - Julesburg Basin[J]. AAPG Bulletin, 1958, 42(11): 2549-2566.
- [8] 旷理雄,郭建华. 潜江凹陷习家口油气聚集区最终可探明储量预测[J]. 石油天然气学报,2006,28(4): 39-41.
- [9] 旷理雄,郭建华. 最终可探明储量预测:以潜江凹陷王广油气聚集区为例[J]. 天然气工业,2006,26(11): 27-30.

(编辑 徐文明)