

石油储量计算的經驗分布变换法

陶 起

(中国科学院地质研究所)

石油勘探开发工作不可避免地要面对地质条件的不确定性,即随机性。因此,在石油资源评价中加强概率统计方法的研究和应用,必然提高评价结果的可靠性。基于这一目的,本文提出一种新的计算方法——經驗分布变换法。

一、体积法中的概率问题

体积法是一类重要的传统石油储量计算方法。它有许多种变异方法。以生油岩体积法中的“氯仿沥青A法”^[1]为例,其计算公式为:

$$Z = k_1 \cdot k_2 \cdot s \cdot d \cdot X \cdot Y \quad (1)$$

其中: Z 是石油地质储量; s 是生油岩有效面积; d 是生油岩密度; k_1 和 k_2 分别是生油岩排烃系数和烃类聚集系数; X 是生油岩厚度; Y 是生油岩氯仿沥青含量百分数。

利用氯仿沥青法计算石油储量,传统的做法是,根据地质资料求出每个参数的平均值,然后按(1)式相乘,求得 Z 。这样计算得出的结果往往是很粗略的。

事实上,生油岩有效厚度 X 和氯仿沥青含量百分数 Y 在空间上的变化很大,而且具有很明显的随机性,单一的平均值无法反映出它们的随机变化特性。所以,必须把它们作为随机变量来处理。而其它参数在一个评价区中的变化不大,可视为常数。

令 $R = k_1 \cdot k_2 \cdot s \cdot d$, 则(1)式可简化为:

$$Z = R \cdot X \cdot Y \quad (2)$$

显然,作为两个随机变量与一个常数的乘积,储量 Z 也是随机变量。

这样,用概率统计方法按(1)式计算石油储量的问题就归结为:在求得 X 和 Y 的概率分布之后,按(2)式求出 Z 的概率分布。

蒙特卡洛法是目下已被广泛采用的解决上述问题的一种概率统计方法^[1,2]。它的基本原理是:先求出 X 与 Y 各自的經驗边缘概率分布,然后在假定 X 和 Y 相互独立的情况下,用蒙特卡洛模拟技术求出 Z 的概率分布。

本文寻求的是解决上述问题的另一种概率统计方法。

二、經驗分布变换法的提出

设 X 与 Y 的总体联合分布密度函数是 $f(x, y)$, 根据随机向量变换技术,由(2)式确定的 Z 的分布密度函数理论上由下式给出:

$$f_z(z) = \frac{1}{R} \int_{-\infty}^{\infty} f(x, \frac{z}{Rx}) \frac{1}{|x|} dx. \quad (3)$$

若 X 与 Y 相互独立, $f_1(x)$ 和 $f_2(y)$ 分别是 X 和 Y 的总体分布密度, 则 Z 的概率分布密度函数的理论计算公式为:

$$f_z(z) = \frac{1}{R} \int_{-\infty}^{\infty} f_1(x) f_2(\frac{z}{Rx}) \frac{1}{|x|} dx. \quad (4)$$

对于 (4) 式, 有人采用正态分布或其它精确分布来逼近 $f_1(x)$ 和 $f_2(y)$, 从而用数值积分或蒙特卡洛法即可算出 Z 的分布。当数据资料十分缺乏时, 这样的处理具有明显的优越性。但对于 $f_1(x)$ 和 $f_2(y)$ 是否确为正态分布, 尚缺乏充足的证据。况且 X 与 Y 的相互独立性也不一定具备。

由于在实际应用中很难确定 $f(x, y)$ 的具体形式, 因此过去一直认为无法用 (3) 式求解 Z 的分布。

T. A. 琼斯^[3]曾提出用数值方法计算统计分布卷积的方法。受他的方法的启示, 本文又进一步利用“经验联合分布密度阵”代替 $f(x, y)$, 从 (3) 式出发, 导出了一套随机向量变换的数值计算过程。由此提出经验分布变换法。

三、经验分布变换法

1. 经验联合分布密度阵

定义 $n \times m$ 阶矩阵

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix} \quad (5)$$

为 X 与 Y 的“经验联合分布密度阵”。其中

$$p_{ij} = \frac{N_{ij}}{N(a_{i+1} - a_i)(b_{j+1} - b_j)}, \quad i=1, \dots, n, \quad j=1, \dots, m.$$

$a_1, \dots, a_n, a_{n+1}$ 和 $b_1, \dots, b_m, b_{m+1}$ 分别是 X 和 Y 的区间分割点; N 是样本容量; N_{ij} 是随机向量 (X, Y) 的样本点 (x, y) 落在小区域 (见图 1)

$$G_{ij} = [a_i, a_{i+1}) \times [b_j, b_{j+1})$$

中的频数, $i=1, \dots, n; j=1, \dots, m$ 。因此, 由上式确定的 p_{ij} 是随机向量 (X, Y) 在小区域 G_{ij} 中的经验分布密度值, 而 $u_{ij} = N_{ij}/N$ 是相应的经验频率值。

类似于二维随机变量的直方图, 经验联合分布密度阵可以用如图 2 所示的“立体直方图”来直观地表示。其中方柱的高是 p_{ij} , 体积是 u_{ij} 。

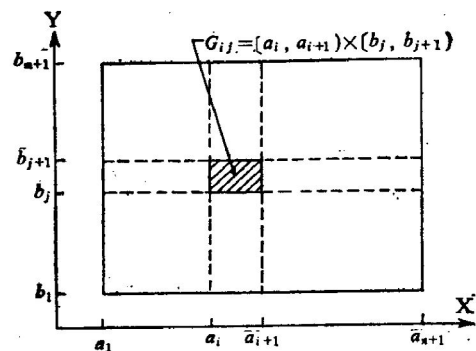


图 1 随机向量 (X, Y) 的分布区域

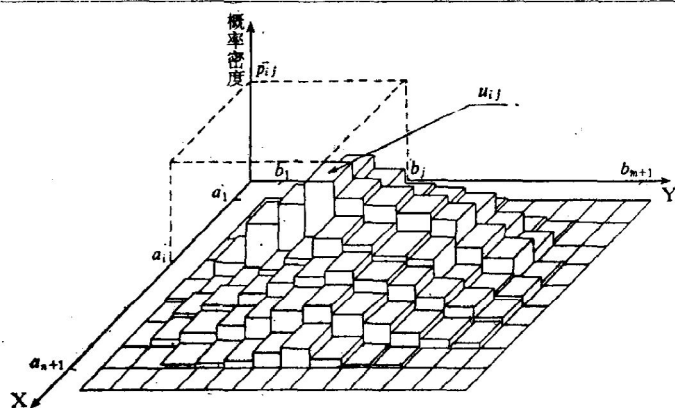


图2 经验联合分布的几何图示——立体直方图

面插值的方法扩大数据量, 或者从等值线图上直接读取大量的网格点数据。

2. 局部石油储量 Z 的分布

设 X 和 Y 分别是某个局部区的生油岩有效厚度和氯仿沥青含量百分数。根据样本数据求得形如 (5) 式的 X 与 Y 的经验联合分布密度阵 P 。用矩阵 P 代替 (3) 式中的 $f(x, y)$, 经过一系列的推导 (推导过程从略), 可得到局部储量 Z 的分布密度数值计算公式:

$$f_z(z) = \frac{1}{R} \sum_{i=1}^n \left\{ p_{ij} \left[\ln a_{i+1} - \ln \left(\frac{z}{R} \right) + \ln b_{j+1} \right] + p_{i,j+1} \left[\ln b_{j+2} - \ln b_{j+1} \right] + \dots + p_{i,j+K-2} \left[\ln b_{j+K-1} - \ln b_{j+K-2} \right] + p_{i,j+K-1} \left[\ln \left(\frac{z}{R} \right) - \ln b_{j+K-1} - \ln a_i \right] \right\} \quad (6)$$

其中, j 和 K 是使下面两个不等式成立的正整数:

$$b_j \leq \frac{z}{Ra_{i+1}} < b_{j+1},$$

$$b_{j+K-1} \leq \frac{z}{Ra_i} < b_{j+K}.$$

计算时, 首先确定 Z 的取值区间 $[z_0, z_M)$ 。可令 $z_0 = R \cdot a_1 \cdot b_1$, $z_M = R \cdot a_{n+1} \cdot b_{m+1}$ 。然后用 M 个等距分割点 $z_1, \dots, z_M$ 将区间 $[z_0, z_M)$ 分割 (M 可取为 100, 间距 $\Delta z = (z_M - z_0) / M$)。利用 (6) 式求出相应的 $M+1$ 个储量分布密度值 $f_z(z_0), f_z(z_1), \dots, f_z(z_M)$ 。再利用下面的递推公式 (7), 算出 Z 的累积概率分布值 F_t 和大于累积概率分布值 G_t :

$$\begin{cases} F_0 = 0, \\ F_t = F_{t-1} + \frac{\Delta z}{2} [f_z(z_{t-1}) + f_z(z_t)], \\ G_t = 1 - F_t, \quad t = 1, \dots, M. \end{cases} \quad (7)$$

3. 两个局部区储量之和的分布

设 Z_1 和 Z_2 是两个不同局部区 (或不同层段) 的石油储量, 它们是相互独立的随机变量。设 $\phi_1(z_1)$ 和 $\phi_2(z_2)$ 分别是 Z_1 和 Z_2 的总体分布密度, 根据随机向量变换技术, 它们的和 $Q = Z_1 + Z_2$ 的分布密度 $\phi_Q(q)$ 理论上由下面的卷积公式确定:

$$\phi_Q(q) = \int_{-\infty}^{\infty} \phi_1(z_1) \phi_2(q - z_1) dz_1. \quad (8)$$

显然, 矩阵 P 仅是总体密度 $f(x, y)$ 的一种近似形式。但因为它根据取自于总体的样本求出的, 并且它无条件地充分利用了样本所提供的信息, 所以, 当样本容量 N 足够大时, 矩阵 P 是总体密度 $f(x, y)$ 的一种很好的近似。当可用的样本容量 N 太小时, 可以根据采样点的相对坐标, 利用平

设已利用(6)式分别求出了 $\phi_1(z_1)$ 和 $\phi_2(z_2)$ 的离散形式:

$$\phi_1(z_1): \begin{pmatrix} g_1 & g_2 & \cdots & g_n \\ z_{11} & z_{12} & \cdots & z_{1n} \end{pmatrix}, \quad (9)$$

$$\phi_2(z_2): \begin{pmatrix} h_1 & h_2 & \cdots & h_m \\ z_{21} & z_{22} & \cdots & z_{2m} \end{pmatrix}. \quad (10)$$

其中, z_{1i} 和 z_{2j} 分别是 Z_1 和 Z_2 的区间分割点, $g_i = f_{Z_1}(z_{1i})$, $h_j = f_{Z_2}(z_{2j})$, $i=1, \dots, n$; $j=1, \dots, m$.

从(8)式出发, 利用(9)和(10), 经过一系列推导(推导过程从略), 可以得出 $\phi_Q(q)$ 的数值计算公式:

$$\begin{aligned} \phi_Q(q) = & \sum_{i=1}^n [g_i h_j (z_{1,i+1} - q + z_{2,j+1}) + g_i h_{j+1} (z_{2,j+2} - z_{2,j+1}) + \cdots + \\ & + g_i h_{j+L-2} (z_{2,j+L-1} - z_{2,j+L-2}) + g_i h_{j+L-1} (q - z_{1i} - z_{2,j+L-1})]. \end{aligned} \quad (11)$$

其中, j 与 L 是使下面不等式满足的正整数:

$$z_{2j} \leq q - z_{1,i+1} < z_{2,j+1},$$

$$z_{2,j+L-1} \leq q - z_{1i} < z_{2,j+L}.$$

计算时, 首先确定储量 Q 的取值区间 $[q_0, q_M]$, 可令 $q_0 = z_{10} + z_{20}$; $q_M = z_{1M} + z_{2M}$. 然后用 M 个等距分割点 $q_1, \dots, q_M$ 将区间 $[q_0, q_M]$ 分割(间距为 $\Delta q = (q_M - q_0)/M$). 利用(11)式求出各分割点相应的储量分布密度值 $\phi_Q(q_t)$, $t=0, 1, \dots, M$. 最后利用下面的递推公式(12), 求出 Q 的累积概率分布值 V_t 和大于累积概率分布值 W_t :

$$\begin{cases} V_0 = 0, \\ V_t = V_{t-1} + \frac{\Delta q}{2} [\phi_Q(q_{t-1}) + \phi_Q(q_t)], \\ W_t = 1 - V_t, \quad t = 1, \dots, M \end{cases} \quad (12)$$

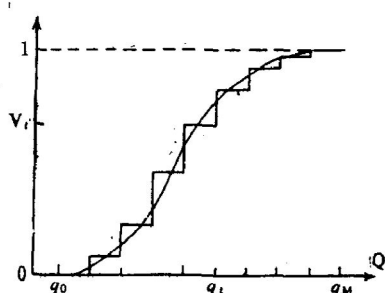


图3 累积分布函数曲线

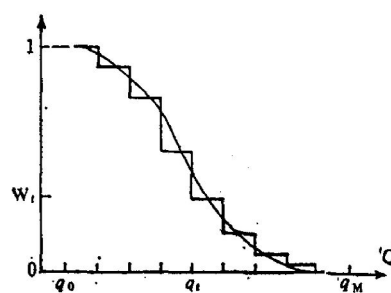


图4 大于累积分布函数曲线

图3和图4分别是 V_t 和 W_t ($t=1, \dots, M$) 构成的阶梯状曲线和它们的近似曲线。出于石油资源评价的特殊需要, 通常采用形如图4的“大于”累积概率分布曲线。

四、应用实例

本文应用经验分布变换法计算了我国S盆地下第三系生油层 T_1-T_2 段的石油地质储量。同时也用蒙特卡洛法作了计算, 以便比较计算结果。

S盆地在早第三纪始新世—渐新世为断陷型构造盆地，在近4000万年时间内堆积了5000—7000米厚的下第三系。其中暗色生油岩厚度占20%—25%。生油岩系沉积后，盆地持续沉降，又连续沉积了1500—2500米厚的新地层。S盆地面积虽小，但生油层厚度大，分布广，转化程度高，烃类成熟的生油区面积占盆地总面积的47%。因此，该盆地具有丰富的油源。

S盆地主要生油层是下第三系 T_1 — T_6 段。表1列出了本文计算时采用的数据。两种方法的计算所用的数据相同。为比较计算结果，蒙特卡洛法采用5种抽样次数分别作了计算。部分计算结果列于表2。图5是用经验分布变换法计算得出的S盆地总石油地质储量的概率分布密度曲线。图6是两种方法得出的总石油地质储量的大于累积概率分布曲线。这些计算结果与其它单位对S盆地的石油资源评价结果基本吻合。

表1 S盆地石油地质储量计算数据

	T_1 段	T_2 段	T_3 段	T_4 段	T_5 段	T_6 段
面积(平方公里)	751.2	1367.2	1183.2	1479.2	1672.8	1506.4
采用数据个数						
暗色泥岩厚度	37	59	63	66	97	38
氯仿沥青A含量百分数	37	59	63	66	97	38

$$d=23\text{亿吨/立方公里}; k_1=0.15; k_2=0.3$$

表2 S盆地石油地质储量部分计算结果

	抽样次数	计算时间	单边区间估计		数字特征				双边区间估计		
			$p=0.9$	$p=0.1$	均值	中位数	标准差 σ	众数 Q_M	$p=0.8$ 的区间	$Q_M \pm \sigma$ 区间 及其概率 p	$Q_M \pm 2\sigma$ 区间 及其概率 p
蒙特卡洛法	2500	2'47"	4.34	8.80	6.49	6.37	1.71	6.27	4.34—8.80	4.56—7.99 $p=0.6831$	3.01—9.71 $p=0.95$
	5000	3'43"	4.32	8.84	6.54	6.42	1.75	6.50	4.32—8.84	4.74—8.25 $p=0.68$	2.98—10.01 $p=0.95$
	10000	7'55"	4.34	8.60	6.40	6.28	1.66	6.06	4.34—8.60	4.39—7.72 $p=0.68$	3.07—9.39 $p=0.94$
	15000	11'14"	4.31	8.60	6.37	6.27	1.66	6.50	4.31—8.60	4.83—8.16 $p=0.67$	3.17—9.82 $p=0.95$
	20000	14'57"	4.31	8.59	6.38	6.29	1.66	6.28	4.31—8.59	4.61—7.94 $p=0.68$	2.97—9.60 $p=0.95$
	平均		4.32	8.68	6.44	6.33	1.69	6.32	4.32—8.68	4.63—8.01 $p=0.68$	3.04—9.71 $p=0.95$
经验分布变换法		1'41"	3.81	9.09	6.24	6.01	2.06	5.40	3.81—9.09	3.33—7.46 $p=0.69$	2.50—9.53 $p=0.92$

单位：亿吨

五、储量计算结果的概率解释

既然石油储量计算的结果用概率分布及其数字特征的形式表示，那么解释这种结果就要按照概率统计的原理，用概率统计的术语。近年来文献中介绍过几种解释的方法，

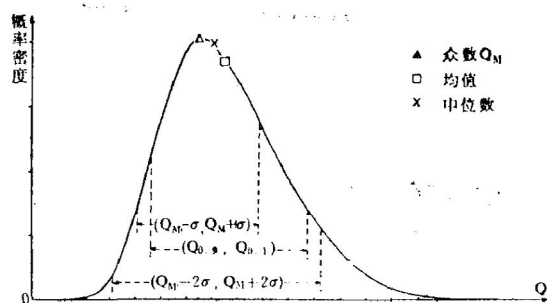


图5 S盆地石油储量概率分布密度曲线

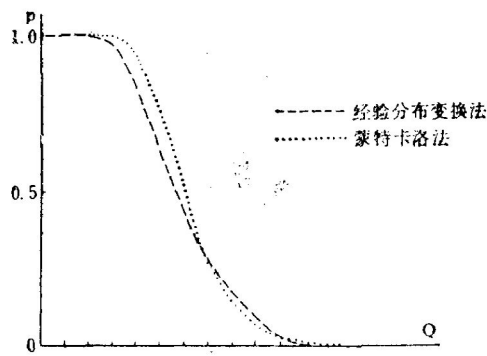


图6 S盆地石油地质储量概率分布曲线

但大都未能充分体现用概率统计方法计算石油储量所具有的优越性。在总结以往方法的基础上, 本文根据概率分布函数的性质, 结合石油资源评价的需要, 提出了一套建议性的解释方法。下面结合对S盆地储量计算结果(经验分布变换法结果)的解释, 来说明这套解释方法。

1. 单边区间估计

设 Q_p 满足 $p = P(Q \geq Q_p)$ 。给定概率值 p 以后, 便可以利用大于累积概率分布函数(形如图6的曲线或形如(9)式的数值表)求出 Q_p 。由此做出估计: “石油储量 Q 大于 Q_p 的概率是 p ”。可称这种估计为“单边区间估计”。如取 $p = 0.9$ 和 $p = 0.1$, 从表2可知, $P(Q \geq 3.81) = 0.9$ 和 $P(Q \geq 9.09) = 0.1$, 即S盆地的石油地质储量大于3.81亿吨的可能性是90%, 而大于9.09亿吨的可能性只有10%。

2. 众数单值估计

众数点使概率分布密度达到最大(见图5), 所以, 储量分布的众数是储量最可能的取值。作为资源量的单值估计, 众数比中位数或均值更具有代表性。从表2可知, $Q_M = 5.4$ 亿吨, 所以, S盆地石油地质储量为5.4亿吨左右的可能性最大。

3. 以众数为中心的双边区间估计

区间 $(Q_M - \sigma, Q_M + \sigma)$ 是以众数 Q_M 为中心, 以标准差 σ 为半径的“双边估计区间”。这种区间的概率 p 可由下式求出:

$$\begin{aligned} p &= P(Q_M - \sigma \leq Q < Q_M + \sigma) \\ &= P(Q \geq Q_M - \sigma) - P(Q \geq Q_M + \sigma) \end{aligned}$$

p 一般为0.6—0.7。概率虽不大, 但区间比较窄, 又位于分布的中部, 所以这种区间估计较为可靠。从表2可知, $P(3.33 \leq Q < 7.46) = 0.69$ 。即, S盆地石油地质储量以69%的概率位于区间(3.33, 7.46)中。还可以如此求出 $Q_M \pm \sigma$ 和 $Q_M \pm 2\sigma$ 区间的位置。

4. 具有指定概率的双边区间估计

在单边估计区间的基础上, 可由下式求出具有指定概率的双边估计区间 (Q_{p_1}, Q_{p_2}) :

$$\begin{aligned} p &= P(Q_{p_1} \leq Q < Q_{p_2}) \\ &= P(Q \geq Q_{p_1}) - P(Q \geq Q_{p_2}) = p_1 - p_2 \end{aligned}$$

显然,对于给定的概率 p ,可以求出许多这种区间。如当 $p=0.8$ 时,可求出 $(Q_{0.99}, Q_{0.19})$, $(Q_{0.95}, Q_{0.15})$, ..., $(Q_{0.38}, Q_{0.08})$ 等等。但应选择那些把众数、中位数和均值包围在中部的区间,才有助于准确地估计石油储量。从表2可知, S盆地的石油地质储量以80%的概率介于3.81—9.09亿吨之间。由图5可知上述区间 $(Q_{0.9}, Q_{0.1})$ 的位置。

后两种区间估计方法的优点是,它们不仅给出取值区间,而且给出了在该区间取值的概率。

六、结 论

1. 经验分布变换法是一种新的计算石油储量的概率统计方法。它可以与蒙特卡洛法同时应用于石油储量计算,以便互相对照、互相补充。对 S盆地石油地质储量的计算结果证明了它的有效性和准确性。

2. 因为经验分布变换法无须大样本抽样或迭代等耗费时间的计算过程,而是一次达到应有的精度范围,因而大大节省了计算时间。从表2可知,与蒙特卡洛法相比较,在精度相近的条件下,经验分布变换法大约可节省50%到80%的计算时间。

3. 因为采用了经验联合分布密度阵进行计算,所以无须假定随机变量间的独立性,从而使储量计算结果的准确性得到提高。根据实际资料所做的相关系数检验和联列表中独立性的检验结果都表明, S盆地生油岩有效厚度 X 与氯仿沥青含量百分数 Y 之间不具有相互独立性。如主观假定它们独立,则将造成计算结果的误差。

4. 石油储量概率分布的解释问题是石油资源定量评价的一个重要问题。应根据概率分布的性质,结合资源评价的需要,用正确的方法进行解释,以便充分发挥概率统计方法的优点。

作为一种新出现的方法,经验分布变换法难免有不足之处和有待于解决的问题。它的正确性还有待于在实际应用中进一步验证。

本文工作得到张文佑教授热情关怀,赵旭东、刘承祚、孙惠文同志给予具体帮助,特此致谢。

(收稿日期 1982年12月29日)

主 要 参 考 文 献

- [1] 赵旭东, 范锦惠, 崔凤文, 1981, 石油资源评价的统计模拟法, 石油勘探与开发, 第1期, P.47.
- [2] Miller, B. M. etc., 1975, Geological estimates of undiscovered recoverable oil and gas resources in the United States. U. S. Geol. Survey Circular 725, P.78.
- [3] Jones, T. A., 1977, A Computer method to calculate the convolution of statistical distribution. J. of IAMG, Vol.9, No.6, P.635.

TRANSFORMATION OF EMPIRICAL DISTRIBUTION— A METHOD FOR PETROLEUM RESERVE EVALUATION

Tao Qi

(Institute of Geology, Academia Sinica)

Abstract

A new method for evaluating petroleum reserve—transformation of empirical distribution (TED)—is studied in this paper.

The TED method can be applied to obtain the probability distribution of petroleum reserve based on the equation $Z=R \cdot X \cdot Y$. Here X and Y are random variables representing certain geological or reservoir parameters, R is a constant coefficient and Z , the amount of petroleum reserve. The main points of the TED method are as follows: (1) the "density matrix of joint distribution" is used to be the approximation of the population distribution density function of X and Y ; (2) a group of numerical computation formulae for transformation of random vector is then derived to produce the curve of probability distribution of petroleum reserve Z .

Correct interpretation of reserve probability distribution is an important problem of quantitative appraisal of petroleum resources. This paper has also proposed a set of suggestive interpretation methods.

我国首届沉积岩成岩作用学术讨论会在成都举行

沉积岩的成岩作用,对油气以及其它多种矿床的形成,起着重要的控制作用。为了交流这一领域的研究成果,进一步提高普查勘探成效,来自石油、地质、煤炭、冶金、化工各部以及中国科学院和高等院校等部门的 54 个单位的代表 130 余人,1983 年 11 月 16 日至 21 日在成都市举行了我国首届沉积岩成岩作用学术讨论会。这次会议是由中国地质学会沉积地质专业委员会与四川省石油学会联合召开的。

会议共收到学术论文 60 余篇,其中大会和小组宣读约 50 篇,它们是我国沉积岩成岩作用研究现状的缩影,其中一部分具有较高的学术水平,受到与会者的称赞。在涉及石油地质科学领域的各论文中,下列成果引人注目:

1. 将碳酸盐岩成岩环境和成岩序列的研究,与孔隙演化的研究结合起来。成都地质学院、西南石油学院以及四川、华北、胜利等油田研究院的科研成果,都比较深刻地揭示了孔隙的演化和储层的分布规律。胜利油田的研究成果认为,该区奥陶系基岩油藏的次生孔隙主要是深层溶解作用所形成,且其形成时期大致和下第三系成油期相同;有的次生孔隙发育带,远离古潜山风化带。地矿部第八普查勘探大队对贵州都匀早泥盆世暗河沉积的成岩系列的研究,既阐明了母岩成岩的地质时代,又指出了暗河充填物强烈白云石化的成因。

2. 将碎屑岩成岩作用的研究,与次生孔隙的形成及其与油气生成、运移和聚集机理结合起来。长庆油田研究院对鄂尔多斯盆地侏罗系、三叠系和第三系低渗透率砂岩储层的次生孔隙与成岩圈闭的研究,成效显著。华北油田研究院的论文指出,冀中地区下第三系次生孔隙发育带与生油成熟门限深度有着良好的配置关系;而次生孔隙的发育,不仅与深(温)度、成岩环境和岩性有关,而且与所处构造部位、构造运动和地下水活动有关。北京大学地质系利用粘土矿物对成岩环境反应敏感这一特点,探讨了粘土矿物从成岩到变质过程的演化规律,提出了成岩作用各阶段的划分标志。

会上,有关同志介绍了国外沉积岩成岩作用研究的进展情况和某些新技术新方法的应用。

代表们对成岩作用各阶段的划分和判别标志,以及一些矿物在成岩作用中的演化机理,进行了热烈讨论。并就今后如何开展这些方面的研究,广泛地交换了意见。

会议期间宣读的主要论文,将在有关学术期刊上陆续发表。

(党仁珊)