

文章编号: 1001-1986(2011)02-0013-05

结合软数据的同位置协同克里金估值新方法

牛文杰¹, 孟宪海², 李吉刚²

(1. 中国石油大学机电工程学院, 山东 东营 257061;

2. 北京航空航天大学计算机学院, 北京 100083)

摘要: 协同克里金法结合软数据可以提高地质体建模中估值的准确性。通过分析传统的结合软数据同位置协同克里金估值过程中存在的问题及原因, 提出并推导了一种新的结合软数据的同位置协同克里金估值方法。该方法采用软数据的期望值以保证外推的可靠性, 利用主变量观测值滤去协同变量的期望值以估计变异函数参数。该方法的特点是: 当待估点位于主变量相关变程之内时, 估值的结果主要反映主变量变化特征; 当待估点落在相关变程之外时, 估值结果的差别主要来源于各自点上协同变量(软数据)的期望值。最后, 用实例验证了新估值方法的优点。

关键词: 同位置协同克里金法; 主变量; 协同变量; 软数据; 硬数据

中图分类号: P634 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-1986.2011.02.003

A new estimation method of collocated CoKrig combined with soft data

NIU Wenjie¹, MENG Xianhai², LI Jigang²

(1. College of Electromechanical Engineering, Petroleum University of China, Dongying 257061, China; 2. School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: CoKriging combined with soft data can improve the accuracy of building geological model. By analyzing the reasons for the problems of the conventional CoKriging and collocated CoKriging, this paper deduces a new collocated CoKriging with soft data. The method uses the expected value of soft data in order to ensure the reliability of extrapolation, the use of the main variables filtered the expected value of co-observations to estimate parameters of variogram. The advantage of the new collocated CoKriging is that the interpolation reflects feature of main variable where the points are within the interrelated range of main variable, and the difference of interpolation comes from expect of each co-variable where the points are beyond the interrelated range of main variable. Finally, examples show the advantages of the new valuation method.

Key words: collocated CoKriging; primary variable; co-variable; hard data; soft data

在地质体建模中, 大量涉及到对所研究变量的估值(或称为插值)过程。协同克里金估值技术(CoKriging)^[1-4]是一种多变量估值技术, 在地质勘探开发过程中, 精确的井位数据(硬数据)往往非常少, 而其软数据(例如地震测线数据)却非常多, 协同克里金法可以结合软数据提高估值的准确性。协同克里金法的简化形式——同位置协同克里金法(Collocated CoKriging)^[5-7]既利用了相关变量的信息, 又将参与估值的协同变量的观测点个数限制为一个(与被估点同位置的观测点), 简化了协同克里金估值过程, 提高了估值效率。实践中发现: 当研究区域不太大, 硬数据相对多时, 同位置协同克里金估

值效果好; 反之, 当研究区域比较大, 硬数据很少, 无法覆盖大部分研究区域时, 同位置协同克里金法估值过程则常常存在一些问题。针对这些问题, 本文拟探讨一种新的结合软数据的同位置协同克里金估值方法, 以期对完善地质体建模有所帮助。

1 同位置协同克里金法估值原理

为简化起见, 这里只给出两变量协同克里金法估值的原理。

假设在研究区域 A 内两个协同区域化变量分别为 $Z_1(u)$ 和 $Z_2(u)$, 它们的观测点分别是:

$$\{Z_i(u_i); i = 1, 2, \dots, n_i\};$$

收稿日期: 2010-05-23

基金项目: 教育部科学技术研究重点项目(108084)。

作者简介: 牛文杰(1967—), 男, 甘肃宁县人, 博士, 副教授, 硕士生导师, 从事计算机图形学、科学计算可视化以及地质统计学研究。

$$\{Z_2(v_j); j=1, 2, \dots, n_2\}。$$

对区域 A 中任意一个点 u_0 的协同克里金估值为：

$$[Z_1(u_0)]^* = \sum_{i=1}^{n_1} \lambda_i Z_i(u_i) + \sum_{j=1}^{n_2} \xi_j Z_2(v_j), \quad u_i, v_j \in A。 \quad (1)$$

式中 $\{\lambda_i; i=1, 2, \dots, n_1\}$ 及 $\{\xi_j; j=1, 2, \dots, n_2\}$ 是待定权系数，由下面方程组解出：

$$\begin{cases} \sum_{i=1}^{n_1} \lambda_i C_1(u_i, u_j) + \sum_{i=1}^{n_2} \xi_i C_{12}(u_i, v_j) + \mu_1 = C_1(u_0, u_j), \\ j=1, 2, \dots, n_1; \\ \sum_{i=1}^{n_1} \lambda_i C_{12}(u_i, v_j) + \sum_{i=1}^{n_2} \xi_i C_2(v_i, v_j) + \mu_2 = C_{12}(u_0, v_j), \\ j=1, 2, \dots, n_2; \\ \sum_{i=1}^{n_1} \lambda_i = 1; \\ \sum_{i=1}^{n_2} \xi_i = 0。 \end{cases}$$

式中 $C_1(u_i, u_j)$ 、 $C_2(v_i, v_j)$ 及 $C_{12}(u_i, v_j)$ 分别为主变量、协同变量的协方差函数及彼此之间的交叉协方差函数； μ_1 、 μ_2 为拉格郎日乘子。

当研究区域比较大时，经常利用待估点邻域 $D \subset A$ 内 n_1 个主变量观测点及 n_2 个协同变量观测点对其协同估值。此时，协同克里金估值可写为：

$$[Z_1(u_0)]^* - m_1 = \sum_{i=1}^{n_1} \lambda_i [Z_1(u_i) - m_1] + \sum_{j=1}^{n_2} \xi_j [Z_2(v_j) - m_2], \quad u_i, v_j \in D \subset A。 \quad (2)$$

式中 m_1 、 m_2 分别为待估点 u_0 邻域 D 内主变量及协同变量的局部期望值。

同样，利用马尔科夫简化模型的同位置协同克里金估计形式^[2]为：

$$\frac{[Z_1(u_0)]^* - m_1}{\sigma_1} = \sum_{i=1}^{n_1} \frac{\lambda_i [Z_1(u_i) - m_1]}{\sigma_1} + \frac{\xi [Z_2(u_0) - m_2]}{\sigma_2}, \quad u_i, u_0 \in D \subset A。 \quad (3)$$

式中 σ_1 及 σ_2 分别为主变量及协同变量在 u_0 处的方差，即： $\sigma_1 = \text{Var}[Z_1(u_0)]$ ； $\sigma_2 = \text{Var}[Z_2(u_0)]$ 。

权系数 $\{\lambda_i; i=1, 2, \dots, n_1\}$ 及 ξ 由下列方程组解出：

$$\begin{cases} \sum_{i=1}^{n_1} \lambda_i \rho(u_i, u_j) + \xi \rho_{12}(0) \rho_1(u_0, u_i) = \rho_1(u_0, u_i), \\ (j=1, 2, \dots, n_1); \\ \sum_{i=1}^{n_1} \lambda_i \rho_{12}(0) \rho_1(u_i, u_j) + \xi = \rho_{12}(0)。 \end{cases}$$

$$\text{式中 } \rho_{12}(0) = \frac{C_{12}(0)}{\sqrt{C_1(0)C_2(0)}}; \rho_1(u_i, u_j) = \frac{C_1(u_i, u_j)}{C_1(0)}。$$

2 实践中存在的问题及其原因

当研究区域比较大，硬数据非常少而无法覆盖大部分研究区域时，结合软数据进行同位置协同估值会发现：

a. 如果利用全部主变量观测点对待估点进行协同估值，则待估点经常落在主变量相关变程之外，由于协同估计时主变量的观测点对估值起主要作用，协同变量的观测点起辅助作用，因而利用式(1)对待估点估值的结果主要是主变量观测点外推的结果。反之，如果利用待估点邻域内的主变量及协同变量的观测点进行协同估值时，又会遇见邻域内无主变量观测点的问题。

b. 同样的道理，由于主变量观测点较少，当待估点局部邻域内不存在主变量观测点时，式(2)、式(3)对待估点局部期望值的获取比较困难，能够得到的也只能是主变量全部观测点对其期望值的外推结果，这一结果往往无法反映局部区域内主变量的变化特征。

通过分析发现，出现上述问题的原因在于没有充分利用软数据所提供的大量信息。具体来说，虽然软数据相对于硬数据精度比较低，但软数据期望函数和硬数据的期望函数一样，都反映了研究变量的变化规律。如果软数据量非常大，覆盖了整个研究区域，那么对于任意一点的软数据期望值的求取就特别方便，在保证克里金估值原理(无偏、估计方差最小的条件)的前提下，充分利用软数据的期望值就可以在在一定程度上保证外推的可靠性。另外，当主变量存在漂移成分时，要同时估计漂移形式与剩余(Residual)变异函数参数，一般比较困难。但从下面关系式(4)及残余变异函数的求取公式知，利用主变量(硬数据)观测值滤去协同变量(软数据)的期望后所得的值去估计变异函数参数，其结果与主变量残余变异函数的估值结果理论上是相同的，由此就可避免上述困难。在估计过程中，将软数据剩余变异函数的参数取值看成硬数据剩余变异函数取值的先验值并加以利用，就可以提高主变量变异函数参数估值的可靠性。

3 新的同位置协同克里金估值方法推导

假设 $\{Z(x), x \in A\}$ 和 $\{M(x), x \in A\}$ 是研究区域 A

内的两个协同区域化变量, 其中 $Z(x)$ 是硬数据(例如: 测井采样点数据), $M(x)$ 是软数据(例如: 地震测线数据)。

$$E[Z(u_i)] = \mu_Z(u_i), u_i \in A;$$

$$E[M(u_i)] = \mu_M(u_i), u_i \in A。$$

和贝叶斯克里金法^[5,8]一样, 将 $M(x)$ 看作是对 $Z(x)$ 的合理猜测, 且

$$\mu_Z(u) = \mu_M(u) + a \text{ (常数)}. \quad (4)$$

注意, 此处 $\mu_M(u)$ 并未按泛克里金法那样假设为低次多项式形式, $\mu_M(u)$ 可根据地质解释或参考物理背景及相关的计算等来确定, 假设为低次多项式只是获取 $\mu_M(u)$ 的一种近似方法。

首先, 重新定义两个区域化变量:

$$\{Z^T(u) = Z(u) - \mu_M(u), u \in A\};$$

$$\{M^T(v) = M(v) - \mu_M(v), v \in A\}。$$

此时, 有 $E[Z^T(u)] = a$ (常数), 则 $Z^T(u)$ 满足平稳假设; $E[M^T(v)] = 0$, $M^T(v)$ 满足平稳假设。

其次, 证明它们可以作为协同区域化变量。因为 $Z(u)$ 、 $M(v)$ 是协同区域化变量, 是相关的, 因此它们之间的交叉协方差函数或交叉变异函数存在。

如果交叉协方差 $C_{ZM}(u, v)$, $u, v \in A$ 存在, 则:

$$\begin{aligned} C_{ZM}(u_i, v_j) &= E\{[Z(u_i) - \mu_Z(u_i)][M(v_j) - \mu_M(v_j)]\} \\ &= E\{[Z(u_i) - \mu_M(u_i) - a][M(v_j) - \mu_M(v_j)]\} \\ &= E\{[Z^T(u_i) - E[Z^T(u_i)]]\{M^T(v_j) - E[M^T(v_j)]\}\} \\ &= C_{Z^T M^T}(u_i, v_j), u_i, v_j \in A。 \end{aligned}$$

即 $Z^T(u)$ 、 $M^T(v)$ 的交叉协方差函数 $C_{Z^T M^T}(u, v)$ 存在且等于 $Z(u)$ 、 $M(v)$ 的交叉协方差 $C_{ZM}(u, v)$ 。

当 $Z(u)$ 、 $M(v)$ 满足二阶平稳(或内蕴)假设时:

$$E[Z(u)] = \mu_Z \text{ (常数)}, E[M(u)] = \mu_M \text{ (常数)}。$$

如果交叉变异函数 $\gamma_{ZM}(u, v)$, $u, v \in A$ 存在, 则:

$$\begin{aligned} \gamma_{ZM}(u_i, v_j) &= \frac{1}{2} E\{[Z(u_i) - Z(v_j)][M(u_i) - M(v_j)]\} \\ &= \frac{1}{2} E\{[Z(u_i) - \mu_M - [Z(v_j) - \mu_M]] \times \\ &\quad \{M(u_i) - \mu_M - [M(v_j) - \mu_M]\}\} \\ &= \frac{1}{2} E\{[Z^T(u_i) - Z^T(v_j)][M^T(u_i) - M^T(v_j)]\} \\ &= \gamma_{Z^T M^T}(u_i, v_j), u_i, v_j \in A。 \end{aligned}$$

当 $Z(u)$ 、 $M(v)$ 含有漂移成分且 $\gamma_{ZM}(u, v)$, $u, v \in A$ 存在时, $\gamma_{ZM}(u, v)$, $u, v \in A$ 应该为残余交叉变异函数:

$$\begin{aligned} \gamma_{ZM}(u_i, v_j) &= \frac{1}{2} E\{[R_Z(u_i) - R_Z(v_j)][R_M(u_i) - R_M(v_j)]\} \\ &= \frac{1}{2} E\{[Z(u_i) - \mu_Z(u_i) - [Z(v_j) - \mu_Z(v_j)]] \times \\ &\quad \{M(u_i) - \mu_M(u_i) - [M(v_j) - \mu_M(v_j)]\}\} \\ &= \frac{1}{2} E\{[Z(u_i) - \mu_Z(u_i) + a - [Z(v_j) - \mu_Z(v_j) + a]] \times \\ &\quad \{M(u_i) - \mu_M(u_i) - [M(v_j) - \mu_M(v_j)]\}\} \\ &= \frac{1}{2} E\{[Z^T(u_i) - Z^T(v_j)][M^T(u_i) - M^T(v_j)]\} \\ &= \gamma_{Z^T M^T}(u_i, v_j), u_i, v_j \in A。 \end{aligned}$$

因此, 不论 $Z(u)$ 、 $M(v)$ 是否满足二阶平稳(内蕴)假设, $Z^T(u)$ 、 $M^T(v)$ 的交叉变异函数 $\gamma_{Z^T M^T}(u, v)$ 均存在, 且等于 $\gamma_{ZM}(u, v)$ 。

以上说明 $Z^T(u)$ 、 $M^T(v)$ 之间是相关的, 可以作为协同区域化变量。

下面推导新的同位置协同克里金估值公式(利用马尔科夫简化模型)。

设 u_0 是研究区域 A 内任意一点, $Z^T(u_0)$ 的线性同位置协同克里金估计有以下形式:

$$[Z^T(u_0)]^* = \sum_{i=1}^{n_1} \lambda_i Z^T(u_i) + \xi M^T(u_0),$$

即:

$$\begin{aligned} [Z(u_0)]^* &= \sum_{i=1}^{n_1} \lambda_i [Z(u_i) - \mu_M(u_i)] + \\ &\quad \xi [M(u_0) - \mu_M(u_0)] + \mu_M(u_0) \end{aligned}$$

由无偏条件 $E[Z(u_0)] = E\{[Z(u_0)]^*\}$, 结合关系式(4)可得:

$$\begin{aligned} a + \mu_M(u_0) &= \sum_{i=1}^{n_1} \lambda_i [a + \mu_M(u_i) - \mu_M(u_i)] + \\ &\quad \xi [\mu_M(u_0) - \mu_M(u_0)] + \mu_M(u_0), \end{aligned}$$

即权系数关系式为:

$$\sum_{i=1}^{n_1} \lambda_i = 1。 \quad (5)$$

利用最小估计方差条件, 结合关系式(5)及马尔科夫简化模型, 可得到待定权系数 $\{\lambda_i; i=1, 2, \dots, n_1\}$ 、 ξ 的求解方程组:

$$\begin{cases} \sum_{i=1}^{n_1} \lambda_i \rho(u_i, u_j) + \xi \rho_{12}(0) \rho_1(u_0, u_i) = \rho_1(u_0, u_i), \\ (j=1, 2, \dots, n_1); \\ \sum_{i=1}^{n_1} \lambda_i \rho_{12}(0) \rho_1(u_i, u_j) + \xi = \rho_{12}(0); \\ \sum_{i=1}^{n_1} \lambda_i = 1。 \end{cases}$$

这时, $\rho_1(u_i, u_j)$ 、 $\rho_{12}(0)$ 分别为 $Z^T(u)$ 的自相关函数及 $Z^T(u)$ 与 $M^T(v)$ 之间的互相关系数。

4 同位置协同克里金估值新方法的性质及优点

新的同位置协同克里金估值方法是按照线性协同克里金原理推导出来的, 因此其满足无偏及估计方差最小条件。

从协同估计的表达式上看, 若所有有限个主变量观测点都参与估值, 当待估点位于主变量相关变程之内时, 估值的结果主要反映主变量变化特征; 对于任意两个落在相关变程之外的待估点来说, 由于主变量所有观测点对它们估值所起的作用相同(相当于外推), 因此这两点估值结果的差别主要来源于各自点上协同变量(软数据)的期望值, 即外推主要反映的是协同变量期望函数的变化规律。

对 $Z^T(u)$ 及 $M^T(v)$ 的变异函数求取可采用最大似然法^[9]。由于 $M(u)$ (软数据) 采样点个数往往非常多, 因此 $M^T(v)$ 的变异函数参数估计结果比较可靠。当求出 $M^T(v)$ 的变异函数参数后, 可将这些参数值作为 $Z^T(u)$ 变异函数参数取值的先验值, 即在最大似然法估计 $Z^T(u)$ 变异函数参数过程中作为迭代求解的初始值, 以提高参数估计的可靠性及效率。

5 应用实例

利用新的同位置协同克里金估值方法对我国玉门油田油区 T3 层顶界深度进行估值绘制等值线图(应用软件见文献[10])。T3 层顶界内测井数据(硬数据)13 个, 而地震测线数据(软数据)共 856 个。图 1、图 2 分别给出的是测井数据点和地震数据点分布图。图 3、图 4 是利用泛克里金法分别对测井数据和地震数据插值得到的深度等值线图; 图 5 是利用原同位置协同克里金法绘制的深度等值线图; 图 6 则是利用本文提出的新的协同克里金法绘制的深度等值线图。

从图 3 可以看出, 测井数据点主要集中在研究区域的左上部, 右下部测井数据点很少, 该区域的深度变化规律主要是外推的结果。另外, 在左中部靠近边界的区域, 等值线反映出深度变化比较平缓。

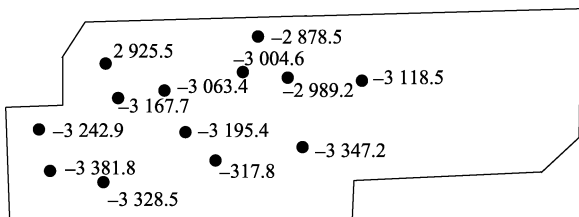


图 1 T3 层测井数据点区域分布图

Fig. 1 Regional distribution of log data points

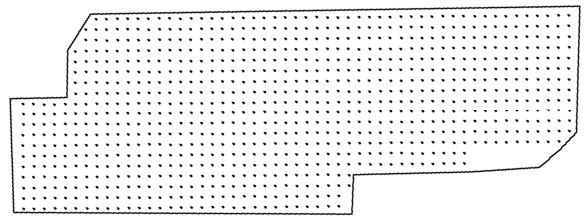


图 2 T3 层地震数据点区域分布图

Fig. 2 Regional distribution of seismic data points

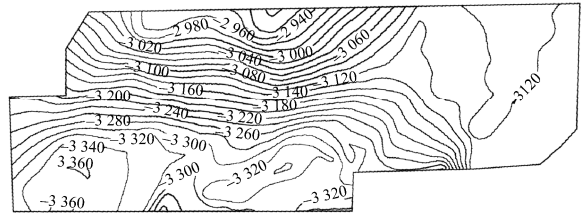


图 3 利用测井点数据绘制的 T3 层深度(m)等值线图

Fig. 3 The depth contour of logging points

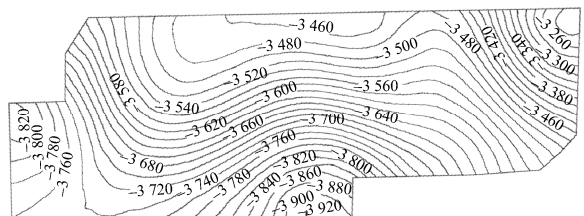


图 4 利用地震点数据绘制的 T3 层深度(m)等值线图

Fig. 4 The depth contour of seismic points

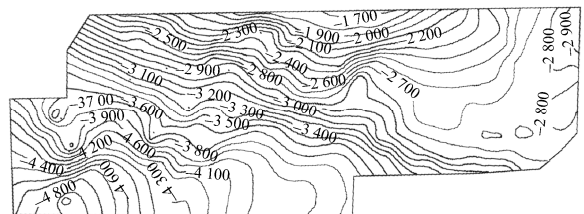


图 5 原同位置协同克里金法绘制的 T3 层深度(m)等值线图

Fig. 5 The depth contour by original collocated CoKriging

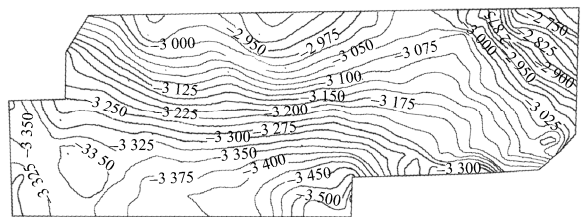


图 6 新的同位置协同克里金法绘制的 T3 层深度(m)等值线图

Fig. 6 The depth contour by new collocated CoKriging

从图 4 中发现, 区域的最右侧有一下降的凹陷, 这在图 3 的等值线中并没有反映出来, 原因在于此处没有测井数据点。图 4 的等值线也没有反映出左中部靠近边界区域深度平缓变化的趋势。

图 5 中, 虽然原同位置协同克里金法利用了软

数据信息,且它与图 3、图 4 相比,左中部靠近边界区域深度变化趋势有所改善(并不明显),但最右侧下降的凹陷特征没有反映出来。其原因是这一凹陷特征是由地震数据反映出来的,这一区域并没有测井数据,所有 13 个测井数据对于这一较远区域内点的期望值外推结果并不会反映这一特征。

图 6 中,在左中部靠近边界区域,深度变化反映出了测井数据的变化规律(与图 3 相似);在最右侧没有测井数据点的区域,也反映出了地震数据的变化规律,等值线明显地反映出凹陷特征。这是因为新的同位置协同克里金法估值对于外推主要反映地震数据点期望值的变化规律。从图 3—图 6 的比较结果来看,图 6 的可靠性比图 3、图 4 及图 5 的均要高。

6 结 论

同位置协同克里金法可以结合软数据提高估值的准确性。当研究区域不太大,硬数据相对多时,同位置协同克里金估值效果好;反之,当硬数据太少,无法覆盖大部分研究区域(同位置)。本文分析了出现这些问题的原因,由此提出了一种新的结合软数据的同位置协同克里金估值方法。该方法基于软数据期望函数和硬数据的期望函数一样,都反映了研究变量的变化规律这一特点,利用了软数据的期望值以保证外推的可靠性,通过主变量观测值滤去协同变量的期望值所得到的剩余以估计变异函数参数,从而解决了当主变量存在漂移成分时,要同时估计漂移形式与剩余变异函数参数困难的问题。从应用实例可以看出,用该方法所得结果比较可靠,与传统方法相比具有一定的优势。但该方法的缺点

是软数据的方差并没有在估值过程中加以考虑,这有待在今后的研究中得以改进。

参考文献

- [1] DEUTSCH C V, JOURNEL A G. GSLIB :Geostatistical software library and user's guide [M]. London : Oxford University Press, 1992 : 231-237.
- [2] 李少华,张昌民,胡爱梅,等. 煤储层孔隙度的协同模拟[J]. 煤炭学报, 2007, 32(9): 980-983
- [3] 张景雄. 野外空间采样的渐进式策略[J]. 武汉大学学报(信息科学版), 2008, 33(5): 441-445.
- [4] 张少良,张兴义,于艳艳,等. 哈尔滨市辖区黑土速效养分空间异质性分析[J]. 土壤通报, 2008, 39(6): 1277-1283.
- [5] 牛文杰. 贝叶斯同位置协同克里金(Bayesian collocated CoKriging) 估值法的研究[J]. 计算机辅助设计与图形学学报, 2002, 14(4): 343-347.
- [6] 刘文岭,夏海英. 同位协同克里金方法在储层横向预测中的应用[J]. 勘探地球物理进展, 2004, 33(5): 367-370.
- [7] JOURNAL W X. Integrating seismic data in reservoir modeling : the collocated cokriging alternative [C]//Proceedings of 67th Annual Technical Conference of the Society of Petroleum Engineers, Paper SPE 24742, 1992 : 546-551.
- [8] HENNING M. Bayesian Kriging-Merging observation and qualified guesses in Kriging [J]. Mathematical Geology, 1987, 19(1): 25-39.
- [9] PARDO I E, DOWD P A. Mlreml : a computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood estimation [J]. Computers & Geosciences, 1997, 23(2): 153-162.
- [10] 牛文杰. 三维数据场可视化克里金建模及其算法的理论和应用研究[D]. 北京: 北京航空航天大学, 2002 : 184-191.

西北矿井水文地质研究院在乌海揭牌成立

近日,由中国煤炭科工集团西安研究院与神华集团共同组建的西北矿井水文地质研究院在乌海市揭牌成立。神华集团目前有 28 对水文地质条件复杂及以上的矿井,不同程度受到水害威胁,水害防治难度巨大。如何有效解决这一问题,是神华集团当前及今后一个时期安全工作的重点和难点。组建西北矿井水文地质研究院,就是要充分利用中国煤炭科工集团西安研究院的技术优势,为神华集团全面构建本质安全型、质量效益型、科技创新型、资源节约型、和谐发展型的“五型企业”,实现煤矿生产“零”死亡的安全生产奋斗目标,在矿井水害综合防治等方面开展安全技术和专业化服务,将“预测预报、有疑必探、先探后掘、先治后采”的防治水原则和“防、堵、疏、排、截”的综合防治水措施落到实处,从而杜绝重大水害事故。

(梅 新)