doi:10.3799/dqkx.2015.160

IDW矿石品位估值结果的不确定性评价

李章林,吴冲龙,张夏林,翁正平,王 平

中国地质大学计算机学院,湖北武汉 430074

摘要:作为一种常用的矿石品位估值方法,距离幂次反比(inverse distance weighting,简称IDW)计算结果的可信度难以评估.对此,提出了一种基于统计学理论的解决途径.将参与IDW估值的样品品位视为指示化阈值,推导得出估值过程中其分配到的权重,即为待估点取这样品值的概率.基于这一结论构建出IDW估计值对应的条件累计分布函数(conditional cumulative distribution function,简称CCDF),从而完成待估位置矿石品位不确定性的建模.以一典型矿区的实际矿体勘探数据为基础进行了对比测试.针对每个估值点的CCDF进行了中位数、期望、方差等信息的提取,验证表明这一方法的计算结果与样品数据的实际情况之间存在较好的吻合度.在一定程度上说明了它在IDW品位估值结果不确定性评价方面的有效性.

关键词:距离幂次反比;不确定性评估;条件累计分布函数;品位插值;矿床地质.

中图分类号:P61 **文章编号**:1000-2383(2015)11-1796-06 **收稿日期**:2014-12-23

Uncertainty Assessment for IDW Ore Grade Estimates

Li Zhanglin, Wu Chonglong, Zhang Xialin, Weng Zhengping, Wang Ping

College of Computer Science, China University of Geosciences, Wuhan 430074, *China*

Abstract: As a typical ore grade estimation method, inverse distance weighting (IDW) has a long-time standing deficiency in reliability evaluation for its estimates, which has seriously limited its practicability. A statistical solution is presented in this paper to address the issue. A key character of IDW, that weight of each participation sample is the probability that the current un-sampled value is equal to the corresponding sample value, is proposed and proved by considering the participation samples as cutoff values during an indicating process. Based on this character, the conditional cumulative distribution function (CCDF) models related to the points to be estimated are constructed to fully characterize the uncertainty of IDW estimates. At last, a case study is carried out using cross validation and a test dataset from exploration engineering of a real ore-body. It is found that the real character of test dataset has been successfully revealed by the medium, expected and variance drawing from the constructed CCDFs, showing that the theoretical analysis and practical application can prove the validity and practicability of the proposed method to some extent.

Key words: inverse distance weighting; uncertainty assessment; conditional cumulative distribution function; grade interpolation; ore geology.

不确定性描述的是可以视为真实值的测量值或估计值的离散程度(Ellison *et al*., 2000),是近年来地质、地理、土壤等众多学科领域广泛关注和研究评价的重要参数(Chilès and Delfiner, 1999; Goovaerts, 2001; Zhang and Goodchild, 2002; Caers, 2011).对未采样点矿石品位值进行估计是矿山日常勘探、生产过程中的关键工作点,也是"数字矿山"建设过程中的重要内容(张夏林等, 2010; 吴冲龙等, 2011).其估值结果的不确定性大小,将直接影响矿山勘探或开采方案设计、品位控制、生产计划定制、风险评估、辅助决策等一系列工作的质量,甚至决定其成败.

距离幂次反比(inverse distance weighting,简称IDW)作为执行这项工作的一类重要方法,具有

基金项目:国家青年自然科学基金项目(Nos. 41202231,40802082);中央高校新青年教师科研启动基金项目(No. CUG110824).

作者简介:李章林(1982-),男,讲师,博士,主要从事三维地学模拟、地质统计学、数字矿山方面的研究工作. E-mail: lizhanglin@126.com

引用格式:李章林,吴冲龙,张夏林,等,2015. IDW矿石品位估值结果的不确定性评价. 地球科学——中国地质大学学报,40(11): 1796-1801.

原理简单、计算效率高和程序实现方便等优点(林连宝, 2003; 李章林和张夏林, 2007). 近年来,IDW 法在矿石品位空间结构表达方面的研究取得了不少成果(Tomczak, 1998; Ahrens, 2006; 牛文杰, 2011; 谭正华等, 2011),而在计算结果可信度评估方面的研究进展却较为缓慢,仅有少量的探索性工作(Achilleos, 2008; 李章林等, 2008a, 2008b).

本文将参照地质统计学法不确定性建模的思路,以经典的统计学理论为基础展开研究. 视待估位置处的矿石品位为随机变量,同时将从指示化变量的角度,推导提出参估样品分配到的 IDW 权值即为待估点取这一样品值的概率. 再根据这一结论构建待估点品位的条件累计分布函数(conditional cumulative distribution function,简称 CCDF),从而完成对估值结果不确定性的评价. 本文在理论分析的基础之上,将以一典型矿山的真实数据为基础,对这一方法的实际应用过程和效果进行详细的实验分析和对比验证.

1　方法模型

1.1　理论背景

1.1.1　IDW 法品位估值的基本原理　距离幂次反比法进行品位估值的基本思想为:距离待估点越近的样品将获得较大的权值;反之,则将获得较小的权值(Shepard, 1968).

设某一矿化区域内:$z(x)$表示待估点 x 的品位值;$z(x_i)$为第 i 个样品点 x 的品位值,d_i 和 λ_i 分别表示这一样品点与待估点 x 之间的距离和分配到的权值;n 为用于估值的样品个数,则 IDW 法计算估计值 $z^*(x)$的表达式:

$$z^*(x)=\sum_{i=1}^{n}[\lambda_i\cdot z(x_i)], \tag{1}$$

其中,计算第 i 个参估样品分配到的权值 λ_i 的经典公式为:

$$\lambda_i=\frac{\frac{1}{{d_i}^p}}{\sum_{i=1}^{n}\left(\frac{1}{{d_i}^p}\right)}, \tag{2}$$

上式中,距离参数 d_i 的幂指数 p 可以是任意自然数或小数,通常的取 2.0.

1.1.2　不确定性评价的基本方法　研究目标或对象的不确定性是目前各学科和行业都较为关注的问题(Chilès and Delfiner, 1999). 以地质统计学为代表,基于统计学理论的方法是当前国内外研究探索这一问题最为成熟和有效的途径之一(Isaaks and Srivastava, 1989). 其基本思路为:将待估点的属性值视为一个随机变量,通过构建其条件累计分布函数实现对这一随机变量的完全刻画,达到对未知点品位的不确定性进行充分评价的目的.

1.2　方法原理

将待估点(x)的品位 $z(x)$视为一个离散随机变量,记为 $Z(x)$,其取值为 $z(x_i)$的概率为 p_i. 以(.)和(.)* 分别表示变量和变量的估计值. 对于式(1)所描述的 IDW 估值过程,有如下两式成立:

$$z^*(x)=E^*[Z(x)], \tag{3}$$

$$\lambda_i=(p_i)^*, \tag{4}$$

可见,从统计学的角度,IDW 法具备以下 2 项基本性质:(1)未采样点的估计值可以视为这一随机变量的期望;(2)估值过程中参估样品分配到的权值可以视为这一随机变量取当前样品值的概率.

假设估值过程中样品品位值已经按由小到大的顺序排列,即对于$\{z(x_1),z(x_2),\cdots,z(x_n)\}$,存在 $z(x_1)\leqslant z(x_2)\leqslant\cdots\leqslant z(x_n)$,这一参估样品数据集对应的 IDW 权值为$\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$,则由式(4)可得:估计待估点 x 位置处品位值小于等于第 k($k=1,2,\cdots,n$)个样品品位值 $z(x_k)$的条件概率 $F[x_0,z(x_k)|(n)]$的公式为:

$$F[x_0,z(x_k)\mid(n)]=\sum_{i=1}^{k}\lambda_i, \tag{5}$$

上式即为 IDW 插值过程中,未采样点矿石品位 $Z(x)$条件累计分布函数的计算公式.

1.3　推导证明

本节将对式(3)和式(4)进行推导. 将 $z(x_i)$记为 z_i,$Z(x)$取值为 z_i 的概率 p_i 记为 $P\{Z(x)=z_i|(n)\}$,则它可以表达成:

$$P\{Z(x)=z_i\mid(n)\}=1\cdot P\{Z(x)=z_i\mid(n)\}+0\cdot P\{Z(x)\neq z_i\mid(n)\}. \tag{6}$$

利用式(7)可以为 $Z(x)$构造出一个以 z_i 为阈值的指示化变量 $\boldsymbol{I}(x;z_i)$:

$$\boldsymbol{I}(x;z_i)=\begin{cases}1, Z(x)=z_i;\\ \qquad\qquad i=1,\cdots,n;\\ 0, Z(x)\neq z_i;\end{cases} \tag{7}$$

上式中指示化随机变量 $\boldsymbol{I}(x;z_i|(n))$的期望 $E\{\boldsymbol{I}(x;z_i|(n))\}$计算公式如下:

$$\boldsymbol{E}\{\boldsymbol{I}(x;z_i\mid(n))\}=1\cdot P\{Z(x)=z_i\mid(n)\}+0\cdot P\{Z(x)\neq z_i\mid(n)\}. \tag{8}$$

对比式(6)和式(8)可知:

$$P\{Z(x)=z_i\mid(n)\}=\boldsymbol{E}\{\boldsymbol{I}(x;z_i\mid(n))\},\quad(9)$$

由上式可知，$Z(x)$取样品值 z_i 的概率即为其对应的指示化变量 $\boldsymbol{I}(x;z_i|(n))$的期望. 另外，对于 0、1 类型的指示化变量，可以直接使用其估计值作为它的期望的估计值(Isaaks and Srivastava，1989；Chilès and Delfiner，1999)，即有下式成立：

$$\boldsymbol{I}^*(x;z_i|(n))=\boldsymbol{E}^*\{\boldsymbol{I}(x;z_a|(n))\}.\quad(10)$$

由式(1)和式(7)，式(10)中的 $\boldsymbol{I}^*(x;z_i|(n))$可以进一步表达成：

$$\boldsymbol{I}^*(x;z_i|(n))=\sum_{i=1}^{n}[\lambda_i\cdot\boldsymbol{I}^*(x_i)]=\lambda_i.\quad(11)$$

由式(9)、式(10)和式(11)可得：

$$P^*\{Z(x)=z_i\mid(n)\}=\lambda_i.\quad(12)$$

由式(1)有：

$$z^*(x)=\sum_{i=1}^{n}[P^*\{Z(x)=z_i\mid(n)\}\cdot z(x_i)]=\boldsymbol{E}^*[Z(x)].\quad(13)$$

由式(12)和式(13)可知，上一节所提出的构建待估点 $Z(x)$条件概率分布函数的基本原理(即式(3)和式(4))成立.

1.4 特性分析

对于一个随机变量，确定了其分布函数就相当于掌握了这个随机变量的全部特征. 因此，通过构建 CCDF 可以实现这一随机变量不确定性的完全评价. 一旦 CCDF 构建完成，就可以通过从中提取各种参数或指标(如方差、置信区间等)实现估值结果不确定性的具体评价. 也可以从各个待估点的 CCDF 上随机提取多个实现，通过不同实现之间的差异程度对估值结果的不确性进行直观表达(Chilès and Delfiner，1999).

这种方法与地质统计学的不确定性评价方法的总体思路是一致的. 因此，其不确定性评价结果与地质统计学的评价结果之间存在一定的可对比性. 例如：本方法和指示克里格法(Isaaks and Srivastava，1989；李章林等，2008a，2008b)，都可以从其构建的 CCDF 函数中提取方差、均值等特征值或多个随机实现，分别研究这些 CCDF 特征值或随机实现与实际情况之间的接近程度，就可以对本方法和指示克里格法的不确定性评价效果进行对比.

由于 IDW 法的参估样品权值具有大于 0 和累计和为 1 的特点，利用式(5)构建未知点 CCDF 模型的过程不会存在类似于指示克里格法需要对其进行各种复杂校正的问题. 同时，所提的不确定性评价的方法，几乎不会为 IDW 法的矿石品位插值过程本身增加任何额外的计算量，仍将具有 IDW 本身的计算机程序实现方便和运行效率高的特点.

另外，注意到式(1)表达的插值过程同样适用于普通克里格法. 因此，所阐述的方法也同样应当适用于普通克里格法插值结果的不确定性评估. 事实上，下文将用到的由 CCDF 提取生成的条件标准差参数式(14)，本质上也就是能够评估普通克里格插值结果不确定性的插值标准差(Yamamoto，2000).

2 实验分析

2.1 实验方法

在交叉验证的方法框架下，以某矿体的实际勘探工程数据为基础，将每个样品数据视为未知点进行 IDW 估值并利用上面阐述的方法对其进行 CCDF 模型的构建. 然后从所构建的 CCDF 中提取相关信息(如中值、均值、方差等)并将其与实际情况进行分析对比，通过 CCDF 的基本信息与实际情况之间的差异程度对所提方法的有效性和实用性进行检验. 实验过程中，将从 CCDF 中分别提取条件标准差(Ellison *et al.*，2000)和估计标准差 σ(Hohn，1999)进行评估，其计算公式分别为：

$$s=\sqrt{\sum_{i=1}^{n}\lambda_i\cdot[z^*(x)-z(x_i)]^2},\quad(14)$$

$$\sigma=\frac{\varphi_{84}-\varphi_{16}}{2},\quad(15)$$

上式中，φ_{84}和 φ_{16}分别是第 84 分位数和第 16 分位数.

2.2 实验数据

实验数据选于我国南方某地区一个金矿的局部矿体，用到的样品数据主要为矿体内部的钻孔和平硐样品. 图 1 对矿体及其样品数据的空间分布情况和品位统计信息进行了描述. 可以看出，整体上样品品位的连续性较好，峰度和偏度分别为 16.82 和 3.08，呈明显的正偏态分布特征.

2.3 实验过程

以 2 作为幂指数、取与当前待估点最近的 8 个样品点进行估值，同时建立每个待估点对应的 CCDF 模型，根据统计学基本知识和式(14)、式(15)，从 CCDF 中提取其期望、中位数、条件标准差和估计标准差信息，对这些信息进行记录并与样品数据的实际品位值及其离散度进行比较，以验证 CCDF 模型构建结果的有效性.

2.4 实验结果及分析

图 2a 记录的是交叉验证结果中期望估计值(式

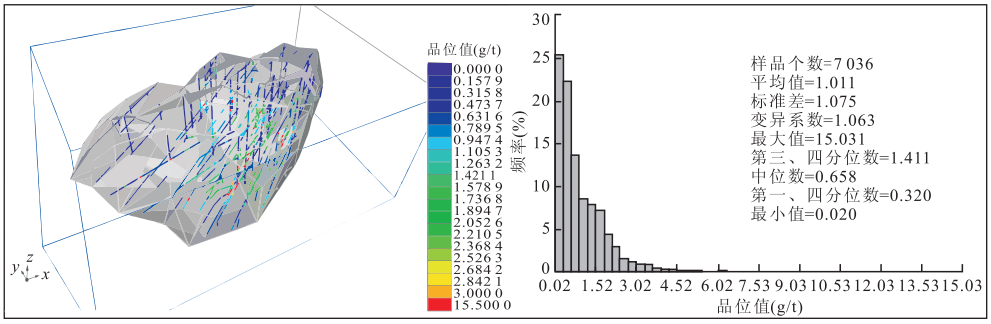

图 1　实验样品数据的空间位置及品位分布情况

Fig. 1　Spatial distribution and summary statistics of the test dataset

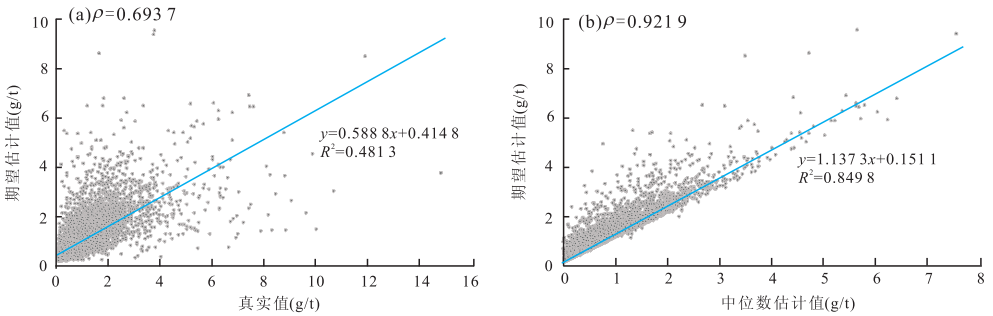

图 2　期望估计值与真实值(a)和中位数估计值(b)之间的关系

Fig. 2　Scatter plots of expected estimates and true values (a) and medium estimates (b)

ρ代表总体相关系数

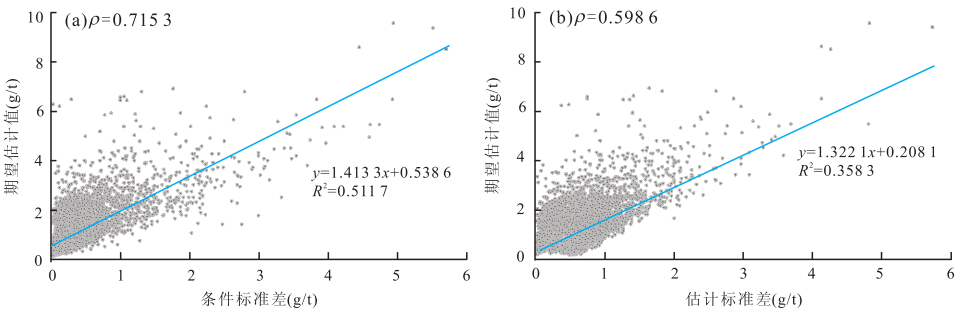

图 3　期望估计值与条件标准差(a)和估计标准差(b)之间的关系

Fig. 3　Scatter plots of expected estimates and conditional standard deviation (a) and estimation standard deviation (b)

ρ代表总体相关系数

(1)和式(3))与真实值的关系,从中可以看出期望估计值能够较好地对样品真实值进行反映. 图 2b 记录的是由 CCDF 得到的期望估计值与中位数估计值之间的关系,可以看到它们之间存在较强的相关性,故这 2 类估值结果在一定情况下都可以作为未知点的估计值使用. 另外,分析图 2b 不难发现,几乎所有的中位数估值结果都明显小于期望估计值,这一特征与图 1 中样品数据呈明显的正偏态分布实际特征是一致的. 因此,图 2 可以从均值和中位数的角度反映出 CCDF 模型构建结果的有效性.

比例效应指的是数据的局部均值与离散程度之间存在的一定的比例关系,是地质数据呈偏态分布情况下的一种常见属性(Yamamoto, 2000; Manchuk *et al.*, 2009). 由图 3 可以得出,由 CCDF 提取得到的条件标准差(式(14))和估值标准差(式(15))都能够与期望估计值之间呈现出较明显的比例关系. 说明这 2 类标准差均能够较好地反映出样品品位的局部离散程度,即能够能对估值结果的不确定性进行衡量. 分别比较其与期望估计值的相关系数可知,条件标准差的不确定性评价结果优于估计标

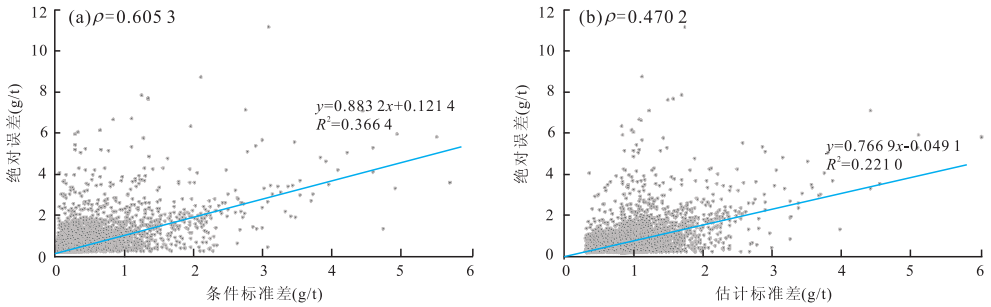

图 4 绝对误差与条件标准差(a)和估计标准差(b)之间的关系

Fig. 4 Scatter plots of estimate true errors and conditional standard deviations (a) and estimated standard deviations (b)

ρ代表总体相关系数

准差的评价结果. 但总体上,图 3 从方差(离散程度)的角度说明了 CCDF 模型构建结果的有效性.

图 4 从条件标准差与估计标准差和绝对误差(期望估计值与真实值之差的绝对值)的关系的角度展示了其作为不确定性评价参数的有效性. 可以看出,无论是条件标准差还是估计标准差,都可以在一定程度上反映出样品品位估值结果真实误差值大小,也进一步说明了估值过程中 CCDF 构建方法的有效性. 虽然与图 3 体现的性质类似,条件标准差对于样品品位估计值局部不确定性的揭示效果要优于估计标准差. 但在理论上,后者只是简单地通过两个分位数之差衡量变量的离散程度,其效果劣于经过严格计算的条件标准差也是容易理解和可以接受的.

结合以上对本次实验过程和效果的分析可知,由 CCDF 构建结果所提取出的 4 个基本参数,即中位数、期望、估值标准差和条件标准差,都能够在一定程度上对样品数据的实际均值、离散程度和偏态分布特征进行反映,从实践的角度证明了所提方法 CCDF 模型构建结果的有效性,实现了 IDW 矿石品位估值结果的不确定性评估.

3 结论

估值结果的不确定性评估对 IDW 矿石品位插值方法具有重要理论和实际意义,本文以经典的统计学理论为基础展开研究,通过构建待估位置矿石品位的 CCDF 完成其不确定性的全面表达:(1)从指示化随机变量的角度,推导并提出了 IDW 估值过程中参估样品分配到的权值可以视为待估点取这一样品值的概率的观点,并基于这一性质阐述了估计值 CCDF 模型的构建方法;(2)以一个矿体的实际勘探工程数据为基础,利用交叉验证技术对所提方法的 IDW 品位估计值不确定性评估效果进行了验证,研究得出所构建的待估点 CCDF 模型,其期望、中位数和方差均能够与样品数据的实际特征保持良好的一致性和吻合度,在一定程度上证明了这种方法的有效性;(3)所提的 IDW 品位插值结果不确定性评价方法具有一定的可扩展性. 虽然本文的研究是以 IDW 矿石品位估值这一特定应用领域为基础而展开,但理论上它还可以扩展应用至其他估值方法,如普通克里格法等;(4)另外,如何提高其实际应用效果是值得进一步研究的问题. 克服 IDW 法的固有缺陷,如选取最优幂指数、降低样品丛聚效应的影响等,可能是研究解决这一问题的有效途径之一.

References

Achilleos, G., 2008. Errors within the Inverse Distance Weighted (IDW) Interpolation Procedure. *Geocarto International*, 23 (6): 429 — 449. doi: 10. 1080/10106040801966704

Ahrens, B., 2006. Distance in Spatial Interpolation of Daily Rain Gauge Data. *Hydrology and Earth System Sciences*, 10 (2): 197 — 208. doi: 10. 5194/hess — 10 — 197—2006

Caers, J., 2011. Modeling Uncertainty in the Earth Sciences. John Wiley & Sons, Ltd., West Sussex, 246.

Chilès, J., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 695.

Ellison, S. L. R., Rosslein, M., Williams, A., 2000. Quantifying Uncertainty in Analytical Measurement (Second Edition). CZTAC, Atlanta, 120.

Goovaerts, P., 2001. Geostatistical Modelling of Uncertainty in Soil Science. *Geoderma*, 103 (1—2): 3—26. doi: 10. 1007/s00477—011—0496—2

Hohn, M. E., 1999. Geostatistics and Petroleum Geology.

Kluwer Academic Publishers, Dordrecht, 235.

Isaaks, E. H., Srivastava, R. M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, Oxford, 592.

Li, Z. L., Wang, P., Zhang, X. L., 2008a. Improvement and Application of Inverse Distance Weighted Method. *Metal Mine*, 328(4): 88－92 (in Chinese with English abstract).

Li, Z. L., Zhang, X. L., 2007. Designing and Realization of Mineral Resources Reserve Calculation Module Using Inverse Distance Square Method. *Geology and Prospecting*, 43(6): 92－97 (in Chinese with English abstract). doi: 10.3969/j.issn.0495－5331.2007.06.017

Li, Z. L., Zhang, X. L., Weng, Z. P., 2008b. Research and Application of Indicator Kriging Method in Ore Body Reserves Calculation. *Express Information of Mining Industry*, 24(1): 11－15 (in Chinese with English abstract).

Lin, L. B., 2003. Assessment and Development of Deposit Resource in Zijinshan Gold Mine. *Nonferrous Metals (Mine Section)*, 55(4): 19－21 (in Chinese with English abstract). doi: 10.3969/j.issn.1671－4172.2003.04.007

Manchuk, J., Leuangthong, O., Deutsch, C., 2009. The Proportional Effect. *Mathematical Geosciences*, 41(7): 799－816. doi: 10.1007/s11004－008－9195－z

Niu, W. J., 2011. Inverse Distance Weighting Method Based on Variogram. *Journal of Donghua University (Natural Science)*, 37(3): 362－367 (in Chinese with English abstract). doi: 10.3969/j.issn.1671－0444.2011.03.020

Shepard, D., 1968. A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. Proceedings of the 23rd ACM National Conference, New York, 517－524.

Tan, Z. H., Jing, Y. B., Wang, L. G., et al., 2011. Improved IDW Method for Ore Grade Estimation Based on Spatial Variation. *Journal of China University of Mining & Technology*, 40(6): 928－932 (in Chinese with English abstract).

Tomczak, M., 1998. Spatial Interpolation and Its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW)-Cross-Validation/Jackknife Approach. *Journal of Geographic Information and Decision Analysis*, 2(2): 18－30.

Wu, C. L., Tian, Y. P., Zhang, X. L., et al., 2011. Discuss on the Theory and Method about Construction of the Digital Mine. *Geological Science and Technology Information*, 30(2): 102－108 (in Chinese with English abstract). doi: 10.3969/j.issn.1000－7849.2011.02.017

Yamamoto, J. K., 2000. An Alternative Measure of the Reliability of Ordinary Kriging Estimates. *Mathematical Geology*, 32(4): 489－509. doi: 10.1023/A:1007577916868

Zhang, J., Goodchild, M. F., 2002. Uncertainty in Geographical Information. Taylor & Francis Ltd., London, 288.

Zhang, X. L., Wu, C. L., Weng, Z. P., et al., 2010. Research and Application of the Digital Mine Software Quanty Mine. *Earth Science—Journal of China University of Geosciences*, 35(2): 302－311 (in Chinese with English abstract). doi: 10.3799/dqkx.2010.031

附中文参考文献

李章林,王平,张夏林,2008a. 距离幂次反比法的改进与应用. 金属矿山,328(4): 88－92.

李章林,张夏林,2007. 距离平方反比法矿产资源储量计算模块设计与实现. 地质与勘探,43(6): 92－97.

李章林,张夏林,翁正平,2008b. 指示克里格法在矿体储量计算方面的研究与应用. 矿业快报,24(1): 11－15.

林连宝,2003. 紫金山金矿资源的评价与开发利用. 有色金属(矿山部分),55(4): 19－21.

牛文杰,2011. 基于变异函数的距离加权反比法. 东华大学学报(自然科学版),37(3): 362－367.

谭正华,荆永滨,王李管,等,2011. 基于空间变异性的 IDW 矿石品位估值改进方法. 中国矿业大学学报,40(6): 928－932.

吴冲龙,田宜平,张夏林,等,2011. 数字矿山建设的理论与方法探讨. 地质科技情报,30(2): 102－108.

张夏林,吴冲龙,翁正平,等,2010. 数字矿山软件(Quanty Mine)若干关键技术的研发和应用. 地球科学——中国地质大学学报,35(2): 302－311.