

自相关—判别分析与矿床统计预测

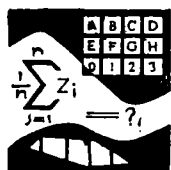
金友渔

(中国地质大学·武汉)

本文提出自相关—判别分析的数学模型并将其用于矿床统计预测之中。以宁芜盆地中段1:5万矿床统计预测为例,说明所提出的方法的应用效果。

关键词: 自相关—判别分析; 主导演化趋势因子; 矿床统计预测

主导演化趋势因子



数学地质

二维空间序列置于平面直角坐标系 TOS 中, 对于网格化数据, 某网格单元(以下简称单元)坐标为 (t, s) , 其邻近的8

个单元坐标如图1所示。

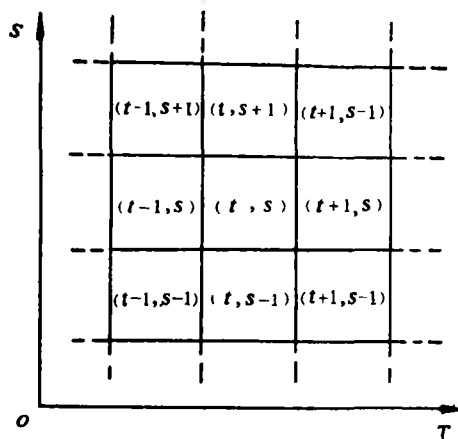


图1 二维空间序列的坐标系

在二维空间序列中, 笔者研究4个特定方向(1: EW, 2: NE-SW, 3: SN, 4: NW-SE)上的多元空间数据的相关性, 设任一单元 (t, s) 处都存在一个第 i 方向上的具有最强一阶自相关系数的 p 个变量的组合, 其形式为

$$u_i(t, s) = \sum_{j=1}^p c_{ij} x_j(t, s) \quad (1)$$

其中 $x_j(t, s)$ 为单元 (t, s) 第 j 变量的中心化变换(原始值减去第 j 变量均值所得之差)值, $c_{ij} (j=1, \dots, p)$ 为线性组合系数。

这种特定的线性组合反映了由 p 个变量构成的地质过程在第 i 方向上的演化趋势^[3], 称这个线性组合为第 i 方向上的演化趋势因子, 4个方向上的演化趋势因子的第一阶自相关系数分别为

$$\rho_1 = \frac{\text{Cov}[u(t, s), u(t \pm 1, s)]}{\sqrt{\text{Var}[u(t, s)] \cdot \text{Var}[u(t \pm 1, s)]}},$$

$$\rho_2 =$$

$$\frac{\text{Cov}[u(t, s), u(t \pm 1, s \pm 1)]}{\sqrt{\text{Var}[u(t, s)] \cdot \text{Var}[u(t \pm 1, s \pm 1)]}},$$

$$\rho_3 =$$

$$\frac{\text{Cov}[u(t, s), u(t, s \pm 1)]}{\sqrt{\text{Var}[u(t, s)] \cdot \text{Var}[u(t, s \pm 1)]}},$$

$$\rho_4 =$$

$$\frac{\text{Cov}[u(t, s), u(t \pm 1, s \mp 1)]}{\sqrt{\text{Var}[u(t, s)] \cdot \text{Var}[u(t \pm 1, s \mp 1)]}}$$

(2)

可以证明, 第 i 个方向上的演化趋势因子就是下列不对称矩阵 A_i 的最大特征值所

对应的特征向量 $C = (c_1, \dots, c_p)^T$, 而因子的第一阶自相关系数等于最大特征值。

$$A_i = \left[\sum^{(0)} \right]^{-1} \sum_i^{(1)} \quad (3)$$

其中 $\sum^{(0)}$ 为各单元 p 个变量的方差协方差阵, 其 j_1 行 j_2 列元素为

$$\begin{aligned} & \sum_{t=1}^{N_t} \sum_{s=1}^{N_s} x_{j_1}(t, s) \cdot x_{j_2}(t, s) \\ & j_1, j_2 = 1, \dots, p \\ & t = 1, \dots, N_t \\ & s = 1, \dots, N_s \end{aligned} \quad (4)$$

其中 N_t 、 N_s 分别为东西、南北方向上的单元数。

$\sum^{(1)}$ 为第 i 方向邻近单元的 p 个变量方差协方差阵, 又称为第一阶自协方差阵, 其 j_1 行 j_2 列元素为 (以 NE-SW 向上为例):

$$\begin{aligned} & \sum_{t=2}^{N_t-1} \sum_{s=2}^{N_s-1} x_{j_1}(t, s) \cdot x_{j_2}(t \pm 1, s \pm 1) \\ & j_1, j_2 = 1, \dots, p \\ & t = 2, \dots, N_t - 1 \\ & s = 2, \dots, N_s - 1 \end{aligned} \quad (5)$$

一般地说, 演化趋势因子具有各向异性, 其中第一阶自相关系数绝对值最大的因子称为主导演化趋势因子。

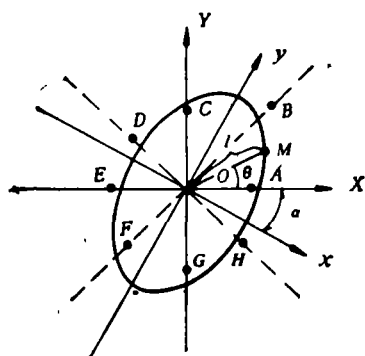


图 2 演化趋势因子各向异性椭圆的拟合

当某种地质过程存在一个主导方向的演化趋势, 并有规则的几何各向异性时, 可采用一个椭圆来表达这种各向异性, 方法如下:

引入二维直角坐标系 XOY (图2), A 、 B 、 C 、 D 、 E 、 F 、 G 、 H 8 个点分别在上述 4 个特定方向上, 它们与坐标原点的距离分别等于该方向演化趋势因子一阶自相关系数的绝对值。设所需拟合的椭圆的长短轴分别平行于 xOy 坐标系的 x 和 y 轴, x 轴正方向为由 X 轴正方向顺时针旋转 α 角而形成。在 xOy 坐标系中, 椭圆可表为:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (6)$$

在 XOY 坐标系中, 通过坐标变换, 该椭圆又可表为:

$$AX^2 + 2BXY + CY^2 = 1 \quad (7)$$

$$\text{其中, } A = \frac{b^2 \cos^2 \alpha + a^2 \sin^2 \alpha}{a^2 b^2},$$

$$B = \frac{\sin \alpha \cos \alpha (a^2 - b^2)}{a^2 b^2},$$

$$C = \frac{b^2 \sin^2 \alpha + a^2 \cos^2 \alpha}{a^2 b^2}$$

设 M 为椭圆上任一点, $OM = l$, θ 为 X 轴正方向反时针旋转到 $\overline{OM}$ 方向上的角度, 则有

$$X = l \cos \theta, Y = l \sin \theta$$

代入 (7) 式并经适当变形可得:

$$\begin{aligned} \frac{1}{l^2} &= A \cos^2 \theta + B (2 \sin \theta \cos \theta) \\ &+ C \sin^2 \theta \end{aligned} \quad (8)$$

引入新变量 $Z_1 = \cos^2 \theta$, $Z_2 = 2 \sin \theta \cos \theta$, $Z_3 = \sin^2 \theta$, 于是拟合椭圆的问题变为利用 A 、 B 、 C 、 D 、 E 、 F 、 G 、 H 8 个点所在的特定方向的方向角 θ 形成 Z_1 、 Z_2 、 Z_3 3 个自变量的 8 个样本值。以这 8 个点所在的特定方向上演化趋势因子一阶自相关系数绝对值平方的倒数为因变量 $1/l^2$ 的 8 个样本值进

行非线性回归（采用麦夸脱迭代法求解），从而确定椭圆参数 a 、 b 和 α ，即要求下列误差平方和达到最小：

$$Q = \sum_{i=1}^8 \left(\frac{1}{f_i^2} - AZ_{i1} - BZ_{i2} - CZ_{i3} \right)^2 \quad (9)$$

经 F 检验，回归方程高度显著时，所拟合的椭圆可以反映演化趋势因子趋势强度的几何各向异性， b 与 a 差异越大，椭圆越扁，主导方向越明显，由解出的 α 可推知主导演化趋势因子的大致方向，并可进而研究

相应地质过程的主导因素，需要指出的是，这种分析只有在拟合度很高时方可作为定性分析的参考，所求出的主导方向仍只是大致的方向，只有当分析的方向数较多时才能有较精确的解。

自相关—判别分析与矿床统计预测

自相关—判别分析的准则为：（1）组间差异尽量地大，组内差异尽量地小；（2）主导演化趋势因子所在的方向（设为第 i 方向）判别得分与演化趋势因子计量差异尽量地小，即要求下列统计量达到极大：

$$I = \frac{(\bar{R}_A - \bar{R}_B)^2}{\sum_{k=1}^{n_A} (R_{kA} - \bar{R}_A)^2 + \sum_{k=1}^{n_B} (R_{kB} - \bar{R}_B)^2 + \sum_{k=1}^N (R_k - \rho_i R_k^{(i)})^2} \quad (10)$$

其中， $R_{kA} = \sum_{j=1}^p \beta_j x_j(t_{kA}, s_{kA})$ 、 $R_{kB} =$

$\sum_{j=1}^p \beta_j x_j(t_{kB}, s_{kB})$ 分别为 A 和 B 组第 k 单元判别得分， (t_{kA}, s_{kA}) 、 (t_{kB}, s_{kB}) 分别为 A 和 B 组第 k 单元坐标， $x_j(t_{kA}, s_{kA})$ 、 $x_j(t_{kB}, s_{kB})$ 为相应单元第 j 变量取值， $R_k = \sum_{j=1}^p \beta_j x_j(t_k, s_k)$ 为整个研究区第 k 单元判别得分， $R_k^{(i)} = \sum_{j=1}^p \beta_j \cdot$

$x_j(t_k^{(i)}, s_k^{(i)})$ 为整个研究区第 k 单元 i 方向上的邻近单元的判别得分， $(t_k^{(i)}, s_k^{(i)})$ 为第 k 单元 i 方向上邻近单元的坐标， n_A 、 n_B 分别为 A 和 B 组单元数， N 为整个研究区单元数， ρ_i 为主导演化趋势因子一阶自相关系数。

由 $\frac{\partial I}{\partial \beta_j} = 0$ ， $j = 1, \dots, p$ ，可获得如下线性方程组

$$\begin{pmatrix} SS_{11} & SS_{12} & \dots & SS_{1p} \\ SS_{21} & SS_{22} & \dots & SS_{2p} \\ \dots & \dots & \dots & \dots \\ SS_{p1} & SS_{p2} & \dots & SS_{pp} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{pmatrix} \quad (11)$$

或 $SS \cdot B = D$ (12)

其中， SS 阵为 A 、 B 两组单元联合方差协方差及二维空间序列第 i 方向上一阶自协方差的 $p \times p$ 矩阵， B 为待定系数的 p 维列向量， D 为 A 、 B 两组单元各变量均值差的 p 维列向量。

SS 阵中第 j 行 h 列元素可表为：

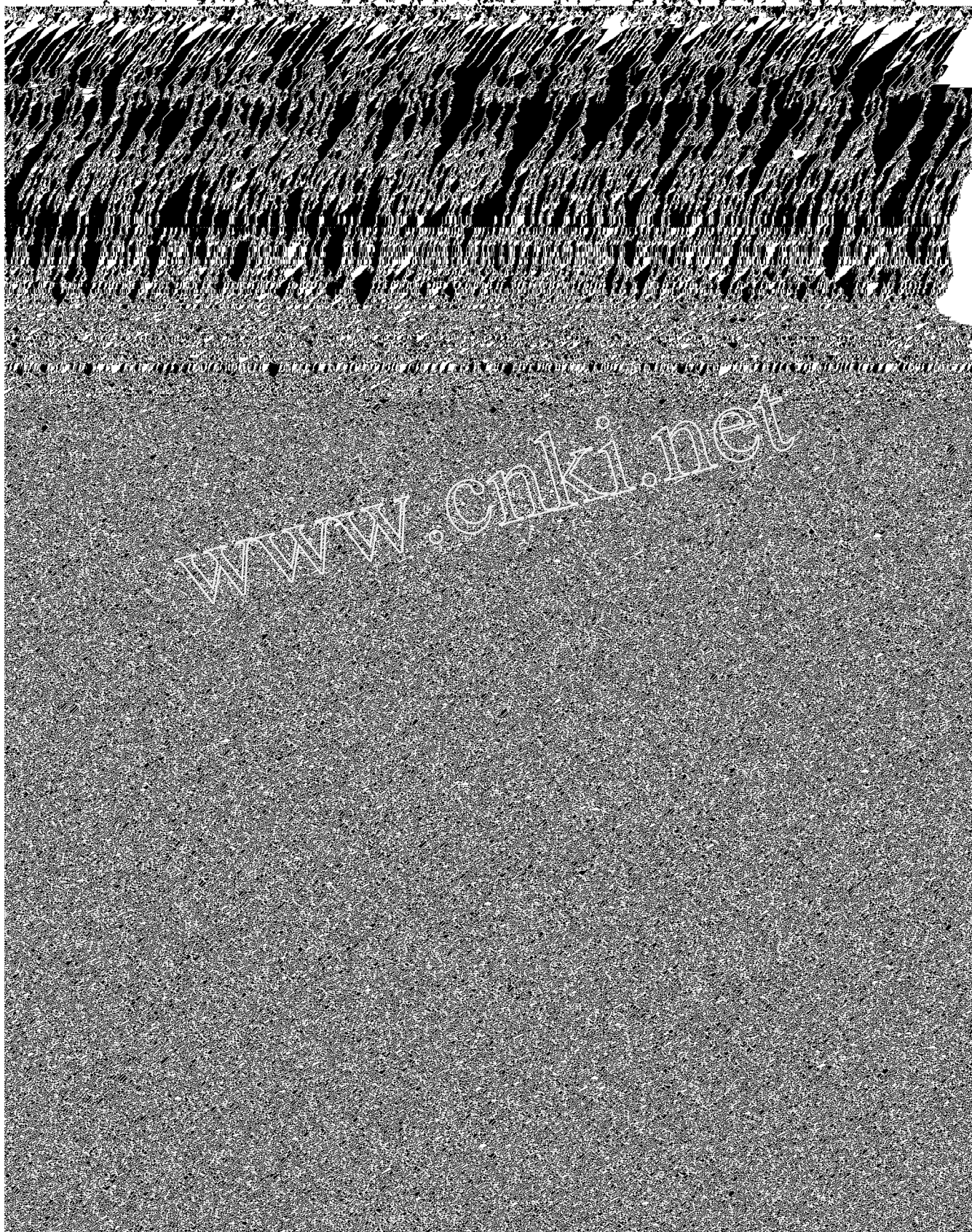
$$\begin{aligned} SS_{jh} = & \left[\sum_{k=1}^{n_A} (x_j(t_{kA}, s_{kA}) - \bar{x}_{jA}) \cdot \right. \\ & \cdot (x_h(t_{kA}, s_{kA}) - \bar{x}_{hA}) + \sum_{k=1}^{n_B} \\ & \cdot (x_j(t_{kB}, s_{kB}) - \bar{x}_{jB}) \cdot (x_h(t_{kB}, s_{kB}) - \bar{x}_{hB}) \left. \right] \\ & + \sum_{k=1}^N x_h(t_k, s_k) \cdot x_j(t_k, s_k) \\ & - \rho_i \sum_{k=1}^N x_h(t_k^{(i)}, s_k^{(i)}) \cdot x_j(t_k, s_k) \\ & - \rho_i \sum_{k=1}^N x_h(t_k, s_k) \cdot x_j(t_k^{(i)}, s_k^{(i)}) \\ & + \rho_i^2 \sum_{k=1}^N x_h(t_k^{(i)}, s_k^{(i)}) \cdot x_j(t_k^{(i)}, s_k^{(i)}) \end{aligned}$$

$$h, j=1, \dots, p$$

(13)

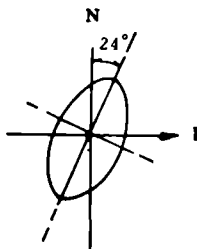
以宁芜盆地铁矿床统计预测为例, 说明

宁芜盆地铁矿床统计预测模型, 宁芜盆地铁矿床统计预测模型, 宁芜盆地铁矿床统计预测模型



宁芜盆地中段4个方向控矿因素演化趋势因子

表 1

方 向		EW	NE-SW	SN	NW-SE	各向异性椭圆
演化趋势因子系数	x_1	$C_{11} = -0.11810$	$C_{21} = 0.19541$	$C_{31} = 0.29338$	$C_{41} = -0.50577$	
	x_2	$C_{12} = -0.13885$	$C_{22} = 0.47757$	$C_{32} = 0.67432$	$C_{42} = 0.64888$	
	x_3	$C_{13} = 0.38779$	$C_{23} = 0.55554$	$C_{33} = 0.00444$	$C_{43} = 0.31525$	
	x_4	$C_{14} = 0.09144$	$C_{24} = 0.84724$	$C_{34} = 0.01253$	$C_{44} = 0.04767$	
	x_5	$C_{15} = -0.10080$	$C_{25} = -0.24976$	$C_{35} = 1.06863$	$C_{45} = 0.77009$	
自相关系数		$\rho_1 = -0.51299$	$\rho_2 = 0.70661$	$\rho_3 = 0.66786$	$\rho_4 = 0.43079$	

宁芜盆地中段铁矿化判别分析

表 2

判别模型	判 别 方 程	$\bar{R}_A$	R_0	$\bar{R}_B$	A 组 错判率	B 组 错判率	C 组 错判率
自相关一判别分析	$R = 0.0312x_1 + 0.0369x_2 + 0.0232x_3 + 0.0728x_4 - 0.0110x_5$	0.0935	0.0479	0.0024	1/7	0/10	3/13
费歇准则下判别分析	$R = 0.2810x_1 + 0.1525x_2 + 0.0917x_3 + 0.2317x_4 + 0.0356x_5$	0.5337	0.2536	0.0576	1/7	1/10	4/13

A组: 单元矿床值对数值 ≥ 3.0 $n_A = 7$, B组: 单元矿床值对数值 < 3.0 $n_B = 10$, C组: 外部检验组(矿床值已知) $n_C = 13$

拟合的各向异性椭圆长轴方向为NNE—SSW向($NE 24^\circ$), 椭圆方程拟合的复相关系数为0.96, 这一方向与该区NNE向闪长玢岩—断裂构造带的方向是一致的。

从单元矿床值对数值 ≥ 3.0 中选出7个单元作为A组, 单元矿床值对数值 < 3.0 中选出10个单元作为B组, 从整个研究区的30对($N = 30$)单元)NE方向上的邻近单元配对)按自相关一判别分析模型进行建模, 并留下13个已知矿床值的单元作为外部检验组(C组), 判别分析结果见表2所示, 为了对比, 同样采用A、B组共17个单元用费歇准则判别分析建模, C组作外部检验, 其结果也示于表2中, 由表2可看出自相关判别分析比普通判别分析效果有所提高。

本文提出的二维空间序列自相关一判别分析, 可使在应用判别分析进行矿床统计预测中引入二维空间相关的信息。即引入反映主导成矿作用方向的空间相关信息的主导演化趋势因子到判别准则之中, 在建立判别方程时将主导演化趋势因子方向上的邻近单元的信息加入, 并以主导演化趋势因子的一阶自相关系数 ρ_i 进行适当地控制, 从而提高了判别分析的效果。

参 考 文 献

- [1] 赵鹏大等,《矿床统计预测》,地质出版社,1983年。
- [2] 杨位钦、顾岚,《时间序列分析与动态数据建模》,北京工业学院出版社,1986年。
- [3] 於崇文等,《数学地质的方法与应用》,冶金工业出版社,1980年。

Statistical Prediction of Mineral Deposit by Autocorrelation-Discriminant Analysis

Jin Youyu

A mathematical model for statistical prediction of mineral deposit has been developed by the author from autocorrelation-discriminant analysis. Taking the work conducted in the middle segment of the Ning-Wu (Nanjing-Wuhu) Basin at a scale of 1:50000 as an example, it shows that the model proposed is able to yield good result in its application to ore-deposit prediction.