

# 基于偏最小二乘通径模型的 武汉市岩溶塌陷危险性预测

冯 永

(中国地质大学工程学院, 湖北 武汉 430074)

**摘要:**目前对岩溶地面塌陷的研究仍主要在宏观和定性的水平上。偏最小二乘通径模型采用一系列一元或多元线性回归的迭代求解,无需对观测变量做特定的概论分布假设,对样本点容量的要求也十分宽松,不存在模型不可识别的问题,是一种实用和有效的线性统计建模方法。本文在分析武汉市岩溶塌陷影响因素的基础上,基于GIS技术,采用偏最小二乘通径模型对武汉市岩溶塌陷危险性进行了预测。结果表明利用该模型可以取得较好的预测结果,在工程实际应用中有一定的借鉴意义。

**关键词:** 武汉; 岩溶塌陷; 偏最小二乘通径模型; 危险性预测

中图分类号: P642.25

文献标识码: A

文章编号: 1000-0844(2008)02-0128-04

## Application of the PLS Path Model to Forecasting Karst Collapse Hazard in Wuhan City

FENG Yong

(Engineering Faculty, China University of Geosciences, Wuhan 430074, China)

**Abstract:** The study of karst collapses is still in macroscopic and qualitative level right now. The PLS Path Model adopts the iteration solution of monogenesis or multivariant linear regression, does not need to hypothesize the probability distribution of observation variables. In this model the requirement of sample point capacitance is very loose, and the problem that the model can not be recognized does not exist, so it is an effective linearity statistics modeling method. In this paper, based on the analysis of the influence factors for karst collapses in Wuhan and GIS technique, the PLS Path Model is used to forecast the hazard potentiality of karst collapse in Wuhan city with a good results. It is found that the application of this model could get good effectiveness and this model should be adopted widely.

**Key words:** Wuhan city; Karst collapse; PLS Path Model; Hazard forecast

## 0 引言

目前对岩溶地面塌陷的研究仍停留在宏观的、定性的研究水平上。尽管如此,既使在目前的认识水平上,还是能够在某些多发区对这种灾害现象的发生条件和分布规律进行宏观研究,对岩溶塌陷危险性进行空间预测,可有望获得巨大的社会效益和经济效益。近年来许多学者对岩溶塌陷危险性评价

方法进行了研究,灰色统计法、神经网络、层次分析-模糊综合评判等最新方法的应用显著提高了预测定量化的精度,取得了较好效果。

偏最小二乘通径模型(PLS Path Model)采用的是一系列一元或多元线性回归的迭代求解。该模型无需对观测变量做特定的概论分布假设;也不存在

收稿日期:2007-08-22

基金项目:武汉市研究创新类计划项目(20066002059)

作者简介:冯 永(1984-),男(汉族),江苏徐州人,在读博士研究生,主要从事岩土工程勘察设计及地质灾害风险评价等方面的研究工作。

模型不可识别的问题;由于采用偏最小二乘方法,对样本点容量的要求也十分宽松。因此该模型是一种实用和有效的线性统计建模方法<sup>[1]</sup>,在岩溶塌陷危险性评价中有着广阔的应用前景。本文在分析武汉市岩溶塌陷影响因素的基础上,基于GIS技术,利用先进的偏最小二乘通径模型对武汉市岩溶塌陷危险性进行预测。

## 1 偏最小二乘通径模型

### 1.1 模型的设定

与结构方程模型类似,偏最小二乘通径模型也是由测量模型和结构模型组成的。通常也把测量模型称为外部模型,把结构模型称为内部模型。

设有 $J$ 组显变量,每组还有 $p_j$ 个变量,则每组显变量可以表示为

$$X_j = (x_{j1}, x_{j2}, \dots, x_{jp_j}) \quad (j = 1, 2, \dots, J) \quad (1)$$

通常假定显变量 $x_{jh}$  ( $j = 1, 2, \dots, J; h = 1, 2, \dots, p_j$ )都是基于 $n$ 个共同的观测点,并且每个变量都是中心化的。每组显变量 $X_j$ 对应隐变量 $\xi_j$  ( $j = 1, 2, \dots, J$ ),并且假定隐变量 $\xi_j$ 是标准化的,即均值为0,方差为1。

#### 1.1.1 测量模型(外部模型)

在测量模型中,一组显变量 $X_j$ 和对应的隐变量之间的关联关系可以用反映方式、构成方式2种形式来表示<sup>[1]</sup>。

#### 1.1.2 结构模型(内部模型)

结构模型描述不同隐变量 $\xi_j$ 之间的因果关系,通常由一组线性方程组来表示,即

$$\xi_j = \sum_{i \neq j} \beta_{ji} \xi_i + \zeta_j \quad (2)$$

其中 $\zeta_j$ 为随机误差项。式(2)说明隐变量之间存在相互制约的关系,可以看成是一个因果模型,这个因果模型必须是因果链,即在因果模型中不存在回环。

### 1.2 模型估计

要对显变量组 $X_j$  ( $j = 1, 2, \dots, J$ )所对应的隐变量 $\xi_j$ 进行估计,可以有2种方法:一种是根据显变量与隐变量间的关系,对隐变量进行计算,这称为外部估计;另一种方法是通过对隐变量之间的关联关系进行计算,这称为内部估计<sup>[1]</sup>。

#### 1.2.1 外部估计

隐变量可以有显变量 $x_{jh}$  ( $j = 1, 2, \dots, J; h = 1, 2, \dots, p_j$ )的线性组合来估计,记该估计量为 $Y_j$ ,且

$$Y_j = \left( \sum_{h=1}^{p_j} \tilde{\omega}_{jh} x_{jh} \right)^* = (X_j w_j)^* \quad (3)$$

其中: $w_j$ 为权重向量;星号\*表示对估计量进行标准化处理(在以下均采用这个记法)。

#### 1.2.2 内部估计

根据结构模型也可以认为,隐变量 $\xi_j$ 还可以通过与之关联的其他隐变量进行估计。这样得到的估计量被称为内部估计量,记为 $Z_j$ ,有

$$Z_j = \left( \sum_{i: \beta_{ji} \neq 0} e_{ji} Y_i \right)^* \quad (4)$$

其中: $\beta_{ji}$ 为式(2)中的方程系数; $e_{ji}$ 为称为内部权数。 $e_{ji}$ 的计算方法为

$$e_{ji} = \text{sign}(r, (Y_j, Y_i)) = \begin{cases} 1 & (r, (Y_j, Y_i) > 0) \\ -1 & (r, (Y_j, Y_i) < 0) \\ 0 & (r, (Y_j, Y_i) = 0) \end{cases} \quad (5)$$

其中: $\text{sign}$ 是符号函数; $r, (Y_j, Y_i)$ 表示外部估计量 $Y_j$ 和 $Y_i$ 的相关系数。

对于权重向量 $w_j$ 的估计方法,本文采用以下模式:

$$w_j = \frac{1}{n} X_j^T Z_j \quad (6)$$

权重向量 $w_j$ 是变量 $X_j$ 和 $Z_j$ 的相关系数。

### 1.3 偏最小二乘通径模型计算步骤

(1) 取向量 $Y_j$ 的初始值等于 $x_{j1}$ ;

(2) 通过式(4)计算 $Z_j$ 的估计值;

(3) 根据 $Z_j$ 的估计值,通过式(6)计算权重向量 $w_j$ ;

(4) 利用计算得到的 $w_j$ ,通过式(3)计算新的 $Y_j$ ;

(5) 再回到第2步,直到计算收敛为止,以最终得到的 $Y_j$ 作为对隐变量 $\xi_j$ 的估计值 $\hat{\xi}_j$ 。

## 2 基于偏最小二乘通径模型的武汉市岩溶塌陷危险性预测

### 2.1 危险性预测模型的建立

利用Mapgis软件,大体上按照规则的正方形单元进行划分(单元格的大小为150 m × 150 m),对于局部边界进行了处理,共计1 633个单元(图1)。选取岩溶基础条件、覆盖层特征、水文地质条件3组作为一级指标,选择这3组一级指标对应的13个指标作为二级指标,构成了武汉市岩溶塌陷危险性评估体系。将一级指标设为隐变量,其相应的二级指标作为显变量(表1),按照对岩溶塌陷危险性贡献大小对各个单元对应的13个显变量进行量化处理(表2)。

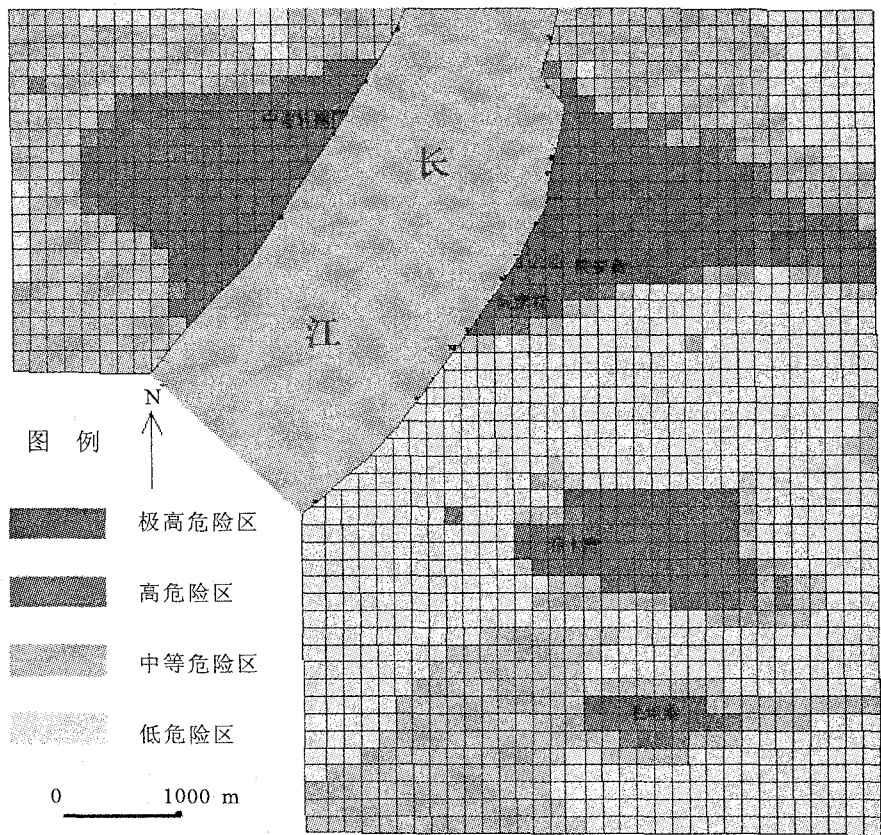


图1 武汉市岩溶塌陷危险性预测成果图

Fig.1 Map of karst collapse forecasting in Wuhan city based on the PLS Path Model.

表1 岩溶塌陷危险性评估显变量组与对应的隐变量

| 隐变量 | 岩溶塌陷基础条件( $\xi_1$ ) | 覆盖层特征( $\xi_2$ )  | 水文地质条件( $\xi_3$ )       |
|-----|---------------------|-------------------|-------------------------|
|     |                     |                   | 孔隙水和岩溶水联系情况( $x_{31}$ ) |
|     | 岩性( $x_{11}$ )      |                   | 枯水位( $x_{32}$ )         |
|     | 岩溶分布类型( $x_{12}$ )  | 覆盖层结构( $x_{21}$ ) | 洪水位( $x_{33}$ )         |
| 显变量 | 岩溶发育程度( $x_{13}$ )  | 覆盖层厚度( $x_{22}$ ) | 地下水波动情况( $x_{34}$ )     |
|     | 与构造关系( $x_{14}$ )   |                   | 距离长江远近( $x_{35}$ )      |
|     |                     |                   | 是否受水井影响( $x_{36}$ )     |
|     |                     |                   | 是否在水域范围内( $x_{37}$ )    |

表2 隐变量量化标准

| 指标量化标准                  | 100       | 60              | 30              | 0         |
|-------------------------|-----------|-----------------|-----------------|-----------|
| 岩性( $x_{11}$ )          | 灰岩        | 白云岩             | 砂岩              | 黏土岩       |
| 岩溶分布类型( $x_{12}$ )      | 覆盖型岩溶     |                 |                 | 埋藏型       |
| 岩溶发育程度( $x_{13}$ )      | 强         | 中强              | 中               | 弱         |
| 与构造关系( $x_{14}$ )       | 强         | 中强              | 中               | 弱         |
| 覆盖层结构( $x_{21}$ )       | 二元结构      |                 |                 | 一元结构      |
| 覆盖层厚度( $x_{22}$ )       | < 22 m    | 22 ~ 26 m       | 26 ~ 30 m       | > 30 m    |
| 孔隙水和岩溶水联系情况( $x_{31}$ ) | 强         | 中强              | 中               | 弱         |
| 枯水位( $x_{32}$ )         | > 20 m    | 17 ~ 20 m       | 14 ~ 17 m       | < 14 m    |
| 洪水位( $x_{33}$ )         |           |                 |                 |           |
| 地下水波动情况( $x_{34}$ )     | 强         | 中强              | 中               | 弱         |
| 距离长江远近( $x_{35}$ )      | < 2 000 m | 2 000 ~ 3 000 m | 3 000 ~ 4 000 m | > 4 000 m |
| 是否受水井影响( $x_{36}$ )     | 有         |                 |                 | 无         |
| 是否在水域范围内( $x_{37}$ )    | 有         |                 |                 | 无         |

对所有显变量进行标准化处理(变换13组变量的均值为0,方差为1),采用反映方式对3组显变量分别做主成份分析,进行唯一维度检验<sup>[1]</sup>。计算结果表明唯一维度条件都通过检验(说明分组合理)。可按照上述方法构造对武汉市岩溶塌陷危险性进行评估的偏最小二乘通径模型,建立显变量与隐变量,以及隐变量之间的路径关系。

由专业分析(层次分析)可以得到一级指标的权重向量,并用此权重向量修正算法中的第3步,可得到各隐变量的数值;进而可以计算隐变量与其相应的显变量之间的相关系数,表明各隐变量与其相应的显变量之间存在较明显的相关性,可建立隐变量对显变量的偏最小二乘回归模型。将各隐变量作为因变量,其对应的显变量为自变量,进行偏最小二乘回归,得到式(7);另外各隐变量之间不存在明显的相关性,所以采用多元回归模型,得到回归方程式(8):

$$\begin{aligned}\xi_1 &= 0.1777x_{11} + 0.5161x_{12} + 0.3175x_{13} + 0.0296x_{14} \\ \xi_2 &= 1.0114x_{21} + 0.2079x_{22} \\ \xi_3 &= 0.9712x_{31} + 0.0286x_{32} + 0.0329x_{33} + 0.0509x_{34} \\ &\quad + 0.046x_{35} + 0.0045x_{36} + 0.0043x_{37} \\ &\quad (7)\end{aligned}$$

$$\xi_4 = 0.5687\xi_1 + 0.1861\xi_2 + 0.5912\xi_3 \quad (8)$$

可以按照参考文献[1]计算此模型的调整复测定系数为0.999,很好的拟和了数据,各种系数与整体模型都通过了显著性检验。

## 2.2 基于 Mapgis 的预测结果

式(7)、(8)中各个变量的系数即为各指标变量的权重,进而可用式(7)、(8)确定单元 $\xi_4$ 值;通过 Spss 软件对所有单元 $\xi_4$ 数值进行直方图统计,可以确定各个危险性等级对应的临界值;利用 Mapgis 技术进行岩溶塌陷危险性评价<sup>[3]</sup>,得到的预测结果如

图1所示。通过以上计算,将研究区岩溶塌陷危险性分为四级:(1)极高危险区域:分布于灰岩地区,岩溶发育程度较高,而且覆盖层具有上粘下砂的二元结构。武汉市已有的岩溶塌陷均发生在该区域。(2)高危险区:分布于灰岩地区,岩溶发育程度虽较高,但覆盖层为一元结构,与极高危险区相比危险性有所降低。部分受开采井影响强烈的单元也分布在该区域内。(3)中等危险区:包括部分黏土岩厚度小于1m的区域,及受水域影响和构造影响比较强烈的区域。该区域应该注意地下水和地表水的维护,避免引起岩溶塌陷。(4)低危险区:没有可溶性灰岩,岩溶程度不发育,覆盖层的富水性也较差,因此发生塌陷的可能性较小。

从以上的危险性预测结果来看,利用该模型可得到武汉市岩溶塌陷比较可靠的预测结果。

## 3 结论

本文利用先进的偏最小二乘通径模型对武汉市岩溶塌陷危险性进行了预测,研究结果表明利用该模型可以取得较好的预测结果,在工程实际应用中有一定的借鉴意义。

### [参考文献]

- [1] 王惠文,吴载斌,孟洁. 偏最小二乘回归的线性与非线性方法[M]. 北京:国防工业出版社,2006.
- [2] 林盛,吴兵福,宁禄乔,等. 基于多个前变量 PLS 算法的一种迭代方法[J]. 系统工程学报,2006,21(4):446-448.
- [3] 陈静,马亚杰,朱庆杰. 人工神经网络模型在岩溶塌陷安全评价中的应用[J]. 地质灾害与环境保护,2005,16(2):139-144.
- [4] 殷坤龙. 滑坡灾害预测预报[M]. 武汉:中国地质大学出版社,2004.
- [5] 郑先昌. 武汉市岩溶塌陷诱发因素分析[J]. 城市勘测,2004,15(2):15-19.