

# 线性回归模型在北京平原地面沉降预测中的应用

范珊珊<sup>1</sup>, 郭海朋<sup>2</sup>, 朱菊艳<sup>1</sup>, 李文鹏<sup>3</sup>

(1. 中国地质大学(北京) 水资源与环境学院, 北京 100083; 2. 中国地质环境监测院, 北京 100081;

3. 中国地质调查局 水文地质环境地质调查中心, 河北 保定 071051)

**摘要:**北京平原的地面沉降日趋严重,地面沉降问题已经成为北京平原区最主要的地质灾害。本文在一系列地下水开采量、地下水水位和地面沉降量实测数据的基础上,运用 Excel 软件建立了地下水开采量与沉降量、地下水水位与沉降量之间的线性回归方程。基于地面沉降量和地下水开采量或地下水水位之间的回归方程对地面沉降量进行预测,并对预测值的可靠程度进行了验证。本文利用所建立的线性回归模型对地面沉降量进行预测,可以为地面沉降的控制提供依据。

**关键词:**线性回归;地面沉降;地下水开采;预测

**文章编号:**1003-8035(2013)01-0070-05

**中图分类号:**P642.26

**文献标识码:**A

## 0 引言

北京作为国家首都,一座国际性的大都市,随着城市规模的不断扩大,经济的快速发展和人口的持续增长,对水资源的需求量不断增加,由于大量开采地下水,造成地下水位大幅下降,导致北京平原形成了大面积的地面沉降区。由于北京平原的范围太广,本文选择地面沉降较严重的天竺地区作为典型研究区域,进而推断整个北京平原地面沉降的发展趋势。

随着数值模拟技术的快速发展,以地面沉降理论为基础,利用计算机建立了大量的地面沉降预测模型,其中应用比较广泛的有灰色预测模型,包括 GM(1,1)模型<sup>[1-2]</sup>和 GM(2,1)模型<sup>[3]</sup>,BP 神经网络模型<sup>[4-5]</sup>以及 MODFLOW 模型<sup>[6]</sup>等。近年来,建立地下水开采与地面沉降的线性回归方程也被广泛地应用于地面沉降的预测<sup>[7]</sup>。这种方法主要是依据最小二乘法原理得到回归方程的系数,建立地下水开采和地面沉降的线性回归方程,从而对沉降量进行预测。可以用于线性回归分析的软件主要有 Excel 和 SPSS,由于 Excel 使用普遍,操作简单,本文的回归分析主要是应用这种方法。

## 1 地下水开采与地面沉降的相关性分析

由于长期超量开采地下水,造成地下水位的大幅度下降,导致含水层上覆土层的孔隙水压力减小,使土层失水压缩固结,从而形成地面沉降。所以,地面沉降与地下水开采存在着一定的相关性。图 1 为北京市天竺地区中深部地层沉降的变化与该层段内地

下水水位变化曲线图,从图中可以看出,地面沉降量的变化趋势与地下水水位变化趋势基本一致。图 2 为累积地面沉降量和累积地下水开采量曲线图(考虑到地面沉降对地下水位下降有一定的滞后性,所以在分析时采用累积沉降量和累积开采量,从而减少滞后带来的影响<sup>[8]</sup>),如图 1 所示,地面沉降量随着地下水开采量的增加而增大。由此可见,地面沉降和地下水开采的相关程度较高。下面就对沉降量和地下水开采量或地下水水位的相关性进行定量分析。

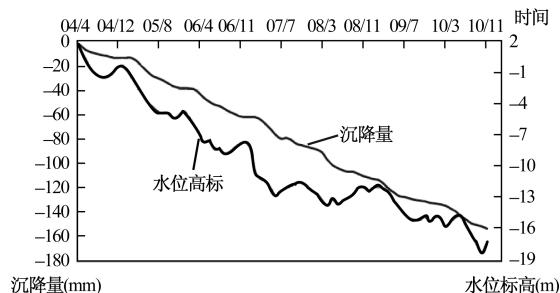


图 1 地面沉降量和地下水位变化曲线图

Fig. 1 Curves of subsidence quantity and groundwater level variation

收稿日期:2012-07-27;修订日期:2012-10-28

**基金项目:**国家重点基础研究发展计划“华北平原地下水演变机制与调控”(2010CB428806);中国地质调查局地质矿产调查评价专项项目(1212011140022)

**作者简介:**范珊珊(1987—),女,河北唐山人,硕士研究生,现于中国地质大学(北京)地下水科学与工程专业攻读硕士学位。

E-mail:shan shanf@126.com

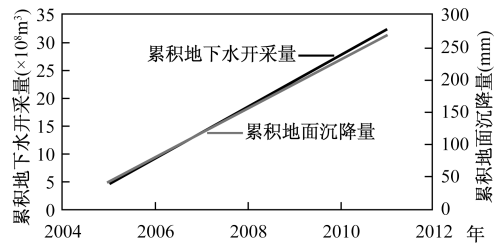


图 2 累积沉降量和累积开采量曲线图

Fig. 2 Curves of accumulative subsidence quantity and accumulative groundwater mining

表 1 为 2005 ~ 2009 年北京天竺地区地面年沉降量<sup>[9]</sup>和地下水开采量<sup>[10]</sup>统计表。

表 1 天竺地面年沉降量和地下水开采量统计表

Table1 Statistical table of Tianzhu subsidence quantity and groundwater mining

| 时间   | 年沉降量<br>(mm) | 累积沉降量<br>(mm) | 地下水年开<br>采量(×10 <sup>8</sup> m <sup>3</sup> ) | 累积开采量<br>(×10 <sup>8</sup> m <sup>3</sup> ) |
|------|--------------|---------------|-----------------------------------------------|---------------------------------------------|
| 2005 | 42.99        | 42.99         | 4.81                                          | 4.81                                        |
| 2006 | 38.07        | 81.06         | 4.75                                          | 9.56                                        |
| 2007 | 40.82        | 121.88        | 4.49                                          | 14.05                                       |
| 2008 | 30.19        | 152.07        | 4.68                                          | 18.73                                       |
| 2009 | 42.46        | 194.53        | 4.52                                          | 23.25                                       |

根据 Pearson 相关系数计算公式

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$
, 设累

积开采量为  $x$ , 累积沉降量为  $y$ , 则  $\sum xy = 10065.32$ ,  $\sum x = 70.4$ ,  $\sum y = 592.538$ ,  $\sum x^2 = 1203.308$ ,  $\sum y^2 = 84242.4$ ,  $n = 5$ , 代入公式得到相关系数为  $r = 0.998797$ 。

根据相关系数的判别规则,  $r > 0$  表明地面沉降量和地下水开采量成正相关, 与地下水开采量和地面沉降量的曲线图(图 2)所呈现的趋势是一致的, 即随着地下水开采量的增加, 沉降量也会增大。当相关系数在  $0.8 \leq r \leq 1$  时, 两个变量之间为高度相关, 所以地面沉降量和地下水开采量的相关性十分显著, 可以建立线性回归方程。

同理利用 Pearson 相关系数的计算公式可以得到地面沉降量和地下水水位的相关系数为  $r = 0.98596$ , 根据以上分析过程, 地面沉降量和地下水水位之间也呈高度相关, 同样适宜建立线性回归方程。

2 线性回归分析

2.1 地面沉降量与地下水开采量的线性回归分析

回归分析是指从一些反映某些变量之间关系的观测值出发, 分析变量间的相关程度, 并建立回归模型来拟合变量间的关系, 从而提高对变量之间关系的认识。根据自变量的数量多少, 回归分析可以分为一元回归分析和多元回归分析; 根据自变量和因变量之间的关系类型, 可以分为线性回归分析和非线性回归分析。本文所选择的是一元线性回归模型来进行研究, 回归分析的主要任务就是根据若干个观测值找出两个变量  $x, y$  之间的线性关系, 从而得到两者间的线性回归方程  $\hat{y} = ax + b$ , 其中  $\hat{y}$  是  $y$  的估计值。求回归方程的过程, 实际上是用回归直线对各个观测值进行拟合的过程, 最常用的方法是最小二乘法。由于最小二乘法的计算过程非常繁琐, 通常借助 Excel 软件的数据分析工具来进行回归分析。

将表 1 中的累积开采量和累积沉降量的一系列数据输入到 Excel 中, 利用回归分析工具得到的分析结果如图 3 所示, 通过分析结果中的数据可以得到线性回归方程的系数, 即斜率  $a = 8.121$ , 截距  $b = 4.156$ 。

|                                                                               |          |          |          |         |                |          |          |           |  |
|-------------------------------------------------------------------------------|----------|----------|----------|---------|----------------|----------|----------|-----------|--|
| SUMMARY OUTPUT                                                                |          |          |          |         |                |          |          |           |  |
| 回归统计                                                                          |          |          |          |         |                |          |          |           |  |
| Multiple R                                                                    | 0.998797 |          |          |         |                |          |          |           |  |
| R Square                                                                      | 0.997595 |          |          |         |                |          |          |           |  |
| Adjusted R Square                                                             | 0.996793 |          |          |         |                |          |          |           |  |
| 标准误差                                                                          | 3.352928 |          |          |         |                |          |          |           |  |
| 观测值                                                                           | 5        |          |          |         |                |          |          |           |  |
| 方差分析                                                                          |          |          |          |         |                |          |          |           |  |
|                                                                               | df       | SS       | MS       | F       | Significance F |          |          |           |  |
| 回归分析                                                                          | 1        | 13988.42 | 13988.42 | 1244.29 | 5.00997E-05    |          |          |           |  |
| 残差                                                                            | 3        | 33.72637 | 11.24212 |         |                |          |          |           |  |
| 总计                                                                            | 4        | 14022.14 |          |         |                |          |          |           |  |
| Coefficients: 标准误差 t Stat P-value Lower 95% Upper 95% Lower 95.0% Upper 95.0% |          |          |          |         |                |          |          |           |  |
| Intercept                                                                     | 4.156173 | 3.571759 | 1.163621 | 0.32871 | -7.210757786   | 15.5231  | -7.21076 | 15.523104 |  |
| X Variable 1                                                                  | 8.12155  | 0.230239 | 35.27444 | 5E-05   | 7.388827066    | 8.854273 | 7.388827 | 8.8542733 |  |

图 3 Excel 回归分析输出结果

Fig. 3 Output results of excel regression analysis

从而建立累积地面沉降量和累积地下水开采量的线性回归方程  $s = 8.121 Q + 4.156$

式中:  $s$ ——累积地面沉降量;

$Q$ ——累积地下水开采量。

图 4 为累积地面沉降量和累积地下水开采量的线性回归直线, 通过图表法得到的线性回归方程与利用回归分析工具得到的方程是一致的。从图 4 中还能够看出回归方程的判定系数  $R^2 = 0.997$ , 这表明自变量对因变量的解释程度较高, 即总体的地面沉降有很大比例是由地下水开采引起的, 判定系数可以用来

衡量回归直线的整体拟合效果。

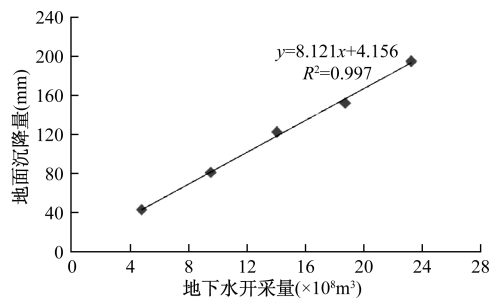


图 4 累积沉降量和累积开采量的线性回归直线

Fig. 4 Linear regression line between accumulative subsidence quantity and accumulative groundwater mining

2.2 地面沉降量与地下水水位的线性回归分析

利用 2004 ~ 2008 年天竺地区中深部地层沉降量和地下水水位的观测数据同样可以得到地面沉降量和地下水位的线性回归方程(图 5):

$s = 6.919 H - 3.257$

式中: $s$ ——中深部地层沉降量;  
 $H$ ——地下水水位。

| SUMMARY OUTPUT                                                       |            |          |          |          |                |          |          |            |  |
|----------------------------------------------------------------------|------------|----------|----------|----------|----------------|----------|----------|------------|--|
| 回归统计                                                                 |            |          |          |          |                |          |          |            |  |
| Multiple R                                                           | 0.98596053 |          |          |          |                |          |          |            |  |
| R Square                                                             | 0.97211817 |          |          |          |                |          |          |            |  |
| Adjusted R Square                                                    | 0.96282423 |          |          |          |                |          |          |            |  |
| 标准误差                                                                 | 6.95197162 |          |          |          |                |          |          |            |  |
| 观测值                                                                  | 5          |          |          |          |                |          |          |            |  |
| 方差分析                                                                 |            |          |          |          |                |          |          |            |  |
|                                                                      | df         | SS       | MS       | F        | Significance F |          |          |            |  |
| 回归分析                                                                 | 1          | 5055.162 | 5055.162 | 104.597  | 0.001993       |          |          |            |  |
| 残差                                                                   | 3          | 144.9897 | 48.32991 |          |                |          |          |            |  |
| 总计                                                                   | 4          | 5200.151 |          |          |                |          |          |            |  |
| Coefficients: 标准误差 t Stat P-value Lower 95%Upper 95%下限 95.0%上限 95.0% |            |          |          |          |                |          |          |            |  |
| Intercept                                                            | -3.257453  | 5.672765 | -0.57423 | 0.606043 | -21.3107       | 14.79582 | -21.3107 | 14.7958176 |  |
| X Variabl                                                            | 6.91974889 | 0.676598 | 10.22727 | 0.001993 | 4.766512       | 9.072866 | 4.766512 | 9.07286617 |  |

图 5 Excel 回归分析输出结果

Fig. 5 Output results of excel regression analysis

图 6 为地面沉降量和地下水水位的线性回归直线,从图中可以得到该回归方程的判定系数  $R^2 = 0.972$ ,表明回归直线的拟合效果较好。

上述两个方程是从面和点两个不同层面上对地面沉降和地下水开采的相关性进行分析,它们的判定系数分别为 0.997 和 0.972,说明回归直线的整体拟合效果都很好。所以,可以分别利用地下水开采量和地下水水位对地面沉降量进行预测。

3 线性回归方程的验证和预测

3.1 地面沉降量与地下水开采量回归方程的验证和预测

根据回归方程  $s = 8.121Q + 4.156$ ,将 2005 ~ 2011 年的地下水开采量代入其中,可以得到对应的

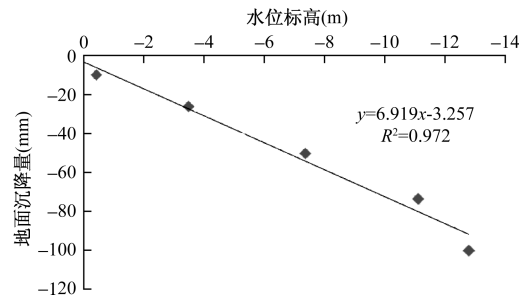


图 6 地面沉降量和地下水位的线性回归直线

Fig. 6 Linear regression line between subsidence quantity and groundwater level

地面沉降量预测值(表 2、图 7)。2010 年和 2011 年是新输入的年份,所以不如前五年的拟合程度好,但总体趋势是吻合的。为了研究预测值的可靠程度,可以引入统计学中的置信区间来进行检验,即在一定置信度下所要预测的变量的可能取值范围。这里取置信度为 95%,根据标准正态分布表可以得到置信度 95%的置信区间为  $(s - 1.96\delta, s + 1.96\delta)$ , $\delta$  是由回归分析得到的标准误差,用来表示变量  $s$  偏离回归直线的误差。

表 2 地面沉降量的真实值和预测值

Table2 The measured and predicted values of subsidence quantity

| 时间      | 2005  | 2006  | 2007   | 2008   | 2009   | 2010   | 2011   |
|---------|-------|-------|--------|--------|--------|--------|--------|
| 真实值(mm) | 42.99 | 81.06 | 121.88 | 152.07 | 194.53 | 247.58 | 285.67 |
| 预测值(mm) | 43.22 | 81.79 | 118.25 | 156.26 | 192.97 | 227.40 | 264.03 |

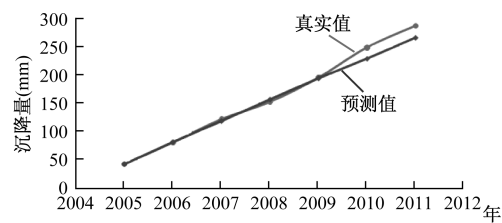


图 7 地面沉降量的真实值和预测值拟合图

Fig. 7 Fitting curves of the measured and predicted values of subsidence quantity

在线性回归直线  $s = 8.121Q + 4.156$  的上下作两条平行直线(图 8):

$s' = 8.121Q + 4.156 - 1.96\delta$  和  $s'' = 8.121Q + 4.156 + 1.96\delta$

从图 8 中可以看出,2010 和 2011 年累积沉降量真实值都落在置信度为 95% 的置信范围内,因此通过线性回归方程得到的沉降量预测值可靠程度较高。

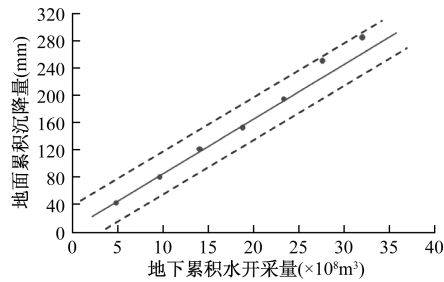


图 8 回归方程的预测值可靠性检验图

Fig. 8 The reliability of predicted values based on regression equations test chart

假定 2012 ~ 2014 年地下水开采量将会维持 2011 年的现状开采水平,那么可以根据回归方程得到预测的累计沉降量,从而算出年沉降量的预测值(表 3)。

表 3 2012 ~ 2014 地面年沉降量的预测值

Table3 Predicted values of subsidence quantity between 2012 and 2014

| 时间      | 2011  | 2012  | 2013  | 2014  |
|---------|-------|-------|-------|-------|
| 预测值(mm) | 38.09 | 38.25 | 38.17 | 40.36 |

从表 3 可以看出,在现状开采条件下,2012 ~ 2014 年的年沉降量在 2011 年的基础上略有增加或基本持平,地面沉降持续发展,累积沉降量逐年增加。

3.2 地面沉降量与地下水位回归方程的验证和预测

将 2004 ~ 2010 年天竺地区中深层地下水水位观测值代入方程,可以得到中深部地层的沉降量预测值(表 4)。沉降量的预测值和真实值的拟合程度较好(图 9)。

表 4 2004 ~ 2010 年中深部地层沉降量的真实值和预测值

Table4 Predicted values of subsidence quantity in middle-deep strata

| 时间      | 2004 | 2005  | 2006  | 2007  | 2008  | 2009   | 2010   |
|---------|------|-------|-------|-------|-------|--------|--------|
| 真实值(mm) | 9.87 | 25.93 | 49.97 | 73.35 | 99.81 | 122.90 | 141.97 |
| 预测值(mm) | 6.14 | 27.33 | 54.03 | 79.93 | 91.49 | 96.63  | 112.82 |

将 3.1 中预测值可靠程度的验证过程同理应用到该方程预测值可靠性的分析过程中,可以得出根据地下水位预测的地面沉降量同样是十分可靠的。进行预测时可以将预计的地下水位标高值输入到方程中,从而预测同层段内的地面沉降量。

4 结论

(1)通过对北京天竺地区地下水开采和地面沉降的关系进行分析,可以看出两者之间具有一定的线

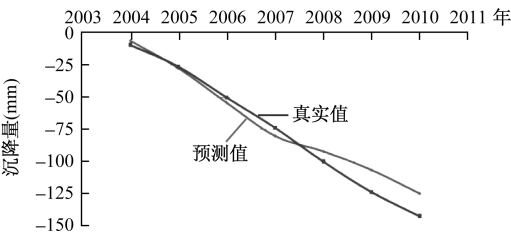


图 9 地面沉降量的真实值和预测值拟合图

Fig. 9 Fitting chart of the measured and predicted values of subsidence quantity

性关系,通过地下水开采量或地下水水位,可以对地面沉降量进行预测。

(2)建立线性回归方程操作简单,工作量小,但是预测误差是不可避免的。由于资料有限,本文建立的预测模型没有考虑地面沉降的其它影响因素如地层岩性、地面荷载等。在以后建立的线性回归模型中,如果可以将多种影响因素一起考虑,那么预测精度将会有很大提高。

(3)由以上分析,北京天竺地区地面沉降的变化趋势与地下水水位变化趋势基本一致,地面沉降量与地下水开采量具有高度的相关性,可以推断出北京平原地面沉降产生的主要原因是地下水的大量开采。

(4)通过模型预测可知,在维持地下水现状开采的条件下,不考虑其它影响因素,北京平原区的地面沉降将会持续发展,整体的地面沉降量仍会继续增加。

参考文献:

[ 1 ] 赵显富,郭淳. 灰色预测模型及其在沉降预测中的应用 [ J ]. 铁路航测, 2002, 28 ( 3 ) : 30 - 32.  
ZHAO Xianfu, GUO Chun. Gary forecast model and it's application in settlement observation [ J ]. Railway Air Survey, 2002, 28 ( 3 ) : 30 - 32.  
[ 2 ] 兰孝奇,严红萍,刘精攀. 灰色系统预测模型在沉降监测中的应用 [ J ]. 现代测绘, 2006, ( 1 ) : 24 - 26.  
LAN Xiaoqi, YAN Hongping, LIU Jingpan. The gary model application for settlement observation [ J ]. Modern Surveying and Mapping, 2006, ( 1 ) : 24 - 26.  
[ 3 ] 郑良飞,折学森. 地面沉降的 GM ( 2, 1 ) 模型预测研究 [ J ]. 中国地质灾害与防治学报, 2007, 18 ( 4 ) : 66 - 69.  
ZHENG Liangfei, ZHE Xuesen. A study to the ground settlement forecast with GM ( 2, 1 ) model [ J ]. The Chinese Journal of Geological Hazard and Control, 2007, 18 ( 4 ) : 66 - 69.  
[ 4 ] 周复旦,赵长胜,高卫东. BP 神经网络模型在采水地面

- 沉降中的应用研究[J]. 测绘科学, 2011, 36(6): 233 – 234.
- ZHOU Fudan, ZHAO Changsheng, GAO Weidong. Application of BP neural network model to ground subsidence of mining water area[J]. Science of Surveying and Mapping, 2011, 36(6): 233 – 234.
- [ 5 ] 金爱善. 采用神经网络模型对天津滨海新区地面沉降预测的研究[J]. 现代地质, 2000, 14(4): 475 – 478.
- JIN Aishan. Prediction of land subsidence in the new coastal region of Tianjin with artificial neural network model[J]. Geoscience, 2000, 14(4): 475 – 478.
- [ 6 ] 宋小军. 基于 MODFLOW 对天津宁河县的地面沉降预测研究[J]. 矿产勘查, 2010, 1(6): 564 – 568.
- SONG Xiaojun. Research on prediction of land subsidence based on MODFLOW in Tianjin Ninghe county [J]. Mineral Exploration, 2010, 1(6): 564 – 568.
- [ 7 ] 马锋, 杨发俊, 陈润桥, 等. 天津市地下水开采对地面沉降影响的多元回归分析[J]. 中国地质灾害与防治学报, 2008, 19(2): 63 – 66.
- MA Feng, YANG Fajun, CHEN Runqiao, et al. Effect of groundwater exploitation on land subsidence in Tianjin using multiple regression analysis method [J]. The Chinese Journal of Geological Hazard and Control, 2008, 19(2): 63 – 66.
- [ 8 ] 谭荣初. 吴江市地面沉降与开采地下水关系的研究[J]. 水资源保护, 2002(2): 48 – 50.
- TAN Rongchu. Relationship between ground settlement and groundwater exploitation in Wujiang City[J]. Water Resources Protection, 2002 (2): 48 – 50.
- [ 9 ] 杨艳, 贾三满, 王海刚. 北京平原区地面沉降现状及发展趋势分析[J]. 上海地质, 2010, 31(4): 23 – 28.
- YANG Yan, JIA Sanman, WANG Haigang. The status and development of land subsidence in Beijing Plain[J]. Shanghai Geology, 2010, 31(4): 23 – 28.
- [ 10 ] 北京市水务局. 北京市水资源公报[R]. 2005 – 2009.
- Beijing Water Authority. Beijing Water Resources Bulletin[R]. 2005 – 2009.

## Application of linear regression model for land subsidence prediction in Beijing plain

FAN Shan-shan<sup>1</sup>, GUO Hai-peng<sup>2</sup>, ZHU Ju-yan<sup>1</sup>, LI Wen-peng<sup>3</sup>

(1. Water Resources and Environment School, China University of Geosciences(Beijing), Beijing 100083, China;

2. China Institute of Geo-Environmental Monitoring, Beijing 100081, China;

3. Center for Hydrogeology and Environmental Geology Survey, CGS, Baoding 071051, China)

**Abstract:** The land subsidence in Beijing Plain tends to exacerbate, and has become the main geology disaster of Beijing plain. Based on measured data of groundwater exploitation, groundwater level and land subsidence, this paper establishes linear regression equations on amount of groundwater mining-subsidence quantity and groundwater level-subsidence quantity by Excel. The subsidence quantity is forecast based on the regression equations, and the reliability of predicted values are verified. The established linear regression models are used to forecast land subsidence, which could provide basis for land subsidence control.

**Key words:** linear regression; land subsidence; groundwater mining; prediction