成果与方法

滑动线性回归模型预测路堤沉降初探

刘红英[1],李朋军[2],赖桂新[2]

(1. 德州市测量队,山东 德州　253000;2. 淮海工学院空间信息科学系,江苏 连云港　222001)

摘要:在路堤沉降预测中,线性模型是应用最为广泛的模型之一。但是随着观测信息的丰富,常规的线性模型的不足凸现,影响了模型的精度,为了克服常规模型的缺点提出了滑动线性回归模型。滑动线性回归模型要求参与建模的数据个数保持不变,新观测的数据取代旧有的数据,从而获得了新的模型参数,通过实例验证,得到了理想的结果。

关键词:路堤沉降;线性回归;滑动模型;预测

中图分类号:U412.22;TU433　　**文献标识码:**A

0　引言

随着我国道路建设的飞速发展,高填方及特殊土等路堤大量出现。在建设过程中,为确保铁路公路建设质量、指导施工和安全运营,相应的铁路和公路路基规范都要求对路堤进行沉降观测。对于沉降数据的处理方法有很多,如统计回归、神经网络、时间序列和小波分析等。通常用建模的精度和预测结果的准确度来衡量数据处理方法的优劣,并且从模型中找出路堤沉降规律,及时发现问题、处理问题。

在生产实践中曾运用滑动线性回归模型解决了不少变形观测问题,该文试将该模型运用到路堤的沉降预测中来,并用实例验证了其效果。

1　滑动线性回归模型的提出

线性回归模型是应用最为广泛的模型之一,也是其他模型应用的基础。在现实世界中,许多变量之间具有线性或近似线性的依赖关系;有些变量之间是非线性的,但经过适当的变换后,新产生的变量之间就具有了近似的线性关系;另外线性关系比较容易处理,也容易达到比较理想的结果。因此在路堤预测中应该首选线性回归模型。

1.1　线性的回归模型

通常把线性回归模型分为多元线性回归和可化为线性回归模型的非线性回归 2 种,而后者主要有幂函数模型、指数函数模型、对数函数模型、双曲函数模型、多项式函数模型,简称为常规模型。文中列出 3 种常规模型[1]。

多元线性回归模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \quad (1)$$

对数函数模型:

$$y = \beta_0 + \beta_1 \ln x \quad (2)$$

多项式函数模型:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m \quad (3)$$

式中:y 表示自变量;x 表示因变量;β 表示回归系数。

1.2　滑动线性回归模型

滑动线性回归模型是为得到更好的建模精度和预测效果而提出的。当信息比较丰富时,如果依然使用常规模型,由于观测误差的积累和数据的异常跳跃,模型受外界影响的程度增加,而用于描述建模精度的单位权方差 σ^2 与残差平方和呈正相关关系,致使建模结果不够理想,同时其模型稳定性不能得到保障,与观测值的走势也不能够很好地吻合,导致预测误差偏大。另外,常规模型不能很好地反映外界因素对沉降观测值的影响和沉降本身变化规律。因此必须设法克服常规模型的缺点。而滑动线性回

* **收稿日期:**2007-12-26;**修订日期:**2008-05-28;**编辑:**陶卫卫

作者简介:刘红英(1975-),女,山东齐河人,工程师,主要从事工程测量工作。

归模型就是应用新陈代谢法则,固定参与模型求解个数的观测值(简称等维),即考虑增加 1 个新数据,剔除 1 个离预测值最远的旧数据,从而获得了模型的新参数,预测下一个数据,如此不断进行建模和预测。滑动模型能充分顾及到事物发生发展的客观规律,因此滑动模型预测精度好,准确度高。另外,参与建模的数据个数也较常规模型少,简化了计算。

2　实例研究

2.1　工程概况

正在建设中的上(海)瑞(丽)高速公路湖南境内(湘)潭邵(阳)段主要为丘陵地貌,高填深挖路基、特殊土(如软土、膨胀土)路基较多,为确保路基施工安全、施工质量和研究沉降变化规律,特别是为防止桥台跳车,应对路桥过渡段的桥背路堤沉降严格控制,对路堤沉降进行了施工监测。实测数据见表1,其中时间的单位是 10 天,以下用“10 d”表示[2]。

表 1　DK39 +098 桥台路堤沉降实测值

时间(10d)	实测(mm)	时间(10d)	实测(mm)	时间(10d)	实测(mm)
1	6.7	10	53.8	19	98.0
2	13.3	11	66.7	20	99.0
3	25.8	12	69.7	21	102.0
4	30.8	13	73.8	22	104.0
5	35.0	14	81.8	23	104.6
6	39.7	15	85.0	24	107.0
7	41.8	16	88.2	25	107.0
8	45.6	17	91.4	26	107.9
9	48.8	18	95.0	27	108.0

2.2　用(可化为)线性的回归模型预测其沉降

通常情况下,在分析采用何种线性模型时,首先要分析散点图,根据表 1 数据绘制出散点图(图 1)。

从图 1 看出,该段路堤沉降呈现为一种非线性关系,另外当 $t=3$,$t=11$ 和 $t=14$ 时,沉降观测值有明显的跃变。因此可以初步认为对数模型(2)和多项式函数模型(3)适合该问题。

2.2.1　常规模型

为了比较建模的效果,首先按照常规模型的算法,获得相关的模型和预测值(表 2,3)。$N<11$ 时,属于贫信息,不宜用滑动线性回归模型。

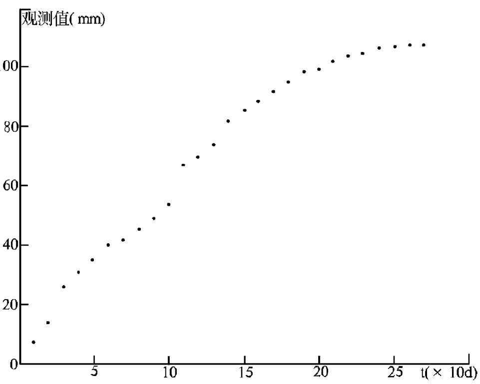

图 1　桥台路堤沉降实测散点图

表 2　常规对数模型

时间(10d)	实测(mm)	预测值(mm)	误差值(mm)	误差率(%)	σ^2
12	69.7	57.3	−12.4	−17.7	23.0
13	73.8	61.6	−12.2	−16.5	33.1
14	81.8	65.7	−16.1	−19.7	41.2
15	85.0	70.2	−14.8	−17.4	55.9
16	88.2	74.3	−13.9	−15.8	66.0
17	91.4	78.1	−13.3	−14.6	73.2
18	95.0	81.6	−13.4	−14.1	78.6
19	98.0	85.0	−13.0	−13.2	83.5
20	99.0	88.3	−10.7	−10.8	87.3
21	102.0	91.1	−10.9	−10.7	88.2
22	104.0	93.9	−10.1	−9.7	89.1
23	104.6	96.5	−8.1	−7.7	89.2
24	107.0	98.8	−8.2	−7.7	87.8
25	107.0	101.0	−6.0	−5.6	86.6
26	107.9	103.0	−4.9	−4.5	84.2
27	108.0	104.8	−3.2	−2.9	81.6

表 3　常规多项式模型

时间(10d)	实测(mm)	预测值(mm)	误差值(mm)	误差率(%)	σ^2
12	69.7	77.5	7.8	11.2	2.7
13	73.8	83.5	9.7	13.2	3.9
14	81.8	86.7	4.9	6.0	5.9
15	85.0	93.6	8.6	10.1	6.0
16	88.2	96.7	8.5	9.6	7.4
17	91.4	98.3	6.9	7.6	8.7
18	95.0	99.6	4.6	4.9	9.3
19	98.0	101.4	3.4	3.4	9.2
20	99.0	103.0	4.0	4.1	9.0
21	102.0	103.2	1.2	1.2	8.9
22	104.0	104.3	0.3	0.3	8.4
23	104.6	105.4	0.8	0.7	8.0
24	107.0	105.5	−1.5	−1.4	7.7
25	107.0	106.5	−0.5	−0.5	7.4
26	107.9	106.4	−1.5	−1.4	7.0
27	108.0	106.4	−1.6	−1.4	6.8

从表 2 和表 3 中可以看出,无论是对数模型还是多项式模型,效果都不理想,尤其是常规对数模型预测值平均误差率竟高达 11.8%,而常规多项式模

型预测值平均误差率也将近5%。同时从反映建模精度的σ^2来看,常规对数模型最大达89.2,最小的也有23.0,平均为71.8;常规多项式模型的$\overline{\sigma^2}$为7.3。很明显,随着观测值的增多,观测误差凸现,模型受外界影响的程度增加。加之受异常跳跃数据的影响,致使预测精度降低,误差率居高不下。

2.2.2　滑动线性回归模型

对于滑动线性回归模型,首先要探讨用于建模观测值的个数N,从图1中可知,当$N=11$时,模型基本稳定,规律比较明显;另外当$N>11$时,对σ^2的影响不大。多项式阶数的确定,依据非线性模型平差原理[3]中的统计量F进行检验:

$$F = \frac{S_{残m-1} - S_{残m}}{S_{残m}/(n-m)} \tag{4}$$

式中,S代表残差平方和;n代表总观测数;m代表回归参数的个数。

根据表1中前11个数据分别得到了二阶、三阶和四阶的残差平方和:112.5581,24.2791,12.0251。计算得,接受原假设,即认为三阶与四阶没有显著区别,故取三阶。同时用三阶的多项式分别计算$N=8\sim15$时σ的值(表4),可以看出$N=11$,σ变化已经很小。

表4　N=8~15对应σ的值

N	8	9	10	11
σ	1.1062	1.6307	1.6528	1.8950
N	12	13	14	15
σ	1.8900	1.8395	1.8215	1.8901

因此,选取表1中11个数据参与建模,分别采用对数模型和多项式模型进行计算,结果如表5和表6。滑动对数模型预测值平均误差率为5.4%,滑动多项式模型预测值平均误差率为2.7%;与其对应的常规模型相比平均预测精度分别提高了6.4和2.1个百分点,可见,滑动模型的效益十分明显。从建模精度σ^2来看,滑动对数模型$\overline{\sigma^2}=11.5$,滑动多项式$\overline{\sigma^2}=2.3$,均小于常规模型的$\overline{\sigma^2}$值。如果仅从建模精度来考虑,常规对数模型误差太大,不适合该问题,但使用了滑动对数模型,随着可选择的数据增多,它也能取得较好的效果。换句话说,滑动模型的建模精度有了大幅度的提高。为直观、方便比较,将表1~3和表5~6绘制成图(图2)。

表5　滑动对数模型

时间(10d)	实测(mm)	预测值(mm)	误差值(mm)	误差率(%)	σ^2
12	69.7	57.3	-12.4	-17.7	23.0
13	73.8	64.0	-9.8	-13.3	24
14	81.8	69.3	-12.5	-15.3	30.0
15	85.0	76.6	-8.4	-9.9	32.0
16	88.2	83.3	-4.9	-5.6	26.2
17	91.4	89.4	-2.0	-2.2	17.7
18	95.0	94.9	-0.1	-0.1	8.3
19	98.0	99.4	1.4	1.4	4.6
20	99.0	102.9	3.9	3.9	3.9
21	102.0	104.6	2.6	2.6	4.7
22	104.0	105.7	1.7	1.6	1.7
23	104.6	107.8	3.2	3.1	1.9
24	107.0	108.8	1.8	1.7	2.1
25	107.0	109.6	2.6	2.5	0.6
26	107.9	110.6	2.7	2.5	1.1
27	108.0	111.1	3.1	2.9	1.3

表6　滑动多项式模型

时间(10d)	实测(mm)	预测值(mm)	误差值(mm)	误差率(%)	σ^2
12	69.7	77.5	7.8	11.2	2.7
13	73.8	85.0	11.2	15.2	3.8
14	81.8	83.8	2.0	2.4	4.8
15	85.0	88.8	3.8	4.5	4.4
16	88.2	88.8	0.6	0.7	3.9
17	91.4	88.0	-3.4	-3.7	3.1
18	95.0	90.3	-4.7	-5.0	3.4
19	98.0	95.0	-3.0	-3.0	3.6
20	99.0	100.9	1.9	1.9	2.4
21	102.0	102.3	0.3	0.3	2.4
22	104.0	102.5	-1.5	-1.5	1.1
23	104.6	106.3	1.7	1.6	0.6
24	107.0	106.9	-0.1	-0.1	0.7
25	107.0	107.6	0.6	0.6	0.2
26	107.9	107.7	-0.2	-0.2	0.2
27	108.0	108.3	0.3	0.3	0.2

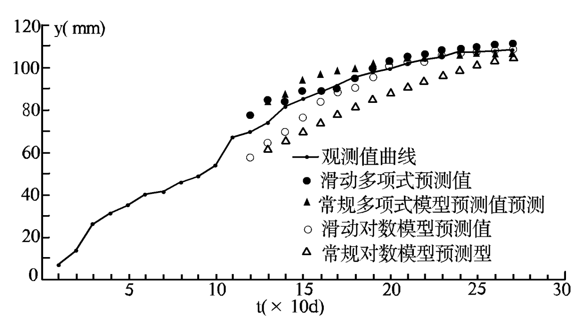

图2　4种模型预测值与预测值走势对比图

综上所述,得出了滑动模型能充分顾及到事物发生发展的客观规律,滑动模型预测精度好,准确度

高。因此决定选取三阶滑动多项式模型为主要模型来研究潭邵段高速公路路堤的沉降规律。

3　结语

线性滑动模型固定了参与模型求解的观测值个数,顾及到路堤滑动变化的客观事实,能有效地预测在一个时间段内路堤沉降量,并且计算量相对常规模型小,建议在今后的生产中推广使用该模型。

参考文献:

[1]　刘大杰,陶本藻. 实用测量数据处理方法[M]. 北京:测绘出版社,2003.

[2]　肖武权,冷伍明. 陆地沉降的时间序列动态预测方法[J]. 铁道工程学报,2004,12(4):52-56.

[3]　王新洲,陶本藻. 高等测量平差[M]. 北京:测绘出版社,2006,42-52.

Brief Predication of Embankment Settlement Based on Dynamic Linear Regression Model

LIU Hong-ying, LI Peng-jun, LAI Gui-xin

(1. Dezhou Measurement Brigade, Shandong Dezhou 253000, China; 2. Huihai Technical College, Jiangsu Lianyungang 222001, China)

Abstract: Linear regression model is one of the most widely used method in the predication of embankment settlement. Accompanying with rich investigation information, the disadvantages of traditional linear model stand out extremely, so the accuracy of prediction is reduced. To overcome deficiencies in the traditional model, regression linear dynamic model is put forward. This model requests that the number of data used to make model is kept unchanged, new data replaced the old data, so the new parameter is gained. Ideal result is gotten through the instance.

Key words: Embankment settlement; linear regression; dynamic model; predication