

# 基于数据驱动模型的深圳填海区地面沉降预测分析

王艳<sup>1</sup>, 邓英尔<sup>2</sup>, 高延超<sup>3</sup>

(1. 深圳市勘察研究院有限公司, 深圳 518000;

2. 成都理工大学地质灾害防治与地质环境保护国家重点实验室, 成都 610059;

3. 中国地质调查局成都地质调查中心, 成都 610081)

**摘要:** 为解决土地资源问题, 深圳市的迅猛发展伴随着填海造陆工程。填海区海相淤泥层属于高饱和软土, 高压缩性, 上部荷载作用及回填土的自重固结作用易导致地面沉降地质灾害的发生, 对填海区地面沉降问题的预测研究也变得至关重要。以空港新城区域 2019 年 5 月至 2022 年 4 月的监测数据为例, 选择多种数学模型对地面沉降进行预测, 包括 NAR 神经网络模型、GM(1,1)模型、多项式回归模型和 ARIMA 模型。预测表明: 4 种模型均能较好地反应出累计沉降量的未来发展趋势, 可为后期地面沉降防治方案及地基处理等问题提供基础依据, 有效降低地面沉降地质灾害的危险性。综合其预测结果的稳定性及精确度等因素, 4 种模型的优劣可按 ARIMA 模型、NAR 神经网络模型、多项式回归模型、GM(1,1)模型顺次排序。

**关键词:** 地面沉降; 数据驱动模型; 预测; 填海区

**中图分类号:** P642.26; P694; TU434 **文献标识码:** A

## PREDICTIVE ANALYSIS OF LAND SUBSIDENCE IN SHENZHEN RECLAMATION AREA BASED ON DATA-DRIVEN MODELS

WANG Yan<sup>1</sup>, DENG Ying-er<sup>2</sup>, GAO Yan-chao<sup>3</sup>

(1. Shenzhen Survey and Research Institute Co., Ltd., Guangdong, Shenzhen 518000, China;

2. State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China; 3. Chengdu Center, China Geological Survey, Chengdu 610081, China)

**Abstract:** To solve the problem of land resources, the rapid development of Shenzhen was accompanied by reclamation projects. The marine silt layer in the reclamation area is a highly saturated soft soil with high compressibility. The effect of upper loads and the self-weight consolidation of backfill can easily lead to the occurrence of geological hazards of ground settlement, and the prediction study of ground settlement problems in reclamation areas has become crucial. Taking the monitoring data from May 2019 to April 2022 in the Airport New City area as an example, various mathematical models were selected to predict the ground subsidence, including the NAR neural network model, the GM (1,1) model, the polynomial regression model, and the ARIMA model. The prediction shows that all four models can better reflect the future development trend of the cumulative settlement, which can provide a basic basis for the prevention and control program of ground settlement and foundation treatment in the later stage, and effectively reduce the risk of geological disasters of ground settlement. Combining the stability and accuracy of their prediction results and

other factors, the advantages and disadvantages of the four models can be ranked in order of the ARIMA model, NAR neural network model, polynomial regression model, and GM (1, 1) model.

**Key words:** land subsidence; data-driven model; prediction; reclamation

## 1 引言

随着改革开放的进程,深圳市开始迅猛发展,自 20 世纪 80 年代起,为了解决土地资源紧张的问题,深圳市开始第一次填海工程。1978 年~1985 年深圳市填海造陆面积约 9.14 km<sup>2</sup>,此后 20 a,填海造陆面积以倍数的增长速度不断增加,在 1996 年~2007 年深圳填海造陆工程达到峰值时段<sup>[1]</sup>。随后,填海造陆面积波动式下降,截至 2020 年深圳市填海面积高达 133.46 km<sup>2</sup>。人们逐步认识到大规模地填海造陆会导致自然海岸线减少、海岸自然景观破坏,沿海区域规划利用结构发生变化,海水净化能力减弱、海洋生态系统失衡等问题<sup>[2]</sup>,如大铲湾-深圳河河口岸段,仅有以红树林生态保护区为主的原生态岸线外,其余岸线均为填海造陆形成的人工岸线<sup>[3]</sup>。填海工程开始变得理性化、全面化,深圳市总体规划(2010~2020)中明确指出:原则上不得进行大规模填海造地活动。填海区的地面沉降问题已引起社会和政府的高度重视,政府成立专项小组,建立重点区域地面沉降的长期监测系统,完善对地面沉降的预测预警机制。

关于地面沉降的预测问题,前人做了诸多工作。针对传统的数学模型,吕传振、孙钟磊等<sup>[4-5]</sup>基于灰色模型开展了地面沉降的预测研究;王琰、范珊珊、黄婧等<sup>[6-8]</sup>基于多元线性回归模型开展了地面沉降的预测研究;朱宝等<sup>[9]</sup>基于 ARIMA 对海岸带地面沉降进行了预测研究;牛全福等<sup>[10]</sup>基于 Kalman 滤波的 ARIMA-NAR 神经网络模型对深基坑沉降进行预测研究。结果表明数据驱动模型能够较好地反应地面沉降的发展趋势,预测数据比较可靠。

然而,为使模型预测能更好地指导工程施工,对潜在地质灾害的发生进行及时预警,选择一种易于实现并且准确的模型至关重要。本文以深圳市空港新城区域为例,基于 ARIMA 模型(Auto Regressive Integrated Moving Average model,差分自回归移动平均模型)<sup>[11]</sup>、NAR 模型(Nonlinear Auto Regressive models,非线性自回归模型)<sup>[12]</sup>、多项式回归模型及灰色理论模型,开展地面沉降的预测研究。探讨一种合适的方法实现深圳填海区乃至其他

软土地区地面沉降的良好预测,准确的预测结果是后期沉降防治方案的基础,亦可为地基处理提供理论依据。

## 2 填海区地质概况及监测数据分析

宝安区及南山区西部作为深圳的主要填海区,上部为第四系全新统(Qh)地层,主要由砂砾、淤泥质砂、黏土、淤泥、淤泥质黏土等组成;下部为第四系上更新统(Qp<sup>3</sup>)地层,主要由砂砾、卵石、粗砂、细砂、黏土、砂质淤泥、有机质淤泥、泥炭等组成。淤泥层属高饱和软土,其具有承载力低、压缩性高、加荷后易变形且不均匀、变形速率大且稳定时间长、渗透性小、触变性及流变性大的特点。回填土属特殊性岩土,具有结构松散、均匀性差、抗剪强度低、透水性强及稳定性差等特点,在施加上部荷载作用加之回填土的自重固结作用易导致地面沉降地质灾害的发生。主要危害对象为地下管网、天然浅基础建(构)筑物、道路及行人等。据空港新城区域 2019 年 5 月至 2022 年 4 月的 35 期监测数据表明,百川路、石围路、滨江大道、重庆路与松福大道交叉口处、会展中心展景路和综合管廊处沉降较为严重<sup>[13]</sup>。典型监测点历史监测数据如图 1。

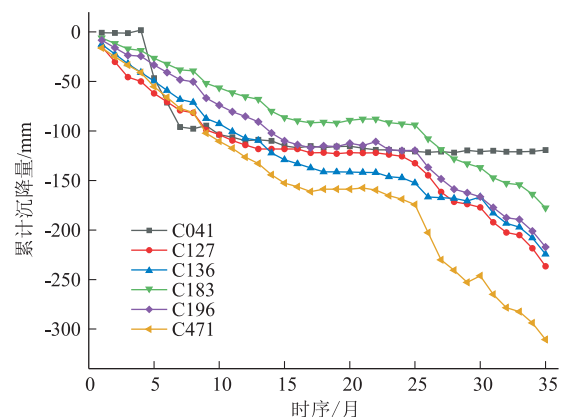


图 1 典型监测点累计沉降量

时间序列的平稳性是预测的关注重点,通过特征根检验法可以判断序列的平稳性<sup>[14]</sup>。特征根检验法是指先拟合确定一组适应性模型,并求取该模型参数组成的特征方程的特征根  $\lambda_i$ ,当  $\lambda_i < 1$  时, $\lambda_i$  满足平稳性条件,认为时间序列为平稳序列,否则即

为非平稳序列。可表示为:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t \quad (1)$$

式中,  $\varepsilon_t$  为冲击值;  $a_j$  表示自回归系数;  $p$  为阶数。

经计算 6 组数据均属非平稳序列, 这意味着, 在预测中, 有必要考虑到其中存在的非平稳特征, 需采用考虑序列非平稳性的数学模型进行预测。并且, 选取的 6 组典型监测点形变速率存在明显差异性拐点(图 2), 以此实现对预测模型适应性的优劣情况进行综合比较。

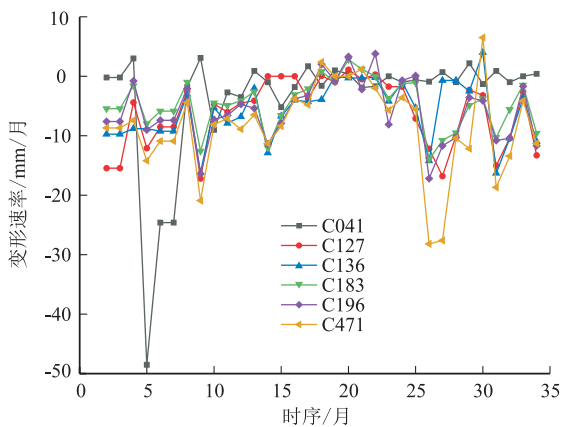


图 2 典型监测点形变速率

## 3 模型构成

### 3.1 ARIMA 模型

随着人们面对越来越多现实生活中的时间序列问题, 经典的差分自回归移动平均模型 (ARIMA) 油然而生。通过对真实时间序列数据运用数学模型进行一种近似的描述, 确定模型的结构及相关参数后, 通过训练集即可实现对未来的预测。模型具有应用范围广且适用于非平稳时间序列预报的特点。ARIMA 模型的基本思想可表述为: 对数据进行平稳性检验, 若属于非平稳序列, 即对原始时间序列进行  $d$  阶差分处理使其平稳化, 然后在确定模型的结构及参数并进行序列预测, 再对预测结果做还原处理即可得到原始序列的预测值。自回归 (Autoregressive, AR) 模型是一种动态模型, 对一组时序  $X$ , AR 模型可表示为:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t \quad (2)$$

式中,  $\varepsilon_t$  为冲击值;  $a_j$  表示自回归系数;  $p$  为阶数。

滑动平均 (Moving Average, MA) 模型一般可写为:

$$X_t = \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t \quad (3)$$

式中,  $b_j$  表示移动平均系数;  $q$  为阶数。

自回归滑动平均 (Autoregressive Moving Average, ARMA) 模型是上述两种模型的合集, 其可写为:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t \quad (4)$$

ARIMA 模型可记作 ARIMA( $p, d, q$ ), 建模思想与灰色模型有类似之处。一个 ARIMA 模型可表达为:

$$a_p(B) \nabla^d X_t = b_q(B) \nabla^d \varepsilon_t \quad (5)$$

式中,  $B$  表示后移算子;  $\varepsilon_t$  为冲击值;  $a_j$  表示自回归系数;  $b_j$  表示移动平均系数;  $p$  为阶数。

蒋金良等<sup>[15]</sup>给出了 ARIMA 模型的 3 种检验方法, 即残差序列白噪声检验; AIC (Akaike information criterion, 最小信息量) 准则、可决系数检验。本文选用 AIC 准则, 模型的复杂程度及其拟合效果可以通过 AIC 准则进行描述。按照下式计算 AIC 值:

$$AIC = -\frac{2L}{N} + \frac{2(p+q+1)}{N} \quad (6)$$

式中,  $L$  为对数似然值。AIC 取值越小, 则代表模型越合理。

### 3.2 NAR 模型

动态的非线性神经网络模型已经成为一种在各领域应用广泛的数学模型, 该模型每一时刻的预测值均是由上一个结果在动态拟合的情况下得到的, 模型具有反馈性和记忆性。主要由输入、隐含、输出和滞后 4 个部分共同组成。在处理时间序列问题, 尤其是非线性复杂问题中凸显出了一定的优势。对于一组时序  $Y$ , 一个 NAR 神经网络模型的可表示为<sup>[16]</sup>:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-d)) \quad (7)$$

式中,  $d$  为延时阶数。

### 3.3 灰色模型

GM(1, 1) 模型是一种广泛应用于各学科领域中的灰色理论模型, 该模型具有对时间序列数据要求低、计算简单的特点, 其首先对预测序列  $X^{(0)}$  作一次累加生成, 得到 1-AGO 序列  $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$ , 其中:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (8)$$

计算得到  $X^{(1)}$  的紧邻生成序列为  $Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)\}$ , 其中:

$$z^{(1)}(k) = 0.5(x^{(1)}(k-1) + x^{(1)}(k)), \quad k = 2, 3, \dots, n \quad (9)$$

计算白化微分方程  $\frac{dx^{(1)}}{dt} + \omega x^{(1)} = u$ ，得到 1-AGO 序列的预测值为：

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{\omega})e^{-\omega k} + \frac{u}{\omega},$$

$$k = 1, 2, \dots, n \quad (10)$$

对 1-AGO sequence 做一次累减还原得到原始数据的预测值为：

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k),$$

$$k = 1, 2, \dots, n \quad (11)$$

### 3.4 多项式回归模型

回归分析模型是一种常用的统计分析方法，在一定规律的情况下，回归模型分为线性和非线性两种。通过建立一个或多个自变量与因变量之间的拟合关系方程是被应用最广泛的拟合算法。多项式回归模型可表示为：

$$y = (\beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n) + \epsilon \quad (12)$$

式中， $\beta_0, \beta_1, \beta_2 \dots \beta_n$  为回归系数； $\epsilon$  为误差项。

## 4 地面沉降预测

### 4.1 预测结果

模型共选用 6 组典型监测点的连续 35 个月监测数据，为便于验证，前 29 期数据作为训练集用于建立模型，后 6 期数据作为模型预测结果的验证集。其中针对多项式回归模型以监测点 C127 与 C183 为例，分别建立三次多项式和四次多项式进行预测（图 3），四次多项式出现过拟合，通过预测结果对比，三次多项式的预测结果也明显更符合实际发展趋势，后文多项式回归模型均选用三次多项式模型。

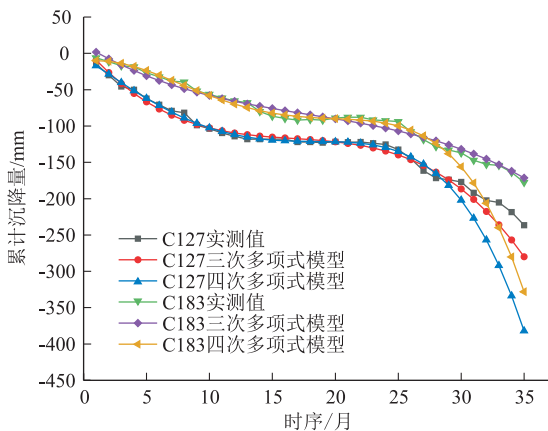


图 3 多项式回归模型预测结果对比

通过 4 种模型展开对 C041、C127、C136、C183、C196 及 C471 监测点的模型建立及预测研究，预测结果如图 4。

### 4.2 预测效果分析

根据模型预测结果对比情况可见表 1~表 2，4 种模型均能较好地反应出序列未来发展趋势，其中 NAR 模型预测结果最好，最小相对误差仅为 0.04%，最大相对误差为 44.05%，相对误差 < 5% 的比例为 44.44%，相对误差 < 10% 的比例为 66.67%。监测点 C041 预测结果的相对误差均 < 2%，误差较大的点主要出现在监测点 C183 的第 32~34 项，误差较大的原因是 NAR 神经网络模型预测结果具有随机性，模型整体稳定性较差。GM(1,1) 模型预测结果次之，最小相对误差为 0.47%，最大相对误差为 45.54%，相对误差 < 5% 的比例为 19.44%，相对误差 < 10% 的比例为 63.89%。多项式回归模型预测结果的最小相对误差为 0.57%，最大相对误差为 38.38%，其中相对误差 < 5% 的比例为 25%，相对误差 < 10% 的比例为 58.33%。ARIMA 模型预测结果的最小相对误差为 0.60%，最大相对误差为 24.07%，其中相对误差 < 5% 的比例为 41.67%，相对误差 < 10% 的比例为 80.56%。

预测结果表明利用 4 种模型对地面沉降的变形预报是可行的。GM(1,1) 模型、多项式回归模型和 ARIMA 模型的预测结果稳定性较好于 NAR 模型，ARIMA 模型的预测精度明显优于 GM(1,1) 模型、多项式回归模型。

## 5 结论

从时间序列预测角度出发，将 NAR 神经网络模型、GM(1,1) 模型、多项式回归模型和 ARIMA 模型用于填海区地面沉降的预测研究，获得如下结论：

(1) 填海区的地面沉降受淤泥层在荷载作用下的压缩变形、填土的自重固结及地下水位变动等多因素影响，根据历史监测数据，其累计变形量时间序列均为非平稳序列。选择合适的数学模型对地面沉降进行预报，可为后期地面沉降防治方案及地基处理等问题提供基础依据，有效降低地面沉降的危险性。

(2) 通过 4 种模型对空港新城区域的 6 组典型监测点累计沉降量的预测研究表明：NAR 神经网络模型、GM(1,1) 模型、多项式回归模型和 ARIMA 模型均能较好地反应出序列未来发展趋势，综合其预测结果的稳定性及精确度等因素，4 种模型的优劣可按 ARIMA 模型、NAR 神经网络模型、多项式回归模型、GM(1,1) 模型顺次排序。

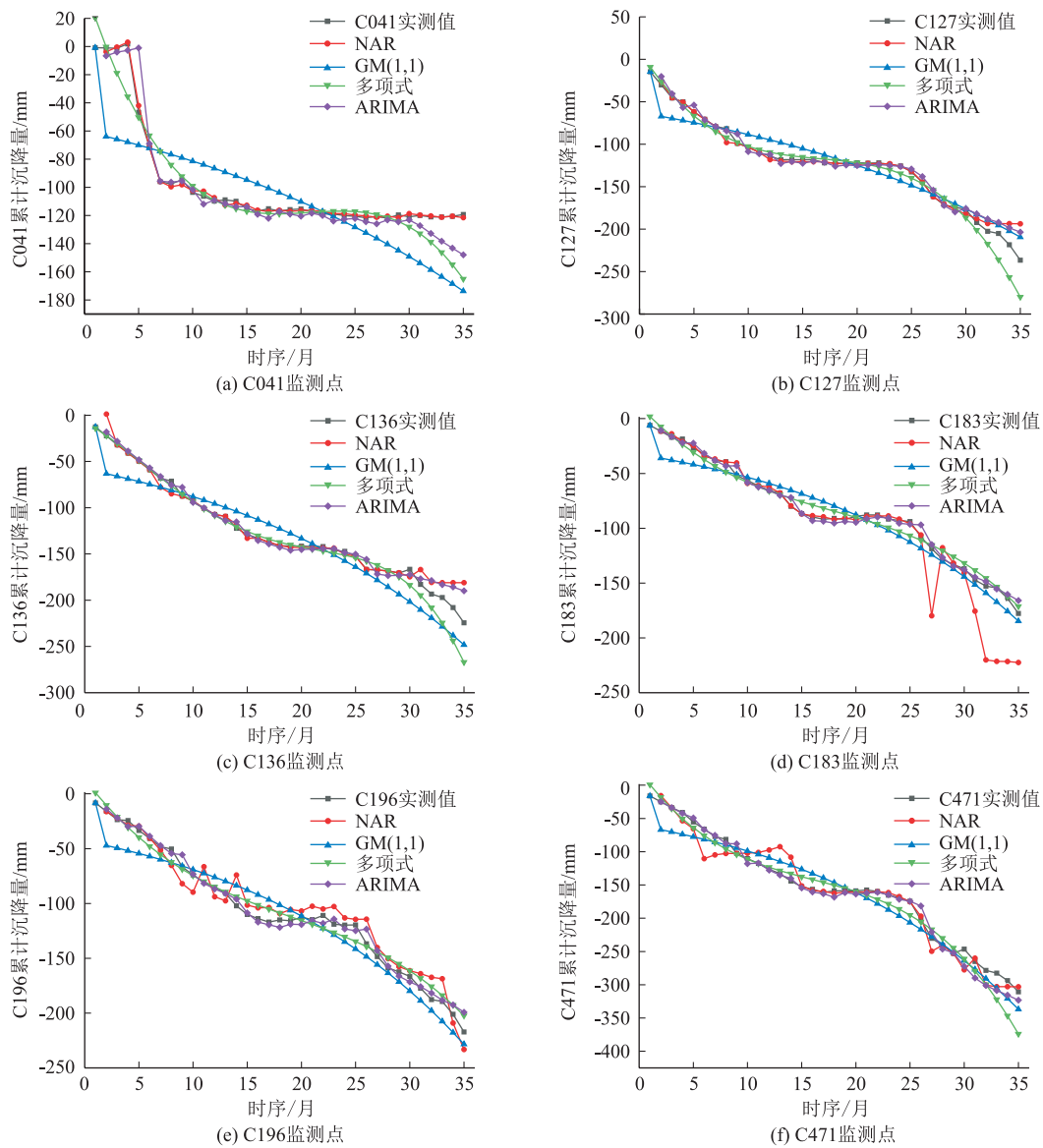


图 4 模型预测结果对比

表 1 几种模型的预测结果

时序/月	C041				C127				C136			
	相对误差/%				相对误差/%				相对误差/%			
	NAR	GM(1,1)	多项式	ARIMA	NAR	GM(1,1)	多项式	ARIMA	NAR	GM(1,1)	多项式	ARIMA
30	1.81	23.43	5.89	1.87	2.91	0.47	5.44	0.71	4.82	21.10	10.28	3.41
31	0.23	28.18	10.73	6.03	2.31	5.03	4.69	5.26	8.78	14.97	6.45	3.34
32	0.54	31.03	14.79	9.74	4.52	6.75	7.40	7.10	6.50	13.39	7.68	7.55
33	0.18	35.06	20.83	14.27	5.61	4.71	15.07	6.36	8.11	15.97	13.92	6.99
34	0.04	39.68	28.43	18.67	11.35	7.38	17.57	9.09	12.96	14.48	17.27	10.75
35	1.83	45.54	38.38	24.07	18.16	11.51	18.33	13.99	19.32	10.60	19.02	15.39
时序/月	C183				C196				C471			
	相对误差/%				相对误差/%				相对误差/%			
	NAR	GM(1,1)	多项式	ARIMA	NAR	GM(1,1)	多项式	ARIMA	NAR	GM(1,1)	多项式	ARIMA
30	2.69	5.28	3.71	0.69	3.11	7.99	3.04	3.02	12.65	7.03	5.97	10.60
31	19.33	2.89	5.99	1.85	7.44	6.38	5.09	0.83	1.88	4.47	5.38	9.37
32	44.05	4.15	4.74	2.75	10.81	5.41	6.37	3.13	7.97	4.44	7.60	8.18
33	43.48	8.37	0.57	0.71	10.93	9.58	2.92	0.60	7.21	8.09	13.99	9.28
34	35.18	7.21	1.17	2.20	3.94	8.26	4.17	4.26	3.11	9.15	18.03	7.41
35	25.15	3.90	3.59	6.66	7.37	5.14	6.85	8.28	2.46	8.43	20.40	4.07

表 2 模型的预测结果对比

模型	最大相对 误差/%	平均相对 误差/%	最小相对 误差/%	相对误差	相对误差
				<5% /个数	<10% /个数
NAR	44.05	9.96	0.04	16	24
GM(1,1)	45.54	12.37	0.47	7	23
多项式	38.38	10.60	0.57	9	21
ARIMA	24.07	6.90	0.60	15	29

## 参考文献

- [1] 谢丽,张振克,刘惠.深圳围填海及其新生成土地开发利用研究[J].皖西学院学报,2015,31(5):134-139.
- [2] 卫诗韵,付东洋,刘大召,等.改革开放 40 年深圳海岸线变化的遥感监测[J/OL].热带地理,2023,43(5):986-1004.
- [3] 王震,赵振业,周连宁,等.深圳市空港新城近岸海域海岸线及围填海变化情况初探[J].环境科学导刊,2019,38(4):91-96.
- [4] 吕传振,安动动.灰色预测模型在地面沉降中的应用[J].测绘与空间地理信息,2021,44(3):73-75.
- [5] 孙钟磊,钱尊岩.基于灰色神经网络 GNNM(1,1)模型的地面沉降预测分析[J].测绘与空间地理信息,2020,43(11):166-168.
- [6] 王琰,吕航,谷复光.多元线性回归方法在地面沉降量预测中的解析及应用[J].安全与环境工程,2021,28(3):156-161.
- [7] 范珊珊,郭海朋,朱菊艳,等.线性回归模型在北京平原地面沉降预测中的应用[J].中国地质灾害与防治学报,2013,24(1):70-74.
- [8] 黄婧,刘亮.基于多项式回归模型的天津港地面沉降预测[J].山西建筑,2021,47(2):67-69.
- [9] 朱宝,罗孝文,吴自银.基于 ARIMA 与 LSTM 的海岸带地面沉降预测方法——以杭州湾地区为例[J].海洋学研究,2022,40(02):53-61.
- [10] 牛全福,李月锋,张曼,等.基于 Kalman 滤波的 ARIMA-NAR 神经网络模型在深基坑沉降监测中的应用[J].兰州理工大学学报,2022,48(2):131-135.
- [11] CONTRERAS J, ESPINOLA R, NOGALES FJ, et al.. ARIMA models to predict next-day electricity prices [J]. IEEE Transactions on Power Systems, 2003, 18(3): 1014-1020.
- [12] Yassin I, Zabidi A, Salleh M, et al.. Malaysian Tourism Interest Forecasting Using Nonlinear Autoregressive (NAR) Model [C]. Proc. IEEE 3rd International Conference on System Engineering and Technology. IEEE, 2013:32-36.
- [13] 谢文军.不同手段和尺度下地面沉降时空变化特征分析——以深圳大空港区为例[J].测绘通报,2022,548(11):138-143.
- [14] 刘罗曼.时间序列平稳性检验[J].沈阳师范大学学报(自然科学版),2010,28(3):357-359.
- [15] 蒋金良,林广明.基于 ARIMA 模型的自动站风速预测[J].控制理论与应用,2008,25(2):374-376.
- [16] Khalil Benmouiza and Ali Chekane. Forecasting hourly global solar radiation using hybrid kmeans and nonlinear autoregressive neural network models[J]. Energy Conversion and Management, 2013,75(2):561-569.

作者简介: 王艳(1992—),女,吉林长春人,硕士,工程师,主要从事水文地质、工程地质及地质灾害防治类工作。

E-mail:1784874092@qq.com

通讯作者: 邓英尔(1967—),男,湖南邵阳人,教授,博士,研究方向:地质工程、地下水、岩土力学与渗流等。

E-mail:dengye6789@sina.com