

谢江丽、李帅、姚远, 2019, 地震应急数据库中人口数据预测——以乌鲁木齐市为例, 中国地震, 35(2), 389~398.

地震应急数据库中人口数据预测 ——以乌鲁木齐市为例

谢江丽 李帅 姚远

新疆维吾尔自治区地震局, 乌鲁木齐 830011

摘要 利用统计年鉴、人口普查数据、遥感影像等资料, 将乌鲁木齐市按区县级行政区域为单位提取近 10 年的人口数据, 确定各区县历年来人口分布变化情况。挖掘研究区内人口-时间变量关系, 建立人口预测模型——GM(1,1)模型, 预测未来 2 年乌鲁木齐各区县人口数据, 以弥补数据库数据因滞后 2 年无法及时更新所导致的数据空缺。

关键词: GM(1,1)模型 人口预测 应急数据 乌鲁木齐

[文章编号] 1001-4683(2019)02-0389-10 [中图分类号] P315 [文献标识码] A

0 引言

“十五”期间, 我国 31 个省(直辖市、自治区)建立了地震应急指挥技术系统(何少林等, 2007), 其中, 地震应急基础数据是整个地震应急指挥系统进行抗震救灾指挥决策的基础和核心, 为地震灾害评估和了解地震灾害影响提供背景数据、为抗震救灾指挥决策提供救灾力量储备数据(姜立新等, 2004)。地震应急基础数据库的数据对地震灾害损失快速评估有决定性影响, 对以往的地震灾害(汶川地震、新疆伽师地震等)分析可知, 其中人口分布情况会直接影响地震灾害损失快速评估的结果。地震应急数据库中的人口数据的更新具有复杂性、艰巨性和长期性的特征, 无法实现随社会发展同步更新。本研究主要通过统计人口数据, 采用定性描述及定量分析模型, 反演出人口等在一定时间和一定地理空间中的分布状态, 以弥补因数据库无法及时更新而出现的数据空缺。

1 研究区数据收集

城市在人类发展中占据着重要的地位, 人作为城市最主要的组成部分, 其不同的分布、集散及组合情况能够反映出城市经济发展、城市规划建设、交通运输等差异状况。乌鲁木齐市城区南边为戈壁, 气候环境恶劣, 因此受地形地貌、资源环境的影响, 乌鲁木齐市城区发展向北边延伸, 如新市区的人口聚集。随着经济水平的提高, 外来人口逐渐向城市聚集, 原有

[收稿日期] 2018-05-30; [修定日期] 2018-09-17

[项目类别] 新疆自然科学基金面上项目(2017D01A66)、中国地震局地震应急青年重点任务(CEA EDEM.201719)共同资助

[作者简介] 谢江丽, 女, 1987 年生, 本科, 工程师, 主要从事地震活动性研究。E-mail: 670198463e@qq.com

的市区规划已不能满足人口增长的需求,天山区、沙依巴克区等老城区必然面临着改造,因此人口变动空间格局具有历史的必然性。

本研究将乌鲁木齐市按行政区划选取天山区、沙依巴克区、新市区、水磨沟区、头屯河区、米东区、达坂城区、乌鲁木齐县等作为研究区,利用《乌鲁木齐统计年鉴》(陈志敏,2008、2009、2010;王伟,2011;黄国森,2012、2013、2014;潘世锦,2015、2016、2017)、人口普查数据来提取人口数据(表1)。

表 1 2007~2015 年乌鲁木齐各区县人口数据

年份	年末人口总万人								人口自然增长率/%							
	天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县	天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县
2015	57.26	55.15	64.12	29.22	26.70	27.09	4.13	6.29	6.40	5.50	5.90	5.60	7.10	9.40	-14.00	11.00
2014	59.96	54.99	60.72	28.73	23.70	28.35	4.16	6.26	7.20	6.80	8.60	6.70	4.60	9.40	7.40	12.10
2013	58.61	55.05	59.69	28.47	23.06	27.67	4.21	6.17	6.52	6.35	7.62	6.01	4.95	8.88	6.15	11.57
2012	57.62	53.75	55.62	30.14	23.02	27.43	4.22	5.98	5.34	4.86	6.40	5.40	5.55	7.81	5.00	18.07
2011	55.79	52.74	52.88	28.20	21.95	27.57	4.44	5.78	5.13	4.68	6.47	5.66	5.55	7.75	5.67	13.83
2010	55.28	52.24	53.46	26.62	13.88	27.64	4.47	9.45	4.41	4.58	5.66	4.72	1.88	0.99	4.93	0.64
2009	55.01	52.49	51.82	26.25	13.74	27.99	4.58	9.31	4.49	4.91	5.69	4.42	3.12	5.61	6.52	9.31
2008	53.99	51.46	51.37	24.43	13.20	28.04	4.60	9.11	5.02	4.65	5.83	5.41	3.85	6.43	7.40	10.67
2007	51.54	52.27	48.32	23.56	12.79	29.58	4.60	8.99	5.27	4.89	5.90	5.36	4.49			8.55

由 2007~2015 年乌鲁木齐市各区县人口分布比例(表2)可见,天山区、水磨沟区、达坂城区基本保持平衡,新市区、头屯河区所占比例有所增加,沙依巴克区、米东区、乌鲁木齐县所占比例有所减少,由此可见,乌鲁木齐市人口逐渐由沙依巴克区、米东区、乌鲁木齐县转移到新市区、头屯河区。

表 2 2007~2015 年乌鲁木齐各区县人口所占比例

年份	比例/%							
	天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县
2015	21	20	24	10	10	10	2	2
2014	22	21	23	11	9	11	1	2
2013	22	21	23	11	9	10	2	2
2012	22	21	21	12	9	11	2	2
2011	22	20	20	11	8	11	2	6
2010	23	21	22	11	6	11	2	4
2009	22	21	21	10	6	12	2	6
2008	23	22	21	11	6	11	2	4
2007	22	23	21	10	5	13	2	4

通过对乌鲁木齐市人口变动情况的分析,可得到以下结论:近 10 年来主城区人口数量呈增长的趋势,但表现出“头屯河区和新市区人口数量增长率大、密度小;天山区、沙依巴克

区人口数量增长率小、密度大”的区域差异格局(毛亚会等,2017)。

依据收集的人口数据,得到乌鲁木齐市各区县人口随时间变化图(2007~2015年)(图1)。由图1可见,相对于五普数据,六普数据波动更大,大部分地区人口呈增加趋势。头屯河区、乌鲁木齐县人口数据在2011年变化较大,经查找资料可知,由于行政区域重新规划,2个相邻区县面积相互调整,导致2010~2011年数据出现阶梯,故头屯河区、乌鲁木齐县人口数据采用2011~2015年人口普查数据。

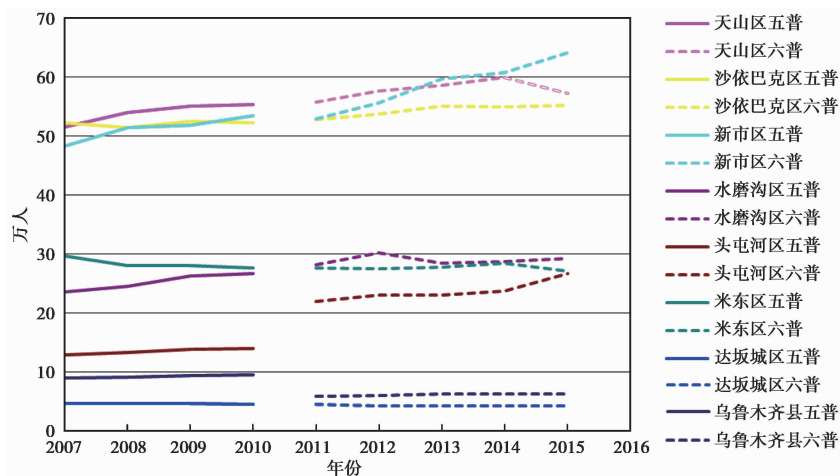


图1 乌鲁木齐各区县人口随时间的分布

2007~2010年为第五次人口普查(简称为五普)数据;2011~2015年为第六次人口普查(简称为六普)数据

2 理论分析与模型建立

选择具有适用性和科学性的人口预测模型是所得预测结果能够符合人口发展趋势的前提。人口预测研究的方法主要有基于微分方程的 Logistic 模型、基于数理统计的线性回归模型、基于矩阵法的 Leslie 模型。总体可概括为一元线性回归模型、多元回归模型、人口增长率法、指数函数法、幂函数法、Leslie 模型、Logistic 模型、灰色系统 GM(1,1)模型、系统动力学等。人口预测方法有其优点和适用范围,可对不同的人口发展变化规律进行预测,因此在选择具体方法时必须结合人口预测的特点。

2.1 灰色系统 GM(1,1)模型

1982年邓聚龙(1990)创立灰色系统理论以研究部分信息已知、部分信息未知的系统,并提出灰色系统预测的概念。其基本思路是,通过对原始数据进行有规则的处理,寻求数据间的内在联系,并对生成变换后的数列建立连续的微分方程动态模型,通过对该模型的求解变换得到预测结果。其中,应用较为广泛的是 GM(1,1)模型,它是一个一阶单序列的线性动态模型,该模型通过五步建模灰色生成或序列算子的作用弱化原始数据的随机性,使其呈现较为明显的特征规律,对于原始数据序列较短且随机性较强的预测具有较好的效果。

对于给定的原始时间序列 $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$, 经过累加生成新的数据序列 $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$, 其中, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $k=1, 2, \dots, n$ 。作均值

序列 $Z^{(1)}(k) = 0.5x^{(1)}(k-1) + 0.5x^{(1)}(k)$, 构造数据矩阵 B

$$B = \begin{bmatrix} -Z^{(1)}(2), & 1 \\ -Z^{(1)}(3), & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n), & 1 \end{bmatrix} = \begin{bmatrix} -0.5[x^{(1)}(1) + x^{(1)}(2)], & 1 \\ -0.5[x^{(1)}(2) + x^{(1)}(3)], & 1 \\ \vdots & \vdots \\ -0.5[x^{(1)}(n-1) + x^{(1)}(n)], & 1 \end{bmatrix}$$

构造数据向量 $Y_N = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$; 则白化形式的微分方程为 $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$,

式中, a 、 b 为待定系数, 分别称为发展系数和灰色作用量, a 的有效区间为 $(-2, 2)$ 。应用最小二乘法可求得: $a = (B^T B)^{-1} B^T Y_N = [a, b]$; 进而得到时间序列预测模型

$$X^{(1)}(k+1) = [X^{(0)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a} \quad (1)$$

式(1)即为一阶累加的拟合模型。

采用灰色系统理论对 2007~2015 年的数据进行拟合, 人口数量拟合结果见表 3。通过计算得到 a 、 b 值(表 4)。

表 5 为精度检验等级参照表, 精度等级指标预测值越小, 则精度越高。由表 5 可见, 该预测结果的相对误差基本都达到一级或接近一级水平。利用上述模型对 2016~2017 年乌鲁木齐各区县人口数据进行预测, 预测结果如表 6 所示。

2.2 一元线性回归的建立

在一定时期内人口的发展, 人口的增长趋势可能会在很短的时间内变化。假设人口发展过程线上任一点基本保持不变, 人口发展速度的切线斜率是一致的, 即每个周期的近似线性形式可以采用一元线性回归方法。一元线性回归也叫时间序列法, 其原理是利用一元线性回归来进行预测的。基本模型为 $y = a + bx$, a 为回归常数, 是回归直线的截距; b 为回归系数, 是回归直线的斜率。根据历史数据(样本)和最小二乘(OLS)可以计算出模型的参数 a 、 b (王宇等, 2016)。

根据 2007~2015 年的原始数据, 使用线性回归方法建立了预测模型(表 7)。由表 7 可见, 显著性概率 sig 大部分为 0.000~0.003, 相关系数为 0.728~0.929, 说明乌鲁木齐各区县人口与年份之间存在显著的线性关系, 各区县的模型效果各不相同, 说明每个区县的人口变化情况各不相同, 影响因子较复杂。因此, 用该模型进行乌鲁木齐各区县人口预测, 表 8 为对乌鲁木齐各区县人口的原始数据拟合结果, 表 9 为运用线性方程得出的 2016~2017 年乌鲁木齐各区县人口数据预测结果。

2.3 不同人口预测模型的结果验证与分析统计

通过对比乌鲁木齐各区县的统计值、灰色系统 GM(1,1)模型预测值和一元线性回归模型预测值曲线图(图 2~9)发现, 灰色系统 GM(1,1)模型能够在更高层次上描述系统的动态过程。灰色系统 GM(1,1)模型与一元线性回归预测相比, 前者适当过滤了偶然的或不完整的信息和数据预测, 提高了预测精度。

将 2017 年乌鲁木齐统计年鉴中 2016 年乌鲁木齐各区县人口数据与灰色系统 GM(1,1)模型预测值和一元线性回归模型预测值进行对比验证(表 10), 由表 10 可见, 灰色系统 GM(1,1)模型预测更接近统计值, 相对误差更小, 其具有较强的实用性和有效性, 为预测乌鲁木齐

表 3 采用灰色系统 GM(1,1) 模型对乌鲁木齐各区人口数据预测拟合

年份	类别	万人数							
		天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县
2015	统计	57.26	55.15	64.12	29.22	26.70	27.09	4.13	6.29
	预测	59.18	55.49	62.97	29.98	25.93	27.51	4.09	6.32
	相对误差	0.03	0.01	0.01	0.02	0.02	0.01	0.00	0.01
2014	统计	59.96	54.99	60.72	28.73	23.70	28.35	4.16	6.26
	预测	58.45	54.91	60.91	29.31	24.68	27.57	4.16	6.22
	相对误差	0.02	0.00	0.00	0.02	0.04	0.02	0.00	0.00
2013	统计	58.61	55.05	59.69	28.47	23.06	27.67	4.21	6.17
	预测	57.73	54.32	58.91	28.67	23.48	27.63	4.23	6.12
	相对误差	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00
2012	统计	57.62	53.75	55.62	30.14	23.02	27.43	4.22	5.98
	预测	57.02	53.75	56.97	28.03	22.34	27.69	4.31	6.02
	相对误差	0.01	0.00	0.02	0.06	0.02	0.00	0.02	0.00
2011	统计	55.79	52.74	52.88	28.20	21.95	27.57	4.44	5.78
	预测	56.32	53.18	55.11	27.41	21.95	27.75	4.38	5.78
	相对误差	0.01	0.01	0.04	0.02	0	0.00	0.01	0
2010	统计	55.28	52.24	53.46	26.62		27.64	4.47	
	预测	55.62	52.62	53.30	26.81		27.81	4.46	
	相对误差	0.01	0.01	0.00	0.01		0.00	0.00	
2009	统计	55.01	52.49	51.82	26.25		27.99	4.58	
	预测	54.94	52.06	51.55	26.21		27.86	4.54	
	相对误差	0.00	0.01	0.00	0.00		0.00	0.01	
2008	统计	53.99	51.46	51.37	24.43		28.04	4.60	
	预测	54.26	51.51	49.86	25.63		27.92	4.62	
	相对误差	0.00	0.00	0.03	0.04		0.00	0.00	
2007	统计	51.54	52.27	48.32	23.56		29.58	4.60	
	预测	51.54	52.27	48.32	23.56		29.58	4.60	
	相对误差	0	0	0	0		0	0	

表 4 乌鲁木齐各区县人口增长 GM(1,1) 模型的参数 a, b

区县	系数	
	a	b
天山区	-0.0124	53.2838
沙依巴克区	-0.0106	50.6860
新市区	-0.0334	47.4167
水磨沟区	-0.0224	24.8192
头屯河区	-0.0495	20.7388
米东区	0.0021	28.0191
达坂城区	0.0175	4.7439
乌鲁木齐县	-0.0164	5.8792

表 5 精度检验等级参照表

精度等级指标预测值	相对误差
一级	0.01
二级	0.05
三级	0.10
四级	0.20

表 6 2016~2017 年乌鲁木齐各区人口数据预测结果

年份	万人数							
	天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县
2016	59.92	56.08	65.11	30.66	27.25	27.46	4.02	6.43
2017	60.67	56.68	67.32	31.35	28.64	27.40	3.95	6.53

表 7 乌鲁木齐各区县人口数据一元线性回归参数

区县	系数						
	a	b	R^2	Sig	常数 t	斜率 t	F
天山区	-1630.780	0.839	0.811	0.001	-5.299	5.481	30.044
沙依巴克区	-909.920	0.479	0.857	0	-6.120	6.479	41.981
新市区	-3603.010	1.819	0.936	0	-9.997	10.151	103.042
水磨沟区	-1430.680	0.725	0.728	0.001	-5.083	5.180	26.834
头屯河区	-2025.550	1.018	0.801	0.400	-3.438	3.478	12.096
米东区	359.074	-0.165	0.392	0.071	2.303	-2.124	4.511
达坂城区	144.981	-0.070	0.932	0	10.091	-9.786	95.767
乌鲁木齐县	-255.594	0.130	0.922	0.009	-5.812	5.950	35.405

表 8 一元线性回归对乌鲁木齐各区县人口数据预测拟合

年份	类别	万人数							
		天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县
2015	统计	57.26	55.15	64.12	29.22	26.70	27.09	4.13	6.29
	预测	59.81	55.27	62.27	30.19	25.72	26.60	3.93	6.36
	相对误差	0.04	0.00	-0.03	0.03	-0.04	-0.02	-0.05	0.01
2014	统计	59.96	54.99	60.72	28.73	23.70	28.35	4.16	6.26
	预测	58.97	54.79	60.45	29.47	24.70	26.76	4.00	6.23
	相对误差	-0.02	0.00	0.00	0.03	0.04	-0.06	-0.04	0.00
2013	统计	58.61	55.05	59.69	28.47	23.06	27.67	4.21	6.17
	预测	58.13	54.31	58.64	28.74	23.69	26.93	4.07	6.10
	相对误差	-0.01	-0.01	-0.02	0.01	0.03	-0.03	-0.03	-0.01
2012	统计	57.62	53.75	55.62	30.14	23.02	27.43	4.22	5.98
	预测	57.29	53.83	56.82	28.02	22.67	27.09	4.14	5.97
	相对误差	-0.01	0.00	0.02	-0.07	-0.02	-0.01	-0.02	0.00
2011	统计	55.79	52.74	52.88	28.20	21.95	27.57	4.44	5.78
	预测	56.45	53.35	55.00	27.29	21.65	27.26	4.21	5.84
	相对误差	0.01	0.01	0.04	-0.03	-0.01	-0.01	-0.05	0.01
2010	统计	55.28	52.24	53.46	26.62		27.64	4.47	
	预测	55.61	52.87	53.18	26.57		27.42	4.28	
	相对误差	0.01	0.01	-0.01	0.00		-0.01	-0.04	
2009	统计	55.01	52.49	51.82	26.25		27.99	4.58	
	预测	54.77	52.39	51.36	25.84		27.59	4.35	
	相对误差	0.00	0.00	-0.01	-0.02		-0.01	-0.05	
2008	统计	53.99	51.46	51.37	24.43		28.04	4.6	
	预测	53.94	51.91	49.54	25.12		27.75	4.42	
	相对误差	0.00	0.01	-0.04	0.03		-0.01	-0.04	
2007	统计	51.54	52.27	48.32	23.56		29.58	4.60	
	预测	53.10	51.43	47.72	24.39		27.92	4.49	
	相对误差	0.03	-0.02	-0.01	0.03		-0.06	-0.02	

表 9 一元线性回归对 2016~2017 年乌鲁木齐各区人口数据预测结果

年份	万人							
	天山区	沙依巴克区	新市区	水磨沟区	头屯河区	米东区	达坂城区	乌鲁木齐县
2016	60.65	55.74	64.09	30.92	26.74	26.43	3.86	6.48
2017	61.49	56.22	65.91	31.64	27.75	26.27	3.79	6.61

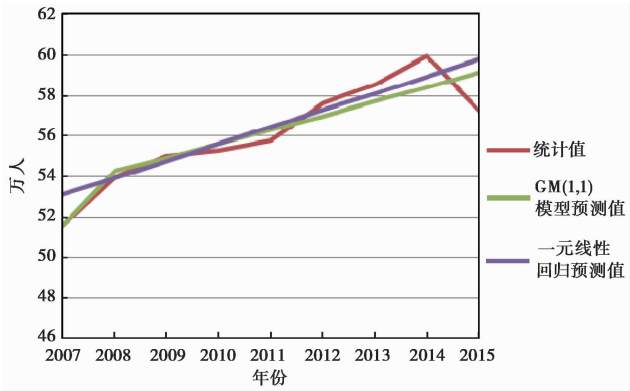


图 2 天山区统计值与预测值

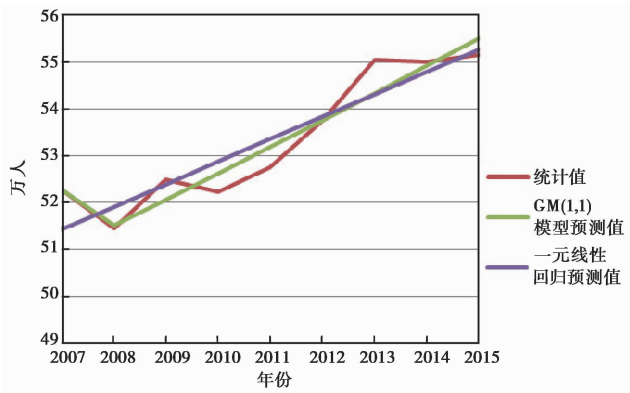


图 3 沙依巴克区统计值与预测值

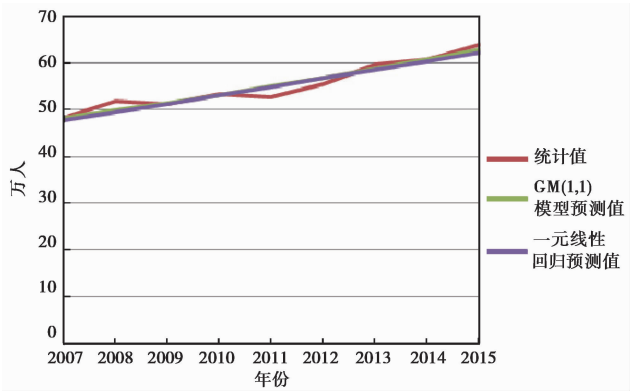


图 4 新市区统计值与预测值

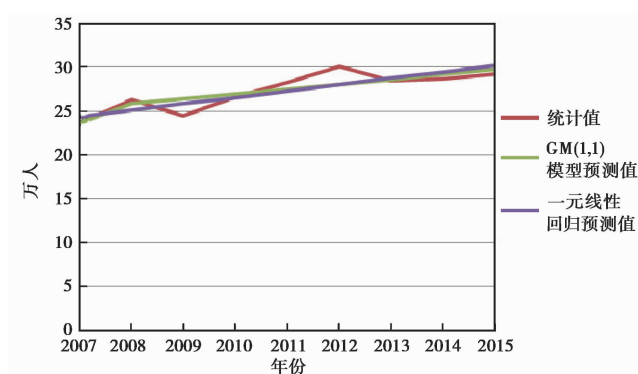


图 5 水磨沟区统计值与预测值

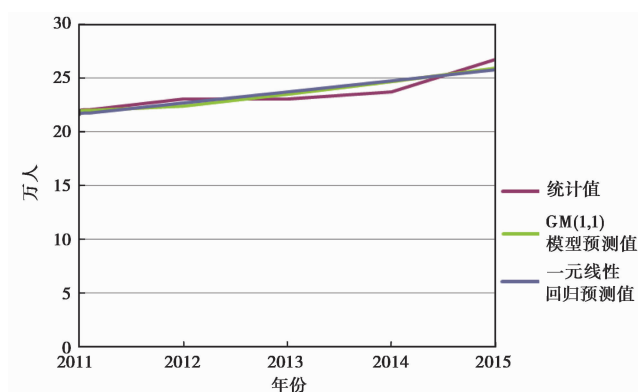


图 6 头屯河区统计值与预测值

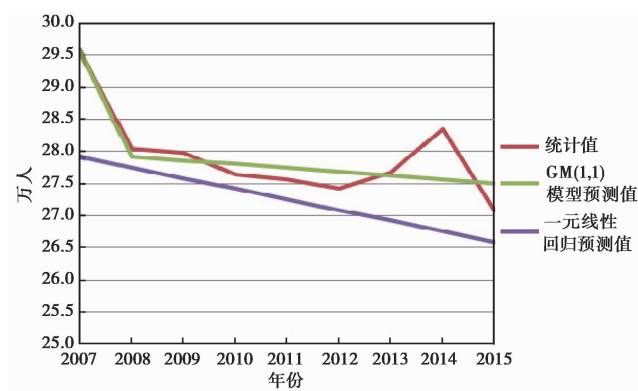


图 7 米东区统计值与预测值

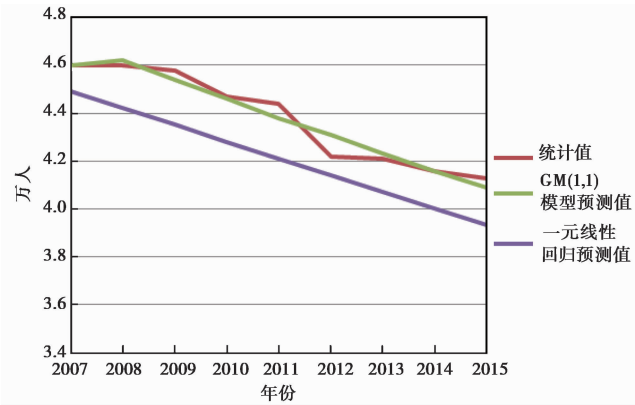


图 8 达坂城区统计值与预测值

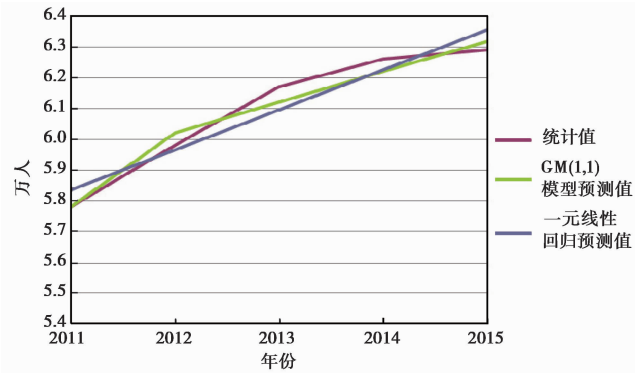


图 9 乌鲁木齐县统计值与预测值

表 10 2016 年乌鲁木齐各区县人口数据和预测数据

区县	统计数据/万人	GM(1,1) 模型		一元线性模型	
		预测结果/万人	相对误差	预测结果/万人	相对误差
天山区	56.88	59.81	0.05	60.56	0.06
沙依巴克区	56.37	55.97	0	55.99	0
新市区	63.10	65.07	0.03	65.06	0.03
水磨沟区	21.59	30.48	0.40	31.43	0.45
头屯河区	22.38	27.25	0.20	26.74	0.19
米东区	26.47	26.62	0	26.03	0.01
达坂城区	4.03	4.02	0	3.67	0.08
乌鲁木齐县	6.30	6.43	0.02	6.48	0.02

齐各区县人口数据的理想模型。

3 结语

地震应急数据库中人口、经济等数据要求每年及时更新,但由于统计数据时间延迟,常使得数据更新周期较长。本研究主要利用数学的方法,从建立增长模型的角度,利用乌鲁木

齐统计年鉴提取基础数据进行预测,以建立乌鲁木齐市一级人口数据增长模型,实现静态人口基础数据在地震发生时的动态计算,使其更接近于人口发展的统计情况;在得到统计数据之前,本研究中得到的人口数据可作为应急数据库中相关基础数据更新的补充手段,或作为进行未来几年震害预测时的参考基础数据。

参考文献

- 何少林、李佐唐、高安泰,2007,甘肃省地震应急指挥技术系统设计与实现,地震地磁观测与研究,28(5),71~79.
- 姜立新、吴天安、刘在涛等,2004,地震现场应急指挥技术系统的结构与实现,地震,24(3),35~41.
- 邓聚龙,1990,灰色系统理论教程,武汉:华中理工大学出版社.
- 陈志敏,2008,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 陈志敏,2009,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 陈志敏,2010,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 王伟,2011,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 黄国森,2012,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 黄国森,2013,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 黄国森,2014,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 潘世锦,2015,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 潘世锦,2016,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 潘世锦,2017,乌鲁木齐统计年鉴,北京:中国统计出版社.
- 毛亚会、余丹林、郑江华,2017,近10年来乌鲁木齐市城市人口变动分析,西北师范大学学报(自然科学版),53(2),118~122、134.
- 王宇、敬莉,2016,基于灰色GM(1,1)模型和一元线性回归的人口发展趋势预测——以新疆南疆三地州为例,西部学刊,(7),54~59.

Prediction of Population Data in Earthquake Emergency Database——a Case Study of Urumqi

Xie Jiangli Li Shuai Yao Yuan

Earthquake Agency of Xinjiang Uygur Autonomous Region, Urumqi 830011, China

Abstract By using statistical yearbooks, census data, remote sensing images, we extracted all population data for the past 10 years at the Urumqi City's administrative region units, and we determined the changes in population distribution over the years in each unit of the study area. By excavating the relationship between population and time variables in the study area, we established a population forecasting model(GM(1,1) model) to predict population data in various districts of Urumqi in the next two years, and to make up for the data gaps that could not be updated in the next two years in the database.

Key words: GM(1,1) model; Population projections; Emergency database; Urumqi