

土地利用规划中建设用地预测方法比较

荣联伟 师学义 杨 静

(中国地质大学(北京)土地科学技术学院,北京 100083)

摘要 为探索提高建设用地预测精度和预测方法比较,该文运用多元线性回归模型、GM(1,1)模型和BP神经网络模型对晋城市建设用地进行预测,预测结果说明三种模型的平均相对误差和预测精度均存在一定差异。研究表明:建设用地预测中应结合数据本身的特点和数据收集的难易程度选择不同模型;多角度多方法进行预测比较,可得到预测精度较高的预测结果;对于建设用地预测方法比较,可以从模型本质特性、数据采用形式和数据获取难易程度等7个方面进行综合比较。该研究为科学选择建设用地预测方法、探索预测方法比较提供一种思路,为土地利用规划修编和土地资源可持续利用提供科学依据。

关键词 建设用地预测;多元线性回归模型;GM(1,1)模型;BP神经网络模型;方法比较

中图分类号 F301.23

文献标识码 A

文章编号 1002-2104(2014)11增-0199-05

建设用地需求量预测是土地需求量预测的核心,科学准确预测建设用地需求量是提高土地利用规划的科学性和可操作性的重要途径。目前建设用地预测方法研究较为广泛,国外对建设用地的预测主要运用GIS技术、CA技术和经济增长理论模型等方法分析,并结合详细调查,在对建设用地规模预测的同时也进行用地布局规划^[1-3]。国内建设用地预测方法主要有定额指标法、分项预测法、C-D生产函数模型、多元线性回归模型、双因素模型、灰色预测模型和BP神经网络模型等^[4-9]。对于建设用地预测方法比较研究仍然较少,比较研究内容还不够充分。罗晔辉和吴次芳对国内外建设用地预测研究方法进行了评述,并结合新一轮土地利用规划修编要求,对建设用地需求预测方法进行理论分析和实证研究^[10]。陈建琼、李红伟、李珊珊从常规方法入手,分别从预测方法原理、优缺点、精度和适用条件等方面,对建设用地需求量预测方法进行了比较研究^[11-13]。为探索提高建设用地预测精度和预测方法多维比较,本文选择多元线性回归模型、灰色GM(1,1)模型和BP神经网络模型等3种常用建设用地预测方法,采用理论分析与实证研究,通过模型本质特性、数据采用形式和历史数据长度影响等7个方面对预测方法综合比较,最后得出研究结论。

1 研究区概况与研究方法

1.1 研究区概况

晋城市位于东经111°55′-113°37′,北纬35°12′-36°00′,处于山西省东南部,区位优势适中,交通便捷,是山西通往中原地区的重要门户。据2005年晋城市统计资料,全市生产总值316.00亿元,一、二、三产业结构为4.7:64.2:31.2;全市总人口为220.54万人,全市土地总面积为942042.85hm²。从土地利用结构来看,全市农用地、建设用地、其他用地面积比例为73.44%、4.85%、21.71%。随着城镇化和工业化进程加快,建设用地规模不断增大,耕地保护与建设用地保障矛盾日益突出,科学预测与有效控制建设用地规模具有重要的现实意义。

1.2 数据来源

研究所用数据来源于《1996-2010年晋城市统计年鉴》和《2008-2010年晋城市国民经济和社会发展统计公报》。根据研究区的自然条件和社会经济条件,选取相关影响因子进行研究。

1.3 研究方法

1.3.1 建设用地总量影响因素分析

根据学者们对建设用地驱动力研究^[4,7],结合晋城市自然和社会经济条件,初步选定总人口、生产总值、二三产业所占比重、全社会固定资产投资、一般预算支出、城镇化

收稿日期:2014-06-21

作者简介:荣联伟,硕士生,主要研究方向为土地资源评价与土地利用规划。

通讯作者:师学义,博士,教授,主要研究方向为土地利用规划与土地利用工程。

基金项目:国土资源部公益性行业科研专项经费项目(编号:201111015-04)。

率、非农人口比重和城乡居民收入等8个建设用地影响因素,运用SPSS20.0进行Pearson相关性分析,再根据建设用地预测模型的数据需要进行综合筛选,得出影响晋城市建设用地主要因素为总人口和全社会固定资产投资。

1.3.2 基于多元线性回归模型的建设用地预测

以总人口(X_1)、全社会固定资产投资(X_2)这2个主要因素作为自变量,建设用地总量(Y)作为因变量,利用SPSS20.0软件分别对晋城市1996-2007年和2002-2007年建设用地总量、总人口和全社会固定资产投资建立多元

线性回归模型。通过多元统计分析可知,两组不同长度历史数据建立的多元线性回归模型,均符合参数估计和假设检验,因此均可用来预测晋城市建设用地总量(如表1所示)。为了便于进行不同历史长度数据对模型预测结果的影响分析,统一采用GM(1,1)模型,并以2002-2007年的总人口和全社会固定资产投资为原始数据,预测2008-2010三年的总人口和全社会固定资产,再将所预测的这2个变量数据分别代入多元线性回归模型,得到2008-2010年建设用地总量的预测值(见表4、表5)。

表1 多元线性回归模型统计分析

历史数据长度	回归方程	拟合优度 R	方差分析 Sig.	D-W 检验值
2002-2007	$Y = 20101.811 + 110.939 * X_1 + 10.127 * X_2$	0.973	0.006	2.940
1996-2007	$Y = -19089.217 + 291.221 * X_1 + 6.283 * X_2$	0.983	0.000	1.742

1.3.3 基于GM(1,1)模型的建设用地预测

以2002-2007年和1996-2007年两组不同历史数据长度的晋城市建设用地总量为原始数据,利用灰色系统软件和Matlab8.0软件,建立GM(1,1)模型,得到模型的参数估计值、时间响应函数和平均相对误差。在预测过程

中,后验差比值 $C < 0.35$,小误差概率 $P > 0.95$,模型精度检验结果均为一级(好),说明两个模型拟合精度较好,可用建立的GM(1,1)模型来预测晋城市建设用地总量。具体灰色系统建模结果如表2,预测结果如表4、表5。

1.3.4 基于BP神经网络模型的建设用地预测

表2 根据不同历史数据长度建立的GM(1,1)预测模型

历史数据长度	间响应函数	平均相对误差	后验差比 C	小误差概率 P	精度检验
2002-2007年	$x(k+1) = 5015537.708474 \exp(0.008967 * k) - 4970055.048474$	0.303 52	0.253	1.000	一级(好)
1996-2007年	$x(k+1) = 3856955.766108 \exp(0.010888 * k) - 3817174.756108$	0.594 89	0.149	1.000	一级(好)

注:表中的平均相对误差为实际统计值与模拟值的平均相对误差。

利用Matlab8.0软件,通过编写BP神经网络程序代码,实现BP神经网络样本训练和模拟预测功能。将晋城市总人口、全社会固定资产投资和建设用地总量进行归一化处理,并将归一化处理后的总人口和全社会固定资产投资作为输入数据,将归一化后的建设用地总量作为输出期望值。通过模型构建、训练方法与参数确定,及样本训练(分别选择2002-2007年和1996-2007年的两个不同历史数据长度作为训练样本)训练结果通过误差性能和误差变化特征曲线进行表征,最后得到建设用地总量模拟值(如表3所示)。通过模型构建、训练方法与参数确定和样本训练,得到基于BP神经网络的建设用地预测模型。将多元回归预测中用灰色GM(1,1)模型预测得到的2008-2010年总人口和全社会固定资产数据,经归一化后作为变量输入,通过BP神经网络模型预测得到2008-2010年晋城市建设用地总量的归一化数据,再经反归一化处理,得到建设用地总量的预测值(见表4、表5)。

表3 原始数据及模拟值的归一化

年份	总人口	固定资产投资	建设用地总量期望值	1996-2007建设用地总量模拟值	2002-2007建设用地总量模拟值
1996	-1.759 3	-0.916 4	-2.148 7	-1.515 4	-
1997	-1.545 9	-0.827 8	-1.227 9	-1.291 2	-
1998	-1.330 9	-0.758 9	-1.117 2	-1.116 4	-
1999	-1.088 3	-0.589 6	-0.825 4	-0.825 4	-
2000	-0.189 7	-0.675 9	-0.377 6	-0.354 1	-
2001	-0.043 4	-0.675 6	-0.186 6	-0.186 4	-
2002	0.140 0	-0.592 0	0.023 3	0.048 2	-1.600 3
2003	0.236 3	-0.478 5	0.173 2	0.147 6	-1.624 1
2004	0.347 3	-0.274 4	0.263 7	0.235 7	-1.617 6
2005	0.469 9	-0.090 5	0.316 1	0.377 6	-1.415 3
2006	0.604 9	0.180 0	0.662 8	0.626 2	-0.979 6
2007	0.736 2	0.585 4	0.837 4	0.847 1	-0.614 6



2 预测结果与模型比较分析

2.1 预测结果

在晋城市建设用地总量影响因子 Pearson 相关性分析基础上,分别利用多元线性回归模型、GM(1,1)模型和 BP 神经网络模型对建设用地规模进行预测。考虑到相关性分析结果、模型对研究数据要求及变量共线性诊断,本文选取总人口和全社会固定资产投资 2 个主要因素来预测

晋城市建设用地规模。为了便于预测和比较,采用分段预测的方法,分别以 1996-2007 年的建设用地规模和相关数据、2002-2007 年的建设用地规模和相关数据预测 2008-2010 年的建设用地规模。根据表 4、表 5 可得晋城市建设用地预测结果,在不同历史数据长度下,多元线性回归模型、GM(1,1)模型和 BP 神经网络模型的平均相对误差和预测精度存在较大差异。

表 4 1996—2007 年数据预测 2008—2010 年的预测值、模型检验及其误差

模型	2008 年预测值及误差 hm ²	2009 年预测值及误差 hm ²	2010 年预测值及误差 hm ²	模型检验	平均相对误差
实际统计值	47 264.91	47 832.65	48 277.45		
多元线性回归模型	47 565.08 -300.18	48 382.61 -549.97	49 393.18 -1 115.73	决定系数 R ² = 0.978	-0.010 93
GM(1,1)模型	47 250.87 14.04	47 676.50 156.15	48 105.96 171.49	后验差比及小误差概率 (C,P) = (0.253 29,1)	-0.002 37
BP 神经网络模型	47 476.00 -211.09	47 954.00 -121.35	48 662.00 -384.55	训练 R = 0.990 9	-0.004 99

表 5 2002—2007 年数据预测 2008—2010 年的预测值、模型检验及其误差

模型	2008 年预测值及误差 hm ²	2009 年预测值及误差 hm ²	2010 年预测值及误差 hm ²	模型检验	平均相对误差
实际统计值	47 264.91	47 832.65	48 277.45		
多元线性回归模型	47 577.15 -312.24	48 275.53 -442.89	49 094.41 -816.97	决定系数 R ² = 0.983	-0.013 65
GM(1,1)模型	47 594.38 -329.47	48 115.40 -282.75	48 642.12 -364.67	后验差比及小误差概率 (C,P) = (0.148 59,1)	0.006 81
BP 神经网络模型	47 137.00 127.91	47 829.00 3.65	48 279.00 -1.55	训练 R = 0.999 9	0.000 92

2.2 平均相对误差与预测精度

平均相对误差:

$$r = \frac{1}{N} \sum_{i=1}^n (Y_i - Y_{ic}) / Y_i \quad (1)$$

平均预测精度:

$$R = \left[\sum_{i=1}^n (1 - |Y_i - Y_{ic}| / Y_{ic}) \times 100 \right] / N \quad (2)$$

式中, Y_i 为第 i 年的理论观察值; Y_{ic} 为第 i 年的预测值; N 为预测年数,本文取 3。

通过公式(1)、(2)可求得不同长度历史数据对应 3 种模型的平均相对误差和平均预测精度,如表 6。

表 6 模型的平均相对误差和平均预测精度比较

历史数据长度	模型	平均相对误差 r	平均预测精度 R
2002-2007	多元线性回归模型	-0.013 65	98.66
	GM(1,1)模型	-0.002 37	99.76
	BP 神经网络模型	-0.004 99	99.50
1996-2007	多元线性回归模型	-0.010 93	98.92
	GM(1,1)模型	0.006 81	99.32
	BP 神经网络模型	0.000 92	99.91

2.3 预测模型比较分析

2.3.1 模型本质特性

从模型特性来看,多元线性回归模型,是通过历史

数据的分析研究,探索经济、社会各有关因素与建设用地总量的内在联系和发展变化规律,并根据对预测期内,本

地区经济、社会发展情况的预测来推算未来的建设用地总量^[14]; GM(1,1)模型的实质是对原始序列作一次累加生成新序列,使生成序列呈一定规律,然后建立一阶线性微分方程模型,求得拟合曲线以对系统(建设用地总量)进行预测;BP神经网络模型是一种按误差逆向传播算法训练的多层前馈网络,它能学习和存贮大量的输入-输出模式映射关系,而无需事前揭示描述这种映射关系的数学方程。

2.3.2 数据采用形式

从数据采用形式来看,多元线性回归模型直接采用原始数据建模,BP神经网络模型虽然直接采用原始数据,但事先需要进行归一化处理,以消除自变量之间、因变量与自变量之间单位和数量级的差异^[15-16];GM(1,1)模型通过累加生成灰色模型,滤去原始数据中可能混入的随机量或其他噪声,从上下波动的时间数列中寻找某种隐含规律,而不直接利用原始数据。

2.3.3 数据获取难易程度

数据获取难易程度取决于模型本质特性,从模型本质特性可知,多元线性模型和BP神经网络模型综合考虑了社会经济因素对建设用地总量的影响,在模型预测之前需要进行建设用地总量的驱动力分析,筛选出主要影响因素后运用模型对原始数据处理,且要求历史数据较为充分,因此这两个模型数据获取的难易程度较大;GM(1,1)模型只需对原始数据序列(建设用地总量)做数学变换,不需要考虑其影响因素,通过收集近期数据就可以进行预测分析^[17-18],因此GM(1,1)模型的数据获取难易程度最小。

2.3.4 预测过程难易程度

从预测过程来看,BP神经网络模型需要进行建设用地总量驱动力分析、MATLAB编程建模、系统模型方法选取和参数设定、原始数据归一化处理、样本训练和结果预测;多元线性回归模型需要进行建设用地总量驱动力分析、运用SPSS进行多元统计分析和结果预测;GM(1,1)模型只需运用灰色系统软件进行简要建模计算和结果预测。因此预测过程难易程度从大到小依次为BP神经网络模型、多元线性回归模型和GM(1,1)模型。

2.3.5 模型预测精度

由表4、表5和表6可知,在相同历史数据长度条件下,利用多元线性回归模型预测建设用地总量的平均相对误差在3种方法中最大,该模型的拟合度较高,预测精度也较高,但预测结果精度在3种方法中最小。在相同历史数据条件下,利用GM(1,1)模型和BP神经网络模型预测建设用地总量的平均相对误差较低,预测值精度较高,拟合一致性较好;对于GM(1,1)模型和BP神经网络模型比较,其平均相对误差和预测精度在不同历史数据长度条

件下有差异。

2.3.6 历史数据长度影响

由表4、表5和表6可知,不同历史数据长度对多元线性回归模型和GM(1,1)模型的平均相对误差和平均预测精度影响最大,而对BP神经网络模型的平均相对误差和平均预测精度影响较小。对于多元线性回归模型,从决定系数 R^2 和历史数据的长度分析来看, R^2 值越大,已知数据的线性关系越好,预测误差越低;历史数据越充分,预测结果越准确。对于GM(1,1)模型,在历史数据较短的条件下,GM(1,1)模型预测的平均相对误差最低,预测值精度最高;在历史数据较充分的条件下,GM(1,1)模型预测的平均相对误差相比较,预测值精度相比较低。对于BP神经网络模型,在不同历史数据长度下,预测的平均相对误差均较低,其预测值精度均较高。

2.3.7 预测模型适用条件

结合表4、表5和表6,并综合以上6个方面的模型比较分析可知,多元线性回归模型适用于大样本历史数据充分,且过去、现在和未来发展模式一致,历史数据基本无突变点的建设用地总量中长期预测。GM(1,1)模型适用于小样本、贫信息、历史数据不充分,且历史数据波动性较大的建设用地总量短期预测。BP神经网络模型适用于不同历史数据长度、不同历史数据波动性水平、目标系统影响因素复杂、平均相对误差和精度有要求的建设用地总量不同时段预测。

3 结论与讨论

通过对建设用地总量预测方法的比较分析可以得出:①应结合数据本身的特点和数据收集的难易程度选择不同的预测模型。②多角度多方法地进行预测方法比较,可得到预测精度较高的建设用地预测结果。③对于建设用地预测方法的比较,可以从模型本质特性、数据采用形式、数据获取难易程度、模型预测精度、历史数据长度影响和模型适用条件7个方面进行方法的综合比较分析和规律总结,为探索提高建设用地预测精度和预测方法比较提供一种思路。

建设用地扩张是一个复杂的系统过程,建设用地总量除了受社会经济条件的影响外,还受自然因素(生态环境承载容量)和政策因素(产业结构调整、土地宏观政策调控)的影响。然而以现有的技术和研究水平,还难以定量研究这两种因素对建设用地总量的影响程度和影响机制。因此,在以后的建设用地总量(规模)预测研究中,根据现有数据和已有资料,合理选用预测模型和方法,在建设用地总量预测后,探讨自然因素和政策因素对建设用地总量的影响机制,并结合情景模拟分析,对建设用地总量预测

值进行修正,得到与实际值较为贴近的预测值。

(编辑:刘呈庆)

参考文献

- [1] Yunyan Du, Wei Wen, Feng Cao, et al. A Case - based Reasoning Approach for Land Use Change Prediction [J]. Expert Systems with Applications 2010 37(8): 5745 - 5750.
- [2] Jian Sun, Lihui Zhang, Chunlu Peng, et al. CA - Based Urban Land Use Prediction Model: A Case Study on Orange County, Florida, U. S. [J]. Journal of Transportation Systems Engineering and Information Technology 2012 12(6): 85 - 92.
- [3] Puertas O L, Henríquez C, Meza F J. Assessing Spatial Dynamics of Urban Growth Using an Integrated Land Use Model: Application in Santiago Metropolitan Area, 2010 - 2045 [J]. Land Use Policy, 2014 38: 415 - 425.
- [4] 姜海, 曲福田. 建设用地需求量预测的理论与方法 [J]. 中国土地科学 2005 19(2): 44 - 51.
- [5] 陈华, 邓传明, 狄秀梅. 建设用地需求量预测的探讨 [C]. 中国土地学会年会论文集 2007: 590 - 597.
- [6] 刘胜华, 詹长根. 基于国民经济和人口发展目标的建设用地需求规模预测研究 [J]. 中国人口·资源与环境 2005 15(5): 48 - 51.
- [7] 邱道持, 刘力, 粟辉, 等. 城镇建设用地预测方法新探 [J]. 西南师范大学学报: 自然科学版 2004 29(1): 146 - 150.
- [8] 王玉波, 唐莹. 省域建设用地需求量预测研究 [J]. 科技管理研究 2009 (12): 161 - 165.
- [9] 刘金国, 张希. 中国城市建设用地需求量预测研究综述 [J]. 国土与自然资源研究 2011 (2): 16 - 18.
- [10] 罗罡辉, 吴次芳. 建设用地需求预测方法研究 [J]. 中国土地科学 2004 18(6): 14 - 17.
- [11] 陈建琼. 建设用地需求量预测方法比较研究 [J]. 合作经济与科 2008 (339): 24 - 26.
- [12] 李红伟, 曾永年. 土地利用规划中建设用地预测模型比较研究 [J]. 水土保持研究 2008 15(2): 56 - 58.
- [13] 李姗姗, 于婧, 高崇辉. 建设用地需求量预测方法比较研究 [J]. 国土资源科技管理 2010 7(3): 49 - 53.
- [14] 陈会洋, 周申立, 唐道甫. 广安市建设用地需求量预测研究 [J]. 环境科学与管理 2007 32(2): 119 - 125.
- [15] 林宁, 游和远. 基于 BP 神经网络的建设用地需求量预测 [C]. 建设社会主义新农村土地问题研究 2006: 408 - 412.
- [16] 曹蕾, 钟菲, 章明. 基于灰色——BP 神经网络的重庆市城镇建设用地规模预测. 江西农业学报 2011 23(8): 188 - 189.
- [17] 罗彩云, 李炳新. GM(1,1) 模型在土地规划数据预测中的应用 [C]. 科学合理用地、人地和谐相处 2008 年学术年会论文集, 2008: 101 - 110.
- [18] 彭宝发. 基于灰色系统模型的城乡建设用地规模预测 [J]. 经济地理. 2007 (27): 999 - 1002.

Comparison of Construction Land Prediction Methods in Land Use Planning

RONG Lian-wei SHI Xue-yi YANG Jing

(School of Land Science and Technology, China University of Geosciences, Beijing 100083, China)

Abstract In order to improve the prediction accuracy of construction land and compare prediction methods, multiple linear regression model, GM(1,1) model and BP neural network model are used to predict construction land in Jincheng. The result of the research indicates that both the relative error and average prediction accuracy of three models have some differences to a certain extent. It is concluded that we should choose different prediction models according to the difficulty of data collection and the characteristics of the data itself; A higher prediction accuracy can be obtained by different aspects and methods comparison. We can compare construction land prediction methods from the aspects of essential characteristics of the model, the forms of the data, the difficulty in data collection and etc. This study provides a new idea for selecting construction land prediction methods and comparing the prediction methods. Otherwise, it could offer the reference for land use planning and sustainable land use.

Key words construction land prediction; multiple linear regression model; GM(1,1) Model; BP Neural Network Model; methods comparison