

几种方法在粮食总产量预测中的对比

姚作芳^{1,2}, 刘兴士¹, 杨飞³, 闫敏华¹

(1. 中国科学院东北地理与农业生态研究所, 吉林 长春 130012; 2. 中国科学院研究生院, 北京 100049;
3. 中国科学院地理科学与资源研究所, 北京 100101)

摘 要: 根据吉林省 1978~2007 年的农业数据, 选定了 14 个影响农业生产的因素作为研究对象, 分别采用了主成分分析法、BP 神经网络法、灰色预测法和逐步回归分析法 4 种分析预测方法, 通过 SPSS 和 Matlab 工具将原始数据进行处理, 得到 4 种不同的预测模型, 进而基于这 4 种模型对吉林省的粮食产量进行预测, 并将各种预测产量和实际产量进行拟合分析。研究结果表明, 拟合性最好的是 BP 神经网络法, 其拟合确定性系数为 0.899; 其次是主成分分析法(拟合确定性系数为 0.834)和逐步回归法(拟合确定性系数为 0.787); 拟合效果最差的是灰色预测法(确定性系数为 0.744)。粮食总产量估算精度最高的是 BP 神经网络法, 达到 93.67%; 其次是主成分分析法, 为 90.45%。

关键词: 主成分分析法; BP 神经网络法; 灰色预测法; 逐步回归法; 粮食产量; 预测

中图分类号: S11*7 **文献标识码:** A **文章编号:** 1000-7601(2010)04-0264-05

粮食是人们维持生存的必需品, 是国家经济发展和社会稳定的基础, 是整个国家安全系统的一个重要部分。中国是一个处于经济转型阶段的人口大国, 粮食安全更不能出现任何闪失, 因为粮食生产的波动必然会引发整个国民经济乃至金融界的波动^[1]。人们在千方百计提高粮食产量的同时, 也希望预知未来一段时间粮食产量的变化情况, 以便为科学决策提供依据, 因此研究粮食产量的预测是非常必要的。吉林省是我国重要粮食生产大省, 吉林省西部约占全国 4% 的耕地每年为国家提供了 10% 的定购粮, 20% 的专储粮, 50% 的出口玉米, 因此准确预测吉林省粮食产量更具有重要战略现实意义。

粮食产量的预测是农业系统的一项重要工作。长期以来, 许多学者在这方面做了大量研究, 形成了多种产量预测方法, 提高粮食产量预测精度一直是研究工作的明确目标和重要方向。由于作物产量的形成是一个复杂的过程, 受气候、土壤、生物以及社会经济等多种因子的影响^[2], 而各种产量预测方法的原理和出发点是不同的, 预测过程中考虑的影响因子侧重点是不一样的; 不同的预测方法往往只能提供某一方面的有用信息, 因而预测的精度是不一样的。另外目前的预测方法大多具有区域差异, 因此针对具体研究区域, 将多种粮食产量预测方法进行对比分析, 选择较优预测方法是很有必要的。本文以吉林省作为研究区域, 基于 1978 年到 2007 年

的吉林省各类农业生产数据, 分别比较计算了主成分分析法、BP 神经网络法、灰色预测法和逐步回归分析 4 种预测分析方法的估算精度, 并将各类预测数据与实际产量进行了拟合, 比较分析了各种方法的拟合效果, 探索较为精确的粮食产量预测模型, 进而对未来粮食生产进行科学预测, 为粮食生产规划、粮食储运、人口控制、合理开发利用农业自然资源以及环境保护等相关政策制定提供科学支持。

1 研究对象与方法

1.1 研究对象

本文以吉林省粮食产量为研究对象, 收集了吉林省 1978~2007 年的农业生产数据。

1.2 研究方法

1.2.1 指标数据的收集 本文选取以下 14 个指标进行分析, 为了文中计算公式标注的方便, 将各种指标进行了编号: Y 为吉林省粮食的年产量; X_1 为乡村劳动力; X_2 为农业机械总动力; X_3 为农村总用电量; X_4 为氮肥折纯量; X_5 为磷肥折纯量; X_6 为钾肥折纯量; X_7 为复合肥折纯量; X_8 为农作物受灾面积; X_9 为农作物成灾面积; X_{10} 为有效灌溉面积; X_{11} 为农村居民家庭人均年纯收入; X_{12} 为农村居民家庭人均年生活消费支出; X_{13} 为单位面积(667m²)农地农药的投入费用。在本研究中 Y 作为自变量, X 都为因变量。

1.2.2 统计计算 依据上述的各类指标数据, 计算

收稿日期: 2009-10-15

基金项目: 中国科学院知识创新工程重大项目(KSCX1-YW-09-13); 农业气候资源评价与高效利用技术研究专项(GYHY200706030)

作者简介: 姚作芳(1982—), 女, 湖南人, 博士生, 研究方向为区域农业与可持续发展研究。E-mail: yaozuofang@163.com。

通讯作者: 刘兴士, 院士, 博士生导师, 主要从事区域农业可持续发展与湿地生态研究。

建立各种预测模型所需的相关统计参数。

1.2.3 建立预测模型 依据相关统计参数,建立各种预测模型:主成分分析预测模型、BP神经网络预测模型、灰色预测模型和逐步回归预测模型。

1.2.4 对粮食产量进行预测 依据4种预测模型计算其产量预测值。

1.2.5 检验得到的预测产量 检验预测值与实际值有无差异性。本文采用2种方法来检验估算结果。一是分析计算各种预测值域实际值的拟合系数,二是计算预测精度:

$$\text{预测精度} = 1 - |(\text{估算值} - \text{实际值})| / \text{实际值} \times 100\% \quad (1)$$

1.2.6 比较分析 比较分析4种预测模型拟合系数和估算精度,依据实际工作需要,从中优选最佳预测模型。

2 结果与分析

2.1 粮食总产量的主成分分析方法估算

主成分分析法就是设法将原来指标重新组合成一组新的互相无关的几个综合指标来代替原来指标。同时根据实际需要从中可取几个较少的综合指标,尽可能多地反映原来指标的信息^[3-5]。

本研究首先对14个指标的30年数据进行标准化处理,以消除不同变量的量纲差异的影响,然后进行主成分变换,通过对14个变量偏相关系数的检验,得知KMO(Kaiser-Meyer-Olkin)检验值为0.805,根据Kaiser度量标准可知,原有14个变量适合进行因子分析。主成分变换之后,前两个主成分分量保留了原14个变量的82.378%的信息,信息量丢失较少,基本可以代替原来的14个变量。其中农作物受灾面积、农作物成灾面积、有效灌溉面积3个指标对第二个主成分贡献较大,而这3个指标均与作物种植面积有关,因此将这个因子命名为作物面积因子,其他11个指标对第一个主成分贡献较大,而其他几个主成分均与粮食生产投入程度有关,因此将第一个主成分命名为生产投入程度因子。

在30年的数据中,随机选取20年的数据用于建立粮食产量的主成分分析估算模型,剩余10年的数据进行模型估算效果检验。利用前3个主成分得到粮食总产量估算模型为:

$$Y = 503.186 \times f_1 - 50.416 \times f_2 + 1800.602 \quad (2)$$

式中, f_1 、 f_2 分别为生产投入程度因子、作物面积因子。利用上述模型估算后10年的粮食预测产量,与实际产量拟合结果如图1所示。由图1可看出,预测值和实际值具有较好的一致性,拟合系数 R^2 达0.

834。后10年粮食产量的平均估算精度为90.45%。

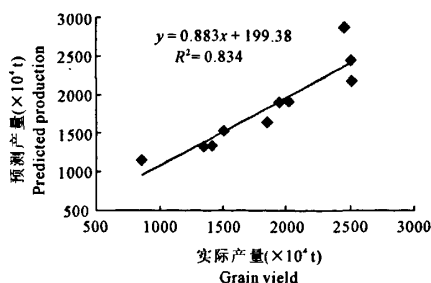


图1 主成分分析法估算的产量与实际产量的对比

Fig.1 The comparison of grain yield and predicted production by principal component analysis

2.2 粮食总产量的BP神经网络方法估算

BP神经网络是一种应用最为广泛的前向网络。BP网络的模型为前向多层网络,由输入层、输出层和若干个隐含层组成。每层都包含若干个神经元,同一层各神经元之间没有相互的连接,相邻层的神经元之间通过权连接。最基本的BP神经网络是三层的前馈网络,分别为输入层、隐藏层、输出层。输入层中有14个输入变量;隐藏层和输出层的传输函数分别为'tansig'和'purelin',输出层为粮食总产量,训练函数采用trainlm,该函数由Levenberg-Marquardt开发,具有运算高效、收敛快的特点^[6,7]。网络目标误差为0.0001,训练迭代次数为1000次,通过多次试验证明,隐藏层神经元个数为12时可以得到较好的粮食总产量估算效果。

在本文30a的数据中,随机选取20年数据用于训练BP神经网络,剩余10a数据用于验证网络训练效果,结果如图2。神经网络估算粮食总产量值和实际产量之间拟合确定性系数 R^2 达0.899,实际产量和预测值一致性较好。后10年粮食总产量的平均估算精度为93.67%。

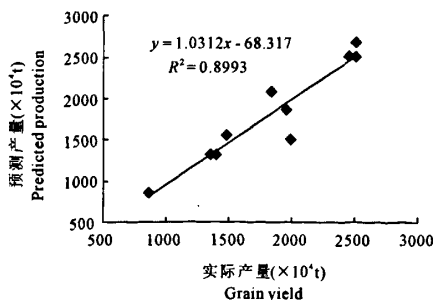


图2 BP神经网络法估算的产量与实际产量的对比

Fig.2 The comparison of grain yield and predicted production by neural network method

2.3 粮食总产量的灰色预测方法估算

灰色 GM(1,1)模型是邓聚龙教授于 20 世纪 80 年代提出来的,是单序列一阶线性动态模型,是一种计算简单、适用性广的预测模型,是用于控制和预测的新理论、新技术,目前已广泛地应用于农业和社会经济等领域^[8-10]。灰色预测的步骤如下:

① 给定原始时间序列,记为: $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$;

② 对原始数据列作一次累加生成 AGO(Accumulated generating operation)为: $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$; 其中: $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) = x^{(1)}(k-1) + x^{(0)}(k), k = 2, 3, \dots, n$

③ 构造累加矩阵 B 和常数项向量 Z :

$$B = \begin{pmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{pmatrix}, Z = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix}$$

建立相应的微分方程模型为:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu$$

式中, a 为发展灰数, μ 为内生控制灰数,按照所确定微分方程的特点,将所建模型称为 GM(1,1) 模型。

④ 根据最小二乘法得到如下公式:

$$\hat{a} = (B^T B)^{-1} B^T Y_n, \text{ 且 } \hat{a} = \begin{pmatrix} a \\ \mu \end{pmatrix}$$

⑤ 求解微分方程,即可得到预测模型:

$$Y_{(k+1)} = (x^{(0)}(1) - \frac{\mu}{a})e^{-ak} + \frac{\mu}{a}$$

式中, $Y_{(k+1)}$ 为第 $(k+1)$ 年粮食的累加值; $x^{(0)}(1)$ 为初始年的粮食产量; k 为年限,取 1、2、3... N 。

根据吉林省 30 年的粮食产量数据,按照时间顺序,每 3 年为一个单元,取每个单元的前 2 年用于模型计算,通过 Matlab 及 SPSS 计算软件得出吉林省粮食产量的灰色预测模型是:

$$Y_{(k+1)} = 23650.54 \times e^{0.046k} - 22736.04 \quad (3)$$

利用上述模型估算剩余 10 年的粮食产量,得到计算结果如图 3 所示。由图 3 可看出,实际产量和预测值的一致性一般,拟合系数 R^2 为 0.7439。剩余 10 年粮食产量的平均估算精度为 80.4%。

2.4 粮食总产量的逐步回归分析方法估算

逐步回归是以多元回归为基础的,利用最小二乘法原理,建立正规方程,求解回归系数的一种回归

方法。其特点为向前选择变量法和向后剔除变量法的结合,每一步只选取一个变量,从而使得到的回归方程更加简化,变量相关性更大。它是研究变量之间关系的一种统计方法,利用它可以给出预测方程,根据一个或几个自变量的值预测因变量的值,并且可以确定这种预测的精度^[11,12],逐步回归的线性方程表达式如下:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

式中, x_p 为自变量(影响粮食的各种因素), Y 为因变量(粮食产量), b_0 为常数项, b_p 为回归系数,也称为偏回归系数,它表示在其他所有自变量不变的情况下,自变量 x_p 每变化一个单位,引起因变量 Y 平均变化的数值。

本文选取粮食产量为因变量,14 种影响因子为自变量,随机选取 20 年数据用于逐步回归计算,通过 SPSS 软件计算得出吉林省粮食产量的线性方程为:

$$Y = 1323.07 + 316.93x_1 + 729.38x_2 - 301.72x_3 + 272.28x_4 - 81.58x_5 - 143.85x_6 - 109.26x_7 + 146.93x_8 - 237.33x_9 - 2853.773x_{10} + 1244.418x_{11} - 1145.399x_{12} + 221.98x_{13} \quad (4)$$

剩余 10 年数据用来验证预测效果,得出预测的产量与实际产量的对比图(图 4)。

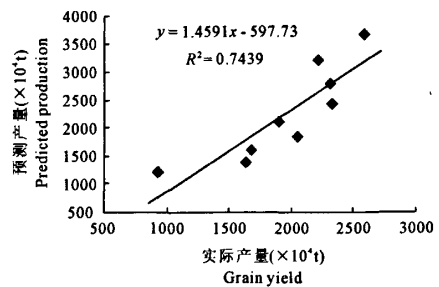


图 3 灰色预测法估算的产量与实际产量的对比

Fig.3 The comparison of grain yield and predicted production by grey prediction model GM(1,1)

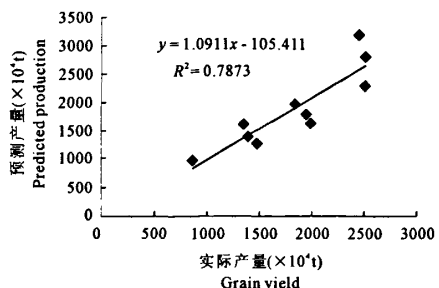


图 4 逐步回归分析法估算的产量与实际产量的对比

Fig.4 The comparison of grain yield and predicted production by stepwise regression analysis

由图 4 可看出,实际产量和预测值的一致性一般,拟合确定性系数 R^2 为 0.7873。剩余 10 年粮食产量的平均估算精度为 86.88%。

2.5 不同预测方法的对比分析

通过表 1,可以发现 BP 神经网络预测值的精度最高,拟合的一致性最好,其次是主成分分析法。这两种模型拟合系数都在 0.83 以上,预测精度都在 90% 以上。在利用这两种模型进行预测之前都将原始数据进行了标准化处理,以消除不同变量的量纲差异的影响。

从预测模型所考虑的影响因子的数目来看:灰色预测要求的数据最少,只需要知道粮食产量的历年数据就行,不需要其他的农业生产数据,实践中可以用该方法来预测粮食产量长期的变化趋势,但不能用于预测其精确值。而主成分分析法、BP 神经网络法和逐步回归法在进行建模时由于考虑了粮食产量的各种影响因子,大大提高了预测精度。通过对比分析发现,BP 神经网络法是对粮食产量进行预测的较为理想的预测方法。

表 1 4 种预测模型的对比

Table 1 The comparison of the four different predicting methods

名称 Name	拟合系数 R^2 Regression	估算精度 Estimation accuracy (%)	所考虑的影响因子个数 Affecting factors number
主成分分析法 Principal component analysis	0.8340	90.45	14
BP 神经网络法 BP neural network	0.8993	93.67	14
灰色预测法 Grey prediction model	0.7439	80.40	1
逐步回归分析法 Stepwise regression analysis	0.7873	86.88	14

3 结 论

本文通过主成分分析法、BP 神经网络法、灰色预测法和逐步回归 4 种预测方法,对吉林省 1978 ~ 2007 年农业生产数据进行分析,研究得出以下结论:

1) 在这 4 种预测模型中,BP 神经网络法的预测精度最高,预测值和实际产量的拟合性最好。BP 神经网络法允许原始的随机数据或数据中含有较多的噪声,这是它区别于其它模型的最大优势,因而任何能用传统的模型分析或统计方法解决的问题,BP 神经网络能处理得更好。在进行粮食产量预测时,

BP 神经网络法是一种非常理想的预测方法,但是在构造神经网络的预测模型时,要注意正确选择影响因素,不要漏掉对预测对象有重大影响的因素,也应舍去不影响预测精度的次要因素。

2) 在利用数学模型对粮食产量进行预测时,考虑的影响因素越多,预测的精度就越高,因而在进行未来粮食产量的预测实践中,应尽可能地考虑各种因素的影响。

3) 灰色模型可以较好地预测未来正常年的粮食产量,只要知道基于时间序列的粮食产量历史数据就能利用灰色模型预测未来正常年的粮食产量,可以用来进行长期预测;BP 神经网络法、主成分分析法、逐步回归预测由于考虑的因素较多,预测精度也较高,是短期预测比较理想的方法,但不适合用于长期预测。

4) 由于粮食产量受各种因素的影响,波动性较大,除了受到上述 14 种因素的影响外,在很大程度上还受国家宏观政策、作物品种、耕作技术等方面的影响,如何更全面地将难以量化的因素也纳入模型中进行考虑分析,从而不断地改进预测模型,提高预测精度,是需要进一步深入研究的工作。

参 考 文 献:

[1] 王介勇,刘彦随.1990 年至 2005 年中国粮食产量重心演进格局及其驱动机制[J].资源科学,2009,31(7):1188—1194.

[2] 何满喜,王桂霞.内蒙古人口耕地粮食系统分析与前景预测[J].干旱地区农业研究,2001,19(1):115—119.

[3] 李永振.主成分分析法在汶川地震预测中的应用[J].四川地震,2009,2(131):21—24.

[4] 杨江民,张凤太.基于主成分分析法的重庆直辖 10 年后农业发展动态研究[J].安徽农业科学,2009,37(19):9199—9201.

[5] 伊元荣,海米提·依米提.主成分分析法在城市河流水质评价中的应用[J].干旱区研究,2008,25(4):497—500.

[6] 周良臣,康绍忠,贾云茂.BP 神经网络方法在土壤墒情预测中的应用[J].干旱地区农业研究,2005,23(5):98—102.

[7] 赵胜利,刘 燕,李书全,等.用 BP 神经网络预测地下水动态[J].河北农业大学学报,2002,25(4):205—207.

[8] 李葆春,马 琦.灰色 GM(1:1)模型在定西县粮食产量预测中的应用[J].甘肃农业大学学报,2005,40(5):660—663.

[9] 谢恒星,张振华,谭春英.灰色预测方法在山东省粮食产量预测中的应用[J].水土保持研究,2006,13(2):257—258.

[10] 赖一飞,郑清秀,章少强,等.灰色预测模型在水运货运量预测中的应用[J].武汉水利电力大学学报,2007,33(1):96—99.

[11] 葛朝霞,薛 梅,宋颖玲.多因子逐步回归周期分析在中长期水文预报中的应用[J].河海大学学报(自然科学版),2009,37(3):255—257.

[12] 骆术斌,董 雯,徐 蕾,等.逐步回归法在黄河下游洪水水位预测中的应用[J].东北水利水电,2008,28(26):50—52.

Comparison of several methods in grain production prediction

YAO Zuo-fang^{1,2}, LIU Xing-tu¹, YANG Fei³, YAN Min-hua¹

(1. *Northeast Institute of Geography and Agro-ecology, Chinese Academy of Sciences, Changchun, Jilin 130012, China;*

2. *Graduate University of Chinese Academy of Sciences, Beijing 100039, China;*

3. *Institute of Geographic Science and Natural Resources, Chinese Academy of Sciences, Beijing 100101, China)*

Abstract: In this study, four predicting methods, namely principal component analysis, BP neural network method, GM(1,1) grey prediction model and regression analysis model, were used and compared for grain production prediction. Based on agricultural data from 1978 ~ 2007 of Jilin Province, fourteen affecting factors were chosen for analysis by SPSS and Matlab tools to build four different prediction models. And these models were also used to predict the grain production in Jilin Province, and comparison was made of the R^2 of each regression line and the prediction accuracy of different models. The studied results showed that BP neural network method performed the best with regression R^2 and estimation accuracy of 0.899 and 93.67%, principal component analysis method performed moderate grain production estimation results, with R^2 of 0.834 and the accuracy of 90.45%, the next is stepwise regression analysis, while grey prediction method was the poorest.

Keywords: principal component analysis; neural network method; GM(1,1) grey prediction model; regression analysis model; grain production; prediction

(上接第 263 页)

Analysis of climatic change and their impacts on agriculture of Atux City in Xinjiang

HU Jiang-ling, Mansuer·Shabiti, Nasiman·Nasiarding

(*College of Geographic Science and Tourism, Xinjiang Normal University, Urumqi, 830054, China*)

Abstract: Based on the statistics of temperature and precipitation from 1960 ~ 2006 in Atux meteorological station, this paper analyzed the annual and seasonal changes of temperature and precipitation in Atux City by employing the methods of regression analysis and trend line analysis. The results indicate that: (1) During the last 47 years, the temperature of the study area shows a trend of increasing, and the increasing rate was $0.13^{\circ}\text{C}/10\text{a}$. From the 1960s ~ 1970s, the temperature in the study area was undulating around the average value and meanwhile decreased slightly. However, from 1970s, it showed larger undulation and increased significantly and kept the same trend with that of the Northwest China. The change of temperature in a year showed seasonal disparities and it had a tendency of increasing in winter, autumn and spring, and decreasing in summer. Among them the temperature changes in winter and spring had more contribution to the annual average temperature change. (2) During the last 47 years, the precipitation of the study area had a trend of increasing, and the average increasing rate was $9.38\text{ mm}/10\text{a}$. This result kept the same trend with that of the Northwest China, too. From the beginning of 1960s to the end of 1970s, the precipitation in the study area was undulating around the average value and hadn't made any significant change. But from 1980s, it showed a larger undulation and the changing trend of wet and dry was distinct. The change of precipitation in a year showed significant seasonal disparities, in which summer had the largest increasing rate, followed by spring, autumn and winter. Among them the increase of the precipitation in summer and spring had more contribution to the annual average precipitation change, while it was smaller in autumn and winter. (3) The climatic warming in the study area is obviously good to the growth of the over winter crops and also to the growth of high temperature crops in its entire fertility period. But in the meanwhile, because the climatic warming would cause soil drought and increase of plant diseases and insect pests, it's not good to the growth of other kinds of crops.

Keywords: Atux City in Xinjiang; climatic change; agricultural production