

基于多元线性回归模型和灰色关联分析的 江苏省粮食产量预测

王春辉,周生路,吴绍华,吴滢滢

(南京大学地理与海洋科学学院,江苏 南京 210046)

[摘要] 针对江苏省近年来粮食产量波动的现状,采用灰色关联模型对 2000~2009 年江苏省粮食产量进行分析. 结果发现,播种面积、化肥施用量、年均气温、年均降水以及农药施用量等 5 个因素是粮食产量波动的主导因素. 然后在灰关联分析的基础上建立多元线性回归方程对 2010~2012 年江苏省粮食产量进行预测与验证,结果表明预测效果良好. 最后,针对现状提出了相关建议.

[关键词] 灰色关联分析,多元线性回归,粮食产量预测,江苏省

[中图分类号] F302.5 [文献标志码] A [文章编号] 1001-4616(2014)04-0105-05

Prediction on Grain Yield in Jiangsu Province Based on Multivariate Linear Regression Model and Gray Relational Analysis

Wang Chunhui, Zhou Shenglu, Wu Shaohua, Wu Yingying

(School of Geographic and Oceanographic Science, Nanjing University, Nanjing 210046, China)

Abstract: According to the situation of grain yield fluctuation in Jiangsu province in recent years, gray relational model analysis was used in the grain yield of Jiangsu province from 2000~2009. The results show that the area of seed sown, amount of chemical fertilizer, annual temperature, annual precipitation and amount of pesticide are the primary factors of grain yield fluctuation. Then multivariate linear regression equation was established on the basis of gray relational analysis to predict and verify on the grain yield in Jiangsu province from 2010~2012, and the result indicated that the predictive effect was excellent. Finally, some relevant suggestions are proposed on the basis of these problems.

Key words: gray relational analysis, multivariate linear regression, prediction on grain yield, Jiangsu province

“民以食为天,食以粮为本”,粮食是人类生存和发展的基础,也是一个国家最重要的战略物资,因此,粮食安全问题一直是人们关注的焦点^[1]. 对于当今世界来说,粮食安全是一个全球性问题,据统计,2009 年全球饥饿人口超过 10 亿,而发展中国家占到了 98.5%^[2]. 在我国,近年来由于种粮的经济效益低,农民种粮的积极性不断下降,大量的农民进入城镇从事第二、三产业,把种粮当成一种副业,而随着人口的增长,粮食刚性需求却呈现稳步增长态势,粮食供求关系趋向于不平衡,我国粮食安全问题的前景也不容乐观^[3,4]. 粮食安全问题的一个重要影响因素在于粮食产量的变化^[5],而产量的变化又受多种因素的影响,一般来讲,可分为自然影响因素和人为影响因素^[6,7],前者包括气温、降水、湿度、自然灾害等,后者则包括粮食作物播种面积、化肥施用量、农业现代化水平等. 在不同地区,这些影响因素又具有不同的权重值,如在一些农业高度发达国家,农业现代化水平是粮食产量的决定因素^[8],而在一些发展中国家尤其是非洲地区,粮食产量的变化依然是自然因素占据主导地位^[9].

江苏省素有中国“三大粮仓”之一的美称^[10],但随着近几十年来人地矛盾的突出,耕地资源日趋减少,导致粮食作物的播种面积受到限制,加上农业环境污染日趋严重,粮食产量出现了一定的波动,因而探

收稿日期:2014-02-20.

基金项目:江苏省国土资源厅科技计划项目(201110).

通讯联系人:周生路,教授,博士生导师,研究方向:土地资源与环境. E-mail:zhoussl@nju.edu.cn

讨其波动原因并预测未来走势对江苏乃至全国粮食安全战略政策的制定具有重大意义. 以往许多学者研究该问题时一般采用单一化模型, 试图找出粮食产量波动的影响因素来进行预测, 而事实上每种模型都具有一定的局限性, 如灰色系统模型中灰关联模型能比较好地找出事件的主要影响因素, 而灰色预测模型对波动较大的曲线进行拟合时精度就较低^[11], 所以在选择模型时应该扬长去短, 可结合多种模型来发挥每种模型的优点从而提高预测精度. 本文针对上述情况, 利用灰色关联分析模型对江苏省 2000~2009 年的粮食产量及影响因子进行分析, 找出粮食产量波动的主导因素, 然后在灰关联分析的基础上利用多元线性回归模型对江苏省 2010~2012 年的粮食产量进行预测与验证, 比较实际值与预测值的大小, 确定回归方程的实用性, 以为江苏省粮食生产战略提供一定的参考.

1 研究方法

1.1 灰色关联分析

灰色关联分析是一种用灰色关联度顺序来描述因素间关系的强弱、大小、次序的系统理论分析方法^[12]. 该分析方法的基本思路是两个变量之间的相似性或一致性越高, 它们之间的关联度也会越高, 反之它们之间的关联度就会越低. 该分析方法具有样本需求量少, 数据分布规律要求低, 精准度高, 工作量少, 定量分析与定性分析结果高度一致等优点^[13].

灰色关联分析的基本原理是:

(1) 确定所要分析数据序列, 并设置一母序列 $x_o(t)$ 和若干子序列 $x_i(t)$, 构成矩阵:

$$Z = \begin{pmatrix} x_o(1) & \cdots & x_o(m) \\ \vdots & \cdots & \vdots \\ x_n(1) & \cdots & x_n(m) \end{pmatrix} \quad (1)$$

式中, 母序列 $x_o(t)$ 表示为分析样本的因变量, 子序列 $x_i(t)$ 表示为分析样本的自变量.

(2) 对矩阵中的数据进行无量纲化处理. 由于母序列和各子序列的单位不同, 所表示的数据含义也不同, 而无量纲化则解决了这一问题, 增强序列之间的可比性, 形成新的母序列 $y_o(k)$ 和新的若干子序列 $y_i(k)$, 计算公式为:

$$y_o(k)' = \frac{y_o(k)}{y_o(1)}, \quad y_i(k)' = \frac{y_i(k)}{y_i(1)}. \quad (2)$$

(3) 计算母序列和子序列之间的绝对差值序列. 计算公式为:

$$\Delta_{oi}(k) = |y_o(k) - y_i(k)|, \quad i = 1, 2, \cdots, n; k = 1, 2, \cdots, m. \quad (3)$$

(4) 计算母序列和各子序列的关联系数. 计算公式为:

$$\xi_{oi}(t) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_{oi}(k) + \rho \Delta_{\max}}, \quad (4)$$

式中, Δ_{\min} 和 Δ_{\max} 表示式(2)中最小值和最大值, $\rho \in (0, 1)$ 为分辨系数, 其取值不影响母序列和子序列的关联度分析, 本文取 0.5.

(5) 计算关联度, 计算公式为:

$$\gamma_{oi} = \frac{1}{n} \sum_{t=1}^n \xi_{oi}(t). \quad (5)$$

最后得到的结果就是母序列与各子序列之间的关联度. 通过分析关联度的大小, 便可以确认母序列与各子序列的拟合度.

1.2 多元线性回归模型

运用多元线性回归预测的基本思路是在确定因变量和多个自变量以及它们之间的关系后, 通过设定自变量参数的回归方程对因变量进行预测. 具体如下:

$$Y = C + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n, \quad (6)$$

式中: Y 表示为粮食总产量, C 和 a 为回归系数, 是待定参数, X 为所选取的影响因素. 多元线性回归方法可分为强行进入法、消去法、向前选择法、向后剔除法、逐步进入法等^[14], 本文运用 SPSS20.0 软件, 对选择的自变量全部进入回归模型, 即强行进入法进行预测. 该模型的优点是方法简单、预测速度快、外推性好等.

2 结果与分析

2.1 粮食产量与影响因素灰关联计算

江苏省粮食产量的变化与多个影响因子相关,本文选取了 10 个对粮食产量相关的影响因子,通过对这 10 个影响因子的灰关联计算,来最终确定影响江苏省粮食产量变化的主导因素,以此为多元回归模型设定参数来进行粮食产量的预测. 首先设定随年份变化的母序列 X_0 表示粮食产量,然后对余下的影响因子分组, X_1 、 X_2 、 X_3 分别代表平均气温、平均降水和受灾面积,分组到自然影响因素; X_4 、 X_5 、 X_6 、 X_7 、 X_8 、 X_9 、 X_{10} 分别代表播种面积、化肥施用量、农药施用量、农业机械总动力、农业劳动力、农村用电量以及农用塑料薄膜,分组到人为影响因素. 通过分组自然因素及人为因素来最终确认哪组因素对粮食生产的影响较大对研究粮食产量变化具有重要意义. 下面对粮食产量与其影响因素进行灰关联计算,首先设置一母序列及若干子序列(表 1). 由表 1 通过式(2)无量纲化处理,得到初始值新矩阵(表 2). 然后对无量纲化处理后的初始值数据,依据式(3)求绝对差值序列,结果见表 3. 接着建立绝对差值矩阵后,找出矩阵中的最大差值和最小差值, $\Delta_{\min}=0$, $\Delta_{\max}=3.1453$,然后将各值代入式(4),得到关联系数矩阵(表 4). 最后,根据式(5),得到关联度值 γ_{oi} (图 1).

表 1 2000~2009 年江苏省粮食产量与其影响因子

Table 1 Grain yield of Jiangsu province and the influencing factor from 2000~2009

年份	X_0	自然影响因素			人为影响因素						
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
2000	3 106.63	15.89	1 080.74	3 411.68	5 304.31	335.45	9.15	2 925.29	1 890.96	314.6	6.51
2001	2 942.05	15.91	1 070.00	2 283.24	4 886.66	338	9.16	2 957.93	1 832.25	345.05	6.61
2002	2 907.05	16.24	963.52	2 288.46	4 882.58	337.53	8.64	2 983.89	1 744.41	423.61	6.88
2003	2 471.85	15.47	1 226.19	3 305.00	4 659.47	334.67	8.79	3 029.10	1 615.49	529.5	6.97
2004	2 829.06	16.30	852.00	836.58	4 774.59	336.80	9.23	3 052.51	1 506.31	679.83	7.04
2005	2 834.59	15.70	1 040.40	2 139.50	4 909.48	340.81	10.33	3 135.33	1 414.83	825.1	7.2
2006	3 096.03	16.40	1 043.30	1 324.07	5 110.8	342.01	9.86	3 278.53	1 323.88	1 011.79	7.51
2007	3 132.24	16.70	1 065.00	1 642.00	5 215.59	342.03	9.68	3 392.44	1 230.28	1 159.03	8.04
2008	3 175.49	15.70	1 036.90	4 83.70	5 267.1	340.76	9.38	3 630.86	1 179.94	1 234.14	8.54
2009	3 230.10	15.90	1 123.50	1 001.63	5 272.04	344.00	9.23	3 810.57	1 120.19	1 316.62	9.43

注:数据来源于 2000~2009 年《江苏统计年鉴》.

表 2 初始值矩阵

Table 2 Initial value matrix

X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	1	1	1	1	1	1	1	1	1	1
0.947 0	1.001 3	0.990 1	0.669 2	0.921 3	1.007 6	1.001 1	1.011 2	0.969 0	1.096 8	1.015 4
0.935 8	1.022 0	0.891 5	0.670 8	0.920 5	1.006 2	0.944 3	1.020 0	0.922 5	1.346 5	1.056 8
0.795 7	0.973 6	1.134 6	0.968 7	0.878 4	0.997 7	0.960 7	1.035 5	0.854 3	1.683 1	1.070 7
0.910 7	1.025 8	0.788 3	0.245 2	0.900 1	1.004 0	1.008 7	1.043 5	0.796 6	2.160 9	1.081 4
0.912 4	0.988 0	0.962 7	0.627 1	0.925 6	1.016 0	1.129 0	1.071 8	0.748 2	2.622 7	1.106 0
0.996 6	1.032 1	0.965 4	0.388 1	0.963 5	1.019 6	1.077 6	1.120 8	0.700 1	3.216 1	1.153 6
1.008 2	1.051 0	0.985 4	0.481 3	0.983 3	1.019 6	1.057 9	1.159 7	0.650 6	3.684 1	1.235 0
1.022 2	0.988 0	0.959 4	0.141 8	0.993 0	1.015 8	1.025 1	1.241 2	0.624 0	3.922 9	1.311 8
1.039 7	1.000 6	1.039 6	0.293 6	0.993 9	1.025 5	1.008 7	1.302 6	0.592 4	4.185 1	1.448 5

表 3 绝对差值矩阵

Table 3 Absolute differences matrix

Δ_{o1}	Δ_{o2}	Δ_{o3}	Δ_{o4}	Δ_{o5}	Δ_{o6}	Δ_{o7}	Δ_{o8}	Δ_{o9}	Δ_{o10}
0	0	0	0	0	0	0	0	0	0
0.054 2	0.043 0	0.277 8	0.025 8	0.060 6	0.054 1	0.064 1	0.021 9	0.149 8	0.068 3
0.086 3	0.044 2	0.265 0	0.015 3	0.070 4	0.008 5	0.084 3	0.013 3	0.410 7	0.121 1
0.177 9	0.338 9	0.173 1	0.082 8	0.202 0	0.165 0	0.239 8	0.058 7	0.887 4	0.275 0
0.115 2	0.122 3	0.665 4	0.010 5	0.093 4	0.098 1	0.132 8	0.114 1	1.250 3	0.170 8
0.075 6	0.050 2	0.285 3	0.013 1	0.103 5	0.216 5	0.159 4	0.164 2	1.710 3	0.193 6
0.035 5	0.031 2	0.608 5	0.033 1	0.023 0	0.081 0	0.124 2	0.296 5	2.219 5	0.157 0
0.042 7	0.022 8	0.527 0	0.025 0	0.011 4	0.049 7	0.151 4	0.357 6	2.675 9	0.226 8
0.034 1	0.062 7	0.880 4	0.029 2	0.006 3	0.003 0	0.219 0	0.398 2	2.900 7	0.289 7
0.039 1	0.000 2	0.746 2	0.045 8	0.014 3	0.031 0	0.262 9	0.447 4	3.145 3	0.408 8

表 4 关联系数矩阵
Table 4 Correlation coefficient matrix

ξ_{o1}	ξ_{o2}	ξ_{o3}	ξ_{o4}	ξ_{o5}	ξ_{o6}	ξ_{o7}	ξ_{o8}	ξ_{o9}	ξ_{o10}
1	1	1	1	1	1	1	1	1	1
0.966 7	0.973 4	0.849 9	0.983 9	0.962 9	0.966 8	0.960 8	0.986 2	0.913 0	0.958 4
0.948 0	0.972 7	0.855 8	0.990 4	0.957 1	0.994 6	0.949 1	0.991 6	0.792 9	0.928 5
0.898 4	0.822 7	0.900 9	0.950 0	0.886 2	0.905 1	0.867 7	0.964 0	0.639 3	0.851 2
0.931 8	0.927 8	0.702 7	0.993 4	0.944 0	0.941 3	0.922 1	0.932 4	0.557 1	0.902 1
0.954 1	0.969 0	0.846 4	0.991 7	0.938 2	0.879 0	0.908 0	0.905 4	0.479 0	0.890 4
0.977 9	0.980 5	0.721 0	0.979 4	0.985 6	0.951 0	0.926 8	0.841 4	0.414 7	0.909 2
0.973 5	0.985 7	0.749 0	0.984 4	0.992 8	0.969 4	0.912 2	0.814 7	0.370 2	0.874 0
0.978 8	0.961 6	0.641 1	0.981 8	0.996 0	0.998 1	0.877 8	0.798 0	0.351 6	0.844 5
0.975 7	0.999 9	0.678 2	0.971 7	0.991 0	0.980 7	0.856 8	0.778 5	0.333 3	0.793 7

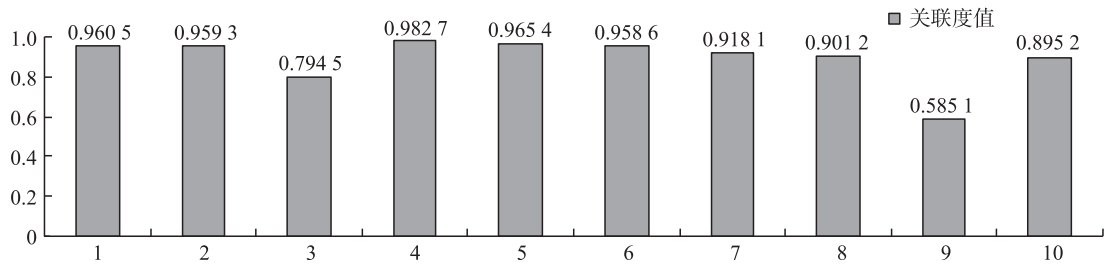


图 1 江苏省粮食产量与影响因子的关联度值
Fig.1 Degree values of grain yield of Jiangsu province and the influencing factor

通过对江苏省粮食产量影响因子的灰关联计算可知,其关联度大小依次是播种面积>化肥施用量>平均气温>平均降水>农药施用量>农业机械总动力>农业劳动力>农用塑料薄膜>受灾面积>农村用电量.可见,2000~2009 年影响江苏省粮食产量的主要因素为人为因素,其中播种面积和化肥施用量起到了主导作用,同时,自然影响因素中的气温和降水对粮食产量的影响也不可忽视,因此,本文针对上述分析,根据相关文献[15],并在尽量减少影响参数的原则下,选取播种面积、化肥施用量、平均气温、平均降水及农药施用量等 5 个因子作为多元线性回归模型中的自变量参数来对江苏省粮食产量进行预测.

2.2 建立粮食产量的多元线性回归模型

在江苏省粮食产量影响因子灰关联计算的基础上,利用多元线性回归模型对粮食产量进行预测,选取 5 个主导因素作为模型参数,利用统计软件 SPSS20.0 中的强行进入法得到回归方程:

$$y=-2\ 897.788+0.946x_1+7.529x_2-7.724x_3-0.680x_4-63.759x_5,$$
 (7)

其中 x_1 、 x_2 、 x_3 、 x_4 、 x_5 分别代表播种面积、化肥施用量、平均气温、平均降水和农药施用量.进行 F 检验,通过计算,可得 $F=25.562$,选择显著性差异 $\alpha=0.01$,根据 $F_\alpha(5,4)$ 查 F 分布显著性检验值 $F_\alpha=15.52$, $F>F_\alpha$. 综上,在 $\alpha=0.01$ 的水平下,回归效果显著.

2.3 粮食产量回归方程的检验

通过建立的粮食产量影响因子的回归模型(7),对江苏省 2010~2012 年的粮食产量的预测值与实际值进行对比与验证,验证结果见表 5.从表 5 可以看出,实际值与预测值相对误差分别只有 0.8%、0.1%和-2.43%,可见预测效果是良好的.

表 5 粮食产量拟合度分析表
Table 5 Analysis table of fitting degree between predicted and actual grain yield

年份	实际值/万 t	预测值/万 t	相对误差/%
2010	3 235.1	3 261.1	0.8
2011	3 307.8	3 311.1	0.1
2012	3 372.5	3 290.4	-2.43

3 结论与建议

通过对江苏省 2000~2009 年粮食产量与影响因子的灰关联计算分析发现,粮食作物播种面积、化肥施用量、平均气温、平均降水及农药施用量是造成粮食产量波动的主导因素;通过构建多元线性回归方程对 2010~2012 年江苏省粮食产量进行验证,发现回归方程误差值较低,预测效果良好.

基于江苏省粮食产量的波动影响因素及预测分析,就促进江苏省粮食产量的稳定性及可持续发展提

出以下建议:

(1)稳定粮食作物的种植面积,提高单产水平.江苏省是我国的一个经济大省,近年来经济发展水平始终位居我国前列,并且随着城市化水平的不断提高,人地关系矛盾日益突出,耕地面积正在不断的减少,使得耕地资源越来越紧张,导致粮食作物种植面积面临挑战.通过灰色关联模型计算发现,种植面积是影响江苏省粮食产量的最关键因素,由此稳定其种植面积是当务之急,而提高粮食作物的单产水平则是在有限耕地资源的前提下提高产量的重要途径.

(2)协调好农业环境与化肥和农药施用量之间的关系,发展绿色农业.江苏省粮食产量的波动在很大程度上受到了化肥施用量的影响,而化肥尤其是氮肥的过量施用可导致农田硝态氮过剩,而硝态氮极易随水迁移,将成为深层土壤和地下水的潜在污染源;另外农药的大量使用,又会造成农田土壤有机物污染加剧.因此,粮食产量的提升不应该建立在环境污染的基础上,而应该走“环境友好型”的绿色农业.

(3)继续提高农业机械化水平.农业机械化水平在一些发达国家对粮食产量的影响已经占据了主导作用,也是衡量一个地区农业现代化水平的关键指标.通过对江苏省粮食产量波动的灰关联分析表明,农业机械化水平对江苏省粮食产量波动虽然也产生了较大的影响,但跟一些发达国家的农业机械化水平相比仍存在较大的差距,因此有待进一步的提高.

[参考文献]

- [1] Godfray H C J, Beddington J R, Crute L R, et al. Food security: the challenge of feeding 9 billion people[J]. Science, 2010, 327(5 967): 812-818.
- [2] 公茂刚. 发展中国家粮食安全研究[D]. 长春: 东北师范大学经济学院, 2010.
- [3] 宋承国. 世界粮食危机与中国粮食安全[J]. 当代经济研究, 2009(2): 53-56.
- [4] Zhou Z. Achieving food security in China: past three decades and beyond[J]. China Agricultural Economic Review, 2010, 2(3): 251-275.
- [5] 何蒲明, 黎东升. 基于粮食安全的粮食产量和价格波动实证研究[J]. 农业技术经济, 2009(2): 85-92.
- [6] 周文魁. 气候变化对中国粮食生产的影响及应对策略[D]. 南京: 南京农业大学经济管理学院, 2012.
- [7] 顾月民. 基于最小一乘准则的中国粮食产量与影响因素的相关性分析[J]. 农业工程学报, 2013, 29(11): 1-10.
- [8] Hornbeck R. Nature versus Nurture: the environment's persistent influence through the modernization of American agriculture[J]. American Economic Review, 2012, 102(3): 245-249.
- [9] 杨艳昭, 吴艳娟, 封志明. 非洲粮食产量波动时空格局的定量化研究[J]. 资源科学, 2014, 36(2): 361-369.
- [10] 谭少华, 倪绍祥. 江苏省粮食安全问题及其对策[J]. 长江流域资源与环境, 2002, 11(4): 305-309.
- [11] 巫锡柱, 宴路明. 粮食生产灰色关联动态分析和预测[J]. 农业系统科学与综合研究, 2005, 11(4): 244-251.
- [12] 邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 2009.
- [13] 成晋松, 吕慧进, 刘玲. 太原市用水量影响因素的灰色关联分析[J]. 水资源与水工程学报, 2012, 23(2): 109-111, 115.
- [14] 高辉. 基于多个回归方程拟合的数据挖掘方法研究与设计[D]. 长春: 东北师范大学软件学院, 2012.
- [15] 陈祺琪, 李君, 梁保松. 基于灰色多元线性回归的河南省粮食产量预测[J]. 河南农业大学学报, 2012, 46(4): 448-452.

[责任编辑: 丁 蓉]