

文章编号:1673-8411 (2016) 03-0098-04

基于 SPSS 的梧州早稻产量预测模型构建

徐芳, 黄帆

(梧州市气象局, 梧州 543002)

摘要:应用界面友好的 SPSS 统计软件及 Microsoft Excel 软件作为建模工具,利用 1983–2014 年梧州市气象局早稻生长发育关键期气象观测资料和市统计局提供的早稻单产资料,采用直线滑动均值和调和权重法模拟外推趋势产量,再通过逐步回归建立气象产量预报方程,得到产量预报值,建立梧州市早稻产量的定量预测模型,通过了 5% 的信度检验,并且 2015 年的预报准确率为 98.6%。

关键词:SPSS; 预测模型; 早稻产量; 梧州

中图分类号:P49

文献标识码:A

Building of early season rice yield prediction model of Wuzhou by SPSS

Xu Fang, Huang Fan

(Wuzhou Municipal Meteorological Service, Wuzhou 543002)

Abstract: Based on the meteorological observation of the early rice growing stage data from 1983–2014 and yield data provided by municipal of statistical bureau, the output of trend extrapolation was obtained by the linear sliding mean and mediate weight method, and meteorological yield forecast equation was established by stepwise regression to get the volume forecast value, then early rice yield quantitative prediction model in Wuzhou was established, which passed the reliability test of 5%, and get the forecast accuracy of 98.6% in 2015.

Key Words: SPSS; predicting model; early season rice yield; Wuzhou

SPSS^[1] (Statistical Product and Service Solutions, 统计产品和服务解决方案) 是当今国际上最新、最流行、最受欢迎的统计软件之一。它最突出的特点是图形菜单驱动界面极为友好, 输出结果美观。SPSS 具有完善的数据输入、编辑、统计分析、报表、图形制作等功能。该软件自带 12 种类型 177 个函数, 基本能够满足气象统计的需要。此外, SPSS 采用类似 EXCEL 表格的方式输入与管理数据, 数据接口较为通用, 能方便地从其他数据库中读入数据。本文以梧州市早稻单产预测为例, 构建早稻产量预

测模型, 以期为提高梧州市早稻产量预报准确率提供一种有效的方法。

1 资料来源及处理方法

1.1 资料来源

利用 1983–2014 年梧州市气象局所属的梧州、蒙山、藤县、苍梧、岑溪等 5 个地面站气象观测资料和市统计局提供的梧州市早稻单产资料。

1.2 处理方法

采用直线滑动均值和调和权重法模拟趋势产

收稿日期: 2016-05-16

基金项目: 由广西梧州市科学研究与技术开发计划项目(2014-30)资助。

作者简介: 徐芳(1967-), 女, 高级工程师, 现从事农业气象服务。

量,再通过逐步回归建立气象产量预报模型,依据作物产量预报模型得到产量预报值。

建模过程的基本方法是^[2]: 首先将梧州市 1983—2014 年早稻单产资料进行处理。去掉产量的时间趋势项,分离出气象产量。其次将气象产量作为因变量、早稻主要生育期(2 月下旬—6 月中旬)气象因子作为自变量,利用逐步回归分析方法,筛选出对早稻产量贡献大的因子建立回归方程。最后在上述分析基础上建立早稻产量预测模型。

2 早稻产量预报模式

制作作物产量预报^[3],要将早稻产量分解为由生产力水平等因素形成的趋势产量和气象因子影响对应的气象产量两部分。其产量预测模式为: $Y=Y_t+Y_m(X_1, X_2, \dots, X_K)+Y_\varepsilon$, 其中 Y, Y_t, Y_m 和 Y_ε 分别表示预测单产、趋势单产、气象单产和随机干扰项, X_1, X_2, \dots, X_K 表示作物生长期气象因子,如降水量、温度、日照时数等。由于随机干扰项 y_ε 构成的因素具有不确定性。通常只有通过历史资料的分析,进行社会调查研究之后,或对产量数据进行订正,或舍弃不用。

2.1 趋势产量模拟

要研究作物产量与气象条件之间关系^[4],必须从实际产量中将生产力水平因素去掉,剩下由气象因子影响的那部分产量与气象要素进行相关分析,才合理。对于产量时间趋势项的模拟方法有多种^[5]^[6],如滑动平均法、重心描绘法、指数法、直线法、正交多项式法、分段直线法等,实际应用中可根据当地具体情况选择采用,本次建模则主要采用滑动平均与调和权重相结合方法,先用滑动平均分阶段建模,后用调和权重减少趋势产量外推的风险性。

分析梧州市早稻实际单产,由图 1 可看出,梧州市水稻实际单产是逐年变化的,上升趋势不明显,因此本次建模首先采用取平滑区间 k 为 7 进行分段建模,应用统计分析软件 spss19.0,在曲线回归模型中,主要选用如下三次曲线或二次曲线计算模拟水稻产量的时间趋势项序列。

$$Y_t(i)=bi(0)+bi(1)t+bi(2)t^2+bi(3)t^3 \quad (1)$$

$$Y_t(i)=bi(0)+bi(1)t+bi(2)t^2 \quad (2)$$

2.1.1 分段模拟在 spss 中的计算方法^[7]

(1)在 SPSS 中建立一新数据文件,第 1 列为 t (年代,设置 1983 年为 1,2014 年为 32,首次计算为 1-7),第 2 列为对应的 y (实际单产),每列为 7 个数

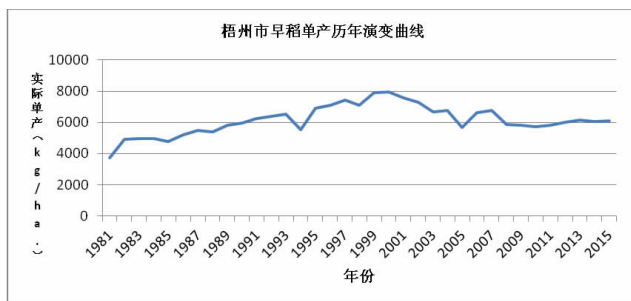


图 1 梧州市早稻单产历年演变曲线

据,表略。

(2) 选择【Analyze】→【Regression】→【Curve Estimation】打开 Curve Estimation(曲线估计)主对话框。将 y (实际单产) 添加到 Dependent (因变量) 列表,将 t (年代) 添加到 Variable (变量),并在 Models (模型) 栏中选择 Cubic (二次曲线)、Quadratic (二次曲线) 项,并选择 Display ANOVA table (显示 ANOVA 表格),然后点击 OK 按钮。

(3)重复上述计算 28 次,得出各线段方程,各线段方程均通过回归分析显著性 F 值的检验。

2.1.2 利用调和权重减少趋势产量外推风险

调和权重法^{[8][9][10]}就是以不同权重估计序列样本对未来趋势的影响,把已知时间序列 y_t 诸点的折线进行平滑并作为时间趋势项的某种近似 f_t ,主要步骤为:

$$C_{t+1}^m = \frac{m_{t+1}}{n-1} \quad (3)$$

$$W_{t+1} = Y_{t+1} - Y_t \quad (4)$$

$$\bar{W} = \sum_{t=1}^{n-1} C_{t+1}^m \cdot W_{t+1} \quad (5)$$

$$\hat{Y}_{t+1} = \hat{Y}_t + \bar{W} \quad (6)$$

即:利用调和权重系数 C_{t+1}^m 和历年趋势产量增量 W_{t+1} ,求得平均增量 \bar{W} ,最后得出趋势产量预报值 $\hat{Y}_{t+1} = \hat{Y}_t + \bar{W}$,权重序列可由 $(t=2, 3, \dots, n-1)$ 确定。由(6)式可以计算梧州市 2015 年趋势产量 $\hat{Y}_{t+1} = \hat{Y}_t + \bar{W} = 6121.302(-6.80755)6114.495$

2.2 用逐步回归分析方法建立气象产量回归方程

(1) 根据表 2 模拟获得的 1983~2014 年早稻趋势产量,代入公式 $Y_m = Y - Y_t$ 即求得各年相应的气象产量 Y_m 。

(2)进行相关分析。利用 1983—2014 年梧州市早稻气象产量与影响早稻分蘖、幼穗分化、乳熟成熟的气象因子在 SPSS 中进行相关显著性分析^[7]:选择【Analyze】→【Correlate】→【Bivariate】,将对话框中的

气象因子逐个调入右边的变量 Yt,单击 OK 按钮。获取 59 个与气象产量相关性显著的气象因子:梧州市所属 5 个气象站的 3 月中旬旬降雨量,4 月上旬旬日照时数,4 月上旬旬平均最高气温、旬平均气温、旬平均最低气温,5 月上旬旬平均最高气温,5 月上旬的旬平均气温、旬平均最低气温,6 月上旬旬平均最高气温,2 月下旬至 6 月中旬的平均最高气温。

(3)进行逐步回归^{[7][11]}。在 SPSS 中建立新数据文件,第 1 列变量为 ym(气象产量),其余各列变量为备选气象因子,记为 Xi (i=1,2… …)。选择【Analyze】→【Regression】→【Linear】→ 打开 Linear Regression(线性回归分析)主对话框。将 Ym(气象产量)添加到 Dependent(因变量)列表,将 xi (I=1,2 ……)(备选气象因子)添加到 Independent(自变量)列表。在 Method(方法)下拉菜单中选择 Stepwise(逐步回归法)。

(4) 获取逐步回归方程模型参数方法。打开 Linear Re.gression:Sta.tistics(线性回归统计:统计量对话框),点选 Estimates(估计值),这将在结果中显示回归系数 B、回归系数的标准误差(SEB)、标准化回归系数(Beta)、回归系数 B 的 t 值及双侧显著性

水平(Sig.)。点选 Model fit(模型拟合),这将在结果中显示被引入模型或被剔出的变量,复相关系数 R、R2、调整 R2、估计值的标准误差及方差分析表。

(5) 打开 Linear Regression:Save (线性回归分析:保存)对话框,点选 Predicted Val.ues(预测值)中的 Unstandardized(非标准化预测值),即模型中因变量的预测值。在 Confidence Interval(置信区间)处输入 95%,即置信度设为 95%。

(6)单击【Continue】【OK】,得到主要结果与逐步回归方程:

由表 1 可知,方程中复相关系数为 0.700,R2 为 0.489,调整 R2 为 0.44,剩余标准差为 166.74。F 值为 9.904,查 F 分布表,有 F>F0.05,表示逐步回归效果显著。由表 2 可得到逐步回归方程 $Ym=410.349-75.458\times X1-42.644\times X2+91.493\times X3$ 。

上式表明,影响梧州市早稻单产的气象因子主要有 3 个:5 月上旬、中旬期间的平均最低气温、4 月上旬的旬平均最高气温、2 月下旬至 6 月中旬期间的平均最高气温。说明 5 月上旬、中旬的平均最低气温偏高,导致夜间呼吸作用增强,不利于营养物质的有效积累^{[1][12][13]},因而对气象产量的贡献是消极的。

表 1 模型汇总

模型	复相关系数 R	R ²	调整 R ²	剩余标准差
3	0.700 ^c	0.489	0.440	166.740

表 2 Anova (方差分析表)

模型		平方和	自由度 (df)	均方	F 值	P 值 (Sig.)
3	回归	826075.855	3	275358.618	9.904	0.000
	残差	861866.207	31	27802.136		
	总计	1687942.063	34			

表 3 回归系数

模型		回归系数		标准化回归系数	B 的 t 值	双侧显著性水平 (Sig) .
		B	标准误差	B _{rta}		
3	常数	410.349	845.389		0.485	0.631
	X ₁ : 蒙山 5 月上旬、中旬的平均最低气温	-75.458	23.247	-.424	-3.246	0.003
	X ₂ : 梧州 4 月上旬旬平均最高气温	-42.644	11.143	-.560	-3.827	0.001
	X ₃ : 蒙山 2 月下旬至 6 月中旬的平均最高气温	91.493	34.757	0.381	2.632	0.013

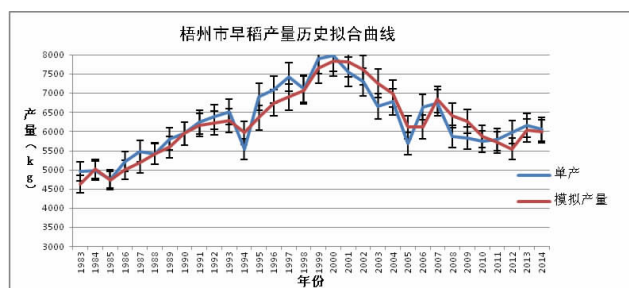


图2 梧州市早稻产量历史拟合曲线(1983—2014 年)

4 月上旬的旬平均最高气温偏高, 对水稻幼苗返青分蘖生长不利, 对气象产量的贡献也是消极的。2 月下旬至 6 月中旬平均最高气温偏高, 说明主要生育期内光照偏多, 对早稻的生长发育是极为有利的^[14], 对气象产量的贡献是积极的。

2.3 早稻历史产量模拟

将 1983—2014 年的气象因子 X_1 、 X_2 、 X_3 代入以上方程, 可得到模拟的水稻气象产量序列, 这样由时间趋势项序列和模拟的水稻气象产量序列, 可得到水稻的模拟产量。计算表明, 水稻产量历史拟合率较高, 均在 91% 以上, 见图 2。

2.4 梧州市早稻产量预测模型与 2015 年产量模拟

由以上得到的利用直线滑动均值和调和权重相结合的方法和逐步回归方程, 可得倒梧州市早稻产量预测模型 $Y_{t+1} = Y_t + \bar{W} + Y_m = Y_t + 410.349 - 75.458 \times X_1 - 42.644 \times X_2 + 91.493 \times X_3$, 其中 X_1 : 蒙山 5 月上旬、中旬的平均最低气温, X_2 : 梧州 4 月上旬的旬平均最高气温, X_3 : 蒙山 2 月下旬至 6 月中旬的平均最高气温。

利用直线滑动均值和调和权重相结合的方法对 2015 年趋势产量进行预报, 趋势产量 $Y_{t+1} = Y_t + 6121.302 + (-6.80755) = 6114.495 (\text{kg} \cdot \text{ha}^{-1})$, 利用逐步回归方程对气象产量进行预报, 气象产量 $Y_m = 410.349 - 75.458 \times 21.9 - 42.644 \times 26.6 + 91.493 \times 24.8 = 108.189 (\text{kg} \cdot \text{ha}^{-1})$; 因此, 2015 年预报产量:

$Y_{t+1} = Y_t + Y_m = 6114.495 + (-108.189) = 6006.3 (\text{kg} \cdot \text{ha}^{-1})$, 与实际产量 $6090 \text{kg} \cdot \text{ha}^{-1}$ 相比, 预报误差 -1.37%, 效果较好。

3 结论

(1) 采用 SPSS 软件进行作物产量预报, 是实际工作中有益的尝试, 它提供了新的技术手段和工作思路。

(2) 利用直线滑动均值和调和权重相结合的方法

法进行趋势产量预报, 用逐步回归分析方法模拟气象产量, 从而得到的梧州市早稻产量预测模型具有较好的预报效果, 预测结果接近实产, 准确率高, 达到了产量预报业务化所需的精度要求, 所挑选出来的 3 个预报因子具有明确的生物学意义。

(3) 在趋势产量的外延预测中, 除原有的滞后增长或增量增长方法外, 还能以不同权重方法去求算各序列样本趋势产量的影响, 大大减少了外推趋势产量的风险性, 且具有明确的生物学意义, 与其它方法比较, 它更符合农业生产的实际情况。

(4) 实际产量预报中, 如何把握趋势产量的分离, 是分析、建立气象产量模式的关键, 也是分析、建立作物产量预报模式的基础。

参考文献:

- [1] 袁立新, 段修荣, 余先超. 用 SPSS 建立自贡水稻产量年景预测模型 [J]. 高原山地气象研究, 2006, 26 (1): 31-33.
- [2] 苏永秀. 广西农业气象产量预报系统 [J]. 广西气象, 1991, (2): 41-44.
- [3] 陈靖, 苏永秀. 广西玉米、花生、柑桔农业气象产量预报技术方法的研究 [J]. 广西气象, 1991, (3): 37-40.
- [4] 周慧琴, 戴小笠, 唐缸艳. 宁夏灌区粮食趋势产量预报方法的探讨 [J]. 宁夏农林科技, 2001, (4): 24-25.
- [5] 代立芹, 吴炳方, 李强子, 等. 作物单产预测方法研究进展 [J]. 农业网络信息, 2006, (4): 24-27, 49.
- [6] 韩永翔, 尹东. 作物产量预报新方法研究 [J]. 干旱地区农业研究, 2002, 20 (3): 124-127.
- [7] 杨善朝, 张军舰. SPSS 统计软件应用基础 [M]. 桂林广西师范大学出版社 (第 2 版), 2010, 3: 70-99.
- [8] 刘文平. 利用调和权重法处理产量预测我省小麦产量效果好 [J]. 山西气象, 1994, 29 (4): 30-32.
- [9] 曲静. 调和权重法在作物产量预报中的应用 [J]. 陕西农业科学, 2011, (5): 109-110.
- [10] 北京农业大学气象教研组. 农业气象统计讲义 [M]. 气象出版社, 1980: 222-225.
- [11] 罗森波, 黄贤潜. 水稻气象产量逐步回归双重分析方法 [J]. 广东气象, 1991, (2): 8-10.
- [12] 廖雪萍, 何燕. 广西种植巴西早稻单季再生稻的适播期分析 [J]. 广西气象, 2006, 27 (1): 39-41.
- [13] 苏志, 涂方旭. 广西春播期低温阴雨的气候变化分析 [J]. 广西气象, 1999, 20 (1): 32-34.
- [14] 涂方旭. 对广西水稻气候区划的探讨 [J]. 广西气象, 2006, 27 (1): 34-36.