

董京铭, 刘瑞翔, 马晨晨, 等. 利用主成分回归方法预估连云港地区水稻气象产量[J]. 江苏农业学报, 2021, 37(3): 606-612.
doi: 10.3969/j.issn.1000-4440.2021.03.008

利用主成分回归方法预估连云港地区水稻气象产量

董京铭, 刘瑞翔, 马晨晨, 郝玲, 史达伟

(连云港市气象局, 江苏 连云港 222000)

摘要: 根据连云港地区1974–2016年5个站点的气象资料和水稻产量资料, 利用主成分回归方法研究该地区水稻不同生长期内, 气象要素与水稻气象产量的关系, 构建预估模型并对其进行检验。结果表明, 抽穗开花期和灌浆成熟期的气温、积温等热力因子对连云港地区水稻气象产量影响最大; 主成分分析的前3个分量为“抽穗开花期因子”、“灌浆成熟期因子”和“移栽返青期因子”, 可以解释连云港地区85%左右的水稻相对气象产量变化; 回归模型组合出3个新的因子, 分别对应抽穗开花期的热力条件、灌浆成熟期的热力条件和移栽返青期的温差, 该模型可以较好地预估连云港地区水稻的气象产量。

关键词: 连云港; 水稻; 气象产量; 主成分回归

中图分类号: S511 **文献标识码:** A **文章编号:** 1000-4440(2021)03-0606-07

Estimation of the meteorological yield of rice in Lianyungang area using principal component regression method

DONG Jing-ming, LIU Rui-xiang, MA Chen-chen, HAO Ling, SHI Da-wei

(Lianyungang Meteorological Bureau, Lianyungang 222000, China)

Abstract: The relationship between meteorological elements and the meteorological yield of rice during different growing periods of rice in Lianyungang region was studied based on the meteorological data and rice yield data of five stations from 1974 to 2016, using principal component regression method. A prediction model was constructed and tested. The results showed that, thermal factors such as air temperature and accumulated temperature at heading-flowering stage and filling-maturity stage had the greatest impact on the meteorological yield of rice in Lianyungang region. The first three components of principal component analysis were “heading-flowering period factor”, “filling-maturity period factor” and “transplanting-returning green period factor”, which could explain about 85% of the relative meteorological yield changes of rice in Lianyungang area. Three new factors were combined by the regression model, which corresponded to the thermal conditions of the “heading-flowering period”, the thermal conditions of the “filling-maturity period” and the temperature difference in the “transplanting-returning green period”, respectively. The model can well predict the meteorological yield of rice in Lianyungang area.

Key words: Lianyungang; rice; meteorological yield; principal component regression

长江流域是中国最大的水稻生产区, 播种面积占全国70%以上, 江苏省水稻种植面积和总产量均

排全国第五位^[1-2]。江苏省以一季稻种植为主, 通常在5月下旬播种, 10月收获^[3]。连云港市是江苏省水稻生产的优势区域, 是全省粮食供给的支柱地区之一, 对保障江苏省农业优势产业的发展、粮食安全具有重要作用^[4]。近年来, 随着全球变暖日益严重, 极端高温事件频发, 水稻抽穗扬花受到影响, 花器官发育不全, 花粉发育不良, 严重影响着江苏省水

收稿日期: 2020-09-21

基金项目: 连云港市气象局科技计划项目(LG202001)

作者简介: 董京铭(1989–), 江苏连云港人, 硕士, 工程师, 主要从事农业气象研究。(E-mail) 792156709@qq.com

通讯作者: 史达伟, (E-mail) 395158610@qq.com

稻的产量^[5-7],另一方面,极端天气事件可能导致温度波动振幅增大,短期严重低温造成的区域性障碍型低温冷害也会水稻产量造成重要影响^[8-9]。

陶炳炎等^[10]发现如果水稻开花时遇到连续 7 d 大于 35 ℃ 的高温,空壳率将升高 10 倍。于堃等^[11]指出常发生高温热害的地区主要分布在淮河以南地区,且 21 世纪以来最为严重。包云轩等^[1]指出江苏省水稻热害主要发生在拔节至孕穗期和抽穗至乳熟期,并且气候偏暖时水稻热害多,产量相应低。陈斐等^[12]认为移栽至灌浆末期的热力因子调控着水稻的气象产量。陈雅子和申双和^[3]采用 Copula 函数验证了夏季高温热害指标与水稻减产之间的相关关系。由此可见温度这一气象因子对水稻的产量有影响。

前人工作大多是研究较大的地理区域,如长江中下游或江苏省,很少针对局部小区域进行研究,但是在农业生产中,大尺度研究得到的结论并不一定适用于局地,所以本研究拟从连云港地区气象站点资料出发,利用主成分回归方法,探究何种气象因子对该地区水稻气象产量具有重要影响,构建回归方程并对其进行检验,为连云港地区农业生产应对气候变化提供帮助与理论依据。

1 材料与方法

1.1 研究材料

本研究使用的资料包括气象资料和水稻产量资料。气象资料为连云港市 5 个气象站点(市区、赣榆、灌南、灌云、东海)的观测数据,包括播种出苗期、移栽返青期、分蘖期、拔节孕穗期、抽穗开花期和灌浆成熟期 6 个水稻生长期,以及相应时期的平均气温、最高气温、最低气温、平均湿度、降水量、日照、温差、积温等物理量。产量资料包括连云港市 5 个县(市区)的水稻气象产量和相对气象产量。资料年限均为 1974–2016 年。

1.2 气象产量、趋势产量与相对气象产量

水稻等农作物产量按照其主要影响因素的不同,可以分解成趋势产量、气象产量与随机误差 3 个分量^[12-14],即

$$Y = Y_t + Y_m + \varepsilon \quad (1)$$

其中, Y 为实际水稻产量, Y_t 为趋势产量, Y_m 为气象产量, ε 为随机误差。趋势产量主要受到经济、耕作技术等要素发展的影响;而气象产量是受到各

类气象要素影响而产生的波动的产量;随机误差可能包括系统偏差或者其他随机噪音产生的产量偏差,相对前两个分量可以忽略。对于趋势产量,可以对实际产量进行滑动平均、低频滤波等方法得到^[15-16]。因此,气象产量可以由实际产量与趋势产量的差表示

$$Y_m = Y - Y_t \quad (2)$$

为消除不同时期不同生产力水平的影响,使得相对气象产量在较长的时间尺度和较广的空间尺度上具有一定的可比性,可以进一步将气象产量与趋势产量的比值定义为相对气象产量 Y_r ,即:

$$Y_r = \frac{Y - Y_t}{Y_t} \times 100\% \quad (3)$$

1.3 主成分回归分析

首先对气象站点的气象记录和水稻相对气象产量间进行相关性分析,从中挑选与水稻相对气象产量最密切相关的若干气象要素作为主要因子。前人研究结果表明,影响水稻相对气候产量的主导因子之间有很高的相关性^[9,17],比如平均气温、日照和积温之间就具有较高的一致性。因此,需要采用一些方法来降低因子之间的相关性,以提高模型拟合的效果。

主成分回归是一种常用的降低因子间相关性的方法。其主要思路为通过主成分分析,对原始自变量进行线性组合,得到对因变量起主导作用的一组相互间不相关的新的自变量;接着对新的自变量进行回归分析,得到回归方程后,再将原始自变量带入,得到最终的模型。

2 结果与分析

2.1 关键气象要素的筛选

将连云港地区 5 个气象站点在水稻 6 个不同生育阶段的 8 项气象观测记录与各地区当年的相对气象产量进行相关分析,所得结果如表 1 所示。在所有 6 个生育阶段的 8 项气象要素中,与相对气象产量相关系数达到 0.01 显著水平的气象要素被认为是影响连云港地区水稻相对气象产量的关键气象要素。由表 1 可知,符合筛选标准的气象要素有以下 9 个:移栽返青期的温差,抽穗开花期的平均气温、最高气温、最低气温、日照和积温,灌浆成熟期的平均气温、最高气温以及积温。其中,抽穗开花期和灌浆成熟期的气温、积温等热力因子在所有关键气象

要素中占的比重很大。前人研究结果也指出由于该地区水稻品种较强的感温性,水稻生长期中移栽至灌浆末期的热力因子对水稻产量有重要影响^[1,9,12]。

因此相关性分析筛选出的 9 个关键气象要素不仅在统计学上与相对气象产量显著相关,同时也与前人研究结果相吻合。

表 1 连云港水稻相对气象产量与各生育期气象因子之间的相关系数

Table 1 Correlation coefficients between the relative meteorological yield of rice and the meteorological factors of each growing period in Li-anyungang

生育期	平均气温	最高气温	最低气温	平均湿度	降水量	日照	温差	积温
播种出苗期	-0.028	-0.036	0.009	0.070	-0.040	-0.018	-0.047	-0.028
移栽返青期	-0.151 *	-0.167 *	-0.035	0.157 *	0.012	-0.104	-0.186 **	-0.151 *
分蘖期	-0.043	-0.049	-0.050	-0.078	-0.171 *	0.001	-0.022	-0.043
拔节孕穗期	0.027	-0.018	0.077	0.111	0.145 *	-0.135	-0.130	0.027
抽穗开花期	0.319 **	0.33 **	0.222 **	0.006	-0.166 *	0.186 **	0.125	0.319 **
灌浆成熟期	0.217 **	0.201 **	0.163 *	0.007	-0.046	0.030	-0.003	0.217 **

* 表示在 0.05 水平显著相关, ** 表示在 0.01 水平显著相关。

2.2 水稻相对气象产量模型

对相关性分析筛选出的 9 个关键气象要素(记为 $X_1 \sim X_9$)进行共线性诊断,诊断结果如表 2 所示。判断共线性有以下几个指标:1. 多个维度的特征根接近 0。2. 条件指数大于 10 且相应的方差比例大于 0.5。根据共线性诊断的结果,维度 D_3 、 D_6 和 D_8 的特征根接近 0,而它们的条件指数大于 10,而且 D_3 与 X_1 、 D_6 与 X_6 以及 D_8 与 X_2 、 X_3 、 X_4 的方差比例均大于

0.5。综合各维度的特征根、条件指数以及与各气象要素的方差比例的结果,所选 9 个气象要素之间存在多重共线性。

对存在多重共线性的变量直接采用线性回归的方法建立模型会导致各变量回归系数的方差过大,影响模型的可靠性^[12]。因此本研究采用主成分回归的方法构建连云港地区水稻相对气象产量的模拟模型。

表 2 主导气象因子间的共线性诊断结果

Table 2 Diagnosis results of collinearity among dominant meteorological factors

维度	特征值	条件指数	方差比例						
			X_1	X_2	X_3	X_4	X_5	X_6	X_7
D_1	7.918	1.000	0	0	0	0	0	0	0
D_2	0.054	12.113	0.06	0	0	0	0.23	0	0
D_3	0.020	19.992	0.62	0	0	0	0.05	0	0.03
D_4	0.006	35.317	0.10	0	0	0	0.04	0	0.27
D_5	0.001	73.921	0.14	0	0	0.02	0.01	0.10	0.13
D_6	4.401×10^{-4}	134.131	0.01	0	0.02	0.00	0.04	0.57	0.37
D_7	2.230×10^{-4}	188.221	0.06	0	0.26	0.14	0.49	0.32	0.19
D_8	2.213×10^{-4}	592.952	0.01	1.00	0.72	0.83	0.14	0.01	0.02

$X_1 \sim X_7$ 为相关性分析筛选出的关键气象要素。

对 9 个关键气象要素进行主成分分析,特征值大于 1 的前 3 个分量(PC1、PC2和PC3)的分析结果如表 3 所示。前 3 个分量累计方差贡献率为 85.25%,因此这前 3 个分量可以解释相对气象产量的主要变化。从表 3 中可以看出,第一分量(PC1)中因子 X_2 、 X_3 和 X_6 所占的比重较高,对应抽穗开花期

的平均气温、最高气温和积温,因此 PC1 主要反映的是抽穗开花期的热力条件,可以称之为“抽穗开花期因子”。类似的,第二分量(PC2)由 X_7 、 X_8 和 X_9 这 3 个灌浆成熟期的因子主导,可以称之为“灌浆成熟期因子”;第三分量(PC3)由移栽返青期的温差因子 X_1 主导,可以称之为“移栽返青期因子”。按照

表3的结果,这3个新因子与9个关键气象要素 $X_1 \sim X_9$ 的关系如下:

$$PC1 = -0.045X_1 + 0.226X_2 + 0.206X_3 + 0.195X_4 + 0.087X_5 + 0.226X_6 + 0.131X_7 + 0.128X_8 + 0.131X_9 \quad (4)$$

$$PC2 = 0.015X_1 - 0.140X_2 - 0.154X_3 - 0.088X_4 - 0.164X_5 - 0.140X_6 + 0.342X_7 + 0.295X_8 + 0.342X_9 \quad (5)$$

$$PC3 = 0.562X_1 + 0.115X_2 - 0.080X_3 + 0.342X_4 - 0.613X_5 + 0.115X_6 - 0.066X_7 - 0.047X_8 - 0.066X_9 \quad (6)$$

将主成分分析所得的前3个分量作为新的因子,构建它们与相对气象产量之间的回归方程,可以得到:

$$Y_r = 2.701 PC1 - 0.110 PC2 - 0.993 PC3 \quad (R^2 =$$

$$0.216) \quad (7)$$

由公式(7)可以看出,3个分量中,PC1(抽穗开花期因子)的系数最大,其次是PC3(移栽返青期因子),PC2(灌浆成熟期因子)的贡献最小。

将前3个分量与各因子的关系[公式(4)、公式(5)和公式(6)]代入公式(7),得到以关键气象要素为自变量的连云港地区水稻相对气象产量回归模型:

$$Y_r = -0.684X_1 + 0.512X_2 + 0.653X_3 + 0.197X_4 + 0.862X_5 + 0.512X_6 + 0.382X_7 + 0.359X_8 + 0.382X_9 \quad (8)$$

表3 连云港地区水稻相对气象产量和关键气象因子的主成分分析结果

Table 3 Principal component analysis results of relative meteorological yield and key meteorological factors of rice in Lianyungang area

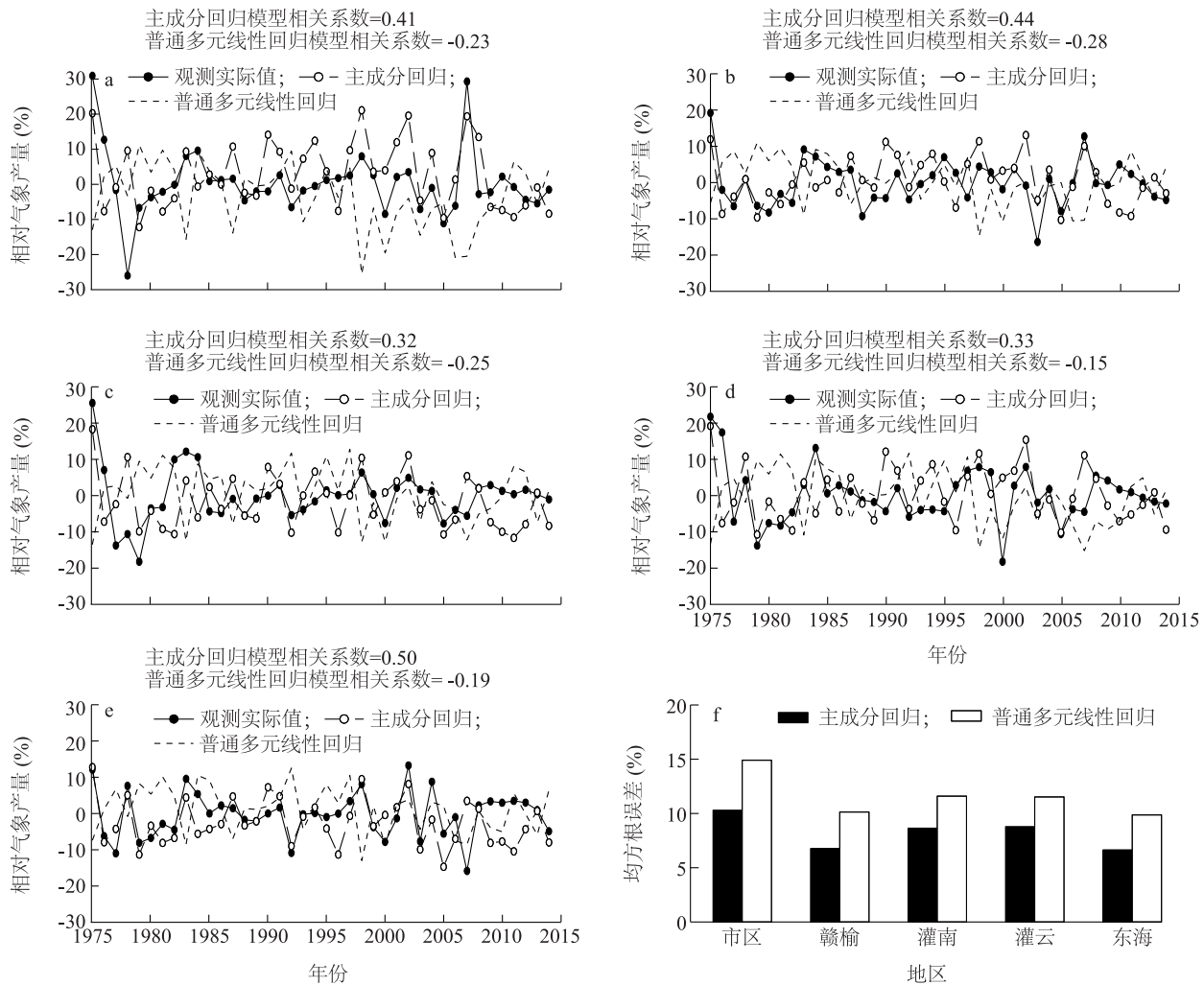
主成分	特征值	累计方差贡献	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
PC1	4.108	45.643	-0.046	0.226	0.206	0.195	0.087	0.226	0.131	0.128	0.131
PC2	2.391	72.212	0.015	-0.140	-0.154	-0.088	-0.164	-0.140	0.342	0.295	0.342
PC3	1.173	85.248	0.562	0.115	-0.080	0.342	-0.613	0.115	-0.066	-0.046	-0.066

2.3 模型对水稻产量的预估效果

首先对建立的连云港地区水稻相对气象产量模型的拟合效果进行评估。将连云港地区这5个县(市区)的9个关键气象要素标准化之后,带入公式(8),得到模型预估的相对气象产量。作为比较,通过普通多元线性回归,建立了另一个相对气象产量与9个关键气象要素间的模型,并且也得到了一组普通多元线性回归模型下的连云港地区水稻相对气象产量的预估值。图1a~图1e展示了1975~2014这40年5个县(市区)水稻的相对气象产量的实际值(实线)、主成分回归模型的预估值(虚线)以及普通多元线性回归模型的预估值(点线)。可以看到,连云港5个县(市区)的主成分回归模型预估的相对气象产量和实际的相对气象产量之间的相关系数均大于0.30;相关系数均通过0.05水平的显著性检验。其中,东海县两者相关系数最大(0.50),而灌南和灌云县的相关系数最小(0.32、0.33);但是普通多元线性回归模型的预估值与实际值的相关系数均为负值。从均方根误差来看(图1f),5个县(市区)的主成分回归模型的预估值与实际值之间的均方根误差均小于0.1,其中赣榆和东海县均方根误

差较小,而市区的均方根误差较大,但也不超过0.1;反观普通多元线性回归模型,5个县(市区)模型预估值与实际值之间的均方根误差几乎都大于0.1。由此可以看出,主成分回归模型相比普通的多元线性回归模型,对连云港地区水稻相对气象产量有比较好的模拟效果,尤其是对东海和赣榆县。

由相对气象产量与实际总产量和趋势产量的关系可知,通过回归方程得到的模型预估的相对气象产量可以结合趋势产量,进一步地代入公式(2)和公式(3),从而对水稻总产量进行预估。图2a~图2e展示了1975~2014年5个县(市区)水稻总产量的实际值(实线)、主成分回归模型的预估值(虚线)以及普通多元线性回归模型的预估值(点线)。由于主成分回归模型对相对气象产量已经有较好的拟合效果,结合与趋势产量的相对关系,可以看到模型预估的5个县(市区)预估总产量与实际总产量之间的相关系数均达到0.90以上,而均方根误差均小于50(小于总产量的10%)。普通多元回归模型对总产量的预估效果较差,不管是相关系数还是均方根误差,普通多元线性回归模型的表现与主成分回归模型有一定差距。



a: 市区; b: 赣榆; c: 灌南; d: 灌云; e: 东海; f: 均方根误差。

图1 主成分回归模型与普通多元线性回归模型对连云港地区水稻相对气象产量预估效果的比较

Fig.1 Comparison on the prediction effect of the relative meteorological yields of rice in Lianyungang area between principal component regression model and common multiple linear regression model

3 讨论

通过主成分回归分析方法,本文构建了水稻生长发育阶段9个关键气象要素与相对气象产量间的模型,并从中得出以下主要结论:(1)关键气象要素中,抽穗开花期和灌浆成熟期的气温、积温等热力因子对连云港地区水稻气象产量影响最大。(2)主成分分析的前3个分量可以解释连云港地区85.25%左右的水稻相对气象产量变化,3个分量分别可以称作“抽穗开花期因子”、“灌浆成熟期因子”和“移栽返青期因子”。(3)主成分回归模型组合出3个新的因子,分别对应抽穗开花期热力条件、灌浆成熟期热力条件和移栽返青期温差,可以较好预估连云

港地区,尤其是东海和赣榆县的水稻相对气象产量。

本文利用连云港地区5个气象站点的观测数据,系统研究了连云港地区水稻不同生长期,气象要素对水稻气象产量的影响,发现“移栽返青期因子”、“抽穗开花期因子”和“灌浆成熟期因子”对水稻气象产量有重要影响。其中,抽穗开花期是指从水稻稻穗于顶端茎鞘中抽出至开花齐穗,此时适宜温度在25~32℃,这一时期如果气温偏低,热力条件较差,水稻容易形成空壳或者瘪谷;但温度32℃以上水稻结实率也会下降。考虑到连云港地区六月初气候平均温度大约26℃左右,因此“抽穗开花期因子”和连云港地区水稻气象产量间基本上应为较好的正相关关系。而灌浆成熟期是开花到谷粒成熟

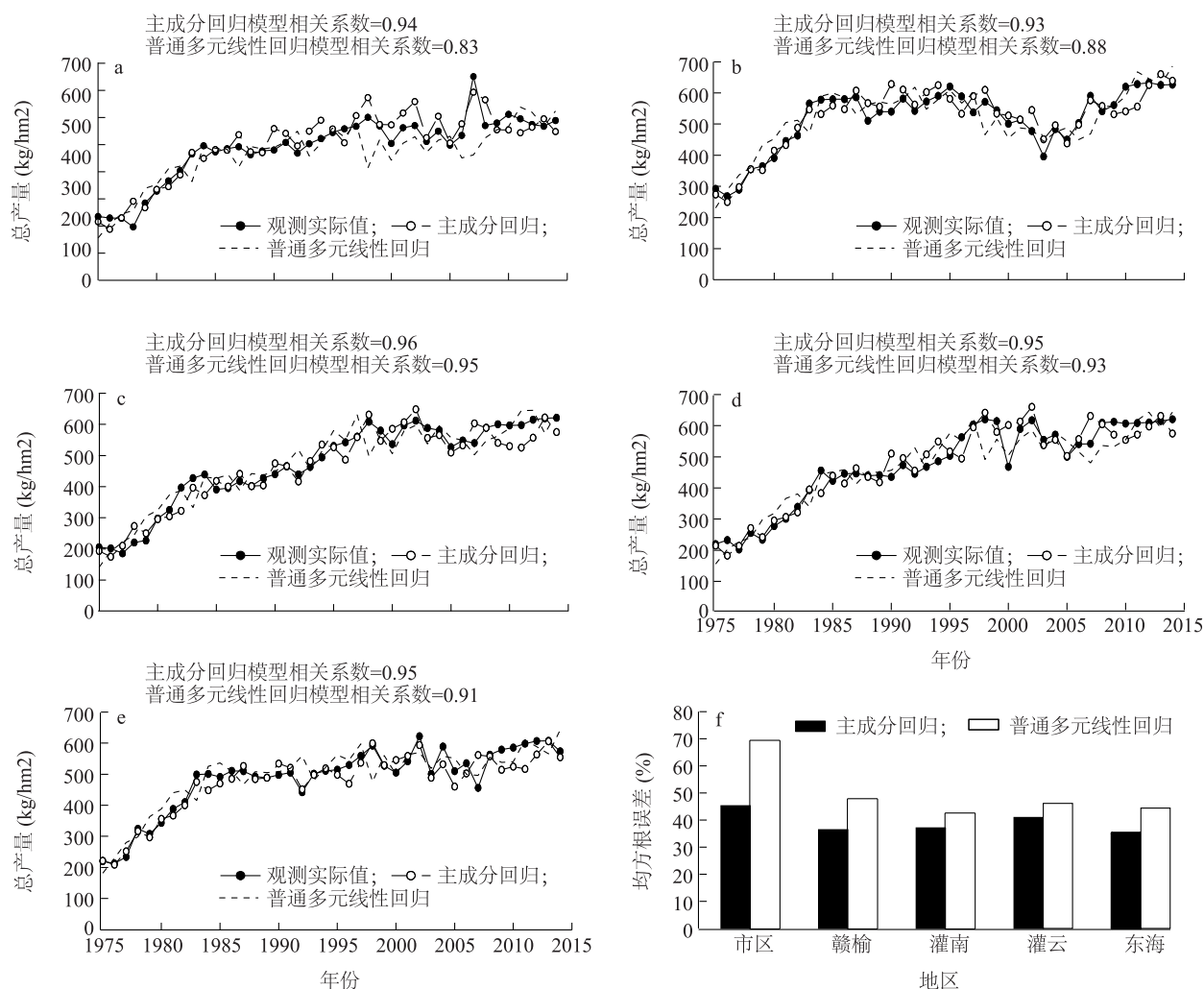


图2 主成分回归模型与多元线性回归模型对连云港地区水稻总产量预估效果的比较

Fig.2 Comparison on the prediction effect of the total rice output in Lianyungang area between principal component regression model and multiple linear regression model

的时期,这一时期对热力条件要求相对较高,适宜温度在23~28℃,温度过低会造成水稻新陈代谢偏慢,结实率降低^[18-20],而过高也会造成水稻本身呼吸消耗增加,结实率也会降低。连云港地区六月至七月的气候平均气温为27℃,因此用线性模型分析这一时期的热力条件与水稻产量的关系可能并不准确,“灌浆成熟期因子”的贡献率很可能在主成分回归的模型中被低估了,这可能给模型带来了一定的不确定性。值得注意的是,移栽返青期的温差这一因子在我们的模型中有这较高的贡献率。移栽返青期主要指水稻移栽到拔节前的这段时期。这一时期的热力条件并不直接影响水稻结实,而是可能通过影响分蘖来影响水稻穗数,从而间接影响最终水稻

产量。另外,正如前面讨论的,线性模型对“灌浆成熟期因子”影响的低估可能间接导致了“移栽返青期因子”影响的高估。尽管主成分回归模型有这样的不确定性,但是同普通线性回归模型相比,主成分回归模型在对水稻产量预估的准确性和因子的可解释性上都有较大的优势。

近年来全球变暖问题深受科学界关注,20世纪中后期以来,中国大陆显著增温^[21-22],江苏省地区的变暖速率基本与全国平均保持一致^[23],Li等^[24]指出全球变暖背景下,东亚夏季风低层环流的增强会导致江淮流域的夏季风降水显著增加,更热更湿的气候环境势必会对该地区水稻的气象产量造成重要影响^[1],这提醒相关部门应尽快制定相应的措施

来应对未来气候变化以及潜在的粮食危机。

参考文献:

- [1] 包云轩,刘 维,高 苹,等. 气候变暖背景下江苏省水稻热害发生规律及其对产量的影响[J]. 中国农业气象,2012,33(2): 289-296.
- [2] 艾玉春,董瑶月,汪吉东,等. 水稻产量主要养分限制因子及养分运筹应对技术研究[J]. 江苏农业学报,2015,31(3): 558-563.
- [3] 陈雅子,申双和. 江苏省水稻高温热害保险的天气指数研制[J]. 江苏农业科学,2016,44(10): 461-464.
- [4] 张正球,胡曙望,范郁尔,等. 连云港市水稻生产制约因素及发展对策[J]. 北方水稻,2013,43(2): 70-72.
- [5] SATAKE T, YOSHIDA S. High temperature induced sterility in indica rice at flowering[J]. Proceedings of the Crop Science Society of Japan, 1978, 47(1): 6-17.
- [6] 李守华,田小海,黄永平,等. 江汉平原近 50 年中稻花期危害高温发生的初步分析[J]. 中国农业气象,2007,28(1): 5-8.
- [7] 任义方,高 苹,王春乙. 江苏高温热害对水稻的影响及成因分析[J]. 自然灾害学报,2010,19(5): 101-107.
- [8] 程勇翔,王秀珍,郭建平,等. 农作物低温冷害监测评估及预报方法评述[J]. 中国农业气象,2012,33(2): 297-303.
- [9] 徐 敏,徐经纬,高 苹,等. 江苏水稻障碍型冷害时空变化特征及敏感性分析[J]. 气象,2015, 41(11): 1367-1373.
- [10] 陶炳炎,汤志成,彭钊安,等. 江苏杂交水稻气候分析和区划[J]. 南京气象学院学报,1981(1): 65-73.
- [11] 于 堃,宋 静,高 苹. 江苏水稻高温热害的发生规律与特征[J]. 气象科学, 2010, 30(4): 530-533.
- [12] 陈 斐,杨沈斌,申双和,等. 基于主成分回归法的长江中下游双季早稻相对气象产量模拟模型[J]. 中国农业气象,2014,35(5): 522-528.
- [13] 秦 剑. 气候变化与昆明粮食生产的研究[J]. 南京气象学院学报,1999,22(4): 671-679.
- [14] 朱自玺,刘荣花,方文松,等. 华北地区冬小麦干旱评估指标研究[J]. 自然灾害学报,2003,12(1): 145-150.
- [15] 房世波. 分离趋势产量和气候产量的方法探讨[J]. 自然灾害学报,2011,20(6): 13-18.
- [16] 孙卫国,程炳岩,杨沈斌,等. 区域气候变化对华东地区水稻产量的影响[J]. 中国农业气象,2011,32(2): 227-234.
- [17] 季彪俊. 影响水稻产量因子的研究[J]. 西南农业大学学报(自然科学版),2005, 27(5): 579-583.
- [18] 沈 航,陈 娇,黎佰杨,等. 低温胁迫对不同杂交水稻组合苗期耐冷性及生理生化指标的影响[J]. 江苏农业科学,2020,48(12): 59-63.
- [19] 田孟祥,张时龙,何友勋,等. 水稻耐低温基因 *bZIP73* 分子标记的开发与验证[J]. 江苏农业学报,2019,35(6): 1265-1270.
- [20] 胡雪琼,刘文英,张明达,等. 云南省水稻低温冷害风险研究[J]. 江苏农业科学,2019,47(18): 283-286.
- [21] 任国玉,初子莹,周雅清,等. 中国气温变化研究最新进展[J]. 气候与环境研究,2005,10(4): 701-716.
- [22] 丁一汇,任国玉,石广玉,等. 气候变化国家评估报告 I: 中国气候变化的历史和未来趋势[J]. 气候变化研究进展,2006,2(1): 3-8.
- [23] 周晓兰,高庆九,邓自旺,等. 江苏气温长期变化趋势及年代际变化空间差异分析[J]. 南京气象学院学报,2006,29(2): 196-202.
- [24] LI Z B, SUN Y, LI T, et al. Future changes in East Asian summer monsoon circulation and precipitation under 1.5 to 5.0 °C of warming[J]. Earth's Future, 2019, 7: 1391-1406.

(责任编辑:陈海霞)