

张 顾,任义方,肖良文,等. 气象因素对江苏地区荷斯坦奶牛产奶量的影响及产奶量预测研究[J]. 江苏农业科学,2020,48(23):150-154.  
doi:10.15889/j.issn.1002-1302.2020.23.030

# 气象因素对江苏地区荷斯坦奶牛产奶量的影响 及产奶量预测研究

张 顾<sup>1</sup>,任义方<sup>2</sup>,肖良文<sup>3,4</sup>,史 潇<sup>1</sup>,罗晓春<sup>1</sup>,敖 意<sup>3,4</sup>

(1. 江苏省气象服务中心,江苏南京 210008; 2. 江苏省气候中心,江苏南京 210008;

3. 江苏桢源应用气象研究院有限公司,江苏南京 211100; 4. 南京卫岗乳业有限公司,江苏南京 211100)

**摘要:**在奶牛品种、生理因素以及管理条件相对稳定可控的情况下,气象因素是影响产奶量的重要因素。利用主成分分析法研究气象因素对产奶量的影响,并与相关分析法结果进行比较,确定影响江苏地区荷斯坦奶牛产奶量的主要气象因素和关键时期;在此基础上,构建考虑气象因素下基于 BP(back propagation)神经网络的产奶量预测模型。结果表明,产奶量与温度、湿度呈明显负相关,与风速、降水量呈负相关,与日照时长、能见度呈明显正相关;主成分分析法得到的 3 个主成分因子(温度因子、天气因子、雨量因子)累积贡献率达 0.794 3;较相关分析法,主成分分析法充分考虑气象因素的协同关系,分析结果更符合实际情况。考虑气象因素条件下基于 BP 神经网络的产奶量预测模型检验样本的决定系数达 0.65,相对误差达 9.48%,精确性、稳定性均较好。研究结果对江苏地区奶牛管理和牛奶生产提供环境控制依据具有重要意义。

**关键词:**荷斯坦奶牛;产奶量;气象因素;主成分分析;BP 神经网络;预测模型

**中图分类号:**S823.9<sup>+</sup>1;S811.1

**文献标志码:**A

**文章编号:**1002-1302(2020)23-0150-05

奶业是江苏现代农业和食品工业发展的重点产业,2018 年末全省存栏 13.4 万头奶牛,生鲜乳年产量 50.0 万 t<sup>[1]</sup>,产奶量是衡量奶牛泌乳性能的主要指标,为保障产奶量稳定,分析并控制其影响因素是基本途径。影响产奶量的因素<sup>[2-4]</sup>主要包括内因(奶牛品种、生理因素等)和外因(气象环境因素、饲养管理因素等),在奶牛品种、生理因素和管理条件相对稳定可控的情况下,气象因素是影响奶牛产奶量的重要因素<sup>[5-6]</sup>。江苏地区气候特点显著,夏季炎热高温,降水集中,雨热同期;冬季寒冷干燥,降水较少;春秋短促,气象条件多变。江苏地区独特的气象条件给奶牛的生长发育和产奶带来诸多影响。

有学者就气象因素与产奶量的关系已进行了诸多研究。Trajchev 等利用线性混合模型,发现高温易导致产奶量大幅下降<sup>[7]</sup>;West 等基于数理统计

法发现,温度、湿度和温湿度指数对产奶量影响较大<sup>[8]</sup>;魏学占等利用相关分析发现,温湿度、气压是乌兰浩特地区春秋季节奶牛产奶量的主要影响因素<sup>[9]</sup>;陈志英等基于多元线性回归定量分析气象因素对黑龙江地区产奶量的影响。结果表明,温度、湿度和日照影响较大<sup>[10]</sup>。由此可见,学者大多利用相关分析和数理统计方法研究气象因素对产奶量的影响,方法多侧重于气象单因素的影响及定性分析,但气象因素之间存在协同作用,最终影响产奶量,因此有必要分析气象多要素对产奶量的影响。

多数研究利用泌乳曲线分析奶牛个体泌乳规律,从而建立产奶量预测模型。Brody 提出简化指数模型拟合泌乳曲线,Wood 考虑环境因素提出不完全伽玛函数模型;王雅春等添加系数和校正值得 Wood 模型进行改进<sup>[11]</sup>,此外泌乳曲线拟合模型还包括逆多项式模型<sup>[12]</sup>、多项式回归模型<sup>[13]</sup>、Wilmink 模型<sup>[14]</sup>等。由此可见,学者主要是针对个体的泌乳规律进行研究,针对某牛场或某地区逐日或逐月产奶量的预测研究较少<sup>[15]</sup>,考虑温度、湿度等气象因素下预测产奶量的报道较少。

综上所述,本试验从产奶量与气象因素的角度,首先利用主成分分析法研究气象因素对产奶量的影响,并传统相关分析的结果比较,确定影响

收稿日期:2020-03-19

基金项目:国家重点研发计划(编号:KJZLJJ 202010);江苏省气象局面上项目(编号:KM201806)。

作者简介:张 顾(1993—),男,江苏扬州人,硕士,助理工程师,从事气象服务和水文气象研究。E-mail:15950505089@163.com。

通信作者:任义方,工程师,从事农业气象研究。E-mail:renyifang2006@126.com。

江苏地区产奶量的主要气象因素和关键时期;然后基于气象因素与产奶量的关系,构建考虑气象因素下的基于 BP(back propagation)神经网络的产奶量预测模型,并分析模型预测准确度。研究结果对江苏地区奶牛管理和牛奶生产提供环境控制依据具有重要意义。

## 1 材料与方法

### 1.1 研究资料

本研究选取南京卫岗乳业有限公司泗洪牧场 2017—2019 年奶量记录资料,资料内容包括荷斯坦奶牛逐日信息数据、逐日牛奶总产量,并根据泌乳奶牛数量计算逐日产奶量(以下简称产奶量);选取泗洪气象站(站码 58135)相应时间序列的气象观测资料,观测要素包括日平均温度、日最低温度、日最高温度、日平均相对湿度、日平均风速、日累计降水量、日最低能见度、日照时长、气压。为保证资料一致性、准确性,共保留 775 组数据。

### 1.2 研究方法

**1.2.1 主成分分析法** 主成分分析(principal component analysis,简称 PCA)是一种将多指标问题转化为较少综合指标的统计方法,基本思路是借助正交变换,将分量相关的原随机变量转换为分量无关的新变量。在处理实际问题时,当变量个数众多且变量之间关系复杂,分析问题难度增加,借助主成分分析法可重新组合成无关的综合变量<sup>[16]</sup>。主成分分析法广泛应用于动植物、生态等方面,近年来在奶牛饲养及其产奶量预测<sup>[17-18]</sup>上也有研究,方法可在不损失太多相关信息的基础上进行全面多性状分析,选择效果突出。

本研究利用主成分分析法分析气象因素对产奶量的影响步骤如下:

(1)采集  $P$  维随机向量的  $n$  个样本构造样本矩。

(2)样本矩标准化变换,公式为:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, i = 1, 2, \dots, n; j = 1, 2, \dots, P. \quad (1)$$

式中: $Z_{ij}$ 为经标准化后的气象因素; $x_{ij}$ 为原始气象因素; $\bar{x}_j$ 为气象因素的均值; $s_j$ 为气象因素的标准差。

(3)对标准化阵  $Z$  气象因素相关系数矩阵计算,公式为:

$$R = [r_{ab}]_{p \times p}; \quad (2)$$

$$r_{ab} = \frac{\sum_{i=1}^n [|(x_{ia} - \bar{x}_a)| |(x_{ib} - \bar{x}_b)|]}{\sqrt{\sum_{i=1}^n (x_{ia} - \bar{x}_a)^2 (x_{ib} - \bar{x}_b)^2}}, a, b = 1, 2, \dots, 9; i = 1, 2, \dots, 775. \quad (3)$$

式中: $R$ 为相关系数矩阵; $r_{ab}$ 为第  $a$  个气象因素与第  $b$  个气象因素的相关系数。

(4)求解相关矩阵  $R$  的特征方程,得  $p$  个特征根  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ 。

确定主成分数量为  $m$  个,使累积贡献率达 80% 左右,且各主成分特征根一般须大于 1。

**1.2.2 BP 神经网络模型** 人工神经网络(artificial neural network,简称 ANN)利用非线性模拟技术拟合输入输出相互关系,BP 神经网络应用最为广泛。BP 神经网络由 Rumelhart 和 Celland 于 1986 年提出,是一种按误差逆传播算法训练的多层前馈网络,算法基于 Delta 学习规则,利用梯度搜索技术,通过反向传播不断调整网络的权值和阈值,最终实现网络的实际输出与期望输出的均方差最小化<sup>[19-20]</sup>。BP 神经网络结构简单、仿真能力强、易于实现,为考虑气象因素的产奶量预测提供了新的途径<sup>[21]</sup>。

本研究采用 3 层网络结构的 BP 神经网络模型(图 1),其拓扑结构包括输入层、隐含层、输出层,每层由一定数量的神经元构成,能够执行简单并行运算。本研究输入层为经分析后的主成分气象因素,输出层为产奶量,关键是确定隐含层层数与节点。若隐含层节点数过少,则 BP 网络获取有效信息少,容错性差;若节点数过多,则 BP 网络样本中非规律信息,泛化能力下降,因此确定合适的隐含层节点数非常重要<sup>[22]</sup>。学者提出大量确定方法,如试凑法、奇异值法和经验公式法等,本研究拟采用 4 种经验公式,并对比选取结果较好的隐含层节点数,公式为:

$$N_{hid} = 2N_{in} + 1; \quad (4)$$

$$N_{hid} = \frac{N_{in} + \max(N_{out}, N_c)}{2}; \quad (5)$$

$$N_{hid} = \sqrt{N_{in}(N_{out} + 2)} + 1; \quad (6)$$

$$N_{hid} = \sqrt{N_{in} + N_{out}} + \alpha, 1 \leq \alpha \leq 10. \quad (7)$$

式中: $N_{hid}$ 为隐含层节点数; $N_{in}$ 为输入层节点数; $N_{out}$ 为输出层节点数; $N_c$ 为目标分类数。

## 2 结果与分析

### 2.1 气象因素对产奶量的影响

气象观测数据经标准化处理、相关系数矩阵计

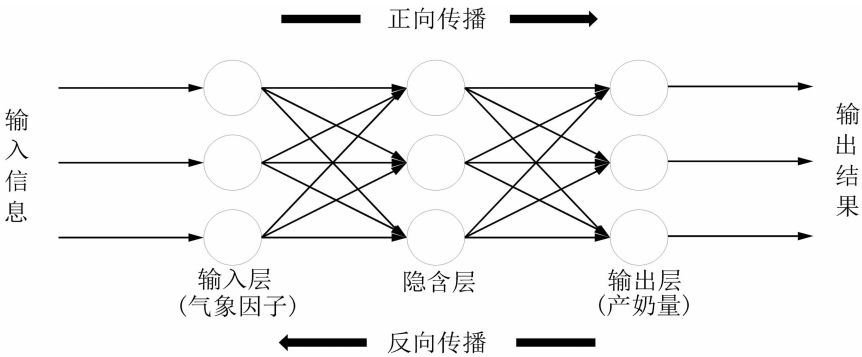


图1 考虑气象因素下的 BP 神经网络产奶量预测模型

算、特征根方程求解等,计算得出相关系数矩阵特征值和贡献率(表 1)。根据主成分因子选取原则,前 3 个成分特征根均大于 1 且累积贡献率为 79.43%,可反映气象因素的大部分信息,因此将前 3 个成分作为主成分,并计算得出其特征向量。程序运行于 MATLAB R2014a 中。

表 1 泗洪站 775 个样本数据相关系数矩阵特征值计算结果

主成分	特征根	贡献率(%)	累积贡献率(%)
第 1 主成分	3.882 7	43.14	43.14
第 2 主成分	2.183 2	24.26	67.40
第 3 主成分	1.109 5	12.33	79.73
⋮	⋮	⋮	⋮

由表 2 可知,在第 1 主成分中,日平均温度、日最低温度和日最高温度载荷值较大且为负值,气压特征值也较大但为正值,江苏地区长时间序列下温度越高,气压越小,可知产奶量与温度呈明显负相关,这是由于荷斯坦奶牛耐冷不耐热,最适宜温度为 10~16℃,当温度超过 25℃时,奶牛出现呼吸急促、食欲不振,产奶量受到影响,因此可将第 1 主成分作为温度因子。在第 2 主成分中,日平均相对湿度较大且为负值、日最低能见度和日照时长较大且为正值,江苏地区能见度与天气良好相关性强,且产奶量与湿度呈负相关、与日照时长、能见度呈明显正相关,这是由于气温超过 25℃时,奶牛主要依靠蒸发散热,湿度越大,与蒸发面水汽压和空气水汽压之差越小,蒸发散热量越小,奶牛产生剧烈热应激,影响产奶量<sup>[23]</sup>,因此可将第 2 主成分作为天气因子。在第 3 主成分中,日平均风速和日累计降水量较大且为负值,可见产奶量与风速、降水量呈负相关,这是因为降水和大风过程常伴随降温、增湿,湿冷的天气易导致产奶量下降,因此可将第 3 主成分作为雨量因子。

表 2 泗洪站气象因素主成分特征向量计算结果

气象因素	载荷值		
	第 1 主成分(F <sub>1</sub> )	第 2 主成分(F <sub>2</sub> )	第 3 主成分(F <sub>3</sub> )
日平均温度(x <sub>1</sub> )	-0.501 8	-0.041 7	0.055 7
日最低温度(x <sub>2</sub> )	-0.487 3	-0.124 4	-0.011 9
日最高温度(x <sub>3</sub> )	-0.497 3	0.020 8	0.108 3
日平均相对湿度(x <sub>4</sub> )	-0.009 2	-0.620 3	0.028 9
日平均风速(x <sub>5</sub> )	-0.077 3	0.202 0	-0.766 2
日累计降水量(x <sub>6</sub> )	-0.061 0	-0.286 7	-0.541 6
日最低能见度(x <sub>7</sub> )	-0.076 1	0.518 1	-0.194 0
日照时长(x <sub>8</sub> )	-0.164 7	0.448 0	0.256 2
气压(x <sub>9</sub> )	0.469 9	0.073 4	0.024 4

为对比传统方法,本研究同时基于 SPSS 22 进行产奶量与气象因素相关分析。由表 3 可知,温度、湿度、降水量、能见度与产奶量的相关关系与主成分分析法所得结果一致,但相关分析中日照时长、气压与产奶量无明显相关性,这与主成分分析得出的结果不一致,分析原因可知相关分析割裂了日照时长、气压与天气、温度的协同关系,而仅仅考虑了产奶量与其相关关系,这与实际情况不符;主成分分析法充分考虑了气象因素之间的关系,分析结果更符合实际情况。根据 3 个主成分的特征向量,可得主成分与气象因素的方程:

$$F_1 = -0.501\ 8x_1 - 0.487\ 3x_2 - 0.497\ 3x_3 - 0.009\ 2x_4 - 0.077\ 3x_5 - 0.061\ 0x_6 - 0.076\ 1x_7 - 0.164\ 7x_8 + 0.469\ 9x_9; \tag{10}$$

$$F_2 = -0.041\ 7x_1 - 0.124\ 4x_2 + 0.020\ 8x_3 - 0.620\ 3x_4 + 0.202\ 0x_5 - 0.286\ 7x_6 + 0.518\ 1x_7 + 0.448\ 0x_8 + 0.073\ 4x_9; \tag{11}$$

$$F_3 = 0.055\ 7x_1 - 0.011\ 9x_2 + 0.108\ 3x_3 + 0.028\ 9x_4 - 0.766\ 2x_5 - 0.541\ 6x_6 - 0.194\ 0x_7 - 0.256\ 2x_8 + 0.024\ 4x_9. \tag{12}$$

表 3 泗洪牧场产奶量与气象因素相关分析结果

指标	平均温度	日最低温度	日最高温度	日平均相对湿度	日平均风速	日累计降水量	日最低能见度	日照时长	气压
Pearson 相关系数	-0.133 **	-0.170 **	-0.111 **	-0.376 **	0.303 **	-0.086 *	0.182 **	-0.024	-0.049
P 值(双尾)	0.000	0.000	0.002	0.000	0.000	0.017	0.000	0.513	0.174

注: \*\*、\* 分别表示在置信度(双测)0.01、0.05 水平上显著相关。

2.2 考虑气象因素下的产奶量预测

本研究构建考虑气象因素下基于 BP 神经网络的产奶量预测模型,为排除时间因素对预测模型的影响,改进样本选取方法。首先将 775 个样本数据随机打乱顺序,然后选取其中 700 个作为训练样本,75 个作为检验样本。模型输入层为经主成分分析后的气象因素,输出层为产奶量,隐含层节点数由经验公式计算并经试凑法验证取 19 个节点。为防止过度训练,在设置训练参数时,迭代次数设定为 1 000,学习率为 0.01,以均方根误差(root mean square error,简称 RMSE)作为训练停止指标,当 RMSE 小于 0.01 时停止训练,完成 BP 神经网络模

型的建模,程序运行于 MATLAB R2014a 中,得到训练样本模拟结果和检验样本预测结果。

由图 2 可知,训练集、验证集和测试集的计算值与目标值之间的相关系数为 0.78、0.67、0.73,700 组训练样本相关系数达 0.75,训练结果优良。由图 3 可知,在检验样本中,考虑部分误差较大值,相对误差≤20% 共有 71 个检验样本,占比 94.67%,相对误差≤10% 共有 48 个检验样本,占比 64%,75 个检验样本的预测值与真实值的决定系数达 0.65,相对误差达 9.48%。由此可知,构建的产奶量预测模型结果精确性、稳定性均较好,同时操作性优良。

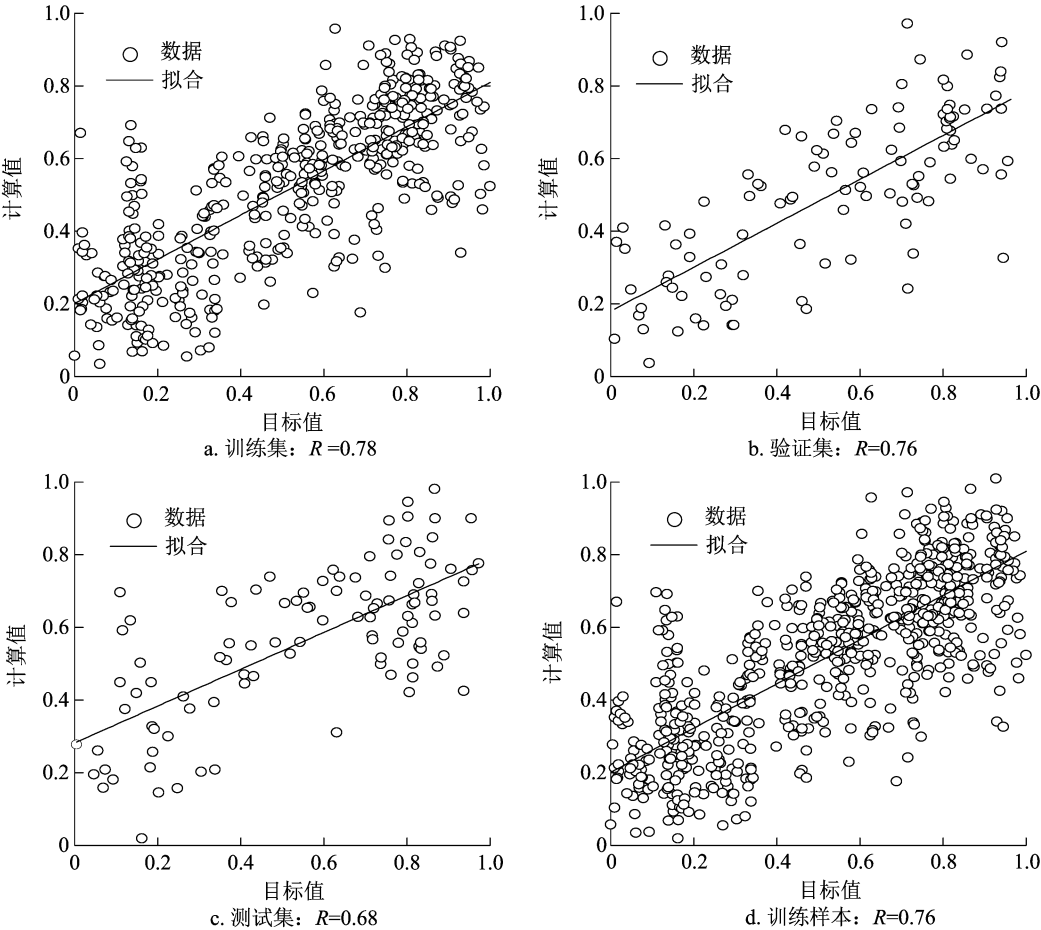


图2 基于 BP 神经网络的考虑气象因素的产奶量训练结果

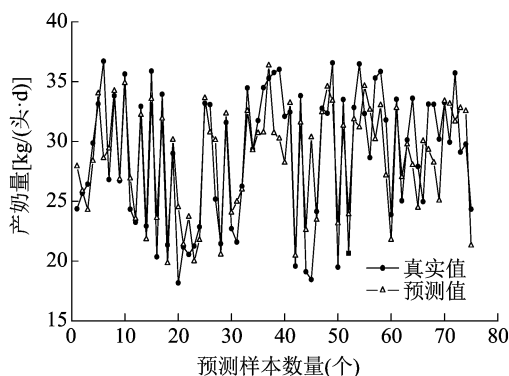


图3 基于 BP 神经网络的考虑气象因素的产奶量预测结果

### 3 结论与展望

本研究利用主成分分析法研究气象因素对产奶量的影响,并与传统相关分析结果比较,确定影响江苏地区产奶量的主要气象因素和关键时期,最终得到 3 个主成分因子,分别为日平均温度、日最低温度、日最高温度和气压所表征的第 1 主成分(温度因子),日平均相对湿度、日最低能见度和日照时长所表征的第 2 主成分(天气因子),日平均风速和日累计降水量所表征的第 3 主成分(雨量因子);产奶量与温度、湿度呈明显负相关,与风速、降水量呈负相关,与日照时长、能见度呈明显正相关;与传统相关分析法相比,主成分分析法更充分考虑气象因素的协同关系,分析结果更符合实际情况。基于气象因素与产奶量的关系,构建考虑气象因素下基于 BP 神经网络的产奶量预测模型,模型预测结果显示检验样本平均相对误差达 9.48%,考虑部分误差较大值,产奶量预测模型精确性、稳定性均较好。

基于主成分分析法确定影响江苏地区荷斯坦奶牛产奶量的主要气象因素和关键时期,在实际生产中,可由牧场管理人员、工作人员与气象服务部门交流对接,及时掌握影响关键时期和转折天气,更好制定和采取奶牛管理和牛奶生产的举措。基于 BP 神经网络的产奶量预测模型中,今后研究除考虑气象因素外,同时可考虑人工因素和管理因素,进一步提高预测准确性。

#### 参考文献:

- [1] 孙宏进. 推动江苏奶业高质量发展[J]. 中国畜牧业, 2019(14): 40-41.
- [2] 朱坤, 刘佳佳, 查满千, 等. 不同速度的音乐对奶牛血液激素水平 and 产奶量的影响[J]. 江苏农业科学, 2017, 45(14): 123-

125.

- [3] 马捷琼, 刘缠民, 陈宏, 等. 徐州荷斯坦牛 *NRAMP1* 基因多态性与产奶量途径分析[J]. 江苏农业科学, 2007(1): 116-118.
- [4] 马燕芬. 影响奶牛产奶量的因素[J]. 农家顾问, 2008(7): 48-49.
- [5] 姚焰础, 江山, 肖融, 等. 热应激对荷斯坦奶牛体温和呼吸的影响[J]. 中国畜牧杂志, 2012, 48(11): 59-62, 75.
- [6] 张志登, 敬盈嘉, 刘影, 等. 热应激对泌乳奶牛生理指标及生产性能的影响[J]. 中国畜牧杂志, 2020, 56(4): 169-172.
- [7] Trajchev M, Nakov D, Andonov S. The effect of thermal environment on daily milk yield of dairy cows[J]. Macedonian Veterinary Review, 2016, 39(2): 185-192.
- [8] West J W, Mullinix B G, Bernard J K. Effects of hot, humid weather on milk temperature, dry matter intake, and milk yield of lactating dairy cows[J]. Journal of Dairy Science, 2003, 86(1): 232-242.
- [9] 魏学占, 唐红艳, 杨忠霞, 等. 气象条件对奶牛产奶量的影响及对策[J]. 内蒙古气象, 2009(3): 23-26.
- [10] 陈志英, 阴妮, 池相河. 气象因素对黑龙江省奶牛产奶量影响分析[J]. 现代畜牧兽医, 2017(7): 44-48.
- [11] 王雅春, 陈幼春, 柏荣, 等. 奶牛泌乳曲线的拟合及其模型参数的遗传分析[J]. 畜牧兽医学报, 1999, 30(5): 399-404.
- [12] 张文龙. 新疆三个品种牛泌乳曲线的分析及产奶量校正系数的制定[D]. 乌鲁木齐: 新疆农业大学, 2013.
- [13] Kocak O, Kiz B. Comparison of different lactation curve models in Holstein cows raised on a farm in the south-eastern Anatolia region[J]. Archiv fur Tierzucht, 2008, 51(4): 329-337.
- [14] Otwinowska-Mindur A, Ptak E, Jagusiak W, et al. Modeling lactation curves of Polish Holstein-Friesian cows. Part I: The accuracy of five lactation curve models[J]. Journal of Animal and Feed Ence, 2013, 22(1): 19-25.
- [15] 李欣, 温万, 王新燕, 等. 基于 ARIMA 模型对宁夏地区奶牛月平均产奶量的预测[J]. 黑龙江畜牧兽医, 2016(23): 99-102.
- [16] 张鹏. 基于主成分分析的综合评价研究[D]. 南京: 南京理工大学, 2004.
- [17] 王新峰, 刘云芳, 李明强, 等. 主成分分析在奶牛产奶量选择中的应用[J]. 中国草食动物, 2002, 22(1): 8-9.
- [18] 陆呈委, 邓廷贤, 谭正淮, 等. 奶水牛乳房线性性状主成分分析研究[J]. 中国牛业科学, 2016, 42(2): 1-3.
- [19] 刘彩红. BP 神经网络学习算法的研究[J]. 西安工业大学学报, 2012, 32(9): 723-727.
- [20] 张永超, 赵录怀, 王昊, 等. 基于环境气体信息的 BP 神经网络苹果贮藏品质预测[J]. 江苏农业学报, 2020, 36(1): 194-198.
- [21] 李友坤. BP 神经网络的研究分析及改进应用[D]. 淮南: 安徽理工大学, 2012.
- [22] 蔡荣辉, 崔雨轩, 薛培静. 三层 BP 神经网络隐层节点数确定方法探究[J]. 电脑与信息技术, 2017, 25(5): 29-33.
- [23] 古少鹏, 高斌战, 袁建琴, 等. 温度风速对奶牛产乳量的影响[J]. 当代畜牧, 2001(4): 10-11.