

基于随机森林的寒区奶牛舍环境因素与产奶量关系研究

丁涛¹ 蒋雷生^{1,2} 施正香^{1,3} 赵洋^{1,2} 马慧⁴

(1. 中国农业大学水利与土木工程学院, 北京 100083; 2. 北京市供水管网与安全节能中心, 北京 100083;
3. 农业农村部设施农业工程重点实验室, 北京 100083; 4. 北京首农畜牧发展有限公司, 北京 100083)

摘要: 寒区奶牛长时间处于非热应激状态, 为了了解此阶段影响产奶量的主要环境因素, 本文根据连续监测的环境数据, 从日均值、百分位值角度研究了牛舍内温度、相对湿度、风速、二氧化碳浓度、氨气浓度、光照强度与奶牛每日头均产奶量之间的关系。同时, 以产奶量为预测目标建立了基于环境因素的随机森林回归模型。结果表明光照强度与二氧化碳浓度是影响产奶量两个较为重要的环境因素, 尤其是在低温时对奶牛生产影响较大。光照强度日均值 250 lx 与二氧化碳浓度日均值 8×10^{-4} 可以较为明显地划分高低产奶量。高产奶量样本点同样集中在光照强度第 90 百分位数高于 800 lx、二氧化碳浓度第 10 百分位数低于 6×10^{-4} 的区域。用于评价回归模型泛化能力的平均决定系数为 0.731 6, 平均均方根误差为 1.037 0 kg。根据结果建议寒区奶牛舍非热应激时期, 至少保证每日不低于 800 lx 的光照 2.5 h, 同时控制二氧化碳浓度高于 6×10^{-4} 的时间不超过 2.33 h。

关键词: 寒区奶牛舍; 环境因素; 非热应激; 产奶量; 随机森林

中图分类号: S811.5 文献标识码: A 文章编号: 1000-1298 (2021) 12-0378-08 OSID: 

Relationship between Environmental Factors and Milk Yield of Dairy Cow House in Cold Region Based on Random Forest

DING Tao¹ JIANG Leisheng^{1,2} SHI Zhengxiang^{1,3} ZHAO Yang^{1,2} MA Hui⁴

(1. College of Water Resources and Civil Engineering, China Agricultural University, Beijing 100083, China
2. Beijing Engineering Research Center of Safety and Energy Saving Technology for Water Supply Network System, Beijing 100083, China
3. Key Laboratory of Agricultural Engineering in Structure and Environment, Ministry of Agriculture and Rural Affairs, Beijing 100083, China
4. Beijing Shounong Animal Husbandry Development Co., Ltd., Beijing 100083, China)

Abstract: Dairy cows in cold regions are in non-heat stress state for a long time. Aiming to understand the main environmental factors affecting milk yield, the relationship between temperature, humidity, wind speed, carbon dioxide concentration, ammonia concentration, light intensity and average daily milk yield of dairy cows was studied from the perspective of daily mean and percentile values based on continuous monitoring environmental data. Meanwhile, a random forest regression model based on environmental factors was established to predict milk yield. The results showed that light intensity and carbon dioxide concentration were two important environmental factors affecting milk yield, especially at low temperature. The average daily light intensity of 250 lx and the average daily carbon dioxide concentration of 8×10^{-4} could obviously classify high and low milk yield. The sample sites with high milk yield were also concentrated in the areas where the 90th percentile of light intensity was higher than 800 lx and the 10th percentile of carbon dioxide concentration was lower than 6×10^{-4} . The mean determination coefficient and mean root mean square error used to evaluate the generalization ability of regression model were 0.731 6 and 1.037 0 kg, respectively. According to the results, it was suggested that during the non-heat stress period of dairy barns in cold areas, at least 2.5 h of light should be guaranteed for no less than 800 lx per day, and the time when carbon dioxide concentration was higher than 6×10^{-4} should be controlled for no more than 2.33 h.

Key words: dairy cow house in cold area; environmental factors; non-heat stress; milk yield; random forest

0 引言

奶牛舍环境对奶牛的生产性能具有决定性影响,当牛舍小气候变化幅度超出奶牛的适应范围时,奶牛的生产力和健康便会受到负面影响^[1]。目前,评价牛舍环境的指标大部分是基于热环境的热应激评价方法^[2],如 THI^[3]、BGHI^[4]、THVI^[5]、HLI^[6]以及不同的体感温度表达式^[7-9],这些指数关注了温度、相对湿度、风速和热辐射,却忽略了光照强度及空气质量的影响,不适用于非热应激季节的奶牛舍环境评估。由于季节特殊性,冬季舍内风速低、有害气体浓度高,特别是二氧化碳浓度和氨气浓度对奶牛的健康和生产有重要影响。研究表明^[10-12],光照使产奶量最大化的明暗交替最佳节点是 16 h 光照,8 h 黑暗,合理的光照能提高产奶量 8% ~ 10%、促进产后发情和受孕。然而针对寒区的气候特点,光照对于奶牛生产的影响没有更加深入地研究。

随机森林算法是数据挖掘分类算法的重要组成部分,在生物信息学、生态学、医学等方面都被广泛应用^[13-14]。针对畜禽舍环境的调控复杂性,国内外已经有采用模糊算法、神经网络等手段实现舍内环境的智能调节^[15-17]。有学者利用奶牛数据和机器学习算法预测奶牛泌乳早期的代谢状态,其中随机森林算法和支持向量机的表现最优^[18]。本文通过对奶牛舍环境参数以及产奶量的连续监测,从不同角度分析其中的变化规律,建立基于环境因素的随机森林回归预测模型,并验证其在寒区奶牛舍的适用性。

1 材料与方法

1.1 数据集

试验于 2018 年 9 月至 2019 年 3 月在吉林省白城市首都农业集团白城牛场进行。牛舍东西长度约 250 m,南北跨度约 27 m,屋顶为彩钢夹芯板双坡屋顶,奶牛采用全舍饲散栏饲养模式,舍内奶牛约 580 头,每日挤奶 3 班。牛舍四周窗户除冬季外均

开启,屋顶小窗常年开启,南面屋顶上每隔 12 m 均匀布置有采光板增加舍内采光,牛舍内无喷淋降温设施。舍内沿东西方向共有 4 排工字钢,东西方向相邻工字钢间隔 6 m。在中间两排工字钢上间隔 12 m 安装 36 寸风机,在躺卧区域上方间隔 15 m 安装 54 寸风机,舍内共安装有 40 台 36 寸风机、32 台 54 寸风机。风机在夏季全部 24 h 运行,秋季白天部分风机在高温时段对称开启,晚上关闭。风机运行个数随温度降低而减少直至全部停止运行,试验期间舍内大部分时间都处于风速较低状态。

试验牛舍由于跨度不大、采光板的设置及非热应激时期太阳高度角较小,前期通过手持 testo540 型光照度测试仪测试奶牛活动区域,结果表明南北光照强度差异不大。牛舍光照为自然光照,舍内在高度 4 m 处间隔 6 m 安装 15 W 节能灯,共安装 80 个节能灯,如图 1 所示,夜间开启节能灯便于人员工作。为了避免距离风机过近数据不准确,因此将传感器安装在相邻两个风机中间的工字钢上,距离地面 1.8 m。舍内各项设施布置均匀,舍外附近无其余建筑物,因此布置 4 个环境数据采集点可以较好地反映舍内环境水平。测点分别距离奶牛舍东门、西门 45 m 及 90 m,舍外布置 1 个采集点 E,位于奶牛舍舍外东南角,如图 2 所示。每个采集点可以采集温度、相对湿度、风速、二氧化碳浓度、氨气浓度、光照强度。传感器为武汉中科能慧公司生产,采样间隔为 10 min,具体参数见表 1。选择泌乳性能相近的 30 头奶牛每日产奶量进行分析。



图 1 试验牛舍内设施布置
Fig. 1 Layout of facilities in experimental cattle house

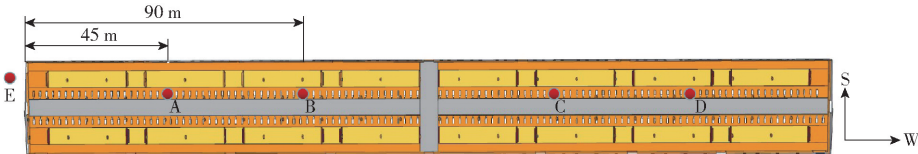


图 2 舍内外环境参数传感器测点布置

Fig. 2 Sampling locations of automatic sensors to measure indoor and outdoor environment parameters

1.2 数据处理

采集的数据往往存在部分缺失、不正确、含有噪声或其他类型的不一致现象^[19]。采用删除法与插补法处理数据缺失的情况,同时利用 Matlab 进行小

波去噪以减少噪声对数据的影响。考虑到环境因素的变化达到一定的阈值才会对畜禽的生产能力产生影响^[20],为了确定当日环境数据的代表值,本文将数据从小到大进行排列取不同百分位值及均值,研

表 1 试验传感器参数
Tab.1 Test sensor parameter list

试验仪器	型号	监测项目	原理	量程	分辨率	准确度
温湿风多要素传感器	NHFSX49 型	温度	能隙式测温	-40 ~ 80℃	0.1℃	±0.2 FS
		相对湿度	电容聚合体测湿	0 ~ 100%	0.1%	±3% FS
		风速	超声波	0 ~ 60 m/s	0.01 m/s	±2% FS
氨气传感器	NH63 型	氨气浓度	电化学反应	$0 \sim 1 \times 10^{-4}$	1×10^{-6}	±3% FS
二氧化碳传感器	NH162 型	二氧化碳浓度	红外检测	$0 \sim 2 \times 10^{-3}$	1×10^{-5}	$\pm 4 \times 10^{-5} \pm 3\% \text{ rdg}$
光照传感器	NHSD10AU2 型	光照强度	热电效应	0 ~ 2 000 lx	10 lx	10dgt ±5% rdg

究其与产奶量的相关关系。

1.3 机器学习

随机森林算法通过自助法在原数据集中有放回地随机抽取 n 个样本集,每次抽取 2/3 样本容量的数据作为袋内数据,建立 m 棵决策树构建随机森林,计算 m 棵决策树回归结果的平均值进行预测^[21]。该方法由于引入“双随机”的思想与组合投票取均值原则,因此对噪声和异常值不敏感、不易陷入过拟合^[22]。

1.3.1 数据预处理

构建机器学习模型时,使用样本的平均误差最小化作为模型的学习准则,则少数类样本的误差在总体误差中的比重会偏小,使得模型对少数类样本的预测性能下降。本文采用合成少数类过采样技术方法,利用随机线性插值法,在两个距离较近的少数样本间创造新的少数类样本来平衡样本。由于各环境因子的取值方式对产奶量的影响程度尚不明确,因此将各环境因子的每日平均值、每日 12 个最大值平均、每日 12 个最小值平均以及第 10、20、35、50、65、80、90 百分位数组成基础自变量集 X ,将产奶量作为基础因变量集 Y 。对自变量和因变量分别采取平方根变换、平方变换、立方变换、自然对数变换、指数变换、倒数变换。经过变换后形成 7 个训练集,每个训练集包含 420 个自变量。

1.3.2 模型评价指标

模型训练的目标是实现对训练样本的最佳拟合,是基于误差最小化原则,确定模型最优参数与结构的过程。选用决定系数(R^2)和均方根误差(E)作为模型评价指标,同时将两者合并为指标 P 以便对比模型性能,其计算公式为

$$P = \frac{1}{2} \left(\frac{R_i^2 - R_0^2}{R_0^2} + \frac{E_0 - E_i}{E_0} \right) \times 100\% \quad (1)$$

式中 R_i^2 ——剔除某自变量后模型训练的决定系数
 R_0^2 ——剔除某自变量前模型训练的决定系数
 E_0 ——剔除某自变量前模型训练的均方根误差
 E_i ——剔除某自变量后模型训练的均方根误差

1.3.3 模型参数优选

随机森林回归模型的构建需要确定决策树数量,对 420 个自变量筛选降维,确定最佳训练集。根据各模型的袋外数据均方误差估计随决策树数量增长的变化曲线确定决策树数量为 120 棵^[23]。基于袋外数据估计的特征重要性,逐步剔除重要性较小的特征,并迭代训练随机森林模型,直到自变量的数量降低到目标个数。最后利用 R^2 和 E 对优选后的模型进行评价。

2 结果与讨论

2.1 环境因素日均值与产奶量之间的关系

试验期间每月的数据变化情况见表 2,目前研究表明奶牛舒适温度区间为 5 ~ 25℃,可以看出在 9—11 月舍内温度处于奶牛舒适的温度范围内时,相对湿度、风速、光照强度、二氧化碳浓度、氨气浓度、产奶量均存在显著性差异。9 月和次年 1 月产奶量近似,期间除二氧化碳浓度外,其他因素均显著变化。2 月和 3 月产奶量近似,除二氧化碳浓度与风速外,其他因素均存在显著性差异。表明在部分时段里二氧化碳浓度不存在显著性差异时,其他环境因素的变化对产奶量没有产生明显地促进或抑制作用。

为了减少自变量从而准确分析其他因素与产奶量的规律,引入温湿指数 (Temperature humidity index, THI),温湿指数计算式为

$$THI = 0.8T + (H/100)(T - 14.4) + 46.4 \quad (2)$$

式中 T ——干球温度 H ——相对湿度

温湿指数与产奶量的关系如图 3 所示。由图 3 可知,THI 与产奶量间并没有明显的变化规律。分析其原因,在 9、10 月舍内风机随温度变化决定开启数量,风机运行影响了舍内气体分布。当温度较低时,牛舍内风速处于较低水平,有害气体浓度增加,光照强度也处于较高的水平。不同环境因素在不同时间段对奶牛的影响不同,因此在同一 THI 水平及不同 THI 水平时,产奶量动态变化。为了确定不同区间内影响产奶量的重要因

表 2 试验期间各月环境数据

Tab.2 Monthly environmental data during test period

月份	温度/℃	相对湿度/%	风速/(m·s ⁻¹)	光照强度/lx	二氧化碳浓度	氨气浓度	产奶量/kg
9 月	(18.8±3.8) ^a	(73±11) ^a	(1.59±0.43) ^a	(166±76) ^a	(61.2±6.4)×10 ^{-5ab}	(3.28±0.75)×10 ^{-6a}	(28.31±1.97) ^a
10 月	(13.0±3.3) ^b	(53±7) ^b	(0.43±0.41) ^b	(215±52) ^b	(70.9±7.5)×10 ^{-5b}	(3.73±0.52)×10 ^{-6ab}	(27.31±1.61) ^b
11 月	(5.8±3.2) ^b	(59±5) ^c	(0.09±0.07) ^c	(136±40) ^{ac}	(84.3±13.3)×10 ^{-5c}	(3.96±0.60)×10 ^{-6bc}	(25.77±1.70) ^c
12 月	(2.6±3.5) ^d	(73±6) ^a	(0.03±0.02) ^c	(112±19) ^c	(97.1±25.5)×10 ^{-5d}	(4.39±0.68)×10 ^{-6bcd}	(26.17±0.47) ^c
次年 1 月	(2.7±2.3) ^d	(71±5) ^{ad}	(0.09±0.03) ^e	(258±37) ^d	(57.2±10.5)×10 ^{-5a}	(4.00±0.76)×10 ^{-6bc}	(28.86±0.66) ^{ad}
次年 2 月	(5.6±4.3) ^e	(67±7) ^d	(0.06±0.03) ^e	(323±98) ^e	(61.4±14.6)×10 ^{-5ab}	(4.68±1.10)×10 ^{-6d}	(29.23±0.56) ^d
次年 3 月	(8.5±6.0) ^e	(52±6) ^b	(0.16±0.06) ^c	(419±30) ^f	(65.6±9.7)×10 ^{-5ab}	(4.46±1.69)×10 ^{-6cd}	(29.66±1.13) ^d

注：同一列数据不同字母表示差异显著($P<0.01$)，数据形式为均值±标准差。

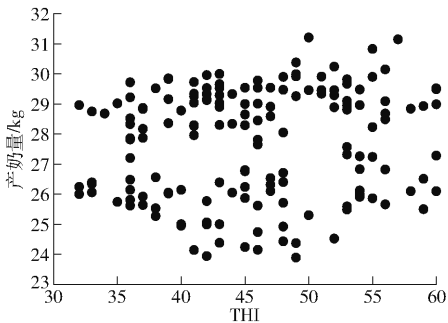


图 3 THI 与产奶量散点图

Fig. 3 Scatter diagram between THI and milk yield

表 3 不同 THI 与产奶量阶段时光照强度与二氧化碳浓度关系

Tab.3 Relationship between illumination intensity and carbon dioxide concentration at different THI and milk yield stage

THI	高产奶量(28~30 kg)		中产奶量(26~28 kg)		低产奶量(24~26 kg)	
	光照强度/lx	二氧化碳浓度	光照强度/lx	二氧化碳浓度	光照强度/lx	二氧化碳浓度
31~40	(260±74) ^{aA}	(5.39±1.14)×10 ^{-4aA}	(141±66) ^{aB}	(7.63±3.43)×10 ^{-4aB}	(117±5) ^{aB}	(9.55±1.67)×10 ^{-4aC}
41~50	(302±94) ^{aA}	(6.34±1.44)×10 ^{-4bA}	(133±36) ^{aB}	(8.99±2.34)×10 ^{-4aB}	(124±38) ^{aB}	(9.17±1.73)×10 ^{-4aB}
51~60	(279±111) ^{aA}	(6.97±0.81)×10 ^{-4bA}	(192±50) ^{bB}	(6.93±0.77)×10 ^{-4aA}	(220±92) ^{bB}	(6.93±1.07)×10 ^{-4bA}

注：同一列数据不同小写字母、同一因素不同列数据大写字母表示差异显著($P<0.01$)。

光照能抑制褪黑激素的分泌,其分泌的交替变化能引发一系列的激素反应,从而提升产奶量^[24]。PETERS 等^[25]研究表明接受 16 h 光照比 9~12 h 光照奶牛产奶量高 6.7%。ESPINOZA 等^[26]研究表明保持 16 h 光照相比 8 h 光照奶牛产奶量高 2.2 kg/d。REKSEN 等^[27]研究结果表明在高纬度地区,夜晚接受光照比未接受光照的奶牛产奶量高 0.5 kg/头。MUTHURAMALINGAM 等^[28]研究了不同光照强度对褪黑激素的影响水平,当夜晚光照强度为 50 lx 时,褪黑激素分泌会受到抑制。目前普遍认为光照强度 200 lx 有助于生产,各类研究明确了光照对产奶量的促进作用。由于各环境因素对奶牛的影响是动态的,本次试验主要分析正常生产状况下多个环境因素与产奶量之间关系,确定不同阶段各环境因素对产奶量的影响程度。夜间光照只用于照明,光照强度包含在每日环境因素日均值中,不影响本文分析结果。

东北地区昼夜温差大,冬季温度较低,牛舍的封

素,根据产奶量及 THI 划分为不同阶段进行分析,结果见表 3。

由表 3 可知,在 THI 处于较低和中等水平时,高产奶量和中产奶量间光照强度、二氧化碳浓度存在显著差异。中产奶量和低产奶量间光照强度不存在明显差异,二氧化碳浓度较高时产奶量低。THI 处于较高水平时,高产奶量和低产奶量样本间光照强度存在显著性差异,二氧化碳浓度不存在显著性差异。表明当二氧化碳浓度较低时产奶量较高,光照对产奶量有进一步的促进作用。

闭措施导致二氧化碳浓度过高,继而引起氧气浓度降低,造成奶牛缺氧^[29]。梅玮等^[30]研究表明牛舍中二氧化碳和氨气浓度过高将不利于奶牛健康和生产性能。目前关于牛舍内二氧化碳浓度的研究多是关于气体分布规律方面的,包括不同牛舍类型、不同季节等方面,并以二氧化碳浓度 1 500 mg/m³为标准进行舍内环境评价^[31-33]。

考虑到随着温度的改变,人工管理会影响舍内环境因素的分布规律。将观测样本划分为高风速与低风速时各环境因素的相关系数见表 4。可以看出,当温度较高时,风机运行的个数与舍内温度有关,因此风速与温度、相对湿度、二氧化碳浓度均存在明显的相关性,各因素与产奶量间存在相关性,但不显著。温度与产奶量相关系数较高,此时评价舍内环境时应以温度为主要因素。当温度较低时,舍内风速一直处于较为稳定的低水平,温湿度呈明显的线性规律。二氧化碳浓度与产奶量相关系数为 -0.532,光照强度与产奶量相关系数为 0.720,二

表 4 高低风速时各环境因素间相关系数

Tab.4 Correlation coefficient between high and low wind speed and various environmental factors

项目	低风速(<0.3 m/s)						
	风速	温度	相对湿度	二氧化碳浓度	氨气浓度	光照强度	产奶量
风速	1	0.261 *	-0.481 *	-0.289 *	-0.106	0.371 *	0.204
温度	0.772 *	1	-0.727 *	0.040	0.374 *	0.319 *	0.179
相对湿度	0.574 *	0.405	1	0.089	0.038	-0.353 *	-0.099
高风速(>0.3 m/s)	二氧化碳浓度	-0.670 *	-0.587 *	1	0.168	-0.524 *	-0.532 *
	氨气浓度	-0.002	0.169	-0.280	1	0.222 *	0.079
	光照强度	-0.188	-0.091	-0.818 *	0.154	1	0.720 *
	产奶量	0.478	0.411	0.108	-0.208	-0.042	1

注:对角线以上为低风速各环境相关系数,对角线以下为高风速各环境相关系数,*表示显著相关。

者均达到显著水平。温度、光照强度、二氧化碳浓度与产奶量的颜色映射图如图 4 所示,可以看出光照强度日均值 250 lx 与二氧化碳浓度日均值 8×10^{-4} 可以明显地划分高低产奶量。因此建议在低温时评价奶牛舍环境时要重点关注光照强度与二氧化碳浓度。

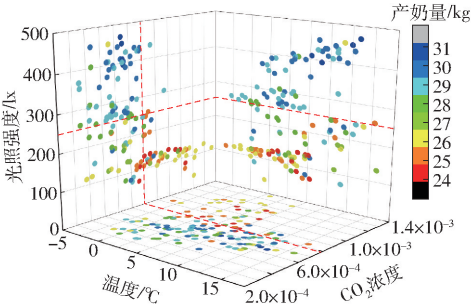


图 4 主要环境因素与产奶量关系

Fig.4 Relationship between major environmental factors and milk yield

表 5 光照强度与二氧化碳浓度非日均值与产奶量相关系数

Tab.5 Correlation coefficient between non-daily mean values of illumination intensity and carbon dioxide concentration and milk yield

百分位数	10	20	35	50	65	80	90
光照强度	0.257 3	0.255 6	0.282 6	0.158 0	0.505 5	0.577 4	0.607 4
二氧化碳浓度	-0.312 0	-0.297 5	-0.275 4	-0.242 6	-0.179 7	-0.104 4	-0.010 7

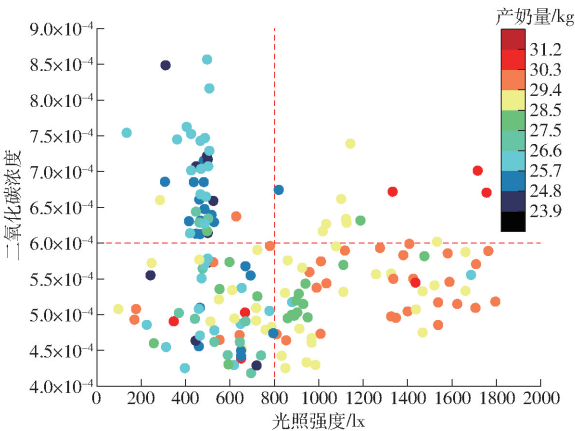


图 5 光照强度与二氧化碳浓度百分位值与产奶量关系

Fig.5 Relationship between percentage of illumination intensity and carbon dioxide concentration and milk yield

2.2 环境因素非日均值与产奶量之间的关系

不同百分位值二氧化碳浓度、光照强度与产奶量间的相关系数见表 5。可以看出二氧化碳浓度与产奶量的相关系数在所有百分位上均呈负相关,并且百分位数越小,两者的负相关程度越高。相反,光照强度与产奶量的相关系数在所有百分位上均呈正相关,且随着百分数增大,两者的正相关程度有增大的趋势。光照强度第 90 百分位数、二氧化碳浓度第 10 百分位数与产奶量关系如图 5 所示。可以看出,高产奶量样本集中在光照强度第 90 百分位数大于 800 lx 且二氧化碳浓度第 10 百分位数低于 6×10^{-4} 的区域。由于采样间隔为 10 min,大于等于第 90 百分位数的数据有 15 个,小于等于第 10 百分位数的数据有 14 个。因此为了提升产奶量,建议奶牛舍每天至少保证 2.5 h 不低于 800 lx 的光照时长。同时,建议控制二氧化碳浓度高于 6×10^{-4} 的时长不超过 2.33 h。

2.3 随机森林回归模型构建

2.3.1 自变量筛选降维

自变量筛选降维时要至少保留各个环境因素的一个自变量,剔除重要性较小的自变量后所训练的模型见表 6。模型编号 1~7 分别表示因变量未作任何变换、平方根变换、平方变换、立方变换、自然对数变换、指数变换、倒数变换所训练的模型;自变量栏中的大写字母 F、W、S、C、A、G 分别表示风速、温度、相对湿度、二氧化碳浓度、氨气浓度、光照强度;ave、max、min 分别代表日均值、每日 12 个最大值平均、每日 12 个最小值平均;数字代表其所在百分位数。可以看出,除了 X 和 1/X 两种形式外,其余形式的自变量均在自变量筛选过程中被剔除。这可能

表 6 筛选自变量后模型性能

Tab. 6 Model performance after filtering independent variables

模型编号	自变量	R^2	E/kg
1	$F_{\text{ave}}, G_{\text{ave}}, W_{\text{max}}, C_{\text{max}}, G_{\text{min}}, A_{10}, S_{65}, C_{65}, G_{65}, 1/G_{\text{ave}}, 1/C_{65}, 1/G_{65}$	0. 898 0	0. 650 9
2	$F_{\text{ave}}, G_{\text{ave}}, W_{\text{max}}, S_{\text{max}}, G_{\text{min}}, A_{50}, C_{65}, G_{65}, 1/G_{\text{ave}}, 1/C_{\text{min}}, 1/C_{65}, 1/G_{65}$	0. 896 2	0. 656 7
3	$F_{\text{ave}}, G_{\text{ave}}, C_{\text{max}}, G_{\text{max}}, G_{\text{min}}, S_{10}, A_{50}, S_{65}, C_{65}, G_{65}, W_{80}, 1/G_{\text{ave}}$	0. 901 2	0. 640 6
4	$F_{\text{ave}}, G_{\text{ave}}, C_{\text{max}}, G_{\text{min}}, C_{65}, G_{65}, W_{80}, S_{90}, 1/G_{\text{ave}}, 1/A_{50}, 1/C_{65}, 1/G_{65}$	0. 897 6	0. 652 1
5	$G_{\text{ave}}, W_{\text{max}}, S_{\text{min}}, G_{\text{min}}, F_{20}, A_{50}, C_{65}, G_{65}, 1/G_{\text{ave}}, 1/S_{10}, 1/C_{65}, 1/G_{65}$	0. 896 6	0. 655 4
6	$A_{\text{ave}}, G_{\text{ave}}, W_{\text{max}}, S_{\text{max}}, F_{\text{min}}, G_{\text{min}}, C_{65}, G_{65}, 1/G_{\text{ave}}, 1/G_{\text{max}}, 1/C_{\text{min}}, 1/G_{65}$	0. 741 7	1. 035 9
7	$F_{\text{ave}}, W_{\text{max}}, G_{\text{max}}, G_{\text{min}}, S_{10}, A_{20}, C_{65}, G_{65}, G_{80}, 1/G_{\text{ave}}, 1/S_{10}, 1/G_{65}$	0. 903 9	0. 631 7

是由于自变量在采用其他形式变换后线性关系不再明显,不利于决策树在分裂时找到合适的分裂点。同时可以看出在多个模型中光照强度、二氧化碳浓度、风速、温度出现次数较多,表明这 4 种环境因素是反映产奶量的重要自变量,其中光照尤为重要。

除了第 6 个模型性能明显劣于其他模型外,其余 6 个模型性能相近。说明除了对 Y 进行自然指数变换, Y 的其他变换形式对随机森林回归模型的构建并无明显影响。这可能是由于自然指数变换会导致 Y 的变换值急剧增大从而导致模型精度降低。因此,本文选用未对因变量进行变换的模型进行后续分析。

表 7 逐步回归确定最佳模型参数

Tab. 7 Step by step regression to determine the best model process

自变量个数	剔除自变量	$P/\%$	R^2	E/kg	自变量个数	剔除自变量	$P/\%$	R^2	E/kg
24	A_{20}	1. 391 6	0. 911 8	0. 605 3	12	$1/C_{65}$	1. 598 9	0. 912 4	0. 603 2
23	S_{65}	0. 993 0	0. 910 6	0. 609 5	11	$1/S_{10}$	1. 499 5	0. 912 1	0. 604 2
22	C_{65}	1. 136 8	0. 911 0	0. 608 0	10	G_{65}	1. 375 5	0. 911 7	0. 605 5
21	S_{max}	1. 057 4	0. 910 8	0. 608 8	9	W_{80}	0. 935 1	0. 910 4	0. 610 1
20	$1/G_{\text{ave}}$	1. 374 2	0. 911 7	0. 605 5	8	G_{max}	0. 332 9	0. 908 6	0. 616 3
19	G_{ave}	1. 678 7	0. 912 6	0. 602 4	7	A_{10}	0. 051 4	0. 907 7	0. 619 2
18	F_{20}	1. 840 4	0. 913 1	0. 600 7	6	$1/A_{50}$	- 1. 130 4	0. 904 0	0. 631 3
17	S_{10}	1. 997 8	0. 913 6	0. 599 1	5	C_{max}	- 2. 948 2	0. 898 3	0. 649 9
16	W_{max}	1. 643 3	0. 912 5	0. 602 7	4	S_{min}	- 6. 999 6	0. 885 0	0. 691 1
15	S_{90}	1. 684 9	0. 912 7	0. 602 3	3	G_{min}	- 14. 872 9	0. 857 3	0. 769 8
14	A_{50}	1. 419 4	0. 911 9	0. 605 1	2	$1/C_{\text{min}}$	- 26. 291 5	0. 813 1	0. 881 1
13	$1/G_{65}$	1. 572 7	0. 912 3	0. 603 5	1	F_{ave}	- 49. 599 6	0. 709 2	1. 099 0

2. 3. 3 优选模型预测效果

将完整样本随机分为 4 个小样本集,选取 1 个作为验证集,其余作为训练集,进行交叉验证评定模型的泛化能力。图 6 为交叉验证过程中较好模型与较差模型的预测性能。全部模型对验证集预测值的平均决定系数 R^2 为 0. 731 6,平均均方根误差 E 为 1. 037 0 kg,说明模型预测值与实际值之间趋势基本一致。两种模型的性能差异主要来自个别样本的较大误差,尤其是对某些产奶量区间的预测精度较低,这可能是由于随机抽取训练集导致某些产奶量区间

2. 3. 2 确定最佳训练集

经过自变量筛选降维后,剩余预测性能较好的 6 个模型中共有 25 个不重复的自变量,将其作为新模型的自变量,以未变换的 Y 作为因变量进行性能评估。根据逐步回归的方法对剩余 25 个自变量进行逐步回归,回归结果见表 7。可以看出,在剩余 17 个自变量时,模型性能达到最高值,随后性能逐步降低,在剩余 6 个自变量时,模型性能低于 25 个自变量建立的预测模型,筛选到最后剩余 G_{80} 变量。考虑到在保证一定模型精度的前提下,尽可能删减自变量个数,因此选取 7 个自变量(G_{80} 、 F_{ave} 、 $1/C_{\text{min}}$ 、 G_{min} 、 S_{min} 、 G_{max} 、 $1/A_{50}$)建立最佳模型。

的样本量过少,模型对该区间预测能力降低所致。两个模型虽然性能上有所差异,不过整体趋势与实际值是一致的。

3 结论

(1)牛舍运行过程中气体分布会受到风机干扰,在高风速时舍内部分环境因素彼此显著相关,但与产奶量的相关性没有达到显著性水平,此时评价舍内环境时主要以温度、相对湿度、风速为主。在低风速时,光照强度日均值 250 lx 与二氧化碳浓

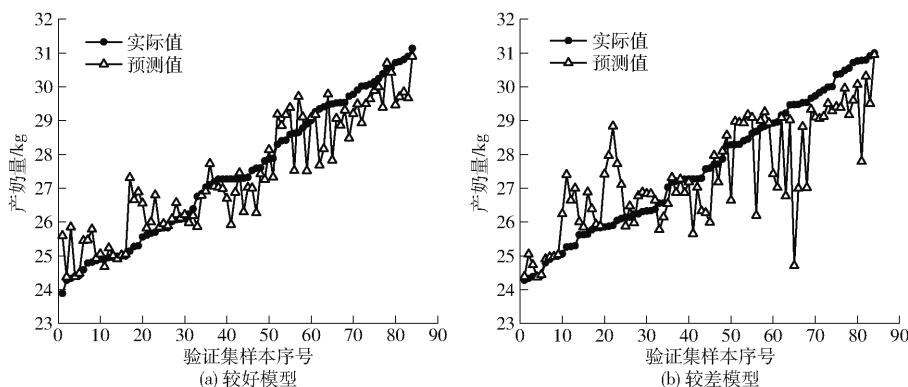


图6 模型预测性能

Fig. 6 Model predictive performance

度 8×10^{-4} 可以明显地划分高低产奶量。

(2) 二氧化碳浓度第 10 百分位数与光照强度第 90 百分位数与产奶量相关性最大, 为了提升产奶量, 建议奶牛舍每天至少保证不低于 800 lx 的光照时长 2.5 h。同时, 建议控制二氧化碳浓度高于 $6 \times$

10^{-4} 的时长不超过 2.33 h。

(3) 将产奶量作为预测指标, 利用随机森林方法建立了基于环境因素的产奶量回归模型。模型泛化能力检验结果表明其具有良好的预测能力。

参 考 文 献

- [1] 解世德. 关于牛舍的小气候(综述)[J]. 国外畜牧学(草食家畜), 1981(1):23-29.
XIE Shide. On the microclimate of the cowshed (review)[J]. Animal Science abroad (Herbivorous Livestock), 1981(1):23-29. (in Chinese)
- [2] 严格齐, 李浩, 施正香, 等. 奶牛热应激指数的研究现状及问题分析[J]. 农业工程学报, 2019, 35(23):226-233.
YAN Geqi, LI Hao, SHI Zhengxiang, et al. Research status and problem analysis of heat stress index in dairy cows[J]. Transactions of the CSAE, 2019, 35(23):226-233. (in Chinese)
- [3] THOM E C. The discomfort index[J]. Weatherwise, 1959, 12(2):57-61.
- [4] BUFFINGTON D E. Black-humidity index(BGHI) as comfort equation for dairy cows[J]. Transactions of the ASAE, 1981, 24(3):711-714.
- [5] 陶秀萍. 不同温湿度条件对肉鸡应激敏感生理生化指标影响的研究[D]. 北京:中国农业科学院, 2003.
TAO Xiuping. Effects of different temperature, humidity and wind conditions on stress sensitive physiological and biochemical indices of broilers[D]. Beijing: Chinese Academy of Agricultural Sciences, 2003. (in Chinese)
- [6] GAUGHAN J B, MADER T L, HOLT S M, et al. A new heat load index for feedlot cattle[J]. Journal of Animal Science, 2008, 86(1):226-234.
- [7] KHELIL-ARFA H, FAVERDIN P, BOUDON A. Effect of ambient temperature and sodium bicarbonate supplementation on water and electrolyte balances in dry and lactating Holstein cows[J]. Journal of Dairy Science, 2014, 97(4):2305-2318.
- [8] BIANCA W. Relative importance of dry- and wet-bulb temperatures in causing heat stress in cattle[J]. Nature, 1962, 195(4838):251-252.
- [9] BEAVER J M, OLSON B E, WRAITH J M. A simple index of standard operative temperature for mule deer and cattle in winter[J]. Journal of Thermal Biology, 1996, 21(5-6):345-352.
- [10] 李云甫, 杨和平, 尤彬, 等. 光照时间对奶牛泌乳和采食性能的影响[J]. 家畜生态学报, 2009, 30(4):37-39.
LI Yunfu, YANG Heping, YOU Bin, et al. Effects of light time on lactation and feeding ability of dairy cows[J]. Journal of Livestock Ecology, 2009, 30(4):37-39. (in Chinese)
- [11] 田允波. 光照对奶牛的影响研究[J]. 甘肃畜牧兽医, 1992(2):29-31.
TIAN Yunbo. Effects of light on dairy cows[J]. Gansu Animal Husbandry and Veterinarian, 1992(2):29-31. (in Chinese)
- [12] HANSEN P J, HAUSER E R. Photoperiodic alteration of postpartum reproductive function in suckled cows[J]. Theriogenology, 1984, 22(1):1-14.
- [13] 曹正凤. 随机森林算法优化研究[D]. 北京:首都经济贸易大学, 2014.
CAO Zhengfeng. Research on stochastic forest algorithm optimization[D]. Beijing: Capital University of Economics and Business, 2014. (in Chinese)
- [14] 师翊, 耿楠, 胡少军, 等. 基于随机森林回归算法的苹果树冠层光照分布模型[J/OL]. 农业机械学报, 2019, 50(5):214-222.
SHI Yi, GENG Nan, HU Shaojun, et al. Illumination distribution model of apple tree canopy based on random forest regression algorithm[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5):214-222. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190525&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.025. (in Chinese)
- [15] 宣传忠, 武佩, 马彦华, 等. 基于自适应模糊神经网络的畜禽舍环境控制系统研究[J]. 内蒙古大学学报, 2013, 44(4):397-403.

- XUAN Chuangzhong, WU Pei, MA Yanhua, et al. Research on environmental control system of livestock and poultry house based on adaptive fuzzy neural network[J]. Journal of Inner Mongolia University, 2013, 44(4):397–403. (in Chinese)
- [16] 刘冠延. 畜禽舍养殖环境智能管控系统设计与实现[D]. 哈尔滨:东北农业大学,2017.
- LIU Guanyan. Design and implementation of intelligent control system for livestock and poultry house breeding environment [D]. Harbin:Northeast Agricultural University, 2017. (in Chinese)
- [17] 何东健,刘建敏,熊虹婷,等. 基于改进 YOLO v3 模型的挤奶奶牛个体识别方法[J/OL]. 农业机械学报,2020,51(4):250–260.
- HE Dongjian, LIU Jianmin, XIONG Hongting, et al. Individual identification method of milking cows based on improved YOLO v3 model[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(4):250–260. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200429&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.04.029. (in Chinese)
- [18] FENLON C, O'GRADY L, DOHERTY M L, et al. The creation and evaluation of a model predicting the probability of conception in seasonal-calving, pasture-based dairy cows[J]. Journal of Dairy Science, 2017, 70(7):32.
- [19] 周英,卓金武,卞月青,等. 大数据挖掘:系统方法与实例分析[M]. 北京:机械工业出版社,2016.
- [20] FORBES J M. Effects of lighting pattern on growth, lactation and food intake of sheep, cattle and deer[J]. Livestock Production Science, 1982, 9(3):361–374.
- [21] 方匡南,吴见彬,朱建平,等. 随机森林方法研究综述[J]. 统计与信息论坛,2011,26(3):32–38.
- FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of random forest methods[J]. Statistics and Information Forum, 2011, 26(3):32–38. (in Chinese)
- [22] 牛志华. 基于 Spark 分布式平台的随机森林分类算法研究[D]. 天津:中国民航大学,2017.
- NIU Zhihua. Research on random forest classification algorithm based on Spark distributed platform[D]. Tianjin: Civil Aviation University of China, 2017. (in Chinese)
- [23] 任才溶. 基于并行随机森林的城市 PM2.5 浓度预测[D]. 太原:太原理工大学,2018.
- REN Cairong. Prediction of URBAN PM2.5 concentration based on parallel random forest[D]. Taiyuan:Taiyuan University of Technology, 2018. (in Chinese)
- [24] 邱伯根. 光照对奶牛生长发育和泌乳的影响[J]. 国外畜牧学(草原家畜),1988(3):36–39.
- QIU Bogen. Effects of light on growth, development and lactation of dairy cows[J]. Animal Science Abroad (Grassland Livestock), 1988(3):36–39. (in Chinese)
- [25] PETERS R R, CHAPIN L T, EMERY R S, et al. Milk yield, feed intake, prolactin, growth hormone, and glucocorticoid response of cows to supplemented light[J]. Journal of Dairy Science, 1981, 64(8):1671–1678.
- [26] ESPINOZA O S, MASAHITO O. Interaction effect of photoperiod management and dietary grain allocation on productivity of lactating dairy cows[J]. Canadian Journal of Animal Science, 2017, 97(3):517–525.
- [27] REKSEN O, TVERDAL A, LANDSVERK K, et al. Effects of photointensity and photoperiod on milk yield and reproductive performance of norwegian red cattle[J]. Journal of Dairy Science, 1999, 82(4):810.
- [28] MUTHURAMALINGAM P, KENNEDY A D, BERRY R J, et al. Plasma melatonin and insulin-like growth factor-1 responses to dim light at night in dairy heifers[J]. Munksgaard International Publishers, 2006, 40(3):225–229.
- [29] 张莹莹,刘春青,刘晓辉,等. 畜禽舍有害气体的来源及其常用消毒方法[J]. 饲料博览,2014(3):11–13.
- ZHANG Yingying, LIU Chunqing, LIU Xiaohui, et al. Sources of harmful gases in livestock and poultry houses and common disinfection methods[J]. Feed Review, 2014(3):11–13. (in Chinese)
- [30] 梅玮,祁巨中,刘英玉,等. 新疆北疆地区某集约化肉牛养殖场春季环境指标评价[J]. 中国畜牧兽医,2015,42(7):1738–1746.
- MEI Wei, QI Juzhong, LIU Yingyu, et al. Environmental index evaluation of an intensive beef cattle farm in North Xinjiang in spring[J]. Chinese Animal Husbandry and Veterinarian,2015,42(7):1738–1746. (in Chinese)
- [31] 王亚男,冯曼,李宏双,等. 坝上地区奶牛舍和犊牛舍冬季有害气体的监测[J]. 中国畜牧兽医,2016,43(6):1635–1640.
- WANG Yanan, FENG Man, LI Hongshuang, et al. Monitoring of harmful gases in dairy and calf barns in Bashang area in winter[J]. Chinese Animal Husbandry and Veterinarian,2016,43(6):1635–1640. (in Chinese)
- [32] 高玉红,李宏双,郭建军,等. 坝上寒区不同建筑类型奶牛舍冬季温热环境评价[J]. 东北农业大学学报,2015,46(12):52–57.
- GAO Yuhong, LI Hongshuang, GUO Jianjun, et al. Evaluation of thermal environment in different building types of dairy barn in Bashang Cold region in winter[J]. Journal of Northeast Agricultural University, 2015, 46(12):52–57. (in Chinese)
- [33] 孙斌. 环境温度与日粮对新疆荷斯坦干奶牛与泌乳牛舍内 CH₄、CO₂ 和 NH₃ 浓度变化的影响[D]. 乌鲁木齐:新疆农业大学,2013.
- SUN Bin. Effects of ambient temperature and diet on CH₄, CO₂ and NH₃ concentrations in Holstein dairy cows and lactating cows in Xinjiang[D]. Urumqi:Xinjiang Agricultural University,2013. (in Chinese)