

DOI: 10.13652/j.spjx.1003.5788.2024.60047

基于随机森林回归模型的小麦粉灰分含量快速测定

刘艳群¹ 肖付刚² 陈彩虹³

(1. 河南工业贸易职业学院, 河南 郑州 450000; 2. 许昌学院, 河南 许昌 461000; 3. 漯河市产品质量检验检测中心(国家肉制品监督检验中心(河南)), 河南 漯河 462000)

摘要: [目的] 实现小麦粉灰分含量的快速精准测定。[方法] 通过预处理小麦原料并深入分析磨粉时间和电导率等关键影响因素, 将这些因素作为特征变量引入随机森林回归模型, 构建小麦粉灰分含量测定模型。通过算术平均值计算得出最终测定结果, 实现小麦粉灰分含量的快速测定。[结果] 该方法与实际结果基本一致, 测定误差低于 0.01 g/100 g, 且重复性波动差距低于 0.01 g/100 g, 平均测定时间为 24 min。[结论] 试验方法具有较高的测量精度和重复性, 显著提升了测定效率。

关键词: 快速测定; 小麦粉; 随机森林回归模型; 灰分含量; 影响因素

Study on rapid determination method of ash content in wheat flour based on stochastic forest regression model

LIU Yanqun¹ XIAO Fugang² CHEN Caihong³

(1. Henan Industry and Trade Vocational College, Zhengzhou, Henan 450000, China; 2. Xuchang University, Xuchang, Henan 461000, China; 3. Luohe Product Quality Inspection and Testing Center, State Meat Product Supervision and Testing Center (Henan), Luohe, Henan 462000, China)

Abstract: [Objective] To achieve rapid and accurate determination of ash content in wheat flour. [Methods] By preprocessing wheat raw materials and analyzing key influencing factors such as milling time and conductivity in depth, these factors were introduced as characteristic variables into a random forest regression model to construct a wheat flour ash content determination model. The final determination result was obtained by calculating the arithmetic mean, achieving rapid determination of ash content in wheat flour. [Results] The method was basically consistent with the actual results, with a measurement error of less than 0.01 g/100 g and a repeatability fluctuation difference of less than 0.01 g/100 g. The average measurement time was 24 minutes. [Conclusion] The experimental method has high measurement accuracy and repeatability, significantly improving the measurement efficiency.

Keywords: rapid determination; wheat flour; stochastic forest regression model; ash content; influencing factors

小麦粉, 作为人类日常饮食的重要组成部分, 其品质优劣直接关系到人们的饮食安全和健康。灰分含量不仅反映了小麦粉中矿物质和微量元素的含量, 还能够一定程度上体现小麦粉的纯净度及其加工过程中的质量控制情况。常规情况下, 在小麦粉的生产和加工过程中, 灰分含量的控制对于确保产品质量至关重要^[1]。灰分含量过高可能导致食品口感粗糙、营养成分不均衡; 而灰分含量过低可能意味着小麦粉中的营养成分不足, 满足不了

人体的营养需求。因此, 小麦粉灰分含量测定的准确性和速度对于保障食品安全、提升食品品质、加强生产企业产品质量精细控制具有重要意义^[2]。

赵二霞等^[3]以小麦粉作为研究对象, 按照灰分含量检验规定, 通过 900 °C 灼烧和 550 °C 灼烧两种方法测定小麦粉灰分含量, 充分考察了灼烧温度对灰分含量测定的影响, 该方法对测试条件要求较高, 需多次确定最佳的测试条件, 致使灰分含量测定效率较低。王晓玉等^[4]以生物质

基金项目: 河南省科技攻关项目(编号: 242102321134)

通信作者: 刘艳群(1978—), 女, 河南工业贸易职业学院讲师, 硕士。E-mail: liuyanqun@hngm.edu.cn

收稿日期: 2024-04-20 改回日期: 2024-09-08

能源化利用作为研究目标,对其灰分含量测定方法进行探究与说明,以 2001—2021 年生物质样本灰分含量测定数据为基础,就高位热值对灰分含量测定结果的影响进行了分析,获得草本植物高位热值与灰分含量呈负相关的结论,通过近红外光谱技术实现了灰分含量的有效测定,该方法应用数据量过多,无法获得精准的灰分含量测定结果。孙晓荣等^[5]以小麦粉品质定量检测作为目标,应用近红外(NIR)光谱技术构建灰分含量的偏最小二乘模型,利用卷积平滑法获得最佳的建模效果,从而获得最精准的灰分含量测定结果;此方法重复性较差,致使灰分含量测定效率较低。郭歌等^[6]以生物质原料为研究对象,应用近红外光谱分析技术提出了一种新的灰分含量测定方法,以 5 个地点 10 种类型的 1 465 份生物质原料样品作为研究数据,通过“筛选分类集合法”划分样本集合,构建近红外光谱生物质样品灰分含量模型;该方法未考虑测定条件等影响因素,致使灰分含量测定精度差。

研究拟提出一种基于随机森林回归模型的小麦粉灰分含量快速测定方法。通过对小麦原料预处理,提取磨粉时间和电导率等特征变量,运用随机森林回归模型集成多个决策树输出,通过算术平均值计算灰分含量,以期实现小麦粉灰分含量的快速精准测定。

1 小麦粉灰分含量快速测定方法

1.1 试验仪器

清洗器:LON30-720C 型,上海比朗仪器制造有限公司;
离子分析仪:LD-FY1 型,山东莱恩德智能科技有限公司;
粉碎机:AEP01 型,山东埃尔派粉体科技股份有限公司;
标准筛:ST-200A 型,山东盛泰仪器有限公司;
电子天平:TCS 型,上海鹰牌衡器有限公司;
真空干燥箱:DZ-1BCIV 型,天津市泰斯特仪器有限公司;
电导率仪:DDB-305A 型,上海阔思电子有限公司;
马弗炉:GW1100-111 型,河北雅阁隆贸易有限公司;
干燥器:SD-YPG200 型,无锡市尚德干燥设备有限公司;
电炉:RT3-240-11 型,淄博正丰电炉有限公司;
离心机:TD4Z-WS 型,长沙湘锐离心机有限公司。

1.2 试验场景搭建

结合小麦粉灰分含量测定需求,搭建试验场景如图 1 所示。

1.3 小麦粉灰分含量测定程序

为了提升研究方法的重复性及应用普遍性,选取多种小麦原料制定测定程序。小麦原料选取结果见表 1。

依据灰分含量测定需求,对小麦原料进行预处理,主要分为两个环节:小麦清洗环节与小麦粉制备环节^[7-10]。

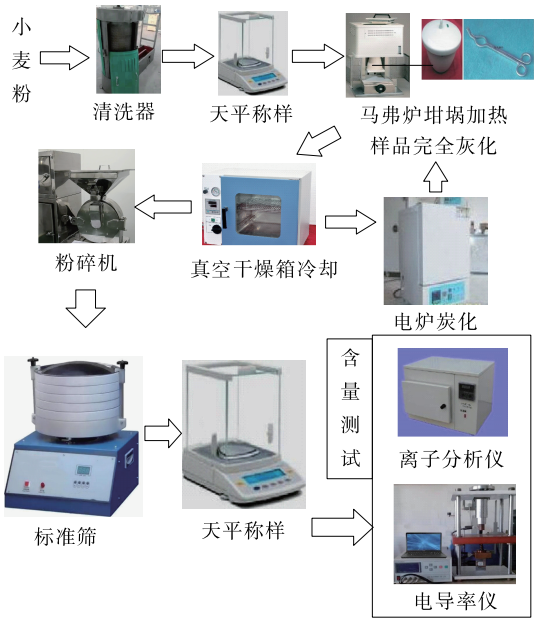


图 1 试验场景示意图

Figure 1 Schematic diagram of experimental scenario

表 1 小麦原料选取表

Table 1 Wheat raw material selection table

品种	特性
郑麦 98	半冬性, 优质强筋
太空 6 号	弱春性, 中筋
新麦 20	弱春性, 中筋
百农 AK58	半冬性中熟, 中筋
兰考矮早 8	弱春性, 中筋
豫麦 18	弱春性多穗型早熟, 中筋
豫麦 34	弱春性, 强筋 1 级
豫麦 49	半冬性, 强筋 2 级
豫麦 50	弱春性中早熟品种, 弱筋
豫麦 53	弱春性, 中筋
周麦 16	半冬性大穗型中熟, 中筋
新麦 18	半冬性中熟品种, 中筋
豫农 949	弱春性, 中筋
西农 979	半冬性中早熟优质强筋品种
洛麦 21	半冬性大穗型中熟品种, 中筋
新麦 208	弱春性, 中早熟, 优质中筋
周麦 19	半冬性大穗型强筋品种
豫麦 49-198	半冬性中熟, 中筋
花培 5 号	弱春性, 中筋
许科 1 号	半冬性中晚熟, 中筋

小麦清洗是小麦粉制备的首要步骤,其目的在于去除小麦表面的杂质和污染物,保证小麦的纯净度,小麦粉制备环节是从小麦中获取小麦粉的关键步骤。

以上述制备的小麦粉为基础,制定小麦粉灰分含量测定程序如图2所示。

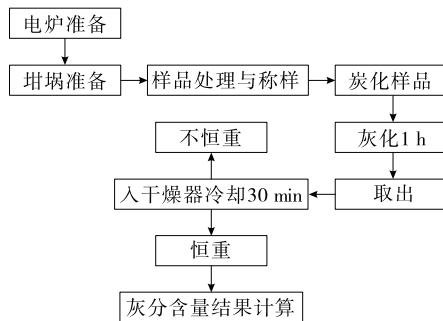


图2 小麦粉灰分含量测定程序

Figure 2 Procedure for determination of ash content in wheat flour

根据小麦粉样品质量和灰分质量,按式(1)计算灰分含量。

$$\zeta = \frac{M_1}{M_0} \quad (1)$$

式中:

ζ ——灰分含量, g/100 g;

M_1 ——小麦粉样品灰分质量, g;

M_0 ——小麦粉样品质量, g。

1.4 灰分含量影响因素分析

灰分含量测定过程中会受到多种因素的影响,尤其是磨粉时间和电导率,若忽视这些因素,势必会降低灰分含量检测测定的精度。磨粉时间是指将小麦研磨成粉所需的时间,其直接决定了小麦粉的细度和均匀性,进而对灰分含量检测产生影响^[11-13]。磨粉过程中,随着磨粉时间的延长,小麦粒逐渐被破碎并细化,麦麸等含灰分较高的部分与麦粉混合的机会增加。因此,实际生产中,需要根据小麦品种、品质以及产品要求等因素,合理控制磨粉时间,以获取理想的灰分含量和品质^[14]。

电导率是衡量物质导电能力的指标,在小麦粉生产过程中,电导率与灰分含量之间存在一定的关联。小麦中的灰分主要由无机盐组成,其在水中溶解后具有一定的导电性。通常情况下,灰分含量较高的小麦粉,其电导率也相对较高^[15]。电导率采用电导率仪进行测量,为了方便,需对电导率仪读数进行处理,表达式为:

$$X = \frac{K}{\alpha} \quad (2)$$

式中:

X ——真实电导率数值, mS/cm;

K ——电导率仪读数, mS/cm;

α ——小麦粉干基质量, g。

1.5 基于随机森林回归模型的灰分含量测定

随机森林回归模型作为一种集成学习方法,具有处

理特征数量巨大的数据、避免模型过拟合、提高预测精度等优势。通过引入随机森林回归模型,可以实现对小麦粉灰分含量的快速、精准测定。将磨粉时间和电导率作为特征变量,引入随机森林回归模型构建小麦粉灰分含量测定模型,从而获得精准的灰分含量测定结果。基于随机森林回归模型的小麦粉灰分含量测定模型如图3所示。每个决策树对应的特征变量——磨粉时间和电导率均不同,其获得的灰分含量也存在一定的差距。由此可见,单一决策树会受到特征变量数值的影响,致使最终的灰分含量测定结果精度无法保障。而构建模型应用了随机森林回归模型,在改变特征变量——磨粉时间和电导率的基础上,获取多个决策树的灰分含量测定结果,通过算术平均值计算方式获取最终的小麦粉灰分含量测定结果,能最大限度提升灰分含量的测定精度^[16]。

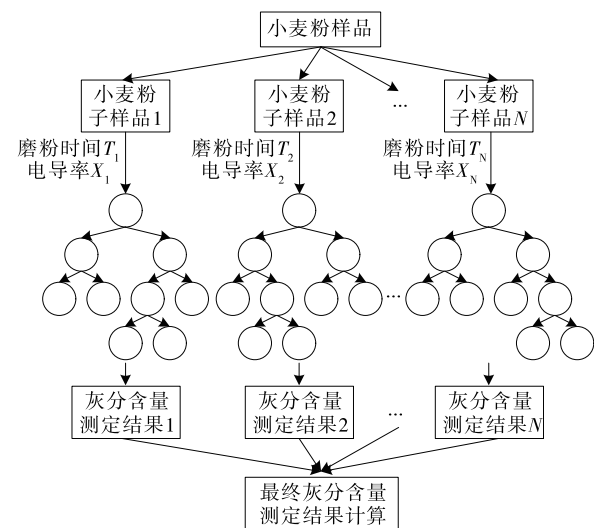


图3 小麦粉灰分含量测定模型示意图

Figure 3 Schematic diagram of ash content determination model for wheat flour

小麦粉灰分含量最终测定结果表达式为

$$\hat{\zeta} = \frac{1}{N} \sum_{i=1}^N \zeta_i \quad (3)$$

式中:

$\hat{\zeta}$ ——小麦粉灰分含量最终测定结果, g/100 g;

N ——随机森林回归模型中决策树的总数量;

ζ_i ——第*i*颗决策树输出的灰分含量测定结果, g/100 g。

2 结果与分析

2.1 试验方法比对

以文献[3]与文献[4]进行小麦粉灰分含量测定比对试验。随机在表1所示的小麦原料中抽取10种作为试验对象,将其制备成小麦粉,磨粉时间为15 min。利用电导率仪获得电导率试验数据见表2,各方法的灰分含量对比如图4所示。

表 2 小麦粉电导率

Table 2 Conductivity of wheat flour		
编号	小麦原料粉	电导率/(mS·cm ⁻¹)
1	百农 AK58	0.33
2	兰考矮早 8	0.52
3	豫麦 18	0.41
4	豫麦 34	0.61
5	豫麦 49	0.81
6	豫麦 50	0.72
7	豫麦 53	0.47
8	周麦 16	0.68
9	新麦 18	0.56
10	豫农 949	0.76

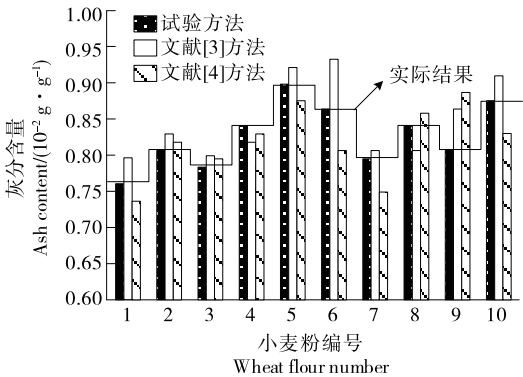


图 4 小麦粉灰分含量测定结果
Figure 4 Determination results of ash content in wheat flour

由图 4 可知,试验方法获得的小麦粉灰分含量测定结果与实际结果基本一致,测定误差 $<0.01\text{ g}/100\text{ g}$,且与小麦粉电导率的数值大小顺序一致。而采用文献[3]和文献[4]方法获得的小麦粉灰分含量测定结果与实际结果的偏差 $>0.1\text{ g}/100\text{ g}$,表明采用试验方法测定的小麦粉灰分含量更加精准。

2.2 波动程度重复性试验

在表 1 中随机抽取一种小麦原料作为试验对象,采用不同方法对其进行 10 次灰分含量测定,通过灰分含量测定结果波动程度来检验设计方法的重复性,小麦粉灰分含量测定结果如图 5 所示。

由图 5 可知,在同一试验对象进行 10 次灰分含量测定背景下,试验方法获得的小麦粉灰分含量测定结果非常接近,波动差距 $<0.01\text{ g}/100\text{ g}$,而对比文献[3]与文献[4]获得的小麦粉灰分含量测定结果之间的差距较大,表明试验方法的重复性更好,获得的测定结果更加可靠。

2.3 不同方法的测定时间

为了进一步验证试验方法的灰分含量测定效率,分别使用试验方法、文献[3]方法和文献[4]方法对小麦粉样品进行制备。选取 3 个不同批次的小麦粉样品,确保样品

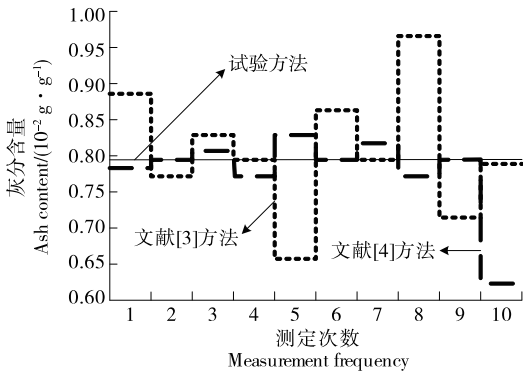


图 5 小麦粉灰分含量测定结果示意图
Figure 5 Schematic diagram of ash content determination results of wheat flour

的均匀性和一致性,并确保样品的处理过程一致。计算每种方法在不同批次下的测定时间见表 3。

表 3 不同方法测定时间对比

Table 3 Comparison of different methods for measuring time		min			
方法	批次 1	批次 2	批次 3	平均值	
试验方法	23	24	25	24	
文献[3]	33	34	35	34	
文献[4]	38	39	40	39	

由表 3 可知,试验方法在 3 个批次的小麦粉样品灰分含量测定中,平均测定时间为 24 min,明显短于文献[3]和文献[4]的,表明试验方法在测定效率上具有显著优势,能够更快速地完成灰分含量的测定。试验方法在不同批次之间的测定时间波动较小,为 23~25 min,显示出较好的稳定性。而文献[3]和文献[4]的测定时间波动范围较大,可能受到更多环境因素的影响。这是因为试验方法使用的随机森林回归模型对环境干扰因素具有较好的容忍度,能够在不同试验条件下保持稳定的性能,提高检测效率。

3 结论

利用随机森林回归模型测定的小麦粉灰分含量数据准确,测定结果与实际值差异较小,表明此方法的准确性高,适用于小麦粉中灰分含量的测定;试验方法的精密度较高,重复测定时波动差距 $<0.01\text{ g}/100\text{ g}$,方法稳定性优于对比方法;试验方法的检测周期明显优于其他对比方法,能够快速准确测定出小麦粉中灰分含量。试验使用的随机森林模型依赖于训练数据集的质量和多样性。如果实际应用中的小麦粉样本与训练集中的样本在品种、产地、加工方式等方面存在较大差异,可能会影响模型的检测准确性。后续将收集更多的小麦粉样本数据,包括

不同品种、产地和加工条件下的样本,以提高模型的泛化能力。

参考文献

- [1] 曹俊英, 包军鹏, 张倩倩, 等. 宁夏不同品种粳米食用品质与理化性质的相关性研究[J]. 粮食与油脂, 2022, 35(7): 29-32.
CAO J Y, BAO J P, ZHANG Q Q, et al. Study on the correlation between the edible quality and physicochemical properties of different varieties of japonica rice in Ningxia[J]. Cereals & Oils, 2022, 35(7): 29-32.
- [2] 张进萍, 何笑丛, 包杰. 不同粉路面粉灰分及其品质特性的关联性研究[J]. 粮食与油脂, 2022, 35(1): 50-52, 77.
ZHANG J P, HE X C, BAO J. Study on the relationship between the ash contents and the related indexes from different flour milling systems[J]. Cereals & Oils, 2022, 35(1): 50-52, 77.
- [3] 赵二霞, 张晓双. 关于 900 °C 与 550 °C 灼烧法测定小麦粉灰分的研究[J]. 现代面粉工业, 2023, 37(3): 30-32.
ZHAO E X, ZHANG X S. Study on determination of ash content in wheat flour by 900 °C and 550 °C burning method[J]. Modern Flour Industry, 2023, 37(3): 30-32.
- [4] 王晓玉, 孙先兵, 汪军, 等. 废弃生物质灰分含量与热值的差异性和相关性及其测定方法[J]. 中国农业大学学报, 2022, 27(9): 160-172.
WANG X Y, SUN X B, WANG J, et al. Differences and correlations between ash content and calorific value in waste biomass and their determination methods[J]. Journal of China Agricultural University, 2022, 27(9): 160-172.
- [5] 孙晓荣, 张晨光, 刘翠玲, 等. 近红外光谱技术对小麦粉品质定量快速检测[J]. 食品科技, 2023, 48(11): 246-252.
SUN X R, ZHANG C G, LIU C L, et al. Application of near-infrared spectroscopy for rapid quantification of wheat flour quality[J]. Food Science and Technology, 2023, 48(11): 246-252.
- [6] 郭歌, 张梦玲, 巩志杰, 等. 基于近红外光谱和复杂样品划分集合的生物质灰分含量模型构建[J]. 光谱学与光谱分析, 2023, 43(10): 3 143-3 149.
GUO G, ZHANG M L, GONG Z J, et al. Construction of biomass ash content model based on near-infrared spectroscopy and complex sample set partitioning[J]. Spectroscopy and Spectral Analysis, 2023, 43(10): 3 143-3 149.
- [7] WU Y H, LI D. Determination of ash content in silicon dioxide filled epoxy-phenolic prepreg using near infrared spectroscopy [J]. Journal of Near Infrared Spectroscopy, 2022, 30(4): 227-233.
- [8] 崔鹏程, 陈鑫, 张涛. 粮食储备库粉尘物性特征参数研究[J]. 粮食与油脂, 2023, 36(9): 66-69.
CUI P C, CHEN X, ZHANG T. Study on characteristic parameters of dust properties in grain storage[J]. Cereals & Oils, 2023, 36(9): 66-69.
- [9] 姜友谊, 刘博伟, 张成健, 等. 利用无人机多光谱影像的多品种玉米成熟度监测[J]. 农业工程学报, 2023, 39(20): 84-91.
JIANG Y Y, LIU B W, ZHANG C J, et al. Multi-variety maize maturity monitoring based on UAV multi-spectral images[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(20): 84-91.
- [10] 梁英凯, 商枫楠, 陈桥, 等. 基于机器视觉与机器学习的火龙果重量估计[J]. 食品与机械, 2023, 39(7): 99-103.
LIANG Y K, SHANG F N, CHEN Q, et al. Dragon fruit weight estimation based on machine vision and machine learning[J]. Food & Machinery, 2023, 39(7): 99-103.
- [11] 张灵枝, 黄艳, 于英杰, 等. 基于近红外光谱技术的六大茶类快速识别[J]. 食品与生物技术学报, 2024, 43(1): 48-59.
ZHANG L Z, HUANG Y, YU Y J, et al. Rapid identification of six major tea categories based on near-infrared spectroscopy [J]. Journal of Food Science and Biotechnology, 2024, 43(1): 48-59.
- [12] 王美雅, 徐涵秋. 超大城市土地覆盖与热环境的随机森林回归模型研究[J]. 遥感技术与应用, 2022, 37(2): 379-388.
WANG M Y, XU H Q. Study on the random forest regression model of land cover and thermal environment in megacities[J]. Remote Sensing Technology and Application, 2022, 37(2): 379-388.
- [13] 杨文丽, 王江龙, 张浩宇, 等. 基于 GC-IMS 技术分析留兰香薄荷及其饲喂草鱼的挥发性组分[J]. 食品与生物技术学报, 2022, 41(1): 84-94.
YANG W L, WANG J L, ZHANG H Y, et al. Analysis of volatile components of spearmint and grass carps fed by spearmint based on GC-IMS[J]. Journal of Food Science and Biotechnology, 2022, 41(1): 84-94.
- [14] 赵龙, 桑国庆, 武玮, 等. 基于随机森林回归算法的山洪灾害临界雨量预估模型[J]. 济南大学学报(自然科学版), 2022, 36(4): 404-411, 423.
ZHAO L, SANG G Q, WU W, et al. Critical rainfall prediction model for mountain torrent disaster based on random forest regression algorithm[J]. Journal of University of Jinan (Science and Technology), 2022, 36(4): 404-411, 423.
- [15] 田静, 陈斌, 陆道礼, 等. 不同分光原理近红外光谱仪光谱标准化方法在小麦粉品质检测中的应用[J]. 中国食品学报, 2022, 22(10): 286-294.
TIAN J, CHEN B, LU D L, et al. Application of spectral standardization of different spectral types of near-infrared analyzers in the quality detection of wheat flour[J]. Journal of Chinese Institute of Food Science and Technology, 2022, 22(10): 286-294.
- [16] 杨习江, 杨昌福, 蒋学乾, 等. 紫花苜蓿粗灰分与矿质元素含量 QTL 定位分析[J]. 草地学报, 2022, 30(2): 312-319.
YANG X J, YANG C F, JIANG X Q, et al. QTL analysis of ash and mineral elements in alfalfa[J]. Acta Agrestia Sinica, 2022, 30(2): 312-319.