

# 猪配合饲料多品质近红外光谱关键变量筛选 与模型建立

王 坤<sup>1</sup>, 吴静珠<sup>1\*</sup>, 王 冬<sup>2\*</sup>, 朱业伟<sup>3</sup>, 韩 平<sup>2</sup>

(1. 北京工商大学人工智能学院, 北京 100048; 2. 北京农业质量标准与检测技术研究中心, 北京 100097;  
3. 北京格致同德科技有限公司, 北京 100124)

**摘 要:** **目的** 基于饲料近红外光谱数据筛选影响猪配合饲料主要品质指标的关键波长变量, 从而建立饲料品质无损快速定量校正模型, 进而提高饲料品质无损快速检测效率。**方法** 采集饲料样品近红外光谱数据并获取水分、粗蛋白、粗脂肪、粗纤维参考值数据; 剔除异常值后采用基于联合 X-Y 距离样本集划分法(sample set partitioning based on joint X-Y distance, SPXY)划分校正集和外部验证集; 基于校正集数据采用蒙特卡罗-无信息变量消除-连续投影算法分别针对 4 个品质指标筛选 25、20、15、10、5 个关键变量, 分别建立校正模型并对外部验证集进行预测。**结果** 针对饲料水分、粗蛋白、粗脂肪、粗纤维所选关键变量个数分别为 15、25、15、15, 模型维数分别为 9、11、10、9, 测定系数分别为 0.8288、0.8605、0.9338、0.8327, 校正均方根误差分别为 0.17、0.81、0.31、0.22, 交互验证均方根误差分别为 0.19、0.93、0.34、0.23, 相对预测性能分别为 2.79、2.38、4.01、2.89。**结论** 通过变量筛选结合外部验证结果表明, 在保证模型准确度的前提下, 所选关键变量数明显少于全谱变量数, 可为提高饲料多品质无损快速定量检测工作效率提供一定的参考。

**关键词:** 猪配合饲料; 近红外光谱; 蒙特卡罗-无信息变量消除-连续投影算法

## Key variables selection and models development based on near-infrared spectra for the multi-qualities in formula feedstuff for swine

WANG Kun<sup>1</sup>, WU Jing-Zhu<sup>1\*</sup>, WANG Dong<sup>2\*</sup>, ZHU Ye-Wei<sup>3</sup>, HAN Ping<sup>2</sup>

(1. College of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China;  
2. Beijing Research Center for Agricultural Standards and Testing, Beijing 100097, China;  
3. Beijing Great-Tech Technology Co., Ltd, Beijing 100124, China)

**ABSTRACT: Objective** Based on the near-infrared spectra of feedstuff samples for swine, to select key wavelength

**基金项目:** 北京市农林科学院科技创新能力建设专项储备性研究课题(KJCX20180409)、科技部国家重点研发计划项目(2017YFD0201607)、北京工商大学食品安全大数据技术北京市重点实验室开放课题(BUBD-2017KF-11)

**Fund:** Supported by the Topic of Reserve Research, Innovation Ability Special, Beijing Academy of Agricultural and Forestry Sciences (KJCX20180409), the National Key R&D Projects of the Ministry of Science and Technology of the People's Republic of China (2017YFD0201607), and the Open Topic of Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University (BUBD-2017KF-11)

\*通讯作者: 吴静珠, 博士, 副教授, 主要研究方向为农产品品质检测技术。E-mail: pubwu@163.com

王冬, 博士, 副研究员, 主要研究方向为振动光谱分析与化学计量学。E-mail: wangd@brcast.org.cn

\*Corresponding author: WU Jing-Zhu, Ph.D, Associate Professor, College of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China. E-mail: pubwu@163.com

WANG Dong, Ph.D, Associate Professor, Beijing Research Center for Agricultural Standards and Testing, Beijing 100097, China. E-mail: wangd@brcast.org.cn

variables for the main quality indices in feedstuff, so as to develop the non-destructive rapid quantitative calibration models of quality indices of feedstuff, and thereby to improve the efficiency of non-destructive and rapid detection.

**Methods** The near-infrared spectra data of feedstuff samples were collected and the specified values of moisture, crude protein, crude fat, and crude fiber were obtained. After outlier elimination, sample set partitioning based on joint  $X$ - $Y$  distance (SPXY) algorithm was used to divide the data set into the calibration set and external validation set. Based on the calibration set data, Monte-Carlo-uninformative variable elimination-successive projection algorithm was used to select 25, 20, 15, 10, and 5 key variables for the 4 quality indices, respectively. Based on the key-variables data, the calibration models were developed and the external validation sets were predicted. **Results** The number of key variables selected for feed moisture, crude protein, crude fat, and crude fiber were 15, 25, 15, and 15 respectively, the number of the model factors were 9, 11, 10, and 9 respectively, the determination coefficients were 0.8288, 0.8605, 0.9338, and 0.8327 respectively, the root mean square errors of calibration were 0.17, 0.81, 0.31 and 0.22 respectively, the root mean square errors of cross-validation were 0.19, 0.93, 0.34 and 0.23 respectively, and the ratio performance deviations were 2.79, 2.38, 4.01 and 2.89 respectively. **Conclusion** It is demonstrated by the results of key variables selection combined with the prediction of external validation set that the numbers of the key variables selected are less than that of the full spectra obviously for the 4 quality indices when the accuracy of the models are ensured, which can provide a certain reference for improving the efficiency of non-destructive and rapid quantitative detection of the quality of feedstuff.

**KEY WORDS:** formula feedstuff for swine; near-infrared spectroscopy; Monte-Carlo-uninformative variable elimination- successive projection algorithm

## 1 引言

我国是全球猪肉消费大国,近几年畜牧养殖业发展迅速,伴随国民收入逐步提升和人口结构转变,在生产方面,生猪出栏量和猪肉产量逐年上升,国内猪肉消费需求将保持持续增长态势<sup>[1]</sup>。其中,规模化养殖企业具有成本优势,生猪成活率较高,饲料成本相对较低,盈利能力相对较强。由此可见,未来我国猪养殖行业的趋势是规模化、规范化,因此饲料的需求量将逐年上升,对于饲料品质的检测也显得越发重要。保证饲料的品质是养殖业不可缺少的一环。面对大量饲料的品质检测需求,提高饲料品质检测效率成为当前亟待解决的问题。

饲料的品质检测主要包括水分、粗蛋白、粗脂肪、粗纤维的定量检测。国标规定的饲料品质检测方法有:GB/T 6435-2014《饲料中水分的测定》<sup>[2]</sup>,应用干燥法测定干燥损失水分占比;GB/T 6432-2018《饲料中粗蛋白的测定凯氏定氮法》<sup>[3]</sup>,用盐酸滴定溶液测定氮含量,乘以 6.25 计算出粗蛋白含量;GB/T 6433-2006《饲料中粗脂肪的测定》<sup>[4]</sup>,应用蒸馏、干燥称量残渣法测定粗脂肪含量;GB/T 6434-2006《饲料中粗纤维的含量测定 过滤法》<sup>[5]</sup>,消煮样品、高温灼烧扣除无关物质的过滤法测定粗纤维含量。采用国标规定方法对饲料品质进行检测的过程需对样品进行离线抽样,且皆为破坏性检测,操作过程繁琐、检测工作效率低,难以满足对于大量饲料多种品质指标的无损快速

检测需求。因此,研发无损、快速、高效的检测方法成为当今饲料行业亟待解决的问题。

随着化学计量学的发展和计算机技术的进步,近红外光谱分析技术<sup>[6-8]</sup>以其快速、无损、高效、环境友好、样品制备简单以及可实现在线检测等特点,目前已经在饲料品质无损快速检测方面有所应用<sup>[9]</sup>。然而,面对大量饲料的多品质检测需求,传统的全谱数据分析逐渐表现出数据量大、计算量大、运算负荷大、仪器效率低等短板。饲料样品的近红外光谱数据中既包含目标物质的组成和结构等有用信息,同时又含有目标物质以外的噪声等干扰信息。面对饲料行业高通量在线检测,如果采用传统的全谱数据分析,很容易因数据量过大导致仪器运算负荷过高,从而造成工作效率降低,间接提高了生产成本。因此,针对待测品质指标进行光谱关键变量筛选成为在线高通量检测过程迫切需要解决的问题<sup>[10-13]</sup>。通过筛选和待测品质指标高度关联的关键变量数据,从而降低仪器运算数据量,进而达到提高仪器工作效率、满足饲料品质检测过程中大量样本多品质同时检测需求的目的是未来饲料检测的趋势。

在关键变量筛选研究方面,很多学者做了有意义的尝试,并取得了一定的成果。Li 等<sup>[14]</sup>基于高光谱成像采用蒙特卡罗-无信息变量消除-连续投影算法(Monte-Carlo-uninformative variable elimination-successive projection algorithm, MC-UVE-SPA)从长波近红外光谱数据中选择最有效的变量,用于测量梨可溶性固形物含量(soluble solid content, SSC);Yuan 等<sup>[15]</sup>采用 4 种变量选择方

法用于近红外光谱(near-infrared spectra, NIRS)变量选择, 鉴定来自中国云南的茯苓的起源; Zhang 等<sup>[16]</sup>研究探索了在 400~1000 nm 波长范围内的高光谱成像技术在砀山梨快速无损预测和糖含量可视化的潜在应用, 应用包括 MC-UVE、SPA、竞争性自适应加权采样(competitive adaptive reweighted sampling, CARS)、遗传算法(genetic algorithm, GA)、CARS-SPA、GA-SPA; Zhang 等<sup>[17]</sup>基于高光谱成像采用 MC-UVE 和 SPA 检测桃的常见缺陷; Wang 等<sup>[18]</sup>基于腐竹的近红外漫反射光谱采用 MC-UVE、随机蛙跳(random frog, RF)和 CARS 算法针对蛋白质、脂质、水分筛选关键变量; Zhang 等<sup>[19]</sup>针对油菜叶片除草剂胁迫下天冬氨酸含量的测定比较了不同光谱预处理方法对变量筛选算法 SPA、MC-UVE、RF 结果的影响; Li 等<sup>[20]</sup>基于可见-近红外(visible-near-infrared, Vis-NIR)光谱采用 MC-UVE-SPA 对梨的不同品种进行无损、准确测定硬度; Wu 等<sup>[21]</sup>针对干紫菜中蛋白质含量的测定比较了 UVE、SPA、UVE-SPA 筛选光谱关键变量的结果。

本研究以猪饲料的水分、粗蛋白、粗脂肪、粗纤维 4 个主要品质指标的无损快速检测为例, 采用近红外光谱仪采集饲料近红外光谱数据, 同时采用参考方法获取饲料样品水分、粗蛋白、粗脂肪、粗纤维含量数据; 采用模型残差值结合光谱杠杆值剔除异常值, 并对剔除异常值后的数据采用基于联合 X-Y 距离样本集划分(sample set partitioning based on joint X-Y distance, SPXY)算法<sup>[22]</sup>分为校正集(calibration set, Calib)和外部验证集(external validation set, ExVal); 基于校正集采用 MC-UVE-SPA 筛选关键变量; 基于所选关键变量, 采用偏最小二乘(partial least square, PLS)回归结合全交互验证算法建立校正模型, 并采用外部验证集对所建模型进行验证。本研究应用变量筛选算法减少参与建模的变量个数, 基于所选关键变量数据建立校正模型并进行模型验证, 可为降低检测仪器运算负荷、提高饲料品质定量检测工作效率提供一定的参考。

## 2 材料与方法

### 2.1 样品与仪器

#### 2.2.1 样品来源

采集市售猪饲料样品 300 个, 分别根据国标方法 GB/T 6435-2014《饲料中水分的测定》<sup>[2]</sup>、GB/T 6432-2018《饲料中粗蛋白的测定 凯氏定氮法》<sup>[3]</sup>、GB/T 6433-2006《饲料中粗脂肪的测定》<sup>[4]</sup>、GB/T 6434-2006《饲料中粗纤维的含量测定 过滤法》<sup>[5]</sup>测定饲料样品水分、粗蛋白、粗脂肪、粗纤维参考值, 对参考值计算样本容量( $n$ )、最小值( $min$ )、最大值( $max$ )、平均值( $ave$ )、样本标准差( $std$ )、极差( $Rx$ )和变异系数( $CV$ )7 个统计量, 详见表 1。

表 1 饲料样品品质指标统计信息

Table 1 The statistics of the quality indices in feedstuff samples

统计量	水分	粗蛋白	粗脂肪	粗纤维
$n$	300	300	300	300
$min$	10.24	12.7	1.93	2.13
$max$	12.61	22.49	7.98	4.94
$ave$	11.99	17.3	3.9	3.61
$std$	0.42	2.11	1.18	0.49
$Rx$	2.37	9.79	6.05	2.81
$CV$	3.50%	12.20%	30.40%	13.60%

#### 2.2.2 仪器与参数

NIR1700 便携式线性渐变分光近红外光谱仪(linear variable filter-near-infrared, LVR-NIR)(美国 VIAVI 公司), 仪器参数如下: 波长范围 908~1676 nm, 光谱变量数 125 个, 光谱中心分辨率 10 nm; 以漫反射方式采集光谱数据, 以聚四氟乙烯白板为参比, 光谱单次积分时间 25  $\mu$ s, 光谱累加 100 次取平均。

## 2.2 数据处理

#### 2.2.1 异常值的剔除

使用化学计量学软件 Unscrambler V9.8 分别对饲料的水分、粗蛋白、粗脂肪、粗纤维 4 个品质指标, 采用 PLS 回归结合全交互验证算法初步建立模型, 根据模型残差值和光谱杠杆值将原始光谱数据中偏离大部分样品的异常样品剔除。

#### 2.2.2 数据分集

在 MATLAB 2016a 环境下自编 SPXY 程序, 将剔除异常值的光谱数据后划分校正集和外部验证集; 以 200 个校正集样品为基准, 校正集和外部验证集的样品数量比约为 2:1。

#### 2.2.3 关键变量的筛选

在 MATLAB 2016a 环境下采用 MC 方法对 200 个校正集样品随机采样, MC 单次采样比率为 0.8, 采样次数为 200, 以提高所筛选关键变量的稳定性; UVE 算法用于消除不提供有效信息的变量, SPA 可有效去除所选变量之间的共线性, 因此本研究将 UVE 和 SPA 相结合从而提高关键变量筛选效率。采用 MC-UVE-SPA 基于校正集数据针对水分、粗蛋白、粗脂肪、粗纤维筛选关键变量, 对每种品质指标所选关键变量个数将通过基于关键变量数据所建模型性能即模型参数确定。

#### 2.2.4 所选关键变量性能的验证

使用化学计量学软件 Unscrambler v9.8 基于所选变量数据采用偏最小二乘回归结合全交互验证算法分别建立饲料 4 种品质指标的校正模型并对相应的外部验证集进行预测。采用测定系数( $R^2_c$ )、校正均方根误差(root mean square

error of calibration,  $RMSEC$ )、交互验证决定系数( $R^2_{CV}$ )、交互验证均方根误差(root mean square error of cross-validation,  $RMSECV$ )对模型进行评价; 结合外部验证集, 采用预测测定系数( $R^2_p$ )、预测均方根误差(root mean square error of prediction,  $RMSEP$ )、相对预测性能(ratio performance deviation,  $RPD$ )对模型预测性能进行评价。

### 3 结果与分析

#### 3.1 异常值的剔除

基于 LVR-NIR 仪采集的饲料近红外光谱如图 1 所示。从图 1 可见, 在 908~1676 nm 范围内, 猪饲料的近红外光谱表现为宽峰, 其中主要有 2 个吸收峰, 分别位于 1190 和 1470 nm。

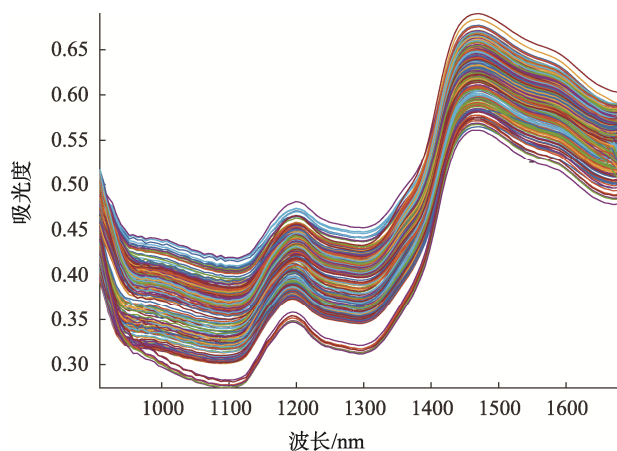


图 1 饲料样品近红外光谱图

Fig.1 NIR spectra of the feedstuff samples

对数据进行中心化预处理, 采用 PLS 回归结合全交互验证算法对饲料样品的水分、粗蛋白、粗脂肪、粗纤维初步建立校正模型, 根据模型残差值和光谱杠杆值剔除异常值。模型残差值和光谱杠杆值的计算公式分别如公式(1)和公式(2)所示。

$$Residual_i = y_i - y_{iP} \quad (1)$$

$$Leverage_i = t_i'(TT)t_i \quad (2)$$

公式(1)中, 针对一个品质指标,  $Residual_i$  是第  $i$  个样品的模型残差值,  $y_i$  是第  $i$  个样品的该品质指标参考值,  $y_{iP}$  是第  $i$  个样品的模型预测值; 公式(2)中, 针对一个品质指标,  $Leverage_i$  是第  $i$  个样品的光谱杠杆值,  $t_i$  是第  $i$  个样品的得分值,  $T$  是模型得分矩阵。当样品的模型残差值超过校正集模型残差值平均值的 2 倍且光谱杠杆值超过校正集光谱杠杆值平均值的 3 倍时, 该样品将作为异常值被剔除。

4 种品质指标剔除异常值后的统计信息见表 2。

表 2 饲料样品剔除异常值后品质指标统计信息

Table 2 The statistics of the quality indices in feedstuff samples after the outliers eliminated

统计量	水分	粗蛋白	粗脂肪	粗纤维
$n$	297	293	295	291
$min$	10.52	12.7	1.93	2.13
$max$	12.61	22.49	7.12	4.94
$ave$	12	17.25	3.85	3.62
$std$	0.41	2.11	1.12	0.48
$Rx$	2.09	9.79	5.19	2.81
$CV$	3.40%	12.20%	29.20%	13.10%

将剔除异常值后的光谱数据建立模型, 4 种品质指标剔除前、后建立校正模型的维数( $N_f$ ),  $R^2_c$ 、 $R^2_{CV}$ 、 $RMSEC$ 、 $RMSECV$  分别如表 3 和表 4 所示。结合表 3 和表 4 分析可知, 剔除异常值后的各模型  $R^2_c$  和  $R^2_{CV}$  皆有不同程度的提高, 而  $RMSEC$ 、 $RMSECV$  皆有不同程度的下降, 说明异常值的剔除有利于提高所建模型准确度, 从而可以为各品质指标关键变量的筛选提供可靠的数据集, 进而可以保证对所选关键变量验证结果的可靠性。

表 3 饲料样品剔除异常值前品质指标统计信息

Table 3 The statistics of the quality indices in feedstuff samples before the outliers eliminated

统计量	水分	粗蛋白	粗脂肪	粗纤维
$N_f$	13	11	9	15
$R^2_c$	0.8506	0.8834	0.9256	0.8158
$R^2_{CV}$	0.16	0.72	0.32	0.21
$RMSEC$	0.8012	0.8558	0.9123	0.7304
$RMSECV$	0.19	0.8	0.35	0.26

表 4 饲料样品剔除异常值后品质指标统计信息

Table 4 The statistics of the quality indices in feedstuff samples after the outliers eliminated

统计量	水分	粗蛋白	粗脂肪	粗纤维
$N_f$	13	11	9	15
$R^2_c$	0.8701	0.8979	0.9360	0.8529
$R^2_{CV}$	0.15	0.67	0.28	0.18
$RMSEC$	0.8317	0.8757	0.9257	0.7981
$RMSECV$	0.17	0.74	0.31	0.21

#### 3.2 数据分集

将剔除异常值后的数据划分校正集和外部验证集,

分集后饲料的 4 个指标的校正集和外部验证集的统计情况如表 5 所示。结合表 5 和表 2, 经过 SPXY 算法划分所得的校正集和外部验证集的统计数据与剔除异常值后的数据具有相似分布趋势, 说明 SPXY 算法分集所得校正集和外部验证集数据集皆有较好的代表性, 外部验证集可为验证基于所选关键变量所建模型的准确度提供可靠的外部验证数据。

### 3.3 关键变量的筛选与验证

采用 MC-UVE-SPA 针对饲料样品各指标校正集数据分别筛选 25、20、15、10、5 个关键变量, 并根据所选关键变量数据建立各指标定量校正模型, 对相应变量数的外部验证集进行预测。

水分全谱模型和基于所选关键变量所建模型结果如表 6 所示, 其中全谱变量数为 125 个。由表 6 可见, 对于水分, 基于所选的 15 个关键变量所建模型的 *RMSEP*

最低, *RPD* 最高,  $R^2_c$  为 0.8288, 同时模型维数比全谱模型的维数减少 4 个, 模型准确度较高; 基于所选 25、20、10、5 个关键变量所建水分模型的准确度相对较低。

粗蛋白全谱模型和基于所选关键变量所建模型结果如表 7 所示, 其中全谱变量数为 125 个。由表 6 可见, 对于粗蛋白, 基于所选的 25 个关键变量所建模型的 *RMSEP* 最低,  $R^2_c$ 、*RPD* 分别为 0.8605、2.38, 模型维数和全谱模型的维数相同, 模型准确度较高; 基于所选 20、15、10、5 个关键变量所建粗蛋白模型的准确度相对较低。

粗脂肪全谱模型和基于所选关键变量所建模型结果如表 8 所示, 其中全谱变量数为 125 个。由表 8 可见, 对于粗脂肪, 基于所选的 15 个关键变量所建模型的 *RMSEP* 最低, *RPD* 在所有粗脂肪模型中最高,  $R^2_c$  为 0.9338, 模型准确度较高; 基于所选 25、20、10、5 个关键变量所建粗脂肪模型的准确度相对较低。

表 5 饲料样品数据分集统计信息  
Table 5 The statistics of diversities of the data of feedstuff samples

统计量	水分		粗蛋白		粗脂肪		粗纤维	
	Calib <sup>*</sup>	ExVal <sup>#</sup>	Calib	ExVal	Calib	ExVal	Calib	ExVal
<i>n</i>	200	97	200	93	200	95	200	91
<i>min</i>	10.52	10.84	12.70	13.06	1.93	2.13	2.13	2.72
<i>max</i>	12.61	12.57	22.49	20.50	7.08	7.12	4.94	4.72
<i>ave</i>	11.95	12.10	17.18	17.42	3.92	3.70	3.65	3.57
<i>std</i>	0.42	0.36	2.17	1.97	1.20	0.92	0.53	0.33
<i>Rx</i>	2.09	1.73	9.79	7.44	5.15	4.99	2.81	2.00
<i>CV</i>	3.5%	3.0%	12.6%	11.3%	30.6%	25.0%	14.5%	9.2%

注: \*为校正集、#为外部验证集。

表 6 饲料水分全谱模型和基于所选关键变量所建模型结果  
Table 6 The full spectra model and the models based on the key variables selected of moisture in feedstuff

统计量	125	25	20	15	10	5
$N_f$	13	10	9	9	8	4
$R^2_c$	0.8830	0.8431	0.8187	0.8288	0.7583	0.5145
$R^2_{cv}$	0.14	0.17	0.18	0.17	0.21	0.29
<i>RMSEC</i>	0.8277	0.8031	0.7758	0.7975	0.7373	0.4833
<i>RMSECV</i>	0.17	0.19	0.2	0.19	0.22	0.3
$R^2_p$	0.8106	0.8088	0.7989	0.8282	0.7574	0.7112
<i>RMSEP</i>	0.16	0.16	0.16	0.15	0.18	0.2
<i>RPD</i>	2.65	2.64	2.57	2.79	2.35	2.13

表 7 饲料粗蛋白全谱模型和基于所选关键变量所建模型结果

Table 7 The full spectra model and the models based on the key variables selected of crude protein in feedstuff

统计量	125	25	20	15	10	5
$N_f$	11	11	10	10	7	4
$R^2_c$	0.8957	0.8605	0.8234	0.7459	0.6136	0.3115
$R^2_{cv}$	0.7	0.81	0.91	1.09	1.34	1.79
$RMSEC$	0.863	0.8159	0.7837	0.7046	0.5683	0.2562
$RMSECV$	0.8	0.93	1.01	1.18	1.43	1.87
$R^2_p$	0.8771	0.7906	0.753	0.7239	0.6064	0.3037
$RMSEP$	0.67	0.91	0.98	1.04	1.24	1.64
$RPD$	3.21	2.38	2.21	2.09	1.75	1.32

表 8 饲料粗脂肪全谱模型和基于所选关键变量所建模型结果

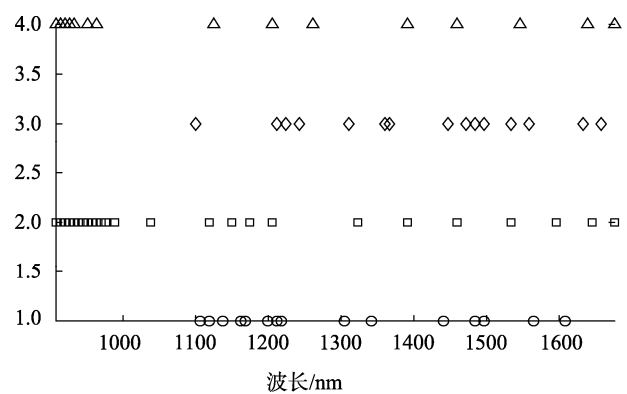
Table 8 The full spectra model and the models based on the key variables selected of crude fat in feedstuff

统计量	125	25	20	15	10	5
$N_f$	9	9	9	10	2	2
$R^2_c$	0.9482	0.9465	0.9387	0.9338	0.2789	0.2789
$R^2_{cv}$	0.27	0.28	0.3	0.31	1.02	1.02
$RMSEC$	0.9366	0.9336	0.9237	0.9196	0.2573	0.2573
$RMSECV$	0.3	0.31	0.33	0.34	1.04	1.04
$R^2_p$	0.8662	0.865	0.8852	0.8975	0.0888	0.0888
$RMSEP$	0.34	0.35	0.32	0.30	1.03	1.03
$RPD$	3.52	3.46	3.77	4.01	1.16	1.16

粗纤维全谱模型和基于所选关键变量所建模型结果如表 9 所示,其中全谱变量数为 125 个。由表 9 可见,对于粗纤维,基于所选的 25 个关键变量和 15 个关键变量所建模型的  $RMSEP$  和  $RPD$  分别相同,  $RMSEP$  在所有粗纤维模型中最低,  $RPD$  在所有粗纤维模型中最高,模型准确度较高;但基于 15 个关键变量所建模型比基于 25 个关键变量所建模型的  $R^2_c$  值高,且基于 15 个关键变量所建模型的模型维数比全谱模型的维数减少 4 个,比基于 25 个关键变量所建模型的维数减少 2 个,因此对于粗纤维,基于 15 个关键变量所建模型更好,基于所选 20、10、5 个关键变量所建粗纤维模型的准确度相对较低。

综上,在保证模型准确度的前提下,本研究针对水分、粗蛋白、粗脂肪、粗纤维所选的最佳关键变量个数分别为 15、25、15、15,4 种品质指标所选关键变量的位置分布如图 2 所示。由图 2 可见,水分模型的关键变量主要集中在 1100~1220 nm、1300~1610 nm 之间;粗蛋白模型的关键变量主要集中在 908~1040 nm、1110~1210 nm、1320~1676 nm 之间;粗脂肪模型的关

键变量主要集中在 1100~1660 nm 之间;粗纤维模型的关键变量主要集中在 908~970 nm、1120~1270 nm、1390~1676 nm 之间。



注: ○水分、□粗蛋白、◇粗脂肪、△粗纤维。

图 2 饲料各品质指标所选关键变量分布示意图

Fig.2 Distribution plot of the selected key variables for the quality indices in feedstuff

表 9 饲料粗纤维全谱模型和基于所选关键变量所建模型结果  
Table 9 The full spectra model and the models based on the key variables selected of crude fiber in feedstuff

统计量	125	25	20	15	10	5
$N_f$	13	11	8	9	9	4
$R^2_c$	0.8528	0.8036	0.7801	0.8327	0.7548	0.3304
$R^2_{cv}$	0.2	0.23	0.25	0.22	0.26	0.43
RMSEC	0.7846	0.763	0.765	0.7128	0.7291	0.2935
RMSECV	0.25	0.26	0.26	0.23	0.29	0.45
$R^2_p$	0.6618	0.7075	0.6356	0.7187	0.7008	0.1583
RMSEP	0.21	0.18	0.22	0.18	0.19	0.31
RPD	2.25	2.89	2.44	2.89	2.79	1.68

#### 4 结论与讨论

本研究采用 LVF-NIR 光谱仪采集 300 个饲料样品的近红外光谱, 根据模型残差值和光谱杠杆值剔除异常值, 使用 SPXY 算法划分校正集和外部验证集, 通过 MC-UVE-SPA 针对各品质指标在校正集中分别筛选 25、20、15、10、5 个关键变量, 基于所选各组关键变量采用 PLS 回归结合全交互验证算法分别建立饲料水分、粗蛋白、粗脂肪、粗纤维校正模型, 并对所建模型进行外部验证。结果表明, 模型的预测准确度与所选关键变量具有一定的关系, 在保证模型准确度的前提下, 针对水分、粗蛋白、粗脂肪、粗纤维所选的最佳关键变量个数分别为 15、25、15、15, 对每个品质指标所选关键变量数明显低于全谱变量数, 因此可以大幅降低建模和预测过程计算量。

研究表明, 针对饲料水分、粗蛋白、粗脂肪、粗纤维筛选少量关键变量后, 基于所选关键变量建立校正模型仍具有较高的准确度和较好的预测性能, 可为提高饲料多品质无损快速定量测定工作效率、间接地降低生产成本提供一定的技术参考, 对其他农业投入品的品质指标无损快速检测工作效率的提高亦具有一定的参考价值。

#### 参考文献

- [1] 刘小红, 张海峰, 陈瑶生. 2018 年生猪产业发展状况、未来发展趋势与建议[J]. 中国畜牧杂志, 2019, 55(3): 115-118.  
Liu XH, Zhang HF, Chen YS. Development status of pig industry in 2018, future development trends and suggestions [J]. Chin J Anim Sci, 2019, 55(3): 115-118.
- [2] GB/T 6435-2014 饲料中水分的测定[S].  
GB/T 6435-2014 Determination of moisture in feed [S].
- [3] GB/T 6432-2018 饲料中粗蛋白的测定 凯氏定氮法[S].  
GB/T 6432-2018 Determination of crude protein in feed-Kjeldahl method [S].
- [4] GB/T 6433-2006 饲料中粗脂肪的测定[S].  
GB/T 6433-2006 Determination of crude fat in feed [S].
- [5] GB/T 6434-2006 饲料中粗纤维的含量测定 过滤法[S].  
GB/T 6434-2006 Determination of crude fiber content in feed-Filter method [S].
- [6] 褚小立, 史云颖, 陈瀑, 等. 近 5 年我国近红外光谱分析技术研究与应用进展[J]. 分析测试学报, 2019, 38(5): 603-611.  
Chu XL, Shi YY, Chen P, et al. Research and application progress of near infrared spectroscopy analytical technology in China in past five years [J]. J Instrum Anal, 2019, 38(5): 603-611.
- [7] 鲁雄, 张力培. 近红外光谱分析技术在饲料行业中的应用[J]. 安徽农业科学, 2016, 44(2): 83-85.  
Lu X, Zhang LP. Application of near infrared spectral analysis technology in feed industry [J]. J Anhui Agric Sci, 2016, 44(2): 83-85.
- [8] 姜训鹏, 雷恒, 李海涛, 等. 不同饲料原料日粮纤维水平的近红外测定方法[J]. 农业机械学报, 2016, (47): 353-358.  
Jiang XP, Lei H, Li HT, et al. Measurement of fiber content in different feed ingredients using near-infrared spectroscopy method [J]. Trans Chin Soc Agric Mach, 2016, (47): 353-358.
- [9] GB/T 18868-2002 饲料中水分、粗蛋白质、粗纤维、粗脂肪、赖氨酸、蛋氨酸快速测定 近红外光谱法[S].  
GB/T 18868-2002 Rapid determination of moisture, crude protein, crude fiber, crude fat, lysine, methionine in feed by near infrared spectroscopy [S].
- [10] Wilcox P, Horton TM, Youn E, et al. Evolutionary refinement approaches for band selection of hyperspectral images with applications to automatic monitoring of animal feed quality [J]. Intell Data Anal, 2014, 18(1): 25-42.
- [11] Sendin K, Williams PJ, Manley M. Near infrared hyperspectral imaging in quality and safety evaluation of cereals [J]. Crit Rev Food Sci, 2016, 58(4): 1-61.
- [12] Baeten V, Pierna JAF, Vermeulen P, et al. NIR hyperspectral imaging methods for quality and safety control of food and feed products: contributions to four European projects [J]. NIR News, 2010, 21(6): 10-13.
- [13] Pierna JAF, Vermeulen P, Amand O, et al. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed [J]. Chemometr Intell Lab, 2012, (117): 233-239.
- [14] Li JB, Xi T, Huang WQ, Zhang BH, et al. Application of long-wave near

- infrared hyperspectral imaging for measurement of soluble solid content (SSC) in pear [J]. *Food Anal Method*, 2016, 9(11): 3087–3098.
- [15] Yuan TJ, Zhao YL, Zhang J, *et al.* Application of variable selection in the origin discrimination of *Wolfiporia cocos* (F.A. Wolf) Ryvarden & Gilb. based on near infrared spectroscopy [J]. *Sci Rep UK*, 2018, 89(8): 1–10.
- [16] Zhang DY, Xu L, Liang D, *et al.* Fast prediction of sugar content in dangshan pear (*Pyrus spp.*) using hyperspectral imagery data [J]. *Food Anal Method*, 2018, (11): 2836–2845.
- [17] Zhang BH, Li JB, Fan SX, *et al.* Hyperspectral imaging combined with multivariate analysis and band math for detection of common defects on peaches (*Prunus persica*) [J]. *Comput Electron Agric*, 2015, (114): 14–24.
- [18] Wang JH, Wang J, Zhang XW, *et al.* Key variables screening of near-infrared models for simultaneous determination of quality parameters in traditional Chinese food (fuzhu) [J]. *J Food Qual*, 2018, (2018): 1–11.
- [19] Zhang C, Kong WW, Liu F, *et al.* Measurement of aspartic acid in oilseed rape leaves under herbicide stress using near infrared spectroscopy and chemometrics [J]. *Heliyon*, 2016, 2(1): e00064.
- [20] Li JB, Zhang HL, Zhan BS, *et al.* Nondestructive firmness measurement of the multiple cultivars of pears by Vis-NIR spectroscopy coupled with multivariate calibration analysis and MCUVE-SPA method [J]. *Infrared Phys Technol*, 2019, (104): 103154.
- [21] Wu D, Chen XJ, Zhu XG, *et al.* Uninformative variable elimination for improvement of successive projections algorithm on spectral multivariable selection with different calibration algorithms for the rapid and non-destructive determination of protein content in dried laver [J]. *Anal Methods UK*, 2011, (3): 1790–1796.
- [22] 王世芳, 韩平, 崔广禄, 等. SPXY 算法的西瓜可溶性固形物近红外光谱检测[J]. *光谱学与光谱分析*, 2019, 39(3): 738–742.
- Wang SF, Han P, Cui GL, *et al.* Near infrared spectroscopy detection of soluble solids in watermelon based on SPXY algorithm [J]. *Spectrosc Spect Anal*, 2019, 39(3): 738–742.

(责任编辑: 李磅礴)

### 作者简介



王 坤, 硕士, 主要研究方向为控制工程。

E-mail: wk840832@163.com



吴静珠, 博士, 副教授, 主要研究方向为农产品品质检测技术。

E-mail: pubwu@163.com



王 冬, 博士, 副研究员, 主要研究方向为振动光谱分析与化学计量学。

E-mail: wangd@brcast.org.cn