

# 南方籼稻米胶稠度近红外光谱分析数学模型的构建与验证

刘红梅, 申 涛, 张文怡, 石习文, 戴 涛, 白 涛, 肖应辉\*

湖南农业大学农学院, 湖南 长沙 410128

**摘 要** 培育优质高产水稻品种是当前水稻育种的重要任务之一; 胶稠度是稻米蒸煮食味品质最重要的指标之一。传统的化学法测定稻米胶稠度, 前处理复杂、过程繁琐、试剂消耗多, 难以满足大批量水稻品种(组合)胶稠度快速无损检测的需求。近红外光谱分析能快速、无损、无污染地分析检测。以 137 份南方稻区主栽或新育成的籼稻品种(组合)为供试材料, 用传统化学法测定其稻米胶稠度并收集其近红外光谱, 以建立南方籼稻米胶稠度近红外光谱分析模型, 并对模型进行校正、检验。用偏最小二乘法(PLS)分别经 20 种数学预处理和 6 种波长段(或组合)建立各自的近红外分析模型。通过比较模型评价指标确定了平滑预处理为最佳预处理方式, 波长段 1 100~1 650 nm 为最优建模波长段。平滑处理模型评价指标: 校准相关系数( $R$ )、检验相关系数( $r$ )、相对百分比偏差(RPD)分别为 0.970 0, 0.964 2 和 3.780 5; 波长段 1 100~1 650 nm 模型评价指标:  $R$ ,  $r$  和 RPD 分别为 0.969 4, 0.963 8 和 3.758 6; 经平滑处理后在 1 100~1 650 nm 波长段建立了稻米胶稠度最优近红外分析模型, 其模型评价指标:  $R$ ,  $r$  和 RPD 分别为 0.979 0, 0.974 1 和 4.419 4; 再用 30 份验证集样本对所得的最优模型进行外部验证, 结果表明: 近红外检测值和化学值绝对误差在 0.198 6~6.502 4 mm, 对其进行配对  $T$  检验,  $p=0.726 > 0.05$ , 表明该近红外模型的检测值与化学值无显著差异, 即应用此近红外模型快速无损检测稻米胶稠度是可行的。本研究结果为优质水稻品种(组合)早世代材料的快速筛选及稻米胶稠度的批量快速分析检测提供了技术支持。

**关键词** 籼稻米; 胶稠度; 近红外模型; 偏最小二乘法

**中图分类号:** O657.33 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)08-2432-05

## 引 言

水稻是世界及我国重要的粮食作物之一, 全球超过一半的人以大米为主食, 我国则有 60% 以上的人口以大米为主食。随着生活水平的提高, 人民对米质的要求越来越高, 因此, 水稻品质改良育种尤其重要。稻米胶稠度与米饭的柔软度和粘稠度有关, 能反映稻米蒸煮食味品质的优劣<sup>[1-2]</sup>。化学方法测定胶稠度操作繁琐、测试周期长、检测过的样品不能继续用于育种研究, 不利于育种工作者对早世代品种的快速筛选<sup>[3]</sup>。近红外分析技术可快速检测有机物含量<sup>[4]</sup>, 能无损地检测稻米品质, 对加快优质稻的育种进程具有重要意义。近红外分析技术在医药、农业等领域已得到广泛的应用<sup>[5-7]</sup>, 而在稻米胶稠度检测方面的研究少有报道。因此, 本研究收集南方稻区主栽或新育成的水稻品种(组合)为材

料, 测定其胶稠度的化学值, 收集其近红外光谱, 采用偏最小二乘法, 建立稻米胶稠度的近红外分析模型, 以期对稻米胶稠度的快速无损检测提供参考。

## 1 实验部分

### 1.1 试验材料

收集 137 份南方稻区主栽或新育成的水稻品种(组合)。

### 1.2 测定项目与方法

#### 1.2.1 稻米胶稠度化学值的测定

按 NY/T 83—2017《米质测定方法》测定稻米胶稠度的化学值。准确称取经过 100 目筛的糯米米粉 100 mg ( $\pm 1$  mg), 加入定制的试管内, 防止米粉粘在试管口壁, 4 次重复。用移液管加入 95% 乙醇麝香草酚蓝溶液 0.2 mL, 并轻微摇动试管, 使米粉充分分散而不沉淀结块; 再加 0.2

收稿日期: 2020-06-28, 修订日期: 2020-10-19

基金项目: 国家重点研发计划项目(2016YFD0101100), 湖南省自然科学基金项目(2016JJ3072), 湖南农业大学大学生创新性实验计划项目(XCX19098, S202010537022)资助

作者简介: 刘红梅, 女, 1979 年生, 湖南农业大学农学院高级实验师 e-mail: liuhongmei@hunau.edu.cn

\* 通讯作者 e-mail: xiaoyh@hunau.edu.cn

mol · L<sup>-1</sup> 的 KOH 溶液 2.0 mL, 并轻轻摇动试管, 勿使米粉沉淀, 将试管立即放入沸水浴中加热 8 min, 试管口盖上玻璃珠, 保持试管内液面低于水浴锅水面, 且在加热过程中米胶高度始终维持在试管长度的 2/3, 不应超过或者溢出; 加热完毕将试管取出, 取下玻璃珠将试管置于试管架上静置 5 ~ 10 min 后, 置于冰浴冷却 20 min, 然后将试管平放于米胶测定箱内 (25 ± 2 °C), 1 h 后以 mm 为单位测量米胶长度。

### 1.2.2 近红外光谱的采集

用近红外光谱分析仪(波通 DA7200)扫描样品获得近红外漫反射光谱值, 扫描波长范围为 950 ~ 1 650 nm, 识别度为 5 nm, 环境温度控制在 25 °C, 每个样品重复扫描 4 次, 对光谱值进行均值化处理。

### 1.3 稻米胶稠度近红外模型的建立

用 The Unscrambler 9.7 (CAMO) 分析软件, 基于偏最小二乘法 (PLS) 分别对校正集样品采用不同数据预处理方式和不同波长段的近红外光谱进行建模。为了防止模型过拟合, 采用交互检验确定最佳主因子数。

#### 1.3.1 校正集样品预处理方式的筛选

将平滑(smoothing)、多元散射校正(multiplicative scatter correction, MSC)、标准正态变量转换(standard normal variate transformation, SNV)、一阶导数(first derivative)、二阶导数(second derivative)等预处理方法单独或组合共 20 种预处理方式分别对原始光谱进行处理, 根据模型评价指标确定最佳预处理方式。

#### 1.3.2 最佳波长段的选择

因样品的近红外光谱在某些波长段明显的吸收峰, 分别对这些表现出明显吸收峰的波长段及其复合波长段进行稻米胶稠度的近红外光谱分析, 通过对比各模型评价指标确定最佳波长段。

#### 1.3.3 模型评价指标

以校准相关系数( $R$ )、校准标准差(SEC)、校准均方根误差(RMSEC)、检验相关系数( $r$ )、检验标准差(SECV)、检验均方根误差(RMSECV)作为衡量稻米胶稠度近红外模型预测效果的主要指标。模型的相关系数越接近 1 越好, 综合使用相对百分比偏差(relative percent deviation, RPD)进一

步对模型预测效果进行评价。若  $RPD \geq 3$ , 说明该模型检测精确度高, 可用于相关组分的检测; 若  $2.5 \leq RPD < 3$ , 说明所建立的近红外模型检测效果一般, 可用于相关组分的定量分析, 但其精度需进一步修正提高; 若  $RPD < 2.5$ , 则说明使用该近红外模型进行分析检测比较困难<sup>[8]</sup>。

#### 1.3.4 模型检验

将用于验证集的 30 份样品(不参与建模)用 DA7200 采集其近红外光谱, 用已经建立的稻米胶稠度近红外分析模型预测胶稠度, 建立预测值和化学值的对比表格, 计算其绝对误差和配对  $t$  测验的  $p$  值, 以判断该近红外模型检测胶稠度的可行性。

## 2 结果与讨论

### 2.1 稻米胶稠度的分布

表 1 为校正集和验证集稻米胶稠度的分布情况。校正集范围为 43.0 ~ 94.0 mm, 平均值为 71.0 mm, 标准差为 12.2 mm。样本胶稠度分布范围较广, 且验证集两端极值在校正集之中, 保证了模型预测值的准确性、实用性。

表 1 稻米胶稠度的化学值统计表

Table 1 The chemical value of rice gel consistency

样本	样品数量	胶稠度/mm	平均值/mm	标准差/mm
校正集	107	43.0~94.0	71.0	12.2
验证集	30	43.5~88.0	70.8	11.5

### 2.2 稻米胶稠度近红外分析的最佳预处理方式筛选

从校正集和验证集综合来看, 由表 2 可知, 采用平滑处理其校正集的校准相关系数( $R$ )、校准标准差(SEC)、校准均方误差(RMSEC)分别为 0.970 0, 2.971 8 和 2.957 9; 验证集的检验相关系数( $r$ )、检验标准差(SECV)、检验均方根误差(RMSECV)分别为 0.964 2, 3.342 1 和 3.227 1, 相对百分比偏差  $RPD = 3.780 5 (> 3)$ , 说明平滑处理的近红外分析模型效果最佳。

表 2 不同预处理方法稻米胶稠度近红外模型的评价指标

Table 2 Evaluation indexes of NIR model of rice gel consistency under different treatments

预处理	主因子	校正集			验证集			RPD
		$R$	SEC	RMSEC	$r$	SECV	RMSECV	
无处理	6	0.970 0	2.973 6	2.959 6	0.962 6	3.313 4	3.298 0	3.699 2
SNV	7	0.947 2	3.919 6	3.901 2	0.936 0	4.304 5	4.284 4	2.847 5
MSC	7	0.947 0	3.928 5	3.910 1	0.934 6	4.349 4	4.329 1	2.818 1
平滑	6	0.970 0	2.971 8	2.957 9	0.964 2	3.242 1	3.227 1	3.780 5
一阶导数	5	0.951 2	3.774 2	3.756 5	0.941 0	4.138 4	4.119 0	2.961 9
二阶导数	6	0.971 7	2.888 6	2.875 1	0.930 3	4.498 7	4.478 3	2.724 2
SNV+MSC	7	0.947 0	3.928 5	3.910 1	0.933 8	4.376 8	4.359 2	2.798 7
MSC+平滑	8	0.954 7	3.638 8	3.621 7	0.941 8	4.112 3	4.094 0	2.980 0
SNV+平滑	7	0.946 3	3.951 3	3.932 8	0.932 7	4.414 3	4.394 0	2.776 5
SNV+1 阶	4	0.943 0	4.069 9	4.050 8	0.928 1	4.552 7	4.531 5	2.692 3
SNV+2 阶	5	0.960 1	3.420 1	3.404 1	0.925 1	4.645 3	4.623 7	2.638 6

续表 2

MSC+1 阶	4	0.943 0	4.070 0	4.051 0	0.932 7	4.409 3	4.388 6	2.779 9
MSC+2 阶	5	0.960 0	3.423 6	3.407 6	0.923 5	4.691 1	4.669 2	2.612 9
平滑+1 阶	5	0.948 5	3.873 1	3.855 0	0.938 3	4.227 7	4.207 9	2.899 3
平滑+2 阶	5	0.951 1	3.777 6	3.759 9	0.930 3	4.488 5	4.468 1	2.730 5
MSC+平滑+1 阶	4	0.939 0	4.205 8	4.186 1	0.927 5	4.570 4	4.549 3	2.681 7
MSC+平滑+2 阶	4	0.944 5	4.017 3	3.998 5	0.932 1	4.430 6	4.410 4	2.766 2
SNV+平滑+1 阶	4	0.939 0	4.206 0	4.186 3	0.927 3	4.574 9	4.554 4	2.678 7
SNV+平滑+2 阶	4	0.944 6	4.013 8	3.995 0	0.928 3	4.546 5	4.526 3	2.695 4
平滑+1 阶+2 阶	6	0.961 4	3.363 4	3.347 6	0.933 0	4.408 2	4.387 7	2.780 5

2.3 稻米胶稠度近红外分析的最优波长段

从图 1 可知校正集样品吸收峰主要分布在 950~1 100, 1 100~1 300和 1 300~1 650 nm, 所以分别对这 3 个波长段及组合 950~1 100, 1 100~1 300, 1 300~1 650, 950~1 300, 1 100~1 650 和 950~1 650 nm 等 6 个波长段的稻米胶稠度近红外光谱经平滑处理后, 用偏最小二乘法(PLS)建立各波长段的稻米胶稠度近红外分析模型。由表 3 可知在波长段(1 100~1 650 nm)的建模效果最好, 其校正集模型评价指标  $R$ , SEC 和 RMSEC 分别为 0.969 4, 3.003 0 和 2.989 0; 验证集模型评价指标  $r$ , SECV 和 RMSECV 分别为 0.963 8, 3.261 1 和 3.245 9, 相对百分比偏差(RPD)为

3.758 6, 吸收峰越高其相关性越大。

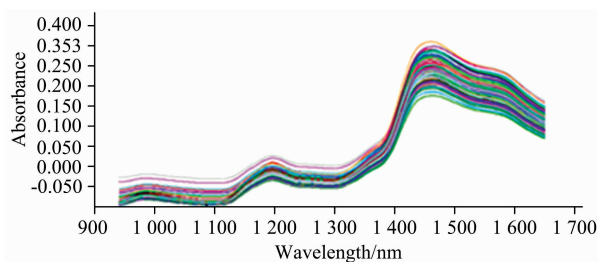


图 1 校正集近红外光谱

Fig. 1 Near infrared spectra of the calibration samples

表 3 稻米胶稠度在不同波长段近红外分析模型的评价指标

Table 3 Evaluation Indexes of near infrared analysis models of rice gel consistency at different wavelength bands

波长段/nm	主因子	校正集			验证集			RPD
		$R$	SEC	RMSEC	$r$	SECV	RMSECV	
950~1 100	7	0.718 4	8.503 9	8.464 1	0.545 8	10.481 4	10.433 1	1.169 4
1 100~1 300	8	0.942 1	4.099 1	4.079 9	0.923 3	4.721 8	4.699 9	2.595 8
1 300~1 650	9	0.953 8	3.672 6	3.655 4	0.942 1	4.101 1	4.082 0	2.988 7
950~1 300	10	0.957 4	3.529 4	3.512 8	0.938 7	4.218 0	4.198 2	2.906 0
1 100~1 650	7	0.969 4	3.003 0	2.989 0	0.963 8	3.261 1	3.245 9	3.758 6
950~1 650	6	0.970 0	2.971 8	2.957 9	0.963 2	3.287 7	3.272 6	3.727 9

2.4 稻米胶稠度近红外分析模型的建立

通过软件分析, 如图 2, 图 3, 剔除残差高的 11 个异常样品后模型评价指标进一步提高, 校正集评价指标  $R$ , SEC 和 RMSEC 分别为 0.979 0, 2.540 6 和 2.527 3; 验证集的评价参数  $r$ , SECV 和 RMSECV 分别为 0.974 1, 2.820 6 和 2.805 8, RPD 为 4.419 4, 提高了模型的准确性。

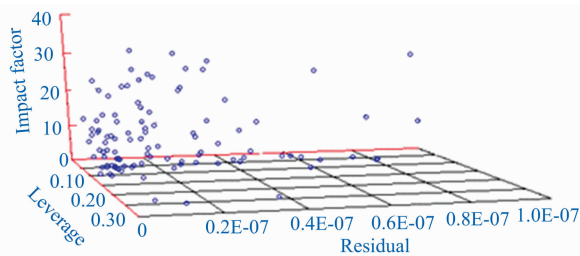


图 2 稻米胶稠度近红外分析的残差和杠杆值

Fig. 2 The leverage and residual of rice gel consistency by near infrared analysis

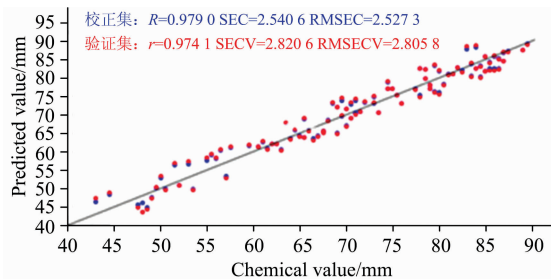


图 3 稻米胶稠度建立的回归模型预测值与其化学值的相关性图

Fig. 3 The correlation of predicted value and chemical value of rice gel consistency established by regression model

2.5 稻米胶稠度近红外模型的检验

用 DA7200 收集 30 份验证集样品的近红外光谱值, 用新建立的稻米胶稠度近红外模型检测其胶稠度, 并将稻米胶稠度的化学值与近红外光谱检测值进行对比分析, 结果如表 4, 其绝对误差范围在 0.198 6~6.502 4 mm, 均符合 GB/T

22294—2008《大米胶稠度的测定》误差标准,胶稠度 40~60 mm 时误差不超过 5 mm,胶稠度大于 60 mm 时误差不超过 7 mm。

对验证集稻米胶稠度近红外检测值和化学测定值进行配

对  $t$  检验,结果如表 5,  $p$  值为  $0.726 > 0.05$ ,表明用此近红外模型检测和化学方法分析得到的结果无显著差异,即用此新建的近红外模型检测稻米胶稠度是可靠的。

表 4 验证集模型预测值与化学测定值

Table 4 Predicted value and chemical value of validation set

样品	预测值	化学值	绝对误差	样品	预测值	化学值	绝对误差
1	48.294 5	43.5	4.794 5	16	75.904 8	73.0	2.904 8
2	52.380 1	47.5	4.880 1	17	70.315 9	74.0	3.684 1
3	57.740 7	53.5	4.240 7	18	71.725 1	74.5	2.774 9
4	54.501 1	56.0	1.498 9	19	70.693 6	76.5	5.806 4
5	55.580 0	57.0	1.420 0	20	75.629 2	77.0	1.370 8
6	57.957 7	60.5	2.542 3	21	78.783 3	78.0	0.783 3
7	57.957 7	62.5	4.542 3	22	76.977 8	79.5	2.522 2
8	63.698 6	63.5	0.198 6	23	74.497 6	81.0	6.502 4
9	67.740 0	64.5	3.240 0	24	84.062 9	81.5	2.562 9
10	66.835 4	65.5	1.335 4	25	77.737 6	82.5	4.762 4
11	68.602 2	67.0	1.602 2	26	87.601 7	83.0	4.601 7
12	73.697 1	68.0	5.697 1	27	83.284 8	84.5	1.215 2
13	64.304 1	68.5	4.195 9	28	83.049 9	85.0	1.950 1
14	71.836 8	70.5	1.336 8	29	82.190 9	86.0	3.809 1
15	73.229 0	72.0	1.229 0	30	89.367 6	88.0	1.367 6

表 5 验证集稻米胶稠度的模型检测值和化学测定值的配对  $t$  检验

Table 5 Paired  $t$  test of model prediction value and chemical value of rice gel consistency in validation set

验证集样本(30 个)						
指标	$t$ 检验	自由度	sig. (双侧)	均值差值	95%置信区间	
					下限	上限
统计值	-0.353	29	0.726	0.643 33	-1.543 17	1.088 35

### 3 结 论

用 DA7200 扫描获得了 137 份南方稻区稻米的近红外漫反射光谱,样品胶稠度范围为 43.0~94.0 mm,覆盖了大部份稻米胶稠度范围。采用偏最小二乘法(PLS)分别经 20 种预处理方式和 6 组波长段下建立近红外分析模型。根据模型评价指标判断最佳预处理方式为平滑处理,最佳建模波长段为 1 100~1 650 nm。对经过平滑预处理的 1 100~1 650 nm 波长段建立的近红外分析模型进行参数评价,校正集的  $R$ , SEC 和 RMSEC 分别为 0.979 0, 2.540 6 和 2.527 3,验证集的  $r$ , SECV, RMSECV 和 RPD 分别为 0.974 1, 2.820 6, 2.805 8 和 4.419 4,模型参数较优。用新建的近红外模型对 30 份验证集样品(未参与建模)进行检测,与化学值进行配对  $t$  检验,近红外检测值与化学值绝对误差在 0.198 6~6.502 4 mm,配对  $t$  检验  $p=0.726 (>0.05)$ ,表明此近红外

分析模型可用于稻米胶稠度的检测。与同组分的其他研究结果相比<sup>[2]</sup>,此近红外分析模型检测效果更佳,其建模评价指标和外部验证结果都更优,能用于稻米胶稠度的分析检测,并能为今后稻米胶稠度的近红外分析的相关研究提供理论参考。

近红外光谱分析在农业领域的研究及应用还有待加强。为了提高模型的准确性,采用了 20 种不同的预处理方式和 6 组波长段处理,发现经平滑处理的 1 100~1 650 nm 波长段建立的近红外分析模型效果最佳,且经验证表明此模型可用于稻米胶稠度的分析检测。是否还有更好的预处理方法?是否在其他波长段包含更多的胶稠度信息?是否其他建模方法建立的模型更精确?这些有待进一步探索。因此,为了获得更准确的近红外分析模型,(1)可以尝试足够多的光谱预处理方法或者开发新的预处理方法以减少光谱的影响因素;(2)可考虑用不同的回归分析模型进行建模以选择更优的建模方法。

## References

- [ 1 ] SHI Lü, ZHANG Xin-yue, SUN Hui-yan, et al(石 吕, 张新月, 孙惠艳, 等). Chinese Journal of Rice Science(中国水稻科学), 2019, 33(6): 541.
- [ 2 ] XU Ming-ya, ZHAO Zi-wei, DU Jing-lin, et al(徐明雅, 赵紫薇, 杜京霖, 等). Journal of Nuclear Agricultural Sciences(核农学报), 2019, 33(5): 936.
- [ 3 ] LIU Hong-mei, XIAO Zheng-wu, SHEN Tao, et al(刘红梅, 肖正午, 申 涛, 等). Journal of Hunan Agricultural University • Science and Technology(湖南农业大学学报 • 自然科学版), 2019, 45(2): 189.
- [ 4 ] Olivares Díaz Edenio, Kawamura Shuso, Matsuo Miki, et al. Food Chemistry, 2019, 286: 297.
- [ 5 ] Irina Torres, María-Teresa Sánchez, María-José de la Haba, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 217: 206.
- [ 6 ] Mazlina Mohd Said, Simon Gibbons, Anthony Moffat, et al. Journal of Near Infrared Spectroscopy, 2019, 27(5): 379.
- [ 7 ] ZHANG Zhi, ZHANG Kai-fei, CUI Ying-tao, et al(张 志, 张开飞, 崔迎涛, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(Suppl.): 63.
- [ 8 ] LI Guan-wen, GAO Xiao-hong, XIAO Neng-wen, et al(李冠稳, 高小红, 肖能文, 等). Chinese Journal of Luminescence(发光学报), 2019, 40(8): 1030.

# Construction and Verification of a Mathematical Model for Near-Infrared Spectroscopy Analysis of Gel Consistency in Southern *Indica* Rice

LIU Hong-mei, SHEN Tao, ZHANG Wen-yi, SHI Xi-wen, DAI Tao, BAI Tao, XIAO Ying-hui\*

College of Agronomy, Hunan Agricultural University, Changsha 410128, China

**Abstract** Cultivating high-quality and high-yielding rice varieties is one of the important tasks of current rice breeding, and gel consistency is one of the most important indicators of rice cooking and eating quality. The traditional chemical method for measuring the gel consistency of rice has complicated pretreatment, complicated process, and high reagent consumption. It is difficult to meet the needs of rapid non-destructive testing of the gel consistency for large batches of rice varieties (combinations). The near-infrared spectroscopy analysis technology can quickly, non-destructively, and non-pollutingly analyze. In this study, 137 *indica* rice varieties (combinations) mainly cultivated or newly bred in southern rice areas were used as the test materials, and their near-infrared spectra were measured by traditional chemical methods to collect their near-infrared spectra to establish near-infrared spectra of southern *indica* rice. Build an analysis model, and then correct and verify the model. The results showed that the partial least squares method (PLS) was used to establish their respective near infrared analysis models after 20 kinds of mathematical preprocessing and 6 kinds of wavelength bands (or combinations). By comparing the model evaluation indexes, it was determined that smooth pretreatment was the best pretreatment method, and the wavelength band of 1 100 to 1 650 nm was the best modeling wavelength band. Evaluation index of smoothing model: calibration correlation coefficient ( $R$ ), test correlation coefficient ( $r$ ), relative analysis error (RPD) were 0.970 0, 0.964 2, and 3.780 5 respectively; wavelength evaluation range: 1 100 to 1 650 nm Model evaluation indicators:  $R$ ,  $r$ , RPD They were 0.969 4, 0.963 8, and 3.758 6 respectively; after smoothing, the best near-infrared analysis model of rice gel consistency was established in the wavelength range of 1 100 to 1 650 nm, and the model evaluation indicators:  $R$ ,  $r$ , and RPD were 0.979 0, 0.974 1, and 4.419 4 respectively; Then used 30 samples to verify the obtained optimal model. It indicated that the absolute error between the near infrared predicted value and the chemical value was 0.198 6~6.502 4 mm, and paired  $t$  test showed that  $p=0.726>0.05$ , indicating no significant difference between the predicted value and the chemical value. The near-infrared model was feasible for rapid non-destructive testing of rice gel consistency. This study provides technical support for the rapid screening of high-quality rice varieties (combinations) in the early generation of materials and the rapid batch analysis of the gel consistency of rice.

**Keywords** *Indica* rice; Gel consistency; Near-infrared model; Partial least squares