

近红外光谱分析中温度影响的修正

孙彦华^{1,2}, 范永涛^{1,2*}

1. 中国科学院上海光学精密机械研究所微纳光子功能材料实验室, 强激光材料重点实验室, 上海 201800
2. 中国科学院大学材料与光电研究中心, 北京 100049

摘要 样品温度变化会对模型预测结果产生影响, 为解决这个问题, 首先, 对同一样品不同温度下的光谱与同一样品相同温度的光谱进行了比较。结果显示, 不同温度下的光谱差异较大。然后研究了样品温度对玉米粗蛋白模型的预测结果的影响, 对随机选取的粗蛋白含量为 6.04% 的样品不同温度采集光谱, 对这些光谱进行预处理消除温度之外的因素对光谱的影响, 将预处理后的光谱代入已建立好的模型中进行预测, 结果显示, 预测结果与实测值之间的差别随着光谱温度与建模温度相之间差别的增大而增大, 最大的误差为 1.12%。为了解温度对模型预测结果的影响, 进而分析了温度与不同温度下的光谱数据之间的关系, 发现在去除了光谱两端噪声较严重的区域后, 不同温度下, 同一波长点处的光谱数据之间存在一定的线性关系。依据这一发现, 文中提出了温度修正理论, 以建模时的光谱为基准光谱, 然后根据温度与光谱之间的线性关系使用线性回归算法对不同波长点的光谱进行一元线性回归, 求出不同温度下的光谱与基准光谱之间的差, 最后将不同温度下的光谱校正为基准光谱, 通过该理论对光谱进行校正之后, 不同温度下的同一样品的光谱之间的差别和修正之前相比已经有了很大改善, 将修正后的光谱代入模型, 大部分预测结果得到了提高, 符合国家标准 $\pm 0.5\%$ 以下的要求。最后使用和建模无关的 34 个不同含量的样品对该温度修正理论进行验证, 光谱修正前后粗蛋白的模型预测值与标准理化值决定系数分别为 0.910 和 0.982, 均方根误差分别为 0.558 和 0.172, 平均相对误差分别为 6.05% 和 1.75%。该温度修正理论从近红外光谱分析的本质进行了温度修正, 为其他样品的温度修正提供了参考, 有利于手持式近红外光谱仪使用的推广。

关键词 近红外光谱法; 温度修正; 线性回归; 粗蛋白测量

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)06-1690-06

引言

近红外光谱分析由于具有快速、无损、高效检测的优点, 被广泛应用于各个领域^[1-4]。随着科学家对于近红外光谱仪研究的深入, 出现了许多新型近红外光谱仪, 并大大拓展了近红外光谱仪的应用场景, 已经由实验室检测走向现场检测^[5-6]。

由于近红外光谱属于分子振动光谱, 温度改变时, 样品内分子的振动加剧或减弱, 分子间的作用力发生变化, 进而影响到分子振动或转动状态在不同能级之间的跃迁, 从而影响分子的振动光谱^[7], 所以当近红外光谱仪由温度精确的实验室环境进入到现场检测环境时, 检测结果易受温度变化的影响。本课题研究发现, 当温度在 0~25 °C 的范围内变化时,

粗蛋白含量为 6.04% 的样品的预测值与实测值之间的最大偏差为 1.12%。已有报道的研究工作也表明样品温度的不同会对检测结果产生影响, 如王冬^[8]等研究了样品温度对复配乳油的近红外光谱定量分析模型的影响; Li^[9]等对不同温度下的花生油、豆油、玉米油的混合物的吸光度光谱进行分析, 结果发现, 温度对光谱的影响随着体积比的增加越来越明显。因此温度对吸光度光谱的影响不容忽视, 在检测过程中需要去除温度的影响。

常用的温度修正方法有化学计量学方法、全局隐含或显式温度补偿^[8]、去除对温度敏感的波长^[10]等。化学计量学方法适用于含水量较大的样品, 具有局限性; 全局隐含或温度补偿方法在建模时需要测量不同温度下的光谱及样品实测值, 加大了工作量; 去除对温度敏感的波长会降低模型的精度。以上方法虽然在特定的条件下可以对温度产生的影响进

收稿日期: 2019-05-21, 修订日期: 2019-09-19

基金项目: 国家自然科学基金青年科学基金项目(61405223)资助

作者简介: 孙彦华, 1992年生, 中国科学院上海光学精密机械研究所硕士研究生 e-mail: syh_26@163.com

* 通讯联系人 e-mail: fyt@siom.ac.cn

行修正,但都存在一定的缺陷;为了解决温度产生的影响,本课题探究了温度与光谱之间的关系,提出了温度修正理论,从近红外光谱分析的本质对温度产生的影响进行了修正,有效提高了模型的预测结果。

1 实验部分

1.1 样品

玉米样品来自上海嘉定区某饲料厂,不同品牌、不同产地、籽粒大小均匀、表面无霉斑的玉米样品总计 97 个。其中 63 个样品作为建模用。将样品使用粉碎机粉碎,过 40 目筛子筛选后,每个样品一分为二,分别装入密封的自封袋中进行保存,做好标记。将其中一份自封袋中样品放在室温环境(24 °C)内一晚,另一份的样品放置在冰箱冷冻室内使样品的温度达到-5 °C,保障玉米样品的温度保持一致,以备用;另外 34 个样品用作验证本温度修正理论的正确性,对其作与建模样品相同的处理。

1.2 手持式近红外光谱采集系统

近红外光谱采集系统为本课题组自行研发的手持式近红外光谱仪,该手持式近红外光谱仪形状区别于目前市场上存在的“手枪状”手持式近红外光谱仪^[11],为“万用表”操作模式,仪器整体结构如图 1 所示,仪器整体大小与手持式万用表相当,可单手操作,非常方便携带。此检测仪采用直接从透明样品杯底部采集光谱的方法,采样过程中样品与检测仪的位置、角度始终保持一致,完全不受使用者操作方式的影响,大幅提高了检测仪的采样精度,并降低了使用难度,解决了现有手持式近红外光谱仪测量结果易受其光学窗口与样品角度及贴合度的细微差别的影响。



图 1 手持式近红外光谱仪

Fig. 1 Handheld near infrared spectrometer

该手持式近红外光谱仪的波长范围在 900~1 700 nm,波长分辨率为 16 nm,积分时间设置为 80 ms,平均次数为 10 次。该仪器自带温度传感器,可以在采集光谱的同时测量该条光谱相对应的样品温度。

1.3 标准理化值测量方法

粗蛋白理化值测量使用凯氏定氮法^[12-13],将玉米样品与硫酸和催化剂一同放入消化炉(HYP-340,上海纤检仪器有限公司)内加热消化,使粗蛋白分解,然后将消化后的样品

使用定氮仪(KDN818,上海纤检仪器有限公司)进行测量,在测量的过程中每个样品做了 3 个平行样,取 3 次测量的平均值作为每个样品的实测值,该方法通过理化分析含氮量来确定样品的粗蛋白含量,耗时相对较长,但精度较高,玉米粗蛋白含量测定结果如表 1 所示。

表 1 玉米粗蛋白的测定结果

Table 1 Statistics results of corn crude protein content measurement

品质参数	样本集	范围	平均值
粗蛋白/%	校正集	5.77~8.5	7.49
	验证集	6.03~8.45	7.44
	预测集	6.34~8.47	7.64

1.4 光谱采集

将手持式近红外光谱仪预热 10 min,先进行暗光谱和参考光谱的采集,然后将常温下的样品装入样品杯,放置在光谱仪上,每次旋转 30°,测量 6 次,6 次测量光谱的平均光谱代表样品真实光谱。

采集不同温度下的样品光谱时,将样品依次快速从冰箱中取出进行装样,然后在样品温度上升的过程中连续进行光谱的扫描并把各个光谱的相应温度使用仪器自带的温度传感器记录下来。

对预测集样品进行光谱采集时,早上 8 点将样品放置在室外,然后在下午 13 点至 14 点之间进行光谱的采集。

1.5 模型的建立

将 63 个样本按照 2:1 的比例分为校正集和验证集,使用室温(24 °C)下采集的样品的光谱建立偏最小二乘回归(partial least squares regression, PLS)模型,该模型校正集和验证集的相关系数(coefficient of determination, R)分别为 0.988 2 和 0.993 0,校正集标准差(root mean square error of calibration, RMSEC)为 0.144,验证集标准差(root mean square error of prediction, RMSEP)为 0.213。

2 结果与讨论

2.1 温度对光谱的影响

不同于台式设备,手持设备经常在户外使用,气温波动范围一般达到-10~35 °C,若不加处理,将对测量精度产生严重影响。待测样品的反射光谱受环境温度影响剧烈,我们将不同温度下的同一样品的吸光度光谱进行预处理,排除噪声及光程差带来的光谱变化,只考虑温度对吸光度光谱的影响,试验结果如图 2(a)所示。图 2(b)为同一样品相同温度下的光谱,由图 2(a)和图 2(b)对比可知,温度对光谱的影响十分明显。

2.2 温度对预测结果的影响

选取粗蛋白含量为 6.04% 的样品,温度从 0 °C 变化至 24 °C 过程中一共采集了 23 条光谱,将这 23 条不同温度下的光谱依次代入建立好的模型中进行预测,预测结果如图 3 所示。从图中可以看出样品温度与建模温度相差越大时预测结

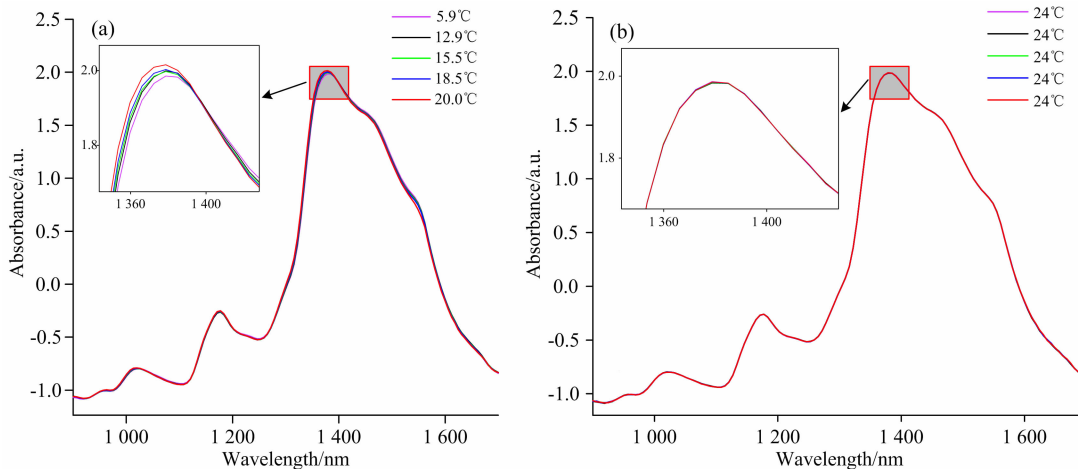


图 2 温度对样品光谱的影响

(a): 不同温度下同一样品的光谱; (b): 相同温度下同一样品的光谱

Fig. 2 The effect of temperature on sample spectrum

(a): Spectra of the same sample at different temperatures; (b): Spectra of the same sample at the same temperature

果与实测值之间的误差也越大, 预测值与实测值之间最大的偏差为 1.12%, 已超出国标所允许范围^[14]。

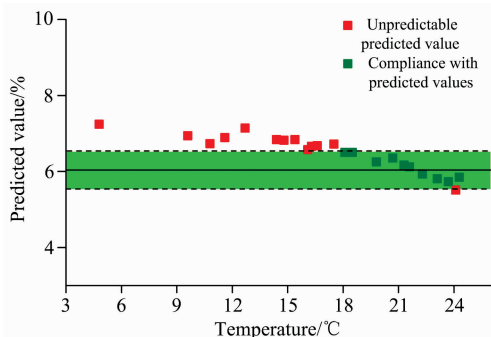


图 3 粗蛋白含量为 6.04% 的样品不同温度下的光谱得到的预测值

Fig. 3 Predicted values of spectra at different temperatures for samples with a crude protein content of 6.04%

2.3 温度修正原理

该温度修正原理是基于吸光度的校正, 宏观上, 温度对样品的吸光度的影响是一个小量; 通过试验对温度和光谱之间的关系进行探究, 试验结果如图 4 所示。

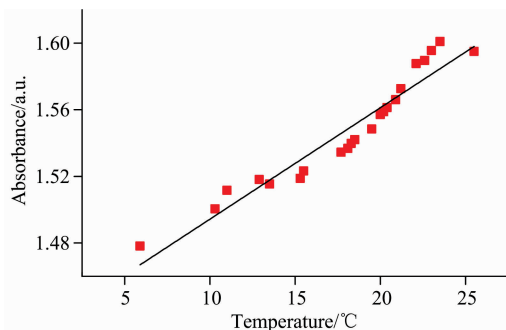


图 4 温度与吸光度的关系

Fig. 4 The relationship between temperature and absorbance

由图 4 可知, 温度与吸光度在一定范围内满足线性关系。基于此, 将同一波长点处不同温度下的吸光度值 $A_{\lambda_i T_i}$ 与温度 T_i 进行一元线性回归

$$A_{\lambda_i T_i} = b_{\lambda_j 0} + T_i b_{\lambda_j 1} \tag{1}$$

式中, $A_{\lambda_j T_i}$ 为同一样品同一波长点处不同温度下的吸光度值, T_i 为不同的温度, λ_j 为不同的波长, $b_{\lambda_j 1}$ 为线性回归系数, $b_{\lambda_j 0}$ 为线性回归常数项, 表示除自变量以外其他因素对因变量的平均影响量。其最小二乘解为

$$\begin{cases} b_{\lambda_j 1} = \frac{n \sum T_i A_{\lambda_j} - \sum T_i \sum A_{\lambda_j T_i}}{n \sum T_i^2 - (\sum T_i)^2} \\ b_{\lambda_j 0} = \frac{\sum A_{\lambda_j T_i} - b_{\lambda_j 1} \sum T_i}{n} \end{cases} \tag{2}$$

以建立模型时样品的温度为基准, 求取要预测样品的温度与基准温度的差值 ΔT_i , 代入式(1)求得由于温度的影响, 吸光度的改变值 ΔA_{λ_j} 。校正之后的吸光度与原始吸光度存在式(3)关系

$$A_{\lambda_j 1} = A_{\lambda_j 0} - \Delta A_{\lambda_j} \tag{3}$$

通过以上操作, 不同温度下的样品吸光度 $A_{\lambda_j 0}$ 将会被修正为统一温度下样品的吸光度 $A_{\lambda_j 1}$, 以 $A_{\lambda_j 1}$ 代入模型进行计算, 一定程度上可以消除环境温度对样品吸光度的影响。

2.4 温度修正试验结果

利用提出的温度修正理论, 以该样品的建模温度为基准, 将不同温度下的光谱分别进行修正, 修正后的光谱图如图 5 所示, 从图中可以看出, 不同温度下的样品的光谱之间的差异和未修正时相比已经有了显著改善。

为了验证温度修正理论的正确性, 还需要将修正后的光谱代入室温(24 °C)下已建立好的的模型进行预测, 预测结果如图 6 所示。通过比较光谱修正前后各个温度下的预测结果, 对温度修正理论的效果进行评价。

图中实横线为样品的实测值, 两条虚线为国家标准允许的误差范围的临界值, 预测值越接近实横线说明模型预测效

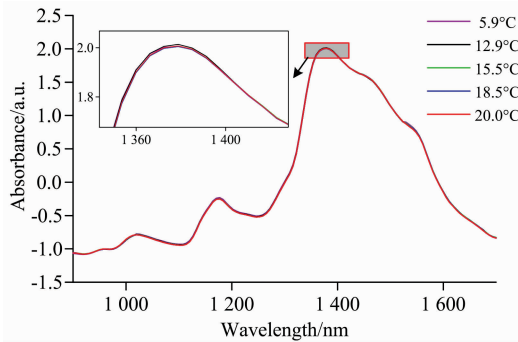


图 5 经过温度修正理论修正后的光谱图

Fig. 5 Spectra obtained by temperature correction theory

果越好，在两条虚线之间的预测值可以满足国家标准的要求。由图 6 可知，样品不同温度下的近红外光谱经过温度修正后的预测结果大部分都已符合国家标准所要求的误差范围，对比图 3 和图 6 发现，粗蛋白含量为 6.04% 的样品的预测值与实测值的误差范围从未修正前的 $\pm 1.12\%$ 下降为 $\pm 0.5\%$ ，由此可以看出，此温度修正理论可以很好地修正温度对近红外光谱产生的影响。

2.5 预测模型验证

为验证本文中提出的温度修正模型的正确性，将采集得到的 34 个不同含量的预测集样品的未经过温度理论修正的

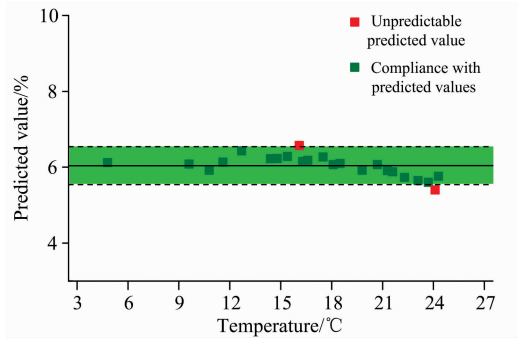


图 6 经过温度修正理论修正后的不同温度下的预测值

Fig. 6 Predicted values at different temperatures obtained by temperature correction theory

光谱和经过温度理论修正后的光谱分别代入已建立好的模型进行预测，分别与利用标准方法测定的理化值进行比较分析。

如图 7 所示，玉米样品粗蛋白含量的模型预测值与标准理化值决定系数分别为 0.910 和 0.982，均方根误差分别为 0.558 和 0.172，最大相对误差 (maximum relative error, MRA) 分别为 15.21% 和 5.01%，平均相对误差 (average relative error, ARE) 分别为 6.05% 和 1.75%。基于提出的温度修正理论可以很好的消除温度对预测结果产生的影响，可以实现手持式近红外光谱仪现场检测的需求。

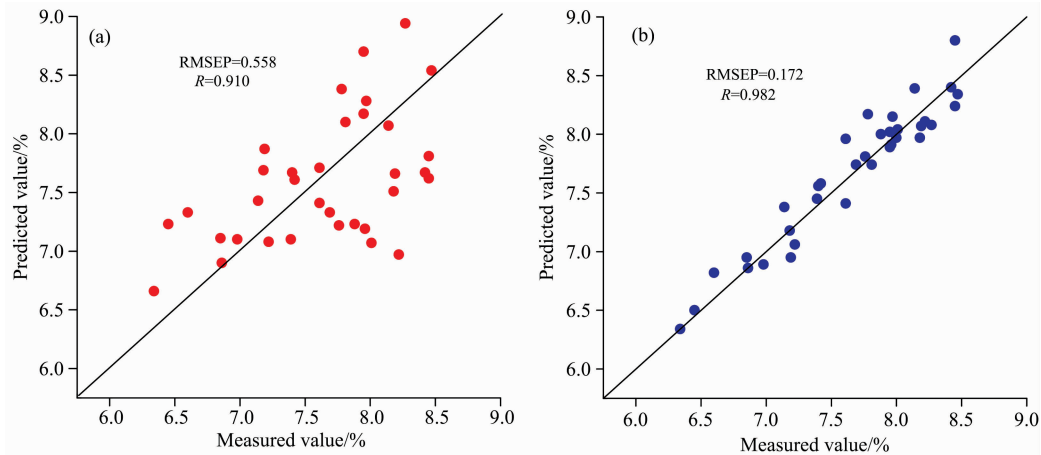


图 7 温度修正前后玉米外部验证模型效果

(a): 光谱未进行修正的预测模型;

(b): 光谱经过温度修正之后的预测模型

Fig. 7 Effect of the external validation model of corn before and after temperature correction

(a): A prediction model in which the spectrum is not corrected;

(b): A prediction model after the spectra corrected by this proposed method

3 结 论

利用自主研制的手持式近红外光谱仪采集玉米样品的近红外光谱，研究样品温度变化对模型预测结果的影响。通过分析发现样品吸光度光谱与温度之间存在线性关系，利用这

一关系提出了温度修正理论，试验结果证明，该温度修正理论可以对光谱进行修正，修正后的光谱代入已建立好的模型预测结果有了很大的改善。最后，使用 34 个样本对温度修正理论进行了检验，光谱修正前后粗蛋白的模型预测值与标准理化值决定系数分别为 0.910 和 0.982，均方根误差分别为 0.558 和 0.172，平均相对误差分别为 6.05% 和 1.75%。表

明, 温度修正理论可以很好的修正温度对样品光谱产生的影响, 可以实现样品的快速无损检测。由于该温度修正理论简单易懂并从近红外光谱分析的本质对温度进行了修正, 所

以方法稳定可靠。该方法不仅可以为修正其他样品温度对模型预测结果的影响提供重要的参考, 而且也可以推动手持式近红外光谱仪的发展。

References

- [1] HUANG Ya-wei, LI Huan, WANG Ruo-lan(黄亚伟, 李 换, 王若兰). *Cereals & Oils(粮食与油脂)*, 2017, 30(7): 2.
- [2] Fu Xiaping, Kim Moon S, Chao Kuanglin, et al. *Journal of Food Engineering*, 2014, 124: 97.
- [3] Jan U Porep, Dietmar R Kammerer, Reinhold Carle. *Journal of Food Engineering*, 2015, 46: 211.
- [4] Shi Z, Ji W, Viscarra Rossel R A, et al. *European Journal of Soil Science*, 2015, 66: 679.
- [5] XU Xiu-qin, CHEN Guo, ZHANG Hao, et al(许秀琴, 陈 国, 章 豪, 等). *Anhui Chemical Industry(安徽化工)*, 2017, 43(4): 7.
- [6] WANG Fan, LI Yong-yu, PENG Yan-kun, et al(王 凡, 李永玉, 彭彦坤, 等). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2018, 49(7): 348.
- [7] REN Dong, QU Fang-fang, LU An-xiang, et al(任 东, 瞿芳芳, 陆安祥, 等). *Near Infrared Spectroscopy and Its Application(近红外光谱分析技术与应用)*. Beijing: Science Press(北京: 科学出版社), 2016. 51.
- [8] WANG Dong, XIONG Yan-mei, HUANG Rong, et al(王 冬, 熊艳梅, 黄 蓉, 等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2010, 38(9): 1313.
- [9] Li Zhe, Zhou Mei, Luo Yongshun. *Talanta*, 2016, 155: 47.
- [10] Wang H L, Peng J Y, Xie C Q, et al. *Sensors*, 2015, 15(5): 11889.
- [11] GUO Zhi-ming, CHEN Quan-sheng, ZHANG Bin, et al(郭志明, 陈全胜, 张 彬, 等). *Transactions of the Chinese of Agricultural Engineering(农业工程学报)*, 2017, 33(8): 245.
- [12] GUO Ying-shi, CAO Xiao-yan, ZOU Hang-jun, et al(郭应时, 曹小彦, 邹杭君, 等). *Food & Machinery(食品与机械)*, 2017, 33(11): 67.
- [13] WANG Qi(王 琦). *Journal of Changzhi University(长治学院学报)*, 2018, 35(2): 59.
- [14] GB/T18868—2002, National Standards of the People's Republic of China(中华人民共和国国家标准). Method for Determination of Moisture, Crude Protein, Crude Fat, Crude Fibre, Lysine and Methionine in Feeds-Near Infrared Reflectance Spectroscopy(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).

Correction of Temperature Influence in Near Infrared Spectroscopy

SUN Yan-hua^{1, 2}, FAN Yong-tao^{1, 2*}

1. Laboratory of Micro-Nano Optoelectronic Materials and Devices, Key Laboratory of Materials for High-Power Laser, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
2. Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract For the problem that the temperature change of the sample affects the prediction result of the model, firstly, the spectrum of the same sample at different temperatures is compared with the spectrum of the same sample at the same temperature. The results show that the spectral difference at different temperatures is large. Then the effect of sample temperature on the prediction of corn crude protein model was studied. Spectral collection of samples with a crude protein content of 6.04% was performed at different temperatures, and near-infrared spectra at different temperatures were pretreated in the same way as those used in modeling, so as to eliminate the influence of factors other than the temperature on the spectra. The pre-processed spectrum is substituted into the established model for prediction. The prediction results show that the difference between the predicted result and the measured value increases as the difference between the spectral temperature and the modeled temperature increases, and the maximum error is 1.12%. In order to solve the influence of temperature on the prediction results of the model, we further analyzed the relationship between temperature and spectral data at different temperatures, and found that after removing the areas with serious noise at both ends of the spectrum, there was a certain linear relationship between spectral data at the same wavelength point at different temperatures. According to this finding, a temperature correction theory is proposed. Taking the spectrum at the time of modeling as the reference spectrum, and then using the linear regression algorithm to perform linear regression on the spectra of different wavelength points according to the linear relationship between temperature and spectrum, the difference between the spectrum at different temperatures and the reference spectrum is obtained. Finally, the

spectra at different temperatures are corrected to the reference spectrum. After the spectrum is corrected by the theory, the difference between the spectra has been greatly improved compared with before the correction. The corrected spectrum is brought into the model, and most of the prediction results are improved, which meets the requirements of $\pm 0.5\%$ of the national standard. Finally, the temperature correction theory was verified by using 34 different samples unrelated to the modeling. The model prediction values and standard physical and chemical value determination coefficients of the crude protein before and after the spectral correction were 0.910 and 0.982, respectively, and the root means square error was 0.558 and 0.172, and the average relative error was 6.05% and 1.75%, respectively. The temperature correction theory has been temperature-corrected from the nature of near-infrared spectroscopy, providing a reference for temperature correction of other samples, which is beneficial to the promotion of handheld near-infrared spectroscopy.

Keywords Near-infrared spectroscopy; Temperature correction; Linear regression; Crude protein measurement

(Received May 21, 2019; accepted Sep. 19, 2019)

* Corresponding author