

# 利用光谱趋势参数快速判定小麦粉 DON 等级的研究

吴威<sup>1</sup>, 祖广鹏<sup>1</sup>, 陈桂云<sup>1</sup>, 徐剑宏<sup>2</sup>, 陈坤杰<sup>1\*</sup>

1. 南京农业大学工学院, 江苏 南京 210031

2. 江苏省农业科学院食品质量与安全检测研究所, 江苏 南京 210014

**摘要** 脱氧雪腐镰刀菌烯醇(DON)是一种经常发生在谷物及其衍生产品中的霉菌毒素,危害人类和家畜的生命健康。开发快速、准确、经济、无环境危害的检测方法是一个亟待解决的问题。本研究定义了可见光-近红外(Vis-NIR)光谱的趋势参数 TP(trend parameter),利用 TP 确定与 DON 浓度最相关的特征波段。文中校正集样本的光谱矩阵行按样本 DON 浓度逐渐增加的顺序排列,矩阵每一列(每一个波段)都对应一个 TP 值,所有样本在某波段下的吸光度在列方向上的递增趋势越强(即 TP 值越大),则此波段下的吸光度与 DON 浓度的相关性就越强,该波段便可以作为评估 DON 浓度的特征波段。研究发现在 666, 1 238 和 1 660 nm 处 TP 出现局部最大值,利用此三个特征波段下的光谱进行二次判别分析 Quadratic Discriminant Analysis(QDA),以此构建的 TP-QDA 模型可以将小麦粉按 DON 污染水平分成轻度( $0 < \text{DON} < 1\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ )、中度( $1\ 000 \leq \text{DON} < 2\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ )、和重度( $\text{DON} \geq 2\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ )三个等级。该模型的整体分类准确率在校正集和验证集中分别为 88.24% 和 86.27%。对比了传统的主成分分析 PCA(principal component analysis)特征波段选取方法,其所构建的 PCA-QDA 模型也将相同的小麦样品分成三个污染等级,整体分类准确率在校正集和验证集中分别为 68.62% 和 72.55%。这些研究结果证实了 TP 选择特征波段的方法在判断 DON 污染水平时优于 PCA 特征波段选取方法,并且 TP-QDA 模型可有效地用于快速对小麦粉的污染等级进行分类,从而减少收储运过程中分析筛选小麦的时间和经济成本。研究结果还有待在更广泛的小麦品种中进行普适性验证。

**关键词** 脱氧雪腐镰刀菌烯醇;趋势参数;二次判别分析;小麦粉

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1565-04

## 引言

脱氧雪腐镰刀菌烯醇(DON),属单端孢霉烯族化合物,通常存在于感染了赤霉病的谷物中。DON 含量超标的食物具有致癌、致畸、致突变的作用,危害人类和家畜的生命健康<sup>[1-2]</sup>。我国在《食品中真菌毒素限量》(GB2761—2011)标准中规定了小麦等制品 DON 的限量指标为  $1\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ <sup>[3]</sup>。

目前, DON 的检测手段主要有液相色谱-串联质谱法、气相色谱-质谱法分析、免疫亲和层析净化高效液相色谱法等<sup>[4-7]</sup>,这些方法技术成熟,检测的灵敏度高,非常适合监管测试,但是这些方法通常步骤繁琐,需要的设备及耗材昂贵,同时会产生一定量的化学废物,对环境造成一定的破

坏。小麦的收储运过程中有大量的样本需要随时检测,以验证原料是否符合规定,这样就亟需有简单、快速、成本低廉的检测方法来响应市场的需求。光谱学的方法在检测领域有着广泛的应用,用近红外光谱检测食品中的水分、蛋白质、脂肪等成分时可以达到很高的精确度<sup>[8-10]</sup>,但在毒素含量等的痕量检测中效果并不理想。虽然如此,在过去的几十年中,研究者们还是不断地尝试利用光谱学方法对谷类作物中的 DON 含量进行检测,以达到简化检测步骤,降低检测成本,以及提高检测效率的目的。Girolamo<sup>[11]</sup>使用傅里叶变换近红外光谱(FT-NIR)识别含有少于  $300\ \mu\text{g} \cdot \text{kg}^{-1}$  DON 的小麦样品,将低 DON( $< 300\ \mu\text{g} \cdot \text{kg}^{-1}$ )和高 DON( $> 300\ \mu\text{g} \cdot \text{kg}^{-1}$ )样品分类,准确度为 69%;Dvorř Āček<sup>[12]</sup>利用 FT-NIR 光谱仪采集的光谱构建偏最小二乘回归和判别分析 PLS-DA 模型,该技术在 DON 浓度低于  $30\ \text{mg} \cdot \text{kg}^{-1}$  的样

收稿日期: 2019-04-04, 修订日期: 2019-08-02

基金项目: 国家自然科学基金项目(U1604234, 31772118), 江苏省农业科技自主创新资金项目(CX(17)1003)资助

作者简介: 吴威, 1978 年生, 南京农业大学工学院副教授 e-mail: wuwei@njau.edu.cn

\* 通讯联系人 e-mail: kunjiechen@njau.edu.cn

品中得到的相关系数较高( $r=0.92$ ),但是这个统计组合不能可靠地区分 DON 浓度低于欧盟卫生限值  $1\ 250\ \mu\text{g}\cdot\text{kg}^{-1}$  的小麦样品; Annalisa, De Girolamo<sup>[13]</sup> 使用阈值为  $1\ 400\ \mu\text{g}\cdot\text{kg}^{-1}$  的主成分-线性判别分析 PCA-LDA 可以将污染小麦成功地分类成低污染和高污染两种。但当模型使用欧盟卫生限值  $1\ 250\ \mu\text{g}\cdot\text{kg}^{-1}$  DON 的阈值时,约有 40% 误分类样品; Peiris<sup>[14]</sup> 采用 FT-NIR 光谱的偏最小二乘(PLS)模型,将 60.8%~82.3% 的低 DON,以及 82.3%~94% 的高 DON 样本正确分类。可见,目前已有的 NIR 检测 DON 含量的研究中,检测的准确率还都不高,尤其是卫生限值附近的 DON 浓度识别率还不是很理想;另外已有的研究多采用全波段 PLS 分析建模,使得采集的冗余数据较多,也影响检测速度;由于小麦在作为食品原料之前还会进行清洗和去皮等处理,这些处理会大大减少 DON 的含量,因此在收储运环节的检测,可以将未处理小麦 DON 的浓度限值由国标的  $1\ 000\ \mu\text{g}\cdot\text{kg}^{-1}$  提高到  $2\ 000\ \mu\text{g}\cdot\text{kg}^{-1}$ ,并且只需要将小麦按污染水平分成几个等级,即可满足原料筛选的需要。基于以上分析,本研究旨在开发一种简单、快速、廉价的 Vis-NIR 判定方法,通过定义趋势参数(TP)选取与 DON 浓度变化趋势最相关的特征波段,构建判别小麦粉中 DON 含量等级的二次判别分析(QDA)模型,从而快速评估小麦样品的 DON 污染等级。

## 1 实验部分

### 1.1 样本制备和光谱采集

实验共采集了 102 个具有不同 DON 浓度的小麦样品,品种为镇麦 168。这些小麦样本在扬花期接种了赤霉菌毒素。样本经过去皮、磨粉、过筛,其粗细程度满足我国标准粉规格。使用近红外光谱仪 InfraXact (Foss, USA) 收集样品光谱。光谱仪配备硅探测器(570~1 100 nm),铟镓砷探测器(1 100~1 850 nm)。样品以 2 nm 间隔在 570~1 848 nm 范围内扫描,数据采集的频率为每次扫描 3 s。每个样品扫描两次,取其平均值。通过 ISIScan (Infrasoft International, Port Matilda, PA, USA) 收集光谱数据。

### 1.2 DON 测定

DON 含量参考值按照《GB 5009.111—2016 食品安全国家标准食品中脱氧雪腐镰刀菌烯醇及其乙酰化衍生物的测定》<sup>[15]</sup> (第二法免疫亲和层析净化高效液相色谱法 HPLC),使用 HPLC 系统 (SHIMADZU, SPD-10A, UV-VIS DETECTOR; LC-10AT VP) 进行定量测定。

### 1.3 数据预处理

利用 Unscrambler® X, v10.1 (CAMO Software AS, Oslo, Norway, 2011) 软件进行数据预处理。所有样本的光谱首先通过 15 点 Savitzky-Golay 函数平滑,然后通过去趋势处理,最后进行乘法散射校正(MSC)。

### 1.4 基于趋势参数的特征波长选取方法

在光谱矩阵中,每一行对应一个样本的光谱,并把样本按 DON 浓度从低到高进行排列。第一列为样本的 DON 浓度,其后的 640 列为 570~1 848 nm 波段下的光谱值(吸光

度)。选取其中所有的奇数行作为校正集光谱,所有偶数行数据作为验证集光谱,校正集和验证集均有 51 个样本。在这样的光谱矩阵中,单波段下所有样本的吸光度从上到下的递增趋势越强,则说明该波段下的吸光度与 DON 浓度的变化趋势越是一致,则该波段可以作为评估 DON 浓度的特征波段。据此,可以为每一个波段定义一个判断递增趋势的趋势参数,以期找到递增趋势最强的特征波段,具体做法如下:

定义一个参数  $V_j$  来衡量第  $j$  ( $2 \leq j \leq 641$ ) 列的递增趋势,设  $x_{ij}$  ( $1 \leq i \leq 51$ ) 是光谱矩阵中第  $i$  个样本的第  $j$  个波段的吸光度。将  $x_{ij}$  与在该波段下的其他样本的吸光度相比较,当  $x_{ij}$  的值小于位于其下面样本  $x_{mj}$  ( $i < m \leq 51$ ) 的吸光度 ( $x_{ij} < x_{mj}$ ),趋势参数就加 1 分 ( $V_{ij} = V_{ij} + 1$ ); 同样,  $x_{ij}$  的值大于其上面样本  $x_{nj}$  ( $1 \leq n < i$ ) 的吸光度 ( $x_{ij} > x_{nj}$ ),趋势参数也加 1 分,反之不加分。将  $x_{ij}$  与其他 50 个样本在该波段下的光谱值依次进行比较,算出其对应的趋势参数  $V_{ij}$ ,最后将该列所有样本的趋势参数相加,就得到该波段下趋势参数值  $V_j = \sum_{i=1}^{51} V_{ij}$ ,较大  $V_j$  值对应的波段可以被选为特征波段。

### 1.5 二次判别分析 QDA 模型

利用 Unscrambler® X, v10.1 软件中的二次判别分析 QDA 模块进行分析建模。提取校正集中特征波段下的光谱,采用两个阈值  $1\ 000$  和  $2\ 000\ \mu\text{g}\cdot\text{kg}^{-1}$ ,将校正集数据分成低、中、高程度三类污染水平,因此 QDA 建模的类别数为 3。假设所有变量的先验概率均相等,权重均为 1.00。所构建的模型在验证集特征波段的光谱下进行模型检验。DON 的分类准确性通过计算被正确分类的样本数占样本总数的百分比来评估。

## 2 结果与讨论

### 2.1 样本中的 DON 含量分布

在 102 个小麦样品中, DON 水平为  $319 \sim 31\ 437\ \mu\text{g}\cdot\text{kg}^{-1}$  不等,通过可见光-近红外光谱法分析小麦样品,并且以 570~1 848 nm 之间的吸光度记录光谱。图 1 显示了 6 种不同 DON 浓度小麦样品的 Vis-NIR 原始光谱,其范围在  $319 \sim 5\ 895\ \mu\text{g}\cdot\text{kg}^{-1}$  DON 范围内。从这些谱图的比较来看,低

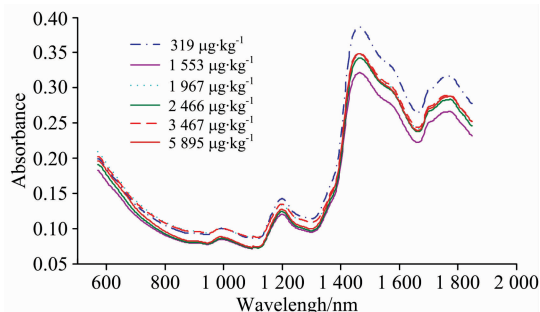


图 1 六种不同 DON 浓度(用高效液相色谱法测量)小麦样品的 Vis-NIR 光谱

Fig. 1 Vis-NIR spectra of wheat samples with six different DON concentrations (measured by high performance liquid chromatography)

DON 水平的小麦样品与含有高 DON 水平的小麦样品具有相同的 Vis-NIR 谱带, 因此表明两种样品中的主要官能团和化学成分是共存的。

### 2.2 选取的特征波段

按照 1.4 节的特征波长选取方法, 计算各个波段对应的趋势参数, 结果如图 2 所示, 可见在 666, 1 238 和 1 660 nm 波段处出现三个局部最大值, 将其提取作为评估 DON 浓度的特征波段。

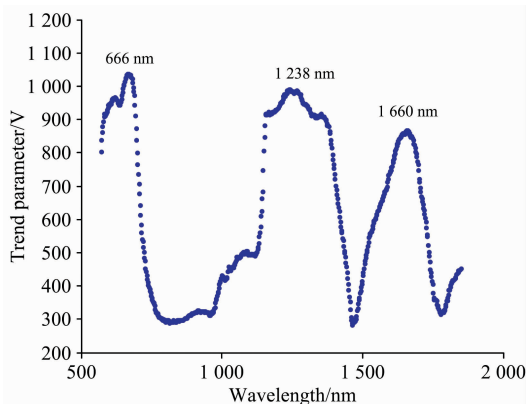


图 2 各个波段下趋势参数分布图

Fig. 2 Distribution of trend parameter of each band

### 2.3 二次判别分析 QDA 分类结果

只提取校正集样本的三个特征波段(666, 1 238 和 1 660 nm)下的光谱。采用两个阈值 1 000 和 2 000  $\mu\text{g} \cdot \text{kg}^{-1}$ , 事先将校正集数据分成低、中、高程度污染。利用 QDA 将校正集样本分成三类, 结果的混淆矩阵如表 1(左侧)所示, 校正过程中整体分类正确率为 88.24%。低和中度污染水平样本都 100% 被正确分类; 重度污染的 31 个样本中只有 6 个样本被

表 1 趋势参数(TP)-二次判别分析(QDA)模型的校正及验证结果

Table 1 The calibration and validation results of Trend Parameter (TP)-Quadratic Discriminant Analysis (QDA) model

参考类别 (HPLC 测量结果)	校正分类结果			验证分类结果		
	轻度	中度	重度	轻度	中度	重度
轻度	4	0	0	2	0	0
中度	0	16	6	1	14	4
重度	0	0	25	0	2	28

## References

[ 1 ] Lessard M, Savard C, Deschene K, et al. Food & Chemical Toxicology, 2015, 80: 7.  
 [ 2 ] Smith M C, Timmins-Schiffman E, Coton M, et al. Journal of Proteomics, 2018, 173: 89.  
 [ 3 ] GB2761—2011 Maximum Levels of Mycotoxins in Foods(食品中真菌毒素限量), 2011.  
 [ 4 ] Michael S, Franz B, Rudolf K. Rapid Communications in Mass Spectrometry, 2010, 20(18): 2649.  
 [ 5 ] Wei W, Liu A, Li M. Cereal & Feed Industry, 2013, 12(1): 60.  
 [ 6 ] XIE Gang, WANG Song-xue, ZHANG Yan(谢刚, 王松雪, 张艳). Journal of Instrumental Analysis(分析测试学报), 2011, 30

误判为中度污染, 剩余的 25 个(80.64%)的重度污染样本被正确分类。利用校正集构建的 TP-QDA 模型去判定验证集合中的样本, 验证集也只选取三个特征波段的光谱, 验证结果如表 1(右侧)所示。验证集总体判断正确率为 86.27%。

主成分分析 PCA 是选取特征波段最有效的方法之一, 谷物中 DON 含量检测的文章中很多都采用了主成分分析<sup>[11-15]</sup>, 从而对数据进行降维。以下将对相同的校正集样本, 采用 PCA 选取特征波段, 建模方式仍为 QDA 分析, PCA-QDA 模型检测的准确率仍在原来的验证集样本中测试, 从而对趋势参数法和主成分分析法选取特征波段的有效性做一一对比。

对校正集全光谱数据进行主成分分析, 选取前 2 个主成分, 解释了 91% 的变量。QDA 分析时采用 PCA 的选取的两个主成分, PCA-QDA 模型的校正和验证集分类结果, 以及其与 TP-QDA 模型检测结果的对比如表 2 所示。可见, 除了验证集中的轻度污染样本, TP-QDA 模型在各个分类中的分类正确率以及整体分类正确率都高于 PCA-QDA 模型, 因此, 趋势参数法选择的特征波段能获得更高的识别率。

表 2 TP-QDA 和 PCA-QDA 模型校正和验证识别准确率对比结果

Table 2 The classification accuracy comparison of TP-QDA and PCA-QDA model in calibration and verification sets

准确率	TP-QDA		PCA-QDA	
	校正	验证	校正	验证
轻度分类准确率/%	100	66.67	50	100
中度分类准确率/%	100	87.50	62.50	81.25
重度分类准确率/%	80.64	81.81	74.19	65.62
整体分类准确率/%	88.24	86.27	68.62	72.55

## 3 结 论

通过定义各个波段下光谱的趋势参数, 可以得到与小麦粉中 DON 浓度最相关的特征波段, 选出的特征波段有助于快速准确地判定小麦粉中 DON 含量的等级。分类模型表现出良好的预测性能, 总体分类率高, 小分类错误率低, 模型简单, 无需昂贵耗材, 能快速分析大量的样品, 所提出的分类模型在筛选小麦样品的脱氧雪腐镰刀菌烯醇含量方面具有实际应用价值, 使得 NIR 光谱学成为监测安全程序的强有力的工具。但是该方法还需要在更多品种小麦样本中进行普适性验证。

- (12): 1362.
- [ 7 ] LI Rui, XIE Gang, WANG Song-xue(黎 睿, 谢 刚, 王松雪). Food Science(食品科学), 2015, 36(6): 206.
- [ 8 ] WANG Fan, LI Yong-yu, PENG Yan-kun(王 凡, 李永玉, 彭彦昆). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(12): 3736.
- [ 9 ] ZHANG Xin-yu, WANG Ying-jie, LIU Ruo-xi(张新玉, 王颖杰, 刘若西). Journal of China Agricultural University(中国农业大学学报), 2017, 22(5): 25.
- [10] LUO Qu, YU Xiu-zhu, ZHANG Jian-xin(洛 曲, 于修烛, 张建新). China Oils and Fats(中国油脂), 2018, 43(3): 136.
- [11] De G A, Lippolis V, Nordkvist E. Food Additives & Contaminants Part A Chemistry Analysis Control Exposure & Risk Assessment, 2009, 26(6): 907.
- [12] Dvořák V, Prohasková A, Chrpová J. Plant Soil & Environment, 2012, 58(4): 196.
- [13] Annalisa D G, Salvatore C, Angelo V. Toxins, 2014, 6(11): 3129.
- [14] Peiris K S, Dong Y, Davis M A. Cereal Chemistry, 2017, 94(4): 677.
- [15] GB/T 5009.111—2016. Determination of Deoxynivalenol and Its Acetylated Derivatives in Food(食品中脱氧雪腐镰刀菌烯醇及其乙酰化衍生物的测定), 2016.

## The Study on Quickly Determining DON Level in Wheat Flour by Trend Parameter of Spectra

WU Wei<sup>1</sup>, ZU Guang-peng<sup>1</sup>, CHEN Gui-yun<sup>1</sup>, XU Jian-hong<sup>2</sup>, CHEN Kun-jie<sup>1\*</sup>

1. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

2. Institute of Food Quality and Safety, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

**Abstract** Deoxynivalenol (DON) is a mycotoxin that often occurs in cereals and their derivatives. It is harmful to the life and health of human and livestock. It is urgent to develop a detection method, which can rapidly, accurately and economically detect DON without environmental hazard. This study defined a Trend Parameter (TP) of the visible-near-infrared (VIS-NIR) spectra. The TP was used to determine the characteristic bands which were most relevant to the DON concentration. In this paper, the rows of spectral matrix of the samples in calibration set were arranged in the order of gradual increase in DON concentration. Each column (each band) of the matrix corresponded to a TP value. Under a certain band, the stronger the increasing trend of the absorbances of all samples in the column direction is (ie, the larger the TP value), the stronger the correlation between the absorbance and the DON concentration in this band is, and this band can be used as a characteristic band for evaluating the DON concentration. The study found that the local maximum of TP appeared at 666, 1 238, and 1 660 nm. The quadratic discriminant analysis (QDA) was performed by the spectra of the three characteristic bands. The wheat flour can be divided into three grades: mild ( $0 < \text{DON} < 1\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ ), moderate ( $1\ 000 \leq \text{DON} < 2\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ ), and severe ( $\text{DON} \geq 2\ 000\ \mu\text{g} \cdot \text{kg}^{-1}$ ) pollution by the constructed TP-QDA model. The overall classification accuracy of the model was respectively 88.24% and 86.27% in the calibration set and verification set. The Principal Component Analysis (PCA) of characteristic bands selection was used to make a comparison. The PCA-QDA model divided the same wheat sample into three pollution levels. The overall classification accuracy rate was 68.62% in the calibration set and 72.55% in the verification one. These findings confirmed that the selection of characteristic bands by TP parameter is superior to the one by the PCA in judging DON pollution level, and the TP-QDA model can be effectively used to quickly classify the pollution level of wheat flour, thereby reducing the time and economic cost of analyzing and screening wheat during the process of acquisition, storage and transport. The results of this study have yet to be tested for universality in a wider range of wheat varieties.

**Keywords** Deoxynivalenol; Trend parameters; Quadratic discriminant analysis; Wheat flour

(Received Apr. 4, 2019; accepted Aug. 2, 2019)

\* Corresponding author