J Am Oil Chem Soc (2021)
DOI 10.1002/aocs.12487

ORIGINAL ARTICLE

Measurements of High Oleic Purity in Peanut Lots Using Rapid, Single Kernel Near-Infrared Reflectance Spectroscopy

Brittany I. Davis[1,2] · Catherine B. Agraz[1,2] · Mark Kline[3] · Emma Gottschall[4] · Michael Nolt[4] · Thomas B. Whitaker[5] · Jason A. Osborne[6] · Erik Tengstrand[7] · Kamil Ostrowski[7] · Rita Teixeira[5] · Jack P. Davis[1,2,8]

Received: 9 November 2020 / Revised: 5 April 2021 / Accepted: 5 April 2021
© 2021 AOCS

Abstract High oleic peanuts have improved shelf life vs. conventional peanuts. Purity (percentage of high oleic peanuts within a lot) is critical to ingredient performance and final lot value. Contamination can result from unintentional mix-ups at the breeder/seed level, improper production handling, or due to physiologically immature high oleic kernels. Therefore, industry groups have established unofficial sampling plans to monitor purity. Assuming equivalent measurement performance and simple random sampling, increasing the sample size decreases variance among replicated sample test results and increases the precision of estimated lot purity. A novel instrument (QSorter Explorer by QualySense AG) using near-infrared reflectance spectroscopy was evaluated for high speed (20 kernels per second) high oleic purity measurements. The study objectives were to assess instrument performance in: (1) measuring oleic acid (%) in runner peanuts and (2) estimating the true high oleic purity of artificially mixed peanut lots. Three grades (Jumbo, Medium, and No 1) of US Runner mini-lots each at seven different contamination levels (0, 5, 10, 20, 30, 50, and 100%) were prepared. Oleic acid (%) of individual kernels was measured by scanning replicated samples of 10, 50, 100, and 500 kernels using the QSorter Explorer. The variance associated with each sample size and lot contamination level on returned purity values is discussed in the context of binomial sampling. Overall, the demonstrated measurement performance and capacity of the QSorter Explorer to process much larger sample sizes suggest this instrument can better identify true high oleic peanut lot purity vs. other currently available technologies.

Keywords NIRS · High oleic peanuts · Seed purity · Sorting · Quality control · QSorter explorer

J Am Oil Chem Soc (2021).

✉ Brittany I. Davis
brittanydavis@jlaglobal.com

1 Technical Services, JLA International, Albany, GA 31721, USA

2 Technical Services, IEH Laboratories, Lake Forest Park, WA 98155, USA

3 Procurement, Hershey, Hershey, 17033, USA

4 Research and Development, Hershey, Hershey, 17033, USA

5 Biological and Agricultural Engineering Department, Statistics Department Food, Bioprocessing and Nutrition Sciences Department, North Carolina State University, Raleigh, 27695, USA

6 Statistics Department, North Carolina State University, Raleigh, 27695, USA

7 Research and Development Division, QualySense AG, Zürich, CH-8152, Switzerland

8 Food, Bioprocessing and Nutrition Sciences Department, North Carolina State University, Raleigh, 27695, USA

Introduction

High oleic peanuts have increased levels of oleic acid and correspondingly decreased levels of linoleic acid compared to conventional oleic peanuts (Moore and Knauft, 1989). This unique fatty acid composition of high oleic peanuts confers improved post-roast resistance to oxidation compared to conventional peanuts, with exact improvements dependent on many factors, including the food matrix, specific fatty acid composition, and storage factors, among others (Braddock et al., 1995; Davis et al., 2016; Reed et al., 2002). Given these shelf life improvements, high oleic peanuts often command a premium price in peanut markets. High oleic peanuts are especially valued in

finished products featuring single-kernel/multi-kernel formats, including various confections, snack bars, and in-shell ballpark style peanuts (Kline, 2016). These products often contain individual kernels or clusters of kernels relatively exposed to the atmosphere, which can lead to rapid degradation in roasted peanut flavor and/or the onset of stale/oxidized flavors. While advanced packaging solutions or addition of synthetic antioxidants may extend shelf life of conventional peanuts in these products, these solutions bring increased costs, environmental concerns, and potential negative consumer perceptions. In contrast, high oleic peanuts provide an excellent solution to deliver great tasty and nutritious peanuts to consumers in these single kernel/cluster formats with an expanded shelf life.

High oleic cultivars of peanut are produced through conventional breeding, typically in tandem with molecular markers, and the genetics and subsequent biochemical pathways for this trait are well established (Chu et al., 2009; Jung 2000a, b; Tonnis et al., 2020). As both conventional and high oleic peanut cultivars are being developed, produced, handled, and shelled, there is opportunity for mixing in the supply chain. Additionally, the exact fatty acid composition of conventional and high oleic peanuts is also dependent on the specific genetics for a cultivar and subsequent seed physiological maturity at harvest, which is largely impacted by agronomic practices and growing environment (Davis et al., 2017; Dean et al., 2020). Given these factors impacting the supply quality and consistency for high oleic peanuts, procurement practices around sampling and testing of potential high oleic lots have evolved. Two measurements are important when specifying and procuring high oleic peanuts: (1) *average fatty acid composition* measured by gas chromatography (GC), including the ratio of oleic acid to linoleic acid and (2) high oleic *purity*, or the frequency of kernels in the lot meeting minimum level of oleic acid (%) to be considered high oleic (Kline, 2016). For average fatty acid composition measurements, standard practice is to comminute 100+ grams of sample, express oil from this ground sample, and measure the fatty acid composition of the expressed oil by GC. The common minimum standard for the ratio of oleic acid/linoleic acid (O/L ratio) for the milled sample to be considered high oleic is a 9/1 (Knauft et al., 2000); however, this ratio can commonly exceed 30/1, and shelf life benefits are continuously observed with increasing O/L ratios above 9.0 (Davis et al., 2016).

GC is the primary/reference method for determining fatty acid composition; however, the method is relatively slow (some hours to prepare samples and obtain results), expensive, and requires a skilled chemist. In the case of purity, where many measurements of individual kernels are required, GC is typically cost and/or time prohibitive. More rapid and less expensive secondary measurements to predict oleic acid (%) among individual kernels have evolved, and these include modified electrophoresis techniques, refractive index, and near infrared (NIR), among others (Chamberlin et al., 2014; Davis et al., 2013). Among these, NIR-based methods are the most rapid, and for peanut breeders, an added benefit with NIR is the method can be nondestructive, meaning the seed can be subsequently recovered for planting (Fox and Cruickshank, 2005; Tillman et al., 2006). Despite the relative speed of single kernel NIR based methods vs. traditional GC measurements, these methods are still limited to practical sample sizes of 10, 50, or 100 kernels when screening potential high oleic lots in the commercial trade. As such, a common purity specification might be 95%, which requires 95 kernels from a random sample of 100 kernels must test high oleic before certifying the bulk lot has acceptable purity (Davis et al., 2017).

The practice of sampling a bulk lot, measuring the sample, and subsequently estimating the composition of the bulk lot from the sample measurement is an important practice in agricultural markets. The statistical framework of various sampling plans for agricultural commodities has been well elucidated and described, for example, when estimating the mycotoxin contamination in bulk lots including peanuts, tree nuts, and corn (Whitaker, 2006). For a given sampling plan, assuming equivalent measurement performance, evaluation of larger sample sizes results in an improved statistical estimate of the bulk lot's true composition; however, these improved estimates come with increased costs that must be balanced against the objectives of the estimate (Whitaker, 2006). Single kernel NIR technologies, alone, or in tandem with single kernel imaging technologies, for measuring, screening, and purifying various commodities including soy, corn, wheat, rice, and peanuts, among others, are becoming increasingly advanced and relevant in modern agricultural systems (Agelet and Hurburgh, 2014; ElMasry et al., 2019). For the current study, a new instrument, the QSorter Explorer (QSE) was evaluated for its potential to measure the oleic acid (%) content of individual runner peanuts at 10–20 kernels s^{-1}, which can in turn be used to estimate high oleic purity of bulk lots. Evaluation of variances among repeated samples of different sizes is compared with the goal to improve the statistical estimate of the bulk lot's true purity.

Materials and Methods

QSorter Explorer

The QSE is a commercial high-speed single-kernel measurement and sorting instrument for analyzing various agricultural products (Armstrong et al., 2017; Rupenyan et al., 2016), including peanut. A high-speed vacuum belt

delivers individual kernels in a fixed position first to a camera for three-dimensional imaging, second for near-infrared reflectance spectroscopy (NIRS) measurement, and third these data are processed by a computer according to predetermined algorithms, and kernels are pneumatically sorted into one of three bins. For the current study, the emphasis is on NIRS measurements of oleic acid (%); however, the QSE can also measure and sort peanuts for moisture content, visually defective kernels, size, and foreign material. Peanuts were measured at a belt speed of 25 kernels per second, with actual pickup rates between 10 and 20 kernels per second. For the NIRS measurements, kernels were illuminated with a tungsten-halogen light source, and light was delivered to the kernels and to the spectrometer by means of optical fibers. The spectrometer was a NIRQuest512 (Ocean Optics, Dunedin, FL, USA) that uses a high-stability, 512-element Hamamatsu InGaAs-array detector with a two-stage thermoelectric cooler and low-noise electronics. The absorbances of the peanut spectra were calculated with the white and dark reference spectra acquired before the peanut spectra to match the behavior of the sorting operations. Spectra were acquired in the range of 900 nm–1700 nm, with 12 ms integration time. The QSE rejects spectra with compromised spectral quality. This occurs as the peanuts' orientations on the vacuum belt will randomly vary, and depending on the orientation, the peanuts can be out of focus for the spectrometer, causing the spectrometer to primarily acquire the background (blue belt) instead. This causes the spectral intensity to drop, and the predictions can be wrong. Therefore, all raw spectra with intensities below a predefined threshold are automatically rejected and the peanuts are sent to Bin 1 where they can be reprocessed if the operator chooses.

NIRS Model Calibration

A variety of conventional, mid, and high oleic peanuts were sourced. All peanuts were runners of US origin from the 2017–2019 crop years sourced from the commercial trade, except some limited mid-oleic Spanish peanuts that were a gift from a peanut breeder. Mid-oleic peanuts are kernels that have an oleic acid value between the range of conventional and high oleic peanut varieties. Oleic acid (%) values for the mid-oleics used in this study ranged from 56% to 72% oleic acid. In addition to different crop years, runner peanuts were also sourced to include different grades (sizes), growing regions, and/or different production regions within growing locations to include robust chemical diversity. Individual peanuts were analyzed in single kernel mode with repetitions to collect spectral data, and kernels were collected in labeled vials after data collection. Reference oleic acid (%) data were collected on kernels after NIRS spectra acquisition. Spectral data were correlated with respective GC reference values to develop a model for predicting oleic acid (%). A total of 699 kernels are included in the calibration model.

NIRS Validation

To validate NIRS predictions of oleic acid (%), a set of 151 kernels were sourced from different crop years (2018 and 2019), market types (runner and Spanish), and various grades from the United States (Table 1). These validation samples were independent of the original calibration samples. Individual peanuts were analyzed in single kernel mode with repetitions to collect spectral data. Three repetitions of spectra for each peanut were acquired to assess repeatability. White and dark reference spectra were acquired at regular intervals (after every 10 peanuts were acquired successfully with 3 repetitions), with 100 replicates of the white and dark reference spectra. The spectra were processed offline to predict the oleic acid (%) of the peanuts and then compared to GC reference data to describe instrument performance. A total of 453 spectra were collected and 40 spectra were rejected as they were below the acceptable spectral intensity threshold at delivery, leaving 413 spectra that were used to assess the performance of the oleic acid (%) model.

Reference Oleic Acid Measurements

After processing the individual kernels in the QSE, the reference oleic acid content of the individual kernels was measured by GC. The samples were prepared for GC by base-catalyzed transesterification, converting extracted peanut oil to fatty acid methyl esters (FAME). Individual

Table 1 One hundred and fifty-one kernels selected for the validation data set

Oil chemistry	Crop year	Market type	Grade	N
Conventional	2018	Runners	Medium	19
	2019		Jumbo	18
	2019		Medium	20
	2019		No 1	10
High	2018	Runners	Jumbo	9
	2018		Medium	9
	2018		No 1	10
	2019		Jumbo	17
	2019		Medium	20
Mid	2018	Spanish	NA	19

Kernels were selected to cover a range of oil chemistry, crop year, market type, and grade. After the QSorter Explorer predictions were collected, all individual kernels were analyzed by gas chromatography—the primary method for determining oleic acid (%).

kernels were placed in glass vials and macerated in the presence of 2 mL of hexanes. The resulting suspension was shaken with 1 mL of methanolic potassium hydroxide and allowed to separate. A portion of the clear hexane layer was placed into a GC vial and diluted with hexanes for GC analysis of the FAME. A Hewlett Packard HP 6890 series GC system with split injection was used to analyze fatty methyl esters. The GC was equipped with a microbore capillary column (Restek Famewax; 30 m × 0.25 mm × 0.25 μm) and helium as the carrier gas (1.5 mL min^{-1}). The oven program included the following temperatures: 180 °C (0.5 min hold), 10° min^{-1} ramp to 190 °C (1 min hold), 10° min^{-1} ramp to 240 °C (8 min hold), until elution of all FAME was complete.

Preparation and Sampling of Mini-Lots

Twenty-one mini-lots of 2019 SE runners were created by gently (to minimize split formation) and thoroughly mixing by mass (4200 g total) conventional and high oleic peanuts to achieve seven contamination levels identified as 0, 5, 10, 20, 30, 50, and 100% contamination of high oleic peanuts with conventional peanuts for each of three separate grades: Jumbo, Medium, and No 1. For each of the 21 mini-lots (three grades and seven contamination levels), eight samples of 500 g were selected from each of the 21 mini-lots and the oleic acid in each peanut was measured by the QSE sorter. For the first replication, only the 0 and 100% contamination mini-lots were specifically processed, to confirm that these mini-lots were in fact primarily composed of high oleic and conventional oleic peanuts, respectively. After confirming this, the remaining mini-lots were selected at random to complete the first replication. For the remaining seven sample replications, the mini-lots (seven contamination levels and three grades) were sampled in random order. After completing a replication, the QSE was cleaned with compressed air and the mirrors for the digital imaging system were wiped with optic paper to clean any dust since peanuts tend to release some degree of debris when processed. Oleic acid (%) measurements for individual kernels for the various samples were automatically transcribed by internal software to *.pre files, which were further analyzed as described next.

Data Analysis

The QSE outputs comma-separated values files that were aggregated and categorized in MS Excel. JMP (Cary, NC, USA) and SAS (Cary, NC, USA) were used for data tabulation, visualization, and statistical analyses. For some small fraction of kernels, two kernels (double) can be randomly picked up on the vacuum belt. In this scenario, the QSE vision algorithm recognizes these as a "double" and the kernels are categorized accordingly by the internal software. In addition to "doubles," spectral outliers are also categorized by the QSE software. For the current analyses, doubles and spectral outliers were excluded. Kernel oleic acid measurements were categorized according to grade (Jumbo, Medium, or No 1), contamination level (0, 5, 10, 20, 30, 50, and 100%), and replication (A–H). After all sample measurements, each mini-lot was subdivided into various sample sizes, namely 10, 50, 100, or 500 kernel counts to study the effect of sample size on variance among samples of a given size and on measurements of oleic acid purity. As every mini-lot, regardless of grade or contamination level, had at least 5000 kernel measurements, 500, 100, 50, and 10 replications of 10, 50, 100, or 500 kernel counts, respectively, were randomly sampled without replacement using the SURVEYSELECT procedure of SAS (SAS, 2019).

Results and Discussion

GC oleic acid (%) reference measurements for the 151 validation kernels are provided in Fig. 1. The reference line at 74% oleic acid is the industry-accepted cutoff, above which a peanut is considered high oleic (Knauft et al., 2000). Conventional peanuts (blue bars in Fig. 1) have oleic acid (%) values from ∼46% to 69%, whereas high oleic kernels (green bars in Fig. 1) range from ∼75% to 83% of oleic acid. These values agree well with previous reports on single kernel GC values (Tillman et al., 2006). There is a clear separation from ∼69% to 75% oleic acid between the two populations: conventional and high oleic peanuts (Fig. 1, top panel). Note that five kernels received as high oleic for this validation exercise were identified by the QSE and later determined *via* GC to actually be conventional oleic, i.e., they were contaminants, and are excluded in Fig. 1. In the bottom panel, mid oleic Spanish kernels (red bars in Fig. 1) were also included. Mid-oleic peanuts can result in early crosses of breeder seed to generate high oleic peanuts, but mid-oleic peanuts are not commonly expected in the commercial trade; nonetheless, these were included to understand the performance of the current oleic NIRS model for the QSE. The addition of the mid-oleic peanuts, which would be considered conventional oleic if encountered in the commercial trade, narrows the natural gap in oleic acid values between conventional and high oleic peanut populations (Fig. 1 bottom panel).

NIRS QSE predictions of oleic acid (%) when plotted against reference data show a positive, linear ($R^2 = 0.78$) association (Fig. 2). Performance statistics for these NIRS validation measurements were calculated (Table 2). A root mean squared error of prediction (RMSEP) of 6.33%

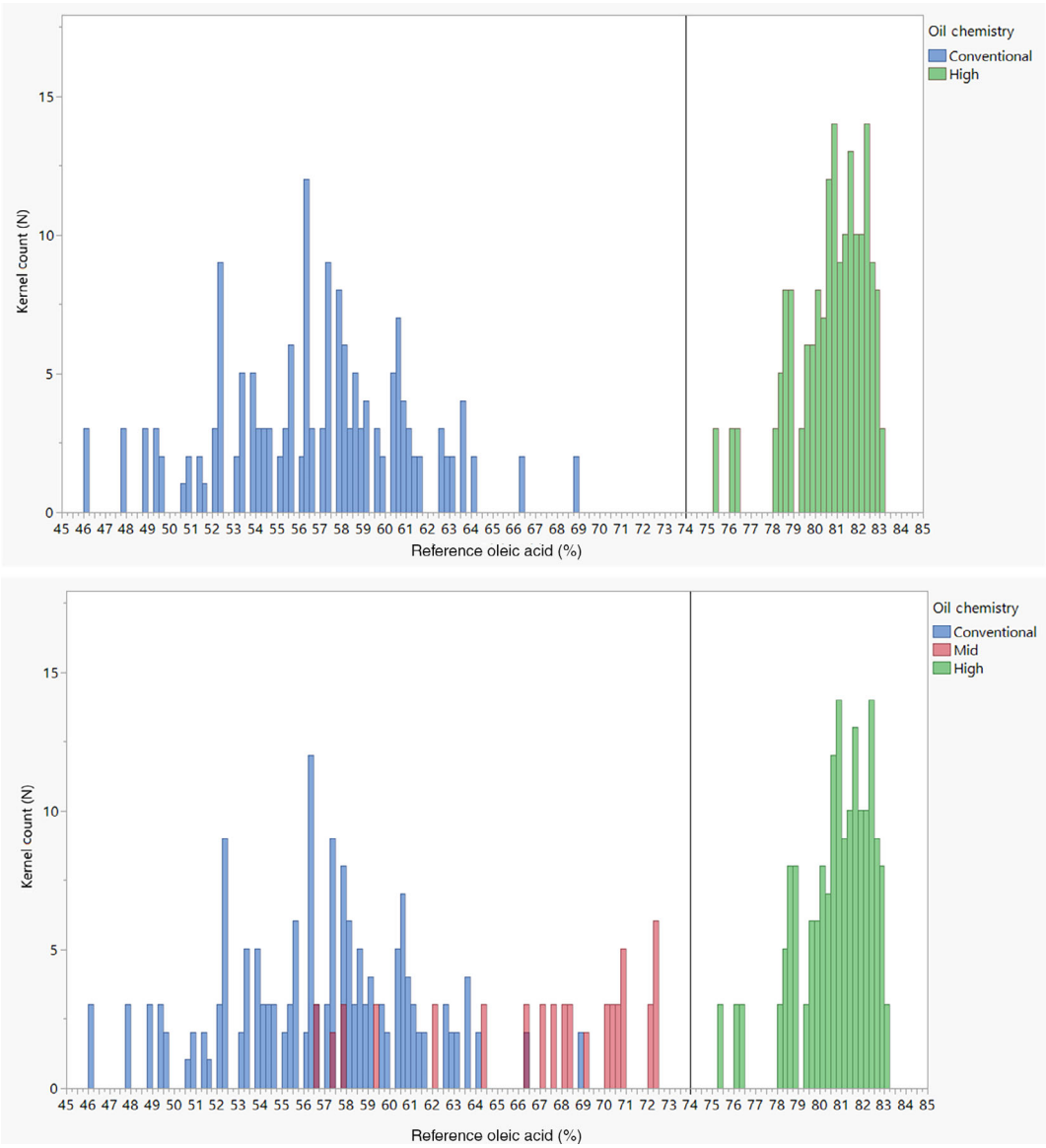

Fig. 1 Reference oleic acid (%) data for high oleic, mid oleic, and conventional kernels used for validation with the five high oleic contaminants removed. Line at 74% oleic acid is the industry-accepted standard, above which a kernel is classified as high oleic

suggests NIRS predictions should be within 12.66% (2 SD) of the reference data 95% of the time. The population of predicted NIRS oleic acid (%) values for conventional oleic (bottom panel) and high oleic (middle panel) kernels is compared with GC reference values (top panel) in Fig. 3. Excluding kernels between 65% and 74% oleic acid, which represented 2% of kernels in the validation set and are largely the uncommon mid oleic peanuts, error rates for various oleic acid (%) thresholds, above which a kernel is classified as high oleic, are provided (Fig. 4). The binary classification error for both conventional and high oleic peanuts is minimized at about 3% (Fig. 4) at a threshold of 66.5% oleic acid. This reference value is provided in Fig. 3, where misclassification in predictions for high oleic kernels is to the left of this line (middle panel) and conventional oleic misclassifications are to the right of this line (bottom panel). By increasing this threshold, it is possible to increase specificity, or the confidence that peanuts above this threshold are truly highly oleic; however, the sensitivity, or true positive rate, will decrease, that is, more high oleic peanuts will be falsely classified as conventional. For example, by moving the threshold up from 66.5% to 74.0%, which is the industry-accepted standard for GC reference data, essentially no conventional peanuts would be falsely classified as high oleic; however, about 12% of true high oleic kernels would be falsely classified as

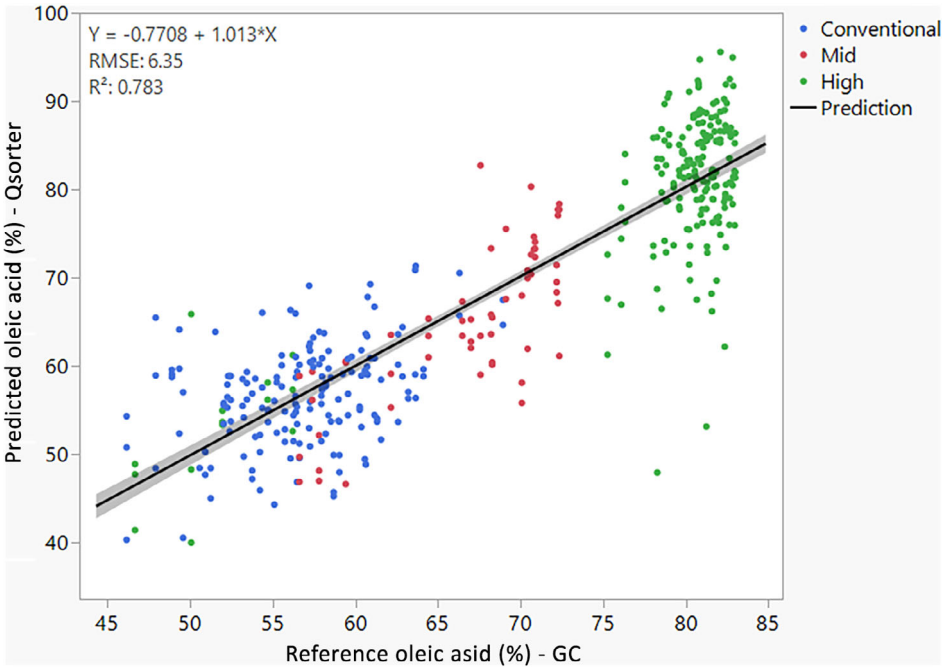

Fig. 2 QSorter predictions of oleic acid vs. reference values of oleic acid (%) measured by gas chromatography

Table 2 Performance statistics for near-infrared reflectance spectroscopy oleic acid (%)

Statistic	Result (%)
Standard error of prediction	6.34; 95% confidence interval: [5.70, 7.14]
Root mean squared error of prediction	6.33; 95% confidence interval: [5.77, 7.08]
Bias	0.11; 95% confidence interval: [−0.91, 1.13]
Repeatability error	4.75; 95% confidence interval: [4.38, 5.20]
Coefficient of determination (R^2)	0.78; 95% confidence interval: [0.71, 0.84]

conventional (Fig. 4). When physically sorting peanuts, users can adjust this threshold for greater specificity or sensitivity depending on the end goal with the sorted peanuts. For example, a breeder may want very high specificity after sorting, i.e., a very pure set of high oleic peanuts, but this will come at some cost of sensitivity/yield, i.e., some of the true high oleic peanuts (belonging to the lower parts of the high-oleic content range) will be incorrectly classified as conventional oleic. For the current study, a threshold of 66.5% oleic acid on the QSE was used to classify peanuts as conventional or high oleic.

The repeatability error was estimated to be 4.75%. Compared to the RMSEP, this indicates that most of the error comes either from variation in the oleic acid in different parts of the peanut or from variation in the spectra. The strongest contributor to this noise is the background signal. The peanuts will have a different orientation when scanned by the spectrometer, which means that different amounts of the background will be in view. This variation in the background contributes to the repeatability error. Since the repeatability error is a significant part of the RMSEP, passing peanuts multiple times will improve the predictions.

NIRS predicted oleic acid (%) values for all grades at the different contamination levels are provided in Fig. 5. Jumbo, Medium, and No 1 data are purple, red, and blue, respectively. These colors were selected to best differentiate the grades; however, the exact values of these histograms are less important than considering the overall trends that can be observed for each grade. Additionally, the threshold value of 66.5% oleic acid, above which these oleic acid predictions were classified as high oleic, is provided for reference (Fig. 5). The overall responses of each grade with changing contamination were similar, resulting in a blend of colors (Fig. 5). When considering the 0% contamination mini-lots, which correspond to high oleic peanuts as received, normal distributions are observed for all grades (Fig. 5) averaging 78–82% oleic acid depending on the exact grade (Table 3). For the 100% contamination mini-lots, which correspond to 100% conventional oleic peanuts as received, normal distributions are again observed for all grades (Fig. 5), averaging 57–59% of oleic acid depending on the grade (Table 3). As the

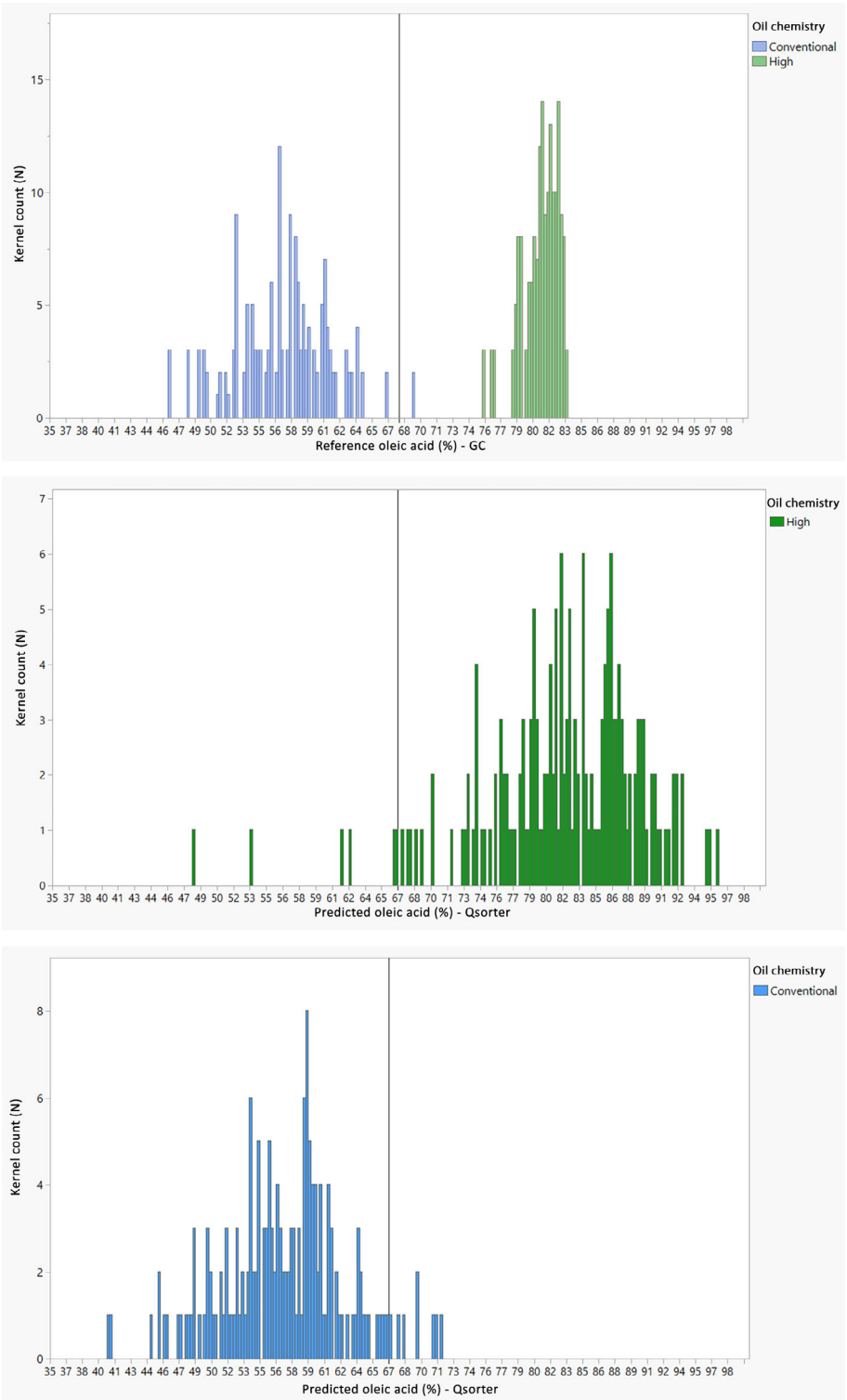

Fig. 3 Gas chromatography reference values (top) and near-infrared reflectance spectroscopy predictions for high oleic (middle) and conventional (bottom) peanuts

contamination level of conventional peanuts increased from 0% to 50%, bimodal distributions emerge for all grades (Fig. 5).

The overall oleic acid data observed in Fig. 5 reflect the calculated average oleic acid (%) values for the different contamination rates and grades presented in Table 3.

Regarding kernel counts in Table 3, it is noted that equivalent masses of the three grades were analyzed, meaning kernel counts are proportionally elevated for the smaller No 1 peanuts, vs. the Mediums, followed by lowest counts for the largest grade analyzed, Jumbo (Table 3). It is also noted for the 0% and 100% contamination mini-lots, which correspond to expected 100% pure high oleic and 100% pure conventional oleic as received, the counts for these samples were slightly different within a grade (Table 3). For example, the 0% contamination Jumbos had 5284 kernels, whereas the 100% contamination Jumbos had 5098 kernels (Table 3), demonstrating the high oleic kernels were slightly smaller for these Jumbo mini-lots. In the case of Mediums and No 1s, the conventional kernels were slightly smaller given the counts at equivalent masses (Table 3). As such, when comparing contamination levels, which were prepared by mass, these differences in count-based purities should be recognized.

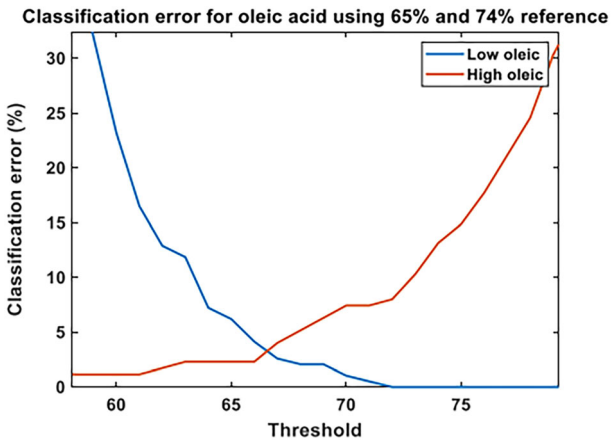

Fig. 4 Error rate for classifying peanuts as conventional or high oleic for the QSorter near-infrared reflectance spectroscopy model when excluding reference values between 65% and 74%

Average purity, or the average frequency of peanuts exceeding the defined threshold of oleic acid (%) to be classified as high oleic, is provided in Table 3 for the three grades and the seven different contamination levels. As the contamination increased within a mini-lot, average purity decreased for all grades. Furthermore, at equivalent contamination levels, Jumbo peanuts tended to have higher purity values compared with Mediums and No 1s. All other factors being equivalent, larger peanuts tend to be more physiologically mature, and as peanuts mature, oleic acid (%) tends to increase (Davis et al., 2017; Dean et al., 2020; Wang et al., 2018). The differences in average oleic acid content (%) and average purity among grades could be, at least partially, attributed to differing maturities. Additionally, the current NIRS model is seemingly less accurate in classifying purity for smaller grades, for example, No 1s, as these kernels will tend to have oleic acid (%) values closer to the threshold due to their inherent immaturity. In the current validation set, when only considering No 1 kernels, approximately 9% of the prediction values for validated high oleic kernels were incorrectly classified, whereas only 3% and 2% of validated high oleic Mediums and Jumbos, respectively, were misclassified. Given this, the purity of

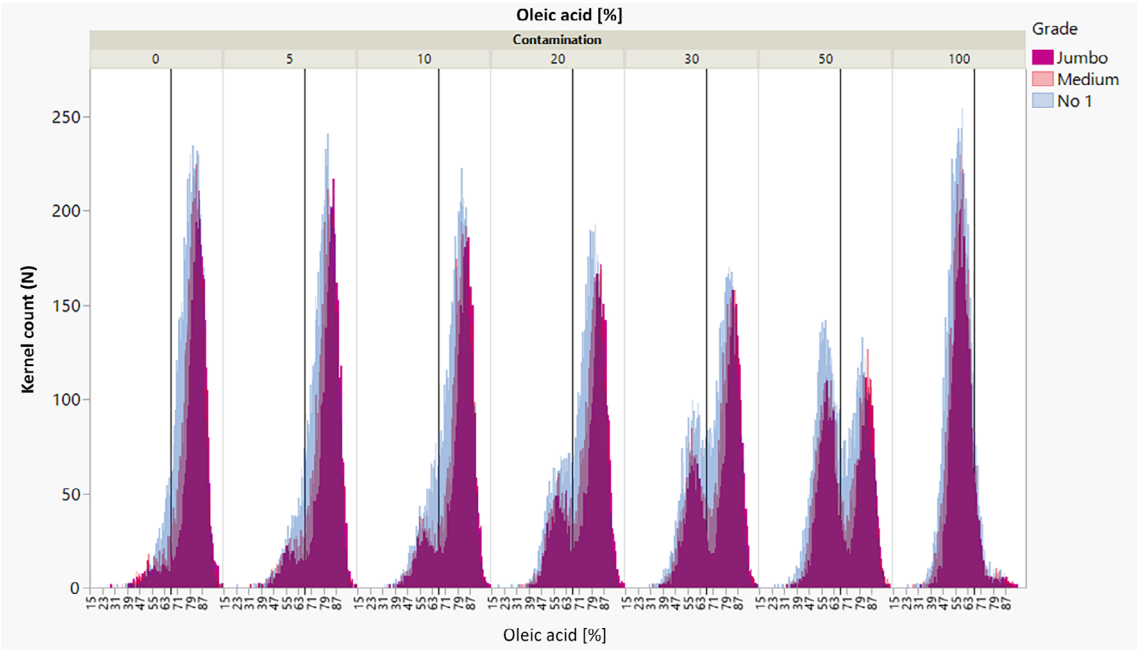

Fig. 5 Oleic acid (%) histograms of seven mini-lots for each of three grades

Table 3 Number of kernels analyzed by the QSorter Explorer, average oleic acid, and purity by grade and contamination

Grade	Contamination	N	Average oleic acid (%)	Purity
Jumbo	0	5284	81.7	94.2
Jumbo	5	5291	81.0	91.2
Jumbo	10	5295	79.6	86.6
Jumbo	20	5290	77.2	78.0
Jumbo	30	5262	75.5	70.8
Jumbo	50	5261	70.7	53.7
Jumbo	100	5098	59.0	10.4
Medium	0	6485	79.6	92.4
Medium	5	6399	78.6	89.0
Medium	10	6362	77.9	85.1
Medium	20	6434	75.2	75.9
Medium	30	6547	73.4	68.5
Medium	50	6557	69.2	51.7
Medium	100	6613	58.0	9.6
No 1	0	8498	78.0	90.4
No 1	5	8446	77.0	86.4
No 1	10	8495	75.9	81.5
No 1	20	8518	73.5	73.6
No 1	30	8559	71.8	66.0
No 1	50	8650	67.0	48.2
No 1	100	8636	57.3	10.1

the 0% contamination of No 1 peanuts summarized in Table 3 could be interpreted as 99.4% pure, or 90.4% returned purity +9% expected measurement error. While a relatively minor portion of the commercial trade, adding more No 1 kernels to the model should improve its performance for this grade, and this will be done in the future.

When considering contamination in high oleic lots, there can be immature high oleic kernels that do not meet the minimum threshold of oleic acid, but from a genetic perspective they are in fact high oleic. However, from an ingredient perspective, these kernels will not provide equivalent shelf life performance vs. true high oleics (Davis et al., 2016). Additionally, there can be contaminants that are clearly conventional oleic given their oleic acid (%) measurements. While beyond the scope of the current paper, in the future, the industry might consider three classifications: (1) true contaminants, (2) physiological immature contaminants, and (3) true high oleics. The QSE could be used to make these classifications. As larger data sets are collected with the QSE technology (or equivalent), the expected optimal purity one might expect given the indeterminate nature of the peanut crop could be better understood and ideal purity targets refined accordingly.

Repeated purity measurements for 10, 50, 100, and 500 kernels samples for each of the mini-lots are provided in Fig. 6. As expected and already discussed, as the contamination rate increased for a given mini-lot, the overall purity decreased. Furthermore, regardless of the mini-lot and its contamination level, as the sample size was increased, the variation (as measured by the variance) in repeated measurements of sample purity decreased. For example, when considering the mini-lot with 20% contamination, which was determined to have an overall purity of 75.9%, returned purity sample values for 10 kernel samples ranged from 30% to 100% purity. In comparison, 100 kernel samples for the 20% contamination mini-lot ranged from about 85% to 66% purity, while the 500 kernel samples ranged from 79% to 73%. As sample size increases, the variance among the replicated average sample purity decreases. As a result, larger sample sizes improved the estimate of the true high oleic purity for a given lot being evaluated. While 10–100 kernel purity checks are currently common, with the QSE, it is possible to sample 500 plus kernels within less than a minute to better understand true lot purity. Although not the focus of the current manuscript, the QSE also predicts moisture for the single kernels in addition to measuring size and visual defects, which can be useful in determining other aspects of peanut quality.

The statistical variances for repeated sample purity measurements at different sample sizes for the different mini-lots and grades were calculated. As purity is a binary classification (oleic acid is more or less than the 66.5% threshold), these variance data were also compared to theoretical binomial variances calculated across a range of purity

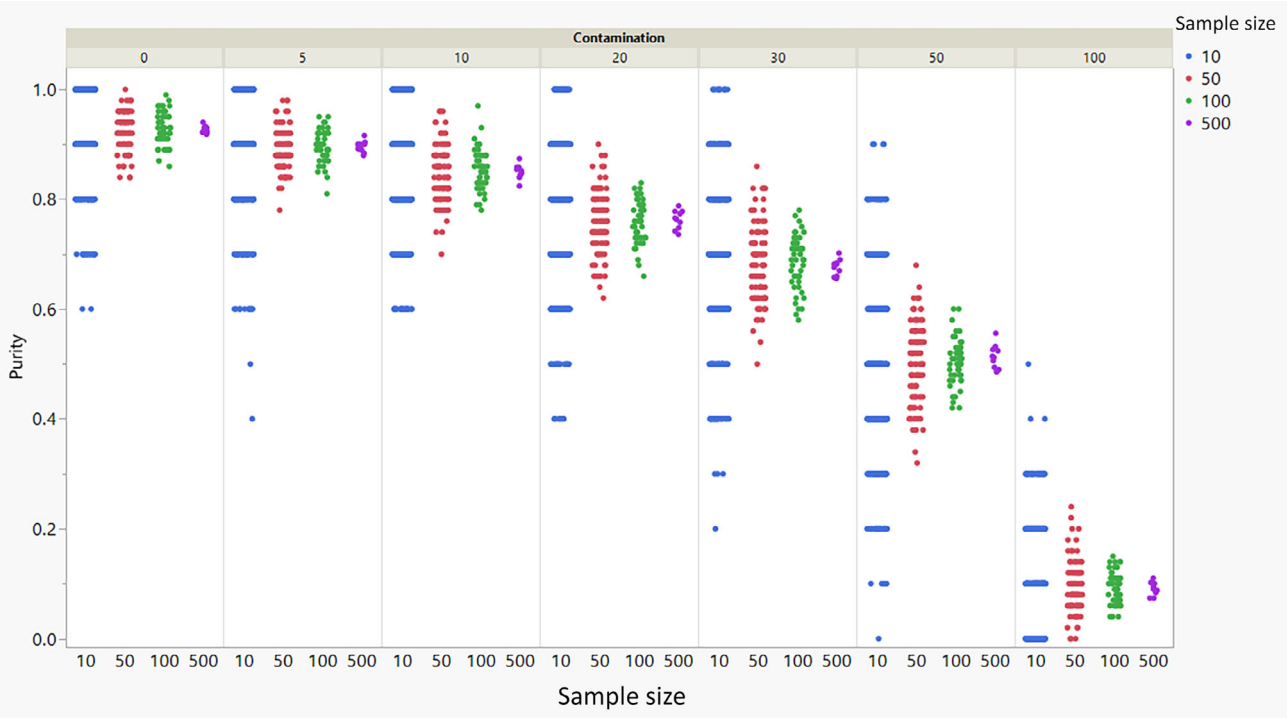

Fig. 6 Purity (frequency of kernels meeting high oleic threshold) for Medium peanuts for different lot contaminations (0–100%). Sample size was 10, 50, 100, or 500 kernels with 500, 100, 50, or 10 replications, respectively

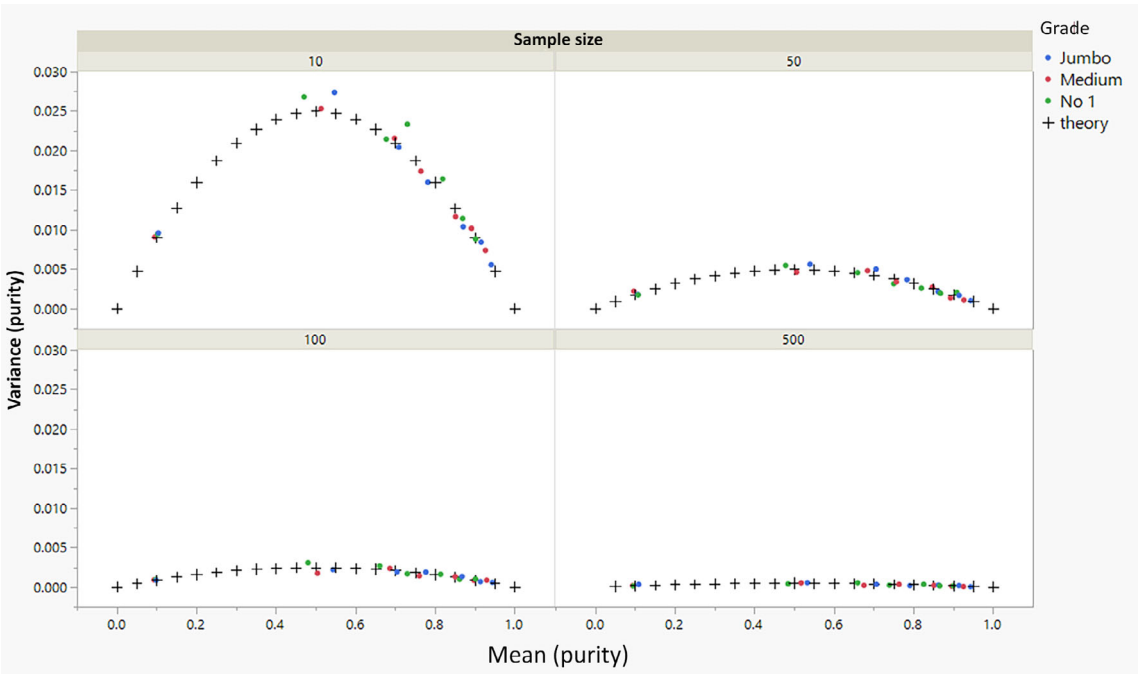

Fig. 7 Variance of purity measurements vs. average purity for seven contamination levels, three grades, and four sample sizes of peanuts. Gray dashed lines are calculated theoretical binomial variances for a range of average lot purity values

values (Fig. 7). The theoretical binomial variances show a parabolic rise, peaking at 50% purity. Maximum variance observed for experimental observations was at 50% contamination regardless of the grade, contamination level, and sample size evaluated. Overall, experimental variances decrease as the sample size increases, and the magnitudes observed agree quite well with theoretical values predicted by Binomial theory (Fig. 7). As a result, the theoretical sampling variance of the number of high-oleic kernels, X, when drawing samples of size n from populations with purity $0 < p < 1$ is given by equation (1) when X is modeled as a binomial random variable:

$$\mathrm{Var}(X/n) = p^*(1-p)/n \quad (1)$$

The observed agreement between empirical and theoretical variances suggests that the binomial distribution can be used to construct/calculate/develop operating characteristic curves. These curves can be used to quantify/assess the effect of sample size on misclassification risk associated with sampling plants to estimate high oleic purity in bulk lots. The binomial distribution has been used to successfully model sampling plans in other agricultural commodities, for example, genetically modified seeds in conventional seed lots (Whitaker et al., 2001).

Summary

The performance of the QSE to predict oleic acid (%) for runner peanuts at 20 kernels s^{-1} was measured and summarized including predicted error rates of misclassification at different oleic acid (%) thresholds. The QSE was then used to measure purity (frequency of true high oleic kernels in a lot) for repeated samples of a given size for various mini-lots of different grades of runner peanuts prepared to have a range of contamination (non-high oleic peanuts). The impact of sample size on the precision among repeated sample purity estimates for various mini-lot contamination levels was demonstrated and variances for repeated sample purity measurements were calculated. Purity data were well described by the binomial distribution, which can be used to evaluate the performance (risk of misclassifying lots) of more advanced sampling plans for determining the purity of potential high oleic peanut lots, including the development of operating characteristic curves. Given the unprecedented single kernel NIRS speeds of the QSE, the instrument will allow buyers/sellers of high oleic peanuts to use much larger sample sizes than currently accessible, to provide a more precise estimate of the true purity of peanut lots being traded in the high oleic market.

Conflict of interest The authors declare that they have no conflicts of interest.

References

Agelet, L. E., & Hurburgh, C. R. (2014) Limitations and current applications of near infrared spectroscopy for single seed analysis. *Talanta*, **121**:288–299.

Armstrong, P. R., Dell'Endice, F., Maghirang, E. B., & Rupenyan, A. (2017) Discriminating oat and groat kernels from other grains using near-infrared spectroscopy. *Cereal Chemistry*, **94**:458–463.

Braddock, J. C., Sims, C. A., & O'Keefe, S. F. (1995) Flavor and oxidative stability of roasted high oleic-acid peanuts. *Journal of Food Science*, **60**:489–493.

Chamberlin, K. D., Barkley, N. A., & Tillman, B. L. (2014) A comparison of methods used to determine the oleic/linoleic acid ratio in cultivated peanut (*Arachis hypogaea* L.). *Agricultural Sciences*, **5**: 227–237.

Chu, Y., Holbrook, C. C., & Ozias-Akins, P. (2009) Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. *Crop Science*, **49**:2029–2036.

Davis, J. P., Leek, J. M., Sweigart, D. S., Dang, P., Butts, C. L., Sorensen, R. B., … Lamb, M. C. (2017) Measurements of oleic acid among individual kernels harvested from test plots of purified runner and spanish high oleic seed. *Peanut Science*, **44**:134–142.

Davis, J. P., Price, K. M., Dean, L. L., Sweigart, D. S., Cottonaro, J. M., & Sanders, T. H. (2016) Peanut oil stability and physical properties across a range of industrially relevant oleic acid/linoleic acid ratios. *Peanut Science*, **43**:1–11.

Davis, J. P., Sweigart, D. S., Price, K. M., Dean, L. L., & Sanders, T. H. (2013) Refractive index and density measurements of peanut oil for determining oleic and linoleic acid contents. *Journal of the American Oil Chemists Society*, **90**:199–206.

Dean, L. L., Eickholt, C. M., LaFountain, L. J., & Hendrix, K. W. (2020) Effects of maturity on the development of oleic acid and linoleic acid in the four peanut market types. *Journal of Food Research*, **9**:1–9.

ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., & Rousseau, D. (2019) Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview. *Sensors*, **19**: 1–32.

Fox, G., & Cruickshank, A. (2005) Near infrared reflectance as a rapid and inexpensive surrogate measure for fatty acid composition and oil content of peanuts (*Arachis hypogaea* L.). *Journal of Near Infrared Spectroscopy*, **13**:287–291.

Jung, S., Powell, G., Moore, K., & Abbott, A. (2000a) The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. II. molecular basis and genetics of the trait. *Molecular and General Genetics*, **263**:806–811.

Jung, S., Swift, D., Sengoku, E., Patel, M., Teule, F., Powell, G., … Abbott, A. (2000b) The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. I. isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. *Molecular and General Genetics*, **263**:796–805.

Kline, M. (2016) Chapter 16 - Manufacturing foods with peanut ingredients. In H. T. Stalker & R. F. Wilson (Eds.), *Peanuts* (pp. 429–445). London: AOCS Press.

Knauft, D. A., Gorbet, D. W., & Norden, A. J., inventors; University of Florida Research Foundation, Inc., assignee. May 16, 2000. Enhanced peanut products and plant lines. US Patent 6063984.

Moore, K. M., & Knauft, D. A. (1989) The inheritance of high oleic-acid in peanut. *Journal of Heredity*, **80**:252–253.

Reed, K. A., Sims, C. A., Gorbet, D. W., & O'Keefe, S. F. (2002) Storage water activity affects flavor fade in high and normal oleic peanuts. *Food Research International*, **35**:769–774.

Rupenyan, A., Sansonne, N., & Dell'Endice, F. (2016) Machine vision combined with near-infrared spectroscopy to guarantee food safety. *Cereal Foods World*, **61**:140–142.

SAS 9.4 User's Guide. (2019) Cary, NC: SAS Institute, Inc.

Tillman, B. L., Gorbet, D. W., & Person, G. (2006) Predicting oleic and linoleic acid content of single peanut seeds using near-infrared reflectance spectroscopy. *Crop Science*, **46**:2121–2126.

Tonnis, B., Wang, M. L., Li, X., Wang, J., Puppala, N., Tallury, S., & Yu, J. (2020) Peanut FAD2 genotype and growing location interactions significantly affect the level of oleic acid in seeds. *Journal of the American Oil Chemists Society*, **97**:1001–1010.

Wang, M. L., Chen, C. Y., Tonnis, B., Pinnow, D., Davis, J., An, Y. C., & Dang, P. (2018) Changes of seed weight, fatty acid composition, and oil and protein contents from different peanut FAD2 genotypes at different seed developmental and maturation stages. *Journal of Agricultural and Food Chemistry*, **66**:3658–3665.

Whitaker, T. B. (2006) Sampling foods for mycotoxins. *Food Additives and Contaminants*, **23**:50–61.

Whitaker, T. B., Freese, L., Giesbrecht, F. G., & Slate, A. B. (2001) Sampling grain shipments to detect genetically modified seed. *Journal of AOAC International*, **84**:1941–1946.