Food Fingerprinting: Using a two-tiered approach to monitor and mitigate food fraud: The rice case study

Terence F McGrath1, Maeve Shannon1, Olivier P Chevallier1, Ratnasekhar Ch1, Fei Xu2, Fanzhou Kong2, Hong Peng2, Ernest Teye3, Selorm Akaba3, DiWu4, Libiao Wu4, Qiang Cai4, Le Nguyen Doan Duy5, Le Van Viet Man5, Arun P Kapil6, Guangtao Zhang2, Mary McBride7, Christopher T Elliott1

1Queens Univ Belfast, Sch Biol Sci, Inst Global Food Secur, Belfast BT9 5AG, Antrim, Northern Ireland

2Mars Global Food Safety Center, Beijing 101407, China

3 University of Cape Coast, College of Agriculture and Natural Sciences, School of Agriculture, Cape Coast, Ghana

4 Yangtze Delta Region Institute of Tsinghua University, Zhejiang, China

5 Ho Chi Minh City University of Technology - VNU-HCM (HCMUT)

6 Green Saffron Spices Ltd, Ireland

7 Agilent Technologies, Santa Clara CA, USA

# ***Abstract:***

Rice is an important staple food that is consumed around the world. Like many foods, the price of rice varies considerably, from very inexpensive for low-quality product to premium pricing for highly prized varieties from specific locations. Therefore, like other foods it is vulnerable to economically motivated adulteration through substitution or misrepresentation of inferior quality rice for more expensive varieties. In this article we describe results of a research project focused to address potential food fraud issues related to rice supplies in China, India, Vietnam and Ghana. Rice fraud manifests differently in each country; therefore, tailored solutions were required. Here we describe a two-tiered testing regime of rapid screening using portable Near Infrared technology supported by second tier testing using mass spectrometry-based analysis of suspicious samples. Portable Near Infrared spectroscopy models and laboratory-based Gas Chromatograph-Mass Spectrometry methods were developed to differentiate between: high value basmati rice varieties and their potential adulterants; six Geographic Indicated protected rice varieties from specific regions within China; various qualities of rice in Ghana and Vietnam; as well locally produced and imported rice in Ghana. Furthermore, an Inductively Coupled Plasma-Mass Spectrometry method was developed to support the Chinese rice varieties methods as well as a Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry method for quality differentiation in Vietnam. This two-tier approach can provide a substantially increased level of testing through rapid screening outside of the laboratory with the reassurance of corroborating mass spectrometry-based laboratory analysis to support decision making.

# Introduction:

Food fraud is defined as the intentional adulteration or misrepresentation of foods and food ingredients for economic gain and is estimated to impact the food industry by $40 billion each year (1, 2). Economically motivated adulteration (EMA) does not just have financial consequences—it can also be a food safety issue with significant public health consequences. Some well-known examples of EMA include the adulteration of dairy products with melamine in China, which sickened nearly 300,000 babies, and killed 6 infants, (3) and the misbranding of industrial–grade rapeseed oil, sold as olive oil in Spain, which caused 20,000 illnesses and at least 300 deaths (4).

The economics of the global rice trade make rice an attractive target for criminals. More than 3.5 billion people depend on rice for as much as 20% of their daily calories (5). Asia accounts for 90% of global rice consumption, but rice is also the fastest-growing food staple in Africa and in Latin America (5). Fraud in the rice supply has been known for some time and typically involves the substitution of lower-quality rice for premium product. For example, substitution of Basmati or Thai Hom Mali rice, which are both prized for their aroma and flavor is common (6, 7),as is the substitution of Wuchung rice (a Chinese geographical indication (GI) protected rice). In a scandal that emerged in 2010, it was reported that ten times more Wuchang rice was sold than was produced (8); West African countries including Ghana, Nigeria and their neighbors have experienced numerous cases of poor-quality rice in their markets. More specifically, expired or poor-quality rice is normally bagged and re-branded with counterfeit labels that consumers associate with high-quality rice.

Here, we describe results from a research project, led by Professor Chris Elliott from the Institute of Global Food Security at Queen’s University Belfast, UK, focused to understand and address food fraud issues specific to rice. An international team from countries that represent most of the major producers, exporters and consumers of rice was assembled, including partners from China, India, Vietnam, Ghana, Ireland, the United Kingdom, and the United States. Fraud in rice is not ubiquitous; different countries face very different challenges requiring tailored solutions. Therefore, we developed a use-case approach for rice fraud in China, Ghana, India and Vietnam to explore the feasibility of using a single testing approach as a solution to address distinct and complex real-world circumstances.

Analytical technologies, including spectroscopic techniques, molecular or genome analysis, and mass-spectrometric methods have been developed and applied for the determination of authenticity in foods (9, 10). For this project, we employed a two-tiered system of testing technologies, for detecting rice fraud in global supply chains. The first tier consists of a screening method, based on a small hand-held (and field-portable) molecular spectroscopic technique that provides end-users with a rapid, low-cost, and easy-to-use “food fingerprinting” technique. This technology can be used by stakeholders across the entire food supply chain to quickly screen samples for fingerprint anomalies. The second tier employs a much more rigorous, laboratory-based confirmatory analysis technique that uses techniques such as Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry (LC-QTOF-MS), Gas Chromatography-Mass Spectrometry (GC-MS) or Inductively Coupled Plasma Mass Spectrometry (ICP-MS) for detection, quantitation, and confirmation of anomolous screening results. The most appropriate technique was selected based on the queston at hand. While the use-cases presented here are focused on rice, it is envisaged that the two-tier approach could be used with a wide range of commodities vulnerable to fraud.

# Experimental Section:

## Rice samples:

The Institute of Global Food Security at Queen’s University in Belfast, UK, has a library of over 1,000 authentic polished white rice samples collected over several years. This library represents different rice varieties and countries of origin. Project partners collected traceable polished white rice samples in-country, over the two year project thus providing not only variability in source of rice but also in harvest, processing batches and storage conditions. All rice samples collected in-country were sub-sampled and shared with the Institute for Global Food Security.

## Instrumentation

### Handheld NIR

Each of the partners in the project received a SCiO instrument for in-country scanning. The SCiO is a small handheld near infrared (NIR) scanner, manufactured by Consumer Physics, measuring 18.8 x 40.2 x 67.7 mm and weighing 35 grams. It works through a phone-based App to record NIR spectra between 740 and 1070 nm. scans take less than 5 sec (11). All spectra were stored in Consumer Physics cloud database. Users access their data through a browser-based App called ‘The Lab’. This App allows users to pre-process data, develop and validate their own models before publishing them for others to use. Project partners shared a single researcher license login that allowed them to store all project data in one account. This researcher license allowed spectral data to be exported from the cloud for manipulation in other software. All participants underwent a one-week training course in instrument use and chemometric model development and validation at the Institute for Global Food Security.

### LC-QTOF-MS

Analysis were carried out using an Agilent Infinity 1290 liquid chromatography system coupled to an Agilent 6545 QToF with a Jet Stream ionization source (Agilent Technologies, Santa Clara, USA) operating in positive mode with the following parameters: mass range, 50−1200 m/z; capillary voltage, 3.5 kV; nozzle voltage, 300 kV; gas temperature, 240 °C; drying gas (nitrogen), 11 L. min-1; nebulizer gas (nitrogen), 35 psi; sheath gas temperature, 320 °C; sheath gas flow (nitrogen), 11 L. min-1; fragmenter, 110 V; All Ion mode; acquisition rate 0.1 scan.s-1.

Chromatographic separation was performed on an Agilent Zorbax Eclipse Plus C18 RRHD column (50x2.1 mm, 1.8 µm) maintained at 45 oC with a flow rate of 0.5 mL.min-1 with the following gradient using water with 0.1 % formic acid and methanol with 0.1 % formic acid as mobile phase A and B, respectively: starting at 99 % A to 95 % A over 1 min, then linear gradient from 95 % A to 5 % A over 6 min, then from 5 % A to 1 % A over 2 min, then returned to initial condition with 99 % A over 0.2 min.

### ICP-MS

Elemental profiling was performed using an ICP-MS (7900) equipped with a concentric micromist nebulizer and a quartz double wall spray chamber, cooled to 2 °C (Agilent Technologies, Santa Clara, USA). Thirty elements (B, Na, Mg, Al, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Rb, Sr, Nb, Mo, Ag, Cd, Cs, Ba, Hg, Pb) were measured. The instrumental setting and operative conditions were adopted from a previous study, with some modifications (12). Briefly, under the helium tune mode, the plasma parameters were 1550 W RF power, 8 mm sampling depth, and 1.16 L·min-1 carrier gas flow (Argon). Cell gas (Helium) flow of 5.0 mL·min-1 was also applied. Multi-element calibration standard 2A (part# 8500-6940) and 4 (part# 8500-6942), environmental calibration standard (part# 5183-4688) and standard solution of Sc (part# 5190-8578) were purchased from Agilent Technologies (Santa Clara, CA, USA). The calibration solution was prepared by mixing and diluting the above-mentioned standards. The certified reference material (CRM) of rice flour (1568b) was purchased from the National Institute of Standards and Technology (Gaithersburg, MD, USA). CRM was measured once every 10 samples to ensure the accuracy of the analysis. And diluted internal standard (IS) solution (1 mg·L-1) of Rh (Agilent Technology, Santa Clara, part# 8500-6945) was mixed with the sample stream before entering the nebulizer in a mixing tee (sample tubing 1.02 mm inner diameter, IS tubing 0.25 mm inner diameter), with the peristaltic pump operating at 0.1 rps. All samples were analyzed in duplicate.

### GC-MS

Samples were analyzed using an Agilent 7890 gas chromatograph coupled to an Agilent 5793C mass spectrometer (Agilent Technologies, Santa Clara, USA). Helium was used as a carrier gas at a flow rate of 1 ml min-1. Samples exposed to Solid-Phase Microextraction (SPME) PDMS/DVB/CARBOXANE fiber was injected into the HP-5 MS capillary column consisting of a stationary phase of 5 % phenyl 95 % methyl polysiloxane in the splitless mode. The injector, ion source and transfer line temperatures were set at 270 °C, 220 °C and 280 °C respectively. The initial oven temperature was held at 50 °C for 1 min, increased to 170 °C for 2 minutes at a rate of 10°C min-1 and finally increased to 280 °C for 1 min at a rate of 30 °C min-1. Full scan mass spectra were acquired at 70 eV from m/z 45-750 at a rate of 1 scan s-1 with an initial solvent delay of 6 min. Overall samples run time was 20 minutes.

## Protocols

### Handheld NIR sample preparation and analysis

Polished white rice was analyzed as whole grains without any further sample preparation. Sufficient sample was dispensed into a glass petri dish to cover the base and to a depth of approximately 10 mm when evenly distributed. Samples were scanned statically from underneath, through the glass, in triplicate with repositioning of the SCiO device between scans to ensure averaged scan data would be representative of the larger sample.

### LC-QTOF-MS sample preparation and analysis

Polished white rice was milled using Hario Skerton Plus Ceramic coffee grinder then sieved using Glenammer sieves 106 mic (800 microns). One hundred milligrams of milled rice was weighed using an analytical balance. The analytes were extracted with 2 mL of water (18.2 MΩ/cm)/LC–MS Chromasolv [methanol](https://www.sciencedirect.com/topics/chemistry/methanol) (4:1, v/v) by vortexing for 10 min at 2500 rpm, followed by sonication for 15 min at maximum frequency. After centrifugation at 10,000 g for 10 min at 4 °C, 0.5 mL of the supernatant was transferred to a fresh tube and solvent evaporated for 8 hours using an Eppendorf Concentrator Plus at room temperature with aqueous setting. Next, the extract was reconstituted in 0.3 mL of ultra-pure water, filtered through a 0.22 μm Costar Spin-X Centrifuge Tube Filter (5,000 g at 4 °C for 10 min) and transferred into vials for the LC-QTOF-MS analysis.

### ICP-MS sample preparation and analysis

0.5 g of rice grains was directly digested using 6 mL of nitic acid in a Teflon digestion vessel. The vessel was placed in a fume hood overnight for pre-digestion and then transferred to the microwave digestion instrument (Anton Paar, Austria). The digestion temperature of 180 °C was gradually reached within 15 min and held for 20 min. After digestion, the solution was cooled to room temperature and diluted to 50 mL with deionized water. In order to avoid cross contamination, precautions were taken by soaking all materials including the digestion vessels in nitric acid (30%, *v*/*v*) for 24 h and then rinsed with deionized water for three times.

### GC-MS sample preparation and analysis

3 grams of rice grains were weighed into a 10 ml head space vial. Samples were incubated with PDMS/DVB/Carboxane fiber for 5 min at 70 °C. Fiber desorption was performed for 2 minutes at 270 oC.

### Model building

#### Handheld NIR

Classification models where built and validated for spectral data using several different applications. Initial models were built using Consumer Physics ‘The Lab’ browser-based interface which allows users to select and apply a set of pre-processing functions and develop models using a Random Forest algorithm. A separate sample set was then used for validation. Project partners also processed the data in various alternative software such as Matlab (MathWorks) and Simca 15 (Sartorius Stedim) which allows for the selection of different pre-processing and classification algorithms. Through this procedure the models that yielded the best internal cross-validation statistics and, where available, the most correctly identified Validation set samples were identified for addressing the specific needs of each partner.

#### LC-QTOF-MS

Raw data files were processed using MassHunter Profinder 10.0 software then exported to Mass Profiler Professional 15 or Simca 15.0 for further data processing/filtering and/or multivariate analysis.

#### ICP-MS

The data was first pre-processed by taking logarithmic transformation and then split into reference and validation sets at a ratio of 80:20. Partial Least Square-Discriminant Analysis (PLS-DA) was implemented in RStudio with additional package “mixOmics” and “ropls”. The PLS-DA model was built with the reference set and optimized using 5-fold cross-validation 10 times. The optimized model was then independently validated using the validation set and the overall accuracy rate was determined.

#### GC-MS

Spectral files were pre-processed using MassHunter software. More than 160 compounds including aldehydes, aromatic compounds, alkanes, alkenes, N-heterocyclic compounds and alcohols where used to differentiate sample groups in this untargeted method**.** Multivariate statistical analysis was performed using Metaboanalyst. Normalised data was used for multivariate analysis to remove the offsets and adjust the importance of high and low abundance metabolites to an equal level. Principal Component Analysis (PCA) was performed to identify clustering patterns. Further, supervised PLS-DA was applied. The developed PLS-DA model was validated using the cross-validation method and its quality assessed on Accuracy, R2 and Q2 scores.

# Results and Discussion

## Use Case 1: India

Basmati rice, known as the ‘queen of fragrance’, is famous for its fragrance and delicate flavor. The aroma is due to the presence of a chemical called 2-acetyl-1-pyrroline, which is 12 times more abundant in Basmati than any other rice (13). Basmati is a healthy ‘Supergrain’. It is gluten-free and low in fat. It contains all eight essential amino acids, folic acid, and is very low in sodium and has no cholesterol in its composition. It has a low to medium glycemic index as compared to other non-Basmati rice, meaning that energy is released at a slower, steadier rate leading to a more balanced level of energy (14).

Basmati is grown only in the foothills of the Himalayan mountain range. This is its GI. The major areas of cultivation of basmati rice are in the states of Jammu & Kashmir, Himachal Pradesh, Punjab, Haryana, Delhi, Uttrakhand and western Uttar Pradesh. There are approximately 36 traditional strains of Basmati, with only about 5 to 6 being traded in the open market at any one time. India is the leading exporter of the Basmati rice to the global market. Major export destinations for 2018-19 were Iran, Saudi Arabia, Iraq, UAE, and Yemen (15). More than 44 million metric tonnes of basmati rice, with an approximate market value of $4.7 billion USD were exported in 2018 (15).

Basmati rice, prized for its unique flavor and fragrance, is vulnerable to fraudulent variety claims and substitution fraud. The main goal of the Indian rice study was to develop a two-tiered method to identify and protect the more expensive Pusa 1121 and Taraori rice varieties; ensuring that they have not been substituted. In total, Green Saffron Ltd provided 1399 samples representing 7 basmati varieties: Pusa 1121 (n= 250); Pusa 1509 (n=149); Sugandha (n=250); Taraori (n=150); Shabnam (n=250); Duplicate basmati (n=200); and Sharbati (n=150)). Pusa 1509 and Sugandha are considered potential adulterants of Pusa 1121, and Shabnam, Duplicate Basmati and Sharbati are considered potential adulterants of Taraori.

### Handheld NIR

Samples were scanned using the SCiO instrument. Sample data was divided into a Reference set (n= 934), to build the models, and a Validation set (n=465) to test the models. Chemometric models were developed using SIMCA 15 to differentiate the varieties.

Multiple pre-processing techniques were applied to spectral data before multiclass models containing Pusa 1121 and its adulterants were built alongside models for Taraori and its adulterants. In total, more than 18 multiclass models were generated and tested using the Validation set. Spectral data that had undergone Standard Normal Variate (SNV), first derivative and Savitzky-Golay (SG) smoothing pre-processing followed by model building using Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) were found to produce the best multi-class models in terms of predicted results for the Validation. The best multiclass models for Pusa 1121 and Taraori, based on the number of correctly identified samples in the Validation set, are presented in Figure 1a and Figure 1b respectively. It is important to note that these figures are a two-dimensional representation of a three-dimensional model. There appears to be overlap between Pusa 1509 and Sugandha in Figure 1a. However, this overlap is lessened when the z-axis is considered. Furthermore, any potential overlap between the two adulterants has no impact on correctly identifying Pusa 1121 as that cluster is well separated from the others. When running the Validation set, none of the adulterants were misclassified as Pusa 1121 and only 2 (both Sharbati) out 200 non-Taraori samples were misclassified as Taraori. Performance improved when binary models were used eg Pusa 1121 vrs adulterants and Taraori vrs adulterants where there was no misclassification of the adulterants when the binary models were tested with the Validation set.

This demonstrates that both multiclass and binary models can be used, with reasonable accuracy, to rapidly screen for adulteration in the field, allowing for suspect samples to be returned to the lab for tier-two testing. In this instance, there are strengths and limitations when using either binary or multiclass models. The binary models clearly identify authentic Pusa 1121 and Taraori rice from their adulterants, but in an adulterated sample, they will not reveal the nature of the adulterant. Multiclass models will identify the adulterant. However, clusters may be closer together, or even overlap, as the OPLS-DA algorithm works to separate multiple clusters from each other. Ultimately the choice of model used will be determined by the end-user. The end-user must determine if identifying samples as authentic (binary model) is sufficient or if identifying the adulterant(s) (multiclass model) will also be required.

### GC-MS Indian rice samples

Volatile Organic Compounds (VOCs) are often treated as an important component in the assessment of rice quality due to the presence of a large variety of aromatic organic carbon-based molecules such as aldehydes, ketones and hydrocarbons. The aroma profile can identify the qualitative differences between different rice samples. Analysis of VOCs is usually performed by either head space or purge and trap with separation and detection by GC-MS. Here, VOC profiles were established, based on the analysis of 10 to 15 samples of each of the seven different varieties of Indian Basmati rice samples, using Head space SPME combined with GC-MS.

PCA analysis was used to differentiate Pusa 1121 from Pusa 1509 and Sugandha rice samples. The first two principal components (PCs) show clear separation between the clusters, Figure 2a. Indeed, they explained 84 % of total variation. PC1 shows the variation of Pusa 1509 from Pusa 1121 or Sugandha. PC2 shows the variation between Sugandha and Pusa 1121 or Pusa 1509. In practical terms this means we can differentiate all three rice varieties from each other. Internal cross validation with an independent data set showed good Q2 (0.98678), R2 (0.99899) and accuracy (1.0). Pusa 1509 and Sugandha samples contains a higher VOC profile than Pusa 1121 rice samples. This forms the basis for differentiation.

PCA of Taraori, Shabnam, Sharbati and Duplicate samples identified clustering of sample groups. The first two PCs show the variation between all clusters, Figure 2b. Indeed, the first 2 PCs explained 64.9 % of the total variation. PC1 shows the variation between Duplicate, Shabnam and Sharbati/Taraori. PC2 shows the variation between Taraori, Sharbati and Shabnam/Duplicate rice clusters. Therefore, based on a combination of the first 2 PCs, we can differentiate all four rice varieties from each other. Internal cross validation with an independent data set shown good Q2 (0.99695), R2 (0.99895) and accuracy (1.0). Duplicate and Shabnam rice samples contain a lower VOC profile content than Sharbati and Taraori rice samples.

### Summary and next steps

A rapid field deployable NIR method has been developed that can distinguish two important basmati rice varieties from their potential adulterants. A proof-of-concept study using GC-MS suggests that this technology is an appropriate laboratory-based tier two supporting confirmation test. The NIR and GC-MS methods are complementary to each other, providing similar identification capabilities but using completely different underlying technologies. They can be used together in a two-tier system of rapid in-field detection with suspicious samples being forwarded to the lab for GC-MS analysis. Additional rice samples from the new season will be shipped, scanned and used to test and improve the database. In order to increase the robustness of the model, other Indian basmati, and non-basmati rice varieties should be added.

## Use Case 2: China

Worldwide, China is the biggest producer and consumer of rice, with approximately 212 million metric tonnes produced and 193 million metric tonnes consumed in 2018 (16). China has a vast history of rice cultivation, with earliest archeology relics among Yangtze river (Diaotonghuan 8,000- 12,000 B.P.) and Pearl river valley (8,200- 13,500 B.P.). There are more than 300 known breeding varieties developed, covering both Indica and Japonica species, among which 132 leading varieties are registered by the Ministry of Agriculture of China (17). Geographically, rice production activities can be found throughout the eastern and southern provinces of the country, especially along the Yangtze and Pearl rivers (18). Varieties grown in different regions have diverse growing seasons, climate and environmental conditions, and there can be between one to three harvests each year. Consequently, huge differences are observed in genetic, physical and chemical components between varieties. These differences result in rich and varied tastes and flavors, which in turn affect the price, which can be a 20-50 fold difference.

The GI system likely originated in 19th Century Europe is a designation assigned to certain products from a specific geographical origin and with qualities or reputations which are associated with historical and culture factors, which the name usually representing the location and commercial brand itself. With rising standards of living, there is a strong demand for GI products in China. There are more than 60 Chinese rice cultivars with GI designations, but none are more prized and known to Chinese people than Wuchang rice. Market demand for Wuchang rice greatly exceeds the amount of Wuchang rice that is actually produced. This situation creates exactly the conditions that drive food fraud, and fraudulent behavior such as partial substitution of high quality Wuchang rice with low quality rice which has been reported in recent years. For example where millions of tonnes of rice produced elsewhere was carried across hundreds of kilometers to the GI area, mixed and packed, or directly substituted with falsified labels (8). Each year, it is estimated that 10-fold more Wuchang rice is available on markets than produced. It is very challenging to accurately identify rice as Wuchang rice by visual inspection. Therefore, there is a great need to develop an easy to use, cost-effective and reliable means to quickly identify and confirm rice GI. We have used NIR spectroscopy to develop GI rice “fingerprints” to prevent adulteration.

In this study, we investigated differentiation of GI varieties, over a large geographic area by establishing authenticity models using several technology platforms. Over 100 polished white rice samples were collected from credible rice processing factories, from five provinces in China. These samples represent 6 different GI varieties, which are among the top 20 most popular rice varieties in China, and have a pricing point at least five-fold of the average rice price: Wuchang (WC) (n=20); Jingshanqiao (JSQ) (n=20); Panjin-Yanfeng (PJ-YF) (n=20); Panjin-Liaoxing (PJ-LX) (n=20); Sheyang (SY) (n=20); and Dongjinxi (DJX) (n=16). These varieties were selected because multi-variable factors can contribute to the study including physical geographic span; induced environmental factors differences (Heilong River, Liao River, Yangtze River and Pearl River); potential genetic differences between species (Dongjinxi and Jingshanqiao are Indica species while the other varieties are Japonica species); and the impact of multiple varieties grown in a smaller scaled GI region (the two varieties in Panjin).

### Handheld NIR

The Chinese rice samples were scanned in triplicate on the SCiO according to the protocol in the experimental section. Pre-processing techniques were applied individually and in combination. In total more than 50 iterations were applied to the data and chemometric models developed, in both ‘The lab’ and Simca 15, to distinguish between the varieties (data not shown).

Using ‘The Lab’, the best Random Forest model, based on internal cross-validation. was obtained after averaging the sample replicates then applying Wavelength Z-score to the full spectral range produced by the SCiO. Table 1 shows the confusion table generated by ‘The Lab’ under these conditions. A confusion table, in this instance, is a tabular representation of predicted result (Y-axis) against known classification (X-axis). It allows us to tell how well a model performs based on internal cross-validation. The ideal situation would show a diagonal green line of predicted results, each box in that green line showing 100 % and all other boxes in the confusion table showing 0 %. This would indicate that everything has been correctly identified. Using Simca 15, the best model, based on internal cross-validation, was generated using an OPLS-DA algorithm on data that had undergone SNV preprocessing (Figure 3). These conditions are different than those used for generating the basmati rice models. However, this is acceptable because we are optimizing, and choosing, the most appropriate conditions to address questions of samples from each country individually. What works best for basmati rice does not necessarily work best to address the questions asked of the Chinese rice samples.

Table 1 shows that during internal cross-validation, all GI varieties can be correctly classified (a diagonal line of green boxes with 100 % correct classified). Figure 3 shows clustering within GI varieties. The Wuchang rice appears to have separated into two sub-clusters. This could be due to different varieties being grown within the GI region or more likely is the result of the limited number of samples used. The robustness of these models can only be tested when additional samples have been collected. However, if either model proves to be robust through a comprehensive validation, then the SCiO could be used for rapid in-field assessment of GI status. Having a robust, reliable, rapid field-based screening technology would not only protect sellers from unfair competition through fraudulent vendors but also protect those consumers who choose to purchase these GI varieties.

### ICP-MS

ICP-MS is a powerful analytical tool that quantitatively determines the level of both metal and non-metal elements. It has the advantages of wide dynamic range, high throughput and relatively easy sample preparation (19). Since the elemental composition in plants is largely determined by factors such as soil characteristics and agricultural practices (20, 21), elemental profiles of plants obtained by ICP-MS have been used to identify their geological origins (22, 23).

Rice samples were analyzed with the Agilent 7900 ICP-MS. The data from the reference set was subjected to PLS-DA in order to build a model to differentiate the six GI rice varieties. As shown in Figure 4, the first two components showed limited differentiation between the classes, accounting for nearly 50% of the total variance. Specifically, PJ-YF, JSQ and DJX could be clearly separated from the others; while PJ-LX, SY and WC were clustered together. Internal cross-validation showed that the model had a decent fitness: R2X= 0.859, R2Y= 0.935, Q2Y= 0.907. Further external validation was conducted using the validation set to determine the accuracy of the model. The overall prediction accuracy for all six GI rice was 88.46%, and the prediction accuracy for each GI rice is shown in Table 2. From the confusion table, in this instance representing predictions for the validation set, there were excellent results for 5 of the 6 GI varieties. However, WC has been misclassified as PJ-LX 75 % of the time. This could be anticipated looking at the clustering of classes in Figure 4 where WC, PJ-LX and SY clusters overlapped. One possible approach to overcome this issue could be to create a separate model to just discriminate between these varieties then run sample scan through two models. The first model would identify PJ-YF, JSQ and DJX and a cluster of the remaining GI varieties whilst the second model would be used specifically to differentiate between PJ-LX, SY, and WC when needed. By zooming into these initially inseparable GI varieties, the second model could potentially improve the differentiation by further exploiting intricate differences between WC, PJ-LX, and SY. If this clustering persists, then additional attempts of variable selection could be made, which had been reported to improve classification predictiveness by removing redundant information and thus highlighting ones with high discrimination power. However, the primary issue to be addressed with all Chinese models is the limited number of samples used in their creation. Therefore, in-depth analysis should be delayed until models have been generated with appropriate sample numbers.

### GC-MS China Rice samples

We analyzed the 6 different Chines rice varieties using the method described in the experimental section. PCA and PLS-DA analysis was performed on the data matrix obtained from GC-MS spectral files. The first two principal components (Figure 5) show variation between rice clusters. Indeed, they explain 42.4 % of the total variation. PC2 shows clear classification of the Panjin groups (Panjin-Yanfeng and Panjin-Liaoxing) from other sample groups (Wuchang, Jinshanqiao, Sheyang and Dongjinxi), Figure 5. PC1 shows the variation between Sheyang and Jinshanqiao. This indicates that we can trace rice varieties based on geographic origin but not necessarily GI variety as indicated by the inability to discriminate between Panjin-Liaoxingand Panjin-Yanfeng. Although not clearly visible in Figure 5, Wuchang rice is differentiated from the Panjin group as well as Sheyang and Dongjinxi GI varieties but not from Jinshanqiao. Internal cross validation with an independent data set show good Q2 (0.85273), R2 (0.95876) and accuracy (0.95).

### Summary and next steps

These Chinese models should be considered proof-of-concept due to the limited number of rice samples used in this study. New samples are being collected to introduce seasonal variability of the GI rice, Non-GI rice varieties are also being collected to broaden the scope. It is clear from the above methods that each of the models have their own uses and limitations, The NIR method remains a useful screening tool but will not give a suitable level of confidence for decision making, meanwhile these limited GC-MS and ICP-MS methods cannot provide a clear solution individually but taken together may provide appropriate answers. The GC-MS models could be useful in determining geographic origin. The ICP-MS method currently works well for some of the varieties. PLS-DA models using algorithms such as Support Vector Machine Learning will be evaluated to separate PJ-LX, SY and WC clusters. The use of Artificial Intelligence to predict results will also be investigated.

## Use Case 3: Ghana

West Africa countries are net rice importers and in 2018, imported more than 7.6 million metric tonnes at a cost of $4 billion USD (24). In Ghana, rice has become a major staple food and consumption is higher than domestic production. Therefore, Ghana imports approximately 60% of the total amount of rice consumed each year (24). This has a huge impact on the local economy. The amount of money Ghana spends each year on rice imports has been rising steadily, increasing from $151M in 2007 to more than $1.2B in 2015, with the majority of imports coming from Thailand, Vietnam, and India. (24-26).

There is considerable variation in the quality of imported rice and there are frequent rumors that poor quality rice is flooding the local markets (27). This claim is worrying because it is extremely cumbersome to identify the integrity status of the rice consignments.

The challenges of authenticating huge imports of rice, coupled with ensuring the safety of poor quality rice has compelled the government of Ghana to introduce a flagship program, known as “planting for food and jobs.” The program has, a special emphasis on rice production, and is forecast to increase rice production from 450,000 metric tonnes to 750,000 metric tonnes by 2020 (28).

In certain circumstances, both local-grown and imported rice are subject to fraud. Many consumers in Ghana show a preference for imported rice over local rice because they believe imported is higher quality. This has created a potential avenue for mislabeling of Ghana rice as imported rice. Conversely, there are consumers who actively seek out Ghana rice to insure they are buying local. Studies have shown that consumers who prioritize local rice do so out of the perception that even though it is inexpensive, it has better nutritional quality (29). Furthermore, the increased production and consumption of local rice builds a more sustainable economy for Ghana and region.

These divergent perceptions regarding quality have facilitated food fraudsters to cheat local consumers who are at the mercy of cumbersome quality control via slow detection methods.

Actual quality attributes including appearance (whiteness and glittering) of uncooked rice and taste/aroma for cooked rice is crucial to consumers in Ghana (29). Consumers also consider cooking quality (notable expansion and speed of cooking), absence of stones, percentage broken and absence of mold (29). These criteria often categorize rice into three distinct groups referred to as high quality, mid quality and low quality. However, beyond this, there are numerous brands in the market with poor quality attributes (mostly not easily identified by the consumer).

### Handheld NIR

Rice samples (n=565) were collected from local millers, and recognized retailers from seven regions in Ghana. Imported rice (from Thailand and Vietnam) was purchased from recognized supermarkets in Ghana. Armed with the quality attributes (appearance, taste, aroma, etc), sensory analysis (1-6 point hedonic scale) was performed and the rice samples were further grouped into three distinct groups: high quality (n=170), mid quality (n=268) and low quality (n=127). Interestingly these grouping correlated to purchase price. Samples were scanned using the SCiO. Data sets were subjected to several preprocessing techniques individually and combined before PCAs were generated in Matlab.

Two major questions were asked; how to differentiate the quality categories of rice and how to identify locally grown Ghana rice from imported rice. PCAs built after MSC pre-processing provides clear cluster trends for each of the questions asked of the Ghana rice samples. For quality grades, the first three PCs explain 98.67% of the variation and whether the rice was locally grown or imported, the first three PCs explain 99.84%.

To build the identification model, the sample set was divided into a Reference set (n=377) and Validation set (188). To avoid bias in selection of samples into the subgroup, members in each subset were selected as follow; for every three samples, two were randomly selected for the reference set while the remaining sample formed part of the validation set (27). The handheld spectroscopic technique coupled with chemometrics was successfully developed for rapid identification of three rice quality categories (High quality, mid quality and low quality) at 87.5 % identification rate in the Reference set and 86.6 % in the Validation set by using multiplicative scatter correction plus support vector machine (MSC-SVM) algorithm. In addition, the technique was employed to differentiate between locally grown rice and imported rice at a rate of 100% for both Reference set (n=339) and Validation set (n=226) by using K-nearest neighbor (KNN) or SVM.

### GC-MS Jasmine rice samples from Ghana

VOC profile of high- and low-quality jasmine from Ghana where analyzed. PLS-DA analysis was performed on the data matrix obtained from GC-MS spectral analysis. Figure 6 shows that the first two components show separation of sample clusters. They explain 71.3 % of total variation. Principal component 1 shows clear separation in high quality and low-quality sample groups. Internal cross-validation by independent data set gave good Q2 (0.97632), R2 (0.99931) and accuracy (1.0) presented in Figure 6. Further permutation test analysis showed model validation with significance (p < 0.01). Hierarchical clustering analysis was further performed and showed clear separate clustering of High-quality jasmine samples from low quality jasmine samples from Ghana.

### Summary and next steps

We have successfully used this handheld spectroscopic technology in marketplaces in Ghana (Cape Coast and Accra) to improve the integrity of rice sold in these locations. Specifically, out of the 30 samples randomly selected and tested nondestructively in the market for either quality grade or imported versus local, 27 samples (90 %) had their quality status correctly identified while 28 samples (93 %) where correctly differentiated in terms of whether they were imported or locally grown. This trial further brings a great relief to consumers and quality control officer alike who are challenged and unable to perform rapid onsite determination of rice integrity. Through this work the attention of national authorities such as FDA - Ghana, and Ghana standard Authority has been drawn for future uptake and collaboration.

## Use Case 4: Vietnam

Vietnam is the third biggest rice exported in the world, following only India and Thailand (30). In 2018 Vietnam exported 6 million metric tonnes of rice with an estimated market value of $3 billion USD (31). The average price of rice is $502 USD per metric tonne (31). Vietnam enjoys two harvest each year, one in early spring and another in late summer.

There are 2 big issues in the Vietnamese rice market. The first issue is the export value. Vietnam does not currently have a national brand name, unlike India (Basmati), or Thailand (Hom Mali). Because Vietnam rice producers believe that the lack of a national brand limits the perceived value of Vietnamese rice exports, Vietnam is trying to build a national brand, but the communication for this brand is not efficient, and the rice importing countries still consider Vietnam rice as low quality. This is a big issue because even though Vietnam exports a lot of rice, the income of rice farmers is still very limited. In addition, some of Vietnam rice varieties are sold at higher prices, for example Dai Thom 8, and some rice manufacturers will mix with lower priced varieties while charging the higher price. This kind of fraud is not good for Vietnamese consumers as they pay high price for the low-quality rice.

In Vietnam the quality of rice is defined by the odor, appearance, palatability and texture of the cooking rice (32). The high-quality rice always has a bright surface, nearly transparent body, soft texture and good cooking quality. The low-quality rice has a hard texture and low cooking quality. The price of high-quality rice is about 20-25% higher than low quality rice. In the local market, there are more than 150 different brands. The most popular brands are always exposed to fraud risk in that fraudulent retailers mix low-quality rice into high value brands in order to get more benefit. The big companies try to protect their brands, but they can’t find an efficient tool for this. Moreover, as rice is not a high value product, using modern methods, with costly high-tech equipment is not a viable choice. Especially when considering that the number of sub-samples needed to be analyzed to protect against fraud events would need to be increased compared to meeting statutory requirements for safety testing. In this instance a two-tier system makes perfect sense where sample screening can be increased using a low-cost portable device while only suspect samples would be taken on to the more expensive mass spectrometry-based methods.

In this study we differentiated between high-quality, mid-quality and low-quality rice. We collected 4 varieties: Dai Thom 8 (n=93); OM 5451 (n=96); OM 6976 (n=55); and IR 50404 (n=68) from 4 different provinces in the Mekong delta. In this collection, IR 50404 is considered as low-quality while OM 6976 is considered mid-quality rice. These varieties are often used to mix with the higher quality rice to lower the production cost but to increase profit for the fraudsters.

### Handheld NIR

Samples where scanned in triplicate on the SCiO. The data were divided into a Reference set (n=179 samples) and Validation set (133 samples). The Reference set was used to develop models in Simca 15 while the Validation set was used as an independent set to evaluate how well the models could accurately predict results. Multiple pre-processing and model algorithms where evaluated. The spectra share very similar absorbance patterns therefore it’s essential that all the spectra undergo the most suitable pretreatment techniques, by applying preprocessing techniques individually and in combination in an iterative approach. The best results, in terms of correctly predicting high-quality, mid-quality or low-quality rice was obtained using an OPLS-DA model on data that had undergone SNV 1st derivative and SG smoothing, in that order, with a correct prediction of 96.6.7 %, 27.3 % and 90.9 % respectively. The model has performed poorly for the identification of mid-quality rice with only 6 of 22 being correctly identified, with 9 samples identified as high-quality and 5 samples could not be clearly assigned to any of three clusters. This was not an unexpected result in that the mid-quality rice cluster bridged the gap between high-quality and low-quality rice varieties. Internal cross-validation of this model shows an R2 of 0.617 and a Q2 of 0.576. When building models excluding the mid-quality rice, we successfully identified high-quality and low-quality rice 92.1 % and 90.9 % respectively. These initial results indicate that the either the multi-class or two-class model could be successfully used as a field deployable screening tool to correctly identify low-quality rice in approximately 90 % of the cases. However, suspect samples and random quality control samples could be sent back to the laboratory for Tier two testing. Additional work would need to be undertaken to see if a model could be optimized for mid-quality rice identification.

### LC-QTOF-MS

This study was designed as a proof of concept to assess further development of a targeted type methodology by initiating a “markers” discovery phase with high resolution mass spectrometry.

Sample set was constituted by 4 different rice varieties (Dai Thom (n=8); OM 5451 (n=7); OM 6976 (n=7); and IR 50404 (n=8) divided in 3 groups based on subjective quality (Low, Medium and High).

Data were processed using Profinder where alignment, deconvolution and then peak piking were carried out. A total of 893 Features were detected and the resulting matrix was exported to SIMCA for multivariate analysis and subsequent evaluation. To assess the general quality of the acquired spectral data (univariate/Pareto scaled) PCA and model assessment were performed. Next, data were mean centered, either Pareto or univariate scaled and grouped into respective classes prior to OPLS-DA. R2 (cumulative), Q2 (cumulative) values and RMSECV were used to determine the validity of the models, with R2 (cum) employed as an indicator of the variation described by all components in the model and Q2 & RMSECV as measures of how accurately the model can predict class membership. All Quality controls were found to be tightly clustered within the center of respective PCA scores plots which indicate good discrimination between different classes of rice, with the first five PCs explained 66% of variation, Figure 7a. An overall three group, Figure 7b, and multiple supervised OPLS-DA models were then generated by comparing rice by quality attribute. OPLS-DA classification models were validated using response permutation (200 iterations) in SIMCA. For instance, an OPLS-DA model between High and Low quality was generated with one latent component and three orthogonal components with resulting R2 = 99.9%, Q2= 92.2 % and RMESECV of 13.3%, Figure 8a. One of the main advantage of this LC-QTOF-MS methodology is to allow more data mining by filtering the main features responsible for the discrimination between group in order to select a limited number of “quality markers” to be transferred to a targeted method on a dedicated routine instrument such as an LC-MS/MS system. For each OPLS-DA model, associated S-plots and variable importance in projection (VIP) plots enabled identification of a set of most promising ions responsible for class separation among all rice varieties Figure 8b. No unique “markers” were identified but a set of putative “markers” to be quantitatively monitor has to be investigated further. More new samples will need to be analyzed in order to validate these markers and then transfer to a proper targeted analysis acting as confirmatory test using for instance a triple quadrupole instrument can be considered. However, this mass spectrometry methodology requires a considerable amount of resources in term of instrument, software and staff time and therefore can only be implemented within a modern well-equipped laboratory.

### Summary and next steps

During the project and an International workshop, the Vietnam team have introduced this new technique to more than 50 stakeholders (rice manufacturers, rice distributors, rice quality control staff) in Vietnam. The stakeholder interest in this approach has been significant and a proposal to develop a method for rapid detection of low-quality rice mixed with high-quality rice is being explored. In addition, some rice manufacturers are also interested in using this technique to protect their GI rice brand name.

# General Discussion

Towards the end of the project, several demonstrations and workshops where held in Ghana, China and Vietnam (Figure 9). Rice industry stakeholders including farmers, processors, regulators, exporters, traders, and consumers, were invited in each country and from neighboring countries. The topic and nature of the workshops generated a lot of interest, with more than 180 participants in the Ghana workshop alone. The purpose of these workshops where to not only to disseminate information on the occurrence of rice fraud but also to demonstrate and discuss the use of the two-tier system, handheld NIR for in-field detection supported by laboratory based mass spec methods, to address the relevant rice authenticity concerns in each country specifically. Feedback was also collected on how Users could make practical use of a two-tier testing regime. Discussions around the two-tier system generated many talking points and excitement grew for the number of potential applications and follow-up activities where instigated.

Both the handheld NIR and mass spec methods have pros and cons. From an end-user perspective, the benefits of the handheld NIR method include: it is rapid, facilitating fast results; low cost, therefore does not significantly add to the cost of food and more samples can be analyzed; provides a degree of consumer confidence; can be performed with minimal training by anyone; and it is not confined to a laboratory. However, models are only as good as the databases used to generate them; and it remains a screening test, therefore legal decisions should not be based on its finding. LC-QTof methods are considered more accurate and can deal with more complex matrices. However, it remains expensive and requires laboratory infrastructure. Method development and analysis are time and resource consuming. However, using LC-HRMS facilitates development of a future confirmatory test such as on a LC-MS/MS system, however there is no guarantee that this can be achieved prior to model development.

It is our belief that combining the screening and confirmation methods into a multi-tiered testing system is the best approach. They complement each other to provide the best solution to the end-user. Screening with the handheld NIR will facilitate increasing the number of samples to be analyzed and provide a fast-positive release while identifying the suspicious samples that are then forwarded for the more sensitive mass spec methods for confirmation.

Through our initial discussions between partners it became clear that a “one size fits all” approach would not work, for example, India had no need for Chinese GI rice determination nor concern of distinguishing high- or low-quality rice. Adding these data streams into the Indian models would have cause unnecessary noise. Tackling each as an individual use case helped to meet country specific needs and this should be the approach taken in the future.

# Conclusion

Different analytical approaches are being used for detection, prevention, and mitigation of food fraud. Different analytical platforms have inherent strengths and limitations which have been well-summarized by many others in numerous reviews. So far, no single technology can be used to address all forms of food fraud, therefore, a combination of different approaches may be needed. The overall objective of this study was to explore the feasibility of using a single testing approach to adulteration in rice. We aimed to determine if the solution could be used to address the distinct and complex real-world circumstances that are encountered with this commodity. We employed a two-tiered approach consisting of a screening method (based on portable NIR) followed by confirmatory analysis, using combinations of LC-MS, GC-MS, and ICP-MS.

Our results demonstrate the value of a two-tiered approach: using orthogonal detection modalities yields considerably more information than any single method. Moreover, consideration should be given to use-case practicalities: will the end-user benefit more from screening approaches, or does the use-case require high-confidence confirmatory analysis?

These results also underscore the need to ensure that “fingerprinting” techniques employ models that have been built using authentic reference samples and adequate numbers of samples – typically at least hundreds samples—reflecting the diversity of the food product. For commodities like rice, consideration must be given not only to the variety of rice, but also to growing conditions, the number of harvests, and seasonal fluctuations. The best models will be built and refined over years.

# Acknowledgements

The authors would like to thank the Agilent Foundation, Mars Global Food Safety Center and the National Natural Science Foundation of China for their support throughout this project.

# References

(1) PwC (2017) Food fraud vulnerability assessment 2020, 18 February <https://www.pwc.com/gx/en/services/food-supply-integrity-services/food-fraud-vulnerability-assessment.html>

(2) Kendall, H., Clark, B., Rhymer, C., Kuznesof, S., Hajslova, J., Tomaniova, M., Brereton, P., & Frewer, L. (2019) *Trends in Food Science & Technology* **94**, 79-90

(3) Guardian, T. (2008) Chinese figures show fivefold rise in babies sick from contaminated milk 2020, 18 February <https://www.theguardian.com/world/2008/dec/02/china>

(4) Gelpi, E., de la Paz, M.P., Terracini, B., Abaitua, I., de la Camara, A.G., Kilbourne, E.M., Lahoz, C., Nemery, B., Philen, R.M., Soldevilla, L., Tarkowski, S., & Syn, W.C.S.C.T.O. (2002) *Environmental Health Perspectives* **110**, 457-464

(5) Seck, P.A., Diagne, A., Mohanty, S., & Wopereis, M.C.S. (2012) *Food Security* **4**, 7-24

(6) Ali, T., Jhandhir, Z., Ahmad, A., Khan, M., Khan, A.A., & Choi, G.S. (2017) *Multimedia Tools and Applications* **76**, 24675-24704

(7) Kukusamude, C., & Kongsri, S. (2018) *Food Control* **91**, 357-364

(8) Rodriguez, L., Li, J., & Sar, S. (2014) Social trust and risk knowledge, perception and behaviours resulting from a rice tampering scandal. in *International Journal of Food Safety, Nutrition and Public Health*, Vol. 5

(9) Medina, S., Perestrelo, R., Silva, P., Pereira, J.A.M., & Camara, J.S. (2019) *Trends in Food Science & Technology* **85**, 163-176

(10) Creydt, M., & Fischer, M. (2018) *Electrophoresis* **39**, 1569-1581

(11) Consumer Physics (2020) Technology 2020, 10 February <https://www.consumerphysics.com/business/technology/>

(12) Hopfer, H., Nelson, J., Collins, T.S., Heymann, H., & Ebeler, S.E. (2015) *Food Chemistry* **172**, 486-496

(13) whfoods.org (2020) What is the difference between basmati rice and regular rice (both the brown and white versions)? 2020, 24 February <http://whfoods.org/genpage.php?tname=dailytip&dbid=365>

(14) Tilda (2020) Why Basmati Is Best 2020, 24 February <https://www.tilda.com/whats-new/why-basmati-is-best/>

(15) APEDA (2015) BASMATI RICE 2020, 24 February <http://apeda.gov.in/apedawebsite/SubHead_Products/Basmati_Rice.htm>

(16) China National Bureau of Statistics (2018) 国家统计局关于2018年粮食产量的公告 2020, 06 March <http://www.stats.gov.cn/tjsj/zxfb/201812/t20181214_1639544.html>

(17) National Rice Data Centre (2019) China Rice Variety and Pedigree Database 2020, 6 March <http://www.ricedata.cn/variety/index.htm>

(18) National Bureau of Statistics (2018) Announcement of the National Bureau of Statistics on the Grain Output in 2018 2020, 6 March <http://www.stats.gov.cn/tjsj/zxfb/201812/t20181214_1639544.html>

(19) Wadood, S.A., Guo, B.L., Zhang, X.W., Hussain, I., & Wei, Y.M. (2020) *Microchem J.* **152**

(20) Chung, I.M., Kim, J.K., Lee, K.J., Park, S.K., Lee, J.H., Son, N.Y., Jin, Y.I., & Kim, S.H. (2018) *Food Chemistry* **240**, 840-849

(21) Zhang, Y., Song, Q., Yan, J., Tang, J., Zhao, R., He, Z., Zou, C., & Ortiz-Monasterio, I. (2010) *Euphytica* **174**, 303-313

(22) Cheajesadagul, P., Arnaudguilhem, C., Shiowatana, J., Siripinyanond, A., & Szpunar, J. (2013) *Food Chemistry* **141**, 3504-3509

(23) Maione, C., Batista, B.L., Campiglia, A.D., Barbosa, F., & Barbosa, R.M. (2016) *Computers and Electronics in Agriculture* **121**, 101-107

(24) Fiamohe, R., Demont, M., Saito, K., Roy‐Macauley, H., & Tollens, E. (2018) *EuroChoices*

(25) Grow Africa (2018) ECOWAS RICE FACTBOOK 2020, 27 February <https://www.growafrica.com/resources/ecowas-rice-factbook>

(26) Business Ghana (2018) Ghana imports over US1,162 billion worth of rice annually-Minister 2020, 27 February <https://www.businessghana.com/site/news/business/167665/Ghana-imports-over-US1-162-billion-worth-of-rice-annually-Minister>

(27) Teye, E., Amuah, C.L.Y., McGrath, T., & Elliott, C. (2019) *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **217**, 147-154

(28) Business Ghana (2019) Rice production in Ghana to reach 750,000 2020, 27 February <https://www.businessghana.com/site/news/general/194744/Rice-production-in-Ghana-to-reach-750,000>

(29) Diako, C., Sakyi-Dawson, E., Bediako-Amoa, B., Saalia, F.K., & Manful, J.T. (2010) *Nature and Science* **8**, 8

(30) Statista (2019) Paddy rice production worldwide 2017-2018, by country 2020, 10th February <https://www.statista.com/statistics/255937/leading-rice-producers-worldwide/>

(31) Asemconnectvietnam.gov.vn (2019) Vietnam’s rice export markets in first nine months of 2019 2020, 5 March <http://asemconnectvietnam.gov.vn/default.aspx?ID1=2&ZID1=8&ID8=91612>

(32) Lu, L., Tian, S.Y., Deng, S.P., Zhu, Z.W., & Hu, X.Q. (2015) *Rsc Advances* **5**, 47900-47908

# Figure Legends

Figure 1Chemometric models to class various basmati varieties of rice. a: Multiclass OPLS-DA model containing 1121 and its two adulterants, b: Multiclass OPLS-DA model containing Taraori and its three adulterants.

Figure 2:VOC Principal component analysis of a: Pusa 1121, Pusa 1509 and Sugandha rice samples. 10-15 samples were analyzed from each group; b: Duplicate, Shabnam, Sharbati and Taraori rice samples. 10-15 samples were analyzed from each group.

Figure 3: Optimal model for Chinese rice samples, produced in Simca 15.

Figure 4 PLS-DA score plot derived from 30 elements in six GI rice.

Figure 5: PCA analysis of Wuchang (WC), Jingshanqiao (JSQ), Panjin-Yanfeng (PJ-YF), Panjin-Liaoxing (PJ-LX), Sheyang (SY) and Dongjinxi (DJX)

Figure 6: Principal component analysis of High quality and low-quality Jasmine rice samples from Ghana. 10 samples were analysed from each sample group.

Figure 7a: Principal component analysis of High quality, Medium quality and low-quality rice samples from Vietnam. b: Multiclass OPLS-DA model containing High, Medium and Low quality.

Figure 8a:Multiclass OPLS-DA model containing high and low-quality rice samples from Vietnam. b: S-plot from Multiclass OPLS-DA model containing high and low-quality rice samples from Vietnam with possible markers of low quality investigated (circle)

Figure 9: Demonstration being performed at a market in Ghana

# Table Legends

Table 1: Confusion table for optimal Random Forest model developed in ‘The Lab’ using Chinese rice samples

Table 2: Confusion matrix derived from external Validation set