



The use of wavelength dispersive X-ray fluorescence (WDXRF) spectrometry and multivariate techniques for assessment of illegal dyes in spices

Adusei E*, Fernandes T, Costa D, Brito JA, Gonçalves LL

Centro de Investigação Interdisciplinar Egas Moniz (CiiEM)- Instituto Superior de Ciências da Saúde Egas Moniz, Campus Universitário, Quinta da Granja, Monte de Caparica, 2829-511 Caparica, Portugal.

Received 10 May 2017; Revised 5 July 2017; Accepted 5 July 2017; Available online 14 July 2017

Abstract

Sudan dyes are carcinogenic azo and diazo compounds that are banned for use in food worldwide including the European Community. The ability of WDXRF spectroscopic technique to predict the types and levels of adulteration of spices suspected to be adulterated with Sudan I-IV, Para Red and Sunset Yellow dyes was investigated. Logistic regression and discriminant analysis classification models were developed to predict the type of adulteration using WDXRF spectral features. Prediction of the levels of adulteration was assessed by using multiple regression analysis. 83% of the 210 adulterated samples were correctly classified by the logistic regression with 90% sensitivity, 75% specificity with a prediction power of 92% into respective adulteration groups. 86% and 90% correct prediction were obtained for discriminant analysis models with 94% sensitivity and 74% specificity. Three multiple regression models were performed for each data set. The full model based on both the Compton and Rayleigh ratios as well as the Compton and Rayleigh scatter intensities revealed to be the best model to predict the levels of adulteration with an adjusted R^2 between 95.1 to 99.0% with errors of estimate between 1.6 to 3.7%. WDXRF technique may represent a promising tool for the screening of Sudan dyes-adulterated spices and can be used as an alternative to classical methods for the determination of Sudan dyes present in food prior to confirmation with more sensitive chromatographic-based methods.

Keywords: Illegal dyes, Sudan dyes, Spices, WDXRF spectrometry

Introduction

Food quality is closely related to its colour and the use of colorants to enhance the aesthetic appeal of food is an old practice (Downham & Collins, 2000). Due to the instability of natural dyes over time, synthetic dyes are used to enhance the colours of food. Sudan dyes (I-IV and Para Red) are a family of synthetic azo and diazo compounds with very extensive industrial and scientific applications such as in textiles, cosmetics, pharmaceuticals, plastics, fuels and staining for microscopy. Due to their low cost, wide availability, brightly coloured nature and prolonged stability, Sudan dyes have been used to impact the colour of culinary spices to maintain their intense red colour. Sudan dyes are classified as category 3 carcinogens meaning they have unclassifiable carcinogenicity to humans (IARC, 2011). These compounds are metabolized to mutagenic and carcinogenic aromatic amines, as observed in animal model studies (Fonovich, 2013; Pan *et al.*, 2012).

Sunset Yellow, a sulphonated analogue of Sudan I, is permitted for use in some foodstuffs in the EU and other countries (Poul *et al.*, 2009). However, Sunset Yellow has been reported to contain Sudan I as a contaminant (Fonovich, 2013) and is also classified as 'category 3' carcinogen (IARC, 2011). It has also been found to be illegally present in foods like tomato sauce above the maximum permissible levels (Dixit *et al.*, 2008; Tripathi *et al.*, 2007, 2010). Sunset Yellow is attributable to cause hepatocellular damage and renal failure (Mekki *et al.*, 1998), reproductive toxicity (Tanaka, 1996) and has immunotoxicity potential in mice (Hashem *et al.*, 2010).

Although reports concerning Sudan dyes-adulterated foods have declined in the past few years (RASFF, 2008), the use of Sudan dyes as food additives is expected to continue as high levels of the

* Corresponding author: eaduseik@outlook.com

dyes have recently been located in other countries, especially in non-branded spices (Mishra *et al.*, 2007; RASFF, 2007).

Several analytical methods have been developed for the determination of Sudan dyes in foodstuffs with most of the methods been chromatography-based coupled with different extraction and detection methods (Fonovich, 2013; Rebane *et al.*, 2010). Although these techniques show high sensitivity, they suffer drawbacks such as time-consuming, labour-intensive, and expensive instrumentation.

The use of spectroscopic methods for the detection of food adulteration has gained attention in recent years due to the advantages that some of these spectroscopic methods offer. These advantages include rapid analysis, non-destructive, easy-to use, and accurate measurements. Combinations of spectroscopic techniques with multivariate methods have proven to be powerful tools for the detection of food adulteration (Ghosh & Jayas, 2009; Karoui & Debaerdemaeker, 2007). They include UV-visible (Di Anibal *et al.*, 2009, 2014), Raman (Di Anibal *et al.*, 2012), and ¹HNMR (Di Anibal *et al.*, 2011). However, these spectroscopic methods have some drawbacks. For instance, UV-visible, Raman and NMR will all require extraction of the dyes from the sample matrices. These introduce additional steps, cost and sometimes raises safety concerns regarding the extraction solvents, when these methods are being used.

Further the ability to predict the levels of adulteration of Sudan (I-IV, Para Red) and Sunset Yellow dyes by the above-mentioned spectroscopic techniques are yet to be reported in literature.

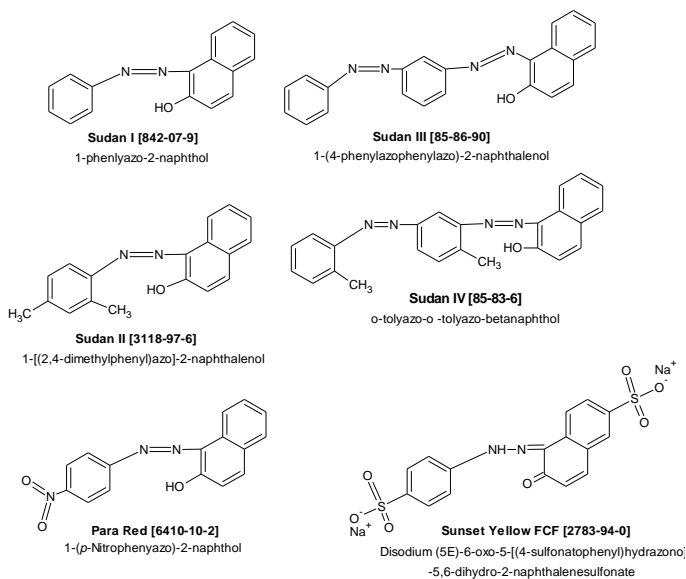


Figure 1. Chemical structures of the azo dyes discussed in this study.

Wavelength-dispersive X-ray fluorescence (WDXRF) spectroscopy is a physical analytical technique which measures the chemical elemental composition of materials both qualitatively and quantitatively based on their characteristic radiation (Bertin, 1978; Jenkins, 1974). WDXRF offers some advantages over other spectroscopic methods. They include multi-elemental analysis, direct measurements on the sample (solid or liquid samples, loose powders or pressed into pellets) without any chemical treatment, avoiding lengthy and laborious process of preparation, particularly in cases of high sample throughput, and sample contamination (Brouwer, 2010; Yellepeddi & Thomas, 2006). The short analysis

period and the ease of use of this technique, combined with low detection limits, are other advantages of WDXRF spectrometry.

Due to the carcinogenicity of Sudan dyes, they have been banned from use as food colorants in many countries and regions including the EU (EFSA Scientific Opinion, 2005). The EU Rapid Alert System for Food and Feed (RASFF) reports of the presence of Sudan dyes in various foods in member countries, since Sudan I was detected in the EU in May 2003 (RASFF, 2005). EU member countries have been entreated to establish uniform controls at the entry points into the Community to monitor entrance of imported food into the EU (Commission Decision 2009). In such instances, screening methods would be appropriate for the control and monitoring of goods or products instead of analytical reports for consignments of imported products into the Community.

This study aimed to develop a feasible, rapid and simple method to predict the type and levels of adulteration of spices suspected of adulteration with Sudan dyes based on spectral features available by this WDXRF spectroscopy technique, namely, the Compton and Rayleigh scatters.

Methodology

Reagents and Samples

Sudan I (catalogue no: 103624), Sudan III (catalogue no: S4131), Sudan IVa (catalogue no: 198102); Para Red (catalogue no: 100994) and Sunset Yellow (catalogue no: 465224) dyes were purchased from Sigma Aldrich (Lisbon, Portugal). Sudan IVb (catalogue no: A12181) dye was purchased from Alfa Aesar® (Lisbon, Portugal). All reagents were of high technical grade. Branded paprika and sweet pepper were purchased from local markets in Portugal.

Sample Preparation

3g total of Sudan I-IV, Para Red and Sunset Yellow-spiked spice samples from 0.5% to 50% (w/w) and 3g of blank spices and the dyes were prepared in triplicates and analyzed. The specific concentrations were 0.5, 1, 2.5, 5, 7.5, 12.5, 15, 20, 25, 30, 35, 40, 45, and 50%. The range encompasses some of the detected adulteration levels reported in literature. Before being analyzed on the WDXRF spectrometer, all samples were homogenized in borosilicate glass mortars and then transferred into high-density polyethylene cups with 35.8 mm diameter assembled with a 4 µm Prolene film.

WDXRF Analysis Parameters

All measurements were done in a 4kW commercial WDXRF system (Bruker S4 Pioneer) with maximum voltage and current set at 60kV and 50mA, respectively. Semi-quantitative standardless measurements were done in Helium mode, using a Rh X-ray tube fitted with a 75 µm Be window and a 34 mm diameter collimator mask. The standardless semi-quantitative mode, unlike the quantitative mode to produce actual elemental concentrations following system calibration, was used during the WDXRF analyses since the composition of the 'supposedly' adulterated spices were unknown.

Cellulose (C₆H₁₂O₆) powder was chosen to simulate the matrix. Spectral data was acquired using SpectraPlus Software Version 1.7.0

Data Analysis

Data analyses were done using STATGRAPHICS Centurion XVI Version 16.1.11 from StatPoint Technologies Inc. The WDXRF

spectral features: Compton & Rayleigh scatter intensities and their derivative ratios (Compton & Rayleigh ratios) were used for all statistical work. These quantities or variables were chosen because they are present in all WDXRF spectra and are easily acquired from WDXRF analyses.

Multivariate Classification Techniques

Logistic regression and discriminant analysis were used to predict the types of adulteration of spices adulterated with Sudan dyes (Sudan I, III, and IV dyes) and non-Sudan dyes (Para Red and Sunset Yellow FCF dyes) prior to the prediction of the levels of adulteration.

Logistic Regression

A total of 210 samples comprising of triplicates of adulterated paprika samples with Sudan I, III, IV, Para Red, and Sunset Yellow FCF dyes with adulteration levels between 0.5% and 50%, were used for the logistic regression. 126 samples belonging to paprika adulterated with either Sudan I, III, and IV were classified as 'Cases' (Sud) with a binary code of '1' whereas the remaining 84 samples adulterated with Para Red and Sunset Yellow FCF (Non-sud) were classified as 'Non-cases' with a binary code of '0' for the logistic regression. All four variables; Compton and Rayleigh intensities, Compton and Rayleigh ratios were used to develop the model for future predictions.

Discriminant Analysis (DA)

Firstly, the 210 samples were divided into two main groups, as was done for the logistic regression. Group 'Sud' comprised of 126 samples made up of Sudan I-, III-, and IV-adulterated paprika whereas the remaining 84 Para Red and Sunset Yellow adulterated paprika samples were categorized as group 'non-Sud'. This classification model (DA model A) was compared with the binary logistic regression model.

Secondly, a third group was introduced by splitting the 'non-Sud' group into two sub-groups making a total of 3 groups or classes for DA model B. The new groups or classes were as follows: class 'Sud' for Sudan I, III and IV-adulterated paprika samples, class 'PRed' for Para Red-adulterated paprika, and class 'SY' corresponding to Sunset Yellow-adulterated paprika samples. Similarly, all four variables which were the Compton and Rayleigh scatter intensities as well as their derivative ratios were used to develop the models.

Multiple Regression Analysis

Three multiple regression models: *Model A* based on the Compton and Rayleigh ratios; *Model B* based on the Rh Compton and Rayleigh scatter intensities; and *Model C* or the full model based both the ratios and the scatter intensities were prepared. Multiple regressions with stepwise backward selection method with using the pre-defined predictors were prepared. A total of 45 data sets per sample, including the blank or pure spices (0% adulteration) were used to prepare the models. Influential points and points with unusual Studentized residuals (>3) were removed from the dataset for optimization of the model.

Result and discussion

WDXRF Spectral information

Compton and Rayleigh Ratios

The Compton ratio is an estimate based on the measured intensity of non-coherent scattering by light elements in the sample

compared to the theoretical intensity based on the simulated matrix (cellulose, $C_6H_{12}O_6$). It is greatly affected by the amount of lighter atomic weight elements present in a sample. A ratio close enough to 1 is desirable. The Compton ratios for the spices and dyes (Figure 2a) indicate that the matrix simulation in the study was credible. This could significantly minimize matrix-mismatched associated errors in the experiment. Paprika and sweet pepper had values closer to 1 which were expected since they are largely composed of cellulose and light elements. Among the dyes, Sudan III and Sudan IV had the highest and lowest ratios of 1.26 and 0.90, respectively. The Compton ratio for Sudan III was expected given the observed 'lighter' nature of the dyes during the experiment. This could suggest that Sudan III dye composed of higher concentration of lighter elements than the other dyes and, the opposite is true for the Sudan IV dye.

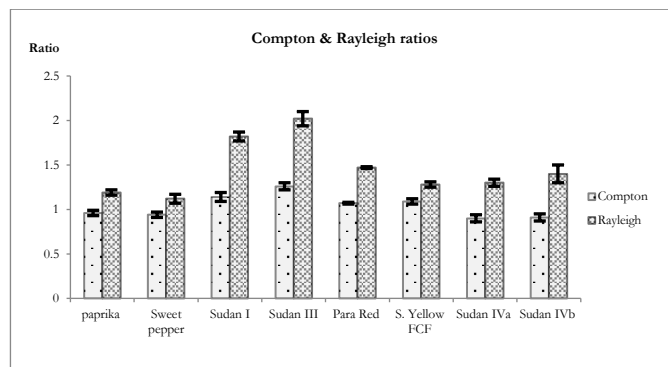


Figure 2a: Compton and Rayleigh ratios of the pure or blank spices and dyes. Paprika and sweet pepper had similar ratios. Sudan III and I dyes had the highest Compton and Rayleigh ratios. Sudan IVa and IVb dyes (from Alfa Aesar® and Sigma-Aldrich®, respectively), had the lowest Compton and Rayleigh among the dyes ratios. A ratio close to one is desirable for matrix simulation. The highest Compton ratio deviations were observed for Sudan I and III dyes whereas Para Red dye had the least deviations for both Compton and Rayleigh ratios.

Similarly, the Rayleigh ratio compares the intensity of coherent scattering due to heavier elements in the measured sample to that of the cellulose matrix. Matrices composed of heavier atomic weight elements (heavier matrix) are more likely to produce Rayleigh scattering than lighter matrices. In relation with the Compton ratio, the Rayleigh ratio gives information regarding the appropriateness of matrix simulation considering heavier elements. As shown in Figure 2a, it was observed that Rayleigh ratios of the spices (paprika and sweet pepper) as well as the dyes, with the exception of Sudan I and III, were closer to 1, with average ratios between 1.12 and 1.47. Paprika and Sweet pepper had average ratios of 1.19 and 1.12 respectively.

These observations support the deduction that the matrix simulation in this study was appropriate as already revealed by the Compton ratios. However, the Rayleigh ratios for Sudan III and I dyes were very high (2.02 and 1.82, respectively) indicating that Sudan III and I contain lesser amounts or concentrations of heavier elements than other spices. This issue can be resolved by the use of a more lighter matrix than cellulose, such as the C-5 sugars (hemicellulose). Sunset Yellow FCF had a low ratio of 1.28 among the dyes followed by Sudan IVa and IVb dyes. These findings were also consistent with the Compton and Rayleigh scattering intensities of the dyes.

Compton & Rayleigh Scatter intensities

The intensity of an X-ray quantum is defined as the number of X-ray quanta per unit time recorded at the detector as pulses per second (Pps) or counts per second (Cps). It contains the information about the concentration of the emitting elements present in a sample.

Sudan I and III had the lowest Rayleigh scatter intensities of 74.5 and 74.6KCps, respectively whereas Sunset Yellow FCF had the lowest Compton scatter intensity of 303.5KCps. In all spices and dyes, the Compton scatter intensities were higher than the Rayleigh scatter intensities, indicating that both spices and the dyes used in the study composed largely of lighter elements. Based on the Compton and Rayleigh ratios (Figure 2a) as well as the scattering intensities of the dyes and spices (Figure 2b), WDXRF exhibits the potential of determining Sudan dyes adulteration in paprika and sweet pepper when the appropriate multivariate classification techniques are applied.

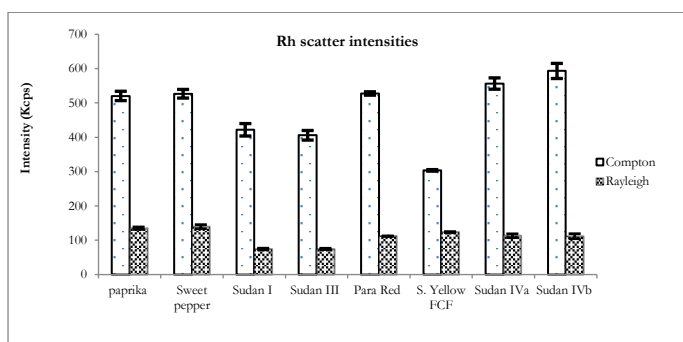


Figure 2b: Rh Compton and Rayleigh scatter intensities of the spices and dyes. Sudan IVa and IVb are Sudan IV dyes from Alfa Aesar® and Sigma-Aldrich, respectively. The highest Compton scattering was observed for Sudan Iva and IVb dyes whereas Sunset Yellow FCF had the lowest Compton scatter intensity among the dyes. Paprika and Sweet pepper had similar Compton and Rayleigh scatter intensities, consistent with the Compton and Rayleigh ratios. Sudan I and III had the lowest Rayleigh scatter intensities, corresponding to the highest Rayleigh ratios in those dyes.

Elemental Compositions of the pure spices (unadulterated paprika and sweet pepper)

The elemental compositions of paprika, sweet pepper and the dyes obtained using the standardless semi-quantitative methods in the WDXRF analysis are shown below. The estimated concentrations as well as the statistical error (instrumental) associated with the measurement are summarized below. It must be emphasised that the concentrations of the elements (ppm) are estimated values whose exact amounts can only be obtained through quantitative analysis following the development of calibration models for the elements present in the samples. However, these estimated concentrations are accurate given the uncertainties (statistical errors) associated with the estimations, especially, elements with uncertainties at 5% and below.

Paprika and sweet pepper had very similar elemental compositions. Ten (10) elements were detected in both paprika and sweet pepper with Br exclusively in sweet pepper. The elemental compositions of paprika and sweet pepper are summarized in Table 1. The most predominant element in both spices was K. The concentrations of Cl, P, Mg, S, Ca, and Si, were also predominant in both spices. However, concentrations of Ca were significantly different between paprika (2,648.3±96.8 ppm) and sweet pepper

(1,713.3±25.2 ppm). Al, Fe, and Cu were present at trace levels in both spices. Trace amounts of Br was also detected in sweet pepper. As already stated, these elemental concentrations are accurate, although they were not measured using a quantitative mode which would require specific calibrations, given the low uncertainties associated with the estimated concentrations (Table 1).

Table 1: Elemental compositions of pure (unadulterated) paprika and sweet pepper

Element	Concentration (x10 ⁴ ppm)		Stat. error (%)	
	Paprika	Sweet p.	Paprika	Sweet p.
K	2.616±0.092	2.921±0.033	0.2	0.2
P	0.357±0.019	0.356±0.004	1.2	1.7
Cl	0.340±0.016	0.362±0.004	1.3	1.3
Mg	0.287±0.018	0.277±0.002	1.6	1.7
Ca	0.265±0.010	0.171±0.003	0.8	1.0
S	0.223±0.014	0.218±0.002	1.1	1.1
Si	0.115±0.008	0.117±0.002	2.8	2.7
Al	0.033±0.002	0.035±0.002	6.9	6.6
Fe	0.027±0.001	0.025±0.0003	1.2	1.2
Cu	0.003±0.0002	0.003±0.0001	4.9	4.9
Br	**	0.005±0.0001	**	2.8

Sweet p refers to sweet pepper and **Stat. error**, the statistical error associated with the estimation of the concentrations of the elements present in the samples by the WDXRF instrument.

Logistic Regression (LR)

Logistic regression was employed as a classification tool to predict the type of spice adulteration prior to the prediction of the levels of adulteration using multiple regression analysis. The omnibus test for the model coefficients of the LR model revealed that the fitted model was statistically significant ($p < 0.05$) at the 95% confidence interval or higher with a Chi-square (χ^2) of 145.4 and degree of freedom (df) of 4. This indicates a significant relationship between the predictors or variables (Compton & Rayleigh intensities and Compton & Rayleigh ratios) fitted in the model at the 95% confidence level. The Hosmer-Lemeshow chi-square test was non-significant ($p > 0.05$) indicating a good fitting of the model supported by the Nagelkerke R^2 which was 0.675.

The overall correct classification of the LR model was 83.3% at 0.5 cut-off value with 88.9% and 75% sensitivity and specificity, respectively (Table 2a). These sensitivities and specificities are very good compared to performances of reported spectroscopic screening methods coupled with multivariate classification, given that, extraction of the dyes from the food matrices is not a prerequisite of the WDXRF technique.

Besides, the 88.9% sensitivity indicates that 88.9% of the spices adulterated with the Sudan dyes (I, III and IV) were correctly classified as Sudan dyes-adulterated. This is very good since most of the reported spice adulteration cases are associated with Sudan dyes.

The Receiver Operating Characteristic (ROC) curve, which summarizes the performance of the fitted LR model at the various cut-offs, considering the sensitivity and specificity of the fitted model (Figure 2c) produced an area of 92.3%. The area under the

curve (AUC), also known as concordance index, c , indicates the prediction power of the fitted LR model.

Table 2a: Logistic Regression classification performance, sensitivity and specificity.

Category	Size	Correct Classification (%)			Sensitivity %	Specificity %
		Case	Non-case	Total		
Case	126	112	14	83.3	88.9	75.0
Non-case	84	21	63			

Sudan I, III, and IV-adulterated paprika samples are classified as 'true' cases whereas paprika adulterated with Para Red and Sunset Yellow dyes (non-Sudan dyes or other dyes-adulterated samples) are known as 'false' cases.

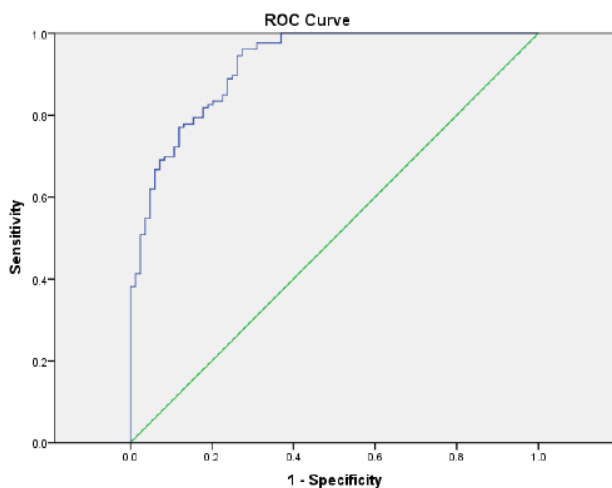


Figure 2c: Receiver Operating Characteristic (ROC) curve of the logistic regression model

Discriminant Analysis (DA)

Two different DA models were developed. *Model A* in which data were classified into two broad classes for comparison with the fore-described LR classification model. Model B was developed to investigate observations that was made in the classification results obtained from model A. Both fitted DA models were statistically significant ($p < 0.05$) at the 95% confidence level.

The single discriminant functions obtained in models A and B were statistically significant at the 95% confidence levels with an Eigen value and canonical correlation of 0.916 and 0.69 for model A. The Eigenvalues were 2.482 and 0.761, canonical correlations of 0.84 and 0.66, with relative percentages of 76.5% and 23.5% for functions 1 and 2, respectively in model B. The clustering of the adulterated paprika samples in both models are shown in Figure 3a whereas Figure 3b shows the plot of the discriminant functions of model B.

The performances of the DA classification models (Table 2b) show excellent sensitivity (94.4%) and good specificity (74%) given that extraction of the dyes from the food matrices are not prerequisite of the WDXRF spectroscopy compared to DA classifications obtained from other spectroscopic techniques combined with multivariate classifications in which extraction was a pre-requisite.

Of particular interest was model A in which the adulterated paprika samples were grouped as was used in the LR. The overall correct classification of model A were 86.2% and 85.7% before and after cross-validation, as opposed to the 83.3% obtained in the LR model. This shows the robustness of the WDXRF technique in

predicting the adulteration types of paprika samples adulterated with Sudan (I, III, and IV) dyes and non-Sudan (Para Red and Sunset Yellow FCF) dyes, irrespective of the classification technique that is employed, although there was improved sensitivity in the DA model than in the LR model. DA model B confirmed the classification of some of the Sudan dyes-adulterated paprika samples as non-Sudan dyes-adulterated and vice versa (Table 2b). These misclassifications could possibly be due to the structural relatedness of Para Red and Sunset Yellow FCF dyes to Sudan I dye as shown in Figure 1.

The classification performances of the LR and DA models reveal the potential of WDXRF to differentiate between paprika samples adulterated with Sudan and non-Sudan dyes with very high sensitivity and specificity.

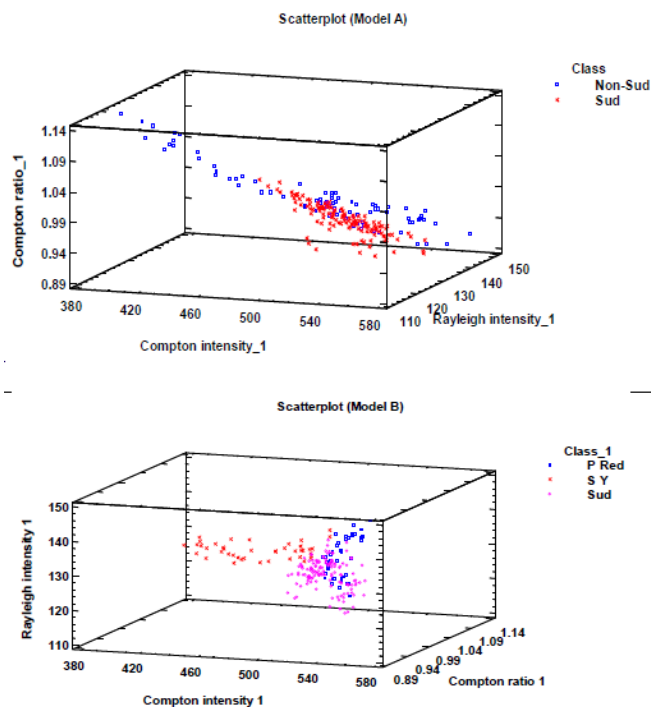


Figure 3a: The scatterplots of Discriminant Analysis (DA) models A (top) and B (bottom) showing the clustering of the various adulteration classes. Sud and Non-Sud correspond to Sudan dyes and non-Sudan dyes-adulterated paprika samples. P Red and SY correspond to Para Red-adulterated and Sunset Yellow-adulterated paprika samples, respectively.

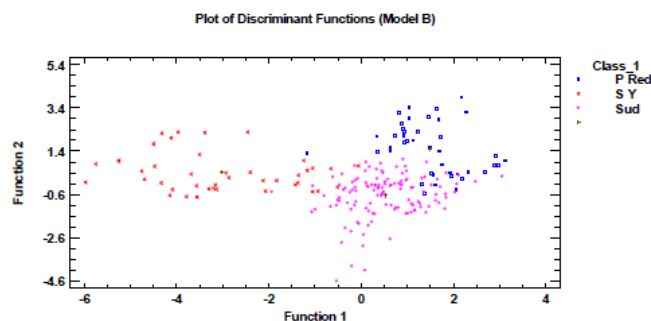


Figure 3b: Discrimination of the adulterated paprika samples in DA model B. Sud, P Red and SY correspond to Sudan I, III, and IV-adulterated, Para Red-adulterated and Sunset Yellow-adulterated paprika samples, respectively.

Table 2b: Classification performance of the Discriminant Analysis models with sensitivity and specificity

Model	Class	Actual Size	Predicted group (%)		Sensitivity (%)	Specificity (%)	
			Sud	Non-Sud			
A	<i>Sud</i>	126	119 (94.4)	7 (5.6)	94.4	73.8	
	<i>Non-Sud</i>	84	22 (26.2)	64 (73.8)			
Cross-validation							
	<i>Sud</i>	126	118 (93.7)	8 (6.3)	93.7	73.8	
	<i>Non-Sud</i>	84	22 (26.2)	62 (73.8)			
			Sud	P Red	SY		
B	<i>Sud</i>	126	114 (90.5)	11 (8.7)	1 (0.8)	90	86.9
	<i>P Red</i>	42	4 (9.5)	37 (88.1)	1 (2.4)		
	<i>SY</i>	42	5 (11.9)	0	37(88.1)		

Sud, *Non-Sud* refer to Sudan dyes-adulterated and non-Sudan dye-adulterated paprika samples, respectively whereas *P Red* and *SY* refer to Para Red-adulterated and Sunset Yellow-adulterated paprika samples, respectively.

Multiple Regression Analysis

Three (3) multiple regression models were prepared for paprika and sweet pepper adulterated with each of the 5 dyes (Sudan I, Sudan III, Sudan IV, Para Rad and Sunset Yellow FCF). Model *A* was based on the Compton and Rayleigh ratios, *B* based on the Compton and Rayleigh scatter intensities whereas model *C* was the full model based on all four spectral quantities (Compton and Rayleigh intensities and their derivative ratios). The best models were selected based on the “goodness of fit” (Adjusted R^2), Mallow’s C_p statistic, Akaike Information Criterion (AIC) and the standard error of estimate (SEE) associated with the models. The variance inflation factors (VIFs) of the predictors in each model were also discussed. Table 3a summarizes adjusted R^2 , Akaike Information Criterion and the standard errors of estimate of the fitted multiple regression models.

The adjusted R^2 statistic indicates the percentage of the observed variability in the percentage of adulteration which is explained by the independent variables in the fitted model. It indicates the goodness of the fitting. The adjusted R^2 values of the models ranged from 53.6% to 90.1%, 88.8% to 98.7%, and 95.1% to 99.0% for models *A*, *B*, and *C*, respectively. Moreover, the standard errors of estimates of the models ranged from 5.3 to 11.5%, 1.9 to 5.6% and 1.7 to 3.7% for *A*, *B*, and *C*, respectively.

Generally, it was observed that there were significant improvements in the goodness of fit and the standard errors from models *A* to *C*. Although there were appreciable degree of fitting for all dyes in paprika and sweet pepper (>50%) in model *A*, the goodness of fit values suggest that the Compton and Rayleigh ratios alone may not be the ideal variables to predict levels of adulteration of paprika and sweet pepper suspected to be adulterated with Sudan dyes. Of particular interest was the prediction of the levels of Para-Red adulteration in the adulterated paprika samples by the three models (Table 3a), where there was a significant increase of about 44% in the adjusted R^2 values between

models *A* and *C*, with corresponding decrease in the standard error of estimate from 7.5% to 2.8%, respectively. Similarly, there were significant increases ranging from 17% to 22%, with the exception of Sunset Yellow adulterated paprika, in the fitting between models *A* and *C*, with corresponding decrease in the standard error of estimate between 8.6% and 3.6%.

The highest fitting and lowest error of estimate were observed consistently for Sunset Yellow-spiked paprika. This could be mainly due to the fact that Sunset Yellow is a hydrophilic dye compared to the other highly hydrophobic dyes; hence the paprika-dye interaction was more appropriate, among other things such as sample preparation errors and instrumental errors.

The Mallow’s C_p statistic measures the bias in a multiple regression model by comparing the total mean squared error with the true variance. An unbiased model has an approximate C_p value of p , where p is the number of coefficients in the model (including the constant). The Mallow’s C_p values for models *A*, *B*, and *C* were 3, 3, and 5, respectively, indicating that all three models were unbiased. The absence of bias in the models indicates that the prediction of the levels of adulteration by the models were accurate.

The Akaike Information Criterion (AIC) is based on the residual mean squared errors with a penalty that grows with increasing number of coefficients in the model. Depending on the selected information criterion, models with the smallest residual errors with as few coefficients as possible is desirable. The observed AICs (Table 3a) ranged from 3.5 to 5.0, 1.4 to 3.6, and 1.3 to 2.9; for models based on the ratios (*A*), scatter intensities (*B*), and both ratios and scatter intensities (*C*), respectively. These findings indicated that the penalties reduced from models based on the ratios to the full models based on both the ratios and scatter intensities.

Table 3a: Summary of multiple regression results for the 3 models.

Adulteration	Adj. R^2 (%)			Mallow’s C_p			AIC			SEE (%)		
	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>
Sudan I in paprika	76.0	94.4	98.2	3	3	5	4.4	2.9	1.8	8.2	4.0	2.5
Sudan III “	68.7	91.6	95.1	“	“	“	4.6	3.3	2.9	9.4	4.9	3.7
Sudan IV “	79.2	89.0	97.1	“	“	“	4.2	3.6	2.3	7.7	5.6	2.8
Para Red “	53.6	88.8	97.1	“	“	“	5.0	3.6	2.3	11.5	5.6	2.8
Sunset Yellow “	90.1	98.7	99.0	“	“	“	3.5	1.4	1.3	5.3	1.9	1.7
Sud I in sweet p.	81.0	95.6	98.1	“	“	“	4.1	3.5	1.4	7.2	3.5	2.3
Sud IV “	79.1	93.3	98.4	“	“	“	4.2	3.1	1.8	7.7	4.4	2.3

A: model based on Compton and Rayleigh ratios; **B:** model based on the Rb Compton and Rayleigh scatter intensities; and **C:** the full model based on both the ratios and intensities of the Compton and Rayleigh scatters. **AIC** is the Akaike Information Criterion and **SEE** is the standard error of estimate. Models *A*, *B*, and *C* had Mallow’s coefficients of 3, 3, and 5, respectively.

The variance inflation factor (VIF) in multiple regression analysis is an indicator of multicollinearity in regression models and computationally reciprocates the tolerance of a regression model. It states the inflation in the magnitude of the standard errors associated with a particular beta weight due to multicollinearity. A maximum VIF of 10 was allowed in this study. Summary of the

VIFs as well as the tolerance in the three models are shown in Table 3b.

The VIFs for the models based on the scatter intensities (1.05-4.22) and ratios (2.96-10.1) were very low compared to the VIFs for the full model (C), indicating the presence of multicollinearity in the full model. However, the multicollinearity in the full model was expected as the Compton and Rayleigh ratios are derivatives of the

Compton and Rayleigh scatter intensities, respectively, obtained by comparisons of the measured scatter intensities of the sample to the theoretical intensity of the simulated matrix. Besides, the ratios and intensities provide different spectral information from the measured samples. The Compton and Rayleigh scatter intensities are based on physical characteristics of the measured samples whereas the ratios take into account the nature of the sample compared to the matrix the simulated matrix (cellulose).

Table 3b: VIFs and tolerances of the multiple regression models.

Adulteration	Model A	Model B	Model C			
			Comp rat.	Ray rat.	Rh Comp	Rh Ray
Sud I pap	4.98 (0.20)	1.06 (0.95)	18.30(0.06)	87.17(0.01)	3.95(0.25)	71.06(0.01)
Sud III “	3.36 (0.30)	1.40 (0.71)	27.75(0.04)	52.68(0.02)	11.72(0.09)	44.13(0.02)
Sud IV “	4.32 (0.23)	1.28 (0.78)	32.10(0.03)	56.01(0.02)	9.31(0.11)	18.27(0.06)
Para Red “	9.39 (0.11)	2.76 (0.36)	24.52(0.04)	50.49(0.02)	6.70(0.15)	23.82(0.04)
S Yellow “	10.13(0.10)	4.22 (0.24)	49.27(0.02)	92.39(0.01)	29.03(0.03)	50.46(0.02)
Sud I swt p	4.59(0.22)	1.10 (0.91)	20.42(0.05)	50.45(0.02)	5.07(0.20)	30.01(0.03)
Sud IV swt p	2.96 (0.34)	1.05 (0.95)	18.99(0.05)	172.15(0.01)	7.75(0.13)	134.41(0.01)

Sud I refer to Sudan I, **Sud III** and **Sud IV** refer to Sudan III and Sudan IV dyes. **Pap** and **Swt p.** refer to paprika and sweet pepper. **Model A** is model based on the Compton and Rayleigh ratios; **Model B:** based on Rh Compton and Rayleigh scatter intensities; and **Model C** is the full model based on both ratios and scatter intensities. **Comp rat, Ray rat, Rh Comp, and Rh Ray** refer to the Compton ratio, Rayleigh ratio, Rh Compton and Rayleigh scatter intensities, respectively. The unbracketed and bracketed figures are the VIFs and tolerances, respectively.

Regarding this, the full model is the best model to predict the levels of adulteration in the spices. Additionally, since the standard errors of estimate and thus the variances in the full models are significantly less than in the model based on the Compton and Rayleigh scatter intensities, the inflation in the variances (VIFs) in the full model may not significantly affect the predictions of the levels of adulteration using the full model. Besides, since the Compton and Rayleigh scatter intensities as well as their derivative ratios are readily available in any WDXRF spectra, both models could be used for the prediction of levels of adulteration and the results compared with each other for any significant differences.

Based on the significantly higher goodness of fit values, coupled with significantly lower standard errors of estimate and information criterion penalties, it is concluded that the best model to predict the levels of adulteration in paprika and sweet pepper adulterated with Sudan (I, III, IV and Para Red) and Sunset Yellow FCF dyes is the full model based on the Compton and Rayleigh ratios with the Rh Compton and Rayleigh scatter intensities. In other words, the Compton and Rayleigh ratios together with the Rh Compton and Rayleigh scatter intensities are the best predictors of levels of adulteration in paprika and sweet pepper adulterated with Sudan and Sunset Yellow FCF dyes.

Conclusion

The abilities of WDXRF to predict the types and levels of adulteration of paprika and sweet pepper adulterated with illegal dyes (Sudan I, III, IV, Para Red and Sunset Yellow FCF) were demonstrated in this preliminary study. Multivariate classifications based on logistic regression and discriminant analysis indicate potential of WDXRF to differentiate between paprika samples adulterated with Sudan dyes and non-Sudan dyes.

The best model to predict the level of adulteration of spices suspected to be adulterated with Sudan dyes is the model which combines the Compton and Rayleigh ratios as well the Compton and Rayleigh scatter intensities which are naturally present in all XRF spectra.

WDXRF seems to be a promising tool for the prediction of levels of adulteration in spices adulterated with Sudan dyes and can be implemented as alternative to conventional screening methods to detect Sudan dyes in spices subject to confirmation by more sensitive chromatographic methods.

Acknowledgements

The authors acknowledge the financial support from the Erasmus Mundus Masters in Forensic Science (European Commission) and Egas Moniz Cooperativa de Ensino Superior, CRL as well as the Ghana Armed Forces for their support.

Conflict of interest

None declared.

References

- ASTA. (2005). *Sudan Red and Related Dyes – White Paper*. Retrieved from <http://www.astaspice.org/food-safety/white-papers/>
- Bertin, E.P. (1978). *Introduction to X-ray spectrometric analysis*. Plenum Press, New York-London.
- Brouwer, P. (2010). *Theory of XRF* (3rd ed.). Almelo, The Netherlands: PANalytical B.V.
- Commission Decision. (2009). Commission Decision 2009/669/EC. *Off J Eur Commun*, (669), 1–22.
- Di Anibal, C. V, Odena, M., Ruisánchez, I., & Callao, M. P. (2009). Determining the adulteration of spices with Sudan I-II-II-IV dyes by UV-visible spectroscopy and multivariate classification techniques. *Talanta*, 79(3), 887–92. doi:10.1016/j.talanta.2009.05.023
- Di Anibal, C., Rodríguez, M. S., & Albertengo, L. (2014). UV-Visible Spectroscopy and Multivariate Classification as a Screening Tool to Identify

- Adulteration of Culinary Spices with Sudan I and Blends of Sudan I + IV Dyes. *Food Analytical Methods*, 7(5), 1090–1096. doi:10.1007/s12161-013-9717-2
- Di Anibal, C. V., Marsal, L. F., Callao, M. P., & Ruisánchez, I. (2012). Surface Enhanced Raman Spectroscopy (SERS) and multivariate analysis as a screening tool for detecting Sudan I dye in culinary spices. *Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy*, 87, 135–41. doi:10.1016/j.saa.2011.11.027d
- Di Anibal, C. V., Ruisánchez, I., & Callao, M. P. (2011). High-resolution 1H Nuclear Magnetic Resonance spectrometry combined with chemometric treatment to identify adulteration of culinary spices with Sudan dyes. *Food Chemistry*, 124(3), 1139–1145. doi:10.1016/j.foodchem.2010.07.025
- Dixit, S., Mishra, K. K., Khanna, S. K., & Das, M. (2008). Bezoate and Synthetic Color Risk Assessment of Fast Food Sauces Served at Street Food Joints of Lucknow, India. *American Journal of Food Technology*, 3(3), 183–191.
- Downham, A., & Collins, P. (2000). Colouring our foods in the last and next millennium. *International Journal of Food Science & Technology*, 35(1), 5–22. doi:10.1046/j.1365-2621.2000.00373.x
- EFSA Scientific Opinion. (2005). Opinion of the Scientific Panel on Food Additives, Flavours, Processing Aids and Materials in Contact with Food on a request from the Commission to Review the toxicology of a number of dyes illegally present in food in the EU Question number EFSA-200. *The EFSA Journal*, 263, 1–71.
- EFSA Scientific Opinion. (2009). Scientific Opinion on the re-evaluation of Sunset Yellow FCF (E 110) as a food additive. *The EFSA Journal*, 7(11), 1330. doi:10.2903/j.efsa.2009.1330
- Fonovich, T. M. (2013). Sudan dyes: are they dangerous for human health? *Drug and Chemical Toxicology*, 36(3), 343–52. doi:10.3109/01480545.2012.710626
- Ghosh, P. K., & Jayas, D. S. (2009). Use of spectroscopic data for automation in food processing industry. *Sensing and Instrumentation for Food Quality and Safety*, 3(1), 3–11. doi:10.1007/s11694-008-9068-7
- Hashem, M. M., Atta, A. H., Arbid, M. S., Nada, S. a, & Asaad, G. F. (2010). Immunological studies on Amaranth, Sunset Yellow and Curcumin as food colouring agents in albino rats. *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association*, 48(6), 1581–6. doi:10.1016/j.fct.2010.03.028
- IARC monographs.(2011). Agents Classified by the IARC monographs, volumes 1-102. Available at <http://monographs.iarc.fr/ENG/Classification/ClassificationsGrouporder/pdf>. [Assessed on December 2013]
- Karoui, R., & Debaerdemaeker, J. (2007). A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chemistry*, 102(3), 621–640. doi:10.1016/j.foodchem.2006.05.042
- Mishra, K. K., Dixit, S., Purshottam, S. K., Pandey, R. C., Das, M., & Khanna, S. K. (2007). Exposure assessment to Sudan dyes through consumption of artificially coloured chilli powders in India. *International Journal of Food Science & Technology*, 42(11), 1363–1366. doi:10.1111/j.1365-2621.2006.01356.x
- Pan, H., Feng, J., He, G.-X., Cerniglia, C. E., & Chen, H. (2012). Evaluation of impact of exposure of Sudan azo dyes and their metabolites on human intestinal bacteria. *Anaerobe*, 18(4), 445–53. doi:10.1016/j.anaerobe.2012.05.002
- RASFF (2005). Rapid Alert System for Food and Feed, Annual Report. Available at: http://ec.europa.eu/food/food/rapidalert/report2005_en.pdf (Accessed March 2014).
- RASFF. (2007). Rapid Alert System for Food and Feed, *Annual Report 2006-07*. SSRN Electronic Journal. Available at http://ec.europa.eu/food/food/rapidalert/report2006_en.pdf [Accessed 15.06.13]. doi:10.2139/ssrn.1152122
- RASFF. (2008). Rapid Alert System for Food and Feed, *Annual Report*. Available at http://ec.europa.eu/food/food/rapidalert/report2008_en.pdf [Accessed 15.06.13].
- Rebane, R., Leito, I., Yurchenko, S., & Herodes, K. (2010). A review of analytical techniques for determination of Sudan I-IV dyes in food matrixes. *Journal of Chromatography. A*, 1217(17), 2747–57. doi:10.1016/j.chroma.2010.02.038
- Tripathi, M., Khanna, S.K., & Das, M. (2007). Surveillance on use of synthetic colours in eatables vis a vis Prevention of Food Adulteration Act of India. *Food Control* 18, 211–219.
- Tripathi, M., Dixit, S., Khanna, S. K., & Das, M. (2010). Intake pattern of synthetic colours by different age and socio-economic consumer groups of Lucknow, India, 3(1), 19–31.
- Yellepeddi, R., & Thomas, R. (2006). New Developments in Wavelength-Dispersive XRF and XRD for the Analysis of Foodstuffs and Pharmaceutical Materials, 21(September), 36–41.