

Document downloaded from:

<http://hdl.handle.net/10251/176260>

This paper must be cited as:

Teixeira Badaró, A.; Amigo, JM.; Blasco, J.; Aleixos Borrás, MN.; Rios Ferreira, A.; Pedrosa Silva Clerici, MT.; Fernandes Barbin, D. (2021). Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta. *Food Chemistry*. 343:1-9. <https://doi.org/10.1016/j.foodchem.2020.128517>



The final publication is available at

<https://doi.org/10.1016/j.foodchem.2020.128517>

Copyright Elsevier

Additional Information

# **Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta**

Amanda Teixeira Badaró<sup>a,b,c</sup>, José Manuel Amigo<sup>d,e</sup>, Jose Blasco-Ivars<sup>c</sup>, María Nuria Aleixos Borrás<sup>f</sup>, Amanda Rios Ferreira<sup>g</sup>, Maria Teresa Pedrosa Silva Clerici<sup>g</sup>, Douglas Fernandes Barbin<sup>a\*</sup>

<sup>a</sup>Department of Food Engineering, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862, Campinas – São Paulo, Brazil

<sup>b</sup>Departamento de Tecnología de Alimentos, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

<sup>c</sup>Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Ctra. CV-315, km. 10,7, 46113, Moncada, Valencia, Spain

<sup>d</sup>IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain

<sup>e</sup>Department of Analytical Chemistry, University of the Basque Country UPV/EHU, P.O. Box 644, 48080 Bilbao, Basque Country, Spain

<sup>f</sup>Departamento de Ingeniería Gráfica, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

<sup>g</sup>Department of Food Technology, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862, Campinas – São Paulo, Brazil

\* Corresponding author; tel: +55 19 3521 0070, E-mail: dfbarbin@unicamp.br

## **Abstract**

Pasta is mostly composed by wheat flour and water. Nevertheless, flour can be partially replaced by different components, as fibers, to provide extra nutrients in the human diet. However, fiber can affect the technological quality of pasta, and may not be properly distributed within the sample. Usually, determinations of important parameters of pasta are destructive and time-consuming. Therefore, the use of rapid and accurate technologies as Near Infrared-Hyperspectral Imaging (NIR-HSI), together with the state-of-the-art machine learning methods, is valuable to improve the efficiency in the assessment of the quality of pasta. The aim of this work was to investigate the ability of NIR-HSI and augmented Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) for the evaluation, resolution and quantification of fiber distribution in enriched pasta. Different strategies are tested in order to evaluate the feasibility of MCR-ALS to obtain information of pasta samples by using images of the raw ingredients (fiber and flour). The results showed percentage of lack of fit (%LOF) lower than 6%, variance explained over 99%, and similarity between the pure spectra and the spectra recovered by the augmented models above 96% and 98% in models using pure flour and control pasta as the initial estimates, respectively. The results found in this work demonstrate the applicability of NIR-HSI coupled to MCR-ALS in the identification of fiber added to pasta samples.

**Keywords:** Pasta; hyperspectral imaging; NIR; spectral unmixing; multivariate curve resolution.

## 1. Introduction

Pasta is a very popular staple food around the world. Its main ingredients are wheat flour and water, although flour can be partially replaced by different components as proteins, antioxidants and fibers in order to provide extra amount of nutrients in the human diet [1]. Recently, fiber has been added to pasta as a nutrient for a healthy diet [2–4]. However, the fortified product has to be in agreement with the food regulations. In case of fiber fortification, the Brazilian food legislation considers a product “source of fiber” if it has at least 2.5 g/100 g and “high fiber” if it has a minimum of 5 g/100 g of product [5]. In Europe, the food regulation considers a product “source of fiber” with 3 g/100 g portion of product and “high fiber” with minimum 6 g/100 g of product [6].

Additionally, the quality parameters of pasta, as cooking time, cooking losses, water absorption, texture, among others, depend on the interaction between starch and proteins during processing, and the addition of fiber or other ingredients in the formulation can affect the structure, taste, and physicochemical properties of the cooked pasta [7]. Moreover, the analysis of these properties is destructive and time-consuming. Pasta firmness and adhesiveness, for instance, are measured with a texture analyzer (TPA) after cooking. This technique is performed by a sequence of compressing and releasing operations on the sample, in order to simulate the chewing process. In the end of the analysis, the remaining sample is discarded [7]. Moreover, these methods do not normally provide any spatial information about the distribution of the elements in the surface of the pasta. Therefore, the use of faster and accurate technologies is required to improve the efficiency in the assessment of the abovementioned characteristics [8].

Vibrational spectroscopy techniques, as near infrared (NIR) spectroscopy, are an alternative for rapid and non-destructive measurements. These measurements are usually taken from one or few single points of the sample, and are considered to represent the entire sample. Nevertheless, the composition of food products is normally heterogeneous and many important information can be lost if the spatial distribution of the ingredients is not considered, affecting the accuracy and the representativeness of the collected data for decision-making during processing. Near infrared spectroscopy has a major

advantage of requiring several measurements for heterogeneous samples. In order to overcome this situation, near infrared spectroscopy can be integrated with hyperspectral imaging devices (HSI). NIR-HSI is an alternative for measurements of quality parameters in food products [9] considering the spatial distribution of the compounds in the measured surface of the product. NIR-HSI has already shown excellent results in qualitative and quantitative determinations of foodstuffs; however, only few works can be found in the application of NIR-HSI in pasta, as the characterization of important parameters of Italian pasta produced by traditional or industrial production parameters [10] or in the discrimination between different production technology approaches [11].

The great amount of information contained in the 3D cube makes the interpretation of the data a cumbersome task, requiring the association with advanced machine learning methodologies to extract relevant and useful information [9,12,13]. Machine learning (a.k.a. Multivariate analysis) compiles a number of algorithms with different purposes like spectral preprocessing, qualitative and quantitative data analysis, etc. [14,15]. When individual pixels are composed by a mixture of the signal of different compounds, the well-known spectral unmixing methodologies become a perfect choice in order to separate the weighted influence of the signal for each compound in the individual pixels [16,17]. Among them, Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) is one of the major algorithms used in hyperspectral image analysis in food [18,19], pharma [20], chemical compounds in biodegradable active films [21], ibuprofen-nicotinamide cocrystal formation [22].

MCR-ALS is a bilinear model that aims at decomposing the hyperspectral image in the spectral profiles and relative distributions for each compound in the sample [23,24]. The apparent simplicity of this model is hampered by the existence of ambiguities (rotational and intensity) that makes the model unstable towards a unique solution. Moreover, food is a complex matrix and NIR is not a very specific method. This promotes that 1) a “pure component” can be, indeed, a mixture of several components or 2) there is not enough spectral information of minor components in the sample (rank deficiency). To solve or, better said, to minimize these two issues, several strategies are normally followed. The first one is the imposition of what is known as natural constraints, being non-negativity

one of the most important ones [25]. Another strategy, if possible, is to augment the images (i.e. analyze several images together, multiset arrangement). Augmenting the image in study can be done with other images of the same nature containing extra information about the pure compounds [20]. Moreover, MCR-ALS has the versatility of adding spectral and spatial information to the original data, and single or multiset arrangement can be used for model development [20,26], in order to reduce the rank deficiency. In the food field, these techniques have been applied in the study of the distribution of different components in white and milk chocolate [18], and in the detection and quantification of adulterants in milk powder [19]. However, no studies regarding pasta, especially fiber-enriched pasta, have been reported to our knowledge.

The aim of this work is to investigate the ability of NIR-HSI together with MCR-ALS for the evaluation, resolution and quantification of fiber in enriched pasta. MCR-ALS models were developed to study the signal contributions and spatial distributions of fiber in the sample surface. Different strategies are tested in order to evaluate the feasibility of MCR-ALS in obtaining information in the pasta samples by using images of the raw ingredients (fiber and flour). Finally, the proposed methodology is an alternative for rapid evaluation and authentication of fiber-enriched pasta.

## **2. Materials and Methods**

### **2.1. Sample preparation**

Dry *fettuccine*-type pasta samples were produced in the Department of Food Technology at University of Campinas (Brazil) according to the methodology proposed by [3], replacing the flour by seven different types of fiber (Figure 2) in four percentages: 2%, 3.5%, 5% and 7% (w/w). For each percentage of fiber, 30 samples of *fettuccine* were produced, totaling 120 samples per type of fiber (Table 1). In addition, 30 samples without fibers were produced (control samples).

**Table 1.** Fiber properties.

<b>Fiber type</b>	<b>Source</b>	<b>Average particle size (<math>\mu\text{m}</math>)</b>	<b>Correspondent pasta sample</b>
<b>Fiber 1</b>	Mix of 20% soluble psyllium and 80% insoluble bamboo	160	Sample 1
<b>Fiber 2</b>	Insoluble bamboo	60	Sample 2
<b>Fiber 3</b>	Insoluble bamboo	145	Sample 3
<b>Fiber 4</b>	Mix of 80% soluble psyllium and 20% insoluble cellulose	160	Sample 4
<b>Fiber 5</b>	Mix of 50% soluble psyllium and 50% insoluble cellulose	160	Sample 5
<b>Fiber 6</b>	Insoluble wheat	60	Sample 6
<b>Fiber 7</b>	Insoluble wheat	145	Sample 7

## **2.2. Image acquisition**

Images of samples were acquired using a SisuCHEMA SWIR hyperspectral camera (Specim Spectral Imaging Ltd, Finland). The system is composed of a high speed push-broom camera that operates in the spectral range of 928-2524 nm with a spectral resolution of 6.3 nm and 320 space channels (156  $\mu\text{m}$  pixel size). The camera is equipped with a tungsten-halogen source and two-dimensional array detectors with 256 wavelength channels. The spectra were acquired with an exposure time of 2.1 ms using a 50 mm lens and a scanning speed of 15.8 mm/s. The measurement was controlled by Evince software (UmBio AB, Sweden). The software automatically subtracted the white (~99% reflectance measured with Spectralon) and dark (0% reflectance) references from subsequently acquired images. To optimize the process, one image included three samples although they were individually analyzed. Images from pure fiber and wheat flour were also taken.

## **2.3. Data analysis: MCR-ALS**

MCR-ALS assumes that the relative concentration and distribution of compounds in a sample of interest are determined according to the pure compounds, based on a bilinear model as showed in eq. 1.

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{eq. 1}$$

Where  $\mathbf{D}$  is the 2D matrix built by unfolding the original hypercube,  $\mathbf{C}$  is the concentration matrix provided by the algorithm,  $\mathbf{S}^T$  is the matrix containing the pure spectra information, and  $\mathbf{E}$  is the matrix expressing the error or variance unexplained by the bilinear model (Figure 1).

The original hypercubes are firstly unfolded into a matrix ( $\mathbf{D}$ ) of  $n$  dimensions (XY) by  $m$  wavelengths ( $\lambda$ ). After that, all matrices are joined and preprocessed with standard normal variate (SNV) prior to MCR-ALS analysis to correct the effects of light scattering.

The first step in MCR-ALS is to determine the number of components, the initial estimates and the correct constraints to be applied in the model development. In this work, two components were taking into account, since they were the major constituents in pasta samples. The initial estimates and the correct constraint are the key to achieve reliable results [27]. The data analysis was carried out considering two different types of initial estimates: (a) pure fiber and pure flour spectral information, and (b) pure fiber and control samples spectral information. The constraints used here are non-negativity for the concentration profiles and normalization of the spectral profiles so all components have the same relative importance in the model [28]. Moreover, closure was applied to the concentration profiles. This assumption can be done in this case, since 1) in NIR the absorbance for each pixel in each wavelength can be assumed to be the sum of the individual absorbance of each compound and 2) neglecting the minimal influence of artifacts, pasta is composed by two compounds. Moreover, equality constraints were also used in the concentration profiles to constraint specific values of the images with pure component to one (corresponding to a 100% relative concentration). These constraints help to reduce effects of rotational ambiguity and, consequently, increasing spectral

and spatial resolution, since they provide some more information about the  $S^T$  and  $C$  matrices [29]. Then, the ALS optimization was performed until the convergence criterion was satisfied (0.1%). A total of 14 augmented MCR-ALS models were developed (2 initial estimates x 7 types of samples). The percentage of lack of fit (%LOF) between the calculated and real results and the percentage of explained variance were considered to evaluate the model performance. Moreover, the histograms of the concentration map generated by the MCR-ALS models were considered to evaluate the concentration in the sample surface; together with the Pearson correlation coefficient between the pure and the recovered spectra of the constituents (similarity). Data analysis, was performed using Matlab software (R2016a, The Mathworks Inc., Natick, MA, USA) and the MCR-ALS 1.0 toolbox, freely available at [www.mcrals.wordpress.com](http://www.mcrals.wordpress.com) (last accessed March, 2020).

### **3. Results and Discussion**

#### **3.1. Control sample and pure flour spectra as initial estimates**

The mean NIR spectra of the control, flour and fiber samples are shown in Figure 3. The spectra of all fiber samples were strongly similar, with a slight variation on the baseline. Control samples had the highest absorbance values, followed by pure flour and, then, pure fiber samples. Additionally, it is possible to observe a peak around 1200 nm, which corresponds to stretching vibration of the C-H bond in the second overtone. Two absorption bands of stretching vibration of the O-H bond in the first overtone of water were also observed around 1450 and 1940 nm. Peaks of absorption around 1780 nm and 2100 nm were also highlighted, corresponding to the stretching vibration of the C-H bond in the first overtone of cellulose and a combination of O-H deformation and C-O stretching vibration, characteristic of starch. Peaks at 2242 and 2294 nm were also noticed, corresponding to N-H stretching vibration. Finally, a combination of C-H and C-C stretching vibration was observed at 2480 nm [30]. It is important to observe the high similarity between pasta/flour and fiber spectra. This will have a negative impact in further MCR-ALS models, making necessary the constraints and additional information mentioned before to obtain reliable models.

## 3.2. MCR-ALS in pasta samples

### 3.2.1. Fiber distribution

MCR-ALS models were developed for each type of sample separately, using the spectra of pure fiber, pure flour and control sample as initial estimates. As an example, the concentration profile of several samples studied in this work considering pure fiber and control sample as the initial estimate are shown in Figure 4. The concentration maps revealed how fibers behaved in each sample. Some fibers showed to be better distributed in the samples, as Fiber 1. However, some of them agglomerated in some parts of the sample, as Fibers 3, 4 and 7. This may be explained by the particle size, composition and/or percentage of fiber added to the samples. The particle size of the fiber shown in Table 1 is the average of different size meshes used for granulometry determination. However, Fibers 3 and 7, for instance, presented bigger particles as can be noticed on Figure 2. This aspect can be noticed in their respective concentration maps (Samples 3 and 7), where red spots (fiber) can be clearly identified.

In addition, the composition of each fiber was different. Fibers 2, 3, 6 and 7 are pure insoluble fibers, while 1, 4 and 5 are a mix of different soluble and insoluble fibers in a different ratio. However, even having soluble fibers in common, Fibers 1, 4 and 5 were differently distributed in their respective samples (Samples 1, 4 and 5). This may be explained by the amount of soluble fiber in each mix, which is 20%, 80% and 50% respectively. Soluble fibers generate a network around the granules of starch [31], so the higher amount of psyllium in Fiber 4 may have led to the agglomerates observed on Sample 4. A similar behavior was observed on the concentration map of Sample 5 (data not shown).

The histograms of samples considering the spectral information of control sample as the initial estimate are shown on Figure 5. As observed, most of the pixels in the samples had fiber concentration between 0 and 0.5. Apart from Sample 5, which had the most spread pixels values, all the samples showed a Gaussian distribution of the pixels. Moreover, Samples 3, 4 and 7 were slightly more disperse than Samples 1, 2 and 6. In addition, it is important to highlight that some samples presented

a large amount of pixels around 1, especially Samples 3 and 7, as seen in their respective concentration maps, indicating presence of pure fiber in some of the pixels.

### 3.2.2. Figures of merit: MCR-ALS validation

The results of %LOF and similarity are reported on Table 2. The LOF percentages of the MCR-ALS models for all samples were lower than 6%, regardless the initial estimate. Moreover, all the models were able to explain more than 99% of the total variance in each sample, suggesting a good performance of the models. The similarity between the pure spectra and the spectra recovered by the augmented models was above 96% and 98% in models using pure flour and control sample as the initial estimates, respectively. As an example, Figure 6 gives an overview of both pure and recovered spectra for Sample 3. These spectral profiles were highly correlated, with a higher deviation between both fiber spectra around 1450 nm, associated to water [30]. This may be associated to the fact that pure fiber used as initial estimate was in its powder form, and the prediction was performed on the pasta samples, which have passed through the drying process.

**Table 2.** MRC-ALS models performance.

<b>Flour as initial estimate</b>			
<b>Pasta Sample</b>	<b>Fiber spectral similarity</b>	<b>Flour spectral similarity</b>	<b>(%) LOF</b>
<b>Sample 1</b>	0.995±0.002	0.964±0.006	4.268±0.321
<b>Sample 2</b>	0.996±0.002	0.969±0.011	3.918±0.457
<b>Sample 3</b>	0.995±0.002	0.969±0.007	4.817±0.481
<b>Sample 4</b>	0.997±0.001	0.965±0.008	4.110±0.296
<b>Sample 5</b>	0.988±0.006	0.968±0.010	4.605±0.449
<b>Sample 6</b>	0.996±0.002	0.966±0.008	4.173±0.313
<b>Sample 7</b>	0.990±0.007	0.9692±0.006	5.159±0.433
<b>Control sample as initial estimate</b>			
<b>Pasta Sample</b>	<b>Fiber spectral similarity</b>	<b>Control sample spectral similarity</b>	<b>LOF</b>
<b>Sample 1</b>	0.994±0.003	0.999±0.001	3.453±0.297
<b>Sample 2</b>	0.995±0.002	0.999±0.001	3.189±0.321
<b>Sample 3</b>	0.995±0.002	0.999±0.001	4.211±0.530
<b>Sample 4</b>	0.996±0.002	0.999±0.001	3.303±0.257
<b>Sample 5</b>	0.986±0.007	0.998±0.001	4.027±0.423
<b>Sample 6</b>	0.995±0.002	0.999±0.001	3.404±0.291

<b>Sample 7</b>	0.988±0.009	0.999±0.002	4.648±0.369
-----------------	-------------	-------------	-------------

### 3.2.3. *Distribution of fiber content in the surface of samples*

The concentration maps resulted from MCR-ALS do not give the exact amount of fiber added to the samples since they provided the signal contribution of each constituent (fiber and flour/control sample) that can be more related to the volume fraction than to the mass fraction [20]. However, they provide the necessary information to establish a threshold and estimate the percentage of fiber in the sample surface. Establishing a proper threshold is not an easy task. In our case, we tested three different values (0.25, 0.30 and 0.40) that we found significant in the concentration maps of enriched-pasta samples (Figure 4) and the corresponding histograms (Figure 5). Pixels with values over the threshold were identified as fiber and those below the threshold was identified as flour/control sample. Figure 7 shows the concentration maps of Sample 3 and its respective images for each threshold. The lower is the threshold value, more pixels are identified as fiber; consequently, the higher is the predicted amount of fiber. Then, the amount of fiber was determined by the ratio in the sample of the area defined as fiber to the area defined as flour/control sample. This parameter was calculated for each sample in the dataset, and, then, averaged (Table 3). Moreover, this step provided a better overview of fiber distribution in the sample surface.

Bearing in mind that this is a semi-quantitative strategy, the results obtained had an acceptable quality in terms of quantitation. The fiber content predicted was plotted against the expected value and the coefficient of correlation ( $R^2$ ) between them are shown in Table 3. Figure 8 shows the fiber content predicted versus expected using fiber and control sample as initial estimates. The differences observed between the actual values and the values obtained by MCR-ALS are due to the fact that during sample production, the fiber characteristics and the particle size interfered in the complete dissolution of the powder in the mixture, being possible the presence of agglomerates of fiber, what may explain the lack of homogeneity in the sample surface. Moreover, from the amount of sample produced, only 30 samples of each type were randomly selected, which did not guarantee that in each

sample of pasta analyzed contained the exactly amount of fiber indicated. This, in fact, emphasizes the importance of this work, since the distribution of fiber within the sample is not homogeneous, which may affect the quality and properties of batch produced in the industry, that may not contain the minimum level necessary, while other parts of the batch may have a higher amount of fiber.

The models with the best performance according to each threshold for each sample are highlighted in dark grey. Overall, Samples 3, 5 and 7 showed the best correlation with the real content, unlike Samples 1, 2, 4 and 6. However, this may not indicate a bad performance of the models, but, as mentioned before, a problem of homogeneity of the sample. In addition, a great dispersion regarding the predicted fiber content among samples with the percentage of fiber, which may have decreased the performance of the models.

**Table 3.** Fiber content predicted.

Expected fiber content – Flour as initial estimate						
Sample	Threshold	R <sup>2</sup>	2%	3.5%	5%	7%
Pasta 1	0.25	0.02	0.3±0.4	0.3±0.4	0.2±0.3	0.6±0.7
	0.30	0.06	0.0±0.0	0.0±0.1	0.0±0.1	0.1±0.2
	0.40	0.04	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Pasta 2	0.25	0.14	0.3±0.4	0.1±0.2	0.1±0.2	1.5±1.6
	0.30	0.01	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.1
	0.40	0.04	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Pasta 3	0.25	0.84	0.4±0.6	6.9±1.9	13.5±3.3	18.6±4.0
	0.30	0.86	0.2±0.3	4.7±1.3	10.5±2.7	14.3±2.7
	0.40	0.83	0.0±0.1	2.5±0.6	6.5±1.9	8.9±2.1
Pasta 4	0.25	0.33	2.3±1.9	3.6±2.9	4.2±3.5	11.6±6.3
	0.30	0.22	1.0±1.3	1.3±1.5	1.5±1.7	5.2±4.3
	0.40	0.30	0.1±0.1	0.1±0.1	0.1±0.2	0.9±0.6
Pasta 5	0.25	0.68	2.3±1.9	6.6±4.2	29.7±9.0	32.2±9.3
	0.30	0.62	1.0±1.4	2.3±2.7	20.6±8.1	22.8±8.5
	0.40	0.50	0.1±0.3	0.3±1.0	6.5±3.5	9.3±6.4
Pasta 6	0.25	0.39	0.1±0.2	0.6±1.3	0.9±2.1	6.1±3.2
	0.30	0.22	0.1±0.1	0.1±0.2	0.3±0.8	1.4±1.6
	0.40	0.35	0.0±0.0	0.0±0.0	0.0±0.0	0.1±0.0
Pasta 7	0.25	0.64	5.1±2.2	5.8±1.2	7.4±1.5	15.0±1.2
	0.30	0.70	2.2±1.3	4.2±1.4	5.4±1.6	10.7±2.6
	0.40	0.78	0.6±0.3	2.1±0.5	2.9±0.7	6.1±1.7
Expected fiber content – Control sample as initial estimate						
Sample	Threshold	R <sup>2</sup>	2%	3.5%	5%	7%
Pasta 1	0.25	0.01	2.0±1.7	1.5±1.8	2.4±4.5	2.5±4.0
	0.30	0.06	0.2±0.2	0.1±0.1	0.1±0.1	0.4±0.4
	0.40	0.26	0.0±0.0	0.0±0.0	0.0±0.0	0.1±0.0
Pasta 2	0.25	0.33	1.7±2.1	1.1±2.0	1.0±2.4	35.9±26.9
	0.30	0.32	0.4±0.8	0.1±0.3	0.4±1.7	16.2±12.1

	0.40	0.13	0.0±0.0	0.0±0.0	0.0±0.1	0.1±0.2
<b>Pasta 3</b>	0.25	0.86	1.1±1.3	10.1±2.9	17.6±3.9	30.0±6.5
	0.30	0.90	0.4±0.9	6.1±1.7	12.7±3.0	21.2±3.4
	0.40	0.89	0.1±0.3	3.2±0.9	7.6±2.0	11.4±1.8
<b>Pasta 4</b>	0.25	0.34	5.5±3.6	8.3±4.7	12.0±8.3	26.1±17.5
	0.30	0.42	1.8±1.8	2.8±2.6	4.4±4.2	22.5±11.6
	0.40	0.33	0.2±0.4	0.4±0.8	0.5±0.8	5.3±3.7
<b>Pasta 5</b>	0.25	0.72	5.2±3.8	12.2±5.0	67.1±23.9	79.7±19.8
	0.30	0.74	2.1±2.9	5.0±3.7	59.0±23.1	75.5±17.9
	0.40	0.60	0.2±0.5	0.6±1.5	33.7±17.1	45.6±23.1
<b>Pasta 6</b>	0.25	0.23	1.6±3.0	1.9±3.2	2.8±4.1	11.3±11.1
	0.30	0.35	0.1±0.1	0.3±0.7	0.7±1.5	8.8±6.6
	0.40	0.18	0.0±0.0	0.0±0.0	0.0±0.0	0.2±0.2
<b>Pasta 7</b>	0.25	0.01	26.1±11.2	8.6±3.9	11.4±3.8	24.6±9.9
	0.30	0.29	10.8±4.8	5.0±1.1	6.8±1.4	16.3±5.8
	0.40	0.73	1.7±0.9	2.7±0.8	3.6±0.9	8.7±1.8

Results of each percentage are reported as averages of thirty samples±StDev.

#### 4. Conclusions

The results found in this work show the applicability of NIR-HSI coupled to MCR-ALS in the identification of fiber added to pasta samples. Homogeneity is a very important feature in the industry to guarantee the quality of the final product. In this work, the homogeneity of pasta samples showed to be a barrier in the quantification of the expected fiber content in the sample. However, the results showed low LOF, high total variance, and great similarity between pure and recovered spectral profiles, denoting good replicability of the models and ability of quantify the amount of fiber in the region of the sample analyzed.

This target could have been addressed by using other approaches (classification models, Classical Least Squares, etc.). Nevertheless, due to the complexity of the signals (high similarity of the fiber and flour/control sample spectra), we found that a step of resolution/unmixing was needed to extract semi-quantitative results. The outputs showed that, even using the raw ingredient spectral information as in pure fiber and flour, it is possible to develop reliable models to identify fiber in pasta. In addition, the concentration maps can be very useful to monitor the homogeneity of the samples. The performance of the models also showed that it is possible to have a qualitative overview of the fiber distribution in pasta samples and, in some cases, it is possible to quantify the amount of fiber in the

sample surface. Thus, this tool presented a great possibility to apply such technique as a qualitative and quantitative method for authentication of fiber-enriched pasta.

## **Acknowledgments**

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and São Paulo Research Foundation (FAPESP) (project numbers 2008/57808-1, 2014/50951-4, 2015/24351-2, 2017/17628-3, 2019/06842-0), and by GVA-IVIA and FEDER funds through project IVIA-51918. The authors would like to thank Nutrassim Food Ingredients company for the donation of the fiber samples, the support provided by Mrs. Cristiane Vidal during NIR-HSI system operation and data processing and Dr. Celio Pasquini for promptly receiving us in the laboratory that he coordinates (Grupo de instrumentação e automação em química analítica, Instituto de química, Universidade Estadual de Campinas, Campinas-SP, Brazil) to data acquisition.

## **Conflict of interest disclosure**

The authors declare no competing financial interest.

## **References**

- [1] P. Kowalczewski, G. Lewandowicz, A. Makowska, I. Knoll, W. B??aszczak, W. Bia??as, P. Kubiak, Pasta Fortified with Potato Juice: Structure, Quality, and Consumer Acceptance, *J. Food Sci.* 80 (2015) S1377–S1382. <https://doi.org/10.1111/1750-3841.12906>.
- [2] F. Muneer, E. Johansson, M.S. Hedenqvist, T.S. Plivelic, K.E. Markedal, I.L. Petersen, J.C. Sørensen, R. Kuktaite, The impact of newly produced protein and dietary fiber rich fractions of yellow pea (*Pisum sativum* L.) on the structure and mechanical properties of pasta-like sheets, *Food Res. Int.* 106 (2018) 607–618. <https://doi.org/10.1016/j.foodres.2018.01.020>.
- [3] A.R. Ferreira, Dry Fettuccine pasta with parcial replacement of *Triticum durum* semolina by

white fibers, University of Campinas, 2018.

- [4] N. Rakesh, C.M. Fellows, M. Sissons, Evaluation of the technological and sensory properties of durum wheat spaghetti enriched with different dietary fibres, *J. Sci. Food Agric.* 95 (2015) 2–11. <https://doi.org/10.1002/jsfa.6723>.
- [5] ANVISA, Resolução RDC N° 54, DE 12 DE NOVEMBRO DE 2012, (2012). [http://portal.anvisa.gov.br/documents/33880/2568070/rdc0054\\_12\\_11\\_2012.pdf/c5ac23fd-974e-4f2c-9fbc-48f7e0a31864](http://portal.anvisa.gov.br/documents/33880/2568070/rdc0054_12_11_2012.pdf/c5ac23fd-974e-4f2c-9fbc-48f7e0a31864).
- [6] European Commission, Nutrition Claims, (2012). [https://ec.europa.eu/food/safety/labelling\\_nutrition/claims/nutrition\\_claims\\_en](https://ec.europa.eu/food/safety/labelling_nutrition/claims/nutrition_claims_en).
- [7] M.C. Bustos, G.T. Perez, A.E. Leon, Structure and quality of pasta enriched with functional ingredients, *RSC Adv.* 5 (2015) 30780–30792. <https://doi.org/10.1039/c4ra11857j>.
- [8] H. Huang, L. Liu, M. Ngadi, Recent Developments in Hyperspectral Imaging for Assessment of Food Quality and Safety, *Sensors.* 14 (2014) 7248–7276. <https://doi.org/10.3390/s140407248>.
- [9] D. Wu, D.W. Sun, Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part I: Fundamentals, *Innov. Food Sci. Emerg. Technol.* 19 (2013) 1–14. <https://doi.org/10.1016/j.ifset.2013.04.014>.
- [10] P. Menesatti, F. Antonucci, F. Pallottino, F.M. Bucarelli, C. Costa, Spectrophotometric Qualification of Italian Pasta Produced by Traditional or Industrial Production Parameters, *Food Bioprocess Technol.* 7 (2014) 1364–1370. <https://doi.org/10.1007/s11947-013-1138-0>.
- [11] P. Menesatti, A. Bucarelli, Quality classification of Italian wheat durum spaghetti by means of different spectrophotometric techniques, in: *Opt. Nat. Resour. Agric. Foods II*, 2007: p. 67610D. <https://doi.org/10.1117/12.735553>.
- [12] E. Assadpour, S. Mahdi Jafari, Advanced Techniques for Hyperspectral Imaging in the Food Industry: Principles and Recent Applications, *Annu. Rev. Food Sci. Technol.* 10 (2019) 231. <https://doi.org/10.1146/annurev-food-032818>.

- [13] J.M. Amigo, *Hyperspectral Imaging*, 1st ed., Elsevier, 2020.
- [14] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images. Essential steps before image analysis, *Chemom. Intell. Lab. Syst.* 117 (2012) 138–148.  
<https://doi.org/10.1016/j.chemolab.2012.05.009>.
- [15] M.M.C. Ferreira, *Quimiometria- Conceitos, Métodos e Aplicações.*, Editora Unicamp, Campinas, São Paulo, Brazil., 2015.
- [16] A. de Juan, Multivariate curve resolution for hyperspectral image analysis, in: J.M. Amigo (Ed.), *Hyperspectral Imaging*, 1st ed., Elsevier, 2020: pp. 115–150.
- [17] J. Nascimento, G. Martín, Nonlinear spectral unmixing, in: J.M. Amigo (Ed.), *Hyperspectral Imaging*, 1st ed., Elsevier, 2020: pp. 151–166.
- [18] X. Zhang, A. de Juan, R. Tauler, Multivariate Curve Resolution Applied to Hyperspectral Imaging Analysis of Chocolate Samples, *Appl. Spectrosc.* 69 (2015) 993–1003.  
<https://doi.org/10.1366/14-07819>.
- [19] D.A.P. Forchetti, R.J. Poppi, Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder, *LWT - Food Sci. Technol.* 76 (2017) 337–343. <https://doi.org/10.1016/j.lwt.2016.06.046>.
- [20] F. Franch-Lage, J.M. Amigo, E. Skibsted, S. Maspoch, J. Coello, Fast assessment of the surface distribution of API and excipients in tablets using NIR-hyperspectral imaging, *Int. J. Pharm.* 411 (2011) 27–35. <https://doi.org/10.1016/j.ijpharm.2011.03.012>.
- [21] L.R. Terra, J. V. Roque, C.C. Pola, I.M. Gonçalves, N. de F.F. Soares, R.F. Teófilo, Study of chemical compound spatial distribution in biodegradable active films using NIR hyperspectral imaging and multivariate curve resolution, *J. Chemom.* 34 (2020) 1–15.  
<https://doi.org/10.1002/cem.3193>.
- [22] S. Ishihara, Y. Hattori, M. Otsuka, MCR-ALS analysis of IR spectroscopy and XRD for the investigation of ibuprofen - nicotinamide cocrystal formation, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 221 (2019) 117142. <https://doi.org/10.1016/j.saa.2019.117142>.

- [23] M. Ghaffari, S. Hugelier, L. Duponchel, H. Abdollahi, C. Ruckebusch, Effect of image processing constraints on the extent of rotational ambiguity in MCR-ALS of hyperspectral images, *Anal. Chim. Acta.* 1052 (2019) 27–36. <https://doi.org/10.1016/j.aca.2018.11.054>.
- [24] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, Spectroscopic imaging and chemometrics: A powerful combination for global and local sample analysis, *TrAC - Trends Anal. Chem.* 23 (2004) 70–79. [https://doi.org/10.1016/S0165-9936\(04\)00101-3](https://doi.org/10.1016/S0165-9936(04)00101-3).
- [25] J.M. Amigo, C. Ravn, Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging, *Eur. J. Pharm. Sci.* 37 (2009) 76–82. <https://doi.org/10.1016/j.ejps.2009.01.001>.
- [26] S. Piqueras, J. Burger, R. Tauler, A. de Juan, Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution, *Chemom. Intell. Lab. Syst.* 117 (2012) 169–182. <https://doi.org/10.1016/j.chemolab.2011.12.004>.
- [27] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, *Anal. Methods.* 6 (2014) 4964–4976. <https://doi.org/10.1039/c4ay00571f>.
- [28] X. Zhang, R. Tauler, Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging, *Anal. Chim. Acta.* 762 (2013) 25–38. <https://doi.org/10.1016/j.aca.2012.11.043>.
- [29] P. Firmani, S. Hugelier, F. Marini, C. Ruckebusch, MCR-ALS of hyperspectral images with spatio-spectral fuzzy clustering constraint, *Chemom. Intell. Lab. Syst.* 179 (2018) 85–91. <https://doi.org/10.1016/j.chemolab.2018.06.007>.
- [30] B.G. Osborn, T. Fearn, P.H. Hindle, Theory of near infrared spectroscopy, in: *Pract. NIR Spectrosc. with Appl. Food Beverage Anal.*, 2nd ed., Longman Singapore Publishers (Pte) Ltd: London, 1993: pp. 13–35.
- [31] C.S. Brennan, C.M. Tudorica, Fresh pasta quality as affected by enrichment of nonstarch polysaccharides, *J. Food Sci.* 72 (2007) S659–S665. <https://doi.org/10.1111/j.1750-3841.2007.00541.x>.

The Mathworks, Matlab. <https://mathworks.com/>, 2016.

## Captions

**Figure 1.** Schematic representation of a MCR-ALS considering two pure components.

**Figure 2.** RGB images of fiber used in samples preparation.

**Figure 3.** NIR spectra of the control, pure flour and pure fiber samples.

**Figure 4.** Concentration maps for four of the seven types of samples studied in this work. The color represents the percentage of fiber in each pixel of the image.

**Figure 5.** Histogram plots of samples. Different colors represent different fiber percentages.

**Figure 6.** Pure and recovered spectra from fiber and control sample (Data pre-processed with SNV).

**Figure 7.** Identification of fiber granules in Sample 3, comparing the 3 thresholds. A) Sample with 2% of fiber, B) Sample with 3.5% of fiber, C) Sample with 5% of fiber, and D) Sample with 7% of fiber.

**Figure 8.** Expected *versus* Predicted fiber content after MCR-ALS analysis (Models highlighted in dark grey on Table 3 for pure fiber and control sample as the initial estimate).