

Analysis of wheat flour-insect powder mixtures based on their near infrared spectra

Eszter Benes^a, Barbara Biró^b, Marietta Fodor^a, Attila Gere^{b,*}

^a Hungarian University of Agriculture and Life Sciences, Institute of Food Science and Technology, Department of Food and Analytical Chemistry, H-1118 Budapest, Villányi út, 29-43, Hungary

^b Hungarian University of Agriculture and Life Sciences, Institute of Food Science and Technology, Department of Postharvest, Supply Chain, Commerce and Sensory Science, H-1118 Budapest, Villányi út 29-43, Hungary

ARTICLE INFO

Keywords:

Entomophagy
Edible insect
Wheat flour
Near infrared spectroscopy
Quality control

ABSTRACT

Insects are gaining more and more space in food and feed sectors, creating an intense scientific interest towards insects as food ingredients. Several papers deal with cereal-based products complemented by insect powder in the past few years. However, adulteration and quality control of such products present some hot topics for researchers, e.g., how can we justify the amounts and/or species of the insects used in the given products? Our paper aims to answer such questions by analysing seven edible insect powders of different species independently. The mixtures with wheat flour were analysed using near infrared spectroscopy and chemometric methods. Not only powders of different species were clearly differentiated, but also mixtures created by different amounts of wheat flour. Prediction of insect content showed 0.65% cross-validated error. The proposed methodology gives an excellent tool for quality control of insect-based cereal food products.

Introduction

In the fields of food, nutrition and environmental sciences, edible insects and entomophagy have become a hot topic during the last decade. Several studies have shown that insects might have a place in the sustainable food supply chain either as feed or as food (van Huis, 2021). The edible species are consumable in every stage of their life cycle, and their nutritional composition is favourable, since most of them are valuable sources of high quality protein, vitamins B and E, iron, zinc and potassium (Kouřimská & Adámková, 2016). According to the existing literature, insect protein has a relatively high biological value, and bioavailability, the latter of which is greatly influenced by processing conditions (Ojha et al., 2021). The fat content of insect species is very diverse, in quantity and also in composition. Some insect larvae contain higher amounts of fats compared to the fully developed specimens, which mainly consist of triacylglycerols and unsaturated fatty acids, e.g. oleic acid, linoleic acid and linolenic acid. However, adults of some species, e.g. several cricket species, also contain higher levels of saturated fatty acids, e.g. palmitic acid and stearic acid (Kemsawasd et al., 2022; Tzompa-Sosa et al., 2014). Insect farming has lower environmental impact (e.g. lower greenhouse gas emission, water and land requirements) compared to other livestock's, furthermore, urban,

indoor, and vertical farming are also feasible methods to rear insects (Nikkhah et al., 2021).

Despite the nutritional and environmental benefits of edible insects, their consumption is not widespread in Western countries: consumers tend to reject insects and insect-based food products as this type of food is unfamiliar, and usually disgusts them (Gere et al., 2018; Ardoin & Prinyawiwatkul, 2021). Their spread in Europe is further hindered by the time-consuming nature of the authorization process. As edible insects are novel foods according to the Regulation (EU) 2015/2283, their placing on the market requires the inclusion in the Union list of authorised novel foods. The first species to be declared safe as food and feed was the yellow mealworm larvae (*Tenebrio molitor* L.), followed by the migratory locust (*Locusta migratoria* L.) as a novel food. These insects are now available in Europe frozen, dried and also in powdered form. According to the latest Scientific Opinion of EFSA on edible insects, house cricket (*Acheta domesticus* L.) is also safe and suitable for human consumption.

The scientific opinion on *Tenebrio molitor* was published by the European Food Safety Authority (EFSA) in January 2021, three years after the starting date of the Regulation's mandatory application (EFSA Panel on Nutrition, Novel Foods and Food Allergens (NDA), 2021). However, if such products have previously been legally placed on national

* Corresponding author.

E-mail addresses: gere.attila@uni-mate.hu, gereattilaphd@gmail.com (A. Gere).

markets, and a novel food application is submitted by 1st January 2019, insects, or insect-based products can continuously be placed on the markets. As a result, consumers in Belgium, Denmark or Finland may already encounter edible insects in the stores.

However, the legal difficulties have not hindered research and experimental product development with edible insects. Scientific publications on product development mostly present the enrichment of different cereal-based products with insect powder, e.g. breads (Kowalski et al., 2022), dry pasta, biscuits (Ayensu et al., 2019; Biró et al., 2020), extruded snacks and crackers (Azzollini et al., 2018; García-Segovia et al., 2020; Ardoin et al., 2021) as well as a bakery product made with oils extracted from insects (Cheseto et al., 2020). Meat products and meat analogues containing insects have also been developed (Scholliers et al., 2020; Smetana et al., 2018).

Due to the growing importance of this new protein source, food adulteration and the violation of the labelling requirements could become a real problem. Therefore, there is a growing demand for accurate and fast species-identifying and quantifying methods. Proteomic methods and PCR tests are currently used to detect the occurrence of insect material and identify the species (Francis et al., 2020; Kim et al., 2019).

Of the aforementioned measurement methods, near-infrared spectroscopy (NIRS) coupled with multivariate data analysis can provide a fast, cheap and non-destructive solution for accurate and fast species-identification and quantification. When using NIR spectroscopy, no or very simple sample preparation is required, and it is possible to analyze samples in their original form without the use of chemicals, which greatly reduces the length and cost of the measurements. The applicability of NIRS has been demonstrated by previous publications on the detection of various insect pests and insect fragments in different cereals such as wheat, rice and sorghum (Biancolillo et al., 2019; Johnson, 2020). NIRS has also been used successfully to analyse the protein content of insect-based energy bars (Beć et al., 2021). NIRS is usually applied with different classification and prediction methods to extract the relevant information from the obtained signals (Benes et al., 2020).

We expect to introduce a methodology which might be used to identify species and their amounts in flour mixtures in a fast, cheap, and easy way. Seven edible insect powders of different species and their mixtures with wheat flour were analysed using NIRS and chemometric tools with an aim to

- i) assess the applicability of NIRS to differentiate the species and their mixtures with wheat flour and
- ii) predict the insect content of the mixtures.

Materials and methods

Sample preparation

Seven edible insect powders of different species were analysed independently and their mixtures with wheat flour. Commercially available insect powder products were purchased from JR Unique Foods Ltd. (Udon Thani, Thailand) (*Acheta domesticus* – AD, *Bombyx mori* – BM, *Brachytripes portentosus* – BP, *Gryllus assimilis* – GA, *Gryllus bimaculatus* – GB, and *Locusta migratoria* – LM), from Kreca Ento-Food BV (Ermelo, The Netherlands) (*Tenebrio molitor* - TM). The developmental stage of the insects used to produce the different powders was pupa for BM, larva for TM and adult for the other five insects (AD, BP, GA, GB and LM). “Nagyititka” all-purpose wheat flour was purchased from GoodMills Magyarország Kft. (Komárom, Hungary). 70 samples were prepared by mixing wheat flour and the different insect powders individually with increasing ratio using five percentage steps from 5 to 50% to 100 g final weight.

Near infrared spectra acquisition

The spectra were obtained using a Bruker MPATM Multipurpose FT-NIR analyser (Bruker, Ettlingen, Germany) in five replicates, over the spectral range of 12500–3800 cm^{-1} . The spectrometer was equipped with an interferometer and a gold-coated integrating sphere. The samples were presented in a rotating quartz cuvette (\varnothing 85 mm) and were measured in diffuse reflectance mode with OPUS 7.2 (Bruker, Ettlingen, Germany) software using a lead sulphide (PbS) detector. Each spectrum was recorded as the average of 32 scans with a spectral resolution of 16 cm^{-1} (scanning speed 10 kHz). To reveal the most significant spectral differences, second derivation and vector normalisation (SNV) pre-treatment techniques were used on the average spectra of each sample. In case of SNV, the scattering is removed by normalizing each spectrum by the standard deviation of the responses across the entire spectral range (Moradi & Huang, 2008; Roberts et al., 2004). A relatively wide range of variability in the sample composition can be observed due to the different insects used in the mixtures. Therefore, the use of SNV is more expedient than multiplicative scatter correction (MSC) because correcting to the mean spectrum may give biased results. The evaluation range was between 9000 and 3800 cm^{-1} to avoid irrelevant data and noise. The assignment of FT-NIR spectra was done through comparison with the literature (Beć & Huck, 2019; Workman & Weyer, 2008).

Principal component analysis (PCA)

PCA was performed to reduce the dimensionality of the data set and also to determine spectral outliers by using F-residual and Hotelling- T^2 values. The spectral data set was randomized and mean centred before the analysis. Three segment cross-validation was used to validate the model (Diana & Tommasi, 2002). Data analysis (plotting, set up, etc.) was done using Unscrambler X 10.4 (CAMO Software, Oslo, Norway).

Linear discriminant analysis (LDA)

LDA was used to classify the mixtures according to the species of insects added to the wheat flour. PCA was applied to reduce the dimensionality the NIR spectral data and based on the eigenvalues and explained variances 20 dimensions (principal component scores) were used during LDA. The spectra were analysed after elimination of water bands, therefore the 9000–7313, 6557–5384, and 4968–3857 cm^{-1} ranges were used for classification purposes. LDA was run using Statsoft Statistica (version 10, Tulsa, OK, USA).

Hierarchical cluster analysis (HCA)

HCA was used to describe the similarities and differences between the NIR spectra of insect species. HCA is an unsupervised method, which does not require any *a priori* class memberships and clusters (or groups) the cases based on a chosen distance metric and agglomeration schedule. Different distance metric (Euclidean, squared Euclidean, Manhattan, Chebyshev etc.) and agglomeration schedule (single-, average-, complete linkage, Ward’s method etc.) combinations were tested and the best combination was chosen using the Silhouette clustering index. An average Silhouette index closer to 1 indicates that the objects are well matched to their own clusters (Rousseeuw, 1987). HCA was run on the NIR spectra of the seven species (100% powders) using Statsoft Statistica (version 10, Tulsa, OK, USA), while Silhouette indices were calculated using R-project (version 4.0.2) (R Core Team, 2020) and cluster package (version 2.1.0) (Maechler et al., 2021).

Partial least squares regression (PLSR)

The obtained NIR data (matrix X) and the measured amount of edible insect powders (matrix Y) were used to develop the prediction model for

the quantitative analysis. During PLS regression, the matrix X has been reduced to only a few factors (rank). The quality of the chemometric model depends on the choice of the correct number of factors needed. Leave-one-out cross-validation was used to determine the optimal statistical parameters of the models. The optimal rank numbers were selected through the root mean square error of the cross-validation (RMSECV). In addition, the performance of the PLSR models was evaluated by considering the coefficient of determination (R^2), the goodness of validation (Q^2), the root mean square error of estimation (RMSEE) and the root mean square error of cross-validation (RMSECV). The value of bias has also been taken into account, which is a systematic deviation of the measured (predicted) values from the true value due to the particular measurement method. The listed qualifying parameters were calculated based on well-known mathematical relationships (Takahama & Dillner, 2015), using OPUS 7.2 (Bruker, Ettlingen, Germany).

Results and discussion

Spectral properties of the original samples

Fig. 1 represent the SNV pre-treated spectra of the original samples. The differences between wheat flour and insect powder samples were clear due to the different nature of the raw materials. Thus, it is probable that the mixing process – especially if the insect powder is present in higher amounts – will affect the spectral data of samples greatly (Fig. 2). The insect powders have not been analysed independently before, but NIR spectroscopy has been used to determine insect contamination of cereals (Biancolillo et al., 2019; Johnson, 2020; Santos et al., 2019). Basically, the spectral properties of the analysed insect samples were similar (Fig. 1), but larger differences were observed in the 6000–5500 cm^{-1} and 5300–5000 cm^{-1} ranges. These variations may be caused by the different protein and fat content and composition of the insects. The spectra of the BM and TM samples are more distinct from the other insect samples, probably due to their different developmental stages. The species TM and BM belong to the orders Coleoptera and Lepidoptera, and the families Tenebrionidae and Bombycidae (ITIS, 2021), which

means that these insects are not so closely related to each other. Considerable variations were also observed for the LM sample, in the spectral range 6000–5500 cm^{-1} , where vibrational regions typically associated with lipids and proteins were found. The absorption peak at 5917 cm^{-1} , which may indicate the presence of aromatic amino acids (phenylalanine, tyrosine and tryptophan) was clearly present in the LM and AD samples, while in the other insect powders it appeared as a shoulder in the spectrum. The spectral characteristics of this region may be affected by the presence of sulphur-containing amino acids such as cysteine and methionine. The peak observed at 8580 cm^{-1} is related to the various unsaturated fatty acids present in the samples, the most abundant of which for insects are oleic acid, linolenic acid, linoleic acid. In addition, the characteristic peaks for aliphatic hydrocarbons (fatty acids) at 4335 cm^{-1} and 4261 cm^{-1} showed a relatively large intensity difference compared to the other samples. In the case of the 5300–5000 cm^{-1} region, the moisture content of the samples can have an impact on the shape of the peak. Examination of the TM spectrum showed that the peak at 5054 cm^{-1} , assigned to the N–H stretching (asymmetric) and N–H in-plane bending combination of CONH₂ groups of proteins, had almost completely disappeared. It is important to note that the spectral properties of the powders can also be affected by the way they are prepared (e.g. lyophilisation).

The spectral characteristics of wheat flour are well described by various scientific papers (De Girolamo et al., 2019; Mishra et al., 2020; Wadood et al., 2019; Zhao et al., 2014). Based on the second derivative spectrum, the absorption bands at 6977 cm^{-1} , 6301 cm^{-1} (first overtone of O–H stretch) and at 5189 cm^{-1} (O–H stretch + O–H deformation) were related to the moisture content. The bands at 8370 cm^{-1} (second overtone of C–H stretch), 7352 cm^{-1} ($2 \times$ C–H stretch + C–H deformation) and 5899 cm^{-1} and 5619 cm^{-1} (first overtone of C–H stretch) were related to lipids, whereas the absorption peaks at 6301 cm^{-1} (O–H stretch), 4758 cm^{-1} ($2 \cdot$ O–H deformation + $2 \cdot$ C–O stretch), 4390 cm^{-1} (O–H stretch + C–C stretch) and 4027 cm^{-1} (C–H stretch + C–C stretch) were associated with starch. Besides, the protein related peaks were at 4844 cm^{-1} (N–H symmetric stretch + amide II) and 4605 cm^{-1} (N–H/C–N/C=O combination band from secondary

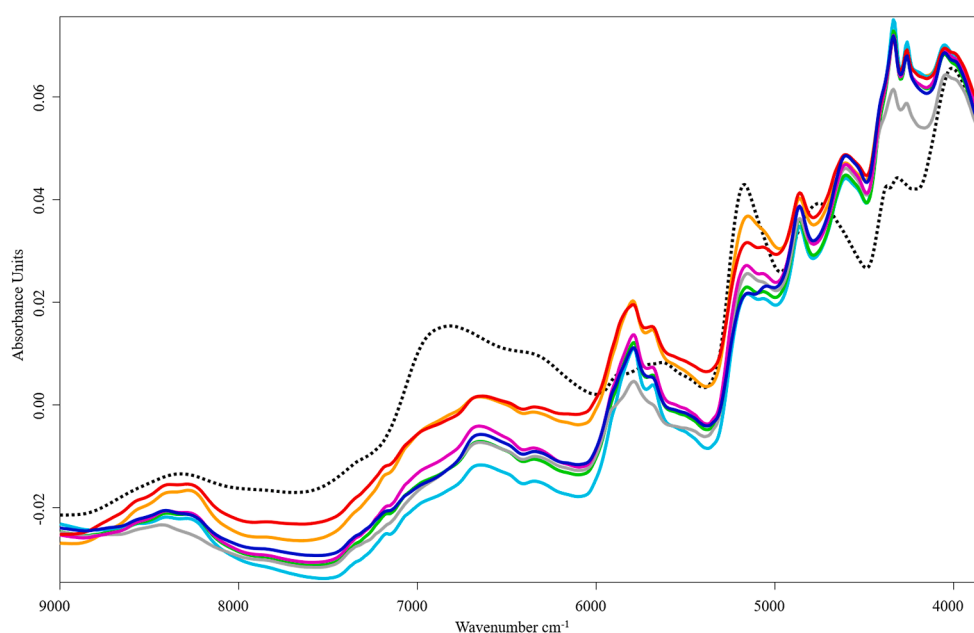


Fig. 1. Standard normal variate (SNV) spectra of wheat flour (black dotted line) and insect powders (● AD, ● BM, ● BP, ● GA, ● GB, ● LM, ● TM). AD – *Acheta domesticus*, BM – *Bombyx mori*, BP – *Brachytrupes portentosus*, GA – *Gryllus assimilis*, GB – *Gryllus bimaculatus*, LM – *Locusta migratoria*, TM – *Tenebrio molitor*.

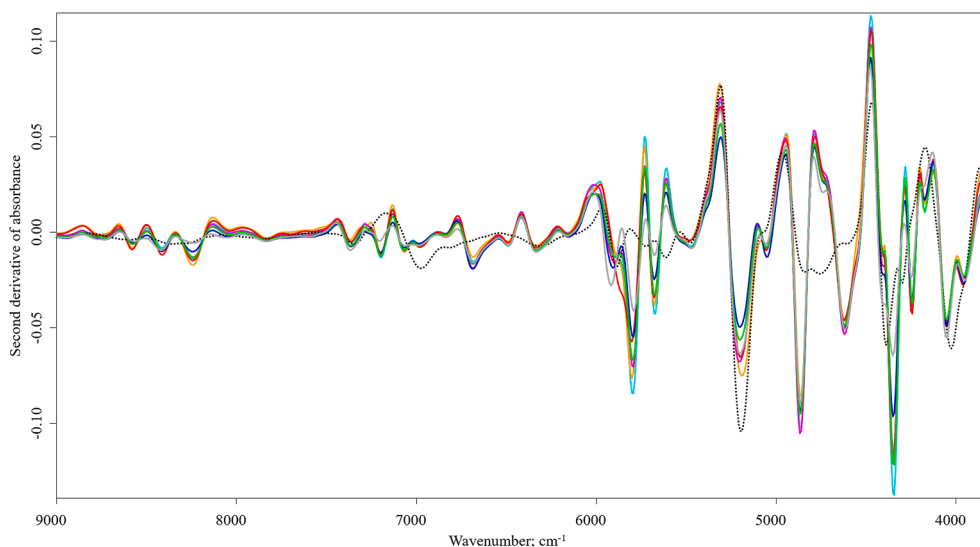


Fig. 2. Second derivative spectra of wheat flour (black dotted line) and insect powders (● AD, ● BM, ● BP, ● GA, ● GB, ● LM, ● TM). AD – *Acheta domesticus*, BM – *Bombyx mori*, BP – *Brachytrupes portentosus*, GA – *Gryllus assimilis*, GB – *Gryllus bimaculatus*, LM – *Locusta migratoria*, TM – *Tenebrio molitor*.

amides in proteins). Fig. 2 shows the second derivative spectra of all samples including the wheat flour samples (indicated with dotted line).

Due to the composition of the insect powders, which consist mainly of proteins, fats, and carbohydrates, absorption peaks associated with these compounds appeared (Table 1). Among these, mainly differences in intensity and smaller peak shifts can be observed. Second derivative and SNV pre-processing techniques were used to reveal the spectral variations among the samples. The spectral differences observed in the LM sample were even more prevalent in the second derivative spectrum. Several factors, such as the species-specific and taxonomic characteristics, developmental stages of the insects, the feeding protocol, etc., can affect the spectrum of powders made from them. The species GB, BP, GA and AD belong to the same taxonomic category, to Orthoptera order and Gryllidae family. The species *Locusta migratoria* also belongs to the Orthoptera order, but it is a member of the Acrididae family (Cigliano et al., 2021).

Effect of different types of insect powders on wheat flour mixtures

PCA provides a preliminary, primarily visual approach to find patterns in NIR spectra. PCA was performed on the raw spectral data to extract information on the major trends among the mixtures and the behaviour of the samples with the increasing amounts of insect powders (Fig. 3). The percentage ratio and the name of the mixed insect were used to indicate sample names, e.g. AD5 – 5% *Acheta domesticus* to 95% wheat flour. The first principal component (PC1) explained 89% of the spectral variation, and the second (PC2) accounted for 9% of the variance of the data. On the scores plot, the samples with lower insect powder content (5–20%) were located relatively close to each other, in the negative side of the scales of PC1 and PC2. As the spectral differences intensified by the addition of increasing quantities of insect powders, the different mixtures became more and more separated. While the spectra of the original insect powders were visibly different, the spectral properties of the flour were more dominant in the mixtures, especially at low mixing levels. Using PCA, even minor differences can be detected, thus providing the possibility to distinguish not only mixing levels but also different insect species from each other, so it can be used as a tool for the detection of food adulteration in the future applications.

According to the results of the scores plot, the different mixtures can also be distinguished from each other based on the insect species and the

levels of mixing. PC2 played a greater role in the discrimination of flour mixtures containing different insect species, while PC1 had a stronger effect on the separation of different levels of mixing. PC1 indicates the mixing levels for the TM and BM samples. Although the insect species were clearly separated from each other, similarities between the samples can be inferred from their proximity on the scores plot. The aforementioned TM and BM samples are on the negative scale of PC2 and mainly on the positive scale of PC1, relatively far away from the other samples. This can be explained by the spectral similarities, which are related to their developmental stage, which is different from other species. Their different taxonomic classification may also be a reason, as the composition of insects also differs on this basis (Rumpold & Schlüter, 2013). In contrast, the LM and AD samples were in the negative range for PC1, while for PC2 they were mostly in the positive scale. However, the separation between the species was clear as they were relatively far apart. The pattern showed that the BP, GA and GB samples tended to be in the positive range when both principal components were analysed. The Hotelling- T^2 test showed that the mixtures prepared come from the same sample population. The confidence interval was defined by the Hotelling- T^2 ellipse at 95% confidence level. Based on the sample residuals for the first two PCs, samples TM50 and GB50 were real spectral outliers. The LM45 and LM50 samples showed high F-residual values, while the BM50 sample showed high Hotelling- T^2 values. Due to the high ratio of insect powder, the spectral properties of these samples changed in such an extent that they become spectral outliers, therefore these were excluded from the development of the prediction models. The mixtures prepared with different types of insect powders could be analysed together, including species in different developmental stages, but the spectral variations also allowed species and mixing levels to be distinguished, making NIRS an excellent tool for detecting food adulteration in future applications.

Results of classification methods

The LDA results showed that the mixtures can be classified with 98.35% accuracy after performing leave-seven-out cross-validation based on the insect powders added to them. Fig. 4a clearly shows that there was an overlap between the BM and TM sample groups, supporting the PCA results. In addition, there was a minor overlap between the AD and GA samples. These could be explained by the fact that the insect

Table 1
The main NIR absorption bands of insect powders.

Wavenumber [cm ⁻¹]	Band assignment	Structure	Species
8580	C—H stretch 2nd overtone, HC = CH	lipid	BM, GA, GB, TM
8418	C—H stretch 2nd overtone, CH ₃	lipid	AD, BM, BP, GA, GB, LM, TM
8240	C—H stretch 2nd overtone, CH ₂	lipid	AD, BM, BP, GA, GB, TM
7200	2 × C—H stretch + C—H deformation, CH ₂	chitin	AD, BM, BP, GA, GB, LM, TM
6688	N—H stretch 1st overtone, NH	amide, protein	AD, BM, BP, GA, GB, LM, TM
6334	N—H 1st overtone, CONHR	protein	AD, BM, BP, GA, GB, LM, TM
5917	C—H Aromatic C—H (Aryl)	protein	AD, LM
5801/5793	C—H stretch 1st overtone, CH ₂	lipid	AD, BM, BP, GA, GB, LM, TM
5676	C—H stretch 1st overtone, CH ₂	lipid	AD, BM, BP, GA, GB, LM, TM
5480	O—H stretch + 2 × C—O stretch	chitin	AD, BM, BP, GA, GB, LM, TM
5198	O—H stretch + O—H deformation	water	AD, BM, BP, GA, GB, LM, TM
5054	N—H <i>asym.</i> stretch + amide II	protein	AD, BM, GA, GB, BP, LM
4869	N—H <i>asym.</i> stretch + amide II	protein	AD, BM, BP, GA, GB, LM, TM
4620	N—H <i>sym.</i> stretch + amide II	protein	AD, BM, BP, GA, GB, LM, TM
4351	2 × amide I + amide III	protein	TM
4345	C—H bending	lipid	AD, BM, GA, GB, BP, LM
4261	C—H Methylene C—H, associated with linear aliphatic R(CH ₂) _n R	lipids	AD, BM, BP, GA, GB, LM, TM
4247	3 X (C—H bending): C—H	polysaccharides, chitin	AD, BM, BP, GA, GB, LM, TM
4052	C—H combination	lipid	AD, BM, BP, GA, GB, LM, TM
3960	C—N—C asymmetric stretching	amide, protein	AD, BM, BP, GA, GB, LM, TM

AD – *Acheta domesticus*, BM – *Bombyx mori*, BP – *Brachytrupes portentosus*, GA – *Gryllus assimilis*, GB – *Gryllus bimaculatus*, LM – *Locusta migratoria*, TM – *Tenebrio molitor*.

powders added to the flour similarly altered the shape of the characteristic peaks, e.g., for AD, GA and LM samples the peak in the 5000–4500 cm⁻¹ region was more significantly distorted by the added powders (Supplementary material). By combining PCA and LDA, mixtures could be clearly distinguished by the insect species added to the wheat flour using only their NIR spectra. The results of the LDA model were validated by randomly assigned cases during which cases were randomly assigned to the original spectra. The mixed image of LDA with randomly assigned cases proved that pattern recognition was successful, and the method can be applied to identify the species of insects in mixtures.

Different combinations of distance metrics and agglomeration schedules were tested and based on their Silhouette indices, squared

Euclidean distance and Ward's method with two clusters received the highest index value (0.873). As HCA (Fig. 4b) was run on the NIR spectra of the seven species (100% powders), the similarities and differences between the powders can be seen. One of the clusters consists of TM and BM, a larva, and a pupa, while the other five species, which have been processed in adult forms, were clustered into one larger group. The agglomeration schedule presents that there are shorter linkage distances within this second cluster. AD and GA are located close to each other, then comes another cricket species (BP). After the three cricket species comes the only locust species (LM) and interestingly GB, a species belonging to the Gryllidae family (similarly to AD, GA, and BP) is linked as last. Nevertheless, it should be noted that since taxonomy classifies groups of organisms according to morphological and functional characteristics, molecular genetic specificities and evolutionary factors, cluster analysis based on the examined insects' NIR spectra presumably does not separate them by their taxonomic classification, but by the chemical structure of their composition.

PLSR prediction models

After removing outliers, PLSR regression was performed on a smaller data set (n = 65) to predict the percentage amount of insect powders in the different mixtures. Two different spectral ranges were used throughout the optimization process: 9000–3800 cm⁻¹ (full range model with 676 variables) and 9000–7313, 6557–5384, and 4968–3857 cm⁻¹ (elimination of water bands). After the removal of regions affected by moisture content, an automatic optimization process was also performed using OPUS 7.2 software (Bruker, Ettlingen, Germany) to decrease the value of RMSECV. The already mentioned spectra pre-processing methods were also tested and analysed. The most important statistical parameters of PLSR models are shown in Table 2. The higher the value of Q² and RPD and the lower the value of RMSECV, the better the performance of the model.

Even the regression on raw spectral data covering the whole range (9000–3800 cm⁻¹) resulted in a high coefficient of determination, but the RMSECV was found to be 1.18%, which is relatively high for mixing levels adjusted for a 5% quantitative difference. Spectrum pre-treatment methods have been applied to reduce irrelevant information carried by the different physical properties of the mixtures and to increase the signal-to-noise ratio. In all cases, the coefficients of determination of the validated models (Q²) were greater than 0.99, which supports the applicability of the NIRS throughout the analysis of insect powder and flour mixtures. The best model parameters were obtained by SNV pre-treatment and by excluding absorption peaks related to the moisture content, followed by automatic optimization of the spectral ranges. Compared to the full spectral range model without data pre-processing, the RMSECV decreased to 0.65% and the value of Q² increased from 0.993 to 0.998. Although the statistical parameters in the latter case proved to be the best, the automatically selected ranges (339 variables) do not necessarily lead to the best results, as important spectral regions may be excluded. The PLSR model using spectra pre-treated by water peak excision and SNV transformation was found to be the most efficient, as the RMSECV was low at 0.78% and the number of variables used in the regression (518 variables). Their range better served the description of the spectral properties of the sample population. Among the samples used for calibration, mixtures with *Bombyx mori* were always among the excluded samples in large numbers. In addition, mixtures with *Tenebrio molitor* were also excluded in several cases, suggesting that it may be appropriate to separate mixtures of insects at different stages of development. It was concluded that NIR spectroscopy is suitable for estimating the amount of different insect powders mixed with wheat flour in the range of 5–50%. The determination and monitoring the mixing ratio could also become an important question in food control in the future, as ensuring the right ratio is important for the consistency of the finished product.

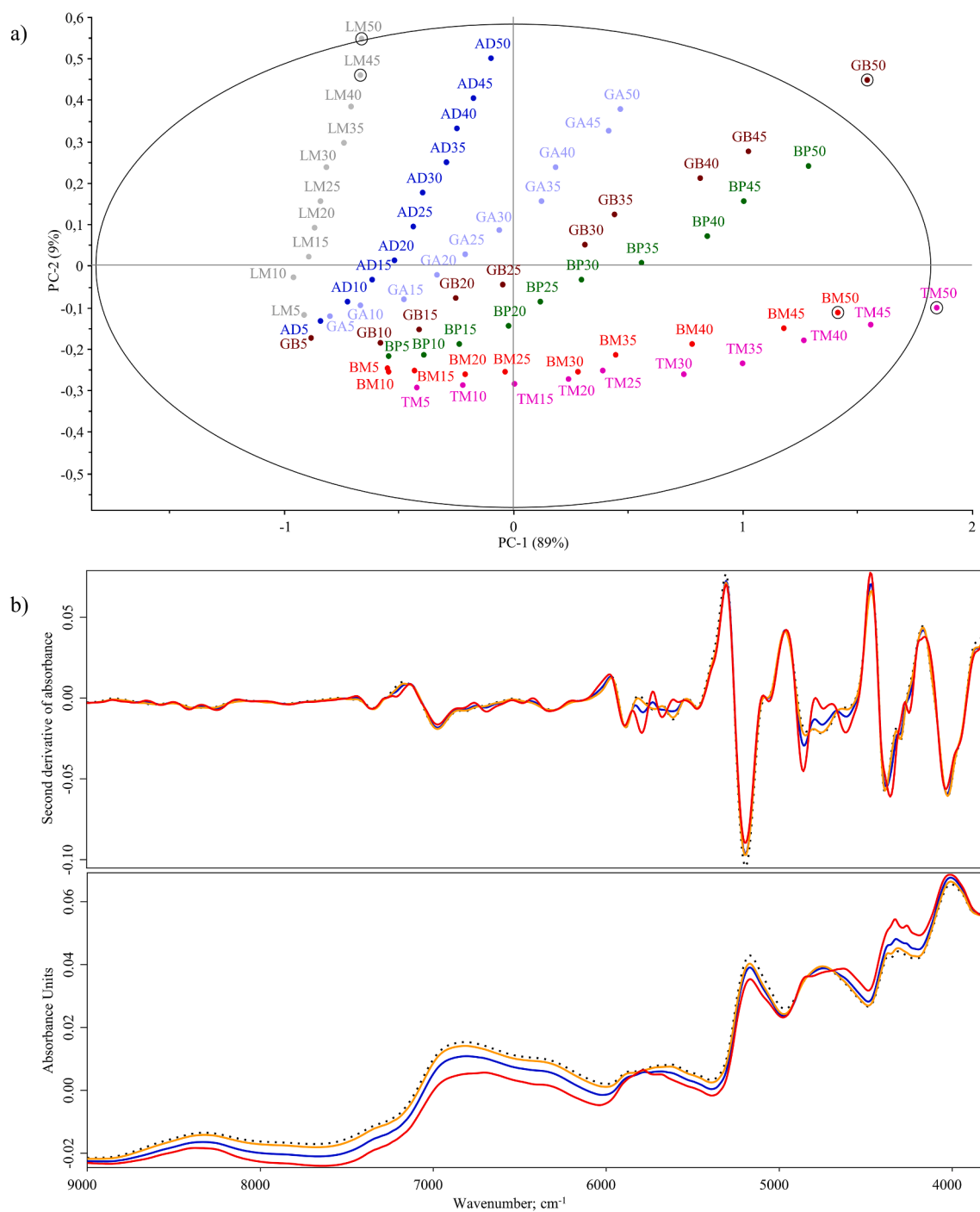


Fig. 3. a) Principal component analysis scores plot of the raw spectral data of the flour – insect powder mixtures (5 – 50%). b) Wheat flour (black dotted line) and its different mixtures with *Acheta domestica* (● 5%, ● 25% and ● 50%). AD – *Acheta domestica*, BM – *Bombyx mori*, BP – *Brachytrupes portentosus*, GA – *Gryllus assimilis*, GB – *Gryllus bimaculatus*, LM – *Locusta migratoria*, TM – *Tenebrio molitor*.

Conclusions

Near infrared spectroscopy coupled with chemometric tools has proven its efficiency in many fields of applications. Insects are gaining more and more space in food and feed sectors due to the falling of legal barriers and to the increasing amount of information on their chemical, microbial, nutritional, and environmental benefits. The presented paper is the first addressing the question of “How can we justify the amounts and/or species of insects used in food products?”. The wide range of analysed species is also unique, as the results of seven species are presented, each mixed with wheat flour using five percentage steps from 5 to 50%

insect content.

The fast and accurate differentiation of powders of different species and their different mixtures with wheat flour suggests that the presented methodology can be used by producers and food safety authorities as well. However, the presented results should not be generalized at this stage. As flours are mixed in many cases, the effect of flour mixtures (e.g., wheat flour mixed with buckwheat flour) should also be addressed later. Further research is needed if such a good differentiation can be done after different technological steps (e.g., fermentation, baking etc.) and with other food types, such as meat products or energy bars. From the modelling point of view, smaller differences were observed regarding

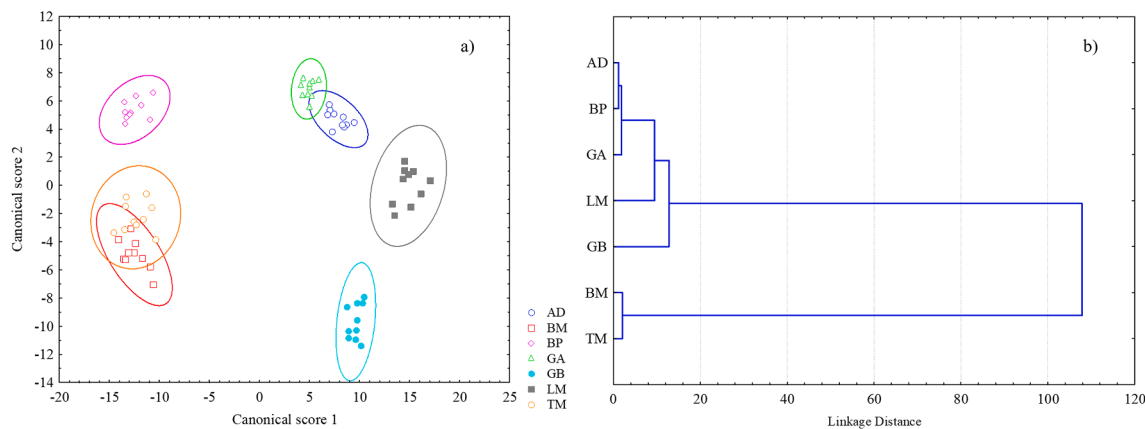


Fig. 4. a) Linear discriminant analysis plot of the flour – insect powder mixtures (5 – 50%) run on their principal component scores. (● AD, ● BM, ● BP, ● GA, ● GB, ● LM, ● TM). b) Dendrogram of agglomerative hierarchical cluster analysis using squared Euclidean distance and Ward's method. AD – *Acheta domesticus*, BM – *Bombyx mori*, BP – *Brachytrupes portentosus*, GA – *Gryllus assimilis*, GB – *Gryllus bimaculatus*, LM – *Locusta migratoria*, TM – *Tenebrio molitor*.

Table 2

Results of partial least squares regression (PLSR) models.

Pre-processing	Spectral range [cm ⁻¹]	R ²	RMSEE [%]	Q ²	RMSECV [%]	Bias	RPD	Rank	Calibration samples
no	9000–3800	0.995	1.05	0.993	1.18	−0.0083	11.6	7	57
SNV	9000–3800	0.997	0.79	0.996	0.89	0.0099	15.7	8	55
FD + SNV	9000–3800	0.997	0.85	0.996	0.97	−0.0183	14.3	8	57
no	9000–7313, 6557–5384, 4968–3857	0.997	0.83	0.995	0.97	0.0030	14.1	10	55
SNV	9000–7313, 6557–5384, 4968–3857	0.998	0.64	0.997	0.78	0.0234	18	10	57
FD + SNV	9000–7313, 6557–5384, 4968–3857	0.998	0.76	0.996	0.93	−0.0227	15	8	56
no	9000–8154, 4482–4266	0.998	0.60	0.996	0.77	−0.0178	16.4	10	54
SNV	9000–7313, 6557–5971, 4173–3857	0.999	0.52	0.998	0.65	−0.0342	21	10	57
FD + SNV	9000–7313, 4968–3857	0.999	0.58	0.997	0.72	−0.0085	18.9	10	59

FD – first derivative; SNV – standard normal variate; RPD – ratio performance deviation; The best prediction is indicated with bold.

the species as expected. From the international literature, we can see that there are high standard deviations between developmental stages and nutritional profiles, the latter can be even observed when the same species comes from different regions (Gere et al., 2019). For example, mixtures with *Bombyx mori* have been identified as outliers which might be caused by the significantly different feed. These indicate that it might be necessary to create different models for groups of species in order to cover a larger range of species and samples. It is assumed that by using a sufficiently large insect and flour sample variance and by increasing the sample size significantly, the model becomes generalizable.

CRedit authorship contribution statement

Eszter Benes: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Barbara Biró:** Conceptualization, Methodology, Investigation, Writing – original draft. **Marietta Fodor:** Methodology, Validation, Writing – review & editing, Supervision. **Attila Gere:** Conceptualization, Formal analysis, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

EB and BB thank the support of Doctoral School of Food Sciences, Hungarian University of Agriculture and Life Sciences. AG thanks the

support of János Bolyai Research Scholarship of the Hungarian Academy of Sciences and National Research, Development and Innovation Office of Hungary (OTKA, contract No. FK137577 and K134260). Supported by the ÚNKP-20-3-II-SZIE-15, ÚNKP-20-3-II-SZIE-23 and ÚNKP-21-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund. The Project is supported by the European Union and co-financed by the European Social Fund (grant agreement no. EFOP-3.6.3-VEKOP-16-2017-00005).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fochx.2022.100266>.

References

- Ayenu, J., Lutterodt, H., Annan, R. A., Edusei, A., & Loh, S. P. (2019). Nutritional composition and acceptability of biscuits fortified with palm weevil larvae (*Rhynchophorus phoenicis* Fabricius) and orange-fleshed sweet potato among pregnant women. *Food Science and Nutrition*, 7(5), 1807–1815. <https://doi.org/10.1002/fsn3.1024>
- Ardoin, R., Marx, B. D., Boeneke, C., & Prinyawiwatkul, W. (2021). Effects of cricket powder on selected physical properties and US consumer perceptions of whole-wheat snack crackers. *International Journal of Food Science and Technology*, 56, 4070–4080. <https://doi.org/10.1111/ijfs.15032>
- Ardoin, R., & Prinyawiwatkul, W. (2021). Consumer perceptions of insect consumption: A review of western research since 2015. *International Journal of Food Science and Technology*, 56, 4942–4958. <https://doi.org/10.1111/ijfs.15167>
- Azzollini, D., Derossi, A., Fogliano, V., Lakemond, C. M. M., & Severini, C. (2018). Effects of formulation and process conditions on microstructure, texture and digestibility of extruded insect-riched snacks. *Innovative Food Science and Emerging Technologies*, 45 (December 2017), 344–353. <https://doi.org/10.1016/j.ifset.2017.11.017>

- Beć, K. B., & Huck, C. W. (2019). Breakthrough potential in near-infrared spectroscopy: spectra simulation. A review of recent developments. *Frontiers Chemistry*, 7, 48. <https://doi.org/10.3389/fchem.2019.00048>
- Beć, K. B., Grabska, J., Plewka, N., & Huck, C. W. (2021). Insect protein content analysis in handcrafted fitness bars by NIR spectroscopy. gaussian process regression and data fusion for performance enhancement of miniaturized cost-effective consumer-grade sensors. *Molecules*, 26(21), 6390. <https://doi.org/10.3390/molecules26216390>
- Benes, E., Bajusz, D., Gere, A., Fodor, M., & Rácz, A. (2020). Comprehensive chemometric classification of snack products based on their near infrared spectra. *LWT*, 133(June). <https://doi.org/10.1016/j.lwt.2020.110130>
- Biancolillo, A., Firmani, P., Bucci, R., Magrì, A., & Marini, F. (2019). Determination of insect infestation on stored rice by near infrared (NIR) spectroscopy. *Microchemical Journal*, 145, 252–258. <https://doi.org/10.1016/j.microc.2018.10.049>
- Biró, B., Sipos, M. A., Kovács, A., Badak-Kerti, K., Pásztor-Huszár, K., & Gere, A. (2020). Cricket-enriched oat biscuit: Technological analysis and sensory evaluation. *Foods*, 9(11), 1561. <https://doi.org/10.3390/foods9111561>
- Cheseto, X., Baleba, S. B. S., Tanga, C. M., Kelemu, S., & Torto, B. (2020). Chemistry and sensory characterization of a bakery product prepared with oils from African edible insects. *Foods*, 9(6). <https://doi.org/10.3390/foods9060800>
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2021). *Orthoptera Species File*. <http://Orthoptera.SpeciesFile.org>. Version 5.0/5.0.
- De Girolamo, A., Cortese, M., Cervellieri, S., Lippolis, V., Pascale, M., Logrieco, A. F., & Suman, M. (2019). Tracing the geographical origin of durum wheat by FT-NIR spectroscopy. *Foods*, 8(10). <https://doi.org/10.3390/foods8100450>
- Diana, G., & Tommasi, C. (2002). Cross-validation methods in principal component analysis: A comparison. *Statistical Methods and Applications*, 11(1), 71–82. <https://doi.org/10.1007/BF02511446>
- Francis, F., Mazzucchelli, G., Baiwir, D., Debode, F., Berben, G., & Caparros Megido, R. (2020). Proteomics based approach for edible insect fingerprinting in novel food: Differential efficiency according to selected model species. *Food Control*, 112 (January). <https://doi.org/10.1016/j.foodcont.2020.107135>
- García-Segovia, P., Igual, M., Noguerol, A. T., & Martínez-Monzó, J. (2020). Use of insects and pea powder as alternative protein and mineral sources in extruded snacks. *European Food Research and Technology*, 246(4), 703–712. <https://doi.org/10.1007/s00217-020-03441-y>
- Gere, A., Radványi, D., & Héberger, K. (2019). Which insect species can best be proposed for human consumption? *Innovative Food Science & Emerging Technologies*, 52, 358–367. <https://doi.org/10.1016/j.ifset.2019.01.016>
- Gere, A., Zemel, R., Radványi, D., & Moskowit, H. (2018). Consumer response to insect foods. *Reference Module in Food Science*. Elsevier, 10.1016/B978-0-08-100596-5.21881-7.
- ITIS. (2021). *Integrated Taxonomic Information System*. <https://doi.org/10.5066/F7KHOKBK>
- Johnson, J. B. (2020). An overview of near-infrared spectroscopy (NIRS) for the detection of insect pests in stored grains. *Journal of Stored Products Research*, 86, Article 101558. <https://doi.org/10.1016/j.jspr.2019.101558>
- Kemsawasd, V., Inthachai, W., Suttisansanee, U., & Temviriyankul, P. (2022). Road to the red carpet of edible crickets through integration into the human food chain with biofunctions and sustainability: A review. *International Journal of Molecular Sciences*, 23(3), 1801. <https://doi.org/10.3390/ijms23031801>
- Kowalski, S., Mikulec, A., Mickowska, B., Skotnicka, M., & Mazurek, A. (2022). Wheat bread supplementation with various edible insect flours. In *Influence of chemical composition on nutritional and technological aspects*. *LWT* (p. 113220). <https://doi.org/10.1016/j.lwt.2022.113220>
- Kim, M. J., Kim, S. Y., Jung, S. K., Kim, M. Y., & Kim, H. Y. (2019). Development and validation of ultrafast PCR assays to detect six species of edible insects. *Food Control*, 103(December 2018), 21–26. <https://doi.org/10.1016/j.foodcont.2019.03.039>
- Kourimská, L., & Adámková, A. (2016). Nutritional and sensory quality of edible insects. *NFS Journal*, 4, 22–26. <https://doi.org/10.1016/j.nfs.2016.07.001>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2021). *cluster: Cluster Analysis Basics and Extensions*. <https://cran.r-project.org/package=cluster>.
- Mishra, P., Woltering, E., & El Harchioui, N. (2020). Improved prediction of 'Kent' mango firmness during ripening by near-infrared spectroscopy supported by interval partial least square regression. *Infrared Physics & Technology*, 110, Article 103459. <https://doi.org/10.1016/j.infrared.2020.103459>
- Moradi, B., & Huang, Y.-P. (2008). Objectification theory and psychology of women: a decade of advances and future directions. *Psychology of Women Quarterly*, 32(4), 377–398. <https://doi.org/10.1111/j.1471-6402.2008.00452.x>
- Nikkhah, A., Van Haute, S., Jovanovic, V., Jung, H., Dewulf, J., Cirkovic Velickovic, T., & Ghnimi, S. (2021). Life cycle assessment of edible insects (*Prototia brevitarsis seulensis* larvae) as a future protein and fat source. *Scientific Reports*, 11, 14030. <https://doi.org/10.1038/s41598-021-93284-8>
- Ojha, S., Bekhit, A. E., Grune, T., & Schlüter, O. K. (2021). Bioavailability of nutrients from edible insects. *Current Opinion in Food Science*, 41, 240–248. <https://doi.org/10.1016/j.cofs.2021.08.003>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>.
- Roberts, C. A., Workman, Jr., J., & Reeves, J. B. I. (2004). *Near-Infrared Spectroscopy in Agriculture* (Vol. 44). Wiley. <https://doi.org/10.2134/agronmonogr44>.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rumpold, B. A., & Schlüter, O. K. (2013). Nutritional composition and safety aspects of edible insects. *Molecular Nutrition and Food Research*, 57(5), 802–823. <https://doi.org/10.1002/mnfr.201200735>
- Santos, P. M., Simeone, M. L. F., Pimentel, M. A. G., & Sena, M. M. (2019). Non-destructive screening method for detecting the presence of insects in sorghum grains using near infrared spectroscopy and discriminant analysis. *Microchemical Journal*, 149, Article 104057. <https://doi.org/10.1016/j.microc.2019.104057>
- Scholliers, J., Steen, L., & Fraeye, I. (2020). Partial replacement of meat by superworm (*Zophobas morio* larvae) in cooked sausages: Effect of heating temperature and insect: Meat ratio on structure and physical stability. *Innovative Food Science and Emerging Technologies*, 66, Article 102535. <https://doi.org/10.1016/j.ifset.2020.102535>
- Smetana, S., Ashtari Larki, N., Pernutz, C., Franke, K., Bindrich, U., Toepfl, S., & Heinz, V. (2018). Structure design of insect-based meat analogs with high-moisture extrusion. *Journal of Food Engineering*, 229, 83–85. <https://doi.org/10.1016/j.jfoodeng.2017.06.035>
- Takahama, S., & Dillner, A. M. (2015). Model selection for partial least squares calibration and implications for analysis of atmospheric organic aerosol samples with mid-infrared spectroscopy. *Journal of Chemometrics*, 29, 659–668. <https://doi.org/10.1002/cem.2761>
- Tzompa-Sosa, D. A., Yi, L., van Valenberg, H. J. F., van Boekel, M. A. J. S., & Lakemond, C. M. M. (2014). Insect lipid profile: Aqueous versus organic solvent-based extraction methods. *Food Research International*, 62, 1087–1094. <https://doi.org/10.1016/j.foodres.2014.05.052>
- van Huis, A. (2021). Prospects of insects as food and feed. *Organic Agriculture*, 11, 301–308. <https://doi.org/10.1007/s13165-020-00290-7>
- Wadood, S. A., Guo, B., Zhang, X., & Wei, Y. (2019). Geographical origin discrimination of wheat kernel and white flour using near-infrared reflectance spectroscopy fingerprinting coupled with chemometrics. *International Journal of Food Science & Technology*, 54(6), 2045–2054. <https://doi.org/10.1111/ijfs.14105>
- Workman, J., & Weyer, L. (2008). *Practical guide to interpretive near-infrared spectroscopy*. CRC Press, 10.1201/9781420018318.
- Zhao, H., Guo, B., Wei, Y., & Zhang, B. (2014). Effects of grown origin, genotype, harvest year, and their interactions of wheat kernels on near infrared spectral fingerprints for geographical traceability. *Food Chemistry*, 152, 316–322. <https://doi.org/10.1016/j.foodchem.2013.11.122>