NIR spectroscopy and geostatistical analysis for modelling spatial distribution of analytical constituents in bulk animal by-product protein meals

J.A. Adame-Siles*[1], T. Fearn[2], J.E. Guerrero-Ginel[1], A. Garrido-Varo[1], F. Maroto-Molina[1], D. Pérez-Marín[1]

[1]*Department of Animal Production, Non-Destructive Spectral Sensor Unit, Faculty of Agricultural and Forestry Engineering, University of Córdoba, Agrifood Campus of International Excellence (ceiA3). Campus Rabanales, N-IV, km 396, Córdoba 14014, Spain.*
[2]*Department of Statistical Science, University College London. Gower Street, London WC1E 6BT, United Kingdom.*
**Corresponding author at: Campus Rabanales, N-IV, km 396, Córdoba 14014, Spain. Tel. +34 957 21 85 55. E-mail address:* g42adsij@uco.es

Abstract

The control and inspection operations within the context of safety and quality assessment of bulk foods and feeds are not only of particular importance, but are also demanding challenges, given the complexity of food/feed production systems and the variability of product properties. Existing methodologies have a variety of limitations, such as high costs of implementation per sample or shortcomings in early detection of potential threats for human/animal health or quality deviations. Therefore, new

proposals are required for the analysis of raw materials in situ in a more efficient and cost-effective manner. For this purpose, a pilot laboratory study was performed on a set of bulk lots of animal by-product protein meals to introduce and test an approach based on Near Infrared Spectroscopy (NIRS) and geostatistical analysis. Spectral data, provided by a fiber optic probe connected to a FT-NIR spectrometer, were used to predict moisture and crude protein content at each sampling point. Variographic analysis was carried out for spatial structure characterization, while ordinary kriging achieved continuous maps for those parameters. The results indicated that the methodology could be a first approximation to an approach that, properly complemented with the Theory of Sampling and supported by experimental validation in real-life conditions, would enhance efficiency and the decision making process regarding safety and adulteration issues.

Keywords: NIRS; Geostatistics; kriging; in situ analysis; real-time control; mapping of analytes.

Introduction

Quality and safety control of foods and feed before they enter the marketplace is a crucial target for all the stakeholders involved (manufacturers, regulatory bodies and agencies, business operators, organizations, etc.). Consequently, establishing proper surveillance plans and monitoring programs across all stages of the production chain becomes a key pillar for ensuring regulatory enforcement and compliance[1]. In a practical sense, guidelines and standards proposed by international organizations (e.g. ISO, FAO, CEN, ISTA) have tackled this goal[2]. However, good practice codes and quality assurance systems are also needed by the agro-food industry, where traceability and quality control of both raw materials and final product are determining factors.
All this poses many challenges of different nature and scope, for instance, logistical, operational and methodological problems. Of particular importance and difficulty is the upstream part of the supply chain, where complete assessment of incoming bulk raw materials is critical for their correct characterization. In this case, normal practice is to perform sampling and analysis as two separate phases. While the former is carried out in product reception or storage areas, the latter generally occurs in a laboratory environment. The control of bulk products has traditionally been governed, however, by an analytical view, which means that the role of sampling has often not been given enough attention. Indeed, sampling errors typically contribute most to the measurement uncertainty, amounting to 10-100 times the analytical errors[3].

The *Theory of Sampling* (TOS) provides definitions and fundamental principles, classified according to the spatial nature of the lot (e.g. 0-D –batch sampling, or 1-D lots –process sampling), for designing representative sampling processes and characterizing and tackling material heterogeneity[4]. The TOS stipulates a systematization of the total sampling errors (TSE) derived from every stage in a sampling procedure. For 0-D lots, they comprise a set of five errors, including both *correct* sampling errors or errors associated with the material alone (the Fundamental Sampling Error – FSE and the Grouping and Segregation Error –GSE) and *incorrect* sampling errors or errors associated with the sampling process (the Increment Delimitation Error – IDE, the Increment Extraction Error – IEE and the Increment Preparation Error – IPE)[5,6]. Within the TOS framework it is possible to control the systematic (bias) and the random parts of the sampling error, assuring an *accurate* (unbiased) and *reproducible* (precise) sampling process and, consequently obtain a *representative* sample[7].

TOS-compliant standards are already available and systematic approaches are outlined in the literature[8,9]. They address methods for eliminating or minimizing as many sampling errors as possible. Moreover, a set of criteria and specifications are described in order for practical sampling to be representative. TOS states the importance of compositing several *increments* from the lot, which demands a statistical selection process that must not compromise the representativity of the mass-reduced sample for analysis. Both selection and mass reduction processes are critical, and TOS establishes

techniques and methods regarding these two features[9,10]. Nevertheless, under real conditions, restrictions and nonstatistical issues (primarily, economical and practical considerations) dominate most of the current sampling standards and protocol designs, where sampling is also commonly considered as a simple process of material collection. The risks of failing to correctly select the final analytical aliquot are, however, decisive. If the increments extracted do not truly represent the original lot, decisions about the lot might be incorrect, regardless of the accuracy of the analytical methods used[11]. Alternative approaches to grab sampling and mass reduction methods are needed, and new schemes within the TOS context addressing the existing constraints of sampling and analysis operations are also essential. These should provide alternatives to overcome drawbacks such as the highly expensive and time-consuming implementation of sampling or limitations on the ability to increase the sampling volume. In addition, the existence of so many steps in the field-to-aliquot pathway and the fact that analytical results are achieved later in time and distant in space from the source (usually in the laboratory) are preventing current methods from providing in situ inspection, management and decision-making solutions. In view of the preceding, the control needs for evaluating bulk raw materials and feedstuffs more efficiently are not currently being met. Accordingly, the proposal and assessment of fast and reliable tools for this evaluation is imperative[12].

Extensive research over recent years on near infrared (NIR) spectroscopy has shown the potential of this technology for carrying out rapid and reliable analysis in a multitude of types of foods and feeds, including heterogeneous materials[13,14]. However, far fewer papers have addressed the NIRS analysis of bulk products with the aim of improving existing sampling plans, even though its integration into these plans as an analytical tool would bring multiple benefits. To begin with, it would enable qualitative and quantitative non-destructive analysis, dispensing with the need for extracting samples from the inspection batch. Near infrared spectroscopy is characterized by requiring only a few seconds to perform each measurement, which would enhance productivity, increase the sampling intensity (allowing a much closer approximation to the target decision unit), and encourage product evaluation in a more cost-effective way. These features contribute to making this technology ideal for providing in situ support to counter, for example, product misbranding, adulteration or safety issues, which are crucial to safeguarding the production chain from inappropriate or unsafe raw materials. Additionally, this methodology can produce a result at each measuring point, allowing this extra spatial information to be fully exploited by tools dealing with spatial patterns of variables. In this context, Geostatistics, a branch of statistics specializing in the analysis and interpretation of spatial continuous data, encompasses a set of useful techniques for recognising and modelling the spatial autocorrelation of the sampled variable. There are many published sources for case studies using geostatistics, most of

them in disciplines related to earth sciences, such as hydrogeology, soil science, geography, ecology or climatology[15–19], but applications to bulk food/feed products, can hardly be found. Nevertheless, geostatistics offers in this context the opportunity to go further in the development and optimization of procedures for estimating the spatial variation of properties of interest, using as inputs the existing observations, in order to make predictions at unsampled points. NIRS plus geostatistics leads to the possibility of mapping the spatial distribution of attributes under study, for an efficient evaluation of bulk lots of raw materials. This could be an important tool for planning sampling strategies and making real-time decisions regarding quality and safety issues.

The aim of this study was to analyse, on a laboratory scale, the spatial behaviour and variability of key analytical constituents (moisture and protein) in feed products, specifically in bulk lots of animal by-product protein (ABPs) meals. For that purpose, a methodology combining NIRS and geostatistical analysis is proposed.

Materials and Methods

Data generation

This pilot study intends to simulate in a laboratory environment the methodology to be implemented for the control of truckloads of bulk lots of ABPs. To this end, two

different case studies were carried out. The first study involved original lots coming from the rendering plant, while the second study assessed two types of adulteration.

Case study 1

Data were collected from a set of 15 batches of ABPs of diverse species composition (Table 1). Bulk lots were placed into a glass container (0.5 m in length, 0.35 m in width and 0.3 m in height) for analysis, where the top surface in each test reached a height of 0.15 m approximately. The measurements were performed in two horizontal planes, parallel to the top surface, at a depth of 0.04 (*layer A*) and 0.12 m (*layer B*). A methacrylate sheet, designed with a grid of 140 holes (10 x 14), was placed on the product surface to precisely position the probe for analysis at the insertion points (Figure 1a). At each point of the grid, the probe was first inserted to reach the layer A where a measurement was taken; then it was introduced deeper to obtain a measurement at layer B.

Table 1. Species composition percentages for each ABPs batch.

LOT	*Species Percentages*				*Reference Chemistry (%)*	
	Poultry	*Pig*	*Cattle*	*Sheep*	*Dry matter*	*Crude Protein (on dry matter basis)*
1	100				88.66	72.11
2	58	42			88.40	71.45
3	64	36			89.01	75.05
4	100				88.74	74.75
5	50	50			86.73	77.08
6	100				88.60	70.14
7	100				88.23	72.55
8	100				90.19	69.77

9	100				89.89	71.16
10	23	60	11	6	90.44	61.64
11	58	42			89.79	73.31
12	100				90.20	70.98
13	100				89.20	71.27
14	100				89.72	70.62
15	100				89.54	70.52

[insert Figure 1]

Case study 2

Different conditions were evaluated in ABPs batches through a new set of tests, in which two sort of adulterations were simulated. First, contamination by moisture was induced in two lots (**1** and **7**): 500 ml of water was poured one day prior to analysis over each corner of the glass container containing lot 7 (Figure 1b), while lot 1 was tested in the same way but using 100 ml of water in each corner and 50 ml in the middle (Figure 1c); as in study 1, layers A and B were measured in both cases. Two further tests addressed the problem of detecting unacceptable heterogeneity due to either poor mixing of products or an irregular composition in incoming bulk lots of ABPs. These consisted of making two different types of mixtures between lots **1** and **5.** In the first case, an accumulation of sample from lot 1 was located at one corner of the container (Figure 1d), while in the second test, two opposite corners were filled with sample from lot 1 (Figure 1e). In these tests measurements were taken only for layer A.

NIRS Analysis

A Matrix-F FT-NIR instrument (Bruker Optics, Germany) was used to measure reflectance spectra in ABPs lots (operating range 834.2-2502.4 nm). The equipment was interfaced to the *Turbido* reflection probe (Solvias AG, Switzerland) (Figure 1a), which consists of a stainless steel body with an insertable length of 300 mm and an outer diameter of 12 mm. The probe is configured with two optical fibers (600 µm core): one illumination fiber and one detection fiber. Two fiber cables (100 m in length) were used to connect the probe to the instrument. The probe end, which is angled at 20º, has a sapphire optical window illuminating a 1.5 mm diameter spot.

First, the noise level of the signal was assessed along the spectral range by applying to the log 1/R data a first derivative pre-treatment, with a single-unit gap and 5 data points smoothing. Visual evaluation found that spectra became noisy at the beginning and at the end of the spectral range, which led to the selection of the spectral range 1386-2033 nm.

In preliminary work a dataset of 346 samples of ABPs from different species (poultry by products meal, pork meal, cattle meal, meat and bone meal and mixture of different species) was analysed with the same instrument but using a detection head for contactless measurements (measurement area of 10 mm; working distance of 10 cm). Prediction equations were developed and different strategies were implemented for validation purposes, including an external validation from a totally independent set of 19 samples. The Modified Partial Least Squares (MPLS) regression method was used,

and four cross-validation groups were established[20]. Combined standard normal variate (SNV) plus Detrend treatments were used for scatter correction[21]. First- and second-derivative treatments were tested: 1.5.5.1, 1.10.5.1, 2.5.5.1 and 2.10.5.1, where the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth is the second smoothing[22]. Further details can be found in [23,24].

The following statistics were used to evaluate and select the best calibration model for each of the study parameters: standard error of calibration (SEC); standard error of cross-validation (SECV); R^2_C (coefficient of determination for calibration) and R^2_{CV} (coefficient of determination for cross validation); ratio of performance to deviation (RPD), i.e. the ratio of standard error of performance to standard deviation, and coefficient of variation (CV)[25].

Results for external validation were evaluated using the validation protocol recommended by [26] and [27] based on the following statistics: standard error of prediction (SEP), standard error of prediction corrected for bias (SEP(c)), bias and R^2_v (coefficient of determination for validation). This statistical process is based on the determination of a known significant error, termed “bias”, and an unexplained significant error, termed SEP(c) (standard error of prediction, bias-corrected). Generally, for calibration groups comprising 100 or more samples, and validation groups containing nine or more samples, the following control limits are assumed: Limit Control SEP(c) = 1.30 x SEC

(standard error of calibration); Limit Control bias = ± 0.60 x SEC (standard error of calibration), minimum 0.6 for R^2v.

On the basis of this preliminary work, the full data set was then transferred to the analysis mode used in this work (*Turbido* reflection probe). For transferring the database, the standardization methodology described in [28,29] was used. Later on, a recalibration procedure was performed for NIR analysis of the target lots by the probe. This calibration allowed a prediction to be obtained for each probe insertion point in the designed grid, where each spectrum was the result of 32 scans with a scanner velocity of 10 kHz and a resolution of 16 cm^{-1}. A set of ten raw NIR spectra collected with the probe from lot 1, displaying typical peaks characteristic of this type of products, can be seen in Figure 2. A measurement of a white reference was taken with a probe-specific spectralon every set of 42 measurements, which meant about every 25-30 minutes in time. In order to avoid cross contamination between measurements from different sampling points, compressed air cleaning was performed after every probe insertion.

[insert Figure 2]

Software OPUS v7.0 (Bruker Optik) was used for spectral acquisition and noise evaluation. WinISI v.1.50 (Infrasoft International, Port Matilda, PA, USA), MATLAB v. 7.8 (The MathWorks, Natick, USA) and PLS Toolbox (Eigenvector Research, Manson, USA) were used for calibration transfer and evaluation, and for obtaining NIR predictions.

Geostatistical analysis

The geostatistical analysis aimed at characterizing and representing the spatial pattern and distribution of moisture and protein content in ABPs lots of case studies 1 and 2. For this purpose, the geostatistical study was developed in two major stages: (a) structural analysis and (b) spatial estimation.

Structural analysis

The data set for each lot and case study was first assessed and prepared for geostatistical analysis. This assessment included the performance of normality tests and skewness and kurtosis calculation[30]. Having examined the distributions, the variographic analysis or spatial correlation analysis involved two steps: variogram estimation and its subsequent modelling.

The semivariogram (variogram is often used synonymously) plots the semivariance, $\gamma(\mathbf{h})$, against the distance between pairs of sample points (usually referred to as *lag* and denoted by **h**), and is defined as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h}) \right]^2 \quad (1)$$

where N(**h**) represents the number of pairs of points separated by the vector **h**, whereas z(**u**) is the variable under consideration depending on the location or vector of spatial coordinates **u**. Graphically, the result of the computation of semivariances via Equation 1 from a sequence of lag classes, varying in length within an interval and in orientation

up to a given tolerance on angle, is termed sample variogram or experimental variogram. This variogram measures the average dissimilarity between z values with respect to the lag classes, and it is a key tool supporting the structural analysis of the spatial continuity and variability of a regionalized variable[31].

The analysis of the spatial continuity typically begins with an omnidirectional variogram[32], for which the directional tolerance (α) is set at 180º, large enough that it could be thought as the average experimental variogram over all directions and in which only the magnitude of **h** is important, so the vector is replaced by the scalar in Equation 1. In this study, omnidirectional variograms were calculated for each case study and ABP lot based on the protein and moisture NIR predictions.

Once the omnidirectional variograms were computed, directional variograms were calculated to detect anisotropy (differences in the autocorrelation structure depending on direction). In this case, four directions were considered, defined by the progression 0, $\pi/4$, $\pi/2$, $3\pi/4$, i.e. 0, 45, 90, 135 degrees, with $\alpha=\pi/4$ (45º), in order to observe possible differences in ranges (*geometric anisotropy*), in scales (*zonal anisotropy*) or in shapes (indicating a trend)[32–34].

Regarding the features and interpretation of a variogram, a typical one shows that the larger the separation distance between observations, the higher the semivariance, or in other words closer points present more similar values than those further apart. This behaviour, however, may only persist up to a certain finite lag distance, beyond which

the variogram levels off. The distance at which the variogram stabilises is called the range (a), and it determines the limit beyond which z(**u**) and z(**u**+**h**) are uncorrelated or spatially independent. The semivariance value at the range is called the sill, which is the upper bound of the variogram and the a priori variance of the process. At the other end of the distance scale, almost every experimental semivariogram produced within the applied science domain is not strictly zero for 0 lag. This discontinuity at the origin of the variogram is called the nugget effect, which means that γ(**h**) does not tend to zero when **h** does. From a purely geostatistical point of view, the nugget effect is commonly assumed as a consequence of a short scale variability. In contrast, the Theory of Sampling has proposed a detailed description on the physical meaning of this effect[35]. The range, the sill and the nugget are the three most important parameters for describing a semivariogram[31–33].

In order to tackle spatial prediction, continuous functions have to be fitted to the experimental values provided by the sample variogram. The problem of estimating attribute values at unsampled locations needs a model of spatial dependence, which allows computing a variogram value at locations different from the existing data points. Therefore, modelling the experimental variogram is the most frequent approach to defining the pattern of spatial continuity. For this purpose, a set of permissible models obeying certain rules and constraints are widely used[31–33,36].

There are basically two types of variogram models: those that reach a plateau, or bounded, and those that do not, or unbounded. The spherical (Equation 2) and exponential (Equation 3) functions, the most common bounded models, and the linear model (Equation 4), as an unbounded function, were used for the fitting of the variogram.

$$\gamma(\mathbf{h}) = \begin{cases} 1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3 & \text{if } h \le a \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

$$\gamma(\mathbf{h}) = 1 - \exp\left(\frac{-3h}{a}\right) \quad (3)$$

$$\gamma(\mathbf{h}) = |\mathbf{h}| = h \quad (4)$$

A more thorough description and discussion on structural analysis, including variogram interpretation or assumptions, constraints and mathematics related to variogram modelling, can be found in[31–33,36].

Spatial estimation

A number of spatial interpolation mechanisms address the estimation of the value of continuous properties at unobserved sites within the area from which the observations originated. In this field, a family of generalized least-squares linear regression algorithms, called kriging[37], have traditionally been used in Geostatistics. Among all the existing interpolation techniques, kriging is characterized by being a highly-accurate and robust method, successfully overcoming the task of describing the relationships

between sample points and computing the estimations at unmeasured locations with reliable results[38].

The basic form of the linear regression estimator $Z^*(u)$ is defined as[36]:

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{i=1}^{n(\mathbf{u})} \lambda_i(\mathbf{u}) \left[Z(\mathbf{u}_i) - m(\mathbf{u}_i) \right] \quad (5)$$

where $\mathbf{u}$ and $\mathbf{u}_i$ are the location vectors for the estimation point and the neighbouring data points, indexed by i; $n(\mathbf{u})$ is the number of data points in a given local neighbourhood or window $W(\mathbf{u})$ centred on $\mathbf{u}$ and used for the estimation of $Z^*(\mathbf{u})$; $\lambda_i(\mathbf{u})$ are the weights, whose relative proportions mainly vary according to the positions of the sampling points and the values of the variogram functions, assigned to datum $z(\mathbf{u}_i)$ interpreted as a realization of the random variable $Z(\mathbf{u}_i)$; $m(\mathbf{u})$ and $m(\mathbf{u}_i)$ are the expected values of the random variables $Z(\mathbf{u})$ and $Z(\mathbf{u}_i)$.

The kriging estimator differs depending on the model adopted for the random function $Z(\mathbf{u})$ itself. $Z(\mathbf{u})$ is decomposed into a trend component, $m(\mathbf{u})$, and a residual component, $R(\mathbf{u}) = Z(\mathbf{u}) – m(\mathbf{u})$. The different kinds of kriging are distinguished according to the model considered for the trend $m(\mathbf{u})$. The methodology in this work is based on ordinary kriging (OK), which is the most robust method and one of the most common type of kriging in practice[33].

Ordinary kriging is frequently associated with the acronym BLUE for “best linear unbiased estimator”. It is “linear” as its estimates are weighted linear combinations of

the available data; it is “unbiased” since the residual component is modelled as a stationary random function with zero mean; it is “best” because it aims at minimising the error variance $\sigma_E^2(\mathbf{u})$. OK accounts for fluctuations of the mean over the entire domain, so limits its stationarity to the local neighbourhood W($\mathbf{u}$), centred on the location $\mathbf{u}$ being estimated, where the mean is deemed unknown. In this case, filtering the unknown local mean by forcing the kriging weights to sum to 1 leads to the OK estimator (Equation 6).

$$Z_{OK}^{*}(\mathbf{u}) = \sum_{i=1}^{n(\mathbf{u})} \lambda_i^{OK}(\mathbf{u}) Z(\mathbf{u}_i) \quad \text{with} \quad \sum_{i=1}^{n(\mathbf{u})} \lambda_i^{OK}(\mathbf{u}) = 1 \tag{6}$$

Further detailed information on spatial data processing and ordinary kriging theory and practice is described in[32,33,36,39,40]. This approach was implemented to form the moisture and crude protein predictions, and produce a continuous map for each parameter and each ABP lot in the previously mentioned case studies.

All geostatistical analyses were carried out in the R environment (version 3.2.1), including the exploratory data analysis, the variographic analysis and the mapping of spatial estimations. The R package *Gstat* was used to develop the methodology [41].

Results and discussion

NIRS calibrations

Results for external validation using the validation set (N=19) showed that relevant statistics for moisture and crude protein (bias, the R^2_V and the SEP(c)) were all within the limits recommended by [26,27] for external validation [24].

Once the results from the previous work were obtained, NIR calibration equations for predicting moisture and crude protein constituents in ABPs with the probe were developed using the dataset of 346 samples.

The calibration for moisture accounted for 76.9% of the variation existing in the set, while the SECV was 0.36% in this case (Figure 3a), equal to the one reported by the previous work[24]. The equation for crude protein presented a coefficient of determination of 86.4% and a SECV value of 2.45% (Figure 3b), similar to the one obtained in [24] (2.4%).

[insert Figure 3]

Taking into account the uncertainty for both critical process quality control parameters (moisture and crude protein) and the chemical variability encountered in ABPs, and also if all the relevant indicators are considered with caution when interpreting the goodness of fit of the model developed[42–44], the results reveal that the NIRS equations displayed a reasonable predictive ability for a process control application, thus confirming their validity for the purpose of the present paper.

Case study 1

Omnidirectional variograms, as a result of the structural analysis at layer A for the same batches involved in case study 2 (lots 1, 5 and 7), are reported in Figure 4. These variograms, with plots of semivariance against distance (mm), are shown for both constituents, crude protein and moisture, and for each ABP lot.

[insert Figure 4]

The experimental semivariograms for crude protein, calculated from NIR predictions at each point of the sampling grid, generally show a steady increase in the semivariance with lag distance. Therefore, linear models were used for the fitting of the variogram of this parameter. In order to fit the model to the experimental values, the R package *gstat* needs to perform iterations based on starting values for the parameters nugget, sill and range. Moreover, one of these three parameters needed to be fixed to use the linear model, so after a number of trials the range was fixed at a value of 180 mm. All protein tests displayed a slight slope with a discontinuity at the origin (Lot 1: 1.32; Lot 5: 2.09; Lot 7: 1.85) (Figure 4a, 4c, 4e), which according to geostatistics may reflect a short scale intrinsic variation[31–33]. However, the TOS provides a more comprehensive conceptualization and interpretation of the nugget effect for this type of application, as it is defined as the sum of all variances in the sampling procedure (correct and incorrect sampling errors) as well as the total analytical error (TAE)[35].

Therefore, both the sampling scheme and the signal acquisition errors when taking measurements with the probe are likely to have had an impact on the experimental

variograms in these cases, producing the discontinuities at the origin. This illustrates the intimate link that should be established between sampling (TOS) and process analytical technologies in order to build an effective and reliable analytical chain[10]. Further research should tackle that link so as to determine and minimise the real impact of these errors on the subsequent steps of the methodology proposed.

Concerning the sample variograms for moisture, the different behaviour that lot 1 shows in this case compared to lots 5 and 7 is noteworthy. The exploratory data analysis for lot 1 indicated a high-kurtosis distribution (data not shown), which could be the reason why the observed effect was an absence of variability in the data and a pure nugget variogram (Figure 4b), i.e. the nugget variance remained constant for all **h** [32,33]. On the other hand, Figure 4d and 4f illustrate semivariograms for moisture with a clear spatial autocorrelation, reaching a plateau in both cases. The theoretical function for the fitting of these bounded variograms was the spherical model, which resulted in nugget values of 0.023 (Lot 5) and 0.05 (Lot 7), and ranges of 145.78 (Lot 5) and 186.27 (Lot 7). These range values are consistent with the species composition of both lots, so the lag at which measurements become spatially independent when the lot consists of a single species (Lot 7) is higher than when the lot comes from a mixture of species (Lot 5).

All the directional variograms calculated in this case study showed that there were no differences in the autocorrelation structure with direction, reaching the same ranges and

sills as the omnidirectional variograms. Therefore, omnidirectional variograms were finally used in the spatial estimation stage.

Once the structural analysis was completed, ordinary kriging was performed from all data points of the grid, so that spatial estimations could be made between them throughout the whole area in each layer of measurements. The continuous surfaces obtained for the crude protein and moisture parameters are shown in Figure 5. A visual inspection of the maps allows the spatial behaviour of the constituents to be inferred in each case. The protein maps mostly display uniformity, as might be expected from the previous structural analysis. This is plausible, given the industrial manufacturing process of the tested products, which results in animal by-product protein meals with a high degree of homogeneity. Conversely, maps representing moisture content show a spatial pattern with more variability, except for lot 1, in which the surface becomes flat between the sampling points due to the pure nugget variogram[32,33]. The results of this case study motivated the performance of a second set of trials, in which the methodology could be evaluated to serve as a surveillance tool for prompt detection of risk zones corresponding to moisture contamination or other adulteration.

[insert Figure 5]

Case study 2

To evaluate the response of a geostatistical study for identifying contamination or adulterations in bulk lots of ABPs, a set of tests was carried out in the manner described

in the methodology (Case study 2). Figure 6 shows the omnidirectional variograms for all tests performed. In this case study, the anisotropy analysis did not reveal evidence of significant differences in the spatial autocorrelation with direction either, which may otherwise have indicated the existence of a trend or any kind of anisotropy. As a consequence, omnidirectional variograms were used again.

Measurements taken at layers A and B, both from contaminated lots 7 (Figure 6a and 6b, respectively) and 1 (Figure 6c and 6d), produced omnidirectional semivariograms with evidence of spatial continuity for the moisture parameter. After fitting spherical models to the sample variograms, the nugget parameters were close to zero in most cases, whereas the range at which the sill is reached varied from 142.95 to 255.16 mm.

[insert Figure 6]

The predictions, obtained using ordinary kriging in the same way as for case study 1, were mapped and the continuous surfaces are shown in Figure 7. If the distribution of water used for performing each test (Figure 1b and 1c) is taken into account, it can be seen how these distributions are accurately characterize by the maps. Figure 7a and 7b, associated with layer A and B of lot **7**, display high values of moisture at the four corners, precisely where the highest concentration of water took place. Furthermore, the maps representing the moisture distributions of lot 1 (Figure 7c and 7d) show how the deeper layer manages to detect and represent the small amount of water collected and located in the middle.

[insert Figure 7]

Finally, in order to evaluate how the geostatistical procedure responded to an irregular composition in lots of ABPs, which might be indicating product mislabelling or adulteration, two more tests based on an adulteration of the lot 5 with sample from lot 1 were carried out. Figure 6e shows the semivariogram from NIR predictions for the crude protein constituent and the first type of mixture (Figure 1d), while Figure 6f displays the omnidirectional variogram for the second sort of contamination (Figure 1e). Linear models were used again to fit a theoretical model to the sample variograms. They showed nugget values of 1.97 and 2.37 respectively. The steeper slope in these two cases compared with those in case study 1 can be interpreted as an indication of the higher level of variability found in case study 2.

The spatial distributions of protein are shown in Figure 7e and 7f. The map for the first mixture shows a different behaviour for the protein values in the upper right-hand corner, which correspond to the location of the lot 5, as Figure 1d indicates. Moreover, the surface representing the second type of adulteration allowed the inference of the original distribution since, as Figure 1e shows, the material from lot 1 was located both in the upper right-hand corner and the lower left-hand corner of the container. As a consequence, the geostatistical study achieved promising results here regarding rapid detection of changes in spatial patterns of the crude protein parameter in ABPs lots, as it also did with moisture.

Conclusions

The results described in this paper represent a first step towards defining a new method for in situ analysis and evaluation of bulk materials, i.e. directly at reception in the agro-food industry before the material enters the production chain. NIRS technology enables the implementation of methodologies for real-time analysis. In addition, it provides many more measurements than existing procedures at lower costs. These characteristics allow the Theory of Sampling (TOS) approach along with geostatistical techniques to be applied in this context to exploit the extra spatial information provided by NIR measurements. Together these provide a framework with significant potential within which fast assessment of spatial distributions of key properties in animal by-product protein meals could be made. This laboratory study suggests it would be worth carrying out further research with regard to the evaluation of the sampling and analytical errors derived from the implementation of this methodology as well as performing validation tests in real situations, for example, in trucks and railway wagons.

Acknowledgements

This research was carried out within the framework of the Research Excellence Program of the Andalusian Regional Government (Project No. AGR-6033) and Research Project INIA (Project No. RTA2012-00063-C02-02). The authors are grateful to RENDER GRASAS, S.L. (Spain) for providing batches of ABPs for this research, and Ms. María

del Carmen Fernández-Fernández and Mr. Antonio López-López for the technical assistance.

References

1. E. Commission. “Regulation (EC) No 882/2004 of the European Parliament and of the Council of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules”. Off. J. Eur. Union. 2004. L 165(April): 1–141.
2. ISO - 6497. Animal Feeding Stuffs - Sampling. International Organization for Standardization, Geneva, Switzerland, 2002.
3. L. Petersen, P. Minkkinen, K.H. Esbensen. “Representative sampling for reliable data analysis: Theory of Sampling”. Chemom. Intell. Lab. Syst. 2005. 77(1–2): 261–277. http://dx.doi.org/10.1016/j.chemolab.2004.09.013.
4. P. Gy. Sampling for Analytical Purposes. 1st ed. John Wiley & Sons, Chichester, UK, 1998.
5. P. Gy. “Sampling of discrete materials—a new introduction to the theory of sampling: I. Qualitative approach”. Chemom. Intell. Lab. Syst. 2004. 74(1): 7–24. http://dx.doi.org/10.1016/j.chemolab.2004.05.012.
6. P. Gy. “Sampling of discrete materials: II. Quantitative approach—sampling of zero-dimensional objects”. Chemom. Intell. Lab. Syst. 2004. 74(1): 25–38. http://dx.doi.org/10.1016/j.chemolab.2004.05.015.
7. F.F. Pitard. Pierre Gy’s Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness, and Statistical Process Control. 2nd ed. CRC Press, Boca Raton, FL, USA, 1993.
8. K.H. Esbensen. DS 3077. Representative Sampling - Horizontal Standard. Danish Standards, 2013.
9. L. Petersen, C.K. Dahl, K.H. Esbensen. “Representative mass reduction in sampling—a critical survey of techniques and hardware”. Chemom. Intell. Lab. Syst. 2004. 74(1): 95–114. 10.1016/j.chemolab.2004.03.020.

10. K.H. Esbensen, P. Mortensen. “Process Sampling (Theory of Sampling, TOS)-the Missing Link in Process Analytical Technology”. Process Analytical Technology. 2nd Editio. Wiley, 2010. Pp. 37–80.

11. C. Paoletti, K.H. Esbensen. “Distributional Assumptions in Food and Feed Commodities—Development of Fit-For-Purpose Sampling Protocols”. J. AOAC Int. 2015. 98(2): 295–300. 10.5740/jaoacint.14-250.

12. F. Cheli, A. Campagnoli, L. Pinotti, V.D. Orto. “Rapid Methods as Analytical Tools for Food and Feed Contaminant Evaluation: Methodological Implications for Mycotoxin Analysis in Cereals”. In: P.A. Aladjadjiyan, editor. Food Production - Approaches, Challenges and Tasks. InTech, 2012.

13. D.C. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero-Ginel, A. Gómez-Cabrera. “Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration”. Anim. Feed Sci. Technol. 2004. 116(3–4): 333–349. 10.1016/j.anifeedsci.2004.05.002.

14. E. Fernández-Ahumada, A. Garrido-Varo, J.E. Guerrero-Ginel. “Feasibility of diode-array instruments to carry near-infrared spectroscopy from laboratory to feed process control.” J. Agric. Food Chem. 2008. 56(9): 3185–92. 10.1021/jf073534t.

15. C. Piccini, A. Marchetti, R. Francaviglia. “Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment”. Ecol. Indic. Elsevier Ltd, 2014. 36: 301–314.

16. P. Goovaerts. “Geostatistics in soil science: state-of-the-art and perspectives”. Geoderma. 1999. 89(1–2): 1–45.

17. J.J.L.L. de Souza, W.A.P. Abrahão, J.W.V. de Mello, J. da Silva, L.M. da Costa, T.S. de Oliveira. “Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil”. Sci. Total Environ. 2015. 505: 338–349.

18. M. Lanfredi, R. Coppola, M. D’Emilio, V. Imbrenda, M. Macchiato, T. Simoniello. “A geostatistics-assisted approach to the deterministic approximation of climate data”. Environ. Model. Softw. Elsevier Ltd, 2015. 66: 69–77.

19. R. Alcaraz, E. Graniel, A.F. Castro, I. Vadillo. Assessment of groundwater contamination in Yucatan Peninsula (Mexico) by geostatistical analysis. B. Andreo, F. Carrasco, J.J. Durán, P. Jiménez, J.W. LaMoreaux, editors. Environmental Earth Sciences. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

20. H. Mark, J. Workman. Statistics in Spectroscopy. Academic Press, Inc, USA,

1991.

21. R.J. Barnes, M.S. Dhanoa, S.J. Lister. “Standard Normal Variate transformation and De-trending of near infrared diffuse reflectance spectra”. Appl. Spectrosc. 1989. 43: 772–777.

22. ISI. The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking. F. NIRSystems/TECATOR, editor. Infrasoft International, Silver Spring, MD, USA, 2000.

23. A. Sanchez-Bonilla, A. Garrido-Varo, F. Maroto-Molina, J.E. Guerrero-Ginel, B. De la Roza-Delgado, D.C. Pérez-Marín. Influence of the fiber optic length on the repeatability of the near infrared spectroscopy signal. 17th International Conference on Near Infrared Spectroscopy. https://proceedings.galoa.com.br/nir-abstracts/papers/, 2015.

24. A. Sanchez-Bonilla. Master Thesis. “Control de calidad on-line en el proceso de fabricación de subproductos de origen animal mediante tecnología FT-NIRS”. University of Córdoba, 2014.

25. P.C. Williams. Implementation of Near-Infrared Technology in the Agricultural and Food Industries. 2nd ed. American Association of Cereal Chemists, Inc., St. Paul, Minessota, USA, 2001.

26. W.R. Windham, D.R. Mertens, F.E. Barton II. “Protocol for NIRS calibration: sample selection and equation development and validation”. Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, nº643. USDA-ARS, US Government Printing Office, Washington D.C., 1989. Pp. 96–103.

27. J.S. Shenk, J.J. Workman, M.O. Westerhaus. Application of NIR Spectroscopy to Agricultural Products. Handbook of Near Infrared Analysis. D.A. Burns, E.W. Ciurczak, editors. Second Edi. Practical Spectroscopy Series, Vol. 27. Marcel Dekker, USA, 2001.

28. E. Zamora-Rojas, D. Pérez-Marín, E. De Pedro-Sanz, J.E. Guerrero-Ginel, A. Garrido-Varo. “Handheld NIRS analysis for routine meat quality control: Database transfer from at-line instruments”. Chemom. Intell. Lab. Syst. Elsevier B.V., 2012. 114: 30–35. 10.1016/j.chemolab.2012.02.001.

29. D. Perez-Marin, A. Garrido-Varo, J.E. Guerrero-Ginel. “Remote near infrared instrument cloning and transfer of calibrations to predict ingredient percentages in intact compound feedstuffs”. J. Near Infrared Spectrosc. 2006. 14(2): 81–91.

30. J.F. Kenney, E.S. Keeping. “Mathematics of Statistics, Pt. 1”. 3rd ed. Van Nostrand, Princeton, NJ, 1962. Pp. 102–103.

31. J.-P. Chiles, P. Delfiner. Geostatistics. Modeling Spatial Uncertainty. John Wiley & Sons, Inc., New Jersey, 2012.

32. E.H. Isaaks, R.M. Srivastava. An Introduction to Applied Geostatistics. Oxford University Press, New York, 1989.

33. R. Webster, M. Oliver. Geostatistics for Environmental Scientists. 2nd ed. John Wiley & Sons, Inc., Chichester, 2007.

34. E. Gringarten, C. V Deutsch. "Teacher's Aide Variogram Interpretation and Modeling 1". Math. Geol. 2001. 33(4).

35. K.H. Esbensen, H.H. Friis-Petersen, L. Petersen, J.B. Holm-Nielsen, P.P. Mortensen. "Representative process sampling — in practice: Variographic analysis and estimation of total sampling errors (TSE)". Chemom. Intell. Lab. Syst. 2007. 88(1): 41–59. http://dx.doi.org/10.1016/j.chemolab.2006.09.011.

36. P. Goovaerts. Geostatistics for Natural Resources Evaluation. Oxford University Press, Oxford, 1997.

37. D.G. Krige. A statistical approach to some mine valuations and allied problems at the Witwatersrand. University of Witwatersrand, 1951.

38. X. Liu, J. Hu, J. Ma. "Quantitative Evaluation of Spatial Interpolation Models Based on a Data-Independent Method". Advances in Data, Methods, Models and Their Applications in Geoscience. InTech, 2011. Pp. 53–70. 10.5772/1133.

39. N. Cressie. Statistics for Spatial Data. Wiley, New York (USA), 1991.

40. D.E. Myers. "Interpolation and estimation with spatially located data". Chemom. Intell. Lab. Syst. 1991. 11(3): 209–228. 10.1016/0169-7439(91)85001-6.

41. E.J. Pebesma. "Multivariable geostatistics in S: the gstat package". Comput. Geosci. 2004. 30(7): 683–691. 10.1016/j.cageo.2004.03.012.

42. A. Garrido-Varo, S. Vega, F. Maroto-Molina, M. José De la Haba, D. Pérez-Marín. "On-site quality control of processed land animal proteins using a portable micro-electro-mechanical-systems near infrared spectrometer". J. Near Infrared Spectrosc. 2016. 24(1): 47. 10.1255/jnirs.1192.

43. K. Esbensen, P. Geladi, A. Larsen. "The RPD myth…". NIR news. 2014. 25(5): 24. 10.1255/nirn.1462.

44. T. Fearn. "The overuse of R2". NIR news. 2014. 25(5): 2014. 10.1255/nirn.1464.

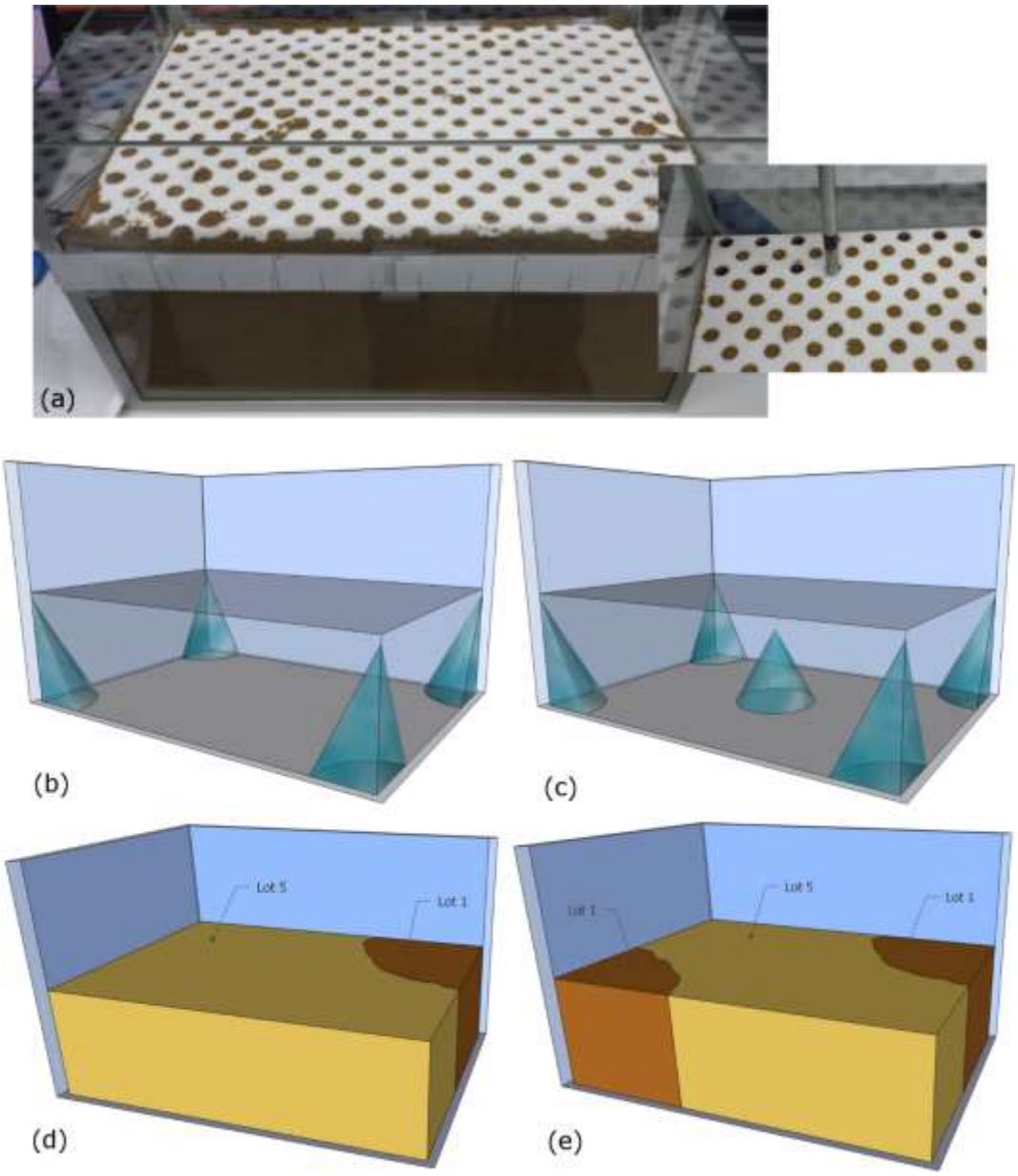

FIGURE 1

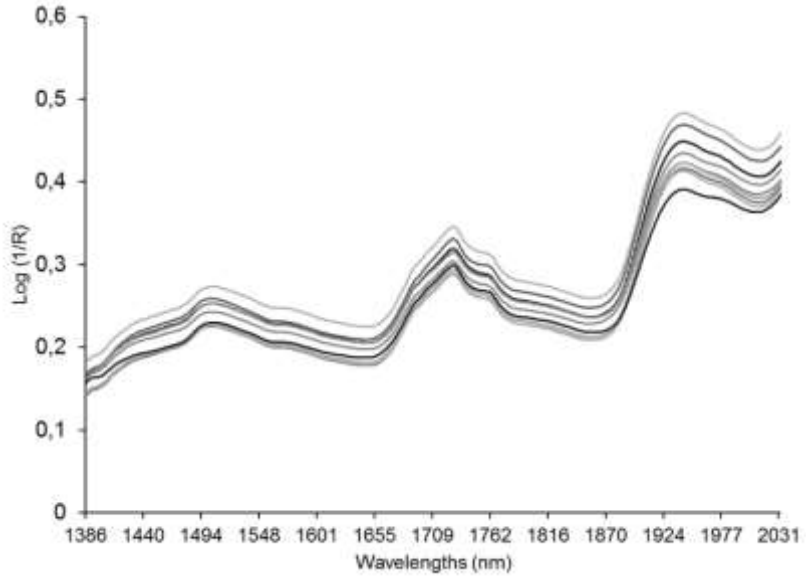

FIGURE 2

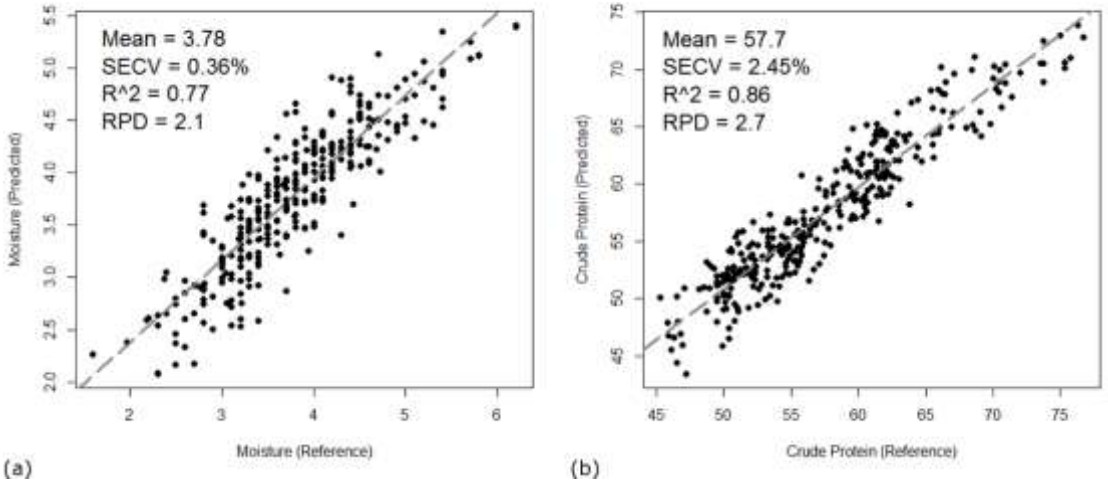

FIGURE 3

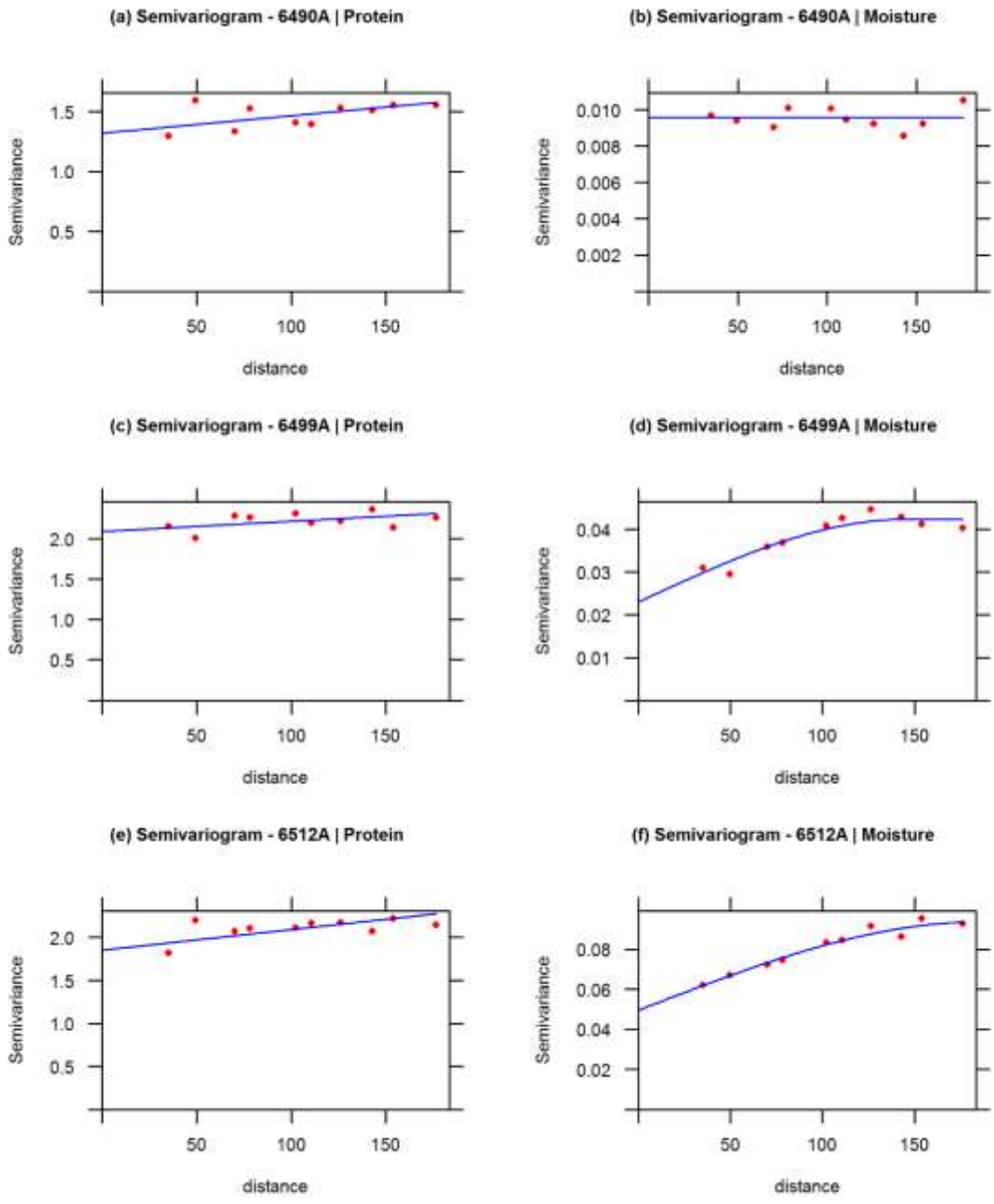
(a) Semivariogram - 6490A | Protein
Semivariance
1.5
1.0
0.5
50
100
150
distance
(b) Semivariogram - 6490A | Moisture
Semivariance
0.010
0.008
0.006
0.004
0.002
50
100
150
distance
(c) Semivariogram - 6499A | Protein
Semivariance
2.0
1.5
1.0
0.5
50
100
150
distance
(d) Semivariogram - 6499A | Moisture
Semivariance
0.04
0.03
0.02
0.01
50
100
150
distance
(e) Semivariogram - 6512A | Protein
Semivariance
2.0
1.5
1.0
0.5
50
100
150
distance
(f) Semivariogram - 6512A | Moisture
Semivariance
0.08
0.06
0.04
0.02
50
100
150
distance

FIGURE 4

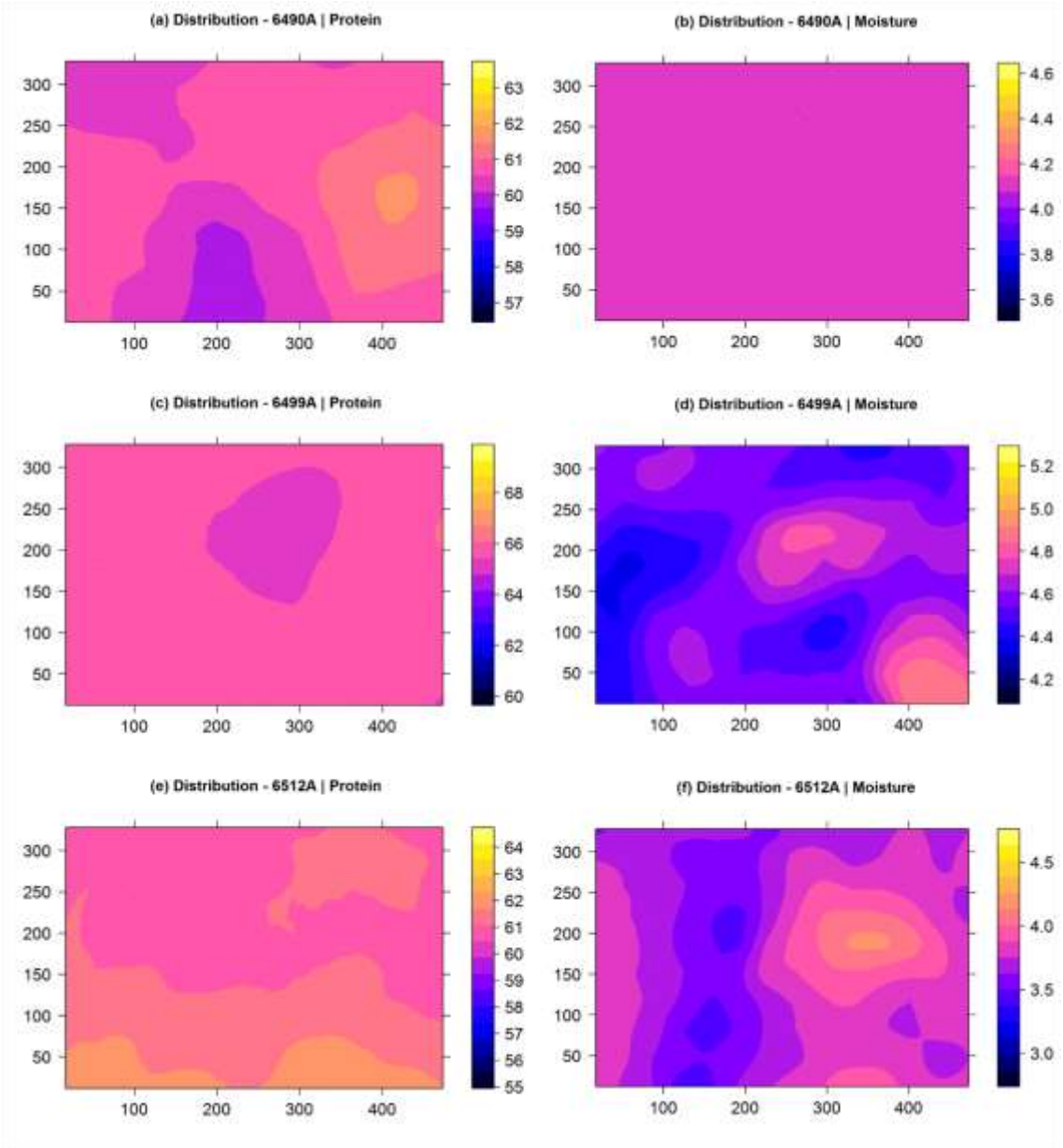
(a) Distribution - 6490A | Protein
(b) Distribution - 6490A | Moisture
(c) Distribution - 6499A | Protein
(d) Distribution - 6499A | Moisture
(e) Distribution - 6512A | Protein
(f) Distribution - 6512A | Moisture

FIGURE 5

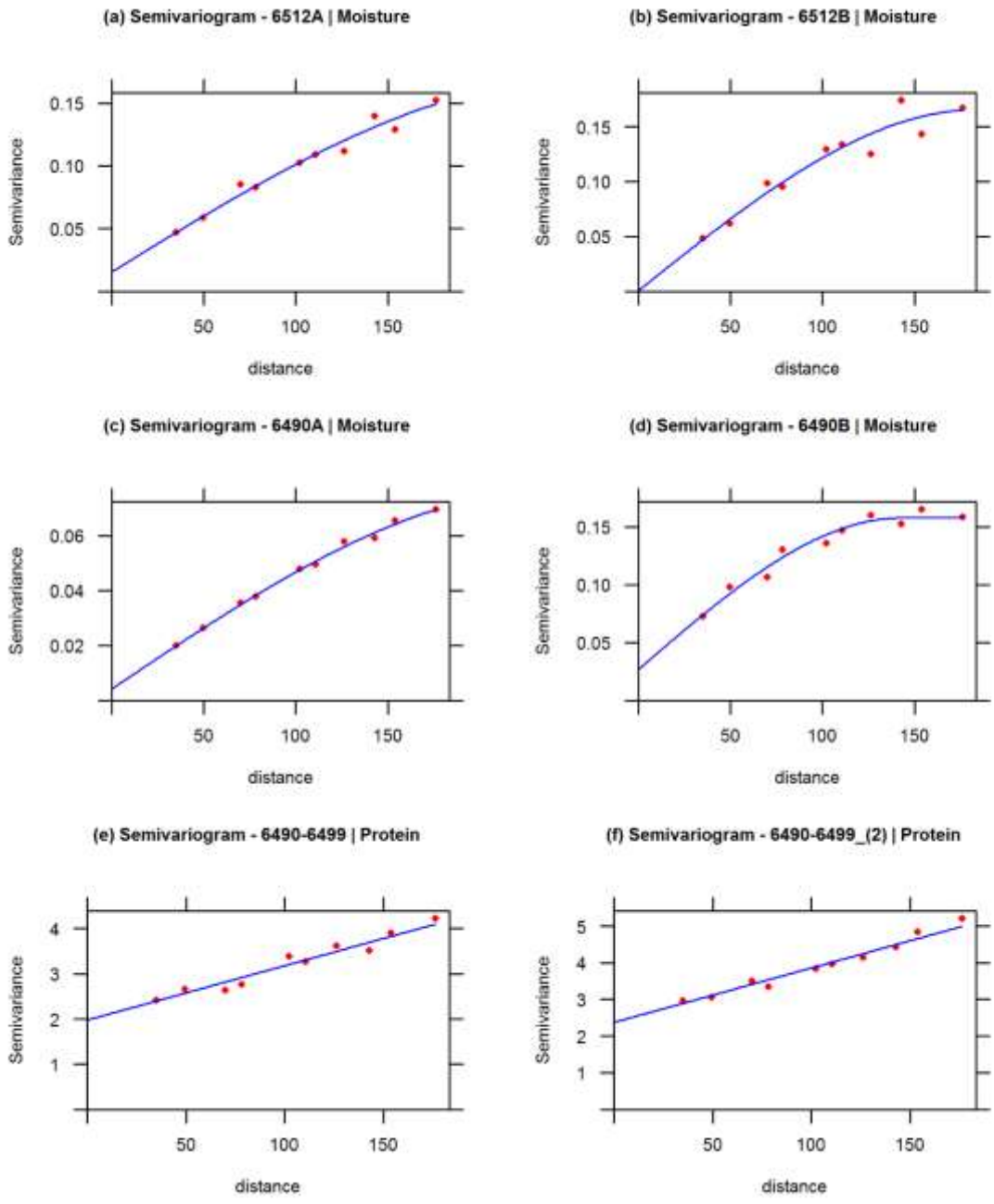
(a) Semivariogram - 6512A | Moisture
(b) Semivariogram - 6512B | Moisture
(c) Semivariogram - 6490A | Moisture
(d) Semivariogram - 6490B | Moisture
(e) Semivariogram - 6490-6499 | Protein
(f) Semivariogram - 6490-6499_(2) | Protein
Semivariance
distance

FIGURE 6

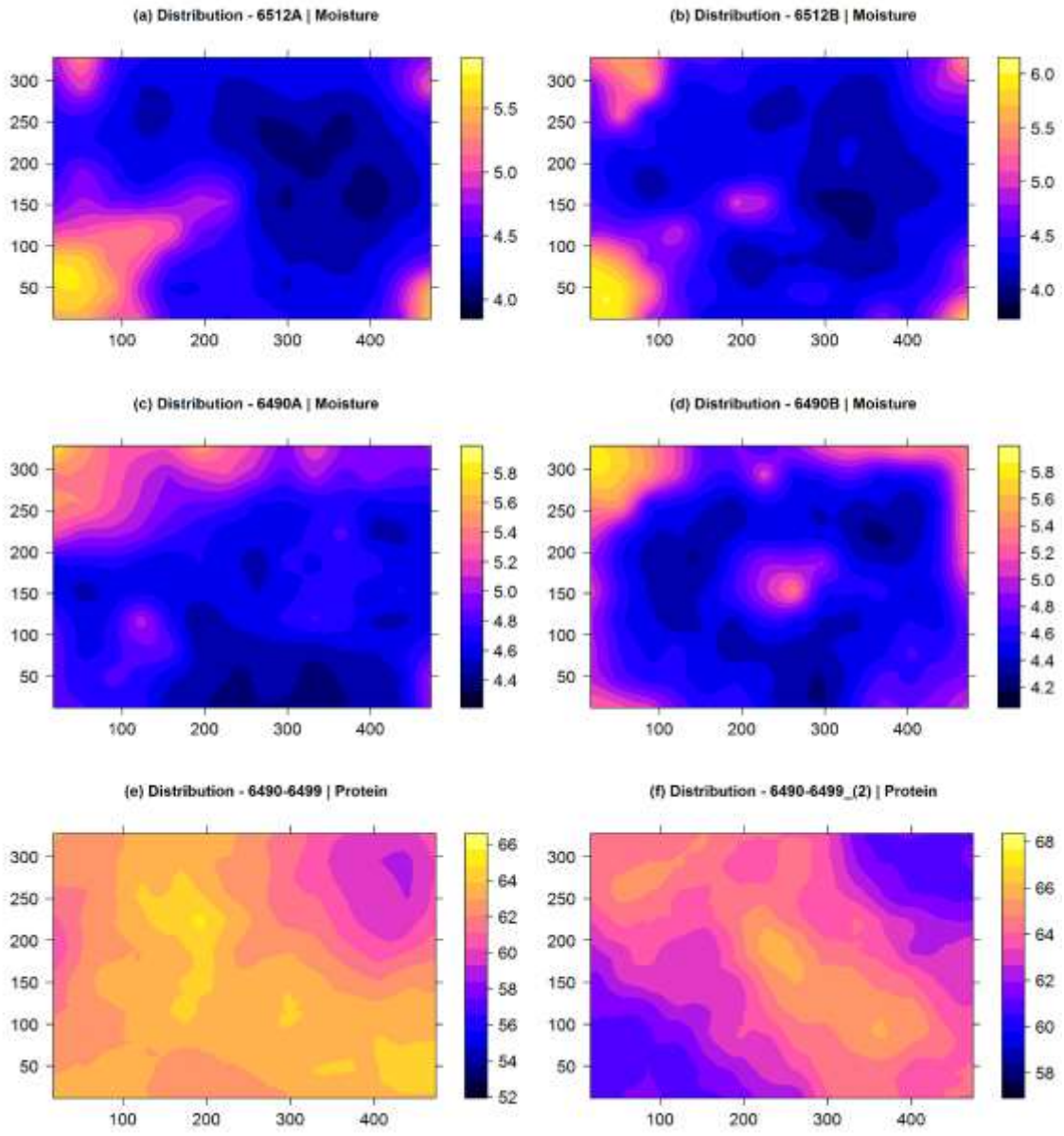
(a) Distribution - 6512A | Moisture
(b) Distribution - 6512B | Moisture
(c) Distribution - 6490A | Moisture
(d) Distribution - 6490B | Moisture
(e) Distribution - 6490-6499 | Protein
(f) Distribution - 6490-6499_(2) | Protein

FIGURE 7