Multistage and adaptive sampling protocols combined with NIR sensors for automated monitoring of raw materials in bulk

J.A. Adame-Siles*[1], J.E. Guerrero-Ginel[1], T. Fearn[2], A. Garrido-Varo[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

A NIR spectroscopy-based real-time monitoring system is proposed to sample and analyse agro-industrial raw materials transported in bulk in a single stage, easing and optimising the evaluation process of incoming lots at reception of agri-food plants. NIR analysis allows rapid and cost-effective analytical results to be obtained, and hence to rethink current sampling protocols. For this purpose, multistage and adaptive sampling designs were tested in this paper, which have been reported (in soil science and ecology) to be more flexible and efficient than conventional strategies to study patterns of clustering or patchiness, which can be the result of natural phenomena. The additional spatial information provided by NIR has also been exploited, using geostatistical analysis to model the spatial pattern of key analytical constituents in Processed Animal Proteins (PAPs). This study addresses the assessment of two kinds of quality/safety issues in PAP lots –moisture accumulation and cross-contamination. After a simulation study, qualitative and quantitative analyses were carried out to make a performance comparison between sampling designs. Results show that sampling densities below 10-15% demonstrated higher estimation errors, failing to represent the actual spatial patterns, while a stratified adaptive cluster sampling design achieved the best performance.

Keywords: Near infrared spectroscopy, Geostatistics, Kriging, adaptive cluster sampling, automatic control system, real-time surveillance

33 **Abbreviations**

ACS	Adaptive Cluster Sampling
ANOVA	Analysis of Variance
Clu	Cluster Sampling
CPcs	Case study testing cross-contamination
CtrA	Criterion to define critical values (10% above the mean)
CtrB	Criterion to define critical values (90^{th} quantile)
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
ISO	International Organization for Standardization
ISTA	International Seed Testing Association
Mcs	Case study testing high moisture content areas
NIRS	Near infrared Spectroscopy
OK	Ordinary Kriging
PAP	Processed Animal Proteins
R2	Coefficient of determination
RMSEP	Root Mean Square Error of Prediction
RPD	Residual Predictive Deviation

SECV	Standard Error of Cross-validation
SRS	Simple Random Sampling
StrACS	Stratified Adaptive Cluster Sampling
TOS	Theory of Sampling

1 Introduction

The EU General Food Law Regulation provides a regulatory framework involving all stages of the food & feed chain (European Commission, 2002). This law constitutes an overarching, strict and comprehensive policy, assuming food & feed safety and quality as a priority. In terms of regulatory compliance, a wide range of rules are enforced for issues such as trade aspects and processing or storage of raw materials. Thus, establishing management programs and surveillance schemes is vital for all stakeholders involved (from public bodies and food & feed operators to consumers). As a consequence, they need to address the implementation and strengthening of monitoring and rapid alert systems, as well as codes of good practice, which help to ensure food & feed safety and quality standards, and improve traceability. For this purpose, food & feed controls along the stages of the chain, before distribution, become a key pillar. In this regard, the EU Official Controls Regulation (European Commission, 2017) acknowledges the importance of auto-controls performed by operators, including private quality assurance systems, as a support for the official controls.

Cooperation between operators and authorities is therefore of major interest (Directorate General for Health and Food Safety, 2017; FEFAC, 2016, 2018), and the agri-food industry has a crucial role to play in this context. Over the last few years, substantial efforts have been made by the scientific community and competent participants (manufacturers, laboratories, authorities, etc.) to develop methods and standards to monitor and control all production steps. Nevertheless, despite the success achieved in some stages of the process, sampling and analysis of raw materials in bulk still remain as demanding tasks. In this case, the challenge is of greater magnitude, being determined not only by the difficulty of dealing with large volume products, but also by the need to design and adapt control protocols (sampling strategies and analysis methods) to the specific requirements of each type of product and lot.

Bulk food and feed sampling is widely understood as a multistep process in which classically a set of primary increments (taken from the lot) is pooled to form a composite sample, then mass-reduced (in various steps) to ultimately get the analytical aliquot with the right size for laboratory analysis (European Commission, 2009, 2013; ISO, 2002, 2009). The importance of obtaining a representative sample as a result of this process should be emphasised. This is indeed considered an essential prerequisite, equally important as the analytical accuracy, to finally draw reliable conclusions (Esbensen, Paoletti, & Thiex, 2015; Kuiper & Paoletti, 2015). The Theory of Sampling (TOS) has emerged as an effective framework to control and minimize errors occurring at all sampling steps involved, thus providing principles for representative sampling (Esbensen, 2013; Esbensen & Mortensen, 2010; Gy, 2004; Petersen, Minkkinen, & Esbensen, 2005).

On the other hand, a number of international organizations such as the International Organization for Standardization (ISO), Codex Alimentarius, the Food and Agriculture Organization of the United Nations (FAO) or the International Seed Testing Association (ISTA) have also attempted to define sampling approaches and procedures for the inspection of bulk materials. Nevertheless, their appropriateness has been discussed. Paoletti and Esbensen (Paoletti & Esbensen, 2015) argue that most of them "lack of guidance on the correct prerequisite design, implementation, and operation of fit-for-purpose sampling plans and sampling procedures". In addition, they underline that sampling plans derived from these standards rely on distributional assumptions which are often neither explicitly described nor verified, and conveniently based on an unjustified randomness assumption for the distribution of the analyte of interest.

Notwithstanding the above, the patterns of a wide variety of phenomena affecting food and feed quality and safety, such as some material properties (e.g. heterogeneity) or the presence of contaminants (bacteria, fungi, etc.), show evidence of aggregation tendencies. Thus, conventional sampling designs may be inefficient to evaluate and detect issues that follow these spatial distributions. Moreover, most procedures conventionally carry out a sample plan fixed before sampling, the negative effects of under-sampling occurring are likely to increase. Consequently, there is a need for sampling strategies that may provide a viable solution in this context.

Adaptive sampling designs, which have not yet been studied for food & feed control purposes, have become well-known in soil and natural sciences as a way of addressing the inherently difficult sampling situations associated with these fields (e.g. mineral exploration or epidemiological studies). An adaptive sampling design is one in which the sample selection procedure depends on the values observed while conducting the survey. On the basis of this

principle, unlike conventional sampling designs, these allow to make decisions during the survey and adaptively increase sampling intensity, so that whenever a condition of interest is satisfied by the observed value of a selected unit, neighbouring sites are then explored and added to the sample. Therefore, adaptive sampling strategies have been reported to improve significantly the effectiveness of sampling effort, as well as the precision of the estimates when trying to infer patchy distributions as well as concentration or density of the aggregation patterns (Thompson, 1990, 2012; Thompson & Seber, 1996).

Besides the improvement of sampling designs, there is still much work to be done as regards the range of shortcomings that are currently linked to sampling of raw materials in bulk. Such limitations as cost, qualified manpower requirements and time constraints are determining factors that often lead to over-simplistic protocols (e.g. grab sampling). They are also characterised by low sample throughput (with a severe mass reduction, from several tonnes to a few grams), compromising the lot-sample representativity, and include too many stages in the field-to-aliquot pathway, which is detrimental to the efficiency of the evaluation process. Furthermore, the incremental samples are always aggregated before analysis and, as a consequence, all information about the distribution and the spatial structure of the characteristic of interest is lost.

Bearing these constraints in mind, research was recently initiated to explore a new approach for the analysis of raw materials in bulk directly inside the transport unit (trucks, containers, etc.), before unloading (Adame-Siles et al., 2017). From an analytical standpoint, this approach is based on near-infrared (NIR) spectroscopy. Considering the benefits of this technology, it has already become a crucial asset for a large number of agri-food industries, which have integrated NIR-based quality-control systems successfully, although still mostly as at-line applications. This work, however, sought to take the analysis stage from the laboratory to the delivery point of raw materials at any agri-food production plant. In addition to the ability to perform rapid and cost-effective analysis, which makes it possible to increase significantly the sample volume, the use of NIR fibre-optic sensors was proposed to obtain an analytical determination for each sampling point. This means that not only can every sampling unit be analysed and recorded separately, but it can also preserve its spatial coordinates. The study subsequently exploited the potential of the extra spatial information obtained, performing a geostatistical analysis in order to recognise and model the spatial structure of key properties of PAPs (Processed Animal Proteins). Therefore, this methodology laid the foundation for rethinking the existing sampling approaches and the evaluation of a real-time NIR-based monitoring system of raw materials in bulk.

The main aim of this paper is to investigate some fit-for-purpose sampling protocols, based on the adoption of multistage and adaptive sampling plans, for the inspection of raw materials in bulk using the above-mentioned methodology. It also pursues a performance comparison between strategies on their ability to characterize the spatial distributions of two quality and safety issues tested in PAP lots, selected as a case study to assess the proposed method.

2 Materials and Methods

2.1 Lots

A total of 8 lots of PAPs, directly received from the rendering plant, were involved in the experimental design of this paper. The set was selected considering the variability in species composition of the lots listed in (Adame-Siles et al., 2017). The selection consisted of the following lots: Lot 1 (100% Poultry), Lot 2 (58% Poultry, 42% Pig), Lot 3 (64% Poultry, 36% Pig), Lot 4 (100% Poultry), Lot 5 (50% Poultry, 50% Pig), Lot 7 (100% Poultry), Lot 8 (100% Poultry) and Lot 10 (23% Poultry, 60% Pig, 11% Cattle, 6% Sheep).

Two types of quality and safety risks were tested simulating a variety of situations. First, the presence of high moisture content areas was evaluated as case study (henceforth referred to as Mcs), since this factor is of great importance as it may be conducive to fungal or bacteriological problems. On the other hand, the adulteration by cross-contamination by PAPs of different nature or category was also addressed as another case study (henceforth referred to as CPcs). To this end, a glass container served to house all the PAP lots for sampling and analysis (Fig. 1A). A sheet of methacrylate with a sampling grid of 14 × 10 points was used as a reference to position the sensor.

Five lots (Lots 1, 2, 3, 4 and 7) were selected to form part of Mcs. The use of a different amount and distribution of water, poured one day prior to analysis, gave rise to the set of tests that constitute this case study. Most of them mainly involved water accumulation at the walls and corners of the container, which are commonly the highest risk areas in lots in bulk. On the one hand, Lot 1 was tested pouring a volume of water of 50 ml in the centre and 100 ml in every corner of the container (Fig. 1B). Secondly, 500 ml of water were added to Lot 2 and located in the centre of the container (Fig. 1C). Thirdly, two discharges of 300 ml were applied to Lot 3 in the north and south-central areas of the container (Fig. 1D). Furthermore, Lot 4 test involved the pouring of 400 ml and 250 ml of water over the north-west and south-east corners, respectively (Fig. 1E). Finally, Lot 7 was tested by pouring 500 ml of water in all corners (Fig. 1F). For these cases, measurements were taken at two different depths (layer U and L).

On the other hand, CPcs was carried out using Lots 1, 5, 8 and 10. Tests under CPcs experiment were designed in order to explore whether the methodology was able to recognise possible regions that did not follow the expected pattern of the lot being tested. For this purpose, three different mixtures were made by varying the distribution of the lots in the glass container. Two tests involved Lots 1 and 5, one of them was carried out positioning sample from Lot 1 in the north-east corner of the container and the rest of the it being sample from Lot 5 (Fig. 1G) and, the other test, locating sample from Lot 1 in the north-east and south-west corners, while the rest of the container was filled with sample from Lot 5 (Fig. 1H). Moreover, a third test involved Lot 10, which was located at the north-east and south-west corners of the container, and Lot 8 was used in the rest (Fig. 1I). In this case study, only layer U was measured.

The set of experiments performed aimed at dealing with borderline cases of both kinds of adulteration, in which risk areas are highly localised, since this served as a starting point to test the limits of the methodology.

2.2 Instrumentation and analysis

NIR analysis was performed by measuring spectra in each test using a reflection probe (Turbido, Solvias AG, Kaiseraugst, Switzerland) interfaced to a Matrix-F FT-NIR instrument (Bruker Optics, Ettlingen, Germany) (834.2–2502.4 nm). The probe features a bundle of two optical fibres (core size of 600 µm) encased in a stainless-steel body (300 mm in length; 12 mm in diameter), whose end splits in two legs (illumination/acquisition), and its tip has a sapphire window of 1.5 mm in diameter. Two fibre-optic cables of 100 m enabled to connect the probe to the instrument.

Measurements were the result of a total of 32 scans, with a resolution of 16 cm^{-1} and a scanner velocity of 10 kHz. A probe-specific Spectralon was used for white reference measurements, which were made after every set of 42. Within the context of a preliminary study (Adame-Siles et al., 2017), firstly, the noise level of the signal was evaluated along the spectral range by applying to the log 1/R data a first derivative pre-treatment, with a single-unit gap and five data-point smoothing. After visual examination, noisy regions were found at the beginning and at the end of the spectral range, leading to the selection of the optimum wavelength range 1386-2033 nm. Subsequently, a standardisation methodology was initiated to transfer a database of 346 samples of PAPs, from which calibration equations had been developed using a different analysis mode (the same instrument was used but coupled to a detection head for contactless measurements). Finally, after a recalibration procedure, calibration equations (whose most relevant statistics are shown in Table 1) were obtained so that an analytical result could be got at every sampling unit using the NIR reflection probe. Moisture and crude protein were the constituents selected as control parameters in Msc and CPsc, respectively.

Spectral measurements were acquired using software OPUS v7.0 (Bruker Optics). Moreover, NIR prediction models were applied by using WinISI v.1.50 (Infrasoft International), Matlab R2018a (The MathWorks Inc.) and PLS Toolbox (Eigenvector Research).

2.3 Sampling study

2.3.1 Preliminary setup

First of all, NIR measurements were made once at every sampling unit of the global grid (N = 140; hereinafter referred as '100% sampling') in all tests of both case studies. This population was designed to be used as analytical reference for the subsequent sampling study. Following this, according to the different sampling plans, samples were then obtained as a subset of N. All plans included some randomness, thus a total of S = 1000 simulation replications were computed in all cases, which enabled a performance comparison among sampling designs.

2.3.2 Sampling designs

2.3.2.1 Multistage sampling

A two-stage sampling design (Cluster-SRS) was tried in this case, with cluster (Clu) sampling at the first stage, and simple random sampling (SRS) at the second stage. For this purpose, N was divided into 14 primary units, each composed of a total of 10 secondary units. A set of four different sampling intensities were then addressed (i = 30, 20, 10 and 5% of the N population).

The approach consisted of selecting 7 clusters randomly (regardless of the sampling intensity attempted) and, at the second stage, a simple random sample of secondary units, which varies depending upon the sampling intensity to achieve: i = 30% (6 units/cluster), i = 20% (4 units/cluster), i = 10% (2 units/cluster) and i = 5% (1 unit/cluster). Figure 2A shows an example of a two-stage sample (i = 5%) selected following this protocol. As noted above, this procedure was run a total of 1000 times for each i (Fig. 3A).

2.3.2.2 Adaptive Cluster Sampling

An adaptive cluster sampling (ACS) design generally includes the following steps: (i) an initial sample of units is selected using some probability sampling scheme; (ii) for every unit in which the observed value meets a given criterion *C*, additional units in some pre-defined neighbourhood of that unit will become part of the sample; and (iii) if any of these units satisfy *C*, their neighbourhoods are then included too, which gradually leads to obtaining what is known as a *network* (a group of adjacent units whose values are all greater than or equal to the critical value). In theory, this process continues until *C* is not met by any unit, which is usually

referred to as *unrestricted adaptive sampling*. Nonetheless, in order to avoid open-ended sampling designs, with the consequent effects on costs and logistics, ACS frequently requires a stopping rule to terminate the sampling process. Moreover, another limitation of the adaptive selection procedures is that they may introduce biases into conventional estimators, so that the need for design-unbiased estimators is emphasised here (such as Hansen-Hurwitz and Horvitz-Thompson estimators) (Thompson, 1990, 2012).

In this paper, three variants of ACS were tested, all with a preliminary sampling intensity of 5%. They differ, however, in the method for selecting the initial sample. On the one hand, the first scheme tried was a pure ACS, i.e. a simple random sample of units was obtained to begin the process (Fig. 2B). In addition, two versions of stratified sampling were examined. The first strategy (StrACS-1) divided the study area into 7 strata: north and south-central strata, 4 regions in the corners of the container and a central stratum (Fig. 2C). The within-stratum sample size of the initial sample was allocated based on proportional allocation. The second approach (StrACS-2) made a different arrangement within the study area, stratifying it into three zones with unequal probability (Fig. 2D). One stratum covered the edge of the container, the second one was contiguous, while the innermost layer constituted the third stratum. In this case, the initial sample (i = 5%, i.e. 7 sampling units) was allocated to strata in proportions, 50% (4 units), 30% (2 units) and 20% (1 unit), decreasing from outer to inner strata (rounding sample size to a whole number). In the stratified protocols, the neighbourhoods were allowed to cross the boundaries of strata, as stratification was only applied to the initial sample.

In order to limit the total sampling effort to a practical level, a stopping rule of level 2 (the neighbourhood exploration procedure was only conducted twice) was used. There exist different possible patterns when defining a neighbourhood of units (e.g. top, bottom, left and right; northwest, southwest, northeast and southeast, etc.). In this paper, the first-order neighbourhood consisted of the initial unit itself and the 8 contiguous boundary units. Moreover, Figs. 2B, 2C and 2D show the distribution of the second-order neighbourhood defined, in which 8 more units are added to the sample.

Concerning the critical value C, two different criteria were applied in this regard. The first criterion, henceforth referred to as *CtrA*, considered as critical those units in which the parameter is equal to or greater than 10% above the mean of initial samples. On the other hand, the second criterion, or *CtrB*, sets C to the 90^{th} quantile of the sample values. Figure 3B represents the simulation procedure followed for the adaptive designs described. As previously stated, a total of 1000 simulation replications were performed in each case.

The algorithms needed to perform the simulation study were developed in RStudio (v 1.1.1463).

2.4 Spatial analysis

The spatial patterns of the analytical constituents considered, moisture (Mcs) and crude protein (CPcs), were analysed using geostatistics. Therefore, a geostatistical study was conducted for every iteration of the simulation process, and took place in two stages: (i) structural analysis; and (ii) spatial estimation.

The first phase, commonly referred to as variography, uses the semi-variogram to characterize and model the autocorrelation structure of data, thus providing a means to quantify the spatial variation of the property of interest (Chiles & Delfiner, 2012). The experimental semi-variogram displays a curve relating the distance between data pairs (lag distance, or **h**) with their semi-variance, a measure that averages squared differences of the variable. Both omnidirectional and directional (0, 45, 90 and 135º) variograms were calculated for each case study in this paper. Variogram modelling was addressed as a subsequent step. This task allows to fit an authorised mathematical function of the distance to the experimental variogram, providing the parametric model needed to compute a variogram value at unobserved sites and to meet the mathematical property of *conditional negative definiteness* (Gringarten & Deutsch, 2001).

The second stage refers to the use of the previously modelled spatial variance to estimate interpolated values between sampling points. For this purpose, one of the most reliable and commonly used kriging estimators, named ordinary kriging (OK), was applied for the mapping of the analytical parameters in this paper. Kriging is a robust family of generalised least squares linear regression algorithms which, based on the results of the structural analysis, achieves to accurately estimate values at unsampled locations. A more detailed description of this geostatistical approach can be found in (Adame-Siles et al., 2017), and more on OK theory and practice in (Cressie, 1991; Goovaerts, 1997a; Isaaks & Srivastava, 1989; Myers, 1991; Webster & Oliver, 2007).

Variographic analysis and spatial interpolation were both implemented in the R environment (version 3.4.3). The gstat R package was used to perform all geostatistical analyses (Pebesma, 2004).

2.4.1 Calculation of the estimation error

The Root Mean Square Error of Prediction (RMSEP) statistic was used to evaluate the performance in each case study:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{j=1}^{N}\left(y_{j,krig} - y_{j,NIR}\right)^2}{N}} \quad (1)$$

where $y_{j,krig}$ are the interpolations obtained by kriging and, $y_{j,NIR}$, the analytical values, for the sampling unit j of the sampling grid N. It is worth noting that for some j units, the $y_{j,krig}$ will be the actual measurement, so the derived errors will be zero in these cases.

Moreover, the mean and the standard deviation of the RMSEP were also calculated for the S=1000 simulations performed:

$$\mu = \frac{1}{S}\sum_{i=1}^{S}\mathrm{RMSEP}_i \quad (2)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{S}\left(\mathrm{RMSEP}_i - \mu\right)^2}{S-1}} \quad (3)$$

In addition, one-way analysis of variance (ANOVA) was performed for every test to examine whether significant differences in log values of the RMSEP were found among the three adaptive sampling designs tried.

3 Results and Discussion

3.1 Data preparation

As a first step, the geostatistical study addressed the variographic analysis of both case studies in order to model the spatial pattern of the constituents of interest. For variography it is recommended to have at least 100 - 150 sampling points to obtain robust results (Webster & Oliver, 2007), therefore semi-variograms were computed from the data set of N sampling points in each case, which would also allow to build a library of structural analyses of PAP lots to help future routine structural assessments based on sample data. As reported by Adame-Siles et al (Adame-Siles et al., 2017), the structural analysis for moisture and crude protein revealed several contrasts between both constituents in their spatial behaviour. Moisture variograms exhibit zero, or close to zero, nugget (the semi-variogram value at the origin), and a linear increase until they reach an asymptote, or ‘sill’ (the semivariance value at which the semivariogram levels off). Nevertheless, crude protein tests generally showed a monotonic increase with increasing lag distance and a positive intercept on the ordinate.

A model was fitted to the experimental variogram. The most common functions available for this purpose were tested and, two mathematical models, linear and spherical, were selected as they provided the best fit for the crude protein and moisture semi-variograms, respectively.

The sampling study started after the variogram modelling stage and was carried out according to the procedure described in section 2.3. Then, every sampling design (Fig. 2) was performed in each test of case studies A and B following the protocol illustrated in Fig. 3. This procedure was executed a total of 1,000 times in each case.

Following the structural analysis, spatial interpolation by ordinary kriging was tackled, using as an input the sample data set resulting from each iteration of the sampling study (i.e. for every sampling design and case study). As a consequence, kriging maps representing the spatial pattern of the constituent were obtained in each case. Finally, the sample sets together with the interpolated values at all 140 points were used in Eq. (1)~~(1)~~ to evaluate the performance and for comparison purposes among designs.

3.2 Performance assessment

3.2.1 Qualitative analysis

Two representative illustrations of the quality and safety risks evaluated in lots of PAPs are shown in Fig. 4 and Fig. 5. The first one depicts moisture spatial distributions associated with one test of Mcs (Lot 1 – layer L), while the second shows crude protein surfaces generated from one test belonging to CPcs (Lot1+5(2)). Both figures represent the outcome of one iteration from the set of S simulations obtained for each sampling design tested.

The mapping of moisture for the '100% sampling' design is shown first in Fig. 4, offering a visual reference against which to compare all other distributions resulting from the sampling designs to be evaluated. As seen in Fig. 1B, the experimental design conceived for this test included the pouring of water at the centre of the container and all four corners. Bearing this in mind, if the effect of irregular water drainage and distribution is taken into account, which varies depending on the depth of the layer under consideration, the spatial distribution in this case managed to illustrate the regions where the higher moisture accumulation took place.

Regardless of the design, sampling intensity plays a crucial role, having a pronounced impact on the results. This effect can be inferred taking into account the kriged maps obtained by the multistage strategy, Cluster-SRS. In this case, four sampling intensities were tested, and a reduction in performance on the mapping of the affected areas can be observed as sample size decreases. Higher sampling intensities, i =30% and (to a lesser extent) i = 20%, generally manage to reproduce the original distribution, whereas the lowest ones (i =10% and i =5%) did

not succeed in portraying all the risk areas, demonstrating a distinct loss of information or smoothing effect compared with the preceding ones.

Figure 4 also shows the spatial distributions achieved from the adaptive sampling designs tested (ACS, StrACS-1 and StrACS-2). Surfaces using *CtrA* (10% above mean) to determine the critical value are depicted in Figure 4.5 to 4.7, while maps based on *CtrB* (90th quantile) are shown in 4.8 to 4.10. On the basis of the stopping rule of level 2 used, both criteria differ in the maximum sampling density they reach in the distributions represented, with *CtrA* (20%) slightly surpassing to *CtrB* (16.4%). In terms of accuracy to characterise the actual moisture distribution, *CtrA* surfaces perform well especially regarding the critical areas present at the edge of the container, whereas all designs but ACS lost the central area. By contrast, *CtrB* maps manage to picture this risk area in the centre, but most (especially ACS) lose information on the edge.

The results for the crude protein test are shown in Fig. 5. In this case, the experimental design consisted of placing two different lots in the container, in such a way that both the lower-left corner and the upper-right corner belonged to Lot 1 and the rest to Lot 5 (Fig. 1H). The surface obtained from the 100% sampling design for the crude protein constituent effectively discriminates the different pattern at the corners, thus in line with the design of the test.

As regard sampling intensity, the resulting maps for the Cluster-SRS sampling design show a decline in performance with lower densities, from which a similar analysis to the previously made can be derived. As was the case for the moisture test, sampling intensities of 30% and 20% generally achieve more faithful crude protein surfaces than 10% and 5% do. Concerning the adaptive sampling designs, it may be seen that those linked to *CtrA* finally remain at the initial sampling density (i = 5%), which means that no critical units were found during the process. In fact, this is because the crude protein parameter ranges for this test from 57.61% to 67.66%, averaging 63.39%, thus making that none of the units can satisfy *CtrA*, i.e. exceed 1.1 times the average value. Taking this into account, only StrACS-1 achieved reasonable results in comparison with the other designs. On the other hand, sampling designs following *CtrB* attained a sample size of 16.4%, generally accomplishing the task of illustrating the actual distribution.

3.2.2 Quantitative analysis

The estimation error was calculated for each test of both case studies (A and B) by the Eq. (1)~~(1)~~. The RMSEP statistic was computed from the comparison between the kriging estimations and the actual NIRS analytical values at each unit of the population (N). The

average value (Eq. (2)~~(2)~~) and standard deviation (Eq. (3)~~(3)~~) of the RMSEP were also calculated for each sampling design from the 1000 simulations performed.

The estimation error values for the multistage sampling design tested in this paper (Cluster-SRS) are shown in **Error! Reference source not found.**~~Table 2~~. In quantitative terms, they support the previous qualitative analysis regarding the effect of the sampling density on the results. Thus, it can be seen that there exists a negative correlation between the sample size and the estimation error. A determining factor contributing to this effect is the declining availability of information with lower sampling intensities, which inevitably leads to bigger errors.

The estimation errors associated to the adaptive sampling designs performed under *CtrA* are reported in Table ~~Table 3~~ (ACS, StrACS-1, StrACS-2), while those carried out under *CtrB* are shown in Table ~~Table 4~~ (ACS, StrACS-1, StrACS-2). Moreover, in order to facilitate assessment and comparison, Figure 6 graphically displays the average values (from the 1000 simulations) of the RMSEP statistic for all the sampling designs tested. Cluster-SRS was included in this figure for i=5%, which corresponds to the initial sampling density for the adaptive designs (the average sampling intensity, for the S=1000 simulations, reached by each adaptive design is also expressed). The results are grouped by test (only layer U is shown for those belonging to Mcs).

If the methods used to determine the critical value are compared, the criterion based on the 90th quantile (*CtrB*) accomplished better results in all cases than the approach considering critical values above 10% over the mean (*CtrA*). The performance gap between both criteria may be the result of several factors. On the one hand, the efficiency of the current industrial manufacturing process of PAPs regularly allows to achieve homogeneous products, which might show as a result a low chemical variability. Additionally, the higher critical value set by *CtrA* may have caused a lack of units meeting this criterion which, in turn, leads to a smaller sample size (with the previously discussed consequences when applying kriging). This has made that in many cases the adaptive sampling designs under *CtrA* remain just as the initial sampling probability scheme used.

It can be noted that the sampling design with the lowest estimation error in all cases is StrACS-1 (Table ~~Table 3~~ and Table ~~Table 4~~). This strategy prevailed regardless of the criterion used for determining the critical value or the case study considered (moisture and crude protein tests), which suggests that, based on the strata distribution defined by this design, it achieved~~s~~ a more effective sample allocation than the rest in the tests performed. On the other hand, when compared ACS and StrACS-2, their estimation error values remained close (both under criteria *CtrA* and *CtrB*), so that the former sometimes outperformed the latter or vice versa, thus no clear evidence was found to help decide between them in this study. For instance, considering

418 *CtrA*, the (mean) error of StrACS-1 for the moisture test involving Lot 1 (layer L) is 0.366,
419 however, the estimation errors of ACS and StrACS-2 in this case are 0.379 and 0.374,
420 respectively. If the protein test using Lot 1+5(2) is considered, it can be observed a similar
421 result, StrACS-1 achieves an error value of 2.119, whereas for ACS is 2.171 and for StrACS-2
422 is 2.165. Taking *CtrB* into account, once again StrACS-1 reaches the lowest errors (0.370 -for
423 Lot 1 (layer L)- and 1.973 -for Lot 1+5(2)), while ACS and StrACS-2 errors are equal for Lot 1
424 (layer L) (both are 0.377), and ACS (2.012) outperforms StrACS-2 (2.038) in the protein test of
425 Lot 1+5(2). The ANOVA results (Table 5) show that there was significant variation in RMSEP
426 values among the three adaptive sampling designs in all cases ($P<0.05$), except for Lot 3 (layer
427 U, CtrA).

428 Adaptive sampling designs have shown that they can be more efficient, flexible and
429 practical in reality than conventional sampling designs under certain circumstances. They also
430 have, however, some practical limitations as can be seen from the results obtained. The
431 selection of a suitable critical value must rely on prior knowledge about the population
432 distribution, otherwise it may not be a trivial task. It is therefore difficult to give a "rule of
433 thumb" for the choice of an optimal criterion, as a high critical value may result in a too low
434 sampling density, while a low critical value might lead to the problem of indefinite sampling.
435 For this reason, a stopping rule is typically required in order to terminate the sampling process
436 when using adaptive sampling.

437 This study suggests that the minimisation of the RMSEP involves several factors. As
438 discussed earlier, sampling density plays a key role in the subsequent geostatistical analysis and
439 estimation by kriging. In addition, constraints of sampling designs should be taken into
440 consideration too when trying to characterize the spatial structure and possible sources of
441 quality/safety issues. Then, it should be highlighted the importance of optimising the sampling
442 intensity/sampling design relationship within the context of the methodology described in this
443 paper. For this purpose, it is the goal of future studies to explore options that may improve this
444 combination, which would lead to better performance and more faithful representations of the
445 spatial surfaces. In this regard, short-term steps may include: (i) the use of a larger initial sample
446 size (when using adaptive designs) to avoid the negative effect of under-sampling; (ii) test other
447 thresholds and methods to set the critical value from the analysis of a robust database of PAP
448 lots; and (iii) test a new set of types of neighbourhood and stopping rules. Moreover, further
449 research should also be conducted as regards the kriging-based approach, so as to improve
450 aspects such as the smoothing effect. The optimization of this process might be explored by
451 replacing kriging with stochastic simulation techniques. Unlike kriging, which provides the
452 'best' local estimates of the variable of interest (without regard to the resulting statistics of those
453 estimates), stochastic simulation aims at reproducing the global statistics and maintaining the

texture of the variation, and these take precedence over local accuracy (Goovaerts, 1997b, 2001; Webster & Oliver, 2007). Therefore, depending on the purpose of the control, stochastic simulation may also help to infer the spatial distribution of the characteristic under study.

4 Conclusions

This study provides a methodology, based on NIR spectroscopy in combination with geostatistical inferential methods, for performing real-time sampling and analysis of raw materials in bulk (as a single operation). This approach makes it possible to explore the evaluation of new sampling protocols that can be more efficient than current strategies.

The results suggest that sampling density plays a major role in the geostatistical process. Overall, results indicate that sampling intensities below 10 - 15% showed poorer performance, failing to reproduce the actual spatial patterns. Furthermore, the sampling design is also key to characterize the spatial structure. In this case, the Stratified Adaptive Cluster Sampling design (StrACS-1) performed better than the rest of protocols tested.

The implementation of the methodology proposed requires an optimal balance between the sampling design, the intensity and the criterion used to determine the critical value in order to minimize the estimation error and ensure reliable results. In this regard, our research group is currently exploring the potential of automating the sampling-analysis process using this methodology by developing a robot unit. First, this could solve some of the existing limitations by significantly increasing the sampling intensity without cost and time implications. In addition, if implemented, this could provide with a rapid and cost-effective monitoring system, which would bring transparency to the supplier-purchaser relationship and benefit both efficiency and the decision-making process.

Acknowledgements

This research was carried out within the framework of the Excellence Project AGR-6033 “On-site incorporation of NIRS technology for the control of products and processes in the animal feed industry” financed by the Andalusian Regional Government, and Project RTA2012-00063-C02-02 “Latest-generation portable NIRS instruments for on-site analysis in the rendering industry”, financed by the INIA and European Regional Development Fund (ERDF). The authors are grateful to Kaura Coproducts, S.L. (Spain) for providing the samples for this research. They are also grateful to Ms. Mª Carmen Fernández and Mr. Antonio López for their help in handling and analysing samples.

References

Adame-Siles, J. A., Fearn, T., Guerrero-Ginel, J. E., Garrido-Varo, A., Maroto-Molina, F., & Pérez-Marín, D. (2017). Near-Infrared Spectroscopy and Geostatistical Analysis for Modeling Spatial Distribution of Analytical Constituents in Bulk Animal By-Product Protein Meals. *Applied Spectroscopy*, *71*(3), 520–532. https://doi.org/10.1177/0003702816683958

Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics. Modeling Spatial Uncertainty*. New Jersey: John Wiley & Sons, Inc.

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

Directorate General for Health and Food Safety. (2017). *Interim Overview Report - Audits of Official Controls in EU-Member States*. https://doi.org/10.2875/202803

Esbensen, K. H. (2013). *DS 3077. Representative Sampling - Horizontal Standard*. Danish Standards. Retrieved from www.ds.dk

Esbensen, K. H., & Mortensen, P. (2010). Process Sampling (Theory of Sampling, TOS)-the Missing Link in Process Analytical Technology. In *Process Analytical Technology* (2nd Ed, pp. 37–80). Wiley.

Esbensen, K. H., Paoletti, C., & Thiex, N. (2015). Representative Sampling for Food and Feed Materials: A Critical Need for Food/Feed Safety. *Journal of AOAC International*, *98*(2), 249–251. https://doi.org/http://dx.doi.org/10.5740/jaoacint.SGE_Esbensen_intro

European Commission. (2002). Regulation (EC) No 178/2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Official Journal of the European Union*, *L 31*(January), 1–24.

European Commission. (2009). Commission Regulation (EC) No 152/2009 laying down the methods of sampling and analysis for the official control of feed. *Official Journal of the European Union*, *L 54*(January), 1–130.

European Commission. (2013). Commission Regulation (EC) No 691/2013 of 19 July 2013 amending Regulation (EC) No 152/2009 as regards methods of sampling and analysis. *Official Journal of the European Union*, *L 197*(July), 1–12.

European Commission. (2017). Regulation (EU) 2017/625 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products. *Official Journal of the European Union*, *95*(1), 1–142.

FEFAC. (2016). *FEFAC vision on Feed Safety Management*. Brussels. Retrieved from https://www.fefac.eu/files/67549.pdf

FEFAC. (2018). *Annual Report 2017-2018*. Brussels. Retrieved from https://www.fefac.eu/files/83625.pdf

Goovaerts, P. (1997a). *Geostatistics for Natural Resources Evaluation*. Oxford: Oxford University Press.

Goovaerts, P. (1997b). Kriging vs Stochastic Simulation for Risk Analysis in Soil Contamination. In R. E. Soares, A., Gomez-Hernandez, J., Froidevaux (Ed.), *GeoENV I — Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics, vol 9.* (pp. 247–258). https://doi.org/10.1007/978-94-017-1675-8

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, *103*(1–2), 3–26. https://doi.org/10.1016/S0016-7061(01)00067-2

Gringarten, E., & Deutsch, C. V. (2001). Teacher's Aide Variogram Interpretation and Modeling 1. *Mathematical Geology*, *33*(4).

Gy, P. (2004). Sampling of discrete materials—a new introduction to the theory of sampling: I. Qualitative approach. *Chemometrics and Intelligent Laboratory Systems*, *74*(1), 7–24. https://doi.org/http://dx.doi.org/10.1016/j.chemolab.2004.05.012

Isaaks, E. H., & Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

ISO. (2002). *ISO 6497. Animal Feeding Stuffs - Sampling*. Geneva, Switzerland: International Organization for Standardization.

ISO. (2009). *ISO 24333. Cereals and cereal products - Sampling*. Geneva, Switzerland: International Organization for Standardization.

Kuiper, H. a., & Paoletti, C. (2015). Food and Feed Safety Assessment: The Importance of Proper Sampling. *Journal of AOAC International*, *98*(2), 252–258. https://doi.org/10.5740/jaoacint.15-007

546 Myers, D. E. (1991). Interpolation and estimation with spatially located data. *Chemometrics and*
547 *Intelligent Laboratory Systems*, *11*(3), 209–228. https://doi.org/10.1016/0169-
548 7439(91)85001-6

549 Paoletti, C., & Esbensen, K. H. (2015). Distributional Assumptions in Food and Feed
550 Commodities—Development of Fit-For-Purpose Sampling Protocols. *Journal of AOAC*
551 *International*, *98*(2), 295–300. https://doi.org/10.5740/jaoacint.14-250

552 Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers &*
553 *Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

554 Petersen, L., Minkkinen, P., & Esbensen, K. H. (2005). Representative sampling for reliable
555 data analysis: Theory of Sampling. *Chemometrics and Intelligent Laboratory Systems*,
556 *77*(1–2), 261–277. https://doi.org/10.1016/j.chemolab.2004.09.013

557 Thompson, S. K. (1990). Adaptive Cluster Sampling. *Journal of the American Statistical*
558 *Association*, *85*(412), 1050–1059.

559 Thompson, S. K. (2012). *Sampling* (Third Ed.). John Wiley & Sons.

560 Thompson, S. K., & Seber, G. A. F. (1996). *Adaptive sampling*. John Wiley & Sons, Inc.

561 Webster, R., & Oliver, M. (2007). *Geostatistics for Environmental Scientists* (2nd ed).
562 Chichester: John Wiley & Sons, Inc.

563

564

565

566

567

568

569

570

571

572

Tables

Table 1. Calibration statistics for predicting moisture and crude protein content (%) in PAP lots.

Constituent	Pre-processing	Mean	SECV	R^2	RPD
Moisture	1,5,5,1	3.78	0.36	0.77	2.1
Crude Protein	1,5,5,1	57.7	2.45	0.86	2.7

SECV: standard error of cross-validation (%); R^2: coefficient of determination; RPD: Residual Predictive Deviation.

Table 2. RMSEP (Mean and Standard Deviation) for Moisture (M) and Crude Protein (CP) tests. Cluster then Simple Random Sampling (CluSRS).

CluSRS	Sampling Intensity (%)	Lot 1 (M)		Lot 2 (M)		Lot 3 (M)		Lot 4 (M)		Lot 7 (M)		Lot 1+5 (CP)		Lot 1+5(2) (CP)		Lot 8+10 (CP)	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Layer U	30	0.161	0.010	0.114	0.009	0.116	0.008	0.137	0.006	0.287	0.016	1.483	0.118	1.578	0.086	1.540	0.050
	20	0.169	0.018	0.105	0.006	0.136	0.007	0.141	0.009	0.250	0.015	1.588	0.058	1.769	0.086	1.626	0.055
	10	0.218	0.020	0.127	0.014	0.158	0.016	0.157	0.011	0.293	0.026	1.777	0.103	2.044	0.135	1.817	0.111
	5	0.275	0.030	0.148	0.018	0.173	0.018	0.173	0.011	0.384	0.072	1.826	0.100	2.038	0.156	1.961	0.127
Layer L	30	0.298	0.009	0.103	0.005	0.100	0.005	0.105	0.007	0.337	0.024						
	20	0.297	0.024	0.104	0.008	0.197	0.012	0.123	0.005	0.401	0.015						

	10	0.379	0.031	0.119	0.009	0.139	0.025	0.125	0.010	0.329	0.035						
	5	0.411	0.034	0.127	0.016	0.149	0.009	0.140	0.009	0.439	0.031						

Table 3. RMSEP (Mean and Standard Deviation) for Moisture (M) and Crude Protein (CP) tests. Adaptive Sampling designs (Criterion A).

	Layer	Lot 1 (M)			Lot 2 (M)			Lot 3 (M)			Lot 4 (M)		
		Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std
ACS	U	7.9	0.259	0.049	5.0	0.143	0.017	5.7	0.176	0.023	5.0	0.172	0.016
	L	11.4	0.379	0.061	5.0	0.131	0.015	5.0	0.167	0.026	5.0	0.144	0.015
StrACS-1	U	8.6	0.235	0.043	5.0	0.138	0.016	5.7	0.169	0.019	5.0	0.171	0.016
	L	12.1	0.366	0.056	5.0	0.127	0.015	5.0	0.152	0.015	5.0	0.142	0.012
StrACS-2	U	8.6	0.249	0.048	5.0	0.146	0.019	5.7	0.173	0.027	5.0	0.172	0.017
	L	11.4	0.374	0.051	5.0	0.137	0.016	5.0	0.166	0.031	5.0	0.143	0.013
	Layer	Lot 7 (M)			Lot 1+5 (CP)			Lot 1+5(2) (CP)			Lot 8+10 (CP)		
		Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std
ACS	U	11.4	0.334	0.074	5.0	1.986	0.245	5.0	2.171	0.218	5.0	1.948	0.160
	L	10.7	0.395	0.065									
StrACS-1	U	12.9	0.304	0.055	5.0	1.943	0.187	5.0	2.119	0.204	5.0	1.920	0.150
	L	10.7	0.371	0.069									
StrACS-2	U	10.7	0.340	0.064	5.0	2.009	0.267	5.0	2.165	0.231	5.0	1.952	0.168
	L	10.7	0.384	0.053									

*Intensity (%): average sampling intensity reached calculated from the S=1000 simulations performed.

Table 4. RMSEP (Mean and Standard Deviation) for Moisture (M) and Crude Protein (CP) tests. Adaptive Sampling designs (Criterion B).

	Layer	Lot 1 (M)			Lot 2 (M)			Lot 3 (M)			Lot 4 (M)		
		Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std
ACS	U	12.9	0.242	0.048	15.0	0.123	0.016	13.6	0.162	0.019	12.9	0.159	0.015
	L	12.1	0.377	0.044	15.0	0.114	0.015	14.3	0.150	0.025	13.6	0.131	0.012
StrACS-1	U	12.9	0.214	0.035	15.7	0.118	0.013	13.6	0.153	0.012	13.6	0.156	0.014
	L	12.1	0.370	0.044	15.7	0.111	0.012	14.3	0.137	0.015	13.6	0.129	0.011
StrACS-2	U	11.4	0.238	0.048	14.3	0.128	0.021	12.9	0.158	0.020	12.1	0.158	0.015
	L	11.4	0.377	0.040	14.3	0.122	0.019	12.9	0.151	0.029	12.1	0.134	0.013
	Layer	Lot 7 (M)			Lot 1+5 (CP)			Lot 1+5(2) (CP)			Lot 8+10 (CP)		
		Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std	Intensity (%)	Mean	Std
ACS	U	12.9	0.328	0.059	14.3	1.834	0.192	14.3	2.012	0.187	14.3	1.790	0.117
	L	12.1	0.374	0.054									
StrACS-1	U	12.9	0.301	0.039	14.3	1.794	0.157	14.3	1.973	0.165	15.0	1.782	0.111
	L	12.1	0.356	0.052									
StrACS-2	U	11.4	0.338	0.057	12.9	1.857	0.205	12.9	2.038	0.200	13.6	1.824	0.144
	L	11.4	0.384	0.051									

*Intensity (%): average sampling intensity reached calculated from the S=1000 simulations performed.

Table 5. One-way ANOVA results (P values) for the three adaptive sampling designs tested (ACS, StrACS-1, StrACS-2) in Moisture (M) and Crude Protein (CP) tests.

Criterion	Layer	Lot 1 (M)	Lot 2 (M)	Lot 3 (M)	Lot 4 (M)
A	U	0	0	0.38	$7.03x10^{-10}$
	L	$3.1x10^{-7}$	0	0.002	0
B	U	0	0	$3.39x10^{-7}$	0
	L	$8.3x10^{-6}$	0	0	0
	Layer	Lot 7 (M)	Lot 1+5 (CP)	Lot 1+5(2) (CP)	Lot 8+10 (CP)
A	U	0	$1.63x10^{-9}$	$1.28x10^{-8}$	$3.16x10^{-6}$
	L	0			
B	U	0	$3.08x10^{-13}$	$3.55x10^{-14}$	$3.31x10^{-14}$
	L	0			

Figures

Fig. 1. Experimental design: (A) Glass container. (B) Lot 1. (C) Lot 2. (D) Lot 3. (E) Lot 4. (F) Lot 7. (G) Lot 1+5. (H) Lot 1+5(2). (I) Lot 8+10.

Fig. 2. Sampling designs: (A) Cluster then SRS, i=5% example. (B) Adaptive Cluster Sampling (ACS). (C) Stratified ACS-1. (D) Stratified ACS-2.

Fig. 3. Strategy for the simulation study. (A) Multistage design. (B) Adaptive designs.

Fig. 4. Spatial surfaces (Lot 1L; Moisture): 100% sampling; Cluster then SRS, intensities: (1) 30% (2) 20% (3) 10% (4) 5%; ACS (CtrA), intensity 20% (5); StrACS-1 (CtrA), intensity 20% (6); StrACS-2 (CtrA), intensity 20% (7); ACS (CtrB), intensity 16.4% (8); StrACS-1 (CtrB), intensity 16.4% (9); StrACS-2 (CtrB), intensity 16.4% (10).

Fig. 5. Spatial surfaces (Lot 1+5(2); Crude Protein): 100% sampling; Cluster then SRS, intensities: (1) 30% (2) 20% (3) 10% (4) 5%; ACS (CtrA), intensity 5% (5); StrACS-1 (CtrA), intensity 5% (6); StrACS-2 (CtrA), intensity 5% (7); ACS (CtrB), intensity 16.4% (8); StrACS-1 (CtrB), intensity 16.4% (9); StrACS-2 (CtrB), intensity 16.4% (10).

Fig. 6. Estimation error values (RMSEP) and average Sampling Intensity (%) reached by each design for Moisture (M) and Crude Protein (CP) tests. Sampling designs: Cluster then SRS (CluSRS); Adaptive Cluster Sampling – CtrA (ACS-A); Stratified Adaptive Cluster Sampling-1 – CtrA (StrACS-1-A); Stratified Adaptive Cluster Sampling-2 – CtrA (StrACS-2-A); Adaptive Cluster Sampling – CtrB (ACS-B); Stratified Adaptive Cluster Sampling-1 – CtrB (StrACS-1-B); Stratified Adaptive Cluster Sampling-2 – CtrB (StrACS-2-B).