Performance comparison of sampling designs for quality and safety control of raw materials in bulk: a simulation study based on NIR spectral data and geostatistical analysis

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

This study exploits the potential of near infrared (NIR) spectroscopy to deliver a measurement for each sampling point. Furthermore, it provides a protocol for the modelling of the spatial pattern of analytical constituents. On the basis of these two aspects, the methodology proposed in this work offers an opportunity to provide a real-time monitoring system to evaluate raw materials, easing and optimising the existing procedures for sampling and analysing products transported in bulk. In this paper, Processed Animal Proteins (PAPs) were selected as case study, and two types of quality/safety issues were tested in PAP lots —induced by moisture and cross-contamination. A simulation study, based on geostatistical analysis and the use of a set of sampling protocols, made a qualitative analysis possible to compare the representation of the spatial surfaces produced by each design. Moreover, the Root Mean Square Error of Prediction (RMSEP), calculated from the differences between the analytical values and the geostatistical predictions at unsampled locations, was used to measure the performance in each case. Results show the high sensitivity of the process to the sampling plan used — understood as the sampling design plus the sampling intensity. In general, a gradual decrease in the performance can be observed as the sampling intensity decreases, so that unlike for higher intensities, the too low ones resulted in oversmoothed surfaces which did not manage to represent the actual distribution. Overall, Stratified and Simple Random samplings achieved the best results in most cases. This indicated that an optimal balance between the design and the intensity of the sampling plan is imperative to perform this methodology.

Keywords: Near infrared spectroscopy; Geostatistics; Kriging; real-time evaluation; in situ monitoring; spatial analysis

1 Introduction

The European policy on quality and safety assurance of foods and feeds is stringent and extensive. Strengthening monitoring schemes and management systems is therefore a strategic goal for public bodies and food/feed operators, which need to develop alert systems and good manufacturing practices to improve traceability, obtain safe products and ensure quality standards. In this context, the existing legislation as regards animal by-products (ABPs) and, in particular, processed animal proteins (PAPs), goes in line with the above framework [1–3].

For PAPs, which are valued as a major component of pet foods, a number of ingredients are available in their manufacturing process (sheep, poultry, pig, bones, feathers, etc.), and the effect of varying their proportions may lead to significant differences in the chemical composition and nutritional value of the final product [4]. This case is evidence of the key role that the industry plays in the agri-food chain. In this respect, the vision of a shared responsibility for food/feed safety and the need to take steps towards closer interaction between operators and authorities has already been expressed [5–8].

Bearing the above background in mind, all participants involved in the food chain (manufacturers, authorities, laboratories, etc), along with the scientific community, have made substantial efforts in response to the challenging task of implementing methodologies to assess raw materials. On the one hand, the importance of the design of private and official food/feed sampling plans should be emphasised. First, sampling of bulk raw materials must be done so that representative sample can be obtained, which is crucial for accurately determining quality and safety parameters. In fact, Kuiper and Paoletti [9] argue that representative sampling is a must, so it needs to be considered as “a prerequisite equally important as the analytical methodology to ensure reliability of final results”.

Bulk food/feed sampling is typically described as a multistep process. First, a set of incremental samples are taken from the lot. These are combined to form an aggregate sample, which is then mass-reduced (possibly in several steps) to obtain the final analytical aliquot intended for laboratory analysis [10,11]. The minimisation of all errors that may arise during this process is of great relevance, and the Theory of Sampling (TOS) provides a fundamental framework to categorise and either eliminate or minimise these errors, thus ensuring sampling representativeness. To this end, a number of scientific studies addressing TOS and TOS-compliant standards are available in the literature [12–15]. Nonetheless, there is still much to be done to prevent sampling procedures of raw materials in bulk from being held back by financial factors, required resources or time constraints. In fact, they usually lead to over-simplistic solutions (e.g. grab sampling), excessive reduction of the sample volume (from several tons -lot; to a few grams -lab aliquot), with the risks that this entails for the lot-sample representativity,

loss of spatial information and sampling procedures that are not tied in with the nature of the product. Thus, improved and cost-effective methods and monitoring tools are needed, which would enable the development of more efficient sampling plans. Moreover, the implementation of fit-for-purpose protocols and real-time decision support systems are still lacking.

From an analytical point of view, Near-infrared spectroscopy (NIRS) can be a crucial asset for the design of food/feed quality assurance systems. NIRS has experienced a strong development over the last few years. Currently, it allows reliable analytical measurements to be made at different steps of the production process. The acceleration of technological innovation has also led to improved instrumentation, which has contributed to the use of NIRS for different purposes (from at-line to in-situ applications) [16]. In the light of this, published research has grown exponentially demonstrating NIRS abilities for analysing a wide variety of foods and feeds, including heterogeneous materials, under diverse conditions. Considering its potential impact, a number of industries have already integrated NIR-based quality-control schemes into their manufacturing processes successfully, mostly in the form of at-line applications [17].

This notwithstanding, the analysis of raw materials in bulk directly from the load of a transport vehicle has not been explored in depth yet. However, NIRS features a range of valuable qualities that can make it ideal for this task. Unlike traditional wet chemistry methods, this technique is capable of performing quantitative and qualitative analysis of intact sample within seconds, thus allowing the volume of sample analysed to be significantly increased. The application of NIRS to this task could resolve some of the constraints of the existing methodologies by performing rapid and cost-effective analysis. On this basis, research was recently initiated to implement a real-time NIRS-based monitoring system for the in-situ characterization of raw materials at delivery points of production plants [18]. The methodology proposed by the authors relies on using NIRS optical probes to sample loads of transport units of products in bulk. A subsequent geostatistical analysis of the observations succeeded in mapping the spatial distribution of key properties of PAPs. Despite this, the study did not fully investigate the sampling stage of the evaluation process, as well as its impact on the representation of the spatial surfaces of the PAP quality/safety attributes.

On the basis of this methodology, this paper aims at making a performance comparison of a set of sampling schemes through carrying out a simulation study. A further goal is to evaluate these plans concerning their ability to spatially characterize the quality and safety issues tested in PAP lots.

2 Materials and Methods

2.1 Samples and experimental design

A set of 8 lots of PAPs coming from a rendering plant was selected for this work from the ones used by Adame-Siles et al [18]. The set 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). These captured the variability of the available batches in terms of species composition.

Two sorts of quality and safety issues were simulated, on one hand the presence of high moisture content areas (issue A), which may act as indicators as they could lead to fungal growth or bacteriological problems, and on the other, adulteration or cross-contamination between products of different nature or category (issue B). A glass container was used to place and analyse each lot and type of issue (Figure 1). Issue A was induced in five lots in different ways. In each case, two layers at different depths were measured and a methacrylate sheet with 10 x 14 sampling points facilitated the positioning of the probe for analysis (first at layer A, and then the probe was inserted deeper into the sample at each point to reach layer B). Lots 1 (Figure 1B), 2 (Figure 1E), 3 (Figure 1F), 4 (Figure 1G) and 7 (Figure 1A) formed part of this particular case study, in which distribution and quantity of water were the two sources of variability among tests. Water was poured 1 day prior to analysis according to the distribution and amount of water that Figure 1 shows for each test. On the other hand, three further tests, involving Lots 1, 5, 8 and 10, addressed the evaluation of issue B. For this evaluation, tests were carried out making three different mixtures between lots, two of them between Lot 1 and Lot 5, varying their distribution (Figure 1C and Figure 1D), and the third one between Lot 8 and Lot 10 (Figure 1H). Measurements were taken for layer A in these tests. Moreover, it is important to point out that the tests performed aimed at exploring the limits of the methodology, which is why both issues were induced in localized areas.

2.2 Instrumentation and data analysis

2.2.1 Near-infrared Spectroscopy analysis

A reflection probe (Turbido, Solvias AG) (Figure 1) was interfaced to a Matrix-F FT-NIR instrument (834.2–2502.4 nm) to measure reflectance spectra in PAP lots according to the experimental design described above. The probe features a stainless-steel body of 12 mm in diameter with an insertable length of 300 mm, and its end has a sapphire window that is capable of illuminating a 1.5 mm diameter spot. The probe is composed of two optical fibers: one for

illumination and one for detection (each one of 600 μm core); while two fiber cables of 100 m in length connect the probe to the instrument.

In every test, a measurement was taken for each probe insertion point of the designed grid, and each spectrum was an average of 32 scans with a scanner velocity of 10 kHz and a resolution of 16 cm^{-1}. White reference measures were taken with a probe-specific Spectralon every set of 42 measurements (every 25-30 minutes approximately).

Within the context of a preliminary study [18], first, 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 standardization 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 for both moisture (issue A) and crude protein (issue B) constituents could be got at every sampling unit using the NIR reflection probe.

Software OPUS v7.0 (Bruker Optik) was used for spectral acquisition. WinISI v.1.50 (Infrasoft International), Matlab R2018a (The MathWorks Inc.) and PLS Toolbox (Eigenvector Research) were used for applying the NIRS prediction models.

2.2.2 Geostatistical analysis

The geostatistical study addressed the analysis of the spatial distributions of moisture (issue A) and crude protein (issue B). The methodology used provides for a two-stage assessment process for each test: (a) structural analysis; and (b) spatial estimation.

First, the structural analysis comprised both the exploratory data analysis and the variographic analysis. The spatial correlation analysis, which aims at describing the relationships between sampled points, was carried out following two steps: the estimation of the semi-variogram and its subsequent modelling. The semi-variogram measures the average dissimilarity between data separated by distance **h**, a vector commonly known as the lag distance or lag, and it is calculated as half the average squared difference between the components of data pairs [19]. The experimental variograms were computed for each test, obtaining omnidirectional and directional variograms (defined by 0, 45, 90 and 135º), in order to analyse the autocorrelation structure and the spatial pattern of the considered constituent in each case. The spherical and linear models were used for the fitting of the experimental variograms, hence obtaining continuous functions

that provide a model of spatial dependence, which is needed to compute a variogram value at unobserved locations [20].

The spatial estimation of the variables under consideration was addressed by using Kriging as the interpolation technique. This family of generalized least squares linear regression algorithms, characterized by being highly accurate and robust, manages to use the combination of weights and values at known locations (from the structural analysis) to estimate the value at unsampled locations, achieving reliable results. The regionalized variable of interest, $Z(\mathbf{u})$, is considered a random function, generically decomposed into a trend component, $m(\mathbf{u})$, and a residual component, $R(\mathbf{u})$. There exist different kinds of kriging estimators, which mainly differ in their treatments of the trend component. Ordinary kriging (OK) was used to interpolate the NIRS predictions for the parameters of interest in this paper. OK, which is one of the most commonly used variants of kriging, is based on the assumption that the mean is unknown and limits its stationarity to the local neighbourhood of the location **u** being estimated. Further details on the implementation of the geostatistical approach in PAPs and on OK theory and practice may be found in [18,21–25].

All geostatistical analyses were carried out in the R environment (version 3.4.3), including the exploratory data analysis, the variographic analysis, and the mapping of spatial estimations. The R package gstat was used to develop the methodology [26].

2.3 Simulation study

2.3.1 General procedure

NIRS measurements were taken once for every sampling location, i.e. obtaining a grid of 10x14 points for each test performed, which were used as analytical reference for this simulation study. Considering this population of N=140 units, the data were then sub-sampled by several procedures to give different sample sizes and distributions and evaluate the loss of information in each case.

A set of four different sampling intensities were tried (i=30, 20, 10 and 5% of the population) for every sampling design (defined in section 2.3.2), making a total of 16 sampling plans (Figure 2). These intensities were chosen so as to facilitate comparison by ensuring that all sampling designs achieved the same sample size n(i).

Moreover, after performing each sampling plan (defined by the sampling design and the sampling intensity), the resulting data sets were all assessed and treated for geostatistical analysis, so that both variographic analysis and subsequent kriging with the sub-sampled data were performed according to the procedure described before. Therefore, this led to obtaining

(for each sampling plan) spatial predictions from the geostatistical models that could be compared with the reference data.

All the sampling plans include some randomness, so in order to carry out a performance comparison of the methodology among sampling plans, a total of R=1000 simulation replications were computed for each one.

2.3.2 Sampling designs

- *Simple random sampling (SRS)*

Most standards, guidelines and regulations dealing with the control of loose feeds consider simple random sampling (SRS) as the selection process of incremental samples when sampling from bulk [10,27].

In this paper, different sets of n(i) units, from the sampled grid (N = 140 points, all having an analytical result available), were randomly selected according to every sampling intensity tried, i (Figure 3A).

- *Stratified random sampling (StRS)*

The study area was partitioned into 7 regions (Figure 3B). These areas were distributed so that the corners, the centre and the lateral walls of the container represent different strata. Therefore, this resulted in a total of 4 corner strata each composed of a grid of 5x4 sampling points, two lateral strata with a grid of 3x6 units respectively, and a central stratum with a grid of 4x6 sampling locations.

The sampling design within each stratum was SRS, with the selections in the different strata being made independently. Furthermore, the same number of units was selected from each stratum to form the final sample size n(i).

- *Cluster sampling (Clu)*

In this work, clusters were conceived to be formed of two sampling units adjacent on the vertical axis (Figure 3C). Consequently, an arrangement of 70 clusters over the study area was available.

A number of clusters, $k \leq n(i)/2 \in \aleph$, were randomly selected according to each sampling intensity i, so that the final sample size n(i) contained the same number of sampling units as the rest of the sampling designs; except for i=5%, in which 3 clusters were selected and therefore n(5%) was made of 6 units, one less than in all other cases.

- *Systematic sampling (Sys)*

Among the possible realizations of this type of sampling scheme, it was decided to simulate one in particular due to its high potential and efficiency to run multiple iterations within the study area. In this case, the approach was to simulate that the NIR fiber-optic probe would follow the moves of a knight over a PAP lot as if it were a chessboard.

Many methods tackle the issue of the knight's tour problem, trying to discover those possible paths or sequence of moves such that every square is only visited once. As a consequence, a number of algorithms and solutions have been found to that problem (brute-force approach, neural network computing, etc.). This paper implements an algorithm based on the Warnsdorff's rule. The algorithm starts by randomly selecting 1 unit of the available sampling grid (N=140), and then proceeds to the adjacent, unvisited unit with the least degree from which the probe will have the fewest onward moves, and so on (Figure 3D). As a stopping rule, the algorithm uses the sampling intensity considered, i.

2.3.3 Performance evaluation

As a measure to compare the performance of the different sampling plans, the Root Mean Square Error of Prediction (RMSEP) statistic was used in this paper:

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

where $y_{j,krig}$ are the predictions obtained for N by applying kriging from the sample selected by each sampling plan and $y_{j,NIR}$ are the NIR analytical values obtained for the sampled grid *N*.

Both the mean and the standard deviation (SD) of this statistic were calculated for the R=1000 simulations performed:

$$\mu = \frac{1}{R}\sum_{i=1}^{R}\mathrm{RMSEP}_i \tag{2}$$

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

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

All algorithms to carry out the simulations of the different sampling plans, together with the calculation of those statistics involved in the performance assessment, were developed in RStudio (v 1.1.1463).

3 Results and Discussion

3.1 Data preparation and analysis

Once all the experimental tests were performed, analysing each case with the reflection probe, the NIRS calibration equations were applied to every unit of the sampling grid (hereinafter referred to as "100% sampling", i.e. N=140 points) of each test. Hence, this made it possible to obtain a NIRS prediction for the parameter of interest, either moisture (issue A) or crude protein (issue B), at each sampling location of the case being analysed.

The first stage of the geostatistical study was then tackled to find the model that best describes the spatial pattern of the constituent in each test. A comprehensive description of the spatial behaviour of these parameters in tests with PAPs can be found in [18]. Overall, the results of the variographic analysis of the tests presented in this paper are in line with those reported in [18]. Thus, the semivariograms, computed from the "100% sampling" data sets in each case study, show differences between the spatial variation of both constituents. On the one hand, crude protein semivariograms generally result in plots with a steady increase in the semivariance with distance, as well as a discontinuity at the origin. In contrast, semivariograms for tests involving moisture adulteration display curves with lower semivariances than crude protein, which show a zero, or close to zero, intercept and rise until they reach a plateau. To infer spatial estimations at unobserved locations from the spatial autocorrelation analysis, theoretical functions are needed for the fitting of the semivariograms. For this purpose, linear and spherical models were generally used for crude protein and moisture constituents, respectively, as they were the mathematical models that provided the best fit in each case.

Based both on the NIRS predictions and the results of the variographic analyses, the simulation study proceeded to perform the different sampling plans, which were implemented based upon the arrangement of the intensities and designs described in the methodology. For each sampling strategy, corresponding to a specific sampling intensity together with the sampling design in question, a routine was run comprising a total of 1000 simulations. The specific sample data set

resulting from every simulation of each test was used as an input for spatial interpolation by ordinary kriging, thus producing a map based on the sample and interpolated values at all 140 points for use in equation (1).

3.2 Spatial distributions

Figure 4 and Figure 5 illustrate representative examples of the two kinds of issues in lots of PAPs tested in this work. While the former shows the spatial distributions for moisture of one test associated with issue A (Lot 1-layer B, Figure 1B), the latter displays crude protein spatial surfaces obtained for one of the tests related to issue B (Mixture of lots 1+5(2), Figure 1D). In both scenarios, the results are derived from the geostatistical study and application of ordinary kriging to the dataset composed of the NIRS predictions (for the constituent under study) at the sampled locations (defined by each sampling plan). Both figures present one random iteration for each sampling plan.

Figure 4 (top of picture) shows first the original distribution achieved by applying OK to the 100% sampling dataset, with the goal of allowing for comparison with the rest of sampling plans. In this plot, the moisture distribution reveals that higher values are concentrated in the corners and the central region of the investigated area, which in fact corresponds to the accumulation of water induced in this test (Figure 1B).

One of the clearest results that can be observed from the maps obtained, taking into account the range of sampling intensities tested, is that in general there is a decrease in the accuracy of mapping risk areas by moisture accumulation with sampling intensity. In this way, while sampling intensities of 30% and 20% frequently manage to portray most of the critical areas, both 10% and 5% intensities give rise to a significant loss of at least one or several of these regions in most cases.

On the other hand, when it comes to sampling designs, simple random sampling and stratified sampling achieve the best spatial surfaces, if the actual moisture distribution in the test performed is considered. Both sampling strategies succeed in depicting more faithfully the moisture concentration profile, if compared to 100% sampling, even when a 10% sampling intensity is applied. Conversely, neither cluster sampling nor systematic sampling were able to find accurate distributions, the former failing to reach reasonable results particularly at 10% and 5% of sampling intensity, whilst the latter did so at 20% and less.

As in the previous case, the crude protein distribution for the 100% sampling dataset of the test appears on the top of Figure 5. This map for the crude protein parameter pictures two distinct areas showing a different pattern from the rest, which is consistent with the composition of the mixture of lots performed in this test (Figure 1D).

The resulting spatial maps are also shown for this case according to the sampling plan used. Their analysis and discussion are analogous to the one previously made. As the sampling intensity decreases, there is a clear fall in the performance regardless of the sampling design, with 30% and 20% sampling intensities achieving better results than 10% and 5%.

Once again, comparing all the maps at a given sampling intensity, both simple random sampling and stratified sampling manage to represent most efficiently the actual crude protein distribution. This notwithstanding, cluster sampling remains close to them here at all cases but 5%, whereas systematic sampling fails again particularly at 10% and 5%.

In interpreting of these results, it should be taken into consideration that kriging techniques tend to overestimate small values and underestimate large values under certain circumstances, i.e. its estimates are less variable than the true values. Moreover, the larger the kriging variance of the estimates on average, the more apparent this smoothing effect becomes [22]. One reason for a larger kriging variance is because sample sites might be too sparse. This may have led to the highly smoothed representations observed when sampling intensities of 10% and 5% were used. A clear effect of this is a significant decrease in their ability to faithfully represent the actual scenario and, thereby, the spatial patterns eventually end up providing misleading information in these cases.

3.3 Spatial prediction

Following the qualitative analysis of the spatial distributions, the calculation of the RMSEP for all the tests performed, whose results are reported in Appendix A (Tables A1-A4), aimed at allowing a quantitative examination of the estimation error values. These tables show the mean and the SD of the RMSEP for every sampling protocol of each test, computed from the 1000 iterations carried out. All these errors were obtained comparing, in each point of the 100% sampling grid, constituent values estimated by OK and the true analytical value (NIRS prediction).

In order to facilitate a performance comparison, figures 6 and 7 graphically summarize the mean and the SD (showed as error bars) values of the RMSEP for the different case studies tested, categorizing the results by groups both according to the sampling strategy used and the sampling intensity in each case (layer A statistics for both issues A and B are shown in Figure 6, while results from layer B, i.e. only for issue A, appear in Figure 7). First, they illustrate how there is a clear negative correlation between the sampling intensity and the observed error of the estimates. This may be explained by considering the nature of kriging. As anticipated, this geostatistical technique consists of a multistep process, which is dependent upon the statistical relationships among the measured points. In this way, kriging not only considers the distance

between the observed locations and the prediction point but also the overall spatial arrangement of the sampled points, to finally derive a prediction. To this aim, the task of uncovering the model of statistical dependence, i.e. the spatial autocorrelation model, to be fitted to the observed points is therefore crucial. Consequently, bearing in mind both the distributions and the error values obtained, the aggregate effect of a smaller sample size and sparser sample sites (linked to lower sampling intensities) along with the aforementioned smoothing effect of kriging might have had a critical impact on the spatial dependence model, hindering its representativeness in these cases and leading to the loss of performance observed.

Overall, it can be noticed that stratified sampling outperforms the rest of the sampling protocols in most cases. In fact, if sampling intensities of 10% and 5% are considered, it is the sampling design accomplishing the lowest estimation error in all the tests performed. For higher intensities, however, stratified sampling along with systematic sampling prevail over the rest with the lowest value of RMSEP. It should also be noted that simple random sampling remains close to the performance of stratified sampling in general. Unlike the other designs, cluster sampling did not succeed in being as appropriate for the purpose of inferring the spatial distributions.

The ANOVA results (Table 2) showed that there was significant variation in RMSEP values among sampling intensities in all cases ($P<0.05$). This variation suggests the important role of the sampling intensity in the results. On the other hand, significant differences were also found in most tests among sampling designs. Nonetheless, the results revealed a few exceptions in this case. No statistically significant difference could be determined for moisture tests involving Lot 7 (layer A and B) and Lot 4 (layer A), along with the protein test from the mixture Lot 1+5. As can be noted from the tables provided, in all these cases, systematic sampling outperforms other designs at intensities of 30% and 20%, whilst stratified sampling does the same at 10% and 5%. Thus, as regards efficiency of the sampling design, no clear evidence was found in these cases to help decide one design over the other.

Based on these results, it should be underlined the importance of carrying out a preliminary thorough analysis of the raw material properties (heterogeneity, risk tolerance limits, etc.), as well as a profound study both of the Total Analytical Error (TAE) and the Total Sampling Error (TSE), which is deeply addressed by the Theory of Sampling (TOS), to draw conclusions from the implementation of the methodology. Thus, it is strongly encouraged to be careful, not only when modelling the spatial autocorrelation present in the lot, but also when interpreting the maps as a result of the representation of the kriged estimates. As can be appreciated from the results, a precise balance between an optimal sampling intensity (taking into account that lower intensities might prove to be insufficient to ensure reliable results) and a fit-for-purpose

sampling design is a necessary requirement to represent as faithfully as possible the spatial distribution of the constituents, avoiding misleading pictures.

4 Conclusions

NIR spectroscopy was used to perform sampling and analysis (as a single step) over a set of tests simulating two quality/safety issues. The results show that spatializing critical parameters of PAPs can provide decision makers with a useful, low-cost reference tool to identify patterns and risk areas non-compliant with quality and safety criteria. As a consequence, this might benefit the supplier-purchaser relationship by improving efficiency and transparency along the process.

The spatial analysis of reference constituents, and the estimation in non-sampled locations by geostatistical inferential methods allowed the mapping of crucial analytical constituents for the evaluation of lots of PAPs. This study has shown that the combination of NIRS and geostatistical analysis can be a powerful tool. Nevertheless, the results reveal that the accuracy of the distributions depends to a great extent on the sampling plan performed, i.e. both on the design and the level of intensity. Among the sampling designs tested, stratified sampling achieved the best results in most cases in both qualitative and quantitative terms, followed by simple random sampling and systematic sampling. In addition, sampling intensities of 10% and 5% of the total sampling grid tested proved to be mostly inefficient to represent the actual distributions.

It should be highlighted that the prediction results highly depend on the quality of the available observations and their spatial relationship. Therefore, the adoption of this methodology must necessarily rely on robust NIRS models together with an optimal sampling plan (striking a balance between strategy and intensity), to finally achieve reliable results. In this regard, further research should be carried out, for instance, to explore more efficient and fit-for-purpose sampling plans, as well as to perform validation tests in real conditions for evaluating products in bulk directly over the transport unit (trucks, trailers, containers, etc.)

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

[1] E. Commission, Regulation (EC) No 1069/2009 Laying down health rules as regards animal by-products and derived products not intended for human consumption and repealing Regulation (EC) No 1774/2002 (Animal by-products Regulation), Off. J. Eur. Union. L 300 (2009) 1–33.

[2] E. Commission, Commision Regulation (EU) 142/2011 implementing Regulation (EC) No 1069/2009 laying down health rules as regards animal by-products and derived products not intended for human consumption and implementing Council Directive 97/78/EC as regards certain samp, Off. J. Eur. Union. L 54 (2011) 1–254.

[3] J. Levic, J. Wallace, W. Oleszek, Animal by-products for feed: characteristics, European regulatory framework, and potential impacts on human and animal health and the environment, (2016) 189–202.

[4] M.J. De La Haba, A. Garrido-Varo, D.C. Pérez-Marín, J.E. Guerrero, Near infrared spectroscopy calibrations for quantifying the animal species in processed animal proteins, J. Near Infrared Spectrosc. 17 (2009) 109–118. doi:10.1255/jnirs.838.

[5] Directorate General for Health and Food Safety, Interim Overview Report - Audits of Official Controls in EU-Member States, 2017. doi:10.2875/202803.

[6] FEFAC, FEFAC vision on Feed Safety Management, Brussels, 2016. https://www.fefac.eu/files/67549.pdf.

[7] FEFAC, Annual Report 2017-2018, Brussels, 2018. https://www.fefac.eu/files/83625.pdf.

[8] European Commission, 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, Off. J. Eur. Union. 95 (2017) 1–142. http://eur-lex.europa.eu/legal-

content/EN/TXT/PDF/?uri=CELEX:32017R0625&rid=1.

[9] H. a. Kuiper, C. Paoletti, Food and Feed Safety Assessment: The Importance of Proper Sampling, J. AOAC Int. 98 (2015) 252–258. doi:10.5740/jaoacint.15-007.

[10] European Commission, Commission Regulation (EC) No 152/2009 laying down the methods of sampling and analysis for the official control of feed, Off. J. Eur. Union. L 54 (2009) 1–130.

[11] European Commission, Commission Regulation (EC) No 691/2013 of 19 July 2013 amending Regulation (EC) No 152/2009 as regards methods of sampling and analysis, Off. J. Eur. Union. L 197 (2013) 1–12.

[12] L. Petersen, P. Minkkinen, K.H. Esbensen, Representative sampling for reliable data analysis: Theory of Sampling, Chemom. Intell. Lab. Syst. 77 (2005) 261–277. doi:10.1016/j.chemolab.2004.09.013.

[13] P. Gy, Sampling of discrete materials—a new introduction to the theory of sampling: I. Qualitative approach, Chemom. Intell. Lab. Syst. 74 (2004) 7–24. doi:http://dx.doi.org/10.1016/j.chemolab.2004.05.012.

[14] K.H. Esbensen, P. Mortensen, Process Sampling (Theory of Sampling, TOS)-the Missing Link in Process Analytical Technology, in: Process Anal. Technol., 2nd Ed, Wiley, 2010: pp. 37–80.

[15] K.H. Esbensen, DS 3077. Representative Sampling - Horizontal Standard, Danish Standards, 2013. www.ds.dk.

[16] 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. 24 (2016) 47. doi:10.1255/jnirs.1192.

[17] V. Cortés, J. Blasco, N. Aleixos, S. Cubero, P. Talens, Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review, Trends Food Sci. Technol. 85 (2019) 138–148. doi:10.1016/j.tifs.2019.01.015.

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

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

[20] E. Gringarten, C. V Deutsch, Teacher's Aide Variogram Interpretation and Modeling 1, Math. Geol. 33 (2001).

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

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

[23] P. Goovaerts, Geostatistics for Natural Resources Evaluation, Oxford University Press, Oxford, 1997.

[24] N. Cressie, Statistics for Spatial Data, Wiley, New York (USA), 1991.

[25] D.E. Myers, Interpolation and estimation with spatially located data, Chemom. Intell. Lab. Syst. 11 (1991) 209–228. doi:10.1016/0169-7439(91)85001-6.

[26] E.J. Pebesma, Multivariable geostatistics in S: the gstat package, Comput. Geosci. 30 (2004) 683–691. doi:10.1016/j.cageo.2004.03.012.

[27] ISO, ISO 6497. Animal Feeding Stuffs - Sampling, International Organization for Standardization, Geneva, Switzerland, 2002.

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. Two-way ANOVA results (P values) for the sampling designs (SRS, Str, Clu, Sys) and intensities (i=30%, 20%, 15%, 5%) tested in Moisture (M) and Crude Protein (CP) tests.

Layer	Sampling	Lot 1 (M)	Lot 2 (M)	Lot 3 (M)	Lot 4 (M)
A	Intensities	$4.16x10^{-6}$	$1.62x10^{-6}$	$3.63x10^{-9}$	$2.14x10^{-5}$
	Designs	0.022	0	0.001	0.6
B	Intensities	$9.18x10^{-9}$	$2.52x10^{-5}$	$1.18x10^{-8}$	$3.93x10^{-7}$
	Designs	0.015	0.001	0.017	0.01
Layer	Sampling	Lot 7 (M)	Lot 1+5 (CP)	Lot 1+5(2) (CP)	Lot 8+10 (CP)
A	Intensities	$4.19x10^{-5}$	$2.02x10^{-6}$	$1.9x10^{-9}$	$3.49x10^{-11}$
	Designs	0.097	0.38	0.001	0.013
B	Intensities	$2.19x10^{-6}$	-	-	-
	Designs	0.08	-	-	-

Figures Caption

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

Figure 2. Procedure to perform the simulation study.

Figure 3. Sampling designs. (A) Simple Random, i=10% example. (B) Stratified. (C) Cluster. (D) Systematic, i= 20% example.

Figure 4. Spatial distributions Lot 1b (Moisture): 100%; Simple Random (1-4); Stratified (5-8); Cluster (9-12); Systematic (13-16).

Figure 5. Spatial distributions Lot 1+5(2) (Crude Protein): 100%; Simple Random (1-4); Stratified (5-8); Cluster (9-12); Systematic (13-16).

Figure 6. Estimation error values (mean and SD of RMSEP; layer A) for the Moisture (M) and Crude Protein (CP) case studies and the different sampling plans. Protocols: Cluster Sampling (Clu), Simple Random Sampling (SRS), Stratified Sampling (Str) and Systematic Sampling (Sys). Intensities: 5% (S05), 10% (S10), 20% (S20) and 30% (S30).

Figure 7. Estimation error values (mean and SD of RMSEP; layer B) for Moisture (M) case studies and the different sampling plans. Protocols: Cluster Sampling (Clu), Simple Random Sampling (SRS), Stratified Sampling (Str) and Systematic Sampling (Sys). Intensities: 5% (S05), 10% (S10), 20% (S20) and 30% (S30).

555

556

557

558

559

560

561

562 **APPENDIX A**

563

564 **Table A.1.** Estimation error values (RMSEP Mean and Standard Deviation). Simple Random
565 Sampling (SRS).

SRS	Sampling Intensity (%)	Lot 1		Lot 7		Lot 2		Lot 3		Lot 4		Lot 1+5		Lot 1+5(2)		Lot 8+10	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Layer A	30	0.149	0.016	0.230	0.025	0.098	0.006	0.122	0.010	0.131	0.011	1.456	0.104	1.585	0.081	1.524	0.074
	20	0.174	0.019	0.266	0.031	0.109	0.008	0.136	0.009	0.144	0.010	1.593	0.099	1.736	0.103	1.660	0.083
	10	0.220	0.032	0.327	0.044	0.125	0.013	0.155	0.014	0.161	0.011	1.809	0.151	1.958	0.147	1.818	0.105
	5	0.262	0.042	0.377	0.057	0.142	0.017	0.174	0.021	0.174	0.015	1.994	0.244	2.171	0.239	1.956	0.169
Layer B	30	0.271	0.026	0.230	0.037	0.093	0.006	0.095	0.011	0.101	0.008						
	20	0.315	0.031	0.277	0.046	0.103	0.008	0.110	0.014	0.113	0.009						
	10	0.370	0.037	0.362	0.054	0.118	0.012	0.139	0.019	0.130	0.010						
	5	0.415	0.045	0.423	0.054	0.132	0.016	0.167	0.026	0.144	0.014						

566

567 **Table A.2.** Estimation error values (RMSEP Mean and Standard Deviation). Stratified
568 Sampling (Str).

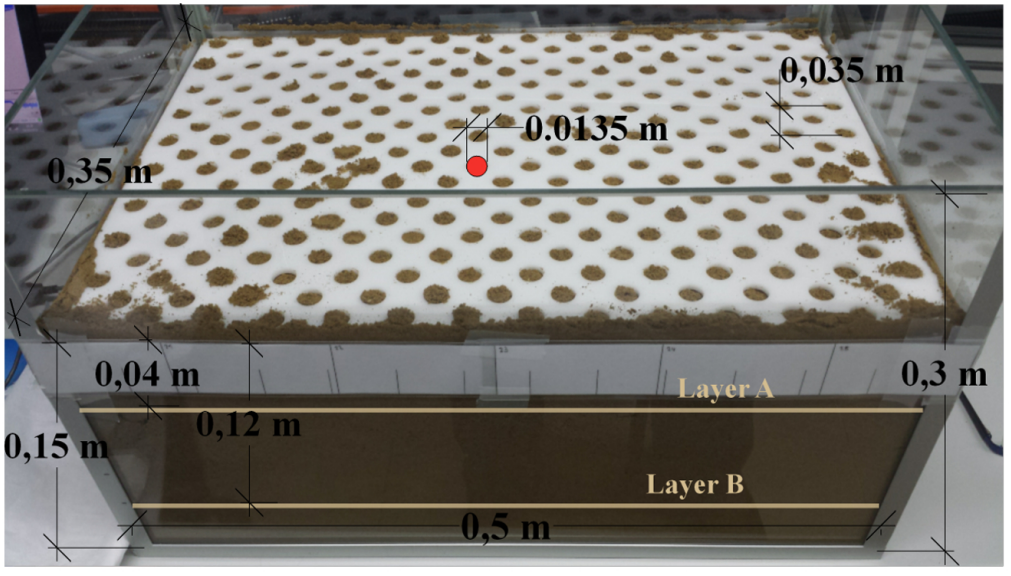
Str	Sampling Intensity (%)	Lot 1		Lot 7		Lot 2		Lot 3		Lot 4		Lot 1+5		Lot 1+5(2)		Lot 8+10	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Layer A	30	0.145	0.013	0.227	0.023	0.098	0.006	0.121	0.010	0.130	0.009	1.442	0.105	1.575	0.078	1.516	0.072
	20	0.169	0.016	0.261	0.023	0.108	0.007	0.135	0.009	0.143	0.010	1.585	0.088	1.723	0.091	1.650	0.072
	10	0.210	0.026	0.312	0.029	0.123	0.012	0.153	0.012	0.160	0.010	1.769	0.104	1.921	0.138	1.806	0.096
	5	0.246	0.034	0.358	0.048	0.137	0.016	0.169	0.018	0.172	0.015	1.941	0.191	2.127	0.202	1.929	0.151
Layer B	30	0.269	0.021	0.223	0.029	0.093	0.006	0.093	0.010	0.099	0.008						
	20	0.309	0.028	0.268	0.038	0.102	0.008	0.107	0.013	0.112	0.009						
	10	0.368	0.034	0.347	0.047	0.116	0.011	0.129	0.014	0.127	0.010						
	5	0.410	0.042	0.414	0.050	0.128	0.014	0.152	0.015	0.143	0.013						

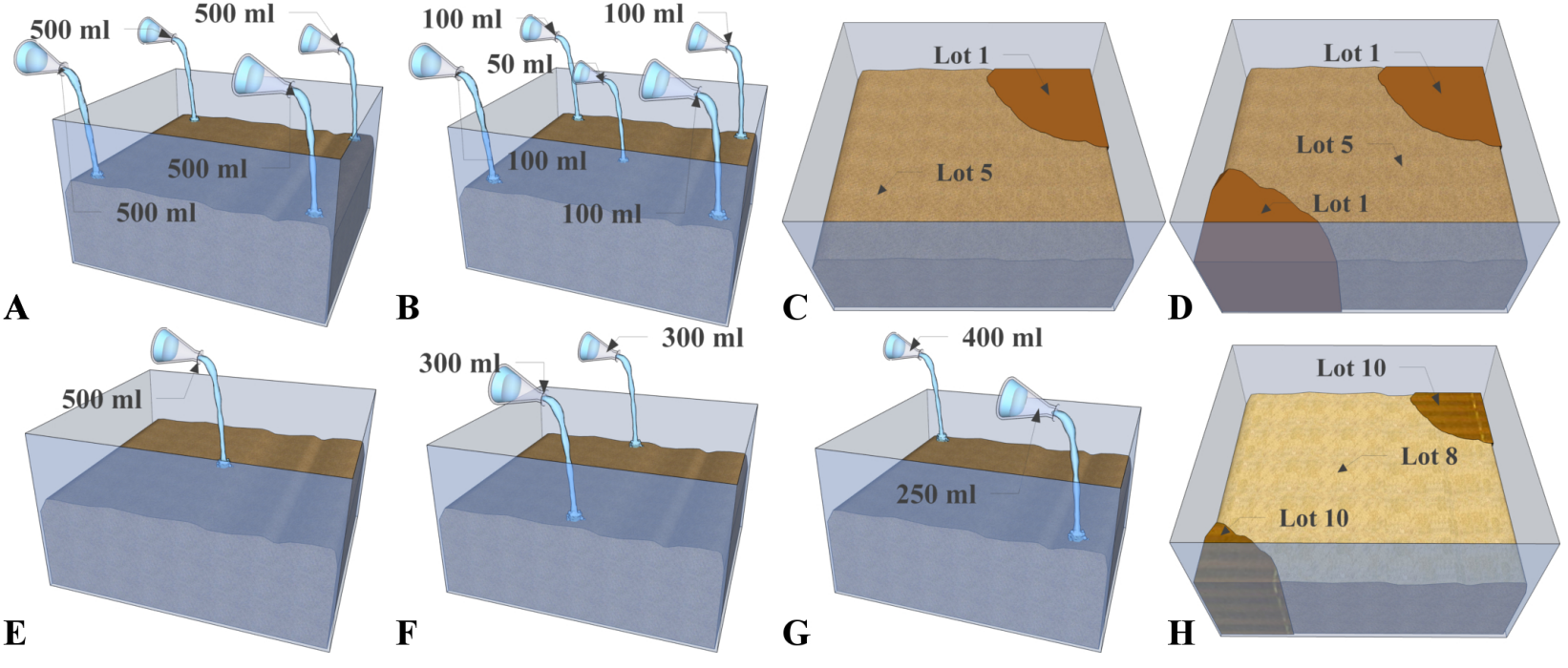
Table A.3. Estimation error values (RMSEP Mean and Standard Deviation). Cluster Sampling (Clu).

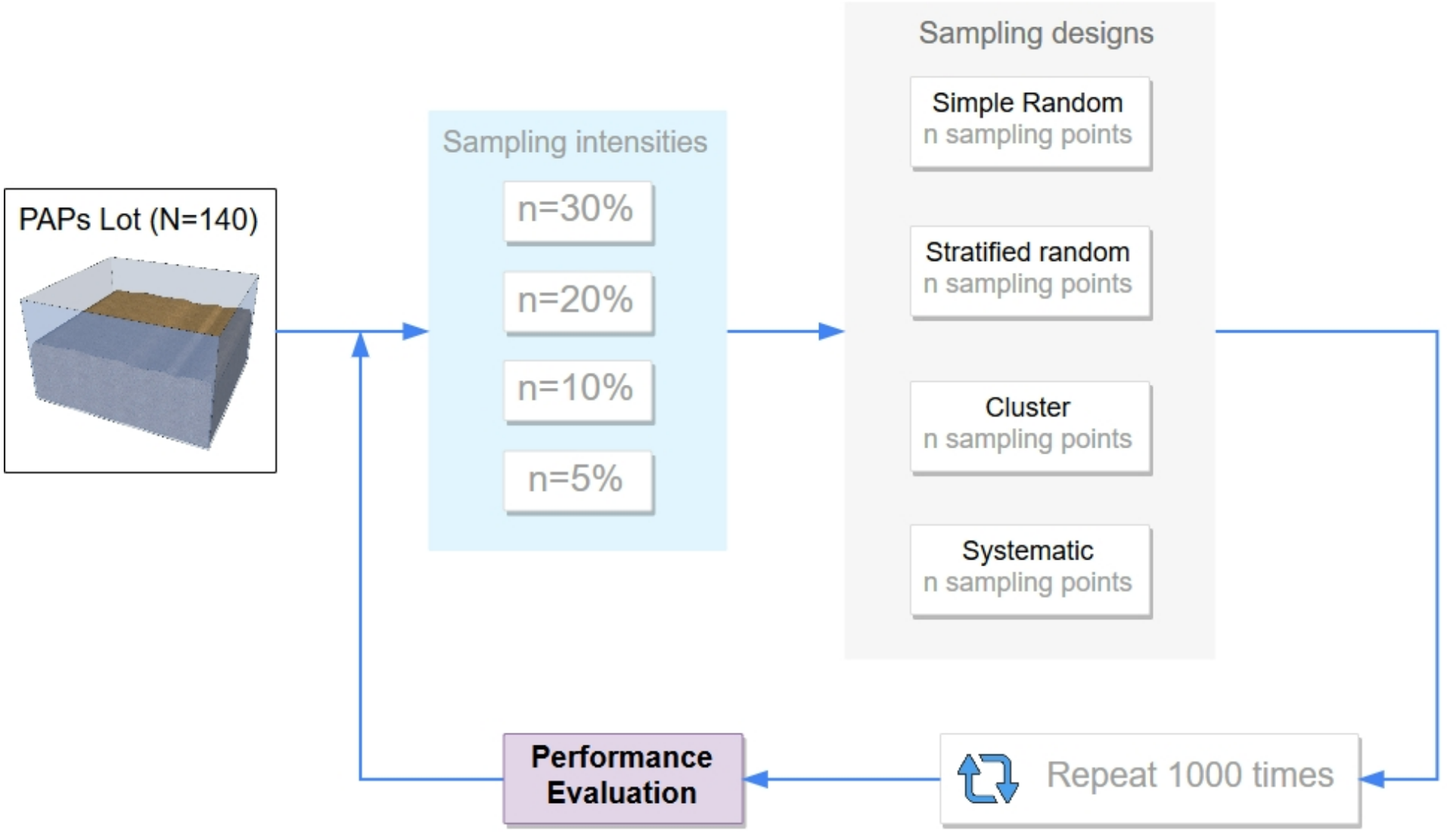
Clu	Sampling Intensity (%)	Lot 1		Lot 7		Lot 2		Lot 3		Lot 4		Lot 1+5		Lot 1+5(2)		Lot 8+10	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Layer A	30	0.170	0.023	0.260	0.034	0.100	0.008	0.124	0.011	0.133	0.011	1.468	0.121	1.606	0.099	1.533	0.077
	20	0.201	0.033	0.301	0.044	0.113	0.010	0.138	0.011	0.147	0.011	1.633	0.126	1.770	0.123	1.674	0.083
	10	0.249	0.042	0.362	0.059	0.133	0.016	0.162	0.019	0.165	0.012	1.873	0.200	2.022	0.168	1.847	0.118
	5	0.308	0.052	0.437	0.072	0.156	0.023	0.191	0.029	0.184	0.023	2.121	0.282	2.292	0.274	1.998	0.194
Layer B	30	0.290	0.034	0.259	0.049	0.096	0.008	0.103	0.016	0.105	0.009						
	20	0.334	0.038	0.317	0.058	0.107	0.010	0.121	0.018	0.118	0.010						
	10	0.391	0.039	0.398	0.057	0.124	0.014	0.156	0.025	0.136	0.012						
	5	0.443	0.055	0.480	0.073	0.144	0.021	0.185	0.032	0.156	0.023						

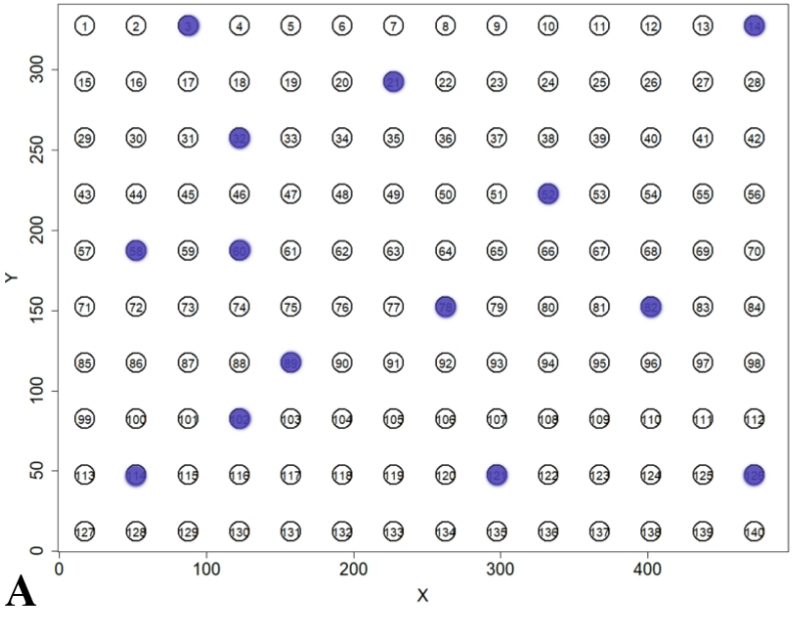
Table A.4. Estimation error values (RMSEP Mean and Standard Deviation). Systematic Sampling (Sys).

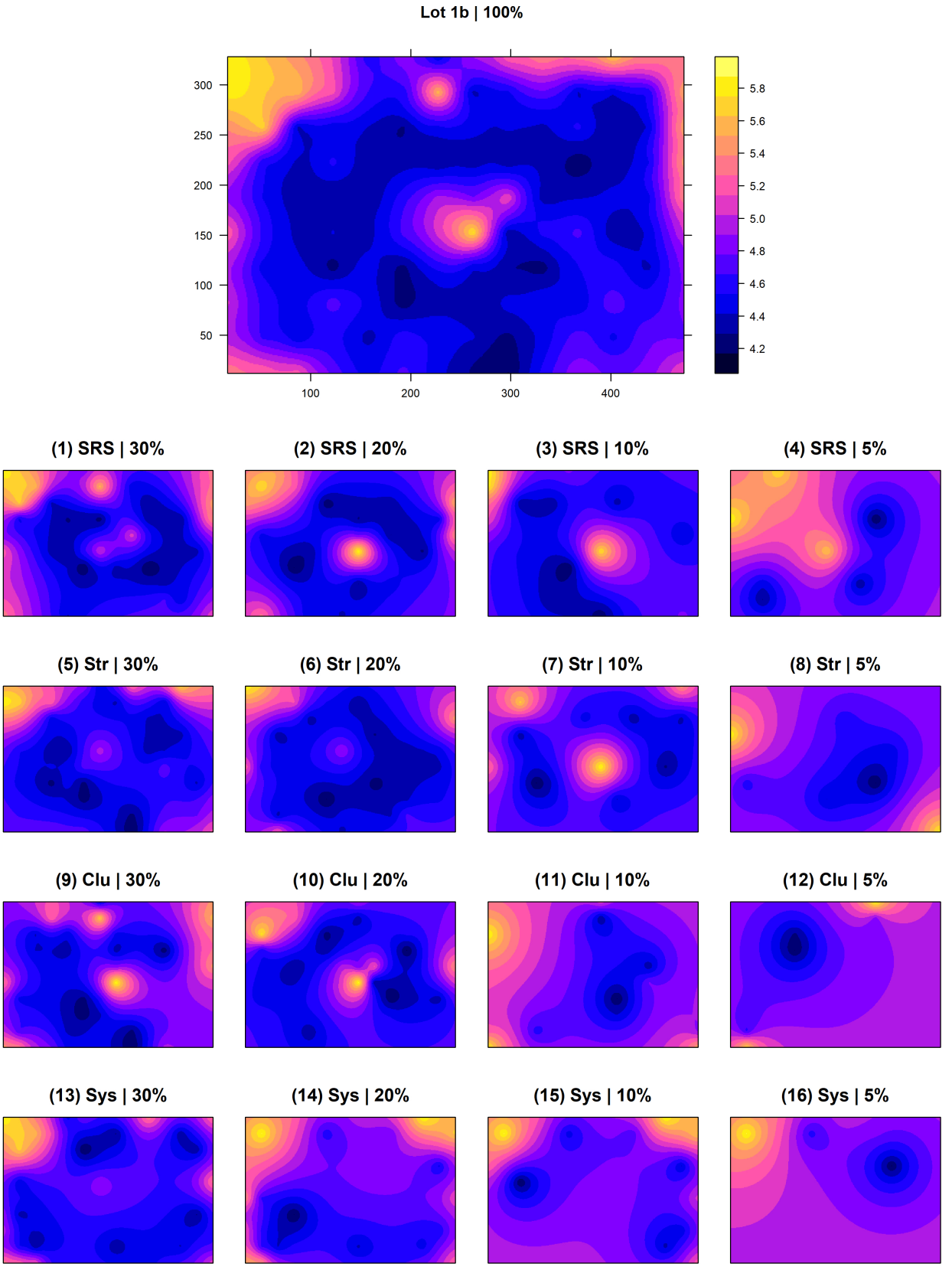
Sys	Sampling Intensity (%)	Lot 1		Lot 7		Lot 2		Lot 3		Lot 4		Lot 1+5		Lot 1+5(2)		Lot 8+10	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Layer A	30	0.153	0.014	0.189	0.013	0.121	0.017	0.113	0.007	0.136	0.010	1.347	0.131	1.655	0.134	1.484	0.051
	20	0.170	0.015	0.219	0.018	0.135	0.013	0.123	0.008	0.149	0.008	1.580	0.075	1.774	0.140	1.631	0.062
	10	0.255	0.079	0.338	0.086	0.146	0.015	0.157	0.026	0.176	0.022	1.952	0.142	2.079	0.217	1.817	0.102
	5	0.356	0.107	0.444	0.112	0.156	0.016	0.209	0.040	0.187	0.030	2.188	0.377	2.372	0.422	1.967	0.214
Layer B	30	0.270	0.020	0.200	0.016	0.119	0.014	0.095	0.006	0.101	0.005						
	20	0.310	0.024	0.241	0.030	0.129	0.012	0.106	0.007	0.108	0.006						
	10	0.395	0.063	0.391	0.087	0.136	0.013	0.162	0.036	0.136	0.015						
	5	0.455	0.088	0.494	0.118	0.141	0.015	0.189	0.035	0.152	0.016						

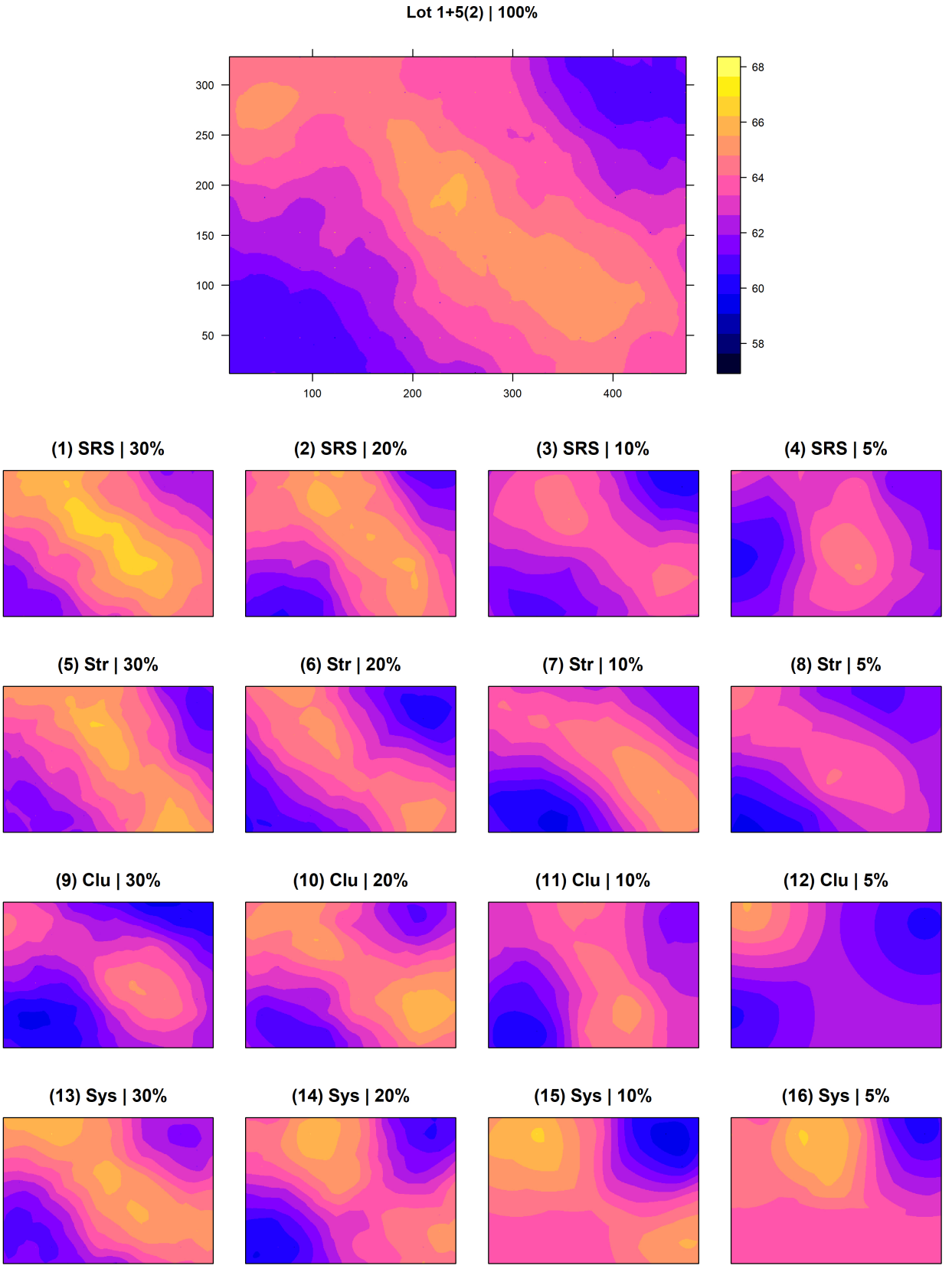

0,035 m
0.0135 m
0,35 m
0,04 m
0,12 m
0,15 m
Layer A
Layer B
0,3 m
0,5 m

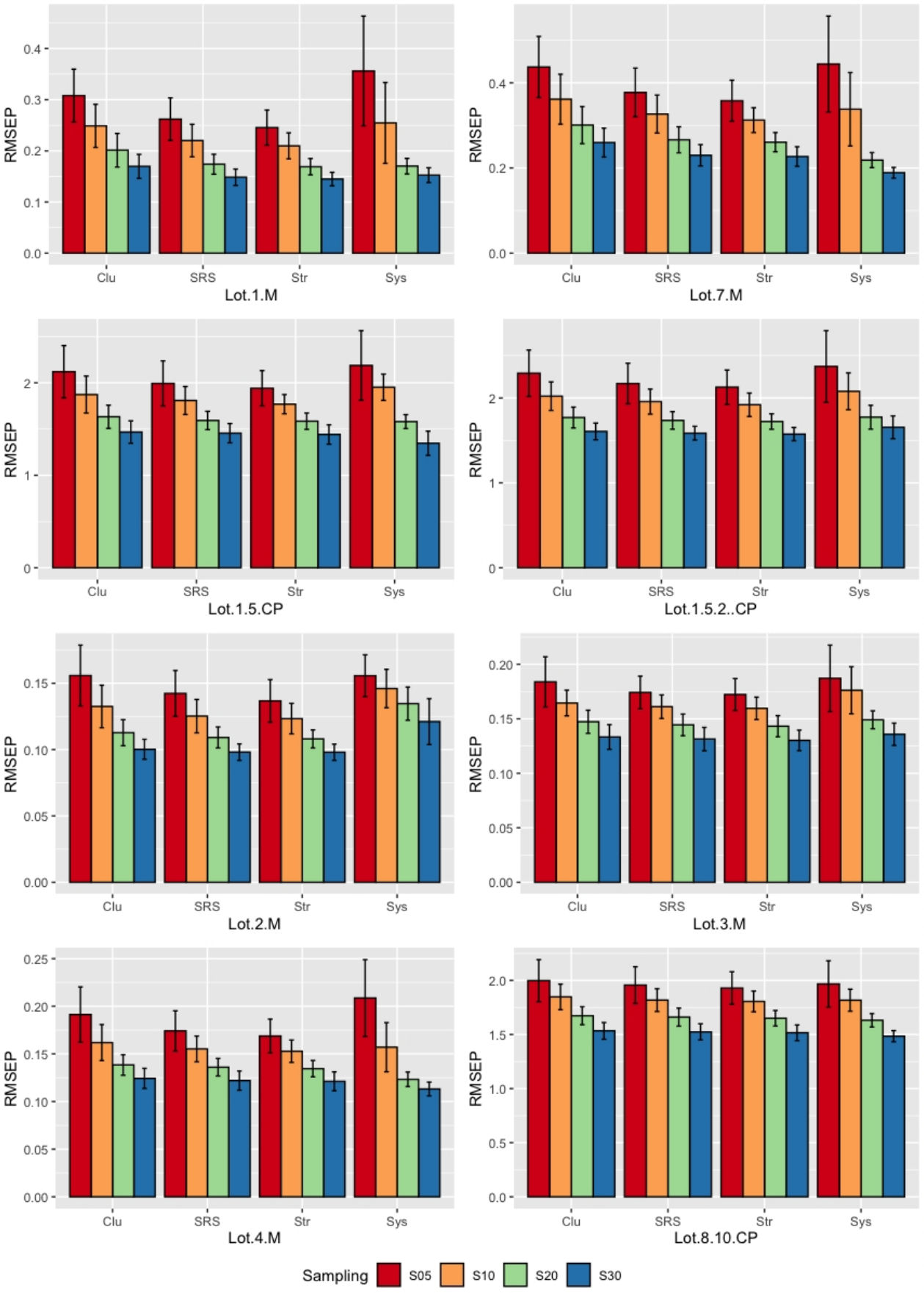

500 ml
500 ml
500 ml
500 ml
A
100 ml
100 ml
50 ml
100 ml
100 ml
B
Lot 1
Lot 5
C
Lot 1
Lot 5
Lot 1
D
500 ml
E
300 ml
300 ml
F
400 ml
250 ml
G
Lot 10
Lot 8
Lot 10
H

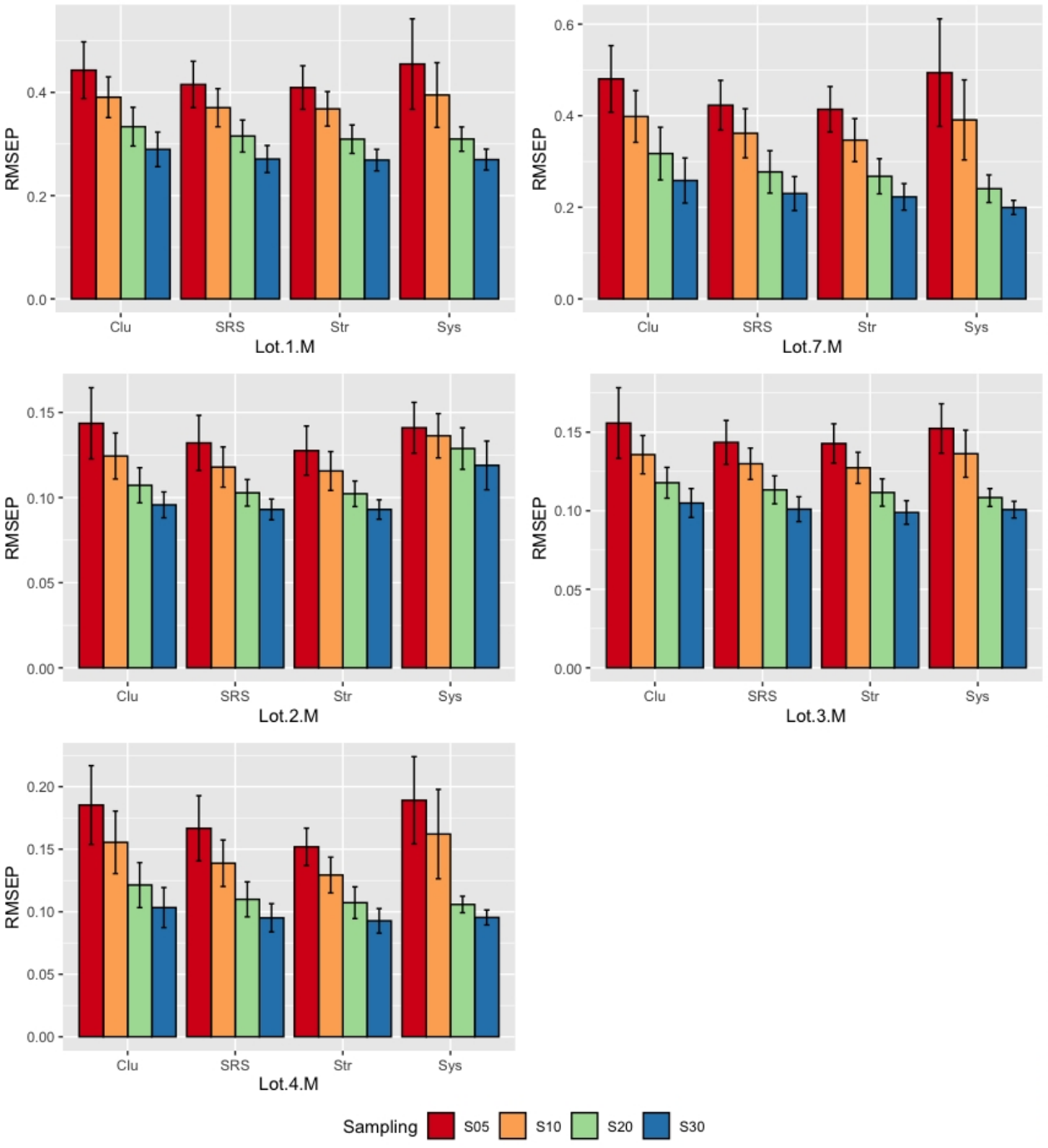

PAPs Lot (N=140)
Sampling intensities
n=30%
n=20%
n=10%
n=5%
Sampling designs
Simple Random
n sampling points
Stratified random
n sampling points
Cluster
n sampling points
Systematic
n sampling points
Repeat 1000 times
Performance Evaluation


A
B
C
D
X
Y


Lot 1b | 100%
300
250
200
150
100
50
100
200
300
400
5.8
5.6
5.4
5.2
5.0
4.8
4.6
4.4
4.2
(1) SRS | 30%
(2) SRS | 20%
(3) SRS | 10%
(4) SRS | 5%
(5) Str | 30%
(6) Str | 20%
(7) Str | 10%
(8) Str | 5%
(9) Clu | 30%
(10) Clu | 20%
(11) Clu | 10%
(12) Clu | 5%
(13) Sys | 30%
(14) Sys | 20%
(15) Sys | 10%
(16) Sys | 5%


Lot 1+5(2) | 100%
300
250
200
150
100
50
100
200
300
400
68
66
64
62
60
58
(1) SRS | 30%
(2) SRS | 20%
(3) SRS | 10%
(4) SRS | 5%
(5) Str | 30%
(6) Str | 20%
(7) Str | 10%
(8) Str | 5%
(9) Clu | 30%
(10) Clu | 20%
(11) Clu | 10%
(12) Clu | 5%
(13) Sys | 30%
(14) Sys | 20%
(15) Sys | 10%
(16) Sys | 5%


RMSEP
Clu
SRS
Str
Sys
Lot.1.M
Lot.7.M
Lot.1.5.CP
Lot.1.5.2..CP
Lot.2.M
Lot.3.M
Lot.4.M
Lot.8.10.CP
Sampling
S05
S10
S20
S30


RMSEP
0.0
0.2
0.4
Clu
SRS
Str
Sys
Lot.1.M
RMSEP
0.0
0.2
0.4
0.6
Clu
SRS
Str
Sys
Lot.7.M
RMSEP
0.00
0.05
0.10
0.15
Clu
SRS
Str
Sys
Lot.2.M
RMSEP
0.00
0.05
0.10
0.15
Clu
SRS
Str
Sys
Lot.3.M
RMSEP
0.00
0.05
0.10
0.15
0.20
Clu
SRS
Str
Sys
Lot.4.M
Sampling
S05
S10
S20
S30