Simulating kernel lot sampling:
the effect of heterogeneity on the detection of GMO contaminations

by

Claudia Paoletti[1]*, Marcello Donatelli[2], Simon Kay[1], Guy Van den Eede[1]

([1]) European Commission
Joint Research Centre, Institute for Health and Consumer Protection (JRC-IHCP)
Via E. Fermi, 1 – 21020 Ispra (VA), Italy

([2]) Research Institute for Industrial Crops (ISCI)
Via di Corticella, 133 - 40128 Bologna, Italy

(*) Corresponding author:
e-mail: claudia.paoletti@jrc.it

Seed Science and Technology

July 2002

(accepted)

Abstract

Guidelines defining kernel sampling strategies for quality analyses have been provisionally adopted for the detection of genetically modified (GM) contamination in kernel lots. However, these guidelines are not specific for GM material detection and are not intended for the sampling of non-uniform distributions, a probable situation with respect to the presence of GM material in kernel lots. An analysis of the problem of non-random distribution, through the investigation of the effectiveness of different sampling techniques in producing representative bulk samples, is presented.

The analysis is based on a two-step modelling procedure: 1) the kernel lot is created, and 2) the lot is sampled to produce a bulk sample. This allows the identification of optimal sampling techniques as a function of specific combinations of population characteristics. For each of 5 levels of GM impurity, varying between 0.1% and 2%, we investigated the effect of 5 levels of stratification (lot size=10^7 kernels). Our results indicate: 1) For every GM level, the higher the heterogeneity level, the more unstable the GM estimate becomes; even modest levels of stratification affect the stability of GM estimates. 2) As the number of increment samples increases, the coefficient of variation (CV) of the estimate decreases. Although the pattern of decrease remains similar across stratification levels, the estimated CV changes: with low levels of stratification, 50 samples are enough to obtain estimates with CV<10%. In the case of modest levels of stratification even 100 samples are not sufficient to maintain CV<10%. In case of strongly heterogeneous lots estimates based on 100 units have CVs around 50%. At the same time, the likelihood of false negative results increases significantly.

Introduction

Background and Legislation

Current legislation in the European Union (EU) requires the labelling of food products, raw or processed, which contain genetically modified (GM) material (EC, 1997; EC 1998; EC 2001). Despite being focused on the demands of the European consumer, this requirement has wider consequences affecting all major food producers, worldwide. Consequently, large scale testing and monitoring programs are being conceived and executed in order to check for compliance with the regulations involved. In addition, there is a strong interest in the schemes adopted for the sampling of food products lots, to ensure that testing of samples is reliable.

Various EU legislative tools regulate the adventitious presence of GM material in food products. This paper focuses on raw kernel bulk materials (seeds and grains) currently regulated under Directive 2001/18 (EC, 2001). Under this Directive, the presence of GMO in unprocessed (bulk) material must be indicated, even in the case of marginal proportion; however, there is no requirement to quantify the amount of GM material (i.e., qualitative evaluation only). No threshold for adventitious contamination is permitted.

Several guidelines defining kernel sampling strategies for quality and purity analyses are available from both standards authorities and national organisations. EU Member States have provisionally adopted such guidelines for the detection of GM contamination in kernel lots (Kay, 2001). However, the suitability of such sampling protocols for the detection of GM material contamination events is questionable for several reasons. First, these guidelines are not specific to the detection of low proportions of GM material, , with the exception of the USDA/GIPSA guideline (GIPSA, 2001). Second, few of the procedures are justified explicitly on the basis of a statistical assessment of the sampling problem. Third, most guidelines recognize that the procedures are not effective for the sampling of non-uniform distributions (Kay, 2001; Lischer, 2001; ISTA, 2001), a situation which may indeed be frequently the case with respect to the adventitious presence of GM material in grain and seed lots. Moreover, all these approaches include three important assumptions:

i) Statistically, the sampling procedure is considered as a single operational step, even though it may comprise a series of individual, independent actions;

ii) No consideration is made for the statistical implications of the typical clustered sampling of kernels (increments) applied in practice to the lot;

iii) The binomial distribution is used to calculate the precision of the estimate. This assumes that the target material is randomly distributed in the lot and, consequently, that each sampling unit (i.e., individual kernels) has an equal probability of being selected during the sampling process.

By contrast, we contend that these three assumptions present a major case for review. Although rarely acknowledged, the above approaches of sampling from a kernel lot imply a multi-stage procedure (Cochran, 1977) that reduces the lot to an analytical sample. Since the selection of kernels cannot - for practical reasons - be performed in an independent manner (i.e. simple random sampling of individual kernels), a bulk sample is first produced by systematically sampling increments from the lot. This bulk sample is then sampled (reduced) via a series of random processes to obtain a laboratory sample (Cross and Majors, 2000). Among these sampling steps applied in practice to any kernel lot, the first one (i.e. the creation of the bulk sample) is crucial because of the specific difficulties of ensuring its representative ness of the true lot properties.

The clustered (incremental) nature of the initial sampling step (lot-to-bulk) may significantly reduce the efficiency of the sampling scheme. The use of a few increment samples cannot provide the same statistical power reachable with thousands of independently sampled individual kernels (Cochran, 1977), a pre-condition to the correct application of the binomial distribution for the estimation of sampling error.

All these difficulties have, so far, frustrated efforts to harmonize and implement suitable sampling strategies for the detection of GM material in kernel lots. At the same time, it is often contended that sampling is the weakest link in the chain of quality control, and thus remains a priority to be addressed.

GM material distribution in large bulk lots

Based upon the assumption of random distribution of GM material within kernel lots, simple random sampling is most generally applied (Remund *et al.*, 2001). Thus, the mean, the standard deviation of the mean and both the producer and consumer risks have usually been estimated according to the binomial, the Poisson, or the hypergeometric distributions, as recommended in ISO standards (ISO, 1995).

However, the assumption of random distribution of impurities in kernel lots is likely to be false. A population (lot) of particulate material (kernels) is always affected by a certain amount of heterogeneity; according to Lischer (2001), the degree of heterogeneity of a lot can be described as a scalar function, for which the state of homogeneity is a limit case. Experimental confirmation of this theory comes from several studies investigating the degree of heterogeneity for several traits in large seed lots (Tattersfield, 1977; Jorgensen and Kristensen, 1990; Kruse and Steiner, 1995). Extensive heterogeneity has been reported for seed lots produced even with large-scale facilities, such as those for grass seed production in the Midwest US. The possible explanation offered by some authors is "*such seed lots are seldom if ever blended by state-of-the-art equipment, but are simply conditioned, bagged, and marketed*" (Copeland *et al.*, 1999).

Attempts to adapt the mathematical properties of the Poisson distribution to events of non-homogeneous GM material distributions have been made (Hübner *et al.*, 2001). However, such approaches may violate inherent assumptions (e.g. normal variance characteristics) required for the use of such tools. Lischer (2001a, b) has attempted to address these issues treating the sampling process as a series of error components that are estimated empirically, taking into account grouping of increments for various levels of heterogeneity. Nevertheless, the approach requires strong assumptions on the expected variability of the various parameters, which are difficult to confirm without experimental data. An additionally known difficulty is given by the expected low percentage of impure kernels per lot; the estimation of such low proportions is a classical problem in sampling theory, usually addressed by the assumption concerning the nature of the distribution of the material (Cochran, 1977). As noted above, however, knowledge of the true likely distribution of GM material in bulk lots is extremely impracticable and costly to obtain.

In summary, it is clear that assuming random distribution of GM material within all lots presents a serious risk, since it encourages solutions to sampling problems that overlook the issue of heterogeneity (Lischer, 2001). Given the recorded high likelihood of non-detectable strata of GM material in kernel lots, an approach that is free of the constraint linked to the assumption of random distribution is essential for the definition of reliable sampling techniques. In the light of all these concerns, a critical evaluation of sampling techniques on a variety of possible populations will provide scientifically reliable sampling recommendations.

The objective of this paper is to assess the effect of heterogeneity on the detection of low levels of kernel traits, such as GM material, in large grain or seed lots. Specifically, we address the problem of non-random distribution by investigating the effectiveness of different sampling rates without the constraints implicit in the assumption of normal (random) distribution.

Materials and Methods

The methodology used in this paper is based on a two-step modelling procedure: first, the kernel lot, from now on named "population" is created, and second, the population is sampled to produce a bulk sample. Creating populations permits the definition of different degrees of stratification of impurities, as opposed to full random distribution of single kernels. This permits the assessment of the efficiency of different sampling rates as a function of specific combinations of population characteristics. In order to test various population structures and different sampling rates , we developed a prototype program – KeSTE, *Kernel Sampling Technique Evaluation* – in Microsoft Excel® 2000. Results presented in this paper are derived from this experimental tool.

Step 1 – Lot simulation

The populations to be sampled are created through the definition of three parameters: the total number of kernels, the percentage of GM kernels, and the level of stratification of GM kernels (see below for detailed definitions).

In order to visualize the population and to impose different degrees of non-random distribution of GM impurities in the lot, we used a cube analogue to spatially define every lot (Figure 1). The cube is subdivided into portions, named "units", with a virtual grid. Each unit contains a specific number of kernels defined by the scale of the grid: the upper limit to the unit size is the size of the lot itself, whereas the lower limit is a unit that contains one kernel. Each unit can be spatially located by its virtual spatial coordinates.

Total number of kernels in the lot (i.e. lot size). According to ISO static sampling protocols (ISO, 1999), the maximum acceptable lot size is 500,000Kg. According to ISTA rules (ISTA Rules Supplement, 1999) the maximum lot size ranges between 10,000 and 20,000 Kg depending upon the species involved, with the exception of maize (40,000 Kg). Such an amount corresponds to approximately ~10^9 kernels in the case of maize (~10^8 according to ISTA rules). The size of our simulated lots was fixed at 10^7 kernels, corresponding to around 3,700Kg of maize kernels, due to restrictions imposed by the program prototype environment. Previous studies (review by Coster, 1993) have indicated that heterogeneity increases with lot size. Therefore, the limited size of our simulated lot (10^7 kernels) compared to the maximum lot size recommended by ISO static sampling protocols (10^9 kernels) provides a conservative estimate of the effects of heterogeneity on kernel lot sampling for accidental GM contamination detection.

Percentage of GM kernels present in the lot. Although in our model this parameter can assume any value larger than 0% (up to 100%, that is a completely contaminated lot), we investigated five possible contamination levels: 0.1%, 0.5%, 1%, 1.5% and 2%. The selection of these values is based on the expectation that the contamination is adventitious and only a low percentage of GM kernels should be present. The total number of kernels and the percentage of GM impurities allow the definition of the average ($\bar{x}_{GM}$) number of GM kernels in contaminated units by assigning an expected mean number of GM kernels per unit, + or – standard deviation (SD) = $\frac{1}{10}\bar{x}_{GM}$.

The total number of kernels present within a unit (i.e. the unit size) is function of the scale of the virtual grid. Results presented in this paper are based on populations with 2,744 units, each containing 3,644 kernels.

Level of stratification of GM kernels. This is a measure of the degree of spatial aggregation of GM kernels within any given lot (i.e. heterogeneity) observed at the level of the sampling unit. In KeSTE the stratification level can range from 100% (uniformity: all N units contain GM kernels) to $\frac{1}{N}$ % (maximum heterogeneity: all GM kernels are located in a single unit). For example, a stratification level of 25% indicates that the GM kernels present in the lot are distributed within ¼ of all the units. In essence, at the unit level, two strata can be present: a "null" stratum – units containing no GM kernels and, a "GM stratum" – units containing GM kernels in a proportion that is function of the level of kernel clustering defined above. Therefore, the smaller the GM stratum is as a fraction of the total lot, the higher the degree of heterogeneity in the population for a given GMO percentage.

The level of stratification, together with the total number of kernels and the percentage of GM impurities present in the population, permits the assignment of a different target number of GM kernels to each contaminated unit. During population creation, GM units can be either randomly allocated in the cube analogue, or clustered with a decreasing probability around one or more sets of coordinates. If we assume that GM kernels from a single source are located within a single unit, GM units randomly distributed within the cube represent multiple sources of GM contamination. In this preliminary study we investigated this latter option.

Step 2 - Sampling of the simulated Lots

Currently adopted sampling protocols entail the production of an initial bulk sample by sampling increments from the lot. Most guidelines do not provide exact recommendations for the number of increments to be sampled (Kay, 2001). However, we assume that current practice dictates a number of increments ranging between 15 and 33

(ISO, 1999). We assess the consequences of implementing different numbers of increment samples (units) on the overall population estimate.

First, in KeSTE we simulated increment samples by sampling entire units, that is all the kernels present in one location (unit) of the virtual lot (cube); a minimum of 5 units progressively increasing by one unit up to a total of 250 units were sampled from each lot. The units are systematically sampled to produce the bulk sample, with a sampling period that is a function of the total number of units present in the lot divided by the number of increments. In other words, we applied systematic sampling, which mimics the approach applied at fixed time intervals during lot off-loading or that of static lot sampling. The results from these simulations were then graphed for inspection and analysis.

Second, we sampled increments from KeSTE populations, again systematically, but this time repeatedly (500 times) at specific sampling rates (10, 20, 50, 100 increments) to estimate the Coefficient of Variation (CV) for different levels of heterogeneity. These results were tabulated.

As noted above, the number of kernels per unit for our simulation is fixed at 3,644 for all the populations we investigated. The guidelines defining kernel sampling strategies adopted for GM testing either do not provide specific recommendations for increment size (ISTA, ISO 13690, WHO/FAO), or recommend an increment size that can range between 0.2Kg and 5Kg depending upon the size of the lot to be sampled (USDA/GIPSA, ISO 542, EU Dir. 98/53, CEN) (Kay 2001). In the case of maize, 0.2Kg corresponds to approximately 550 kernels, whereas 5Kg of maize correspond to approximately 14,000 kernels; for soybeans, the range is 1,000 - 25,000 kernels. Our increment of 3,644 kernels is, therefore, well within the range of current practice.

Results and Discussion

Results are presented for five levels of GM contamination: 0.1%, 0.5%, 1%, 1.5%, and 2%. Table 1 summaries the simulation conditions explored in this paper. For each level of GM impurity we investigated the effect of five levels of stratification, 73%, 36%, 15%, 7%, and 4%, which represent different kernel lot heterogeneity scenarios. For each stratification level the expected mean number of GM kernels per unit is also shown in column three.

In Table 2 GM contamination estimates and their associated coefficients of variation (CV) are shown for two impurity levels: 0.1% and 1%. Although we explored a wide range of possible sampling rates (from 5 to 250 units per population), results based on 10, 20, 50 and 100 units per population are presented.

Our results clearly indicate that for any given sampling rate, the coefficient of variation of the GM estimate increases as the level of stratification decreases (Table 2). In other words, the smaller the proportion of units containing GM kernels (that is the higher the degree of heterogeneity), the more unstable the GM level estimate becomes; for example, for 50 increments the CV increases from 10% (73% stratification) to 72% (4% stratification). This is true for every GM contamination level we investigated, indeed due to the relatively stable levels of GM material per unit within the GM stratum, the variability of each estimate is not much affected by the GM impurity level in our simulation. Of note is that even modest levels of heterogeneity, such as stratification 36% (i.e., all GM material contained in ~ 1/3rd of all units), have a strong effect on the variability of GM level estimates (Table 2), with a CV of approximately 30% for 50 increments.

Another factor affecting the variability around GM estimates is the number of increments used to obtain the estimates themselves, i.e. the sampling rate. Specifically, our results indicate that as the number of sampled units (increments) increases, the variability

(CV) decreases. The pattern of decrease remains similar across stratification levels, although the estimated CV changes. This is true for all the impurity levels we investigated (Figures 2a to 2e). In slightly heterogeneous distributions (stratification 73%) 50 increments are enough to obtain an estimate with a CV < 10%. However, in populations with modest levels of stratification (36% or 15%) even 100 increments are not sufficient to maintain the CV below 10%, leading to strong possibility of wrong GM impurity estimates. In the case of highly heterogeneous populations (stratification 7% and 4%) the problem is even worse, and estimates based on 100 increments have CVs around 30% and 57%, respectively.

Our model reveals another possible detrimental consequence of insufficient sampling rate in case of stratified material distribution, that is the risk of obtaining false negative results. This risk applies to all the contamination rates we investigated in this study; although as an example we focus on a 1% contamination level. The increased probability of obtaining false negative results can be observed comparing Figure 2a-e at low sampling rates (less than 100 increment samples): as the heterogeneity of GM particles distribution within the population increases, the frequency of false negative estimates increases as well. In other words, the sampling approach lacks statistical power. This distortion of GM level estimate can be reduced and even practically eliminated by increasing the sampling rate to an appropriate level, which will depend upon the level of stratification present in the population. For example the risk of false negatives results in case of 1% GM contamination and a stratification level = 4% (figure 2e) is practically null if more than 100 increment samples are sampled, although in our example (population size of 10^7 kernels, 2744 units) this corresponds to >350,000 kernels and a sampling rate of approximately 1/27, much too high to be considered economically viable.

In summary, the approach we used to investigate non-uniform distributions is based on a two-step procedure that allows first, to create virtual kernel lots with different levels of impurities stratification and second, to test different sampling rates. With this approach a large number of kernel lots, with defined population characteristics, can be simulated without imposing any constraint on the distribution of GM kernels. As a result of this flexibility, the effects of different levels of heterogeneity and numbers of increment samples have been assessed, through simulation, on the accuracy and suitability of different sampling approaches for the detection of GM particles within kernel lots.

In particular, the analysis run allowed an initial screening of sampling schemes in the case of non-uniform (stratified) distributions of impurities in kernel lots. Our simulation results show that current procedures for the procurement of bulk samples, as stipulated in international guidelines, are sensitive to non-uniform distribution of impurities. In cases of heterogeneous GM material distribution, bulk samples have a high probability of not correctly representing the lot and/or even not containing any kernels with the trait under investigation (false-negative results). Indeed, even modest levels of stratification make the currently recommended number of increment samples ineffective in coping with this heterogeneity. Although the approach demonstrated in this paper is a tool to evaluate sampling schemes for the detection of GM material in large kernel lots, it could also be applied to purity testing for other types of kernel traits.

While the simulation of non-uniform distributions can be extended to deal with more complex scenarios of GM impurity (multiple sources, different strata densities, within-stratum GM kernel distribution modelling), these exploratory results issue a clear warning with respect to the unconditional acceptance of standardized sampling procedures in absence of the knowledge of GM material distribution in kernel lots.

Acknowledgements

The authors are grateful to Peter Lischer, Michael Kruse and Marc De Loose for useful and critical discussions on kernel sampling approaches.

References

Cochran, G. W. (1977). Sampling Techniques. 3rd Edition. J. Wiley & Sons Inc., NY, USA.

Copeland, L.O., Liu, H. and Schabenberger, O. (1999). Statistical aspects of seed testing. The Seed Technology Newsletter, 73, 1 pp20.

Coster, R. (1993). Seed lot size limitation as reflected in heterogeneity testing. *Seed Science and Technology*, **21**, 513-520.

Cross, R.F., Majors, R.E. (2000). Reducing sample size and obtaining representative samples. LC-GC Europe, October 2000: 727-733.

EC Directive N° 18. (2001). Directive of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC – Commission Declaration. *Official Journal* L106, 17/04/2001 pp1-39.

EC Regulation N° 1139. (1998). Council Regulation of 26 May 1998 concerning the compulsory indication of the labelling of certain foodstuffs produced from genetically modified organisms of particulars other than those provided for in Directive 79/112/EEC. *Official Journal* L159, 03/06/1998 pp4-7.

EC Regulation N° 258. (1997). Regulation of the European Parliament and of the Council of 27 January 1997 concerning novel foods and novel foods ingredients. *Official Journal* L043, 14/02/1997 pp1-6.

Hübner, P., Waiblinger H-U., Pietsch K. and Brodmann P. (2001). Validation of PCR methods for the quantification of genetically modified plants in food. Journal of AOAC International, **84**, 1855-1864.

ISO-13690. (1999). Cereals, pulses and milled products – Sampling of static batches. International Organization for Standardization, Genève, Swizerland.

ISO-2859/0. (1995). Sampling procedures for inspection by attributes. Part 0: Introduction to the ISO-2859 attribute sampling system. International Organization for Standardization, Genève, Swizerland.

ISO-2859/0. 1995. Sampling procedures for inspection by attributes. Part 0: Introduction to the ISO-2859 attribute sampling system. International Organization for Standardization, Genève, Swizerland.

ISO-542. (1990). Oilseeds – Sampling. International Organization for Standardization, Genève, Swizerland.

ISTA. (2001). Position paper concerning the increase or repeal of ISTA maximum seed lot size. *International Seed Testing Association News Bulletin*, **122**, 10-11.

ISTA. (1999). International Rules for Seed Testing. Supplement, *Seed Science and* Technology, **27**, 98-107.

Jorgensen, J. and Kristensen, K. (1990). Heterogeneity of grass seed lots. *Seed Science and Technology*, **18**, 515-523.

Kay, S. (2001). Comparison of sampling approaches for grain lots European Commission Report, Code EX21167/A, Joint Research Centre (JRC) Publication Office.

Kruse, M. and Steiner, A.M. (1995). Variation between seed lots as an estimate for the risk of heterogeneity with increasing ISTA maximum lot size. ISTA 24th Congress Copenhagen, Seed Symposium, Abstracts 21.

Lischer, P. (2001)a. Sampling procedures to determine the proportion of genetically modified organisms in raw materials. Part I: Correct sampling, good sampling practice. *Mitt. Lebensm. Hyg.*, **92**, 290-304.

Lischer, P. (2001)b. Sampling procedures to determine the proportion of genetically modified organisms in raw materials. Part II: Sampling from batches of grain. *Mitt. Lebensm. Hyg.*, **92**, 305-311.

Remund, K., Dixon D., Wright D., Holden L. (2001). Statistical considerations in seed purity testing for transgenic traits. *Seed Science Research*, **11**, 101-120.

Tattersfield, J.G. (1977). Further estimates of heterogeneity in seed lots. *Seed Science and Technology*, **5**, 443-450.

USDA-GIPSA. (2001). Sampling grains for the Detection of Biotech Grains. http://www.usda.gov/gipsa/

Figure 1: Spatial definition of a lot as a cube. The grid defines portions of the cube, the units. See text for details.

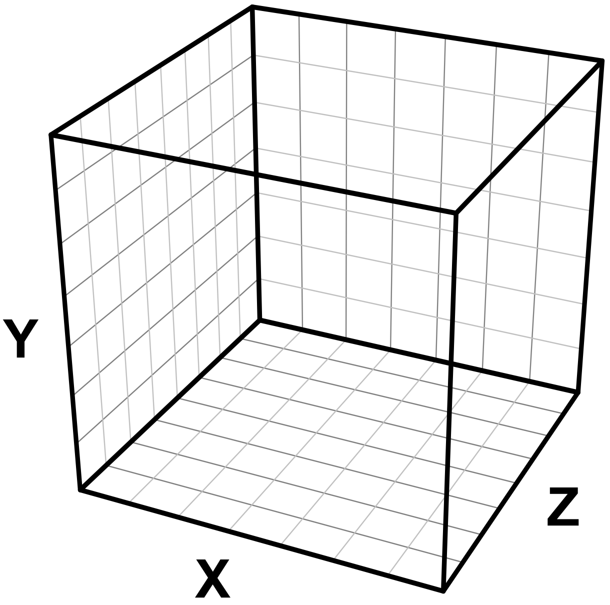

Table 1: Summary of simulation conditions.
Lot size=10^7 kernels; Units/Lot=2.744; Kernels/Unit=3.644

% GM kernels	Level of Stratification	Expected mean Number GM kernels/Unit
0.1	73%	5
	36%	10
	15%	25
	7%	50
	4%	100
0.5	73%	25
	36%	50
	15%	125
	7%	250
	4%	500
1	73%	50
	36%	100
	15%	250
	7%	500
	4%	1.000
1.5	73%	75
	36%	150
	15%	375
	7%	750
	4%	1.500
2	73%	100
	36%	200
	15%	500
	7%	1.000
	4%	2.000

Table 2: Coefficient of Variation (CV) associated to GM contamination estimates based on 10, 20, 50, and 100 increments, respectively. Results are shown for 5 levels of stratification and 2 contamination rates.

%GM kernels	Number of increments drown from the seed lot	Stratification				
		73%	36%	15%	7%	4%
0.1	10	21%	40%	76%	111%	150%
	20	13%	30%	60%	78%	111%
	50	10%	19%	33%	45%	72%
	100	7.8%	15%	22%	35%	53%
1	10	19%	40%	76%	108%	151%
	20	15%	29%	49%	87%	118%
	50	9.4%	18%	27%	47%	70%
	100	6.4%	15%	18%	31%	57%

Figure 2a: GM impurity estimates based on different numbers of units. GM impurity level=1%.

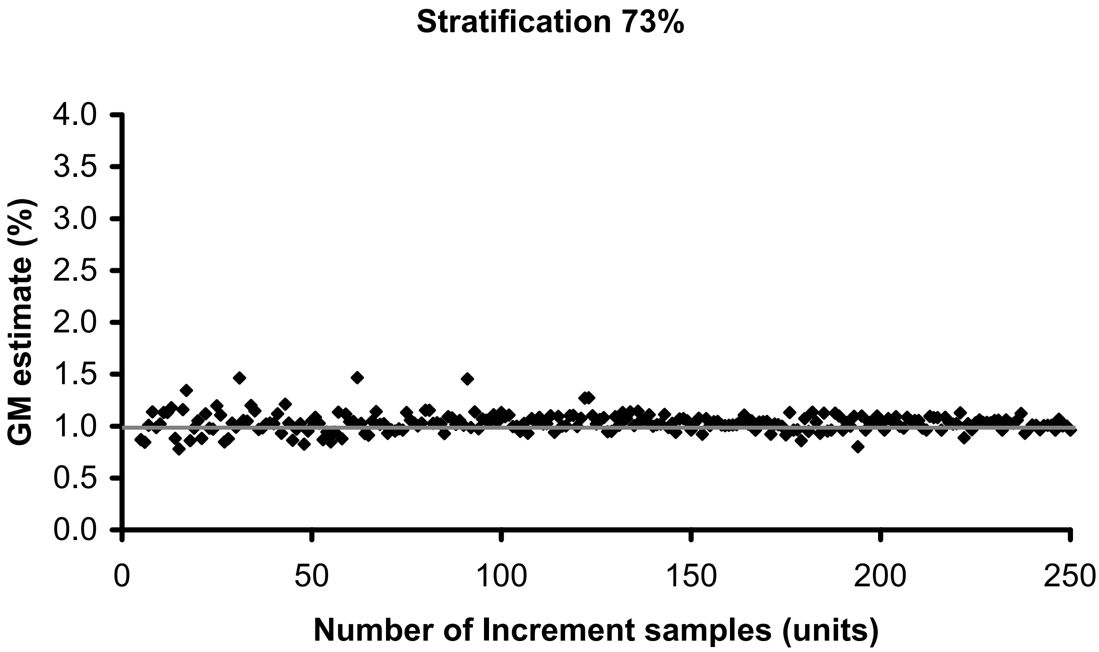

Figure 2b: Gm impurity estimate based on different numbers of units. GM impurity level=1%.

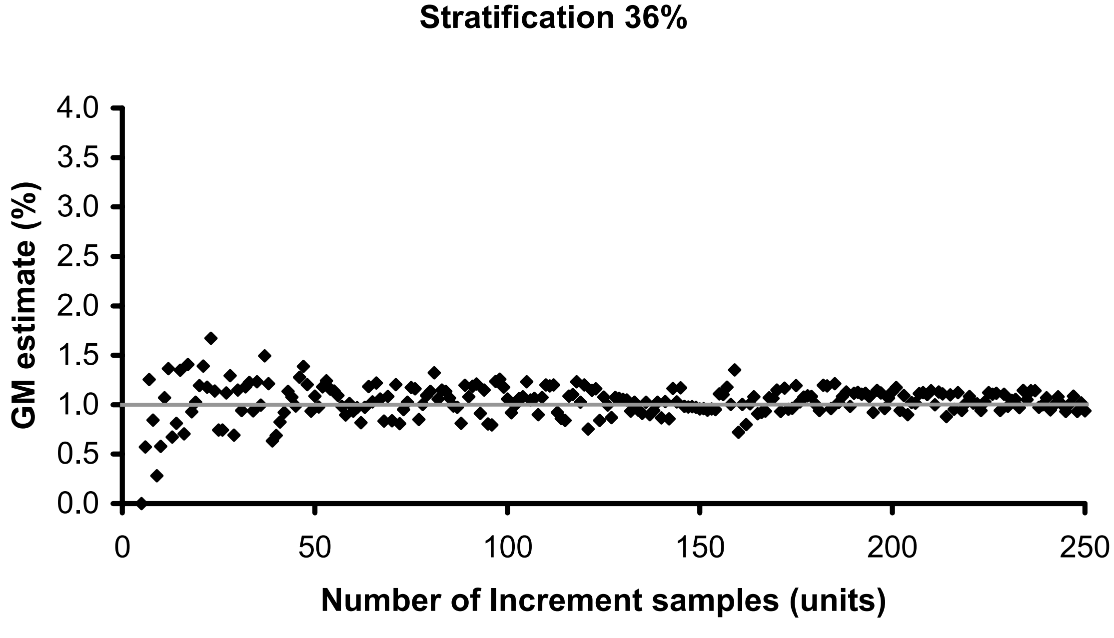

Figure 2c: GM impurity estimates based on different numbers of units. GM impurity level=1%.

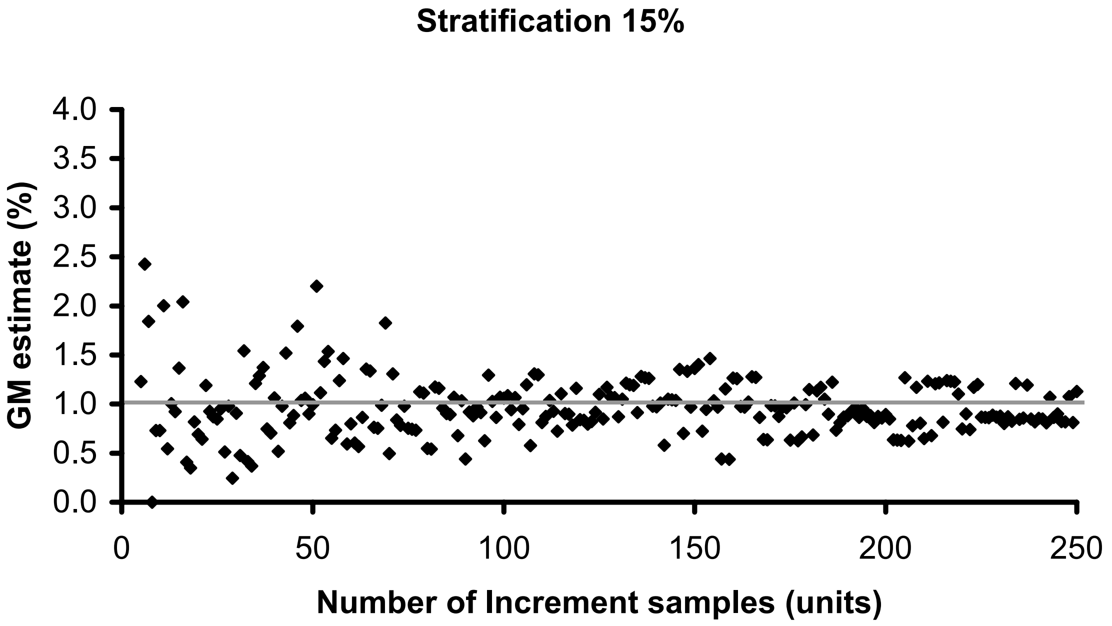

Figure 2d: GM impurity estimates based on different numbers of units. GM impurity level=1%.

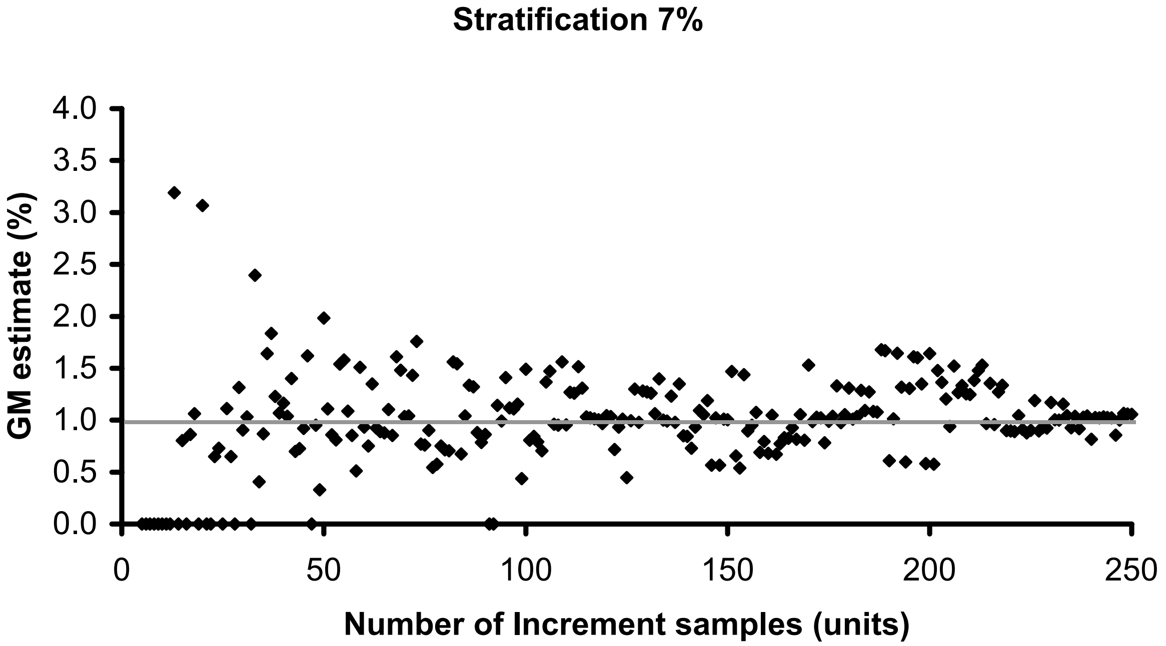

Figure 2e: GM impurity estimates based on different numbers of units. GM impurity level=1%.