RC25392 (WAT1306-063) June 24, 2013
Mathematics

IBM Research Report

Mycotoxin Testing in Food-Stock Lots

Ramesh Natarajan, Mary Helander, Janusz Marecki, Bonnie Ray
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

MYCOTOXIN TESTING IN FOOD-STOCK LOTS

Ramesh Natarajan, Mary Helander, Janusz Marecki and Bonnie Ray
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY, 10598.

June 19, 2013

Abstract

Mycotoxins are toxic metabolites which are a by-product of the fungal diseases that are prevalent in many commodity food crops such as grains, nuts, fruits and legumes. Due to food safety concerns, the maximum acceptable level of individual mycotoxins in various food stocks is subject to national and international regulations, and to specific food-industry norms. The development of acceptance-sampling protocols to ensure that individual suppliers and individual food-stock lots conform to these regulations and norms is therefore of considerable interest. In this report, we provide the relevant background on mycotoxin testing, and consider the need for further development of the theory and practice to include the following topics: (a) quantification of sampling and measurement uncertainty, (b) composite sampling and sub-sampling, (c) non-normal sampling distributions, and, (d) reduced-risk plans for variables sampling.

1 Introduction

The prevalence of fungal diseases in commodity food crops such as grains, nuts, fruits and legumes is attributable to the initial dispersal and attachment of spores from the external environment, followed by the spread of this infestation under conducive growth conditions. The onset and spread of fungal diseases can potentially occur at any point in the planting, growth, harvesting, storage and transportation stages of the agricultural processing pipeline for the food crop.

In turn, these fungal diseases lead to the production of specific toxic metabolites known as mycotoxins, whose presence above certain concentration levels in food stocks intended for human or animal consumption is of great concern from a food-safety perspective. Furthermore, even at apparently benign levels, the presence of mycotoxins detracts considerably from the nutritional value and retail quality of the contaminated food stock.

The presence of the precursor fungal molds and diseases, and the concomitant high levels of mycotoxins in food stocks, are both associated with a number of enabling factors. These factors include, the conditions of moisture, relative humidity, temperature, substrate and pH levels encountered during the growth, harvesting and storage of the crop; the associative growth with other fungi and micro-organisms in the contaminated food-stock lot; and, the stress and damage to the crop caused by weather, insects, fertilizer and mechanical harvesting. The high mycotoxin levels, once established, continue to persist even if the precursor fungal disease has been treated and eliminated from the food stock [42, 10].

The detection of high mycotoxin levels in food stocks is a worldwide problem, and hundreds of problematic mycotoxin varieties have been identified. Some of the prominent varieties that are frequently tested for include aflatoxin (AFL), deoxynivalenol (DON), zearalenone (ZEN), fuminosin (FUM) and ochratoxin (OCH). The precursor fungal molds that are often of greatest concern include Aspergillus, Penicillium and Fusarium [7]. In the U.S., the Food and Drug Administration (FDA) has established regulatory limits for the maximum concentration of individual mycotoxins in certain food stocks, and these limits typically range from parts per million (ppm) to parts per billion (ppb); for instance, the regulatory limit for AFL in grain stocks is 20 ppb [20]. In other cases, the FDA has established advisory limits; for instance, the recommended maximum limit is 1 ppm for DON [18], and 2-4 ppm for FUM [19], in food stocks intended for human consumption. The Grain Inspectors, Packers, Stockyards Association (GIPSA), which is an agency of the U.S. Department of Agriculture (USDA), provides reference material on fungal diseases and mycotoxins [21], and guidelines for mycotoxin testing in grain stocks (see `http://www.gipsa.usda.gov/GIPSA`). In addition, various commercial vendors, including Romer Labs (`www.romerlabs.com`), Neogen Corporation (`www.neogen.com`), VICAM (`www.vicam.com`), among others, who are providers of testing kits, sample preparation equipment, and laboratory analytical services, also provide prac-

tical guidance on many aspects of the sample preparation and analytical procedures for mycotoxin testing.

Individual processed-food manufacturers have also adopted stringent norms and practices at their manufacturing facilities to identify and eliminate any incoming food-stock lots with high mycotoxin levels. An important concern is that many mycotoxins are heat-resistant chemical compounds, whose levels cannot be reduced by the usual pasteurization and sterilization steps that are used to eliminate microbiological contaminants during food processing. The difficulties associated with mycotoxin remediation, therefore, make it especially critical to identify and reject any unacceptable food-stock lots before entry into the manufacturing facility. Consequently, manufacturing facilities that implement the Hazard Analysis Critical Control Point (HACCP) program, invariably locate a control point specifically for mycotoxin testing at the inlet grain-supply dock. (HACCP is a well-known and widely-adopted set of guidelines provided by the U.S. Department of Agriculture (USDA) for the monitoring of any biological, chemical, or physical agents at various stages during food processing steps, which in the absence of proper controls at these stages, may lead to the release of a contaminated product that causes illness or worse.)

Some of the salient issues that must be addressed by processed-food manufacturers for the testing, identification, and elimination of mycotoxin-contaminated food stocks are as follows:

First, the transfer of fungal spores from the environment to the crop which initiates the precursor fungal infestation can occur from the seeds, soil and air during planting and growth, or from the containers and equipment during storage and transportation. Similarly, the subsequent spread of the fungal disease and the concomitant increase in the mycotoxin levels can occur at any point in the agricultural processing pipeline where favorable growth conditions are encountered. For early identification, therefore, it is desirable to carry out the mycotoxin testing at all the locations in the field, storage and supply network. However, this supply network may not be fully known in some instances, and when known, the network locations may often be inaccessible, and in practice, it may only be feasible to perform the required testing at the entry point of the food-stock lots to the manufacturing facility.

Second, the detection of mycotoxin levels near the maximum acceptable limit requires careful sampling, sample preparation and analytical measurement, which must be performed by trained field personnel using specialized equipment and experimental controls, with further calibration and confirmatory testing under even more rigorous laboratory conditions. The size of the bulk material in the food-stock lots can range from hundreds to thousands of tons, with a wide variety of storage formats including bags, bushels and bins, and transportation formats including trucks, railroad cars and ships. Given the measurement requirements, and the size and variety of the bulk materials, it is indeed a challenge to develop a consistent and standardized protocol for obtaining compact, representative samples of the bulk material for mycotoxin measurements.

Third, a substantial portion of the *inter-lot* variations in the mycotoxin levels can be traced to

and linked with the conditions associated with the provenance of each individual food-stock lot. However, this provenance is rarely tracked in a reliable manner from farm to factory, even though it would enable the mycotoxin testing protocols to be customized to the actual or perceived risk. In addition, often, there may be significant *intra-lot* variations in the mycotoxin levels due to the localized nature of the fungal growth during lot storage and transportation (e.g., due to the pattern of moisture infiltration within a grain silo). These intra-lot variations may often be characterized by a non-uniform and clumped spatial distribution of the mycotoxin levels, and by a stratification between the grain and dockage components of the bulk material. The presence of such clumping and stratification effects is important for designing appropriate sampling and measurement protocols for mycotoxin testing.

Fourth, mycotoxin testing protocols typically comprise of three distinct stages, viz., the collection of sample-increments of the bulk material which are combined into a composite sample, the homogenization and reduction of this composite sample, and the analytical measurement on the reduced sub-samples. Each of these three stages will contribute to the uncertainty in the final mycotoxin measurement. However, the magnitude of these individual sources of uncertainty is often not known, and there are few guidelines for reducing the measurement error, keeping in mind the relative time and cost for each distinct stage in the overall testing protocols.

The outline of this report is as follows. Section 2 provides background information on the typical mycotoxin measurement and testing protocols, along with a brief overview of the relevant acceptance-sampling theory in this context. Section 3 considers certain additional topics, such as, sampling and measurement uncertainty, composite sampling, non-normal sampling distributions, and risk-based criteria for acceptance-sampling methods; in our opinion, these topics are of considerable relevance for the further development of mycotoxin testing. Section 4 concludes with a few observations on the implementation of mycotoxin testing programs.

2 Background

This section describes a typical set of protocols for mycotoxin testing, and discusses the relevant acceptance-sampling theory in this context. Our emphasis here is on those aspects of the theory that motivate the topics that are considered further in Section 3. A broad framework for mycotoxin testing is provided by the theory of sampling for bulk particulate materials [38, 44], which is concerned with the following two issues that are common to many different situations involving bulk sampling and quality measurement: first, obtaining representative samples for physical or analytical measurements, and second, identifying all the sources of uncertainty in the final measurement accuracy. A good overview of acceptance-sampling theory for measurements in bulk materials can be found in the papers by Duncan [11, 12, 13]. The specific case of mycotoxin testing within this

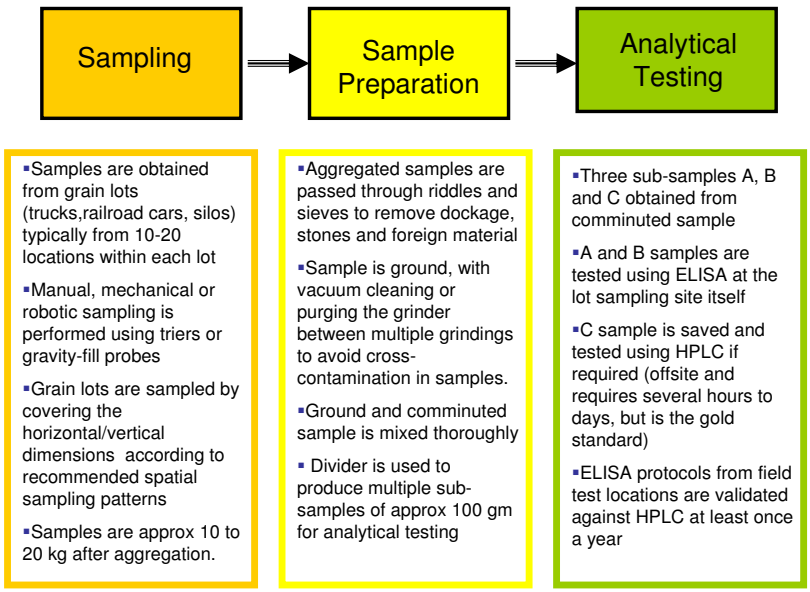

Figure 1: Typical Set of Measurement Protocols for Mycotoxins

framework has also been described in several places, most notably in several papers by Whitaker and co-authors [25, 26, 27, 46].

A wide range of applications in analytical testing [17] and microbiological testing [4] share the important characteristic that the sample collection step is far less expensive and time-consuming compared to the sample preparation and analytical measurement step. For this class of applications, therefore, it is costly and impractical to carry out a separate measurement for each individual sample increment, and as a practical alternative, the individual sample increments are aggregated and blended into a single composite sample, which is then reduced and sub-sampled for the analytical measurement. The use of sample composites in this way significantly reduces the time and cost of testing, but does not directly provide an estimate the intra-lot sampling variability, which is required in order to be able to use these measurements in an acceptance-sampling protocol.

2.1 Overview of Measurement Protocols

Figure 1 is a schematic of the different stages for implementation of mycotoxin testing in a foodstock lot such as a truck load of shelled corn. The three stages depicted here are, respectively, stage I – the collection of sample increments from the bulk material to form a sample aggregate; stage II – the homogenization and reduction of the resulting sample aggregate; and stage III – the subsampling and analytical measurement step. In general, the exact details of each stage will depend on various factors, including the specific mycotoxin being tested, the nature, size and storage format

of the food-stock lot, and the time, cost and accuracy requirements of the analytical measurement protocols.

In the bulk sampling stage, or Stage I, a representative aggregate sample is obtained by extracting individual sample increments that are roughly of sub-kilogram size from different spatial locations within the food-stock lot. These individual sample increments are obtained using specialized sampling probes when the bulk material is stationary, or using continuous extractors when the bulk material is being transferred; in either case, the resulting sample increments are chosen to be representative of the bulk material in terms of the weight, composition and particulate strata. It is often the case that the mycotoxin contamination may often be found in clumps and clusters, or preferentially segregated into certain particulate strata, rather than uniformly or randomly distributed throughout the bulk material [23, 46]. This phenomenon is a result of the selective spread and growth of the precursor fungal disease due to the local conditions of moisture infiltration and saturation within the bulk material, and the relative inability of certain constituents within the crop to withstand the fungal infestation. As a result, in this case, many of the sample increments extracted from the bulk material may have no measurable mycotoxin contamination, while a few sample increments may have exceedingly high levels. The resulting increase in sample variability, which is a consequence of the clumping and preferential segregation, may require many more sample increments to be collected in order to obtain a satisfactory acceptance-sampling plan.

In the sample compositing stage, or Stage II, the aggregate sample obtained in Step I is blended and homogenized, often using special-purpose machines which intersperse the grinding and milling steps with frequent mixing and redistribution. In particular, frequent mixing may be needed to reduce the variations that can arise due to the preferential segregation of the mycotoxin contaminant within certain sieve grade components of the milled aggregate. For example, in one study [2], the finest grade of the milled sample contained a disproportionate fraction of the mycotoxin contaminant, and this grade tended to be retained as a residue that jammed the parts of the grinding device, resulting in large variations in the measured mycotoxin levels unless this residue was also removed and well mixed into the final milled aggregate.

In the analytical measurement stage, or Stage III, a few sub-samples of roughly hundred grams in size are obtained from the homogenized sample aggregate in Stage II, and certain extraction and analytical techniques are used to measure the individual mycotoxin levels. Although omnibus tests are not widely used, certain subclasses of mycotoxins have very similar characteristics and may be jointly measured in a single test. A recent review [40] discusses many of the analytical techniques used for mycotoxin measurement, such as Thin Layer Chromatography (TLC), High-Pressure Liquid Chromatography (HPLC), and Enzyme-Linked Immunosorbent Assay (ELISA), for which standardized protocols and certified testing kits are widely available. The most widely-used technique for onsite mycotoxin measurement is ELISA because of its speed and simplicity,

but the HPLC technique is often regarded by practitioners as the "gold standard" for calibration and confirmatory testing.

Finally, based on these measurements, certain decisions are made on whether the food-stock lot should be accepted, rejected, or tested further. The simplest protocol, which is often the default in food safety applications [24], is to reject the food-stock lot if even a single measurement exceeds the maximum allowable limit. However, this approach ignores the totality of the distribution of the measurement values, which can be very useful for carefully estimating and reducing the acceptance-sampling risks, as described in detail in Section 2.2. In fact, the uncertainty and distribution of the measurement values plays a key role in the acceptance-sampling theory, and a careful review of the relevant protocols may often suggest appropriate, cost-effective approaches for reducing this uncertainty; these approaches may include the use of multiple aggregate samples, more sample increments per aggregate sample, more careful compositing, homogenization and reduction of the aggregate samples, replication of the various steps in the measurement protocols, and frequent calibration and confirmatory testing of the onsite measurement techniques.

2.2 Overview of Acceptance Sampling

In industrial quality testing, acceptance-sampling methods are typically used when exhaustive testing is not possible due to excessive time and cost, and because the sampled contents may be rendered unusable by the testing process itself [41]. In these situations, the accept/reject decision on the lot quality must be made after only testing a small fraction of the lot contents. The tacit assumption is that the lot contents have the same origin and provenance, and that a statistically-representative sample is chosen for quality testing.

The performance of an acceptance-sampling plan is given by its operating characteristic (OC) curve, which is shown schematically in Figure 2. Typically, the acceptance-sampling plan is chosen to provide a fixed performance at two specific values of the lot quality variable (we assume here without loss of generality that large values of this variable correspond to poorer quality); viz., a lower value denoted by A, for which the lot should be accepted with some high probability $(1-\alpha)$, and an upper value denoted by R, for which the lot should be rejected with some high probability $(1-\beta)$ (with $\alpha+\beta \leq 1$). These two values A and R are referred to as the "acceptable quality level" (or AQL), and the "rejectable quality level" (or RQL), respectively. Similarly, α is referred to as the "producer's risk" which is the probability that a lot at the AQL will be rejected by the sampling plan, while β is referred to as the "consumer's risk," which is the probability that a lot at the RQL will be accepted by the sampling plan. The special case $A = R$ often arises in practice, for which the acceptance sampling plan should ideally lead to an OC curve that drops sharply at $A = R$; however, since $\alpha+\beta = 1$ in this case, it is not possible to simultaneously reduce the producer's and consumer's risk independently.

The remaining discussion in this section is limited to a a single-stage, non-rectified, acceptance-

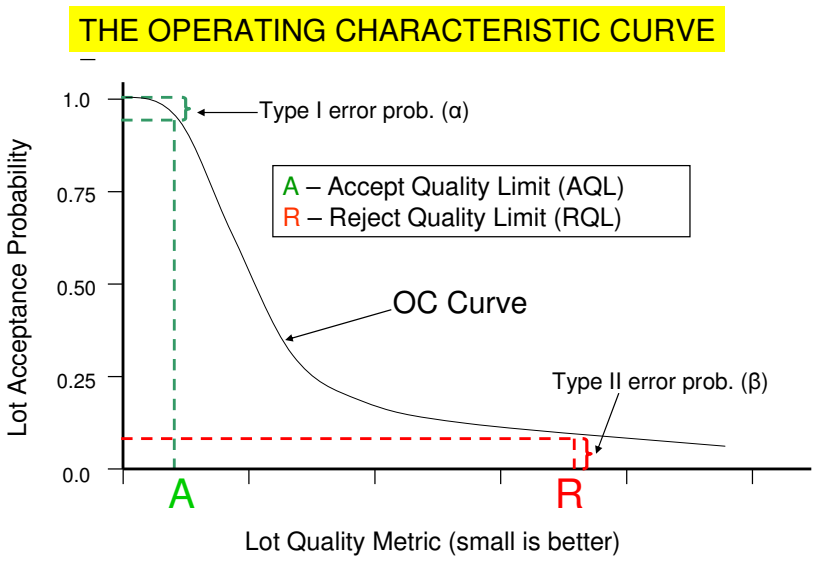

Figure 2: Schematic Operating Characteristic Curve

sampling plan (in the terminology of [41]). For mycotoxin testing applications, we can assume that the size of the aggregated sample increments is very small relative to the bulk material in the foodstock lot, so that the removal of the sampled material has no impact on the size and composition of the remaining bulk material. Further, we initially assume that the analytical measurements are separately performed for each individual sample increment, since the standard acceptance-sampling theory primarily deals with this case (this assumption is clearly not appropriate for mycotoxin testing, where sample compositing is ubiquitous and essential).

Acceptance sampling plans fall into two broad categories [41], viz., attributes sampling (AS) which are typically used when only qualitative measurement outcomes are available (e.g., whether or not the measurement exceeds a reference value), and variables sampling (VS) which are typically used for quantitative measurement outcomes (e.g., the actual magnitudes of the measurements). The AS and VS plans are both relevant for mycotoxin testing, however, when the pre-requisites and assumptions are valid, then the VS plans make relatively more efficient use of the measurement data; therefore, since this is invariably the case, only the VS plans are considered in further detail here.

The basic theory for the VS plans as described below, requires that the individual sample-increment measurements Y_i are independent random variables that follow a normal distribution $Y_i \sim \mathcal{N}(\mu, \sigma^2)$. This assumption may appear stringent, but in many cases, a preliminary transformation of the measurement data is often sufficient to induce normality. For instance, the Box-Cox

family of transformations defined by

$$Y_i^{(\xi)} = \begin{cases} (Y_i^{\xi} - 1)/\xi & \xi \neq 0, \\ \log Y_i & \xi = 0, \end{cases} \tag{1}$$

which includes the log and power-law transformations as special cases, are widely used for this purpose. Many statistical packages provide routines for obtaining the maximum-likelihood estimates of the Box-Cox parameter ξ in (1), and for evaluating the normality of the resulting transformed variable (e.g., by examining Quantile-Quantile plots of the transformed variable on a normal probability graph).

For the VS plan, the measurements Y_i for some n independent sample increments are taken, and the sample increment mean and variance are obtained as $\bar{Y} = (1/n)\sum_{i=1}^{n} Y_i$ and $S^2 = \sum_{i=1}^{n}(Y_i - \bar{Y}^2)/(n-1)$ respectively. The lot is considered to be of acceptable quality if $\bar{Y} + kS < U$ with respect to the specified threshold value U, for some value k. We only consider the case where U is a specified upper threshold value, since this is the relevant case for mycotoxin testing. Suitable values for (n,k), which define the VS plan, can be obtained by specifying the desired performance criteria for the corresponding OC curve. Let $\theta = P(Y_i > U)$, where $0 < \theta < 1$ is the exceedance probability with respect to the threshold value U (note that since $Y_i \sim \mathcal{N}(\mu, \sigma^2)$, we have $2\theta = \mathrm{erfc}((U-\mu)/\sqrt{2}\sigma)$ in terms of the complementary error function). The lot acceptance probability based on the sample increment measurements Y_i is then given by $P(\text{accept}) \equiv P(\bar{Y} + kS < U)$, as shown schematically in Figure 3, with reference to the distributions of Y_i and $\bar{Y}$ respectively.

From sampling theory for the normal distribution [41], it is well known that the random variable $Z = \sqrt{n}(\bar{Y} - \mu)/\sigma$ has a standard normal distribution $Z \sim \mathcal{N}(0,1)$, and the random variable $V = (n-1)S^2/\sigma^2$ has a Chi-squared distribution with $n-1$ degrees of freedom $V \sim \chi^2_{n-1}$. Furthermore, Z and V are independent random variables, for any real value δ, the ratio $t = (Z+\delta)/\sqrt{V/(n-1)}$ is a random variable with a non-central t-distribution $t \sim T(n-1,\delta)$ with $n-1$ degrees of freedom and non-centrality parameter δ, whose corresponding cumulative distribution function is given by $F_T(x; n-1, \delta)$. Letting $z_{1-\theta} = \sqrt{2}\mathrm{erf}^{-1}(1-2\theta)$ denote the quantile function for the standard normal distribution expressed in terms of the inverse error function, we specifically consider the case $\delta = -\sqrt{n}z_{1-\theta}$ for which $t \sim T(n-1, -\sqrt{n}z_{1-\theta})$.

The acceptance probability for the OC curve is then given by $P(\bar{Y} + kS < U) = P(t < -k\sqrt{n}) \equiv F_T(-k\sqrt{n}; n-1, -\sqrt{n}z_{1-\theta})$. The appropriate values for (n,k) for a VS plan consistent with the producers's risk (α, θ_A), and consumer's risk (β, θ_B) respectively, can then be obtained by noting that the AQL θ_A is the exceedance probability for which the probability that the lot should be accepted is greater than $1-\alpha$, and the RQL $\theta_B > \theta_A$ is the exceedance probability for which the probability that the lot should be accepted is less than β. Thus, given specific values for $(\alpha, \theta_A, \beta, \theta_R)$, suitable

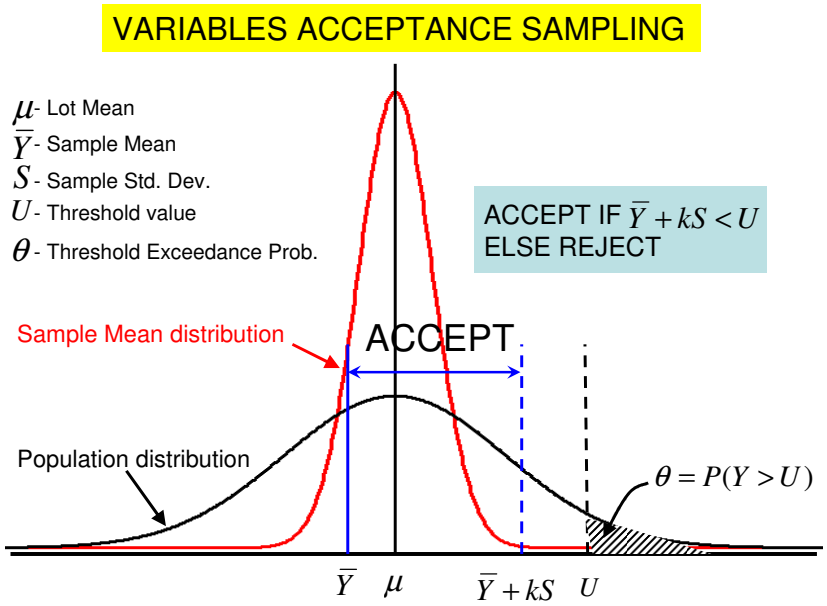

Figure 3: Schematic for Variables Sampling

values for n and k for the appropriate VS plan can be evaluated from the two conditions,

$$F_T(-k\sqrt{n}; n-1, -\sqrt{n}z_{1-\theta_R}) < \beta, \quad (2)$$

$$F_T(-k\sqrt{n}; n-1, -\sqrt{n}z_{1-\theta_A}) > 1-\alpha. \quad (3)$$

The required quantile functions for the non-central t-distribution are available in several statistical software packages (some packages only consider the case $\delta \geq 0$, in which case (2)–(3) can be rewritten using the identity $F_T(-t; n-1, -\delta) = 1 - F_T(t; n-1, \delta)$, which follows from the symmetry properties of the non-central t-distribution).

The theory described here for the OC curve of a VS sampling plan has other useful applications to mycotoxin testing as well. For example, it is possible to specify tolerance bounds for the quantiles of the normal distribution $\mathcal{N}(\mu, \sigma)$ from which the n samples Y_i are taken. Note that if μ and σ are known, then the $1-\theta$ quantile is given by $U = \mu + z_{1-\theta}\sigma$, where $z_{1-\theta}$ is the quantile function for the unit normal distribution. In practice, μ and σ are unknown, but tolerance bounds for U can be obtained in terms of the sample mean $\bar{Y}$ and standard deviation S in the form $\bar{Y} + k_{1-\theta}S$. For instance, the quantity $\bar{Y} + \underline{k}_{1-\theta}S$ is a lower bound to the quantile U with confidence level $1-\alpha$, where $\underline{k}_{1-\theta}$ is obtained from $F_T(-\underline{k}_{1-\theta}\sqrt{n}; n-1, -\sqrt{n}z_{1-\theta}) = 1-\alpha$. Similarly, from the symmetry of the normal

distribution, $\bar{Y} + \overline{k}_{1-\theta} S$ is a upper bound to the quantile U with confidence level $1-\alpha$, where $\overline{k}_{1-\theta}$ is obtained from $F_T(\overline{k}_{1-\theta}\sqrt{n}; n-1, -\sqrt{n}z_\theta) = 1-\alpha$. For mycotoxin testing, it is possible to specify suitable values for α and θ, and based on the sample measurements Y_i, obtain these tolerance bounds for U, which may be of considerable interest for evaluating the food safety risks.

3 Topics in Mycotoxin Testing

In this section, we consider four topics that are of considerable relevance to the further development of the theory and practice of mycotoxin testing in food-stock lots. These topics are briefly introduced here, and subsequently discussed in greater detail in the individual subsections.

First, the basic acceptance-sampling theory for the VS plans assumes that the individual sample-increment measurements are independent and identically distributed. However, in practice, there may be biases and correlations between the sample increments due to spatial, temporal and batch effects. Furthermore, it is often necessary to distinguish the contribution of each individual step to the overall variability of the sampling and measurement protocols. A given acceptance-sampling plan may no longer be appropriate when different analytical measurement techniques are used, or when there are variations in the equipment, processes and human expertise across testing sites. For example, an increase in the measurement uncertainty will shift the OC curve for a given acceptance-sampling plan so as to increase the producer's risk at the AQL [31, 32]. The correlation structure of the sample measurements, and the absolute and relative magnitudes of the sampling and measurement errors are therefore important considerations for the specification of the acceptance sampling plan.

Second, as noted in Section 2.1, the sample preparation and analytical measurement steps in mycotoxin testing are invariably more expensive and time-consuming than the sample-increment collection step, and the individual measurement of each sample increment is quite impractical. Therefore, a composite sample is prepared by aggregating the individual sample increments and homogenizing the resulting aggregate, and then one or more sub-samples are taken from this aggregate for the analytical measurement. This compositing process is equivalent to a "physical averaging" of the individual sample increments, and provides the same information on the sample mean as would have been obtained by averaging the individual analytical measurements on the distinct sample increments. However, after compositing it is no longer possible to estimate the sample variance required for the VS plan in Section 2.2, and furthermore, this physical averaging process also introduces an additional source of uncertainty in the overall measurement process which must be quantified.

Third, the assumption that the sample-increment measurements are normally distributed is often not valid, and although the normality-inducing tranformations described in Section 2.2 can be used to develop the appropriate acceptance sampling plans, there are certain situations where it may be

necessary to directly account for the non-normality of the individual sample increments. As a case in point, in the presence of clumping and stratification effects for the mycotoxin contamination in the food-stock lot, many of the individual sample increments will have no contamination at all, while a few remaining sample increments will have a positive-valued distribution of the contamination. This bi-modal, non-negative and non-normal distribution for the sample-increment measurements impacts the acceptance-sampling plan in a fundamental way, since the theory for obtaining the OC curves for the VS plans in Section 2.2 is then no longer directly applicable.

Fourth, in mycotoxin testing applications it is often difficult to sharply demarcate the acceptable and unacceptable values for the mycotoxin levels. Rather, there is a range of values over which the mycotoxin levels become progressively unacceptable, and in such cases, it is useful to define an intermediate "marginally acceptable" category. For example, in the VS plan, it may be desirable to introduce two threshold values – an upper value U_2 above which any measurements Y_i are considered unacceptable, and a lower value $U_1 < U_2$ with any measurements Y_i between U_1 and U_2 considered marginally acceptable. The introduction of a marginally-acceptable range of values for Y_i leads to acceptance sampling plans that are also suitable for other food-safety applications [24] besides mycotoxin testing, where an additional level of control is often necessary to ensure the usual consumer's risk, while at the same time, precluding even the smallest possibility of accepting a food-stock lot with the quality variable significantly exceeding the RQL.

The four topics outlined above are discussed in much greater detail below.

3.1 Uncertainty Quantification

In general, the mycotoxin level measurements obtained using the protocols in Section 2.1 will differ from the actual mean level in the food-stock lot, due to the presence of systematic biases and random errors in the various steps of the measurement process. An estimate of the various biases and the overall measurement uncertainty is therefore desirable in order to ensure that the measurements are sufficiently reliable and accurate, and in order to evaluate and manage the acceptance-sampling risk.

More specifically, the quantification of measurement uncertainty has a two-fold objective: first, to identify the relative contributions of the different sources of variability to the overall measurement uncertainty, and second, to evaluate the impact of this uncertainty on the acceptance-sampling plans, given the specification of the desired producer's risk and consumer's risk.

In particular, the relative magnitudes of the different sources of uncertainty in the measurement process can be used to identify the specific process steps that have the greatest potential for reducing the overall measurement uncertainty. It should be kept in mind that the cost and time for any additional testing to reduce the measurement uncertainty, may be quite different in the sampling, sample preparation and analytical measurement stages of the measurement process.

Figure 4 provides a schematic of the mycotoxin measurement process, and lists some of the fixed and random effects covariates that can be used to model the mycotoxin measurement data.

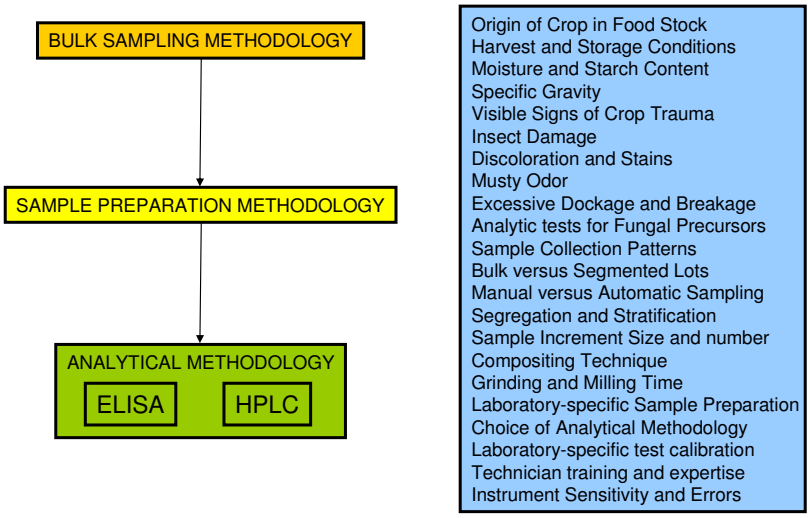

Figure 4: Fixed and Random Effects in Mycotoxin Measurement

The consideration of the fixed effects is taken up in greater detail in Section 3.2. From the acceptance sampling perspective, the primary interest is in the random effects associated with the bulk and composite sampling process, the sub-sampling and sample preparation, and the analytical measurement steps, which are responsible for the *intra-lot* variability, and thereby, directly impact the performance and risk of the acceptance sampling plan for individual food-stock lots.

The quantification of measurement uncertainty for an abstract measurement system is described in [17], and a discussion that is specialized to applications involving microbiological methods is described in [4]. These references are concerned with two basic approaches, as described below.

The first approach, termed the "bottom-up" or "modeling" approach, uses a model-based specification to characterize the uncertainty at each step of the measurement protocols (this process can be formalized using methods such as Ishikawa diagrams, e.g., [41]). Subsequently, the individual uncertainties can be combined across these steps, for example, using independence assumptions, to obtain the overall measurement uncertainty. Some of these individual steps, particularly the laboratory-based analytical methodologies have been extensively studied, and therefore their uncertainty characterization is well known, while other steps, such as the sampling step or the compositing step, may be much less studied and less well characterized. A framework for the identification and modeling of the various steps in the measurement protocols can be carried out using the "error mechanism" classification described in the theory of bulk sampling (e.g., [38, 44]).

The second approach, termed the "top-down" or "empirical" approach, is based on analyzing the measurements obtained with a suitable experimental design that is chosen to identify the sys-

tematic biases and random variations at the various stages in the measurement process (this design could possibly also include measurements obtained at different testing sites and laboratories, or using multiple, different protocols and techniques). A careful experimental design, which may include the use of replication in various steps of the measurement process, will make it possible to obtain measurement data sets that can be analysed by analysis-of-variance (ANOVA) techniques, in order to estimate the relevant contributions of the individual variance components to the overall measurement uncertainty.

Our focus is primarily on this latter "top-down" approach for the characterization of sampling and measurement uncertainty. The analysis that is required in this case can be simply illustrated for the case of a single food-stock lot with a mean mycotoxin level μ, with the following model that is based on a nested experimental design in which Y_{ijk} denotes the measured value from the k'th analytical measurement on the the j'th sub-sample from the i'th sample increment. Then $\mu = E(Y_{ijk})$, where

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}, \tag{4}$$

with $i = 1,\ldots,I,\ j = 1,\ldots,J_i$ and $k = 1,\ldots K_{ij}$. Here $N = \sum_i^I N_i, \quad N_i = \sum_j^{J_i} \sum_k^{K_{ij}} N_{jk}$, where N is the total number of measurements. Similarly, α_i denotes the sampling error in the i'th sample, β_{ij} denotes the sub-sampling error in the j'th nested subsample, and ε_{ijk} denotes the measurement error in the k'th nested analytical measurement. The individual error terms are assumed to be independent random variables with zero mean and constant variance across the levels, i.e., $\mathrm{Var}(\alpha_i) = \sigma_\alpha^2$, $\mathrm{Var}(\beta_{ij}) = \mathrm{Var}(\beta_{ij}) = \sigma_\beta^2$ and $\mathrm{Var}(\varepsilon_{ijk}) = \mathrm{Var}(\varepsilon_{ijk}) = \sigma_\varepsilon^2$ respectively. Then $Y_{ijk} \sim \mathcal{N}(\mu, \sigma_Y)$, and from the independence assumptions in (4), it follows that the overall measurement variance $\sigma_Y^2 = E(Y_{ijk} - \mu)^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2$. This model may also be used with composite samples instead of sample increments, although the compositing procedure should be such that an identical number of sample-increments is used for all the composite samples under consideration. However, this model in (4) is further generalized in Section 3.2 to also include non-uniformity in the compositing step.

The respective variance components in (4) can be estimated by a nested ANOVA procedure using measurement data obtained from a suitable experimental design. A balanced design would require $J_i = J, K_{ij} = K$, but it may be impractical to obtain the data for such a design, particularly given the costs of obtaining a large number of measurements, and given the vagaries of the data collection process which may lead to missing measurements in the original balanced design. Therefore, an unbalanced design is invariably used in the data analysis step, with sufficient measurement coverage at each level of the nested hierarchy in order to perform the variance components estimation while keeping the data collection costs manageable. However, the balanced design is advantageous, since with the additonal assumption that random effects are all normally distributed, i.e., $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$, $\beta_{ij} \sim \mathcal{N}(0, \sigma_\beta^2)$ and $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, the significance tests for the variance component estimates can be carried out using the F-test (the details of the estimation procedure, and

the differences in the ANOVA procedures for balanced versus unbalanced designs are discussed for example in [39]).

A special case of the model in (4) can be used to evaluate the batch effects in the sample increments for VS plans. These batch effects may arise because the food-stock lot itself may be provided in a segmented form (e.g., bins, bags and truck-loads), with only a small number of batches being selected for further sampling. In this case, the samples may exhibit significant within-batch correlation, which should be taken into account in the application of the VS plans in Section 2.2. This situation can be modeled in (4) by omitting the term involving ε_{ijk}, e.g., so that $Y_{ij} = \mu + \alpha_i + \beta_{ij}$, with the within-class correlation denoted by $\rho = \sigma_\beta^2/(\sigma_\alpha^2 + \sigma_\beta^2)$. It can be seen that when $\rho = 1$, so that the between-batch variance is negligible and the individual measurements are essentially independent with $Y_{ij} \sim \mathcal{N}(\mu, \sigma_\beta)$, the sample size is effectively N. On the other hand, when $\rho = 0$, the individual measurements within each batch are essentially redundant so that $Y_{ij} \sim \mathcal{N}(\mu, \sigma_\alpha)$ and the effective sample size is I. The impact of the within-class correlation ρ on the acceptance-sampling criterion in the VS plans, and in particular on the one-sided tolerance limits in Section 2.2, has been analysed using the notion of an "effective sample size" for each ρ, whose values interpolate between the two limits I for $\rho = 0$, and N for $\rho = 1$ [43].

An example of an unbalance nested design is shown schematically in Figure 5. This design can be used to estimate the required variance components, and more accurate estimates may be obtained by increasing the replication of measurements at each level in this design. In addition, any bias due to calibration differences in the analytical measurement techniques (e.g., ELISA and HPLC) can also be estimated from this design.

One approach for quantifying the measurement uncertainty is through spiking experiments, such as the one described in [2], where a fungal outbreak of *penicillium verrucosum* was incubated in a barley consignment for a period of several weeks, leading to levels of Ochratoxin A in the lot that exceeded 20 ppb. The experimental design for the measurement analysis was primarily intended in this case to examine the effect of manual versus automatic sampling methods, and independent sample increments were obtained using both techniques. In either case, the sample increments were combined into 8 separate composite aggregates, and for each aggregate the subsequent sample preparation and the analytic measurement steps were replicated. The results were analysed to estimate the variance components using a robust ANOVA procedure [17], with the overall uncertainty range estimated to be 0-42 ppb for manual sampling, and 12-28 for automatic sampling. This study showed that automatic-sampling techniques reduced the measurement errors attributable to the sampling, and unexpectedly, as a consequence the overall uncertainty was then dominated by the analytical measurement step. Although spiking experiments are useful for providing estimates for the relevant variance components in the mycotoxin measurement process, the experiments are difficult to carry out, and the mycotoxin contamination process under these controlled conditions may not be directly comparable to that encountered in the field.

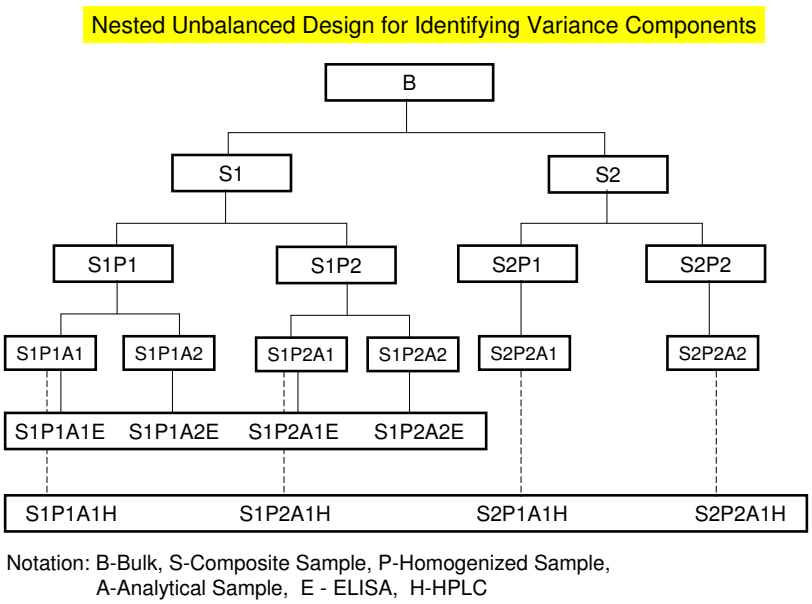

Figure 5: Schematic for Unbalanced Nested Design for Variance Components Estimation

An experimental design that can be used to obtain the variance components for measurements on the same lot across different sites is described in [35], and is based on the fact that the mycotoxin testing is also sometimes performed when a food-stock lot leaves a delivery facility, in addition to the usual case when the lot arrives at a receiving facility. Therefore, it often makes sense for these two facilities to exchange their sub-sampled fractions from the same lot, so that their respective measurements can be mutually calibrated and verified. This design is illustrated schematically in Figure 6, and the resulting analysis can be used to determine any location bias, and to determine the relevant variance components in the measurement process.

3.2 Composite Sampling

In mycotoxin testing, as in many other analytical testing applications, the cost for acquiring an individual sample increment is small, while the cost of an individual analytical measurement is high. Therefore, mycotoxin testing protocols often incorporate composite sampling, in which the individual sample-increments obtained from the food-stock lots are aggregated, homogenized and sub-sampled for the analytical measurement. The compositing process may be regarded as equivalent to physical averaging of the individual sample increments. The resulting composite sample estimate for the mean mycotoxin level is equivalent to separately averaging the measurements from the individual sample increments. Although this composite sample estimate is obtained at a far reduced cost, there are several trade-offs to be considered. First, there is the loss in precision whenever the measurement error is of the same order as the sampling error, so that using the composite

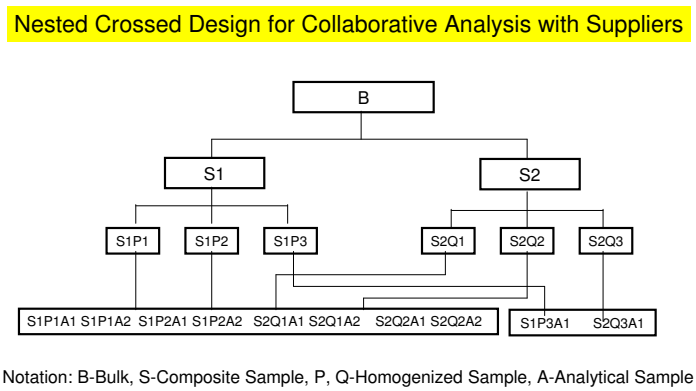

Figure 6: Schematic for Crossed/Nested Design for Variance Components Estimation

sample measurement can lead to increased risks in the acceptance-sampling protocols. Second, the distribution of the individual sample-increment measurements is no longer available to estimate the sample variance, which is required for implementing a VS plan for acceptance sampling. Third, it is no longer possible to include the relationship between the individual sample-increment measurements and other auxiliary variables, such the spatial locations from which the respective sample increments were extracted, making it difficult to identify the spatial correlation structures and "hot spots" in the food-stock lot.

However, with some modifications, many of these limitations in the basic composite sampling procedure can be removed, while retaining to a large extent the benefit of the reduction in measurement costs.

For example, even though the use of the VS plan for acceptance sampling is no longer straightforward when compositing is used, in practice, the required variance estimates can be recovered from multiple, independent composite samples [8, 12], in order to implement a suitable VS plan. The relevant statistical theory for bulk sampling using composite samples is covered from an acceptance sampling perspective in [5, 41], and the estimation of compositing and measurement errors using techniques based on mixed effects models is described in [11, 29, 15].

An additional source of variability in composite sampling arises from the fact that, despite best efforts, the size of each individual sample increment in the aggregated composite may not be identical, and furthermore, even when the sizes are identical, the composite sample may be imperfectly blended, so that any sub-sampled measurement aliquots will contain dissimilar proportions and contributions from the individual sample increments. In either case, the resulting sub-sample that is taken for measurement may be regarded as a weighted combination of the individual sample increments, with the weights also being regarded as random variables. The use of random weights is a novel aspect of the theory of composite sampling, and the relevant mathematical developments can be found, for example, in [16].

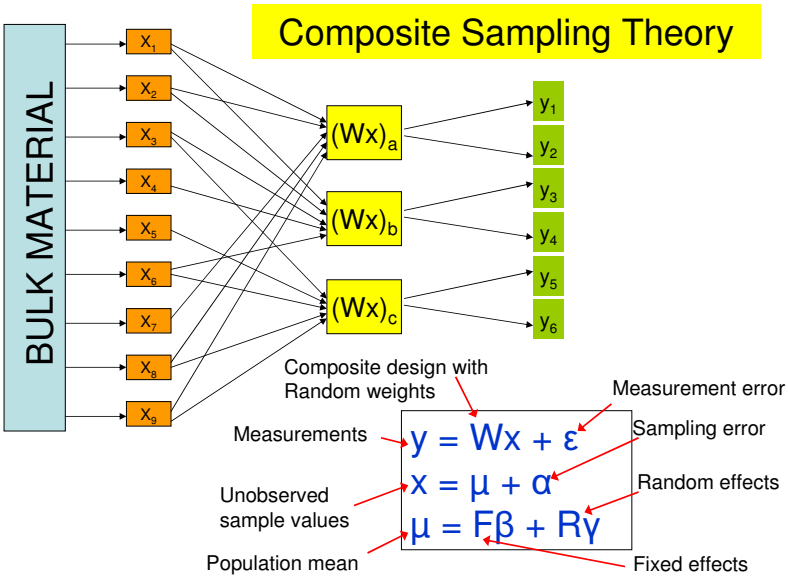

Figure 7: Schematic of Linear Mixed-Effects Model for Composite Sampling

Recent developments in composite sampling are described in the monograph [37], in which a principled approach based on linear mixed-effects models is described, which can incorporate all the sources of variability described in Figure 4. Using this approach, any fixed effects associated with the individual sample increments can be modeled and estimated simultaneously with the estimation of the variance components for the bulk sampling, compositing and sub-sampling, and the analytical measurement stages respectively.

This mixed-effects model can be written in the following hierarchical form

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}+\boldsymbol{\varepsilon}, \tag{5}$$

$$\boldsymbol{x} = \boldsymbol{\mu}+\boldsymbol{\alpha}, \tag{6}$$

$$\boldsymbol{\mu} = \boldsymbol{F}\boldsymbol{\beta}+\boldsymbol{R}\boldsymbol{\gamma}, \tag{7}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is the vector of composite sample measurements, $\boldsymbol{W} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{x} \in \mathbb{R}^m$ are the weight matrix for the composite sample and the vector of unknown sample-increment measurement values respectively, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the measurement error, and $\boldsymbol{\mu},\boldsymbol{\alpha} \in \mathbb{R}^m$ respectively denote the population mean and sampling error for the individual sample increments respectively. In turn, the population mean $\boldsymbol{\mu}$ depends on certain fixed and random effects, where $\boldsymbol{F} \in \mathbb{R}^{m\times p}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ denote the design matrix and parameters for the fixed effects, and $\boldsymbol{R} \in \mathbb{R}^{m\times q}$ and $\boldsymbol{\gamma} \in \mathbb{R}^q$ denote the design matrix and parameters for the random effects. Here, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are mutually-independent random

vectors with mean zero and variance $\boldsymbol{\Sigma_\alpha}, \boldsymbol{\Sigma_\gamma}, \boldsymbol{\Sigma_\varepsilon}$ respectively. The entries of $\boldsymbol{W}$ may be deterministic, or may be random when it is necessary to take into account the variability of the compositing process as described above. In either case, $\boldsymbol{W} \in \mathbb{R}^{n\times m}$ has unit row sums $\boldsymbol{W1} = \boldsymbol{1}$, and the row vectors of $\boldsymbol{W}$ are denoted by $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ respectively, with the j'th entry of $\boldsymbol{w}_i$ being zero if the corresponding individual sample increment is not a part of the i'th composite sample. In the random case, $E(\boldsymbol{w}_i) = \boldsymbol{\omega}_i$, $\mathrm{Var}(\boldsymbol{w}_i) = \boldsymbol{\Sigma_{\omega_i}}$ and $\mathrm{Cov}(\boldsymbol{w}_i, \boldsymbol{w}_j) = \boldsymbol{\Gamma_{w_i w_j}}$ for $i \neq j$, and we note that $\boldsymbol{\omega}_i^T \boldsymbol{1} = 1$ and that $(0, \boldsymbol{1})$ is an eigenvalue-eigenvector pair for $\boldsymbol{\Sigma_{\omega_i}}$ and $\boldsymbol{\Gamma_{w_i w_j}}$.

In [37] the motivation for the model in (5)–(7) is given as follows. The unobserved sample-increment measurements are given by the sampling distribution with mean $\boldsymbol{\mu}$ and sampling variance $\boldsymbol{\Sigma_\alpha}$ where the population mean $\boldsymbol{\mu}$ in turn depends on certain fixed effects (e.g., supplier identity, measured moisture content, or visible damage indicators) represented by $\boldsymbol{F}$ and certain random effects (e.g., food-stock lot identifier, originating supplier identifier) represented by $\boldsymbol{R}$; the relevant data for modeling these fixed and random effects is either part of the auxiliary data, or is easily obtained from simple measurements. Furthermore, the composite sample measurement $\boldsymbol{y}$ is assumed to be some weighted combination of the individual sample increments, with added measurement variance $\boldsymbol{\Sigma_\varepsilon}$.

By eliminating the unobserved quantities in (5)–(7), we obtain

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Wb} + \boldsymbol{\varepsilon}. \tag{8}$$

where $\boldsymbol{X} = \boldsymbol{WF}, \boldsymbol{Z} = \boldsymbol{W}, \boldsymbol{b} = \boldsymbol{R\gamma} + \boldsymbol{\alpha}$ (note that $E(\boldsymbol{b}) = 0, \mathrm{Var}(\boldsymbol{b}) = \boldsymbol{R}^T \boldsymbol{\Sigma_\gamma R} + \boldsymbol{\Sigma_\alpha}$), so that when $\boldsymbol{W}$ is a fixed matrix, this form is equivalent to the well-known Laird-Ware representation for mixed-effects models, and can be fitted using standard statistical procedures available in several software packages (including the `lmer` package in R [3]).

The formulation in (5)–(7) can be applied to fairly general experimental designs such as those shown in Figure 7. The simplest case is when the sample composite comprises of a single sample increment; in this case, the model is equivalent to (4). The relevant estimation methods and examples based on this methodology will be described in future work.

3.3 Non-Normal Distributions and Variables Sampling

In this section, we consider the impact of non-normality in the distribution of the sample increment measurements on the design of VS plans for acceptance sampling in mycotoxin testing. In particular, we consider the compound Poisson-Gamma distribution, which has previously been used in the literature to model the distribution of composite sample measurements for mycotoxin levels [23, 26, 46], but as shown below, this distribution can also be used to model the distribution of the individual sample-increment measurements.

More generally, we consider the random variable $S = Y_1 + Y_2 + \ldots + Y_N, N \geq 0$, where Y_i and N are independent random variables, and the individual Y_i are identical random variables which are

conditionally independent given N. The support of Y_i is the positive real axis, while the support of N is the set of non-negative integers. The random variable S can be used to model the mycotoxin level in a sample increment, since N can be regarded as the number of contaminated grain kernels in a sample increment, and Y_i can be regarded as the non-zero mycotoxin level in a contaminated grain kernel.

The explicit form of the density function for the compound Poisson-Gamma distribution can now be obtained by making specific distributional assumptions about the random variables N and Y_i. If $N \sim Po(\lambda)$ is a Poisson random variable with parameter λ, the cumulant function of S is then given by

$$K_S(t) = \log E(e^{tS}) = \lambda e^{K_{Y_i}(t)} - 1, \tag{9}$$

from which we obtain

$$E(S) = \lambda E(Y_i), \qquad \text{Var}(S) = \lambda[(E(Y_i))^2 + \text{Var}(Y_i)]. \tag{10}$$

Furthermore, if we assume that $Y_i \sim Ga(\alpha, \beta)$ is a Gamma random variate, then

$$E(S) = \lambda\alpha/\beta, \qquad \text{Var}(S) = \lambda\alpha(\alpha+1)/\beta^2. \tag{11}$$

From (11), it can be seen that the random variable $S \sim Tw_p(\mu, \phi)$ is a Tweedie distribution of degree p [28], with $\text{Var}(S) = \phi(E(S))^p$ where $1 < p < 2$, and in terms of the parameters of the underlying distributions, we have

$$p = \frac{\alpha+2}{\alpha+1}, \qquad \mu = \frac{\lambda\alpha}{\beta}, \qquad \phi = \frac{\lambda^{1-p}(\alpha/\beta)^{2-p}}{2-p}, \tag{12}$$

and conversely,

$$\alpha = \frac{2-p}{p-1}, \qquad \beta = \frac{1}{\phi(p-1)\mu^{p-1}}, \qquad \lambda = \frac{\mu^{2-p}}{\phi(2-p)}. \tag{13}$$

The explicit form of the density function corresponding to (9) can be obtained by noting that

$$p_S(y) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} p_{Y_i}^{(n)}(y), \tag{14}$$

where $p_{Y_i}^{(n)}(y)$ is the density function of the n-fold convolution of $p_{Y_i}(y)$, with $p_{Y_i}^{(0)}(y) = 1, p_{Y_i}^{(1)}(y) = p_{Y_i}(y)$, and

$$p_{Y_i}^{(k)}(y) = \int p_{Y_i}^{(k-1)}(y-x) p_{Y_i}(x) dx, \tag{15}$$

in which the integral is taken over the range of Y_i.

These convolutions can be explicitly evaluated in the case of the Poisson-Gamma distribution, where $p_{Y_i}(y)$ in (14) is the $Ga(\alpha, \beta)$ density function, by noting that

$$K_{Y_i}(t) = \log\left(\frac{\beta}{\beta - t}\right)^{\alpha}, \tag{16}$$

so that for $Y_i \sim Ga(\alpha, \beta)$, we have

$$K_{Y_1+\ldots+Y_n}(t) = \log\left(\frac{\beta}{\beta - t}\right)^{n\alpha}, \tag{17}$$

since $p_{Y_i}^{(n)}(y)$ is the $Ga(n\alpha, \beta)$ distribution. Therefore, noting that

$$\lambda\beta^{\alpha} = \frac{1}{\alpha}\left(\frac{1+\alpha}{\phi}\right)^{1+\alpha}, \tag{18}$$

we obtain

$$p_S(y) = \begin{cases} e^{-\lambda}, & \text{if } y = 0, \\ e^{-(\lambda+\beta y)} y^{-1} \sum_{n=1}^{\infty} W_n, & \text{if } y > 0, \end{cases} \tag{19}$$

where

$$W_n(y) = \left[\frac{(1+\alpha)^{1+\alpha} y^{\alpha}}{\alpha\phi^{1+\alpha}}\right]^n \frac{1}{n!\Gamma(n\alpha)}. \tag{20}$$

An important aspect of the use of the compound Poisson-Gamma distribution for the sample-increment measurements, is that this distribution is closed under convolution in an important special case, i.e., if $S_1 \sim Tw_p(\mu, \phi)$ and $S_2 \sim Tw_p(\mu, \phi)$, then the random variable $1/2(S_1 + S_2) \sim Tw_p(\mu, \phi/2)$. This property also applies to weighted sums, i.e., if $S_i \sim Tw_p(\mu, \phi/\omega_i)$, then setting $\omega = \sum_{i=1}^{n} \omega_i$, the random variable $S = (1/\omega)\sum_{i=1}^{n} \omega_i S_i \sim Tw_p(\mu, \phi/\omega)$. Therefore, if the individual sample increment measurements follow identical Tweedie distributions, then the composite sample obtained by combining these sample increments will also follow a Tweedie distribution [23, 26, 46].

The maximum-likelihood estimation requires the evaluation of the density function for Poisson-Gamma distribution, as well as the derivatives of this density function with respect to the parameters to be estimated. This evaluation can be carried out either by direct summation of the infinite series in the explicit form of the density function in (19)–(20), or by numerical evaluation of the Fourier inversion integral for the characteristic function in (17), which makes it possible to carry out the required maximum-likelihood computations using standard statistical packages [14]. This methodology provides several advantages over alternative methods such as the method of moments used in [23, 26] for fitting compound Poisson-Gamma distributions to mycotoxin measurements, since in particular it can be extended to incorporate covariate effects (e.g., using Generalized Linear Models [28]). The estimation of mixed-effects covariates, and the inclusion of composite-sampling variance components, has been well studied for the case of the normal distribution as described in Sections 3.1–3.2, and the extension of this theory to the compound Poisson-Gamma distribution would be of considerable relevance for mycotoxin testing applications.

The density function for the compound Poisson-Gamma distribution is shown fitted to measurement data obtained from an experiment in which 9 trucks from a given supplier were sampled and measured on the same day using composite samples with 10 sample increments. The resulting

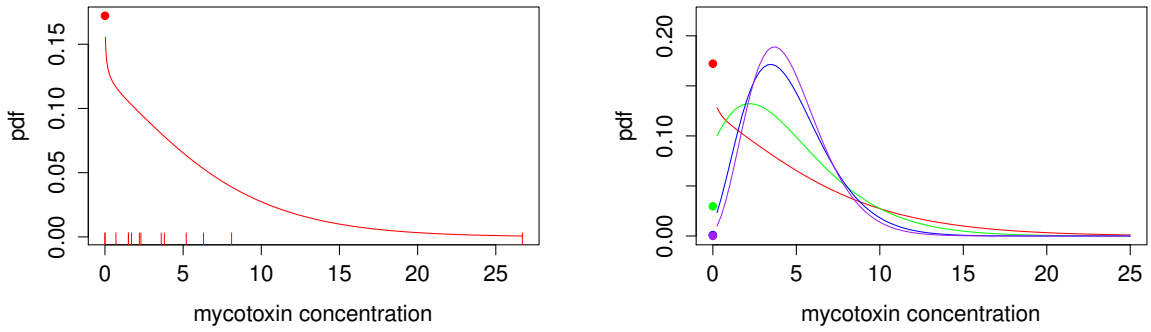

Figure 8: Compound Poisson-Gamma Distribution fitted to Mycotoxin Measurements from composite samples of 10 (red) sample increments (left), and corresponding densities that would be observed for composite samples with 20 (green), 40 (blue) and 50 (purple) sample increments (right).

model fitted using maximum-likelihood, had values $\mu = 4.542857, \phi = 2.454988, p = 1.524490$ for the Tweedie parameters. The resulting density function is shown in Figure 8, along with the corresponding density functions that would be obtained for composite samples with 20, 40 and 50 sample increments respectively. These distributions show significant non-normal behavior, and therefore, particularly for low mean concentrations and a small number of sample increments, the assumption of a normal distribution for the composite sample measurements does not appear to be justified.

The development of VS plans for acceptance sampling relies crucially on the assumption of a normal distribution for the sample-increment measurements, as described in Section 2.2. It is known that the use of the normality assumption, when it is not justified, leads to an under-estimation of the acceptance-sampling risk. The extension of the theory for VS plans to non-normal distributions has been widely studied. For example, the modifications required for taking into account the skewness and kurtosis in a non-normal sampling distribution is considered in [45]. An extension of VS plans to the case of the Inverse Gaussian distribution (which is a special case of a Tweedie distribution with $p = 3$) has been considered in [1, 9]. A similar extension to the specific case of the compound Poisson-Gamma distribution would be of considerable interest and relevance to the mycotoxin testing problem.

3.4 Marginal Quality in Variables Sampling

In mycotoxin testing, there is often no single value for the upper threshold U above which the sample-increment measurement Y_i would be considered unacceptable. Rather, moderate values of Y_i may raise quality concerns that might warrant further measurements, whereas much larger values

of Y_i may reflect serious and unacceptable food-safety concerns leading to immediate rejection of the food-stock lot. In practice, the upper threshold value U is typically taken from the regulatory or advisory guidelines, but practical concerns arise when the measured mycotoxin levels are close to this threshold value. These situations, which involve measurements that may be regarded as marginal or inconclusive in some sense, may lead to additional sampling and retesting of the lot which is invariably carried out in an *ad hoc* manner without a formal acceptance-sampling justification.

The standard 2-class or single-threshold VS plans described in Section 2.2 are quite inadequate for addressing the marginal quality issue, since these plans do not have a "soft" upper threshold for the sample-increment measurements. Indeed 3-class plans have been developed for this purpose, and are widely used in other food-safety testing applications [6, 33], but less frequently for mycotoxin testing.

These 3-class plans have many of the familiar requirements and properties of the equivalent 2-class plans, and for instance, the 3-class VS plan requires the measurements to follow a normal distribution (possibly after a preliminary normality-inducing transformation), as discussed in Section 2.2. However, the 3-class VS plans, by incorporating marginal quality measurements in the acceptance sampling plan, can provide greater flexibility for controlling the risk characteristics of the VS plan. To illustrate this, note that for a specified upper threshold U, there are an infinite number of distributions for measurements $Y_i \sim \mathcal{N}(\mu,\sigma)$ for different sets of parameters (μ,σ) that have the same exceedance probability $\theta = P(Y_i > U)$. The 2-class VS plans for all these distributions would be identical, but their risk profiles would be quite different, as illustrated by the following example given by [33], which describes a quality variable whose log-transformed measurements are normally distributed, with two possible parameter sets $(\mu_1 = 4, \sigma_1 = 1)$ and $(\mu_2 = 2, \sigma_2 = 3)$. The two distributions have the same exceedance probabilities with respect to the upper threshold value $U = 5$, which are denoted by $\theta_1 \approx 0.17$ and $\theta_2 \approx 0.17$ respectively. But changing the upper threshold value to $U = 7$ results in the exceedance probabilities $\theta_1 \approx 0.0013$ and $\theta_2 \approx 0.048$ respectively. In this situation, the 2-class plans with $U = 5$ may not provide risk protection for very large values of the quality variable exceeding $U = 7$, whereas the 3-class plans make it possible to control the risk at each of these threshold values, as described below.

Another situation where 3-class sampling plans are of considerable interest is when the incoming food-stock lots have been classified into groups, say by supplier type. If the mycotoxin level measurements for a particular group are consistently close to the threshold value, then additional testing may be warranted for this group, with careful attention to the risks involved. For instance, the food-safety risk from this supplier may be low, but there may be a long-term market and business downside from the poor retail quality of the accepted food-stock lots. It is difficult to address this marginal quality issue using a 2-class VS plan, even though *ad hoc* approaches such

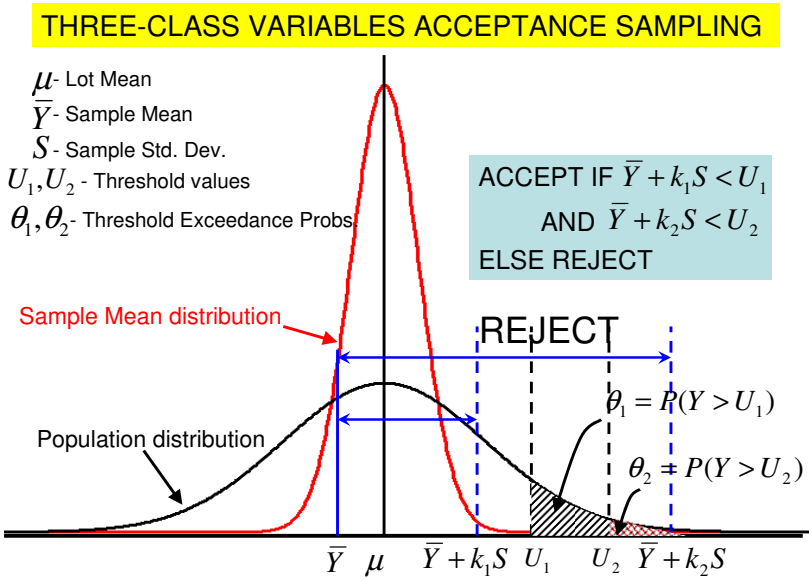

Figure 9: Schematic for 3-Class Variables Sampling plan

as retesting of marginal food-stock lots are often used in this situation without formal theoretical justification.

In addition to basing acceptance-sampling plans on the mycotoxin level measurements, a variety of other factors may also be incorporated into a composite score for the assessment of marginal quality. For example, it may often be useful to sample and test for the precursor toxicogenic fungi and fungal spores, since the the microbiological tests for the former can be carried out with high specificity and sensitivity [7] (although, however, the absence of a toxicogenic precursor fungus is no guarantee of low levels of mycotoxin contamination in a food-stock lot). The ability to include these other measurements as part of a marginal quality criterion is another advantage in the specification of the 3-class VS plan.

In order to obtain the OC curves for the 3-class VS plan, let U_1 and $U_2 > U_1$ denote the two upper threshold values corresponding to the marginal and unacceptable quality measurements respectively, and let by $\theta_1 = P(Y_i > U_1)$ and $\theta_2 = P(Y_i > U_1) < \theta_1$ denote the corresponding exceedance probabilities respectively (see Figure 9). The sample mean $\bar{Y}$ and sample standard deviation S are obtained as described in Section 2.2 for a sample of size n. The acceptance probability in this case is then given by

$$P(\bar{Y} - k_1 S < U_1 \text{ and } \bar{Y} - k_2 S < U_2 | \theta_1, \theta_2) = P(t_1 < -k_1\sqrt{n} \text{ and } t_2 < -k_2\sqrt{n} | \theta_1, \theta_2), \qquad (21)$$

where, for $i = 1, 2$, $t_i = (Z + \delta_i)/\sqrt{V/(n-1)} \sim T_{n-1,\delta_i}$ are random variables with a non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\delta_i = -\sqrt{n} z_{\theta_i}$, with z_{θ_i} being the quantile function for the standard normal distribution corresponding to the cumulative

probability $1-\theta_i$. The acceptance probability in (21) can be evaluated from the joint distribution of t_1 and t_2 which is a special case of a bivariate non-central t-distribution derived from a bivariate normal distribution with correlation 1 (see (22) below).

Suitable values for the three unknowns (n,k_1,k_2) in the 3-class VS plan can be obtained by specifying the acceptance probability for the OC curve at three points. One possibility is to choose two of these points to have the same producer's and consumer's risk as the equivalent 2-class plan, i.e., for $\theta_1=\theta_A,\theta_2\longrightarrow 0^+$, where θ_A is sufficiently small, this acceptance probability should be some value $1-\alpha$ for small α so that the lot is accepted with high probability; similarly, with $\theta_1=\theta_R,\theta_2\longrightarrow 0^+$ where $\theta_R>\theta_A$ is sufficiently large, the acceptance probability should be some small value β, so that the lot is rejected with high probability. Finally, if $\theta_1\longrightarrow\theta_C,\theta_2=\theta_C$ where θ_C is sufficiently large, the acceptance probability should be some small value γ; in general $\theta_C<\theta_B$, which implies that even a moderately high exceedance probability for unacceptable quality should immediately lead to the lot rejection irrespective of the magnitude of the exceedance probabilities for the marginal quality.

Then, in this case, the required conditions for the OC curve can be written in the form (i) $P(\text{accept}|\theta_1=\theta_A,\theta_2\longrightarrow 0^+)>1-\alpha$, (ii) $P(\text{accept}|\theta_1=\theta_R,\theta_2\longrightarrow 0^+)<\beta$, and (iii) $P(\text{accept}|\theta_1\longrightarrow\theta_C^+,\theta_2=\theta_C)<\gamma$. Therefore, given specific values for $\theta_A,\theta_R,\theta_C,\alpha,\beta,\gamma$, the quantities (n,k_1,k_2) can be determined from these three conditions. In fact, in this special case, the conditions (i) and (ii) are the same as that for the 2-class variables plan so that the relevant methods for that case can be used to obtain (n,k_1) independently from k_2; subsequently, the value for k_2 consistent with condition (iii) can be obtained in a separate step, again using methods similar to that for the 2-class variables plan for fixed n.

Aside from the special case described above, the general case for 3-class plans can be developed, which require the coupled evaluation of three specified conditions (similar to (i), (ii) and (iii) above) in order to determine suitable values for (n,k_1,k_2). This, in turn, requires the evaluation of the acceptance probability in (21), for which an explicit integral representation is given in [36, 33]; setting $f=n-1$, this representation takes the form

$$P(t_1<\delta_1 \text{ and } t_2<\delta_2|\theta_1,\theta_2)=\frac{2}{\Gamma(f/2)}\int_0^\infty \mathcal{F}(x,\delta_1,\delta_2,\theta_1,\theta_2)x^{f-1}e^{-x^2}dx, \tag{22}$$

where

$$\mathcal{F}(x,\delta_1,\delta_2,\theta_1,\theta_2)=\Phi(t_1x\sqrt{2/f}-\theta_1)I_{[x\le R]}(x)+\Phi(t_2x\sqrt{2/f}-\theta_2)I_{[x>R]}(x), \tag{23}$$

where Φ denotes the distribution function for the unit normal, I denotes the indicator function, and

$$R=\frac{(\theta_1-\theta_2)\sqrt{f/2}}{(\delta_1-\delta_2)}. \tag{24}$$

A series expansion of (22) in terms of elementary functions is given in [36, 33], but a straightforward numerical procedure using generalized Gauss-Hermite quadrature for (22) is also quite suitable evaluating the relevant acceptance probabilities (this procedure requires the integrand can be extended as an even function to negative values of x, and the range of integration in (22) to be changed to the entire real axis). This evaluation can then be used to obtain suitable (n,k_1,k_2) for 3-class VS sampling plans with specified performance criteria; as discussed above, these plans are of considerable interest in applications that incorporate marginal quality considerations.

4 Summary

We have introduced certain topics that are of considerable relevance for the further development of the theory and practice of mycotoxin testing in the food-processing industry. These topics include: the quantification of sampling and measurement uncertainty, the analysis of composite sampling protocols, the study of non-normal sampling distributions for the quality variable, and the incorporation of marginal quality criteria and risk measures in variables sampling plans.

The implementation and enhancement of mycotoxin testing protocols at food-processing facilities requires a strong cross-functional and cross-organization consensus between all stakeholders, including food-stock suppliers, process quality teams, product manufacturing teams and plant management. In addition, it requires good familiarity with the recommedations and advisories provided by external entities such as industry groups, and national and international regulatory bodies.

From the perspective of a food-processing manufacturer, some of the considerations that arise in implementing continuous improvement in the mycotoxin testing programs are as follows:

1. Mycotoxin testing is often carried out across multiple sites within an organization, including sites which may be located in different states, countries and even continents. The nature of the incoming food-stock lots, and the specific testing protocols and performance criteria at these sites, may vary considerably, and these aspects should be carefully considered in the evaluation and improvement of mycotoxin testing programs across the entire organization.

2. Periodic site audits of the mycotoxin testing programs are highly desirable, and the evaluation criteria should be customized to the specific site requirements (for example, at certain sites, the incoming food-stock lots may be segmented into bags and bushels, which would require alternative procedures for sampling and analysis that take into account the contribution of the inter-segment variations to the overall measurement uncertainty).

3. A partnership with the long-term suppliers including individual farmers, farmer co-operatives and grain storage and elevator companies, is highly desirable. This would enable the mycotoxin testing to be performed at an earlier stage in the supply chain, leading to more effective problem isolation and corrective action. Furthermore, such a partnership would promote col-

laborative practices such as the exchange of samples and sample measurements, which will contribute to the consistency, fairness and efficacy of the mycotoxin testing program.

4. A periodic review and evaluation of the analytical methodologies should be carried out to incorporate new techniques that may reduce the time and cost, and improve measurement accuracy. However, careful consideration should be given to the trade-offs that may be involved. For instance, certain omnibus measurement techiques may simplify the testing for multiple mycotoxins, but the specificity of the test results for the individual mycotoxins may be potentially reduced, leading to increased risks in the existing acceptance-sampling protocols.
5. A formal communication process should be implemented for the various stakeholders in the mycotoxin testing program; these stakeholders include, the management (to define the business goals of the mycotoxin testing process); the research team (to design the testing and sampling plans keeping in mind the business needs and program costs); the testing team (to execute the appropriate lot sampling and sample preparation procedures and onsite measurements); and, the laboratory team (to standardize and calibrate the onsite measurements, and to ensure timeliness in performing the confirmatory and gold-standard measurements).

5 Acknowledgements/Notes

We thank Robert C. Baker (MARS Inc.) and Randall M. Ford (MARS Petcare) for many helpful conversations related to this material.

Software for some the techniques described in this paper (e.g., normality-inducing transformations, two-class and three-class acceptance sampling plans, tweedie model fitting, etc.) is available from the first author.

References

[1] M. S. Aminzadeh, *Inverse Gaussian Acceptance Sampling Plans by Variables*, Communications in Statistics - Theory and Methods, Vol. 25(5), pp. 923-935 (1996).

[2] M. G. Andersson, E. V. Reiter, P. A. Lindqvist, E. Razzazi-Fazeli and P. Häggblom, *Comparison of Manual and Automatic Sampling for Monitoring Ochratoxin A. in Barley Grain*, Food Additives and Contaminants, pp. 1-10 (2011).

[3] D. Bates, *Linear Mixed Model Implementation in lme4*, `http://cran.r-project.org/web/packages/lme4/vignettes/Implementation.pdf` (2011).

[4] *Best Practices in Microbiological Methods*, Final Report and Executive Summaries from the Association of Official Analytical Chemists International Presidential Task Force on Best Practices in Microbiological methods, `http://www.fda.gov/Food/ScienceResearch/LaboratoryMethods/ucm124900.htm` (2006).

[5] C. A. Bicking, *The Sampling of Bulk Materials*, Materials Research and Standards, Vol. 7(3), pp. 95-116 (1967).

[6] D. F. Bray, D. A. Lyon and I. W. Burr, *Three-class attributes plans in acceptance sampling*, Technometrics, Vol. 85, pp. 575-585 (1973).

[7] L. B. Bullerman and H. Gourama, *Toxicogenic Fungi and Fungal Toxins*, Chap. 43, Compendium of methods for microbiological examination of foods, Fourth Edition (eds. F. P. Downes and K. Ito), American Public Health Association, pp. 439-446 (2001).

[8] J. M. Cameron, *The Use of Components of Variance in Preparing Schedules for Sampling of Baled Wool*, Biometrics, Vol. 7(1), pp. 83-96 (1951).

[9] T. K. Das, A. Gosavi and K. M. Kanchibhatta, *Optimal Design of Plans for Acceptance Sampling By Variables with Inverse Gaussian Distribution*, Communications in Statistics - Simulation and Computation, Vol. 31(3), pp. 463-488 (2002).

[10] M. E. Doyle, *Food Safety Risks of Mycotoxins*, Food Protection Trends, Vol. 33(3), pp. 167-171 (2013).

[11] A. J. Duncan, *Bulk Sampling: Problems and Lines of Attack*, Technometrics, Vol. 4, pp. 319-344 (1962).

[12] A. J. Duncan, *Some Measurement Error Considerations in Bulk Sampling with Special Reference to the Sampling of Fertilizer of Attack*, Symposium on Bulk Sampling, Sixty-First Annual Meeting Papers of the American Society for Testing Materials, pp. 3-11 (1959).

[13] A. J. Duncan, *Some Statistical Analysis with Respect to Compositing in the Sampling of Bulk Material*, Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research Development and Testing, pp. 389-407 (1969).

[14] P. K. Dunn and G. K. Smyth, *Evaluation of Tweedie exponential dispersion model densities by Fourier Inversion*, Statistics and Computing, Vol 18(1), pp. 73-86 (2007).

[15] D. A. Edelman, *Three-stage Nested Designs with Composited Samples*, Technometrics, Vol. 9(3), pp. 373-389 (1967).

[16] R. S. Elder, W. O. Thomson and R. H. Myers, *Properties of Composite Sampling Procedures*, Technometrics, Vo. 22(2), pp. 179-186 (1980).

[17] Eurachem/CITAC Guide, *Measurement Uncertainty Arising from Sampling. A Guide to Methods and Approaches*, First Edition, `http://www.eurachem.org` (2007).

[18] Food and Drug Administration, `http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/NaturalToxins/ucm120184.htm` (2010).

[19] Food And Drug Administration: Compliance Program Guidance Manual, Mycotoxins in Domestic and Imported Foods, http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/ChemicalContaminantsandPesticides/ucm109231.htm (2001)

[20] Food And Drug Administration: Compliance Program Guidance Manual, Mycotoxins in Domestic and Imported Foods, http://www.fda.gov/downloads/Food/GuidanceComplianceRegulatoryInformation/ComplianceEnforcement/ucm073294.pdf (2008).

[21] Grain Inspection, Packers and Stockyards Administration (GIPSA) Report, *Grain Fungal Diseases and Mycotoxin Reference*, http://archive.gipsa.usda.gov/pubs/mycobook.pdf

[22] International Organization for Standardization, *Acceptance Sampling Plans and Procedures for the Inspection of Bulk Materials* (2000).

[23] F. G. Giesbrecht and T. B. Whitaker, *Fitting Negative Binomial and Tweedie distributions to Aflatoxin in Peanut Lots (Method of Moments): Investigations of the Problems of Assessing Aflatoxin Levels in Peanuts*, Biometrics, Vol. 54, pp. 739-753 (1998).

[24] B. Jarvis, *Statistical Aspects of the Microbiological Examination of Foods*, Second Edition, Elsevier (2008).

[25] A. S. Johansson, T. B. Whitaker, F. G. Giesbrecht, W. M. Hagler Jr., J. H. Young and D. T. Bowman *Testing Shelled Corn for Aflatoxin. Part I: Estimation of Variance Components*, Journal of Association of Official Analytical Chemists International, Vol. 83(5) pp. 1264-1269 (2000).

[26] A. S. Johansson, T. B. Whitaker, F. G. Giesbrecht, W. M. Hagler Jr. and J. H. Young, *Testing Shelled Corn for Aflatoxin. Part II: Modeling the Observed Distribution of Aflatoxin Test Results*, Journal of Association of Official Analytical Chemists International, Vol. 83(5) pp. 1270-1278 (2000).

[27] A. S. Johansson, T. B. Whitaker, F. G. Giesbrecht, W. M. Hagler Jr. and J. H. Young, *Testing Shelled Corn for Aflatoxin. Part III: Evaluating the Performance of Aflatoxin Sampling Plans*, Journal of Association of Official Analytical Chemists International, Vol. 83(5) pp. 1279-1284 (2000).

[28] B. Jorgensen, *The theory of exponential dispersion models*, Monographs on Statistics and Applied Probability, Vol. 76, Chapman and Hall, London (1997).

[29] K. Kussmaul and R. L. Anderson, *Estimation of Variance Components in Two-stage Nested Designs with Composite Samples*, Technometrics, Vol. 9, pp. 373-389 (1967).

[30] G. J. Lieberman and G. J. Resnikoff, *Sampling Plans for Inspection by Variables*, Journal of the American Statistical Association, Vol. 50, pp. 457-516 (1955).

[31] H. Melgaard, *Sampling in Pharmaceutical and Chemical Industries*, in Encyclopedia of Statistics in Quality and Reliability (2008).

[32] H. Melgaard and P. Thyregod, *Acceptance Sampling by Variables under Measurement Uncertainty* in Frontiers in Statistical Quality Control 6, (eds.) H-J. Lenz and P-T. Wilrich, Physica Verlag, Heidelberg, pp. 47-57 (2001).

[33] P. A. Newcombe and O. B. Allen, *A Three-Class Procedure for Acceptance Sampling by Variables*, Technometrics, Vol. 30(4), pp. 415-421, 1988.

[34] G. Ozay, F. Seyhan, A. Yilmaz, T. B. Whitaker, A. B. Slate and F. G. Giesbrecht, *Determining the Variability Associated with Testing Shelled Corn for Aflatoxin Using Different Analytical Procedures in Louisiana in 1998*, Journal of Association of Official Analytical Chemists International, Vol. 90 (4) pp. 1036-1041 (2007).

[35] D. B. Owen, *A Special Case of a Bivariate Non-Central t-Distribution*, Biometrika, Vol. 52(3/4), pp. 437-446, 1965.

[36] G. P. Patil, S. D. Gore and C. Taillie, Composite Sampling, Springer, New York (2011)

[37] F. F. Pitard, Pierre Gy's Sampling Theory and Sampling Practice, CRC Press, Boca Raton (1993).

[38] S. R. Searle, *Topics in Variance Component Estimation*, Biometrics, Vol. 27(1), pp. 1-76 (1971).

[39] A. Rahmani, S. Jinap and F. Soleimany, *Qualitative and Quantitative Analysis of Mycotoxins*, Comprehensive Reviews in Food Science and Food Safety, Vol. 8, pp. 202-250 (2009).

[40] E. G. Schilling and D. V. Neubauer, Acceptance Sampling in Quality Control, Second Edition, Chapmand and Hall/CRC, New York (2009).

[41] K. K. Sinha, Detoxification of mycotoxins and food safety, Marcel Dekker, Inc., New York (1998).

[42] F. W. Scholz and M. Vangel, *Tolerance Bounds and C_{pk} Confidence Bounds Under Batch Effects*, in Advances in Stochastic Models for Reliability, Quality and Safety (eds. W. Kahle, E. von Collani, J. Franz and U. Jensen), Birkauser, Boston (1998).

[43] P. L. Smith, A Primer for Sampling Solids, Liquids and Gases - Based on the Seven Sampling Errors of Pierre Gy, SIAM Publications, Philadelphia (2001).

[44] K. Takagi, *On Designing Unknown-Sigma Sampling Plans Based on a Wide Class of Non-Normal Distributions*, Technometrics, Vol. 14(3), pp. 669-678, 1972.

[45] T. B. Whitaker, A. B. Slate, M. Bruno Doko, B. M. Maestroni and A. Cannavan, Sampling Procedures to Detect Mycotoxins in Agricultural Commodities, Springer (2010).