Report to CRC for National Plant Biosecurity

CONFIDENTIAL

Conceptual Statistical Framework for Stored Grains Sampling

Project - CRC 30086

David Elmouttie[1,2] and Grant Hamilton[1,2]

[1] Discipline of Biogeosciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland, Australia, 4001.

[2] Cooperative Research Centre for National Plant Biosecurity, LPO Box 5012, Bruce, ACT 2617, Australia.

25th November 2010

Background

Insects have been significant pests in stored grain products since the first storages were developed thousands of years ago (Rees 2004). Storage structures provide the ideal environment for insect populations to flourish. This is primarily due to the makeup of storage structures, as they provide a high resource environment protected from external conditions (Toes *et al.* 2005, Nansen *et al.* 2008).

There are a number of orders of insects associated with stored grains, including beetles (Coleoptera), moths (Lepidoptera), psocids (Psocoptera), bugs (Hemiptera) and parasitic wasps (Hymenoptera) (Rees 2004, Hagstrum and Subramanyam 2006). Many of these species now display global distributions due to thousands of years of grain trading (Rees 2004, Hagstrum and Subramanyam 2006). Many species associated with stored grains have developed into significant pests, and if left uncontrolled can cause significant commodity losses, through direct consumption and spoilage (Subramanyam and Hagstrum 1996, Rees 2004, Hagstrum and Subramanyam 2006). The presence of insects within grain storages can also lead to a number of secondary losses, most importantly a reduction in trade due to their presence, and also increased risk of mould growth and a reduction in grain quality (Rees 2004, Jeffries 2000). Management of stored product insects is therefore critical within grain storages throughout the world (Subramanyam and Hagstrum 1996, Rees 2004, Hagstrum and Subramanyam 2006).

Since the first crude storages were developed a number of advances have been made to minimise the impact of insects on grain supplies (Hagstrum and Subramanyam 2006). Improvements to storage structures themselves have been crucial in minimising the extent and frequency of invasion and colonisation of insect pests into storages (Hagstrum and Subramanyam 2006). The development of chemical and non-chemical control methods to suppress and eradicate insect populations once established has also been an important development in modern grain storages (Hagstrum and Subramanyam 2006). These advances have lead to substantial reductions in pest populations within storages and a reduction in the damage they cause.

Stored product insects however, remain problematic within storages globally (Herron 1990, Collins *et al.* 2000, Stejskal *et al.* 2003, Hagstrum and Subramanyam 2006, Flinn *et al.* 2010).

Commodity losses in developed countries remain significant and have been estimated to be between 5-10% of total grain stored (Adam *et al.* 2006). Thus, at all stages of grain production and supply there is significant impetus to ensure that insects are not present within grain (Collins *et al.* 2000, Stejskal *et al.* 2003, Hagstrum and Subramanyam 2006, Flinn *et al.* 2010).

An integral component to the management of grain supplies is the early detection of stored product pests (Hagstrum *et al.* 1985, Nansen *et al.* 2008). This has lead to the development of a variety of sampling and monitoring strategies (Hagstrum *et al.* 1985, Jeffries 2000, Nansen *et al.* 2008). Sampling and monitoring strategies have been implemented throughout the grains supply chain in an attempt to ensure that insects are detected prior to infestations spreading throughout grain stores leading to significant losses (Hagstrum *et al.* 1985, Hagstrum and Subramanyam 2006, Nansen *et al.* 2008). Further, significant emphasis has been placed on establishing effective sampling techniques to detect insect pests prior to trading grain domestically and internationally (Wilkin 1991, Jeffries 2000).

The first sampling programmes developed for the grains industry were developed primarily to ensure grains being traded internationally met the requirements of importers, with regards to the presence of insect pests (Jeffries 2000). In Australia for example, the first official sampling programmes were initiated as part of the 1963 Export Grains Regulations, as a direct response to growing concern from international buyers with the levels of insect within Australian grain shipments (SCA Working Paper 1981, Love *et al.* 1983, Jeffries 2000, Pheloung and Macbeth 2000). Similarly, other large grain producing nations such as Canada and the USA have developed sampling protocols based on the need to ensure traded grains are free from insect pests (Jeffries 2000).

Early sampling programmes were developed in an *ad hoc* manner as a direct response to buyer concerns. Consequently, protocols were based on pragmatic constraints in the grain supply chain rather than being developed within a robust statistical framework (SCA Working Paper 1981, Jeffries 2000). This has lead to the development of sampling protocols that differ significantly between nations (Jeffries 2000). For example, Australia, the USA and Canada have sampling protocols that differ considerably (Jeffries 2000). Differences in sampling protocols include the quantity of grain sampled per tonne, the frequency of sampling events and the number of subsamples taken (Jeffries 2000). The lack of consistency between international sampling protocols relates to sampling programmes being developed independently and based on the mechanisms and storages used to transport and hold grain in each country (Jeffries 2000).

A number of countries have however investigated the efficacy of their sampling strategy at detecting insects at a given density (Jeffries 2000). Existing sampling protocols have generally been found to be adequate to achieve the desired level of detection and hence they have changed little (Hunter and Griffiths 1978, Wilkin 1991). The statistical models used to examine sampling protocols however, were typically derived from Binomial models and developed to describe sampling procedures in quality assurance and manufacturing, rather being developed specifically for biological systems (Stephens 2001). In such situations, aggregation and rare events are typically not a major concern, and hence the Binomial distribution can safely be used to describe the process. The variation in behaviour of a biological pest is rarely considered and numerous potentially erroneous assumptions are made, particularly relating to species distribution, which may lead to inaccurate results.

Utilising sampling programmes developed for alternative purposes within biological systems therefore poses significant risks. A critical concern is that a sampling plan developed for one system may not adequately represent the distribution of sampling units within the system within which it is to be used. In this circumstance, either type one errors (*e.g.* concluding that grain lot contains insects when it does not) or type two errors (*e.g.* concluding that a grain lot is free from insects when it is not) may occur. In the first case, unnecessary costs may be incurred through the application of treatments when none are needed. While as an immediate cost this may not be high, the advent of Phosphine resistance in many different pest species means that the real future cost (the loss of Phosphine as a treatment option for grains pests) is considerable. In the second case, grain can be lost through consumption and spoilage, but most importantly, consignments can be rejected at a considerable cost to the exporter and to the exporting country. Clearly the risks in adopting an inadequate sampling programme can be very large for individual farmers, bulk grain handlers and exporters.

Developing sampling programmes for biological systems is complex, as target species are often mobile, cryptic or exist within environments that are not easily sampled (Green and

Young 1993, Royle and Nichols 2003, Garrard *et al.* 2008, Elmouttie *et al.* 2010). Although grain storages appear to be homogeneous, significant variation exists throughout any storage which in turn influences the distribution of insects (Lippert and Hagstrum 1987, Cuperus *et al.* 1990, Flinn *et al.* 2010). For example, temperature and moisture profiles will vary depending on season, position within the grain stack, grain quality and climatic conditions at harvest (Rees 2004, Hagstrum and Subramanyam 2006). Hence, insects commonly display varied distributions within and between storages (Hagstrum *et al.* 1985, Lippert and Hagstrum 1987, Cuperus *et al.* 1990, Flinn *et al.* 2010). Further, insect distribution can also be influenced by the interspecific associations among stored grain beetles with the density and distribution of certain species being influenced by the presence of others (Nansen *et al.* 2009).

Hagstrum *et al.* (1985) developed a novel approach for sampling insects within grain storages. Unlike previous sampling protocols developed for grain based on a binomial distribution, the method was based on the Negative binomial probability function, which has seen widespread use in ecology to describe aggregated distributions. Thus rather than an implicit assumption of homogeneity, the use of a negative binomial accounted for insect clumping behaviour within the grain bulk. This significantly improved detection estimates. Although the negative binomial probability function can adequately describe aggregated distribution it does not adequately describe rare events, however (Green and Young 1993). In such instances the Poisson probability function provides a better approximation (Green and Young 1993, Stephens 2001).

Insect distributions within stored grains can vary significantly between high and low levels of aggregation and from high to low densities depending on environmental conditions (Rees 2004, Hagstrum and Subramanyam 2006). It is unlikely therefore to find a single, generic probability distribution which adequately represents the range of conditions which may be present of a grain storage and supply network. Sampling programmes must therefore be developed around a flexible framework that is able to encompass the innate variation that exists within the system. An alternative statistical approach to sampling as developed by Elmouttie *et al.* (2010) is presented below. Unlike previous statistical models developed for grain sampling the method presented is based on two distinct probability functions. The model considers that insects may;

a) be heterogeneously distributed throughout grain bulk and;
b) insect densities within grain bulk may be low.

Conceptual Statistical Model

The model presented by Elmouttie *et al.* (2010) is based on a two step process, first determining the probability of sampling the infested portion of a grain lot and second determining the probability of returning a positive sample when the sample is drawn from the infested portion of the lot. The implication of this is that not all samples from an infested section of the grain lot will be positive.

An essential assumption of the model is that a grain lot can be separated into two distinct non-contiguous components, an infested portion and a portion free of infestation. Furthermore, it is assumed that the contaminants are homogeneously distributed throughout the infested portion of the consignment, according to the Poisson distribution.

The model identifies the following variables

p = the proportion of the lot which is infested

λ = the rate of contamination per (kg) in the infested part of the lot

n = number of samples drawn from the lot

w = the weight of each sample from (*n*) in kilograms

Let *X* denote the number of samples (from *n*) that originate from the infested portion of the grain lot. Then the number of samples that originate from this infested portion of the lot can be calculated using the binomial distribution $X \sim B(n,p)$ from which follows:

$$P\ (X = x) = \binom{n}{x}\ p^x(1-p)^{n-x}$$

For each sample that comes from the infested part of the lot, the probability of detecting an insect depends on the rate of contamination (λ). Let *A* be the number of insects in the sample conditional on the sample having come from the contaminated part of the lot:

$$P(A = a|X = x) = \frac{e^{-xw\lambda}(xw\lambda)^a}{a!}$$

However, of key interest is the situation where no contamination is detected, that is when *a* = 0. In this situation we get,

$$P(A = 0|X = x) = e^{-xw\lambda}$$

Consequently, summing over all possible values for *X* results in the unconditional probability

$$P(A = 0) = \sum_{i=0}^{n} P(X = i)P(A = 0|X = i)$$

$$= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} e^{-iw\lambda}$$

$$= \sum_{i=0}^{n} \binom{n}{i} \left(pe^{-w\lambda}\right)^i (1-p)^{n-1}$$

$$= \left(1 - p + pe^{-w\lambda}\right)^n$$

The final step in the equation is derived from the Binomial theorem -

$$(a+b)^n = \sum_{i=0}^{n} a^i b^{n-i}$$

Therefore the probability of detection is given:

$$P(A > 0) = 1 - P(A = 0)$$

$$= 1 - \left(1 - p + pe^{-w\lambda}\right)^n \text{ (1)}$$

The model (equation 1) is dependent upon four distinct parameters which influence the probability of detection. To date these model parameters have been based on limited empirical data and simulated data. It is therefore important that robust parameter estimates are developed to maximise the effectiveness of the model.

When implementing a model for use over an industry that spans a large geographic area and utilises varied techniques for storage and treatment of product it is also imperative that the natural variation which exists within the system is determined. Therefore model validation and parameter estimation needs to occur throughout the system such that the model can be tailored for specific areas when required. The remainder of this review will outline the approach that will be implemented to:

a) Develop model parameter estimates
b) Validate model on large scale storages (bunkers)
c) Conduct broad scale model validation and parameter estimation across the Australian grain production area
d) Consider various case studies which outline the utility of the model in various scenarios

Developing Parameter Estimates

Methodology for developing single point parameter estimates

First we consider a method to develop parameter estimates at a single point in time. The proportion of a lot infested (p), and the rate of infestation (λ), will vary in relation to a number of factors, such as, geography, season, and grain temperature and moisture. Estimates for these parameters can be developed, using either prior information from literature and previous studies, expert opinion or from intensive field sampling. Although some estimates exist for these parameters in the literature (*e.g.* Hagstrum *et al*. 1985, Nansen *et al.* 2009) these estimates are derived from storages in the USA where conditions and storage structures differ significantly. Intensive field sampling therefore is the most robust method for informing parameter estimates that are representative of local conditions.

To develop initial parameter estimates sampling from a single storage type (bunkers) sampling from a single geographic location will be undertaken. A large number of sample points (*e.g.* 20 - 25) will be drawn from each bunker selected. To account for potential stratification of insects, three depth profiles or strata will be sampled (upper, middle and lower strata). In total 60-75 samples will be drawn from each bunker. This sampling intensity constitutes a sampling protocol far greater than any prescribed rate or sampling that has previously occurred, although it should be noted that this sampling intensity is purely for effective parameter estimation, rather than as a suggested ongoing pest sampling strategy. Each sample will be individually sieved using standardised sampling equipment. For each sample the number of each species (major grain pests) will be recorded. This sampling design allows for estimates of the proportion of the lot infested for each species to be determined by:

$$p = \frac{number\ of\ infested\ samples}{total\ number\ of\ samples}$$

Estimates for the average rate of infestation for each species can be determined by:

$$\lambda = \frac{total\ number\ of\ individuals}{total\ number\ of\ infested\ samples}$$

Model validation

Once robust parameter estimates have been developed it is important to validate the model in a 'real world' scenario. The model will be validated by populating equation 1 with the mean parameter estimates for p and λ for each species, and solving to determine the number of samples n at given sample weight w required for various detection probabilities (*ie.* 0.75, 0.9). A success would be given if an insect is detected in any given sample. Sampling would be then undertaken at predicted intensity at random points in bunkers that had similar characteristics such as the same geographic area, similar storage and treatment histories and similar periods of the year. This method was used by Elmouttie *et al.* (2010) to developed parameter estimates with which the model was subsequently validated.

Broad scale parameter estimation

At any given time, numerous microclimatic factors such as temperature, humidity and moisture will vary within grain storages, in turn influencing the distribution of insects. Therefore although point parameter estimates provide a basis for model validation, the usefulness of such an approach over a large production area and between seasons will depend upon how similar the conditions are to those were where the estimates was taken.

The Australian grain production region is very broad and this encompasses widely differing climatic conditions. For a statistical sampling model to be applicable over such a wide production region, it is necessary to estimate parameters over multiple geographic regions and multiple seasons to account for the extent of variation within the system. Although it is not feasible to account for all geographic and seasonal variation by sampling the entire storage network, distinct geographic areas within the Australian grain production and supply chain can be identified to be representative of the entire system:

- North Eastern (Emerald, Goondiwindi)
- Central Eastern (Narrabri, Dubbo, West Wyalong)
- South Eastern (Horsham, Marong)
- Sothern Central (Cummins, Jamestown)
- North Western (Geraldton, Morawa, Wongan Hills)
- South Western (Albany, Katanning)

These regions represent areas with distinct, climatic condition and production regimes. Sampling over multiple seasons within each region would provide estimates related to local conditions but also a sampling programme tailored for specific periods. Once estimates are developed for the regions identified above a sampling programme can be optimised to maximise detections whilst minimising sample effort.

Developing sampling programmes

Effective sampling programmes are based on robust statistical frameworks and should consider the true distribution, or at least a good estimate of the distribution, of what is being sampled. The model developed by Elmouttie *et al.* (2010) provides a statistical framework which better conforms to the known distribution of insects in grain bulk. Further, the methodology for developing single point and broad scale parameter estimates increases the applicability of the model, considering local conditions. However, models not only have to be based on a statistical robust framework and realistic data, but also consider the pragmatic constraints to sampling for end users.

The methods described above to provide parameter estimates will produce a substantial amount of data for each region. It is therefore important to synthesis the data into a framework which is easily understood and implemented. As outlined above, climate will have a significant effect on insect densities. Climatic conditions vary over a geographic distribution and betweens seasons. The variation in insect density due to climatic variation across season and geography can therefore be summarised as different ***risk periods***. For example, based on relevant past research, it may be found that a continued period of *x* days above temperature *y* will lead to higher insect densities. Sampling can therefore be based around a series of decision thresholds to define risk periods, based on known growth rates and local conditions.

To satisfy the practical and statistical component of a sampling programme sampling intensity will be constructed around a risk matrix. Using a simple matrix design, sample intensity could be predetermined, such that for any given time of the year, within any given region the optimal sampling intensity is known. For example, consider sampling intensity or effort to be related to three risk categories (*ie.* high, medium, low). Risk in this scenario would be based on parameter estimation unique to different geographic regions. In this way, for any given region during a particular season the optimal sampling intensity can be determined simply.

Case studies

Three case studies will be be developed during the project to illustrate the effectiveness and flexibility of different components of the sampling strategy within the Australian grains production industry.

Case Study 1 – Comparing Statistical Sampling Models

In this case study the statistical sampling model developed by Elmouttie *et al.* (2010) will be compared to alternative sampling models. Models will be compared using empirical data (already collected) for multiple insect species. The aim of this case study is to determine the limitations and benefits of the model proposed by Elmouttie *et al.* (2010), when compared to alternative models for various insect densities and distribution patterns.

Case study 2 – Developing a statistical framework that considers alternative risk factors

This case study will utilise the data generated in the broad scale parameter estimation component of the project. The aim of the case study is to develop mean parameter estimates for particular regions and then ground truth the model within each region, against model predictions.

Case Study 3 – Sampling for sale verses control

Currently grains sampling models developed for the Australian production region are based on a zero tolerance threshold. That is sampling is conducted to ensure no insects are present within grain bulk. This concept stems from the fact that most sampling strategies were developed primarily to determine insect densities prior to sale or at intake. Sampling for management however, does not necessarily need to confirm to such thresholds. In this case study the model developed by Elmouttie *et al.* (2010) will be extended to consider alternative thresholds of insect density. The case study will consider the consequence of not treating if insect are detected and when management strategies (*ie.* fumigation) should be initiated.

Conclusions

The conceptual outline presented above provides an alternative sampling model which better adheres to the known distribution of insect within stored grains. This statistical model, coupled with the parameter estimation techniques described, will provide a more accurate method for detecting insects in Australian grain storages. Further the broad scale model validation process and the development of a simple risk matrix will give end users a simple, user friendly method to determine sample intensity which is based on a robust statistical framework.

References

Adam, B.D. Phillip, T.W. and Flinn, P.W. (2006). The economics of IPM in stored grain: Why don't more grain handlers use IPM? Proceedings of the 9th International Working Conference on Stored Product Protection, Campinas, Brazil.

Collins, P.J. Nayak, M.K. and Kopittke, R. (2000). Residual efficacy of four organophosphates on concrete and galvanizes steel surfaces against three liposcelid psocid species (Psocoptera: Liposcelidae) infesting stored products. Journal of Economic Entomology. (3, 1357-1363.

Cuperus, G.W., Fargo, W.S., Flinn, P. W. and Hagstrum, D.W (1990). Variables affecting capture of stored-grain insects in probe traps. Journal of the Kansas Entomological Society. 63, 486-489

Elmouttie, D., Kierimeier, A. And Hamilton, G. (2010). Improving detection probabilities in stored grain. Pest Management Science. 66, 1280-1286

Flinn, P.W., Hagstrum, D.W., Reed, C. and Phillips, T.W. (2010). Insect population dynamics in commercial grain elevators. Journal of Stored Products Research. 46, 43-47.

Garrard, G.E., Bekessy, S.A., McCarthy, M.A. and Wintle, B.A. (2008). When have we looked hard enough? A novel method for setting minimum survey protocols for flora and fauna. Austral Ecology. 33, 986-998

Green, R. H. And Young, R.C. (1993). Sampling to detect rare species. Ecological Applications. 3, 351-356.

Herron, G.A. (1990). Resistance to grain protectants and phosphine in Coleopterous pests of grain stored on farms in New South Wales. Journal of The Australian Entomological Society. 29, 183-1989.

Hunter, A.J. and Griffiths, H.J. (1978). Bayesian approach to estimation of insect population size. Technometrics. 20, 231-234.

Jefferies, G.M. (2000). Review of grain sampling inspection methodology. Department of Agriculture, Fisheries and Forestry, Australia.

Hagstrum, D.W. Milliken, G,A. and Waddell, M.S. (1985). Insect distribution in bulk-stored wheat in relation to detection or estimation of abundance. Environmental Entomology. 14, 655-661.

Hagstrum, D. W. (1987). Seasonal variation of stored wheat environment and insect population. Environmental Entomology. 16, 77-83.

Hagstrum, D.W. and Subramanyam, B. (2006). Fundamentals in stored-product entomology. AACC International Press, St Paul, Minnesota.

Lippert, G.E. and Hagstrum, D.W. (1987). Detection or estimation of insect population in bulk-stored wheat with probe traps. Journal of Economic Entomology. 80, 601-604

Love, G., Twyford-Jones, P. and Woolcock, I. (1983). An economic evaluation of alternative grain insect control measures. Australian Government Publishing Service. Canberra, Australian Capital Territory.

Nansen, C., Meikle, W.G., Campbell, J., Phillips, T.W. and Subramanyan, B. (2008). A binomial and species-independent approach to trap capture analysis of flying insects. Journal of Economic Entomology. 101, 1719-1728.

Nansen, C., Flinn, P., Hagstrum, D., Toews, M.D., and Meikle, W.G. (2009). Interspecific associations among stored-grain beetles. Journal of Stored Product Research. 45, 254-260.

Pheloung, P. and Macbeth, F. (2000). Export inspection: adding value to Australia's grain. Australian Post Harvest Technical Conference. 2000.

Rees, D. (2004). Insects of stored products. CSIRO Publishing. Melbourne, Victoria.

Report of the SCA Working Party on the Infestation of grain. (1981).

Royal, J.A. and Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. Ecology. 84, 777-790.

Stejskal, S., Hubert, J., Kučerová, Z., Munzbergová, Z., Lukáš J. and Žd'árková E. (2003). The influence of the type of storage on pest infestation of stored grain in the Czech Republic. Plant Soil Environment. 49, 55-62.

Stephens, K.S. (2001). The handbook of applied acceptance sampling: plans, principles and procedures. American Society for Quality, Milwaukee, Wisconsin.

Toews, M.D., Phillips, T.W. and Payton, M.E. (2005). Estimating populations of grain beetles using probe traps in wheat filled concrete silos. Environmental Entomology. 34, 712-718

Wilkin, D.R. (1991). An assessment of methods of sampling bulk grain. HGCA Project Report No. 34. Home Grown Cereals Authority, London.