Geostatistical analysis of the spatial distribution of mycotoxin concentration in bulk cereals

Monica Rivas Casado, David J. Parsons, Richard Mark Weightman, Naresh Magan, Simona Origgi

Geostatistical analysis of the spatial distribution of mycotoxin concentration in bulk cereals

Journal:	*Food Additives and Contaminants*
Manuscript ID:	TFAC-2008-241.R1
Manuscript Type:	Original Research Paper
Date Submitted by the Author:	15-Dec-2008
Complete List of Authors:	rivas casado, monica; cranfield university, natural resources Parsons, David; Cranfield univeristy, natural resources Weightman, Richard; ADAS UK Ltd, Centre for Sustainable Crop Management Magan, Naresh; Cranfield University, Applied Mycology origgi, Simona; Food Standards Agency
Methods/Techniques:	Statistical analysis
Additives/Contaminants:	Mycotoxins
Food Types:	Cereals and grain
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Figure1_TFAC_2008_241_R.emf Figure2_TFAC_2008_241_R.wmf Figure3_TFAC_2008_241_R.emf Figure4_TFAC_2008_241_R.emf	

Geostatistical analysis of the spatial distribution of mycotoxin concentration in bulk cereals

M. Rivas Casado[a*], D.J. Parsons[a], R.M. Weightman[b], N. Magan[a], S. Origgi[c]

[a] Cranfield University, Cranfield, Bedford, MK43 0AL, UK

[b] ADAS UK Ltd., Centre for Sustainable Crop Management, Boxworth, Cambridge, CB23 4NN, UK

[c] Food Standards Agency, Aviation House, 125 Kingsway, London, WC2B 6NH, UK

[*]Corresponding author. Email: m.rivas-casado@cranfield.ac.uk

Abstract

Deoxynivalenol (DON) and ochratoxin A (OTA) in agricultural commodities present hazards to human and animal health. Bulk lots are routinely sampled for their presence, but it is widely acknowledged that designing sampling plans is particularly problematical because of the heterogeneous distribution of the mycotoxins. Previous studies have not explicitly looked at the interactions between the spatial distribution of the mycotoxin and the strategy used to take samples from bulk. Sampling plans are therefore designed on the assumption of random distributions. The objective of this study was to analyse the spatial distribution of DON and OTA in bulk commodities with geostatistics. This study was the first application of geostatistical analysis to data on mycotoxins contamination of bulk commodities. Data sets for DON and OTA in bulk storage were collected from the literature and personal communications, of which only one contained data suitable for geostatistical analysis. This data set represented a 26 t truck of wheat with total of 100 sampled points. The mean concentrations of DON and OTA were 1342 μg kg^{-1} and 0.59 μg kg^{-1}, respectively. The results showed that DON presented spatial structure whilst OTA was randomly distributed in space. This difference between DON and OTA probably reflected the fact that DON is produced in the field, whereas OTA is produced in storage. The presence of spatial structure for DON implies that sampling plans need to consider the location of sample points in addition to the number of points sampled in order to obtain reliable estimates of quantities such as the mean contamination.

Keywords:- Geostatistical analysis; deoxynivalenol; ochratoxin A; bulk; cereals.

Introduction

Designing sampling plans for mycotoxins is particularly problematic because of the heterogeneous distribution of these contaminants in bulk lots of different commodities (Stroka *et al*., 2004, Schatzki 1995a and 1995b and Jewers *et al*., 1988). For example, contamination within a lot may result from a very small proportion of grains or kernels containing high concentrations of the mycotoxin. Because of the heterogeneous nature of mycotoxins, the problems lead to uncertainty over both how many samples to take and how to decide on the sampling positions. Normal practice is to aggregate incremental samples taken from a bulk commodity and test a sub-sample to estimate the mean concentration of the mycotoxin. As a result, information on the spatial variability and distribution of the mycotoxin is lost.

Extensive work on mycotoxin sampling has been carried out by the group of Whitaker (Whitaker and Wiser, 1969; Whitaker *et al*., 1994; Whitaker *et al*., 1998; Whitaker *et al*., 2000; Whitaker, 2003; Whitaker, 2004; Whitaker, 2006). These studies concentrate on the relative magnitude of errors relating to sampling and the processing of samples but they provide little or no information on the interaction between the spatial distribution of the mycotoxin and the strategy used to take samples from bulk. It is usually assumed that the samples are independent, but this may not be the case if spatial structure is present.

If spatial autocorrelation is present, it is necessary to design sampling programmes that take the spatial distribution into account, to reduce the probability of falsely classifying a batch below or above the acceptance level (Macarthur *et al*., 2006). For example, hot spots of deoxynivalenol (DON) found in wheat may differ from other mycotoxins in wheat or other commodities because it is mainly generated in the field rather than in storage, which may have an effect on its distribution in subsequent storage and transport the mycotoxin is distributed on a loaded truck. Sampling methods must be both representative and practicable (Stroka *et al*., 2004) so the error associated with the sampling protocol selected is reduced (Whitaker, 2006).

A few studies have looked at the effect that sample size has on determining the mean concentration of DON and ochratoxin A (OTA) in a bulk commodity based on individual incremental samples (e.g. Biselli *et al.*, 2005). Other studies have looked at how aggregation can be characterised using a variety of statistics. For example, Oerke *et al*. (2006) computed Lloyd's index of patchiness for the incidence of several *Fusarium* species in a wheat field. Lloyd's index tests distributions of counts for their similarity to a Poisson distribution, and

does not consider the distance between samples. Similarly, Wilhelm and Jones (2005) compared frequency distributions of *Fusarium* head blight incidence in fields with binomial and β-binomial distributions at scales from 1 m to mesoscale (several counties). Agreement with the β-binomial would indicate greater aggregation. The also performed lag correlation tests using Moran's I, which was one of the first measures of spatial autocorrelation to be developed. Slight non-randomness and autocorrelation were seen at some scales in some cases. Schmale *et al*. (2005), studying spore deposition, had insufficient data points to use lag correlation measures, so applied nonparameteric methods (SADIE statistics and Mantel tests), while noting that the data sets were small even for these methods, and found that spore counts above the canopy were usually random, but deposition counts were often aggregated. However, these approaches give only limited consideration, if any, to the relationship between distance and autocorrelation, and have all considered field data, rather than bulk commodities.

Geostatistics is a branch of statistical science that deals with the spatial structure of the variables under study and has the advantage over other methodologies of accounting for the spatial autocorrelation of the sampled variable, in this case mycotoxin concentration. In geostatistics the spatial variation is considered random and is modelled through a stochastic process. Geostatistics is based on the variogram calculation, a plot that relates the distance between any two sampled points with their semivariance. Effective and efficient sampling strategies can be designed through the characterisation of the variogram parameters: the range, the sill and the nugget. Examples of studies where the variogram was used to analyse the spatial pattern of diseases are provided by Orum *et al*. (1999) and Rekah *et al*. (1999). The former used the variogram to determine the trends and distribution of the aflatoxigenic species *Aspergillus flavus* in soil. The latter used it to study the spatial pattern of *fusarium* crown and root rot in tomatoes.

Geostatistics has also been used as a tool to investigate the spatial dynamics of plant disease propagation. Chellemi *et al*. (1988) used geostatistics to examine the spatial pattern of a population of plant pathogens and diseased plants. Stein *et al*. (1994) applied geostatistics for the analysis of the spatio-temporal pattern of downy mildew pathogen (*Peronospora parasitica*) in cabbage to predict the disease at any point in time, to develop optimal sampling patterns for future assessments, to calculate the expansion rate of the disease and to determine the source of the initial inoculum in space and time. Gottwald *et al*. (1996) analysed the spatial, temporal and spatio-temporal dynamics of citrus tristeza virus (CTV) in Valencia (Spain) to determine the likely rates of disease increase and spread.

As part of a project for the UK Food Standards Agency, six different sources of data were obtained through the literature review and personal contact (Parsons et al, 2007). Table 1 shows a description of each of the data sets considered for analysis. Each set of data was first assessed for its suitability for geostatistical analysis based on the number of points available for the analysis. Webster and Oliver (1992) suggest that variograms computed with fewer than 50 data points are of little value and that at least 100 points are needed for a soil variable to be analysed. The only reference on the number of points required for the variogram calculation of mycotoxins is provided by Stein *et al.* (1994) who determined that 49 observations were necessary to estimate the spatial variogram of downy mildew pathogen in cabbage. Similar requirements would be expected for DON. Those data sets with fewer than 50 points were not considered for analysis. Of the six data sources, only the one collected by Biselli *et al*. (2005) was for mycotoxins in grain and contained sufficient data points with their coordinates.

This study focuses on using a geostatistical approach to characterise the spatial distribution of DON and OTA in the data collected by Biselli *et al*. (2005).

Materials and Methods

Data set

The data set contained DON and OTA results obtained from a 26 t lot of wheat on a truck in Germany. The truck was 2.5 m wide and 10 m long. The data were recorded from 100 points at 20 x 5 grid positions at 0.5 m spacing in the horizontal plane through the truck, using a 5 aperture probe sampler. The probe was vertically inserted into the load in the centre of each grid cell to take a single incremental sample containing grain from 5 depths. Each incremental sample was mixed and sub-sampled before the DON and OTA concentrations were measured (Biselli *et al*., 2005). Thus the final sets of values described the DON and OTA concentrations in a two dimensional horizontal plane aggregated over the depth of the lorry.

Geostatistical approach

Geostatistics describes the spatial autocorrelation among sampled points based on the semi-variogram, a plot that relates the distance between any two points in the space with their semivariance (Figure 1). The semi-variogram shows how similar any two points separated by a distance *h* are: if the semi-variance is small, the points are closely correlated. The distance *h* is known as the lag distance or lag. Each variogram is characterised by three parameters

known as the range, the sill and the nugget (Figure 1). The range is the distance at which the spatial autocorrelation between any two points is lost. The sill is the semi-variance value that corresponds to the range. The nugget is the intrinsic variance of the data (e.g. measurement and sampling error).

The geostatistical analysis included three phases: (i) data preparation, (ii) variogram calculation and (iii) analysis of the spatial structure. Appendix 1 provides a more detailed explanation of the methodology.

The selected data set was prepared for geostatistical analysis. The methods require that the data are normally distributed and stationary, that is they do not exhibit spatial trends. For this purpose, the data were analysed for linear and quadratic trends in the coordinate variables. The data set was considered to present a trend if the model explained more than 20% of the variance. Normality was tested visually by plotting the histogram, box plot (showing median, quartiles, extremes and outliers), and Q-Q plot (quantile-quantile plot, which should be a straight line for a normal distribution). The skewness and kurtosis (Kenney and Keeping, 1961) were calculated. The normal distribution has a skewness of 0 and a kurtosis of 3. If the data do not have an approximately normal distribution, they may be transformed using one of a range of standard transformations to give a close approximation to normality (Webster and Oliver, 2000; Sokal and Rohlf, 1995).

The empirical variogram was calculated in 4 directions defined by azimuth 0°, 90°, 45° and 315° to detect anisotropy (differences in the variogram depending on direction). Azimuth 0° was along the axis of the truck facing forward and angles were measured clockwise in the horizontal plane. The lag distance, azimuth tolerance and maximum distance of analysis were selected according to the results obtained from a sensitivity analysis on these variables. The exponential and spherical models (Webster and Oliver, 2000) were used for the fitting of the variogram. The spatial structure of the mycotoxin was assessed according to the sill, range and nugget values obtained for each variogram.

Results

The OTA data were strongly skewed with mode 0 and therefore could not be transformed to a normal distribution. They were fitted by an exponential distribution with mean 0.57 μg kg^{-1} and standard deviation of 1.13 (Figure 2). Table 2 summarises the descriptive statistics obtained for OTA. The variogram was calculated by treating OTA as an indicator

(present/absent) variable. The results showed a pure nugget variogram: that is, there was no evidence of spatial structure. Figure 3a shows the distribution of OTA in space: OTA presented a random spatial distribution of foci of contamination which is consistent with the variogram results obtained.

The DON data were log transformed (natural log) to meet the normality requirement; Table 2 summarises the descriptive statistics of the original and transformed data. The skewness and kurtosis for the transformed data were close to 0 and 3 respectively, which indicated that the assumption of normality could be accepted. Figure 3c shows the spatial distribution of the log transformed values of DON concentration. In the trend analysis, the linear model accounted for 2.2% of the variance and the quadratic model accounted for 18.9%, which were both below the level at which the data would be considered to present a trend.

The minimum distance between points in the data set was 0.5 m, due to the sampling scheme. The sensitivity analysis on the lag distance showed that a lag distance of 0.7 m provided a good compromise between the number of pairs of points for each lag distance and the number of points in the variogram.

The anisotropy analysis showed that the variogram calculated with azimuth tolerance 60° was significantly different along (azimuth 0°) and across the truck (azimuth 90°). A second analysis was carried out to determine if the anisotropy was due to differences in the spatial pattern of DON or lack of points to compute the variogram across the sampled area. The results showed that the number of pairs of data points per lag distance was significantly smaller when calculating the variogram for azimuth 90°: the apparent anisotropy was an artefact of the smaller number of points across the truck. Therefore, the omnidirectional variogram was used.

The maximum number of pairs of data points for each lag distance was reached at about 4 m for the majority of the azimuths when analysing the longitudinal direction. This indicated that the maximum distance of analysis for the variogram should be around 4 m. The final omnidirectional variogram was calculated with lag distance step 0.7 m and maximum distance 4 m. The spherical model was found to fit better than the exponential.

The results obtained (Figure 4) showed that there was spatial correlation for DON concentration in the selected batch. The variogram was defined by a 4.35 m range, a 0.07 sill and a 0.013 nugget. The non-zero nugget indicates that there was a small amount of variation that either was not captured by the sampling strategy applied or was intrinsic to the

measurements of DON concentration. The variogram sill is generally assumed to be equal to the variance of the population (Barnes, 1991). The sill was consistent with the variance (0.055) obtained from the standard deviation in Table 2.

Discussion

This study was the first application of geostatistical analysis to data on mycotoxin contamination of bulk agricultural/food commodities. Previous studies have used statistics such as the Moran's *I*, SADIE and Mantel's tests and the dispersion index (Wilhelm and Jones, 2005 and Schmale, 2005) to try to detect non-random distributions of fungi or mycotoxins. Most of these ignore the spatial component. Geostatistics provides an explicit characterisation of spatial autocorrelation, so is capable of quantifying effects that would not be detected by other methods. Geostatistics also allows prediction of the value of the variable under study at non measured locations using interpolation techniques such as kriging.

The data set came from a truck load of wheat that was selected for sampling because a high level of contamination was found. The results for this need, therefore, to be treated with caution when attempting to generalise to other situations.

The data set showed clear evidence of spatial structure for DON, but none for OTA. This difference may reflect the fact that DON is mainly produced in the field by a widespread organism, whereas OTA is typically produced in localised 'hot spots' in storage, but this needs to be tested in other data sets, including other commodities. Biselli *et al*. (2005), using classical statistics, also found that the OTA was present in hot spots and had higher variability than DON.

The analysis of the DON data set showed that there was a significant spatial correlation to the distribution of DON mycotoxins in stored grain. The presence of spatial structure implies that samples cannot always be assumed to be independent and that sampling plans need to consider the location of sample points in addition to the number of points sampled in order to obtain reliable estimates of quantities such as the mean contamination and the variance.

The spatial pattern was lost for lag distances greater than 4 m. If this is typical, it has two different implications for sampling, depending on its purpose. If the intention were to characterise the spatial variation, sampling strategies with larger spacing would not detect it. Conversely, points closer than 4 m apart would be autocorrelated, introducing errors or biases into the calculation of sample statistics. Further research should quantify the benefit of

sampling protocols that account for the spatial structure of the mycotoxin when determining levels of contamination to minimise these errors.

Surveillance sampling always aggregates the incremental samples before analysis, so all information about the distribution of contaminants is lost. In order to provide better data for the design of sampling protocols and risk management, there is a need for more good quality data sets in which the incremental samples are analysed and recorded separately. Furthermore, in order to determine the importance of spatial structure in mycotoxins and its potential effect on sampling, experimental sampling should be based on regular grids and record the sample coordinates. On the basis of experience in soil sampling, data sets should contain of the order of 100 points or more (Webster and Oliver, 2000; Webster and Oliver, 1992).

Acknowledgements

This research was funded by the Food Standards Agency Project CO3055. We wish to thank Dr. Scarlett Bisselli (EUROFINS, Hamburg, Germany) for giving permission to use the data analysed here.

Appendix

Geostatistics is based on the variogram model which relates the distance between any two points in the space with their semivariance. Consider a transect along which observations of a variable Z are taken at regular intervals. The positions are denoted as x_i and the value of the observation is $z(x_i)$, i= 1,2…n. The variance of the differences between all the pairs of points at a lag distance h apart can be calculated as follows:

$$\gamma(h) = \frac{1}{m(h)} \sum_{i=1}^{m(h)} (z(x_i) - z(x_i + h))^2 \qquad (1)$$

where $m(h)$ is the number of pairs of data points separated by lag distance h and $\gamma(h)$ is the average variance of all pairs of data points separated by lag distance h. The per-observation variance or semi-variance $\hat{\gamma}(h)$ is half the variance $\gamma(h)$.

The semi-variance is a measure of the similarity between points at a given lag distance. The smaller the semi-variance, the more alike the points are. The graph of semi-variance against lag distance is the experimental semi-variogram (Figure 1). The semi-variogram shows how similar any two points separated by a lag distance h are. In general, the closer any two points are, the more similar their value is. In the empirical semi-variogram the semi-variance increases with lag distance up to a distance a, called the range, at which the semi-variance remains constant. The range is the point at which the autocorrelation between points becomes 0 and marks the limit of spatial dependence: points further apart are spatially independent. The semi-variance value at the range is called the sill (c). The sill is the maximum of the empirical semi-variogram and is the *a priori* variance, σ^2, of the process. The semi-variance at lag distance 0 is called the nugget (c_0) and identifies the measurement error and the variations that occur over lag distances less than the shortest sampling interval (Webster and Oliver, 2000). The sill, the range and the nugget are the three parameters that characterise the semi-variogram.

The model fitted to the experimental semi-variogram is called the empirical semi-variogram (Figure 1) and is a simplification of the experimental variogram. The model is fitted using one of the 'authorised' functions (Webster and Oliver, 2000). The spherical (equation 2) or exponential (equation 3) functions are two of the most frequently used models in geostatistics so these models were selected for the variogram fitting in this study. The fitting of the model was done using ordinary least squares.

$$\hat{\gamma}(h) = \begin{cases} c_0 + c\left(\dfrac{3h}{2a} - \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3\right) & \text{for } h \le a \\ c_0 + c & \text{for } h > a \end{cases} \qquad (2)$$

$$\hat{\gamma}(h) = \begin{cases} c_0 + c\left(1 - \exp\left(h/r\right)\right) & \text{for } h \le a \\ c_0 + r & \text{for } h > a \end{cases} \qquad (3)$$

where c_o is the nugget, c is the sill, a is the range, h is the lag distance and r is a distance parameter that defines the spatial extent of the model. The exponential model approaches its sill asymptotically and therefore, does not have a finite range. Generally, r or effective range is assumed to be the lag distance at which the semi-variance equals 95% of the sill variance, which is approximately $3r$.

These definitions extend to the case of two-dimensional data by measuring the lag distance *h* in all directions instead of along the transect. However, even if the data are sampled using a regular grid, the distances between points are not all multiples of the grid size, because all possible pairs of points are considered, not just those lying on the same row or column. In this case, the semi-variance for a given lag distance is estimated by using all the points separated by distances within a certain tolerance of the required lag distance. In effect, the lag distance axis of the variogram is divided into a series of discrete intervals whose width is the lag tolerance.

For the empirical semi-variogram model to be fitted accurately it is necessary to have sufficient points in the experimental semi-variogram, but for the semi-variance value to be accurate requires sufficient pairs of observations contributing to each point. Increasing the lag tolerance will reduce the first of these, but increase the second. Therefore, the choice of tolerance is a compromise between the accuracy of model fitting and the accuracy of the estimation of the semi-variance.

It is also possible to consider the effect of direction within a two-dimensional data set, by considering pairs of points separated from one another at a specific angle (for example at 45° to the grid rows). The angle relative to some reference direction is called the azimuth. If the variograms differ with azimuth, the data are said to be anisotropic or to present anisotropy. For example, the spatial structure of contamination in a wheat field might differ parallel and perpendicular to the prevailing wind direction or the tramline direction. In a similar way to lag tolerance, the points included for each azimuth are defined by the azimuth tolerance. If the azimuth tolerance is 180°, all directions are included and the variogram is said to be omnidirectional.

The number of points used to estimate each point in the variogram depends on the lag tolerance, as described above, and similarly on the azimuth tolerance. It may also depend on the azimuth itself, either because the sample space has different extents in each direction, or because of the geometry of the sampling grid. Finally, there is a maximum distance beyond which the number of pairs decreases significantly. A sensitivity analysis may be carried out to quantify the effects of these four parameters.

For the geostatistical analysis to be effective the variable under study must be normally distributed, second-order stationary and must present no trend. Data that are not normally distributed may be transformed to achieve normality. A second-order stationary process is characterised by a mean, variance and covariance that depend only on the

separation between points and not on their absolute positions. A systematic component in the spatial variation is an indication of a trend. The trend must be removed from the data when identified so the geostatistical analysis is only carried out for the residuals after subtracting the trend.

References

Barnes RJ. 1991. The variogram sill and the sample variance. Math Geol, 23(4): 673–678

Biselli S, Bruer J, Persin M, Schuh M. Syben M. 2005. Investigation of variability associated with testing lots of wheat kernels for deoxynivalenol and ochratoxin A (case study truck). Paper presented at the World Mycotoxin Forum., 2005. Proceedings of the 3rd World Mycotoxin Forum; Noordwijk, The Netherlands.

Chellemi DO, Rohrbach KG, Yost RS, Sonoda RM. 1988. Analysis of the spatial pattern of plant-pathogens and diseased plants using geostatistics. Phytopathology. 78(2): 221–226.

Gottwald TR, Cambra M, Moreno P, Camarasa, E, Piquer J. 1996. Spatial and temporal analyses of citrus tristeza virus in eastern spain. Phytopathology. 86(1): 45–55.

Jewers K, Bradburn, N, Sharkey AJ. (1988) aflatoxin distribution studies on a 4 tonne batch of maize. Int Biodeterior. 24(4–5): 393–398.

Kenney JF, Keeping ES. 1962. Mathematics of Statistics, Part 1, 3rd ed. Princeton (NJ): Van Nostrand. Section 7.12, Kurtosis; pp. 102–103.

Macarthur R, Macdonald S, Brereton P, Murray A. 2006. Statistical modelling as an aid to the design of retail sampling plans for mycotoxins in food. Food Addit Contam. 23 (1): 84–92.

Oerke EC, Meier A, Dehne HW. 2006. Spatial distribution of Fusarium *spp*. causing head blight in wheat fields. *Proceedings of the 9th European Fusarium seminar, Wageningen.* 19–22. September 2006.

Orum TV, Bigelow DM, Cotty PJ, Nelson MR. 1999. Using predictions based on geostatistics to monitor trends in *Aspergillus flavus* strain composition. Phytopathology. 89(9): 761–769.

Parsons D, Rivas Casado M, Magan N, Dyer C, Weightman R. 2007. Development of representative sampling plans for mycotoxins in foods using distribution modelling. Final Report to the UK Food Standards Agency on Project CO3055. Wolverhampton: ADAS UK Ltd.

Rekah Y, Shtienberg D, Katan J. 1999. Spatial distribution and temporal development of fusarium crown and root rot of tomato and pathogen dissemination in field soil. Phytopathology. 89(9): 831–839.

Schaafsma AW, Tamburic-Ilincic L, Hooker DC. 2005. Effect of previous crop, tillage, field size, adjacent crop, and sampling direction on airborne propagules of *Gibberella zeae*/*Fusarium graminearum*, Fusarium head blight severity, and deoxynivalenol accumulation in winter wheat. Can J Plant Pathol. 27 (2): 217–224.

Schatzki TF 1995a. Distribution of aflatoxin in pistachios. 1. Lot distributions. J Agr Food Chem. 43(6): 1561–1565.

Schatzki TF. 1995b. Distribution of aflatoxin in Pistachios. 2. Distribution in freshly harvested pistachios. J Agr Food Chem. 43(6): 1566–1569.

Schmale DG, Shah DA, Bergstrom GC. 2005 Spatial patterns of viable spore deposition of Gibberella zeae in wheat fields. Phytopathology. 95(5): 472–479.

Sokal RR, Rohlf FJ. 1995. *Biometry: the principles and practice of statistics in biological research.* 3rd edition. WH Freeman and Co.: New York. 887 pp. ISBN: 0-7167-2411-1.

Stein A, Kocks CG, Zadocks JC, Frinking HD, Ruissen MA, Myers DE. 1994. A geostatistical analysis of the spatio-temporal development of downy mildew epidemics in cabbage. Phytopathology, 84(10): 1227–1238.

Stroka J, Spanjer M, Buechler S, Barel S, Kos G, Anklam E. 2004 Novel sampling methods for the analysis of mycotoxins and the combination with spectroscopic methods for the rapid evaluation of deoxynivalenol contamination. Toxicol Lett. 153(1): 99–107.

Webster R, Oliver MA. 1992 Sample adequately to estimate variograms of soil properties J Soil Sci. 43: 177–192.

Webster R, Oliver MA. 2000. Geostatistics for environmental scientists. Chichester: John Wiley and Sons, Ltd.

Wilhelm KP, Jones RK. 2005. Meso- and microscale patterns of Fusarium Head Blight in spring wheat field in Minesota. Phytopathology. 89(5): 474–479.

Whitaker TB, Wiser EH. 1969. Theoretical investigations into the accuracy of sampling shelled peanuts for aflatoxin. J Am Oil Chem Soc. 46(7): 377–379.

Whitaker TB, Giesbrecht FG, Wu J, Hagler WM, Dowell FE. 1994. Variability associated with sampling, sample preparation and chemical testing for aflatoxin in farmers stocks of peanuts. J Assoc Offic Anal Chem. 77(1): 107–116.

Whitaker TB, Trucksess M, Johansson AS, Giesbrecht FG, Hagler WM Jr, Bowman DT. 1998. Variability associated with testing corn for fumonisin. J Assoc Offic Anal Chem. 81(6): 1162–1168.

Whitaker TB, Hagler WM Jr, Giesbrecht FG, Johansson AS. 2000. Sampling, sample preparation, and analytical variability associated with testing wheat for deoxynivalenol. J Assoc Offic Anal Chem. 83(5): 1285–1292.

Whitaker TB. 2003. standardisation of mycotoxin sampling procedures: an urgent necessity. Food Control. 14 (4): 233–237.

Whitaker TB. 2004. Mycotoxins in food: detection and control (eds. Magan N, Olsen M). Cambridge: Woodhead Publishing Ltd. Chapter 4, Sampling for mycotoxins: pp 69–87.

Whitaker TB. 2006. Sampling foods for mycotoxins. Food Add Contam. 23(1): 50–61.

Xu XM, Parry DW, Nicholson P, Thomsett MA, Simpson D, Edwards SG, Cooke BM, Doohan FM, Monaghan S, Moretti A, Tocco G, Mule G, Hornok L, Béki E, Tatnell J, Ritieni A. 2008. Within field variability of Fusarium head blight and its associated mycotoxins. *European J. Plant.* Pathol. 120(1): 21-34.

Tables

Table 1. Description of data sources collected for analysis and criteria for rejection ([a] insufficient measurements available; [b] lack of coordinates identifying the location of each measurement, [c] the original data were not available)

Source	Characteristics
Xu *et al*. (2008).	[ab] Measurements of *Fusarium* head blight were taken at five study sites. At each site 16 quadrants of 0.5 m x 0.5 m were randomly selected along a W-shape walk.
Biselli *et al*. (2005).	The data were recorded from 20 x 5 grid positions at 0.5 m spacing in the horizontal plane through a truck containing wheat, using a 5 chamber probe sampler. After aggregation and subsampling, DON and OTA concentrations were measured.
Oerke *et al*. (2006).	[a] Measurements of 8 *Fusarium* species in wheat were recorded at 5 x 6 grid points on a skewed grid with spacings of 12 x 18 m.
Schaafsma *et al*. (2005).	[a] A total of 68 wheat fields were sampled for *Fusarium*. In each field a transect was selected. DON content was measured from samples of wheat heads randomly hand-harvested from nine traps equidistantly distributed along the transect.
Wilhelm and Jones (2005).	[c] *Fusarium* head blight data in wheat field were collected at four different sampling resolutions: mesoscale, full-field, microscale and adjacent-scale. For mesoscale study 100 points were collected at sixty fields. For the full-field scale nine fields were analyzed with a total of 45 points per field. For the microscale analysis three different grid resolutions were analysed. For the adjacent-head scale a total of 60 consecutive wheat heads were sampled in each of the each of the eight plots.
Prof. P. Battilani, Catholic Univ. of Italy, Piacenza, Personal Communication	[b] A total of 10 measurements of *F. verticillioides* and fumonisins were taken in a 10 m x 10 m area along the diagonals of each of three maize storehouses. Two different sub-samples were measured at each point at two different depths. The samples were aggregated and only the average for the three storehouses were reported.

Table 2. Descriptive statistics for the data analysed.

Statistic	**OTA** μg kg^{-1}	**DON** μg kg^{-1}	**DON transformed**
Number of points	100.0	100.0	100.0
Minimum,	0.05	830.0	6.72
Maximum	8.6	2655.0	7.88
1st Quartile	0.05	1121	7.02
3rd Quartile	0.52	1508	7.31
Median	0.2	1272	7.14
Mean	0.59	1342	7.17
Standard deviation	1.13	339	0.23
Skewness	4.36	1.19	0.48
Kurtosis	30.58	7.77	3.2

Figures

Figure 1. Diagram showing the variogram parameters (i.e. range, sill and nugget). The diagram shows both the experimental (dots) and the empirical (line) variograms.

Figure 2. Ochratoxin A concentration: histogram of observations and fitted exponential distribution.

Figure 3. Spatial distribution of (a) OTA, (b) DON and (c) $\log_e$ transformed DON. All distances are in metres. Data collected are represented by a cross and each cell has been coloured according to the measured value.

Figure 4. Spherical variogram for the $\log_e$ transformed DON data (omnidirectional variogram) calculated with lag distance 0.7 m and maximum distance 4 m.

434