# Final Report

# PCN Soil Sampling

**Ref: 1100018**

**Report Authors: Adrian Roberts, Kim Davie, Jon Pickup**

Report No. 2018/5

*AHDB Potatoes is a division of the Agriculture and Horticulture Development Board.*

Additional copies of this report and a list of other publications can be obtained from:

**Publications**

AHDB Potatoes
Stoneleigh Park
Kenilworth
Warwickshire
CV8 2TL

Tel: 02476 692051
Fax: 02476 478902
E-mail: Potatoes.Publications@ahdb.org.uk

Our reports, and lists of publications, are also available at potatoes.ahdb.org.uk

# CONTENTS

# 1. SUMMARY

## 1.1. Aim

The aim of this project was to assess different sampling strategies for potato cyst nematode (PCN) based on our knowledge of PCN distributions in British field situations.

There is a key difference in sampling approach according to the purpose of the sampling. This is either:

- detection of an early stage infestation; or
- estimation of the PCN population in later stage infestations.

The effectiveness of the range of sampling strategies in common use was compared. From our analysis of these strategies, combined with an understanding of the purpose of the soil sampling, we have provided recommendations for effective sampling and methods of laboratory analysis. We also advise on whether any further investigative work is required to fill gaps in our knowledge that prevent significant conclusions from being reached.

## 1.2. Methodology

Mathematical models were used to represent the features of PCN distributions over a field. The work is based on the assumption that sampling is carried out on 1ha units. A review of potential mathematical models for the spatial distribution of PCN in a field has been carried out. The Been and Schomaker model (developed in the Netherlands) was used to simulate early stage infestations, based on cysts/kg. Other distributional models, less specific to PCN, were used to study sampling under later stage infestations, based on eggs/g and after fitting to example data sets. We used fields with known distributions of PCN and applied three different models to these. We looked at the likelihood of detection of cysts under different intensities of infection as well as the precision of the density estimate. We investigated the use of different core sizes, patterns of sampling in the field and the final soil sample sizes, including the merits of subsampling of soil samples in the laboratory.

## 1.3. Key findings

- The work is based on the assumption that sampling is carried out on 1ha units. We did not examine the impact of scaling up to sampling on larger areas (e.g., 4ha units). There is insufficient field data on PCN distributions in GB to draw firm conclusions on the impact of using larger unit areas. The guidance from the Nematicide Stewardship Programme (NSP) was that the use of 1ha blocks is accepted in the industry and we focused on this approach.
- The purpose of the soil test, whether for detection or for quantification of infestation, is critical in determining how soil samples should be taken and how they should be subsampled in the laboratory.

- As statutory measures prohibit seed potato production in PCN-infested land, we concluded that the models developed by Thomas Been and Corrie Schomaker describing the highly aggregated foci (models that underpin the NemaDecide decision support programme) are the best available for describing early infestations in previously uninfected land. These distribution models have been used to evaluate sampling strategies aimed at the detection of PCN infestations.
- For detection of early stage infestations the key factor is the amount of soil tested at the laboratory. The amount of soil tested at the laboratory is generally more important for detection than the number of cores taken.
- A different approach was taken for the quantification of PCN populations within infested land. We collated available field data from a variety of sources and standardised these data sets for modelling purposes. We used a range of mathematical models that have been previously used to describe similar data, and established that the zero-inflated negative binomial distribution provided the best fit to PCN distributions in PCN-infested fields used for ware potato production in Britain.
- Based on a relatively small set of field sampling data, fields with lower levels of infestation seem to be generally more heterogeneous (patchy). This relationship was built into our models.
- For estimation of population levels (in eggs/g of soil) in later stage infestations, the number of cores becomes more important than the amount of soil tested, but only if the eggs are widely dispersed in many cysts. If the eggs are aggregated into relatively few cysts, the amount of soil tested also becomes important, particularly at lower levels of infestation.
- At a PCN population level of 10 eggs/g, which is sometimes quoted as a threshold for management decisions, the volume taken to create a subsample from a soil sample consisting of 50 cores of 11.8 ml makes minimal difference if it is assumed that each egg is to be found in a separate cyst. Our calculated confidence limits are 7.2 to 12.8 eggs/g for both 100 and 400 g subsamples. However, if the eggs are aggregated into cysts containing 10 eggs per cyst, the confidence intervals increase to 6.7 to 13.4 eggs/g for a 100 g subsample. For an aggregation of 50 eggs/cyst, the confidence intervals increase to 4.9 to 15.1 eggs/ml for a 100 g subsample. For an aggregation of 200 eggs/cyst, the confidence intervals are 0.9 to 19.1 eggs/ml for a 100 g subsample. At this level of aggregation, a 400 g subsample would reduce the confidence interval to a more acceptable 5.0 to 15.0 eggs/ml.
- For Britain, there is a general acceptance that populations of *G. pallida* decline by about 20% p.a., whereas the decline rate of *G. rostochiensis* is slightly higher at 30%. Based on such assumptions, and a conservative initial cyst content of 300 eggs/cyst, an aggregation of 200 eggs per cyst may be expected with a *G. pallida* infestation and a potato crop rotation of 1 year in 3, and an aggregation of 50 eggs per cyst may be expected with a *G. rostochiensis* infestation and a potato crop rotation of 1 year in 6.

- Based on the assumption that the aim is to detect a population level of 10 eggs/ml and that the maximum confidence interval should be 5 to 15 eggs/ml, the minimal recommended laboratory subsample size should be 200 g. Where *G. pallida* is present, and rotations are less than 1 year in 6, a 400g subsample is preferable.
- Further work to evaluate typical numbers of eggs per cyst and how this relates to infestation levels in British fields is required. However, given the variation in field cropping histories and PCN management practices it may not prove possible to generalise beyond estimations based on species and decline rates.
- Previous studies have shown that the depth at which soil is sampled is not important provided the sample is taken from within the top 20 cm, as cysts have been found to have a uniform distribution at this level.
- Spatial sampling recommendations have been based on the distributions found in the Dutch studies of Been & Schomaker. Insufficient spatial data were available from British fields for drawing any conclusions on recommended sampling patterns based on data from British data, particularly from ware potato production.
- In principle, sampling on a grid pattern is more effective than a W-pattern. However, for a relatively small increase in error, using a W-pattern can reduce the time spent and hence the costs of sampling, with a 4 armed W-pattern a sensible alternative. If infestation foci are present and elongated in the direction of cultivation, the orientation of the W-pattern is important.

## 1.4. Practical recommendations

Prior to taking a soil sample the objective of the sampling should be established:

### 1.4.1. Detection of early stage infestations

Sampling for detection is appropriate for ware land with an unknown history of PCN, land which has not been in potato production for a long time, or land which has previously had a low infestation of PCN but control measures, such as the use of resistant cultivars, have been applied.

In these cases the purpose of sampling is to determine the presence or absence of PCN, as in the case of seed land. The more intensively a field is sampled, the greater will be the probability of detecting low level PCN infestations.

The standard rate for land destined for seed production specified in the European Council PCN Directive (2007/33/EC) is 100 cores to make up a 1500 ml sample from each ha unit, with the whole sample being processed.

Detection of a low population of cysts, e.g. less than 500,000 cysts per hectare is unlikely unless very high volumes of soil are tested. Table 1 (below)

summarises the volume of soil generated using a standardised sampling protocol (13mm diameter corer, inserted to 25cm depth with 49 cores per hectare) and the impact of sub sample size on the probability of detection. Choosing to have only a small amount (e.g., 200ml) of the soil that has been collected tested for PCN has a substantial impact on the probability of detecting the pest.

***Table 1*:** Comparison of the impact of sub sampling in the laboratory on the probability of detection of 5,000,000 cysts per hectare.

| Corer size (diameter mm) | Depth (cm) | Cores per ha | Approx volume (ml) of soil collected (per ha) | Approx dry weight soil (g) / ha* | Amount (g) of sample processed | Probability of detection |
|---|---|---|---|---|---|---|
| 13 | 25 | 49 | 1625 | 2600 | 2600 | 67% |
| 13 | 25 | 49 | 1625 | 2600 | 1000 | 48% |
| 13 | 25 | 49 | 1625 | 2600 | 600 | 38% |
| 13 | 25 | 49 | 1625 | 2600 | 400 | 31% |
| 13 | 25 | 49 | 1625 | 2600 | 200 | 20% |
| 13 | 25 | 100 | 3319 | 5310 | 400 | 39% |
| EU Directive - employed for seed potato land (included for comparison): | | | | | | |
| 18 | 6 | 100 | 1500 | 2400 | 2400 | 81% |
| 11 | 4 | 100 | 400 | 640 | 640 | 40% |

* Based on a bulk density of 1.6g/cm$^3$

For detection a grid pattern is more effective that a W-pattern and this should be used if the cost does not make it prohibitive.

## 1.4.2. Estimation of the PCN population in later stage infestations.

For estimation purposes, the greater the number of cores the more confidence can be given to the level of eggs per g found in the soil. A minimum of 49 cores per hectare should be taken with a recommendation to take as many as feasible, e.g. taking 200 instead of 50 cores per ha would reduce the confidence limits associated with a population level of 10 eggs/g from ±2.8 to ±1.5, providing that eggs are evenly distributed within a sample. Where there is a high level of infestation in the soil (e.g. over 50 eggs/g), then fewer cores may be required to obtain sufficient accuracy, in part, due to the increasing levels of uniformity of the PCN distributions within the field.

The amount of soil tested for estimation is also important once the aggregation of eggs within cysts is considered. We would therefore not recommend taking a subsample of less than 200g where *G. rostochiensis* is present and rotations are at least one in 6. A minimum of 400g subsample should be used if *G. pallida* are present and where rotations are less than one in six.

According to the literature, the depth to which cores are taken will have little effect providing they are within the top 20 cm of soil. The time of year that soil samples should be taken is dependent on practicalities and the purpose for sampling. For either detection or quantification of PCN prior to planting a potato crop the optimal time for sampling will be in the autumn or winter before the crop is sown to best determine the impact of PCN on that crop.

Volunteers/ground keepers in a field can maintain or increase PCN numbers and if they have been present a minimum of 400g should be processed by the laboratory.

# 2. INTRODUCTION

General introduction

Potato cyst nematodes (PCN) *Globodera pallida* and *G. rostochiensis* are considered to be one of the greatest concerns for economic potato production in Britain. PCN feed on the roots of potato plants, compete directly with the plant for resources and limit root growth. For this reason plants are often stunted and low yielding, particularly when plants are facing additional stresses. Severe infestations of PCN can lead to plant death.

With every susceptible potato crop that is grown, PCN can multiply up to 100-fold (Evans and Kerry, 2007) and will, if not appropriately managed, lead to land becoming unsuitable for the commercial production of potatoes. To determine whether PCN are present in land and, if present, to assess the species and population level, field soils are sampled and tested for PCN. In a one hectare field, the top 20 cm depth of soil equates to 2 million litres of soil, only a very small proportion of which can feasibly be tested. PCN are unevenly distributed on a horizontal scale across the field. This aggregated distribution reduces the probability of detection. Therefore, the challenge for soil sampling is how to take a soil sample that best represents the PCN status of the field for submission for laboratory testing without the cost becoming prohibitive. The larger the sample, the more expensive the test will be. Field sampling costs are mostly affected by the time taken to draw the sample, this will depend on how many cores and the pattern used to traverse the field. Finding low level and patchy PCN infestations in a field is particularly challenging. A single cyst in a 400 ml sample of soil drawn from 1 ha roughly equates to a population of PCN of 5 million cysts/ha, a level which is likely to have taken several field generations (typically at least 4, depending on rotation periods and varietal susceptibility) to reach a detectable level. Earlier detection of PCN would require soil sampling at a rate that is at least an order of magnitude more intensive and with a proportionate increase in laboratory costs. Therefore, it is unrealistic to expect to be able to reliably detect PCN populations at very early stages of infestation. Fortunately, detection of PCN by soil sampling is feasible before the pest can cause significant economic damage. Once present in fields, the tools to manage PCN and provide commercial yields of potatoes are available, so monitoring and estimating population levels of PCN becomes critically important in underpinning the management decisions required to

sustainably manage PCN populations in infested land and to minimize the economic impact of PCN on the potato crops grown in that land. Therefore there are two purposes for testing fields for PCN: detection i.e. is PCN present; and quantification, i.e. how much PCN is present.

Testing fields for PCN involves two key elements: soil sampling and the subsequent laboratory analysis. Both are viewed as hugely important issues by British growers and agronomists and there is a feeling among them that there is room for improvement in both areas. In particular, there has been a lack of clear guidance on the best method to sample fields. Several sampling strategies are deployed across the industry and the relative merits of the different approaches have not been critically assessed. This report seeks to address these concerns. Through the commissioning of this report, the industry (largely via consultation with the Nematicide Stewardship Programme (NSP)) seeks guidance over the potential to standardise soil sampling.

In terms of PCN testing prior to seed potato production, the sampling criteria are already laid out by the European Council PCN Directive (2007/33/EC). Both species of PCN are listed by EPPO on the A2 list, i.e. quarantine pests that are locally present in the EPPO region. Accordingly, Directive 2007/33/EC requires the place of production for seed potatoes to have been tested and found free from PCN. Therefore, sampling prior to seed production is targeted at detection – there is no tolerance level for PCN, so there is no requirement to estimate population levels. Directive 2007/33/EC recognises that although a quarantine organism, PCN is present in a significant proportion of the ware potato production area of the EU and therefore it would be unreasonable, given the measures available to manage PCN, to prohibit potato production on infested land. There is a requirement to monitor PCN populations in ware potato land to provide the necessary information to underpin management programmes and to evaluate the effectiveness of such programmes.

**Therefore, the focus for this report is the sampling of ware potato land, both for the presence of PCN and for the monitoring and management of PCN populations where there are known infestations**. One question asked is whether a single soil sampling protocol can be provided that would work for all fields and for all levels of infestation. If so, at what depth and size should sample cores be, how many cores must be taken to provide accurate population estimation, does the pattern of sampling matter, how much soil should be taken and how much of this soil should be tested – can the sample be subsampled in the laboratory for example? The industry would also like to know when the most appropriate time to sample is - whether this is shortly before the crop is planted, after it is harvested or at some other interval.

When sampling a field for PCN, the smaller the size of the block or area from which the sample is drawn, the greater will be the discrimination between areas of differing PCN incidence. Consultation with the NSP consortium indicated that in Britain sampling on a scale of a 1 ha block is feasible. We have therefore based the vast majority of the work of this project on the assumption that we are sampling a unit of 1 ha.

Factors influencing PCN population dynamics

The development of cyst numbers from year to year at a single site depends on a number of factors, including species of PCN, potato cultivar, rotation length grower control interventions, temperature and the initial PCN population (van den Berg *et al*., 2006; Trudgill *et al*., 2014; Kaczmarek *et al*., 2014; Evans, 2015).

**Varietal resistance/species of PCN**: Resistant cultivars can help to maintain populations at low levels (Turner and Fleming, 2002). The population of PCN will increase on every susceptible crop that is planted, with potential increases of over 50-fold occurring on each host crop. Evans & Kerry (2007) state a potential increase of 100-fold, but for the purposes of this report we prefer to use a more conservative estimate of 50-fold. Resistant cultivars limit PCN population increases. The 1-9 scoring scale published by EPPO and encapsulated in the EU PCN Directive (EPPO bulletin, 2006) categorizes cultivars according to their resistance, with an increase in the resistance score of 1 point equating to a 50% reduction of the PCN increase. Thus if a 50-fold increase from the initial population is recorded on a susceptible cultivar (Score of 2), growing a cultivar with a resistance score of 3 will be expected to result in a 25-fold increase, a score of 4 to result in a 12.5-fold increase, and so on. Highly resistant cultivars with a score of 9 result in a PCN population that is below 1% of final population on a susceptible cultivar and can be expected to produce a reduction in the original PCN population in excess of 50%. There are a range of cultivars with resistance to one or both species of PCN, with British growers having access to many cultivars with high levels of resistance to *G. rostochiensis*, but only recently have commercial cultivars with equivalent resistance to *G. pallida* started to become available. Many popular British cultivars, e.g. Maris Piper have H1 gene resistance to *G. rostochiensis* (score of 9), but no resistance to *G. pallida* (score of 2), therefore knowing the species of PCN present is essential.

**Initial PCN population/varietal tolerance**: At higher initial populations of PCN, damage to the plants caused by the nematodes will limit the extent to which the PCN population will increase. The damage caused to the plant will also limit the commercial yield from the crop. Some cultivars which are highly tolerant of PCN damage, e.g. Cara, are capable of producing vigorous growth and maintaining yield at moderately high PCN populations. Such tolerant cultivars are also 'successful' in allowing PCN to develop to exceptionally high population levels. The PCN population is eventually self-limiting on all cultivars; if the density is high enough the host plant will die (Been & Schomaker 2006). Tolerant cultivars can withstand much higher PCN population levels than intolerant cultivars.

**Decline rates**: The rate of decline is well studied (Oostenbrink, 1950; Whitehead and Turner, 1998; Trudgill *et al*., 2014). Long rotations are often used as a tool for managing PCN, taking advantage of the natural hatch of c. 30% of the population that occurs each year in the absence of a host plant. Different studies have suggested vastly different decline rates: Devine *et al*. (1999) suggested that the number of viable eggs each year is reduced by about

10%, but in sandy soils decline rates of as high as 60% per annum have been recorded (Cole and Howard, 1962). This spontaneous hatch is generally cited as the reason for the natural decline of PCN in an infested field. This decline can vary with differences in soil composition, soil type and other environmental factors, including aeration and moisture, and is therefore difficult to predict. *G. pallida* is generally thought to reduce more slowly than *G. rostochiensis*. Although populations will decline in terms of eggs per volume of soil, cysts can be very long-lived, so the decline is rarely matched in terms of cysts per volume of soil.

**Soil management practices – Nematicides, trap crops, sterilants and biofumigants**: A range of options to reduce PCN populations in addition to relying on natural decline are available to the grower in the form of these additional management tools. Nematicides are usually used at planting so as to have maximum effect in controlling the juveniles emerging from cysts in response to chemical exudates from the roots of the growing potato crop. Treatments can prevent early damage to root systems of the plants and help to achieve a healthy yield. Nematicides reduce the initial population invading the plant roots, but those nematodes that are successful in invading roots still undergo a rapid multiplication on susceptible hosts. To manage the post-harvest PCN population within a field, a cultivar with resistance to the PCN present in the field should be used. Sterilants and biofumigants act on the dormant eggs within the cysts to reduce the population. Success rates for sterilants typically vary from a 50-90% reduction in the PCN population which is highly beneficial to the grower, particularly in reducing yield loss from a subsequent potato crop. However, as with granular nematicides, if the subsequent potato cultivar is susceptible to the PCN present, the consequent 50-fold population increase will more than negate any potential benefit in terms of PCN population management. Trap crops can stimulate a hatch of the eggs within the cysts without providing a host suitable for population increase, but are not widely practicable for use in Great Britain at present.

<u>Development of PCN infestations</u>

PCN can be introduced into a field by a number of means. Introduction by contaminated seed has been traditionally considered as the most likely means, with population build up around the original introduction site forming a focus within the field. This could occur more than once due to the simultaneous introduction of several sites of infection produced by planting the same infected seed lot, thus leading to the development of several foci within one field. Repeated cultivation of the field will lead to an increase in the spread from the original populations. Subsequent introductions will also lead to the build-up of additional foci.

Other potential means of introduction are from water courses, either by rivers and streams carrying cysts from infested land and subsequently flooding onto potato land or through use for irrigation, or through run-off from neighbouring infested fields. In these cases the resulting field distribution is likely to be different from those introduced by the planting of infested seed. Strong winds can also be responsible for large-scale movement of PCN, particularly in areas

where soils can be dry for prolonged periods. PCN can also be introduced on the feet of animals or with the movement of contaminated machinery. The means of introduction will determine the initial population that will be bulked up by each subsequent susceptible crop.

The population of PCN will increase with every susceptible crop that is planted. Shorter rotations allow the population to increase more rapidly, with potential 50-fold increases occurring on each host crop and shorter intervals between host crops providing less time for natural population decline. Shorter rotations are also more likely to intensify the aggregated distribution of PCN in the field as there is less time for foci to be dispersed around the field during general cultivation of the land. Typical rotations for ware crops in Britain are around five years (Minnis *et al*., 2002). In the Netherlands, where highly aggregated populations of PCN have been recorded, many infestations are on relatively recently reclaimed land so it is likely that the PCN were originally introduced with seed and then increased under frequent rotations. Typically potatoes are grown every three years in the Netherlands, potentially increasing PCN populations to detectable levels in as few as four rotations, and producing the highly aggregated distributions described by Been and Schomaker (2000). Once PCN are so abundant, spread by other means within field becomes more likely – and the PCN distributions are likely to become less heterogeneous.

In Britain, some PCN infestations are known to be 70 or more years old, e.g. SASA has evidence from fields near cities in Scotland suggesting potato crops were grown on almost annual rotations during the Second World War to meet with demand and limited availability of transport. Consequently, populations of PCN reached very high levels, inevitably leading to poor yields. The land has subsequently been brought back into potato production with the PCN population managed by long rotation. In these cases the expectation is of far more homogeneously distributed populations of cysts. Extensive cultivation of the land over many years will lead to a widespread distribution of PCN across the field. In reality there will be many distributions of PCN with levels of aggregation that are intermediate between these two extreme scenarios.

## Reasons for soil sampling and current recommendations

### *PCN testing for land used for seed potato production*

There are several reasons for testing fields for the presence of PCN. Where it is intended that land will be used for seed potato production, it is a statutory requirement that this land is tested under the European Council Directive (2007/33/EC). Where any viable PCN cysts are found, this land cannot be used for seed potato production. For this purpose there is a standard sampling rate of 1500 ml/ha with a minimum of 100 cores/ha preferably taken using a grid sampling pattern of not less than 5 metres in width and not more than 20 metres in length. The sampling rate can be reduced to 400 ml/ha (still taking 100 cores/ha) where there is a lower risk of finding PCN, i.e. rotations of potato crops of at least 7 years or where there is a testing history of PCN freedom. These sampling rates can be further reduced for large fields. So, for fields to be

sampled at the standard rate, an average sampling rate is calculated on the basis of 1500 ml for each of the first 8 ha, with 400 ml for each additional ha. For fields to be sampled at the reduced rate, an average sampling rate of 400 ml for each of the first 4 ha plus 200 ml for each additional ha is calculated. Prior to the introduction of Directive 2007/33/EC, soil sampling rates for seed were typically 600 ml per unit of up to 4 ha, which in Scotland equated to an average sampling rate of 185 ml/ha.

*PCN testing for land used for ware potato production*

In the case of ware crops planted on PCN infested land, the responsible use of nematicides is strongly encouraged via nematicide stewardship schemes. Under these schemes, fields should have tested positive for PCN prior to application of nematicide. For this purpose, the main nematicide producers have provided their own soil sampling recommendations (see their respective stewardship guides).

## Threshold values for PCN management

Two basic factors should be considered in relation to PCN Management: the protection of the crop; and the protection of the land.

**Protection of the crop**: The point at which PCN begin to cause damage in a crop is dependent upon the population level/viability of the cysts, the tolerance of the potato cultivar, the soil type and other environmental factors and therefore threshold values for PCN are difficult to set. Many threshold recommendations are based on ADAS advice originally provided in the 1970s which suggests that where no viable cysts are found the field is safe for potato cultivation, but the land should be sampled again prior to the next crop. Where 1-10 eggs/g are found they considered this a low value, although at above 5 eggs/g they recommended the use of a nematicide. They classed 11-60 eggs/g as a moderate infestation and over 60 eggs/g a high infestation.

**Protection of the land**: A single crop of a susceptible cultivar has the capacity to increase the PCN population by over 50-fold, i.e. taking the population from the limits of detection to highly damaging in one cropping season. Therefore, to protect the land, it is highly advisable not to grow cultivars susceptible to the species of PCN present without taking additional measures to mitigate against the increase.

## Describing the distribution of PCN within fields

The distribution of PCN within a field has a critical influence on the effectiveness of any sampling scheme. Distributions that are more highly aggregated require more intensive sampling than less aggregated or random distributions. The spatial nature of the aggregation will determine the optimum approach to determining the frequency and distribution of sampling points (cores).

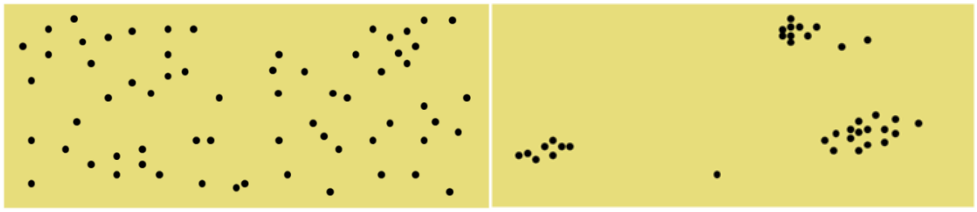


**Figure 1:** *Left panel: random distribution. Right panel: aggregated distribution*

Gilligan (1988) sets a baseline for any spatial pattern as a random distribution. A random pattern means that the chance of a cyst being found at any particular point in the field is neither dependent upon the positions of other cysts in the field nor the position within the field. There can be deviations from randomness in two directions; towards more aggregated distributions and towards more regularly dispersed patterns. In the first case cysts are more likely to occur in the vicinity of other cysts, and in the latter case they are less likely to be near other cysts. Gilligan describes different statistics that can be used to highlight different types of patterns. Because of the nature with which fields become infested, PCN infestations often occur in highly aggregated foci. At very high infestation levels these foci can have a visible effect on crop growth.

Photo of PCN infestation (courtesy of Andy Barker)

Been & Schomaker (2006) clarify that patterns should be linked to scale. This varies from a very fine scale, e.g. mapping the distribution of cysts in and around the roots of a single potato plant, through to field level and then at a landscape level. The type of pattern seen will vary according to the scale but the nature of PCN means that some aggregation is expected at all levels. Note that at a field level, the tendency of cysts to cluster around potato plant roots would tend to induce a regular pattern (reflecting the distribution of the plants). However this is likely to be dominated by the tendency of PCN to build slowly from foci, being spread by cultivation.

The nature of the PCN life cycle means that unassisted movement of the nematodes is slow. PCN are only capable of moving short distances, 10 - 50

cm per year, without assistance from external factors, e.g. cultivation processes such as ploughing. Where there is a predominant direction to this movement over the years, the resulting hotspot is usually skewed in the direction of cultivation (Schomaker & Been, 1999). New foci can arise in the field subsequently through further contamination within, or between, fields.

Methodology for describing and analysing spatial point patterns in a general context is given by Diggle (2003), for soil-based pathogens and diseases by Nicot *et al*. (1984); Campbell & Noe (1985); Gilligan (1988); Madden *et al*. (2007); and specifically for PCN by Evans *et al*. (2003); and Been & Schomaker (2006).

In modelling populations of PCN in a real field, although soil samples may have been taken from points all over the field (see below), information may be stored at different levels of complexity:

a) the total count for the area;
b) counts for each point sampled but with no location information;
c) counts for each point sampled with the corresponding locations.

In the case of a) we can learn little about the degree of aggregation (or heterogeneity) of the PCN. Levels b) and c) are termed sparse sampling and intensive mapping respectively (Diggle, 2003). In the case of sparse sampling, it is possible to learn about the degree of heterogeneity of the PCN population. In the case of mapped counts, it is also possible to study the pattern of the infestation. Note that the costs for obtaining information increases from level a) to level c).

In the case of sparse count data, indices can be used to highlight the presence of heterogeneity or aggregation (Campbell & Noe, 1985; Gilligan, 1988). A commonly used method is to compare the variance of the counts with the mean count (Madden *et al*. 2007). In the case of a random distribution of PCN, the probability distribution of counts is Poisson and the variance will be similar to the mean. If the variance is greater than the mean, the implication is that the distribution is aggregated.

If the distribution of counts is aggregated then there are several potential probability distributions that may fit. A commonly used distribution is the negative binomial distribution (McSorley & Parrado, 1982; Gilligan, 1988; Madden *et al*., 2007). Others use the Neyman Type A distribution (McSorley & Parrado, 1982; Campbell & Noe, 1985; Gilligan, 1988) but the availability of statistical software for modelling this distribution is much more restricted than it is for the negative binomial. Neither of these distributions allow for the possibility that some parts of the field may be absolutely free from PCN. To encompass this possibility, it may be necessary to allow for an excess of zero counts compared to e.g. a negative binomial distribution. A potential solution lies in the zero-inflated negative binomial distribution (e.g. Denwood *et al*., 2008), a distribution that doesn't seem to have used in the context of PCN until now.

If mapped count data are available then several methods are possible for identifying spatial patterns (Nicot *et al*., 1984; Noe & Campbell, 1985; Madden

*et al*., 2007). The simplest is just to produce plots representing the raw counts and their field positions. However, to infer a picture of the PCN levels across the whole field, it is necessary to model the spatial pattern. This can be done through autocorrelation methods, perhaps involving the statistical method of spectral analysis, (Campbell & Noe, 1985) or geostatistical methods involving kriging (Wyse-Pester *et al*., 2002; Evans *et al*., 2003).

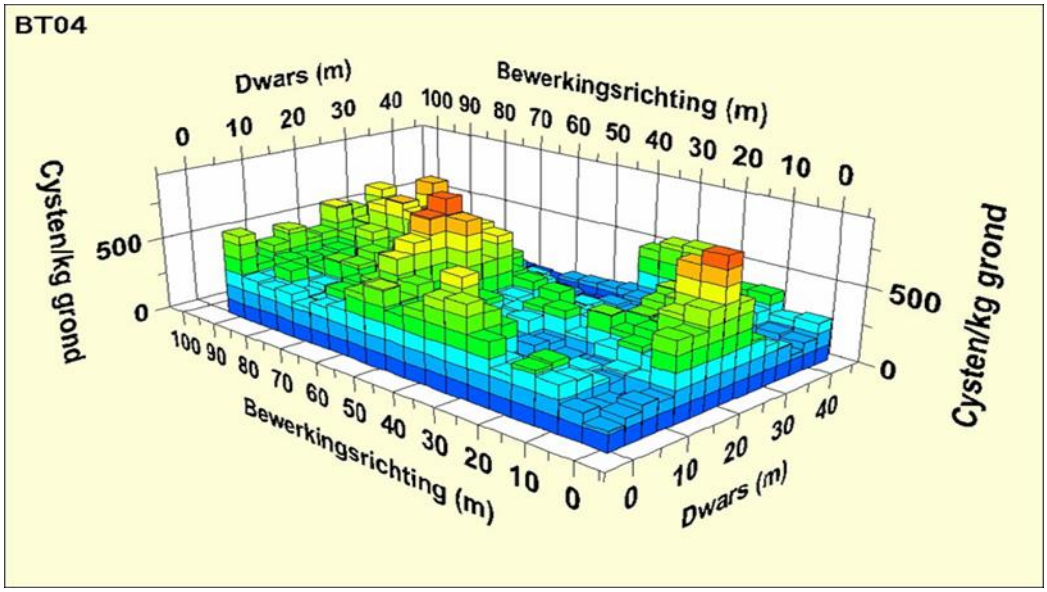



**Figure 2:** *Typical PCN distribution in the Netherlands (courtesy of Thomas Been) Translation: bewerkingsrichting – distance in direction of cultivation; dwars – distance across cultivation; cysten/kg grond - cysts/kg of soil*

Whilst these statistical spatial methods are good for illustrating and understanding the pattern of PCN infestation in individual cases, they may be less useful in setting out generally applicable models which can, in turn, drive decisions on sampling strategies. Alternatively, Schomaker & Been (1999) proposed a more mechanistic model that has the benefit of using the mapping information, whilst also forming the basis for understanding the benefits of different sampling strategies (Been & Schomaker, 2000).

The Schomaker & Been model builds in an understanding of the nature of nematode infestations, i.e. that populations build from initial points of arrival and that cysts are moved relatively slowly and predominantly by cultivation. They reviewed infestations in many potato fields in the Netherlands, finding that infestations were “approximately lozenge-shaped and cysts densities decreased exponentially away from the focus centre, but more slowly in the length than in the width direction”. They constructed a relatively simple model for such distributions. Two parameters, defining the rates of decline in the direction of and perpendicular to the direction of cultivation, were fitted in each field. Estimates for these parameters have been used for subsequent evaluation of sampling schemes. Thus with the model, a field infestation, can be defined simply by the number of hotspots and density of PCN at their centres

(the central population density, CPD). This method has been the core of the Dutch PCN decision-making system (now incorporated into NemaDecide) over recent years and is well described (e.g. Been & Schomaker, 2000, Been, Schomaker and Molendijk, 2007). Note that this model describes the "medium-scale" distribution of PCN. For the small-scale distribution (approximately 1 $m^2$), a negative binomial distribution is used to account for the high local variability in PCN counts. Again the authors fitted these models to the infested fields, leading to an estimate of the degree of aggregation that could then be used in sampling simulations.

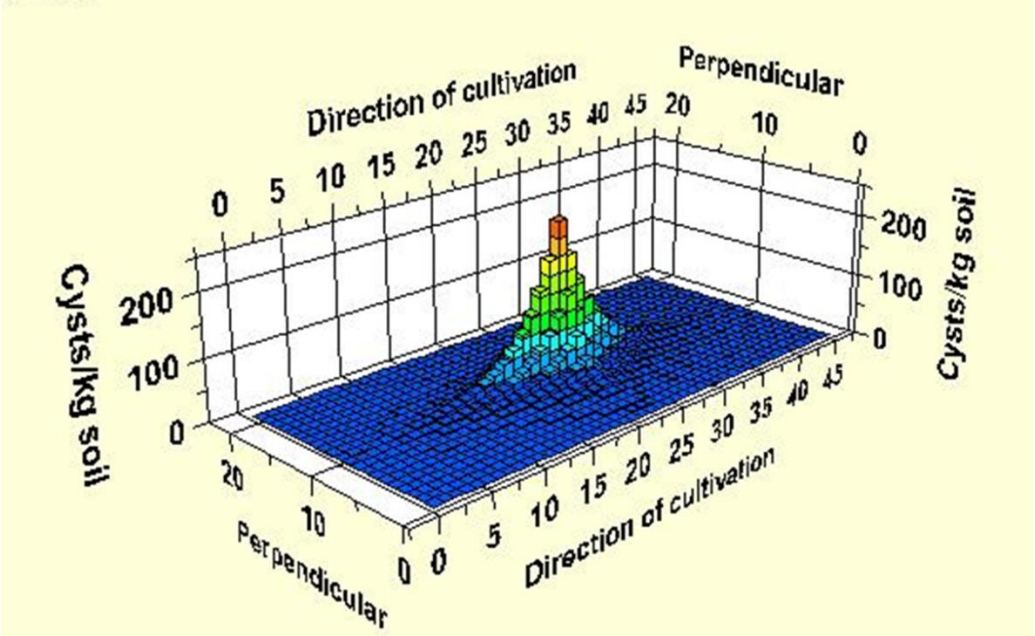



**Figure 3:** *Modelled PCN infestation in field (NemaDecide)*

Whilst the Schomaker & Been model provides a convenient basis to explore sampling options, it is not certain how well it fits typical British crop conditions, particularly for infestations in ware potato land. This is investigated further in this project.

Sampling

General background on sampling for aggregated distributions is given by Madden *et al*., (2007) and for PCN in particular by Turner (1993); Haydock & Evans (1994); Lane & Trudgill (1999) and Hockland *et al.* (2016). Sampling schemes need to take into account likely levels of aggregation of the population (Boag & Neilson, 1994); the greater the level of aggregation the greater the sampling intensity required to achieve the same chance of detection or level of precision.

The key driver for any sampling scheme is the purpose of the sampling. In the case of PCN the key options are the detection of live cysts or the estimation of

the population (Lane & Trudgill, 1999). Analysis of soil samples can give basic information to growers such as the presence or absence of cysts, or they can give much greater detail depending on how they are taken and how they are tested. At early stages of infestation, detection of cysts is of greater importance as most cysts are likely to be viable. Dead cysts (cysts containing no viable eggs) are likely to become much more prevalent once PCN has become more widespread and several crops have been grown on the infested land. Estimating the PCN population level, i.e. viable cyst contents (eggs) then becomes of greater importance. Older infestations are likely to be more dispersed by the greater extent of cultivation that has occurred since the infestation became established. Thus, distributions characterised by the Schomaker & Been model may be more representative of early infestations, or of early population development, i.e. shortly after detection becomes possible.

For the production of seed potatoes, detection is the main objective and seed production is prohibited by the EU PCN Directive on fields in which PCN has been found following the stipulated sampling and testing procedures (Pickup *et al*., 2015). Sampling for detection may also be appropriate for fields for which there is no prior history of infestation, since population levels may be negligible or very low. Potential sampling programmes may be judged by the probability that particular population levels and distributions would be detected (Turner, 1993; Been & Schomaker, 2000).

Estimation of PCN populations provides a basis for choice of management options (e.g. rotation period, choice of cultivar, nematicide treatments). Sampling schemes for quantification can be judged by the probability that the outcome will provide accurate estimates of the true population. Ideally the choice of sampling scheme should be made on the basis of the costs of testing and management options versus the benefits in terms of yield increase and implications for the future use of the land. However, in practice the quantification of all of these costs can be difficult, especially the implications of the post-harvest PCN population for the future use of the land.

Sampling schemes for estimating populations can be limited to a simple estimate of the population level across all of a field, or they can aim to map the infestation over a range of progressively finer levels of resolution. Costs generally increase as more individual soil samples require separate testing. However greater resolution does provide the possibility to identify the highest levels of infestation in the field (‘hot spots’) and provide the grower with the option of making savings by tailoring management options with greater geographical precision. However, following consultation with the Nematicide Stewardship Programme (NSP), we were informed that precision treatments of PCN ‘hotspots’ are unlikely and management decisions are generally made on a crop area basis, i.e. typically 1ha, or even larger units of 4-8 ha.

In sampling a field the following components can be varied:

- the method for soil sampling. For now it is assumed this is by taking cores and samples taken using an auger (Turner, 1993);
- the volume of soil taken with each core;

- the depth of soil sampled
- the number of core samples;
- the position of the core samples in the field: sampling pattern or walking plan (Turner, 1993; Boag & Neilson, 1994; Been & Schomaker, 2006);
- whether the primary samples are subsampled for PCN testing. Whilst subsampling would reduce precision to some degree, this may allow sampling of more soil from the field for the same cost giving an overall increase in precision if the PCN is aggregated in the field.
- when the sampling takes place. Lane & Trudgill (1999) suggested that this is best done before and after cultivation to monitor changes but after the soil has been mixed by cultivation.

The current project looks particularly at how the following components may be optimised given the purpose of the sampling:

1) Number of cores used
2) Volume of cores
3) Size of soil sample
4) Size of any subsample

Aims

The aims of this project were to investigate which PCN distribution models fit British PCN infestations and whether such distributions vary in any predictable manner, e.g. less aggregation with higher population levels. Using the information available on British PCN distributions, simulations were used to compare sampling schemes under different scenarios. Consideration was given to assessments of the most appropriate statistical models for the analysis of typical PCN distributions in fields used for ware potato production in the UK. To do this we investigated observed levels of spatial variation from field distributions, including various trial sites. We look for factors that may have an effect on the field distribution and aggregation of PCN. We assessed the effect that altering the sample size, core size and core numbers had on the likelihood of detection. We also assessed the impact that altering these variables had on the accuracy of quantification of PCN population. We have provided recommendations for the most appropriate methods for soil sampling for PCN under different scenarios and explain the reasoning behind these recommendations.

# 3. MATERIALS AND METHODS

Estimating PCN Populations: Cysts v eggs

The detection of one or more cysts in a soil sample provides information that PCN are present in the land, or at least have previously been present in the

land should the cysts found contain no viable eggs. Extraction methods are targeted at isolating cysts from soil, on the highly reasonable assumption that only the eggs within the cysts will retain the necessary viability to hatch into a 2nd-stage juvenile capable of invading a potato root. When sampling when no host crop is being grown, it is safe to assume that eggs only exist in cysts. However, many factors will determine how many cysts within a field will be viable and how many will be dead. Table 2 collates data collected by SASA from pre-crop soil tests over a period of nearly 40 years, which shows no clear relationship between the numbers of dead and live cysts.

***Table 2:*** *Pre-planting soil tests for seed potato production positive for PCN cysts 1973-2010, SASA archive data*

| | | Viable PCN Cysts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dead PCN Cysts | 0 | 1 | 2 | 3 | 4-5 | 6-10 | 11-20 | 21-50 | 51+ | Total |
| **0** | | 674 | 169 | 62 | 39 | 13 | 6 | 6 | 1 | 970 |
| **1** | 1523 | 318 | 135 | 61 | 29 | 16 | 5 | | | 2087 |
| **2** | 542 | 161 | 97 | 47 | 42 | 30 | 3 | 3 | 2 | 927 |
| **3** | 318 | 106 | 80 | 40 | 43 | 21 | 7 | 1 | | 616 |
| **4-5** | 332 | 128 | 114 | 54 | 50 | 60 | 13 | 3 | 2 | 756 |
| **6-10** | 331 | 141 | 109 | 60 | 83 | 94 | 45 | 21 | 6 | 890 |
| **11-20** | 219 | 95 | 70 | 60 | 79 | 78 | 47 | 20 | 9 | 677 |
| **21-50** | 182 | 47 | 68 | 52 | 81 | 82 | 68 | 63 | 24 | 667 |
| **51+** | 94 | 45 | 38 | 52 | 74 | 64 | 56 | 88 | 94 | 605 |
| **Total** | 3541 | 1715 | 880 | 488 | 520 | 458 | 250 | 205 | 138 | 8195 |

Table 2 indicates that even in seed land, where findings of PCN are assumed to be a consequence of more recent infestations, dead cysts have been more regularly encountered than live cysts – of the 8195 positive soil tests, 970 (12%) had live but not dead cysts, whereas 3541 (43%) had dead but not live cysts.

For ware land, continual cropping with potatoes on infested land, particularly using susceptible cultivars, will generate large numbers of new cysts. As cysts are capable of remaining in the soil for extremely long periods (over 80 years is probably not uncommon) the viability of the cysts present is highly variable. For successful management, growers need a PCN population estimate which is indicative of the potential of the nematodes to invade the roots of the planted crop. Consequently, most laboratories will provide an estimation of the number of eggs per gram of soil based on the soil sample provided.

Therefore, we conclude that for the purpose of seed production, where the purpose of sampling is to determine whether PCN are present or absent, the aim is to establish that the sample contains a cyst containing a live egg. If the sampling is for ware potato production, the aim is more likely to be the provision of a population estimate. In this case, the number of cysts present is less important than an assessment of the population of eggs that they contain. In summary, for detection, cysts with live content is the target, whereas for population estimation, estimates of eggs per gram of soil are the target.

Eggs per unit volume versus eggs per unit weight

When collecting soil from a field for a soil sample, it is easier to work by volume of soil rather than the weight of the soil. Due to the wide variation in soil types, particularly the relative proportions of mineral and organic material, and the variation in water content, there is no direct relationship between volume and weight of soil. Furthermore, storage of samples can result in compaction, particularly where samples are sat on top of each other. There are arguments to support both methods of assessing population levels. Some nematologists prefer to use volume because it seems intuitively more likely that the volume of soil surrounding the roots of a healthy potato plant will vary less than the weight of that same amount of soil. However, for the purpose of this study, in order to compare like with like we have chosen to use eggs per gram (dry weight of soil) rather than eggs per ml. For conversion between these measures we have used a factor of 1.6 kg per litre.

Sampling – Analytical methods

In sampling for PCN, the aim is to optimise the number and size of cores; how and where to place these (e.g. usual options are either a rectangular grid or to walk the field using a zig-zag or W-pattern); and what size of soil subsample to use in the laboratory for egg or cyst counts. In all considerations for sampling, the costs of any additional sampling should be balanced against benefits of greater detection or accuracy of results.

A key difference in sampling approach is identified according to whether detection of an infestation is required or whether an estimation of the level of the infestation is needed.

For early stage infestation (e.g. fields for seed production or where PCN has not previously been found) detection should be the priority. Potential sampling schemes were compared on the basis of their ability to detect early stage infestations. Practical sampling schemes should have a high probability of detecting emergent infestations. A standard scheme for comparison is the sampling protocol set out within the EU Directive, which was developed for this purpose and has been implemented across the EU for pre-crop sampling prior to seed potato production.

For later stage infestations, estimation of the population levels and therefore the potential impact such populations have on potato production is of greater importance. The estimated level of infestation can then be used for making management decisions to keep the infestation under control. Here egg counts are more relevant, as they are more strongly linked to crop damage and the development of PCN populations for future cultivation. Different sampling schemes can be compared through the precision of estimation. Precision is measured through the standard error of the estimate of egg/g. In practice, to allow easy comparison between different infestations, we divide this standard error by the eggs/g to give a coefficient of variation (CV).

Where possible, direct calculations have been used to compare different sampling schemes based on formulae developed here. If this is not feasible, computer simulation has been used. In all cases it was assumed that an area of 1ha was being sampled and that the shape of this area was a square (this is only really relevant for comparisons of different sampling patterns).

Models for PCN counts

Mathematical models were used to represent the features of PCN distributions over a field. Fitting the models to real data provided estimates of key parameters. These models were then used to compare sampling schemes.

We looked at three types of models. These differ in the degree to which they represent the spatial aspects of the PCN distributions (so how PCN is aggregated over a field) and ultimately how they may be used in practice.

***Schomaker and Been model***

This is a partly mechanistic model based on the way that infestations tend to develop from single points of introduction. Essentially foci of infestation are represented by an exponential model which occupies a diamond-shaped area of the field. The longer axis of the focus lies in the direction of predominant cultivation. The focus is defined by the central population density (CPD) in cysts/kg and parameters defining the fall-off from the centre in two directions (Schomaker and Been, 1999).

This model requires many data sets with detailed sampling to fit the key parameters. As sufficient data are not available from British fields, we rely on parameter estimates from Been and Schomaker's work, namely length (*L*) and width (*W*) parameters of 0.77 and 0.55 and a common coefficient of aggregation of 70.

This is by its nature a model suitable for early stage infestation. Once an infestation has developed beyond this stage we might expect a rather more complex array of potential distributions.

Being a fully spatial model, it is possible to compare different patterns for laying out cores in a field, as well as investigating the effect of changing numbers and sizes of cores and the amount of soil subsampled for testing. Here we do that under the assumption of a single focus in a one hectare square field. This maximises the degree of aggregation for a given overall population density and thus gives a conservative view on sampling.

Simulation is required to calculate probabilities of detection. Software has been written for this purpose in R (R Core Team, 2015). The main version works where the cores are laid out in a grid, following the description of the approach in Been & Schomaker (2000). Here the position of the focus is systematically moved around the centre of the field for one grid cell; this saves computation whilst maintaining precision. Another version of the simulation software is used

when comparing the grid layout with W-patterns. In this case the focus is laid out at random in the field without overlapping the edge.

The Been and Schomaker model is described in terms of the peak density of the focus. However it is possible to link the CPD with the mean density over a field area. For a single focus there is a simple relationship between the CPD, *P*, and mean density, *μ*:

$$P - \ln P = 1 + \mu(A.\log(L).\log(W)/4)$$

where *L* and *W* are the length and width parameters defining the gradients of the focus (Been and Schomaker, 2000) and *A* is the field area being sampled.

Figure 4 illustrates this relationship when *L* is 0.77 and *W* is 0.55.

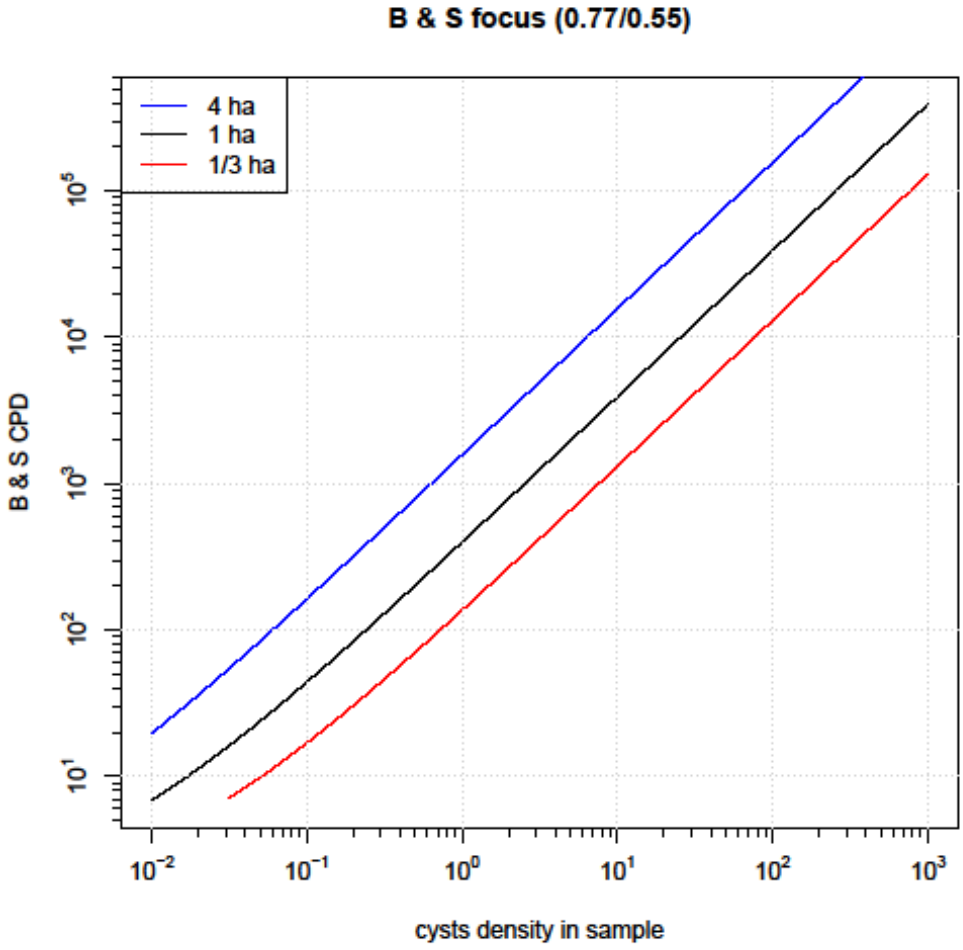



***Figure 4:*** *Relationship between the peak density of a single Been & Schomaker focus with the overall cyst density in field areas of 1/3 ha, 1 ha and 4 ha.*

### *Geostatistical models*

Geostatistical models can be used to describe the spatial distribution of PCN without the mechanistic backbone of the Been and Schomaker approach. Thus they may be more suitable for older infestations which don't fit the Been and Schomaker foci-based model. They have been used in previous studies (e.g. Evans *et al*., 2003)

These models take into account the location of the counts. They model how similar neighbouring points are dependent upon the distance they are apart. The semi-variogram is a useful tool to understand the strength and form of this relationship with distance and to potentially identify suitable models. In principle, this type of model could be used to investigate the merits of different layouts of core sampling points in a field. However there are currently too few data sets with the required level of information to substantially develop such investigations at this stage.

### *Distributional models*

Distributional models describe PCN counts without taking into account location or scale. They simply show the form of the distribution: the relationship between variance and mean, the skewness and the fatness of the tails (kurtosis).

The following distributions were considered suitable candidate models for PCN count data that originated from British fields (see Table 4):

- Poisson distribution
- Negative binomial distribution : allows for greater variability than that expected with a Poisson distribution (heterogeneity)
- Neyman type A distribution: also allows for heterogeneity, but not as widely used and much harder to work with (due to lack of software)
- Zero-inflated versions of above distributions: allows for the presence of more zero counts than the above distributions would allow for, in recognition that significant parts of fields may be free from PCN.

These distributions were fitted to the example data sets. The fits of these distributions were compared using the Akaike Information Criterion (AIC); this was not feasible for the Neyman Type A distribution due to the lack of suitable software. The absolute fit of each distribution was reviewed using quantile-quantile plots.

We have derived formulae for the coefficient of variation for the estimate of eggs/g based on these models. Here we show those for the zero-inflated negative binomial and zero-inflated Neyman type A distributions.

Zero-inflated negative binomial CV

$$\sqrt{\frac{1}{\lambda V}+\frac{p_z}{n(1-p_z)}+\frac{I_{znb}}{n(1-p_z)}}$$

Zero-inflated Neyman type A CV

$$\sqrt{\frac{1}{\lambda V}+\frac{p_z}{n(1-p_z)}+\frac{I_{NA}}{\lambda n l}}$$

$\lambda$ is the eggs/g over the field – to be estimated
$n$ is the number of cores sampled
$l$ is the size of the cores (g)
$V$ is the size of the soil subsample tested by the lab (g). Note that if there is no subsampling, this becomes *nl* (number of cores times their size)
$p_z$is the proportion of observations with zeroes that aren't due to the negative binomial or Neyman type A part of the model. Note this proportion may differ between the two models
$I_{nb}$is the heterogeneity index for the zero-inflated negative binomial distribution (parameterised so that larger values indicate greater heterogeneity and zero means no heterogeneity)
$I_{NA}$ is the heterogeneity index for the zero-inflated negative binomial distribution. Larger values indicate greater heterogeneity and zero means no heterogeneity.

These formulae differ in the last of the three elements within the square root. In both formulae the first term becomes smaller with higher egg density or greater volumes of soil subsampled for testing. Also the second term becomes smaller with more cores. It also becomes larger with an increasing zero count element of the model. For both models, the third term increases with greater heterogeneity and decreases with more cores. However this term also increases with increasing $p_z$ for the negative binomial, whereas for the Neyman type A model it decreases as the expected number of eggs in the soil sample increases (this depends on the egg density, number of cores and the size of the cores).

The effect of these different factors on the CV is evaluated more fully later based on parameters estimated from the data sets. Approximate confidence intervals for the eggs/g estimate can be produced straightforwardly e.g. a 95% confidence interval is approximately $\hat{\lambda}(1 \pm 1.96\mathrm{CV})$.

Whilst the above rationale results in relatively straightforward formulae, it is based on some assumptions. One key assumption relates to subsampling. With subsampling it is assumed both that a representative subsample is obtained and that eggs are uniformly distributed throughout the drawn sample. At this time, there is little information available to inform best subsampling practice. However it is clear that eggs will not be uniformly distributed throughout the sample as they are aggregated within cysts. When subsampling, the level of this aggregation will affect the precision of the estimate of egg density.

Building this aggregation into the CV calculations requires knowledge of the number of eggs per cysts and also how this varies across the field. Useful information on the number of eggs per cyst is limited and it may not be practical to estimate this.

From pre-crop soil tests, which are usually tested 6 years since the previous crop (although cysts may be younger than 6 years old if groundkeepers have

persisted in the field), the range of eggs per cyst in viable cysts can vary from 1 to over 300 and there is no predictive relationship between numbers of eggs and numbers of cysts. Table 2 shows a highly variable ratio of live to dead cysts, Also, from a study of the impact of the cultivation of PCN-resistant cultivars on PCN populations, we found that the ratio of live to dead cysts, and hence the mean eggs per cyst, is highly influenced by the cultivar used for planting (Table 3). Cultivation of a highly resistant variety had a much more marked effect of depleting the PCN population in terms of eggs/g than it did in reducing the number of cysts. Cultivation of a variety with a lower level of resistance increased both cyst and egg counts. Consequently the mean values for eggs/cyst were much higher in the samples from where the less resistant variety had been grown. Viability of cysts can vary considerably, e.g. a population of 10 eggs/ml could consist of 1000 cysts per litre, each with 10 eggs per cyst, or could consist of a similar number of cysts per litre, but with all the eggs present in just 100 cysts (at 100 eggs/cyst) and the remaining 900 cysts containing no eggs.

***Table 3:*** *Eggs per cyst pre-planting and post-harvest in a* G. pallida *infested field where cultivars with different* G. pallida *resistance scores were planted. Results from analyses of 400 ml soil samples. Part 1-14 = different parts of the field.*

| | | Pre-planting | | | Post-harvest | | |
|---|---|---|---|---|---|---|---|
| | | Cyst count | Eggs/g | Mean eggs/cyst | Cyst count | Eggs/g | Mean eggs/cyst |
| Cultivar with resistance to *G.pallida* Score 9 | Part 1 | 1 | 0.03 | 18.0 | 3 | 0.01 | 1.0 |
| | Part 2 | 8 | 1.29 | 112.3 | 3 | 0.36 | 50.3 |
| | Part 3 | 16 | 2.92 | 126.5 | 28 | 0.32 | 5.0 |
| | Part 4 | 23 | 4.13 | 123.9 | 8 | 0.11 | 5.9 |
| | Part 5 | 11 | 1.46 | 91.2 | 13 | 0.14 | 4.8 |
| | Part 6 | 24 | 3.70 | 102.6 | 17 | 0.22 | 4.9 |
| | Part 7 | 61 | 11.76 | 124.9 | 35 | 0.31 | 3.7 |
| Cultivar with resistance to *G.pallida* Score 3 | Part 8 | 61 | 11.76 | 124.9 | 119 | 35.17 | 120.6 |
| | Part 9 | 3 | 0.06 | 12.0 | 29 | 6.20 | 103.4 |
| | Part 10 | 2 | 0.20 | 63.5 | 5 | 1.18 | 110.4 |
| | Part 11 | 0 | 0.00 | 0.0 | 0 | 0.00 | 0.0 |
| | Part 12 | 5 | 0.02 | 2.2 | 21 | 11.11 | 245.2 |
| | Part 13 | 1 | 0.09 | 56.0 | 1 | 0.39 | 175.0 |
| | Part 14 | 25 | 5.87 | 145.3 | 44 | 12.00 | 111.2 |

In the absence of available information with which to predict the number eggs per cyst likely to be present within a field, we have focussed on exploring the extent to which this aggregation affects precision of our estimates, looking at 10, 50 and 200 eggs per cyst. Formulae to express typical distributions of eggs/cyst have not yet been developed; this might be possible but might involve approximation. Instead computer simulation has been used on the following basis:

- The number of eggs in the sample (r) are randomly generated using a zero-inflated negative binomial distribution.

- Given the number of eggs per cysts assumed (1, 10 or 50), compute the effective number of cysts in the sample = n (by rounding the number of eggs divided by the number of eggs per cyst).
- work out the actual eggs per cyst r/n.
- simulate a random subsample by generating a binomial random sample based on n and the proportion of soil subsampled - this gives the number of cysts in the subsample c.
- calculate the number of eggs in the subsample =round(c * r/n).

Note this assumes that the number of eggs per cyst is constant from cyst to cyst. While this is unlikely, we do not currently have information to support an alternative model.

***Summary***

The Been and Schomaker model was used to simulate early stage infestations, based on cysts/kg. The distributional models were used to study sampling under later stage infestations, based on eggs/g and after fitting to example data sets. Work here with geostatistical models was limited to reviewing semi-variograms for suitable example data sets.

Datasets

The data sets of British fields available for this study are summarised in Table 4. All have egg counts. Note that the eggs counts for Dataset 5 were inferred from PCR measurements of DNA.

***Table 4:*** *Data sets from British fields available for this study*

| *Source* | *Field* | *Size* | *Units* | *Unit size* | *Distance between units (m)* | *Cores per unit* | *Soil sample* | *Subsample tested* | *Pf/Pi* | *Date* | *Potential for spatial modelling* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset 1 | 10 fields | 2 to 22 ha | 1 to 202 | 6m strips (0.1 to 0.3 ha) | 6 | Approx. 200 | 1kg | 1000g | Pi | Feb/March 2014 | Yes |
| Dataset 2 | "Trial 2" | 5.7 ha | 6 | ~ 1 ha strips | ~ 25 | 40 | 1-1.5kg | 100g | Pi | Early Summer | No |
| | | | 5 | ~ 1 ha squares | 100 | 50 | 1-1.5kg | 100g | Pi | Sep 2011 | No |
| | | | 21 | ~ ¼ ha squares | 50 | 50 | 1-1.5kg | 100g | Pi | Aug 2011 | No |
| | "Trial 3" | ~20 ha | 20 | ~ 1 ha squares | 100 | 50 | 1-1.5kg | 100g | Pi | May 2002 | No |
| | | | 459 | Point sampling | ~19 | 1 | 200g | 200g | Pi | March | Yes |
| Dataset 3 | Whole | ~ 14 ha | 4 | 3.5 ha strips | 50 | 50 | Unknown | 100g | Pf | Oct 2011 | No |
| | | | 14 | 1 ha squares | 100 | 50 | Unknown | 100g | Pi/Pf | March 2014 & 2015 | No |
| | Detail | 480 sq m | 48 | 10 sq m | 3 | 50 | Unknown | 100g | Pi | May 2014 | Yes |
| Dataset 4 | Whole | 13.4 ha | 13 | 1 ha squares | 100 | 50 | 2kg | 100g | Pi | May 2013 | No |
| | | | 55 | ¼ ha squares | 50 | 50 | 2kg | 100g | Pi | Aug 2013 | Yes |
| | Detail | ¼ ha | 9 | 17 x 17 m squares | 17 | 50 | 2kg | 100g | Pi | Oct 2013 | No |
| Dataset 5 | Wide | ¼ ha | 44 | ¼ ha squares | 50 | 25 | 400g | 400g | Pi | March 2014 | Yes |

# 4. RESULTS

Densities, heterogeneity and fit of distributions in example data sets

The comparative fit, represented by the delta AIC, of different distribution models to the data sets is presented in Table 5: the smaller the value the better the fit. The best fitting model for each dataset has a value of 0 – any model having a delta AIC value within five or six points can be considered to provide a similar fit. Note that at this time no software was available for computing fits for the Neyman type A.

***Table 5:*** *Comparative fits for different distributions using Akaike Information Criterion (AIC)*

| | | | Delta AIC | | | |
|---|---|---|---|---|---|---|
| Source | Field/sampling | Units | Poisson | Negative binomial | ZI Poisson | ZI Negative binomial |
| Dataset 1 | A | 66 | 5478.4 | 30.3 | 2938.2 | 0 |
| Dataset 1 | B | 51 | 5506.5 | 22.8 | 2417.4 | 0 |
| Dataset 1 | C | 202 | 7418.4 | 13.2 | 1185.6 | 0 |
| Dataset 1 | D | 86 | 5497.6 | 11.1 | 2274.8 | 0 |
| Dataset 1 | E | 144 | 35832.4 | 17.8 | 22085.7 | 0 |
| Dataset 1 | F | 24 | 52309.9 | 0 | 48074.0 | 0.7 |
| Dataset 1 | G | 14 | 7504.8 | 2.2 | 5984.8 | 0 |
| Dataset 1 | H | 13 | 173653.3 | 0.6 | 132146.4 | 0 |
| Dataset 1 | I | 10 | 9248.0 | 0 | No zeros | |
| Dataset 1 | J | 21 | 147473.9 | 0.2 | 109393.5 | 0 |
| Dataset 2 | Trial 2 blocks 1/4ha | 21 | 7893.6 | 8.8 | 5797.1 | 0 |
| Dataset 2 | Trial 2 blocks 1ha | 5 | 1102.6 | 1.7 | 551.8 | 0 |
| Dataset 2 | Trial 2 strips | 6 | 2856.8 | 0 | No zeros | |
| Dataset 2 | Trial 3 blocks 1ha | 20 | 4927.5 | 5.4 | 2657.7 | 0 |
| Dataset 2 | Trial 3 points | 472 | 227083.3 | 166.2 | 94228.1 | 0 |
| Dataset 3 | 3.5 ha strips | 4 | 6920.9 | 0 | No zeros | |
| Dataset 3 | blocks 1ha - Pf | 14 | 8590.4 | 8.7 | 5037.8 | 0 |
| Dataset 3 | blocks 1ha - Pi | 14 | 30217.0 | 4.0 | 22430.5 | 0 |
| Dataset 3 | detail 10m2 - Pi | 48 | 22217.9 | 0 | No zero | |
| Dataset 4 | Trial block 1/4 ha | 53 | 58617.2 | 31.2 | 25456.6 | 0 |
| Dataset 4 | Trial block 1ha | 13 | 12493.2 | 3.8 | 6774.8 | 0 |
| Dataset 4 | Trial detail | 9 | 788.4 | 0 | No zeros | |
| Dataset 5 | Full Trial | 44 | 1113166.9 | 0 | No zeros | |

It is clear that the zero-inflated negative binomial model provided the best overall fit, although there were data sets where there were no zeroes and the occasional case where the fit of the negative binomial was a little better. However neither the Poisson nor the zero-inflated Poisson fitted nearly as well as the negative binomial distribution. This was due to the substantial heterogeneity in counts.

There was no comparison with the (zero-inflated) Neyman type A. Software was not easily available, so progress on this would have involved substantial work. Later we evaluate whether this omission was critical.

The absolute fit of the zero-inflated negative binomial model (or negative binomial where there are no zeroes) was evaluated using quantile-quantile plots. For most data sets the fit seemed adequate taking into account the numbers of observations and ignoring the occasional outlier.

One exception was the point sampling experiment at Dataset 2 Trial 3. The Q-Q plot is shown in Figure 5. In this case the high values were greater than expected, implying perhaps bimodality or a longer tail than expected under the model. Figure 6 shows the spatial distribution of counts for this case. Whilst it seems there was one main focus and the mean egg density seems quite low at 0.8 eggs/g, that focus was much larger compared to that expected with a Been and Schomaker focus.

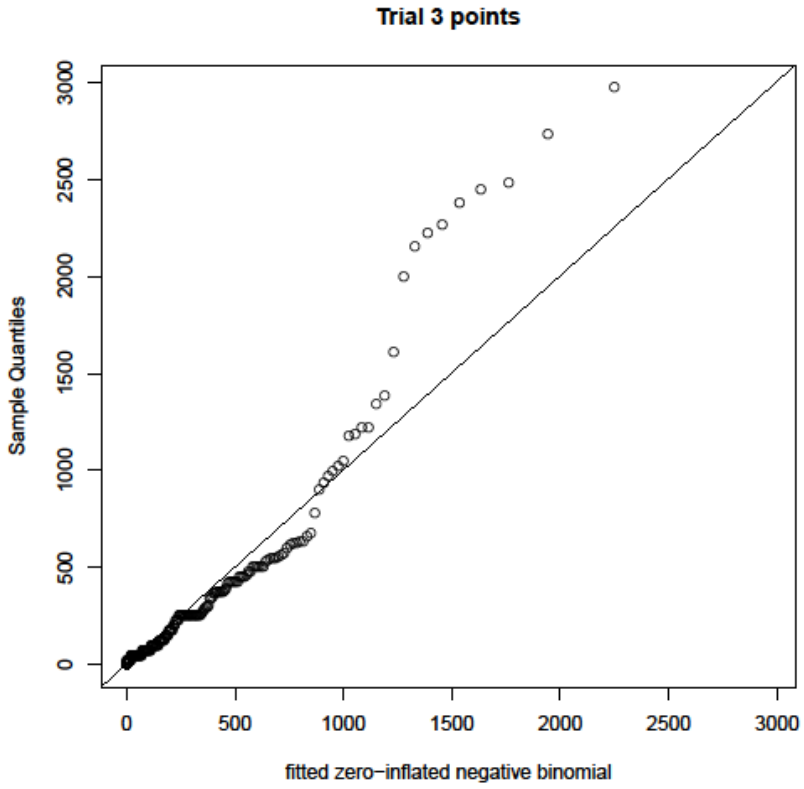



***Figure 5:*** *Quantile-quantile plot for fit to a zero-inflated negative binomial distribution for Dataset 2 Trial 3 Point Sampling. Compares actual egg counts versus expected under given distribution (from Kerry, Barker & Evans, 2003)*

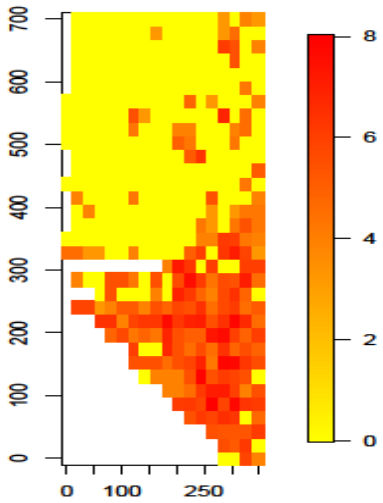



***Figure 6:*** *Dataset 2 Trial 3 Point Sampling. Scale is eggs/g*

Table 6 shows parameter estimates for the negative binomial and zero-inflated negative binomial for each case. In this table, $p_z$ is the proportion of observations with zeroes that aren't due to the negative binomial. $I_{ND}$ is the dispersion index, i.e. the variance of counts over the mean count. Larger values indicate greater heterogeneity and unity means no heterogeneity. $I_{nb}$ is the heterogeneity index for the zero-inflated negative binomial distribution (larger values indicate greater heterogeneity and zero means no heterogeneity). $I_{znb}$ is the heterogeneity index for the zero-inflated negative binomial distribution. The better fitting model is highlighted in bold. It can be seen that inclusion of zero-inflation reduces the heterogeneity index considerably. However the degree of zero-inflation varies considerably between data sets.

Figure 7 shows how the heterogeneity index for the negative binomial distribution relates to the egg density. A spline fitted trend is included. It seems that heterogeneity tends to decline with high rates of infestation, although further datasets would be required to validate this. Figure 8 shows how the level of zero inflation and heterogeneity for the zero-inflated negative binomial distribution relates to egg density. Both the heterogeneity index $I_{znb}$ and the zero proportion $p_z$ seem to decline with egg density. There was not a clear relationship with the scale (the distance between units) but this is likely to be in part due to insufficient information.

***Table 6:*** *Levels of infestation, zero-inflation and heterogeneity*

| Source | trial | units | distance between units (m) | Eggs/g | $p_z$ | $I_D$ | $I_{nb}$ | $I_{znb}$ |
|---|---|---|---|---|---|---|---|---|
| Dataset 1 | A | 66 | 6 | 0.07 | 0.24 | 106.3 | 2.7 | **0.7** |
| Dataset 1 | B | 51 | 6 | 0.09 | 0.29 | 107.7 | 3.7 | **0.8** |
| Dataset 1 | C | 202 | 6 | 0.01 | 0.88 | 110.7 | 46.1 | **1.2** |
| Dataset 1 | D | 86 | 6 | 0.03 | 0.49 | 109.4 | 7.3 | **1.4** |
| Dataset 1 | E | 144 | 6 | 0.13 | 0.29 | 330.2 | 4.9 | **1.8** |
| Dataset 1 | F | 24 | 6 | 1.02 | 0.00 | 4210.3 | **2.7** | 1.9 |
| Dataset 1 | G | 14 | 6 | 0.36 | 0.14 | 869.2 | 2.6 | **1.1** |
| Dataset 1 | H | 13 | 6 | 4.34 | 0.29 | 28128.0 | 7.9 | **2.6** |
| Dataset 1 | I | 10 | 6 | 2.83 | 0.00 | 915.3 | **0.5** | |
| Dataset 1 | J | 21 | 6 | 2.70 | 0.25 | 10453.9 | 7.1 | **3.1** |
| Dataset 2 | Trial 2 blocks 1/4ha | 21 | 50 | 5.02 | 0.10 | 383.4 | 1.6 | **0.7** |
| Dataset 2 | Trial 2 blocks 1ha | 5 | 100 | 2.50 | 0.20 | 230.0 | 3.2 | **0.6** |
| Dataset 2 | Trial 2 strips | 6 | ~25 | 11.67 | 0.00 | 684.6 | **0.4** | |
| Dataset 2 | Trial 3 blocks 1ha | 20 | 100 | 1.13 | 0.40 | 390.4 | 6.2 | **1.1** |
| Dataset 2 | Trial 3 points | 472 | ~19 | 0.82 | 0.58 | 475.6 | 12.3 | **1.2** |
| Dataset 3 | Trial block 1/4 Ha | 53 | 50 | 5.51 | 0.43 | 1415.3 | 8.3 | **0.9** |
| Dataset 3 | Trial block 1Ha | 13 | 100 | 6.00 | 0.31 | 1063.2 | 5.7 | **1.3** |
| Dataset 3 | Trial detail | 9 | 17 | 20.67 | 0.00 | 493.5 | **0.2** | |
| Dataset 4 | NW 3.5 ha strips | 4 | 50 | 44.50 | 0.00 | 2517.6 | **0.5** | |
| Dataset 4 | NW blocks 1ha - Pf | 14 | 100 | 5.29 | 0.21 | 720.2 | 3.4 | **0.7** |
| Dataset 4 | NW blocks 1ha - Pi | 14 | 100 | 18.07 | 0.14 | 2256.4 | 3.1 | **1.1** |
| Dataset 4 | NW detail 10m2 - Pi | 48 | ~3 | 41.04 | 0.00 | 501.4 | **0.1** | |
| Dataset 5 | Trial | 44 | 50 | 84.79 | 0.00 | 39004.4 | **0.6** | |

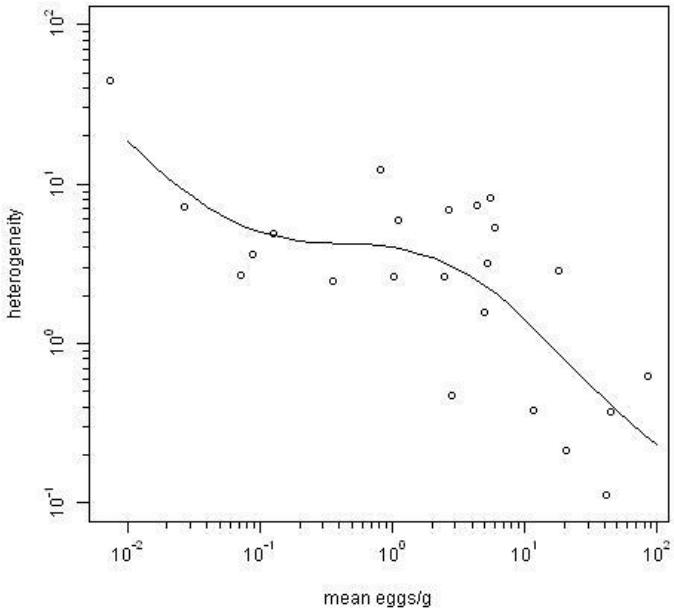



***Figure 7:*** *heterogeneity index, $I_{nb}$, for the negative binomial distribution against egg density*

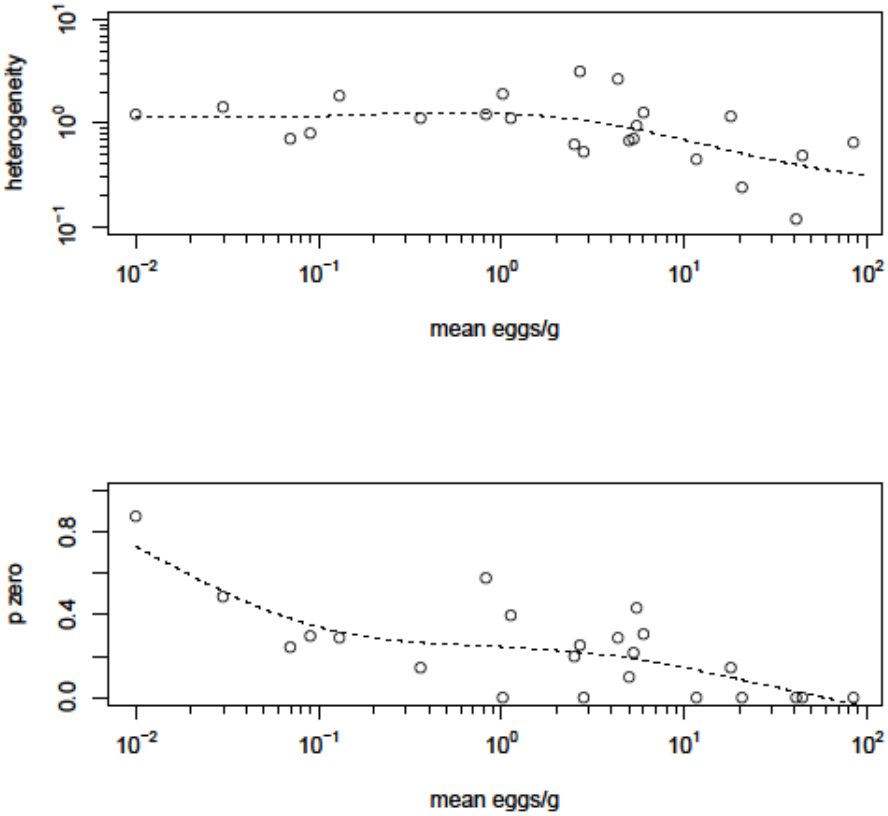



***Figure 8:*** *zero-inflation proportion and heterogeneity index, $I_{znb}$, for the zero-inflated negative binomial distribution against egg density*

Spatial analysis

The spatial aspect of heterogeneity was reviewed for example data sets using semi-variograms applied to the log of the eggs counts. Most of the example data sets are too small (in terms of sampling points) to define the nature of the spatial autocorrelation apart from identifying whether it exists or it's negligible. In some cases it is possible to do this separately according to the direction. Table 7 summarises the results for those fields with sufficient data points (>24).

***Table 7:*** *Data sources for spatial analysis*

| Source | Trial | No samples | Eggs/g | Dimensionality | Variogram |
|---|---|---|---|---|---|
| Dataset 1 | A | 66 | 0.07 | 1D | Negligible spatial effect |
| Dataset 1 | B | 51 | 0.09 | 1D | Spatial effect |
| Dataset 1 | C | 202 | 0.01 | 1D | Spatial effect |
| Dataset 1 | D | 86 | 0.03 | 1D | Slight spatial effect |
| Dataset 1 | E | 144 | 0.13 | 1D | Spatial effect |
| Dataset 2 | Trial 3 points detail 10m2 - Pi | 472 | 0.82 | 2D | Spatial effect predominantly in one direction |
| Dataset 3 | | 48 | 41.04 | 2D | Spatial effect predominantly in one direction |
| Dataset 4 | block 1/4 Ha | 53 | 5.51 | 2D | Negligible spatial effect |
| Dataset 5 | ¼ ha | 44 | 84.79 | 2D | Negligible spatial effect |

The degree of spatial effect reflects the degree to which nearby samples have similar egg counts (spatial aggregation). Note that it is possible to have heterogeneity in the counts (greater variability than expected) without this being due to spatial aggregation. There were marked differences between these data sets in the degree to which spatial effects were seen. To some extent this may be related to the degree of heterogeneity. This, combined with an understanding of the variety of ways in which an infestation could develop, **leads to the conclusion that there wouldn't be a single spatial model that could describe the spatial heterogeneity present in British fields with established infestations.** In addition, we believe that the use of distributional models should be sufficient to answer questions on sampling strategy for such fields, other than information about how cores should be located (e.g. grid or W pattern). So this would mean that there is little point in looking at further data sets with this level of detail for the purpose of improving sampling processes.

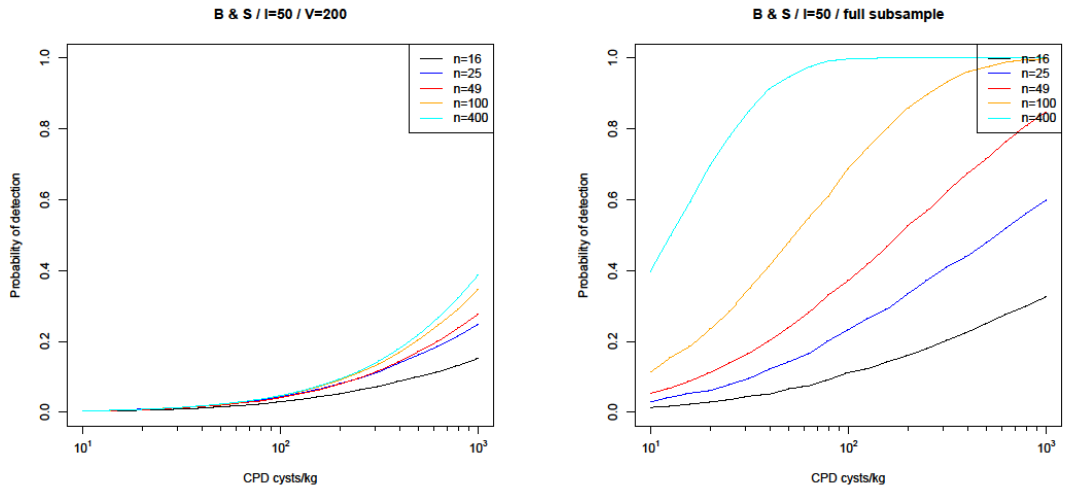
Sampling for early stage infestations

Table 8 shows the results from simulations based on the Been and Schomaker model with a single focus in a 1ha square field area. Comparisons of numbers of cores, core size and subsample size are based on a square grid of sampling points. This clearly demonstrates that for populations of 500,000 or fewer cysts per hectare, the chances of detection are very low. In this situation, the likelihood of detecting a cyst using 50 cores and testing 200mls of soil is approximately 3%, this rises to 21% using the EU higher sampling rate. It is only once the level of cysts per hectare reaches 20 million that there is a 95% chance of detection at the highest EU rate, but even at this level there is only a 41% chance of detection with 50 cores and a 200ml sample. NB The EU sampling rate was originally calculated based on the detection of 4 foci per ha (three at a CPD of 50 cysts/kg, and one at a CPD of 100 cysts/kg).

***Table 8:*** *Detection probability using Been and Schomaker model with one focus. The sampling protocol described by DuPont (2016) has been used as an example of current industry practice. It comprises: taking 50 cores of 1 cm diameter per hectare to a depth of 10-15 cm to provide a maximum sample size of 1 kg, from which a 200 ml subsample should be processed. The EU higher rate is 100 cores and 1500mls of soil while the EU lower rate is 100 cores with 400ml of soil.*

| Cysts/ha | Cysts/litre | CPD Cysts/kg | | DuPont | Eu Higher | Eu lower |
|---|---|---|---|---|---|---|
| | | | No of cores | 50 | 100 | 100 |
| | | | Core size (ml) | 11.78 | 15 | 4 |
| | | | Volume tested (ml) | 200 | 1500 | 400 |
| 500,000 | 0.25 | 66.2 | | 0.03 | 0.26 | 0.08 |
| 1,000,000 | 0.5 | 127.9 | | 0.05 | 0.41 | 0.14 |
| 2,000,000 | 1 | 250.7 | | 0.09 | 0.58 | 0.23 |
| 5,000,000 | 2.5 | 617.8 | | 0.18 | 0.81 | 0.40 |
| 10,000,000 | 5 | 1228.8 | | 0.28 | 0.92 | 0.55 |
| 20,000,000 | 10 | 2450.3 | | 0.4 | 0.98 | 0.70 |
| 50,000,000 | 25 | 6113.4 | | 0.58 | 1.00 | 0.86 |
| 100,000,000 | 50 | 12217.7 | | 0.716 | 1.00 | 0.93 |

Figure 9 shows the effect of varying the number of 50 ml cores on the probability of detection in relation to CPD, using different subsample sizes. If a subsample of 200 g is used for laboratory analysis (Figure 9a), the effect of changing the number of cores is only moderate. However if all the soil is tested the probability of detection is substantially better (Figure 9b) and the benefit of increasing the number of cores is much larger as more soil will be tested. Note this may be impractical; 400 cores of 50 ml produces 20 litres of soil.




(a) Subsample of 200 g (b) Subsample using all soil

***Figure 9:*** *Effect of number of cores on the probability of detection with core size of 50ml*

Figure 10 shows the effect of changing the number of cores whilst maintaining the same sample volume (V = 1500ml). The cores are taken on a square grid pattern of 5 x 5, 7 x 7, 10 x 10, 14 x 14 and 20 x 20 cores, with the core size adjusted accordingly to maintain a 1500 ml sample volume. Increasing the number of cores taken has clear benefits in terms of increasing the probability of detection, although beyond 100 cores/ha the resulting improvement is relatively low.

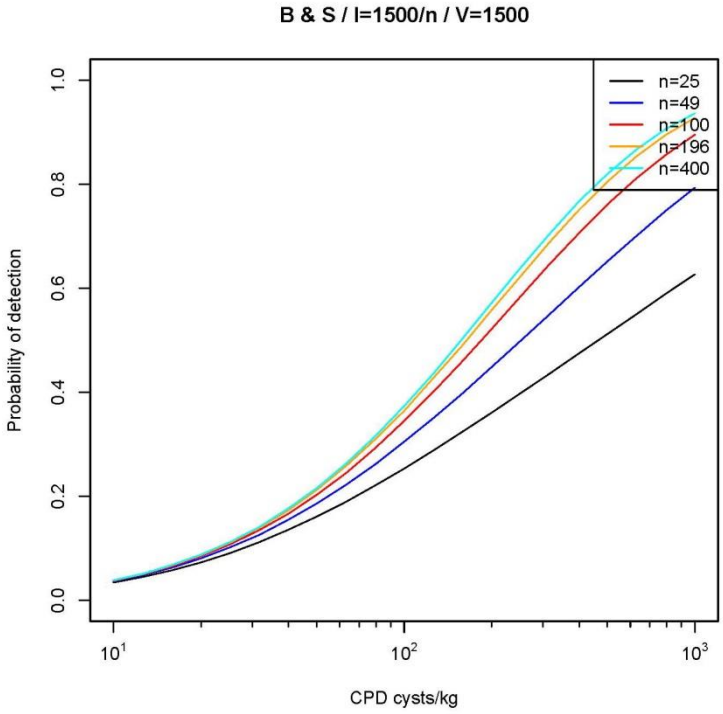



***Figure 10:*** *Effect of number of cores on the probability of detection with a fixed sample volume of 1500 ml*

Figure 11 shows the effect of the core size (4 ml to 100 ml) on the probability of detection in relation to CPD, again using a subsample of 200 g for the laboratory analysis. The core size seems fairly unimportant if subsampling for testing in the laboratory is occurring.

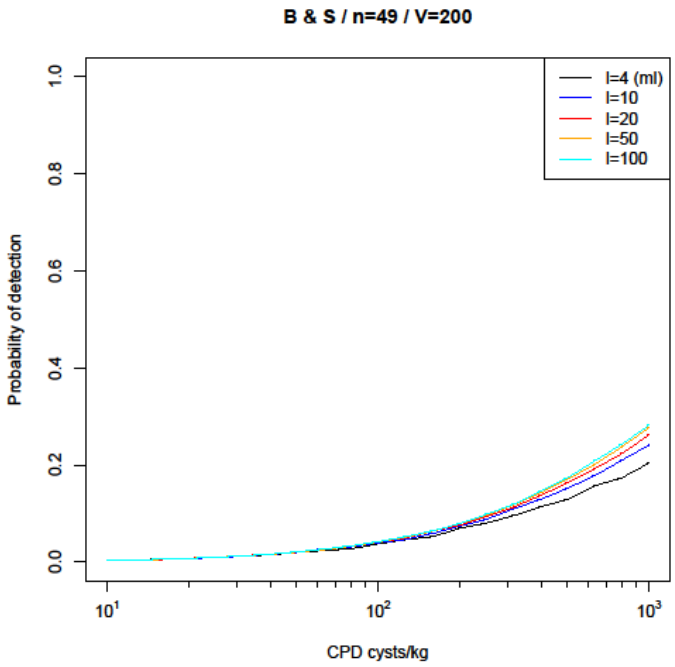



***Figure 11:*** *Effect of core size on the probability of detection with 49 cores and a subsample of 200g*

The effect of increasing the amount of soil subsampled for laboratory testing is shown in Figure 12 (100 g to 2000 g). It is clear that it is highly beneficial to test more soil.

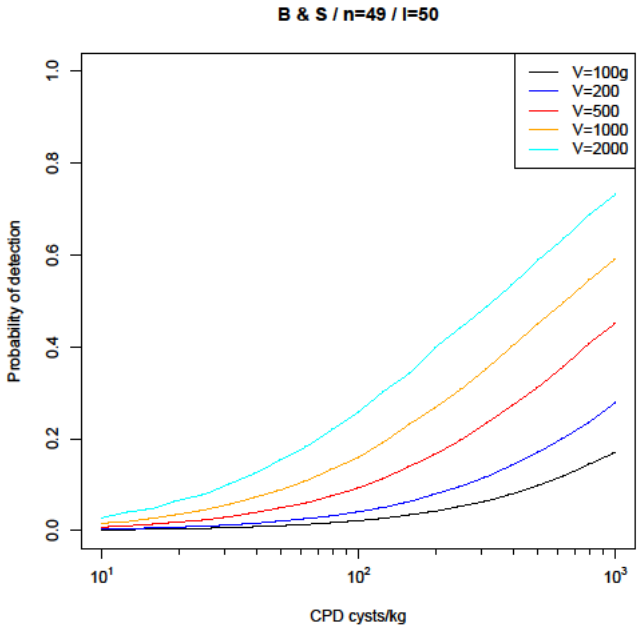
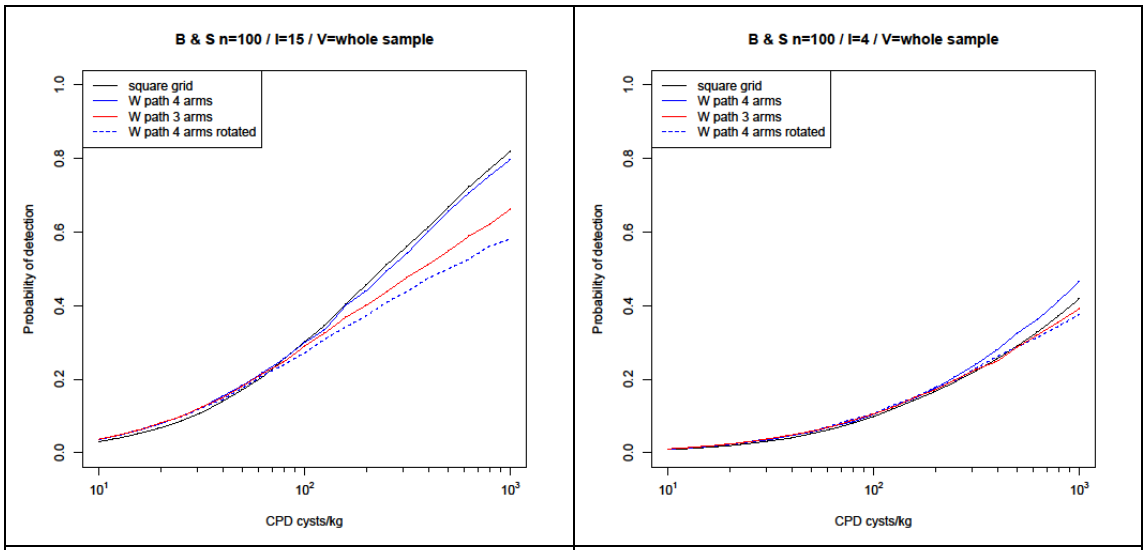



**Figure 12:** *Effect of soil subsample size on the probability of detection with 49 cores and cores of 50 ml*

**So in summary, to increase probability of detection the key factor is the amount of soil tested at the laboratory. It is better to increase the soil available for testing through increasing the number of cores rather than the size of cores.**

Next the effect of the layout of core sampling positions is evaluated (Figure 13). The simulation is based on 100 cores of 15 ml (EU higher rate for seed) and 100 cores of 4 ml (EU lower rate for seed). Soil subsampling is not used. A square grid pattern for the 100 cores is compared to a W-pattern with 4 arms and a W-pattern with 3 arms. For completeness the 4-arm W-path is included at right angles to original.




(a) EU higher rate | (b) EU lower rate

***Figure 13:*** *Effect of core sampling pattern on the probability of detection with 100 cores*

When 15 ml cores are used, the grid gives the highest chance of detection above population densities of 100 cysts/kg but the benefit over the 4-arm W-pattern is minimal. However the 3-arm W is slightly better at low densities. With 4 ml cores, the 4-arm W-pattern has a very small advantage at higher infestation levels. However care is needed here for two reasons. Firstly and more importantly, with the W-pattern it is critical that the orientation is correct: the rotated version performs considerably worse especially with 15 ml cores at higher infestation levels. Secondly results may be slightly different depending on the details of how the paths are laid out on the field and how this relates to the simulation algorithm; many variations are possible.

These results are not entirely intuitive, so we will explore this further. Figure 14 shows a Been and Schomaker focus in a 1 ha field. It is very small and has a distinct orientation. It is helpful to think in terms of effective coverage of path. This is the proportion of the field where a focus can be detected. This depends on the size and orientation of the focus, core size, etc. plus any threshold of detection we might apply. This can be done by placing foci in every single point in the field, then working out the probability of detection. Figure 15 shows an example with different sampling patterns. The rotated W-pattern loses out because the orientation of the arms coincides with the length of the focus.

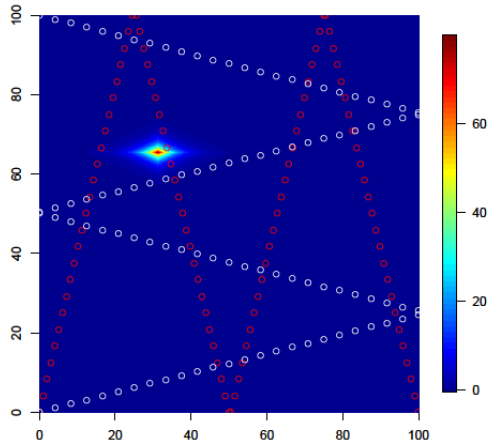
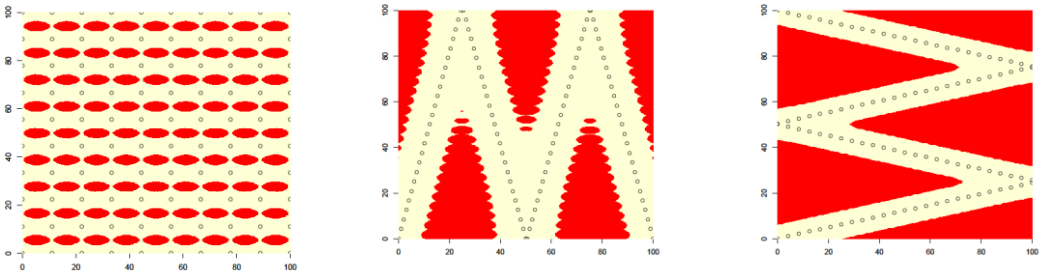



***Figure 14:*** *Example of Been and Schomaker focus with W-patterns*



(a) Square grid (b) 4 arm W (c) Rotated 4 arm W

***Figure 15:*** *Coverage of different walking patterns with 100 cores. The red colour indicates areas where foci cannot be detected and the circles mark the sampling points*

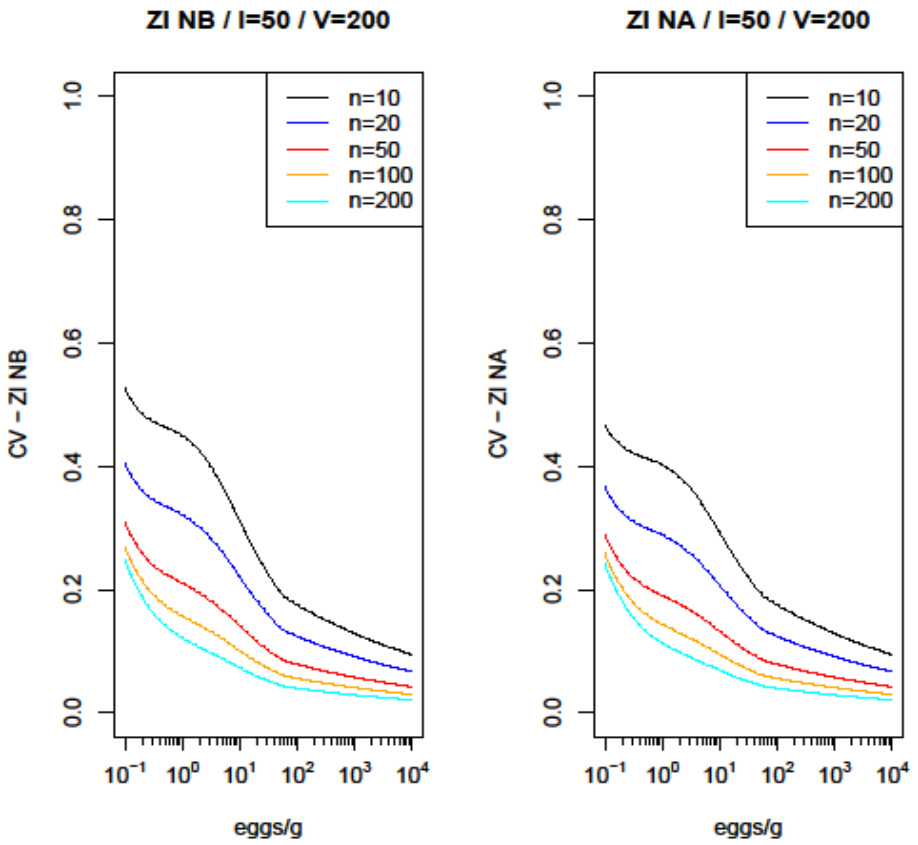
Although the performance of the grid is mixed compared to the W-pattern in these simulation examples, it should be better at picking up larger infestations when sufficient soil is sampled. Importantly it is also much more robust than the W-pattern as it is less dependent on orientation. Also note that these simulations were carried out on a square field.

Sampling for later stage infestations

The levels of zero-inflation and heterogeneity for the negative binomial distribution were set according the egg density using the spline curves fitted in Figure 8. At the time of writing, no software was available for fitting the zero-inflated Neyman type A distribution, so for the calculations below, we used the same level of zero-inflation and set $I_{zNA}$=$l\lambda I_{znb}$ to give an equivalent level of heterogeneity. Figures 16, 17 and 18 show how the coefficient of variation (CV) depends on the egg density, the number of cores and their volume and the subsample of soil tested for both zero-inflated distributions.

It was clear that the number of cores is very important (Figure 16), but the core size is almost immaterial (above 4 ml) (Figure 17 – note the lines in both figures are co-incident – the line for the 100 ml core size obscures the others). Assuming one egg per cyst, soil subsampling has little effect when the egg density is above 2 eggs/g (Figure 18). At 1 egg/g there is a small and diminishing benefit in evaluating a greater volume of soil.

Therefore, if there were an equal distribution of eggs and cysts, i.e. one egg per cyst, a laboratory subsample of 100 g of soil would be sufficient for population estimation, providing that this subsample is representative of the whole sample, i.e. thoroughly mixed and that the egg density is thought to be above 1 egg/g. It is not possible to infer from these distribution models which sampling pattern works best; the models ignore spatial aspects of the distribution of PCN. However sampling principles would imply that a square grid would be the most prudent choice, in the absence of knowledge about the potential pattern of PCN in the field.




(a) Zero-inflated negative binomial

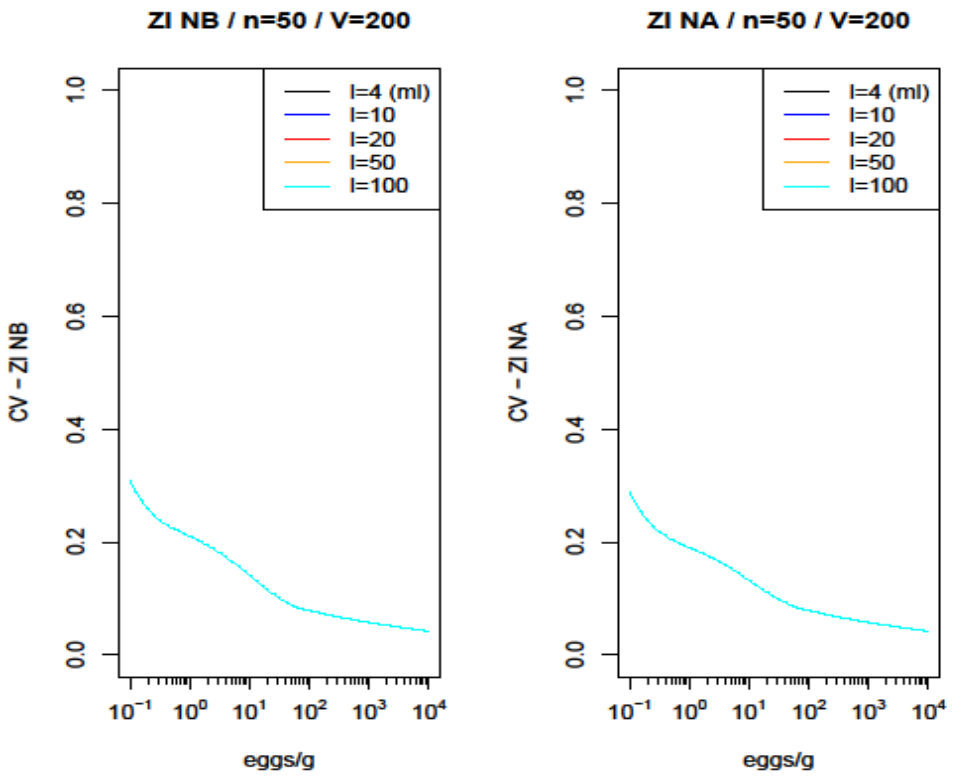
(b) Zero-inflated Neyman type A

***Figure 16:*** *Effect of number of cores on the coefficient of variation for the egg density with core size of 50 ml and a subsample of 200g, assuming 1 egg/cyst. Two PCN distributions are compared: the zero-inflated negative binomial and the zero-inflated Neyman type A*

***Table 9:*** *Zero inflated negative binomial distribution: approximate 95% confidence intervals over a range of population levels for four reference soil sampling procedures based on one egg per cyst*

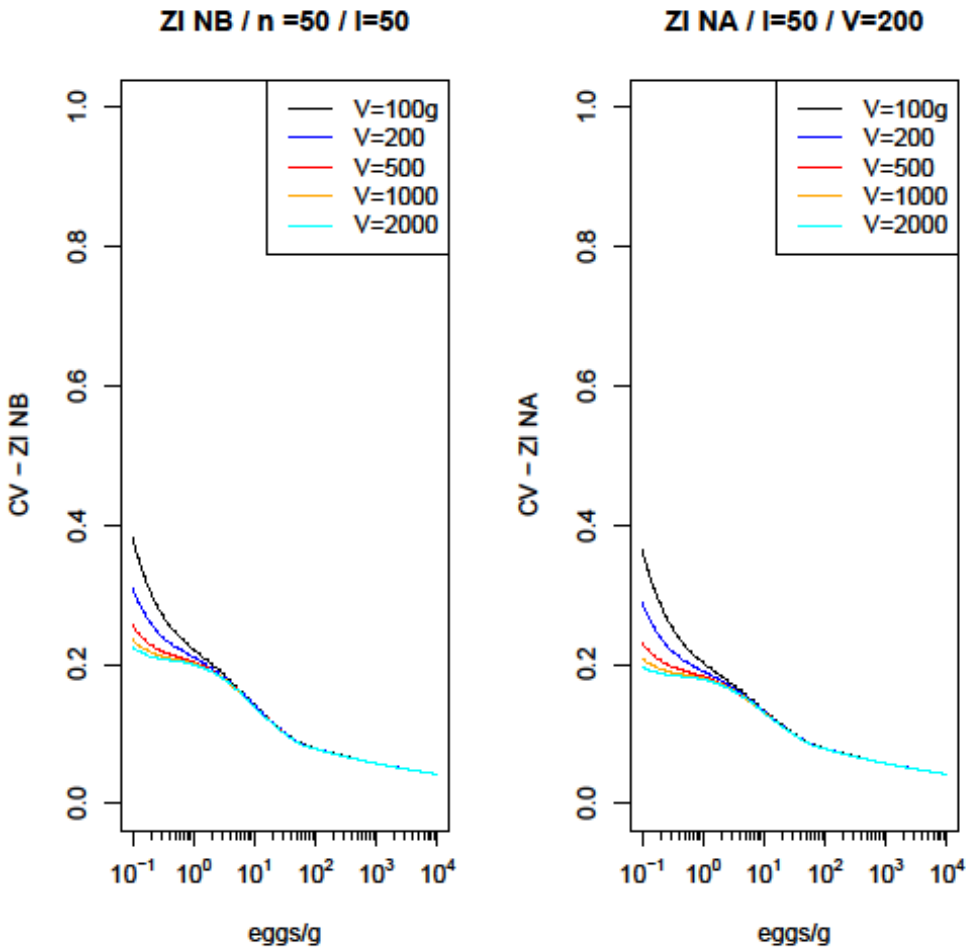
| Eggs/ha | Eggs/g | | DuPont | | DuPont | | Eu Higher | | Eu Lower | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | No cores | 50 | | 50 | | 100 | | 100 | |
| | | Core size (ml) | 11.78 | | 11.78 | | 15 | | 4 | |
| | | Weight tested (g) | 200 | | 100 | | 1500 | | 400 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^{9}$ | 0.1 | | 0.039 | 0.161 | 0.025 | 0.175 | 0.066 | 0.134 | 0.057 | 0.143 |
| $3 \times 10^{9}$ | 0.2 | | 0.098 | 0.302 | 0.081 | 0.319 | 0.139 | 0.261 | 0.128 | 0.272 |
| $6 \times 10^{9}$ | 0.5 | | 0.28 | 0.72 | 0.26 | 0.74 | 0.36 | 0.64 | 0.34 | 0.66 |
| $1 \times 10^{10}$ | 1 | | 0.59 | 1.41 | 0.56 | 1.44 | 0.72 | 1.28 | 0.71 | 1.29 |
| $3 \times 10^{10}$ | 2 | | 1.24 | 2.76 | 1.21 | 2.79 | 1.47 | 2.53 | 1.46 | 2.54 |
| $6 \times 10^{10}$ | 5 | | 3.36 | 6.64 | 3.33 | 6.67 | 3.86 | 6.14 | 3.84 | 6.16 |
| $1 \times 10^{11}$ | 10 | | 7.23 | 12.77 | 7.20 | 12.80 | 8.06 | 11.94 | 8.04 | 11.96 |
| $3 \times 10^{11}$ | 20 | | 15.5 | 24.5 | 15.4 | 24.6 | 16.8 | 23.2 | 16.78 | 23.2 |
| $6 \times 10^{11}$ | 50 | | 41.3 | 58.7 | 41.2 | 58.8 | 43.9 | 56.1 | 43.85 | 56.2 |
| $1 \times 10^{12}$ | 100 | | 84.6 | 115.4 | 84.5 | 115.5 | 89.1 | 110.9 | 89.09 | 110.9 |
| $3 \times 10^{12}$ | 200 | | 172 | 228 | 172 | 228 | 180 | 220 | 180 | 220 |
| $6 \times 10^{12}$ | 500 | | 438 | 562 | 438 | 562 | 456 | 544 | 456 | 544 |
| $1 \times 10^{13}$ | 1000 | | 887 | 1113 | 887 | 1113 | 920 | 1080 | 920 | 1080 |

Table 9 shows the confidence values for different rates of eggs per gram using different testing methods taking in to consideration the number of cores, the size of the cores and the volume of soil tested. The higher the number of eggs per g the more reliable the estimate in relative terms. Increasing the number of cores makes a greater difference to the reliability of the results than increasing the volume of soil tested.




(a) Zero-inflated negative binomial (a) Zero-inflated Neyman type A

***Figure 17:*** *Effect of core size on the coefficient of variation (CV) for estimating PCN populations over a range of egg densities using 50 cores and a laboratory subsample of 200g, assuming 1 egg/cyst. Two PCN distributions are compared: the zero-inflated negative binomial and the zero-inflated Neyman type A. NB The lines in both figures are co-incident.*




(a) Zero-inflated negative binomial (b) Zero-inflated Neyman type A

***Figure 18:*** *Effect of amount of soil subsampled for laboratory analysis on the coefficient of variation (CV) over a range of egg densities using 50 cores of 50 ml, assuming 1 egg/cyst. Two PCN distributions are compared: the zero-inflated negative binomial and the zero-inflated Neyman type A*

***Table 10:*** *Zero-inflated negative binomial distribution: approximate 95% confidence intervals over a range of population levels for four subsample sizes: 50 to 400 ml. Based on one egg per cyst.*

| Eggs/ha | Eggs/g | No cores | 50 | | 50 | | 50 | | 50 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Core size (ml) | 11.78 | | 11.78 | | 11.78 | | 11.78 | |
| | | Weight tested (g) | 50 | | 100 | | 200 | | 400 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^9$ | 0.1 | | 0.003 | 0.197 | 0.025 | 0.175 | 0.039 | 0.161 | 0.048 | 0.152 |
| $3 \times 10^9$ | 0.2 | | 0.052 | 0.348 | 0.081 | 0.319 | 0.098 | 0.302 | 0.108 | 0.292 |
| $6 \times 10^9$ | 0.5 | | 0.22 | 0.78 | 0.26 | 0.74 | 0.28 | 0.72 | 0.29 | 0.71 |
| $1 \times 10^{10}$ | 1 | | 0.52 | 1.48 | 0.56 | 1.44 | 0.59 | 1.41 | 0.60 | 1.40 |
| $3 \times 10^{10}$ | 2 | | 1.17 | 2.83 | 1.21 | 2.79 | 1.24 | 2.76 | 1.25 | 2.75 |
| $6 \times 10^{10}$ | 5 | | 3.28 | 6.72 | 3.33 | 6.67 | 3.36 | 6.64 | 3.38 | 6.62 |
| $1 \times 10^{11}$ | 10 | | 7.13 | 12.87 | 7.20 | 12.80 | 7.23 | 12.77 | 7.25 | 12.75 |
| $3 \times 10^{11}$ | 20 | | 15.3 | 24.7 | 15.4 | 24.6 | 15.5 | 24.5 | 15.47 | 24.5 |
| $6 \times 10^{11}$ | 50 | | 41.1 | 58.9 | 41.2 | 58.8 | 41.3 | 58.7 | 41.33 | 58.7 |
| $1 \times 10^{12}$ | 100 | | 84.4 | 115.6 | 84.5 | 115.5 | 84.6 | 115.4 | 84.61 | 115.4 |
| $3 \times 10^{12}$ | 200 | | 172 | 228 | 172 | 228 | 172 | 228 | 172 | 228 |
| $6 \times 10^{12}$ | 500 | | 438 | 562 | 438 | 562 | 438 | 562 | 438 | 562 |
| $1 \times 10^{13}$ | 1000 | | 887 | 1113 | 887 | 1113 | 887 | 1113 | 887 | 1113 |

Table 10 illustrates that the confidence values for estimation of eggs per g is not greatly affected by the size of the subsample of soil used particularly when at higher population densities when aggregation of eggs into cysts is ignored.

***Table 11:*** *Zero-inflated negative binomial distribution: approximate 95% confidence intervals over a range of population levels for four coring rates: 25 to 200 cores per ha. Based on one egg per cyst.*

| Eggs/ha | Eggs/g | No cores | 25 | | 50 | | 100 | | 200 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Core size (ml) | 11.78 | | 11.78 | | 11.78 | | 11.78 | |
| | | Weight tested (g) | 100 | | 100 | | 100 | | 100 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^9$ | 0.1 | | 0.014 | 0.186 | 0.025 | 0.175 | 0.031 | 0.169 | 0.035 | 0.165 |
| $3 \times 10^9$ | 0.2 | | 0.056 | 0.344 | 0.081 | 0.319 | 0.095 | 0.305 | 0.104 | 0.296 |
| $6 \times 10^9$ | 0.5 | | 0.19 | 0.81 | 0.26 | 0.74 | 0.30 | 0.70 | 0.33 | 0.67 |
| $1 \times 10^{10}$ | 1 | | 0.42 | 1.58 | 0.56 | 1.44 | 0.66 | 1.34 | 0.72 | 1.28 |
| $3 \times 10^{10}$ | 2 | | 0.92 | 3.08 | 1.21 | 2.79 | 1.41 | 2.59 | 1.54 | 2.46 |
| $6 \times 10^{10}$ | 5 | | 2.68 | 7.32 | 3.33 | 6.67 | 3.78 | 6.22 | 4.08 | 5.92 |
| $1 \times 10^{11}$ | 10 | | 6.09 | 13.91 | 7.20 | 12.80 | 7.97 | 12.03 | 8.50 | 11.50 |
| $3 \times 10^{11}$ | 20 | | 13.6 | 26.4 | 15.4 | 24.6 | 16.7 | 23.3 | 17.58 | 22.4 |
| $6 \times 10^{11}$ | 50 | | 37.7 | 62.3 | 41.2 | 58.8 | 43.7 | 56.3 | 45.46 | 54.5 |
| $1 \times 10^{12}$ | 100 | | 78.2 | 121.8 | 84.5 | 115.5 | 89.0 | 111.0 | 92.07 | 107.9 |
| $3 \times 10^{12}$ | 200 | | 160 | 240 | 172 | 228 | 180 | 220 | 186 | 214 |
| $6 \times 10^{12}$ | 500 | | 412 | 588 | 438 | 562 | 456 | 544 | 469 | 531 |
| $1 \times 10^{13}$ | 1000 | | 840 | 1160 | 887 | 1113 | 920 | 1080 | 943 | 1057 |

Table 11 shows that increasing the number of cores taken continues to have an effect of reducing the confidence interval even at the highest levels of infestation.

The effect of aggregation of eggs within cysts on estimation accuracy based on subsample size.

For simplicity of modelling, the previous section to determine the most important factors influencing population estimation was based on a uniform distribution of eggs throughout the sample. This is only possible if eggs were independent from cysts or were found at a rate of one egg per cyst. The following graph and tables display the confidence intervals if the aggregation within cysts is increased to 10, 50 or 200 eggs per cyst.

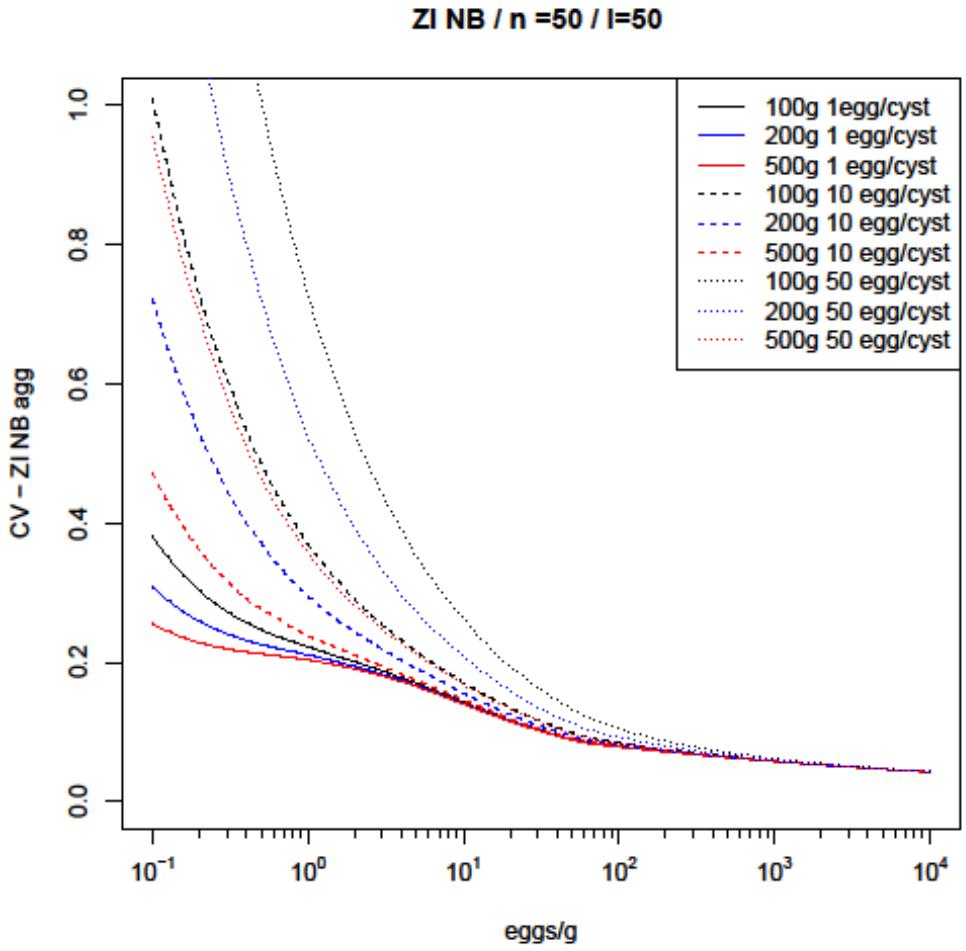



***Figure 19:*** *Effect of aggregation into cysts combined with soil subsampling for testing on the coefficient of variation for the egg density with 50 cores of 50 ml. Based on 100,000 simulations except when 1 egg/cyst is assumed (direct calculation)*

**Figure 19 shows how the precision of egg density estimates is reduced when aggregation into cysts is taken into account. Note that this only affects subsampling; if the whole sample is tested then there is no reduction**. The higher the eggs per cysts, the greater the CV. The increase in the CV is larger with lower egg densities and with smaller subsamples. Without aggregation, 100 g, 200 g and 400 g subsamples have similar CVs over and above about 2 eggs/g. This convergence occurs at higher densities when aggregation is taken into account.

Table 12 shows the approximate 95% confidence intervals under different levels of eggs/cyst. Note that the approximation used for the confidence interval becomes poor when the CV is high and this is why some confidence intervals have negative lower limits.

***Table 12:*** *Zero-inflated negative binomial estimation of approximate 95% confidence limits based on subsampling*

*(a) Ignoring aggregation of eggs in cysts (equivalent to 1 egg per cyst – calculated CVs)*

| Eggs/ha | Eggs/g | No cores | 50 | | 50 | | 50 | | 50 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Core size (ml) | 11.78 | | 11.78 | | 11.78 | | 11.78 | |
| | | Weight tested (g) | 50 | | 100 | | 200 | | 400 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^9$ | 0.1 | | 0.003 | 0.197 | 0.025 | 0.175 | 0.039 | 0.161 | 0.048 | 0.152 |
| $3 \times 10^9$ | 0.2 | | 0.052 | 0.348 | 0.081 | 0.319 | 0.098 | 0.302 | 0.108 | 0.292 |
| $6 \times 10^9$ | 0.5 | | 0.22 | 0.78 | 0.26 | 0.74 | 0.28 | 0.72 | 0.29 | 0.71 |
| $1 \times 10^{10}$ | 1 | | 0.52 | 1.48 | 0.56 | 1.44 | 0.59 | 1.41 | 0.60 | 1.40 |
| $3 \times 10^{10}$ | 2 | | 1.17 | 2.83 | 1.21 | 2.79 | 1.24 | 2.76 | 1.25 | 2.75 |
| $6 \times 10^{10}$ | 5 | | 3.28 | 6.72 | 3.33 | 6.67 | 3.36 | 6.64 | 3.38 | 6.62 |
| $1 \times 10^{11}$ | 10 | | 7.13 | 12.87 | 7.20 | 12.80 | 7.23 | 12.77 | 7.25 | 12.75 |
| $3 \times 10^{11}$ | 20 | | 15.3 | 24.7 | 15.4 | 24.6 | 15.5 | 24.5 | 15.47 | 24.5 |
| $6 \times 10^{11}$ | 50 | | 41.1 | 58.9 | 41.2 | 58.8 | 41.3 | 58.7 | 41.33 | 58.7 |
| $1 \times 10^{12}$ | 100 | | 84.4 | 115.6 | 84.5 | 115.5 | 84.6 | 115.4 | 84.61 | 115.4 |
| $3 \times 10^{12}$ | 200 | | 172 | 228 | 172 | 228 | 172 | 228 | 172 | 228 |
| $6 \times 10^{12}$ | 500 | | 438 | 562 | 438 | 562 | 438 | 562 | 438 | 562 |
| $1 \times 10^{13}$ | 1000 | | 887 | 1113 | 887 | 1113 | 887 | 1113 | 887 | 1113 |

*(b) including aggregation of eggs in cysts – 10 eggs per cyst (100,000 simulations)*

| Eggs/ha | Eggs/g | No cores | 50 | | 50 | | 50 | | 50 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Core size (ml) | 11.78 | | 11.78 | | 11.78 | | 11.78 | |
| | | Weight tested (g) | 50 | | 100 | | 200 | | 400 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^9$ | 0.1 | | -0.179 | 0.379 | -0.098 | 0.298 | -0.041 | 0.241 | -0.002 | 0.202 |
| $3 \times 10^9$ | 0.2 | | -0.200 | 0.600 | -0.084 | 0.484 | -0.008 | 0.408 | 0.045 | 0.355 |
| $6 \times 10^9$ | 0.5 | | -0.15 | 1.15 | 0.02 | 0.98 | 0.14 | 0.86 | 0.21 | 0.79 |
| $1 \times 10^{10}$ | 1 | | 0.05 | 1.95 | 0.28 | 1.72 | 0.42 | 1.58 | 0.51 | 1.49 |
| $3 \times 10^{10}$ | 2 | | 0.56 | 3.44 | 0.86 | 3.14 | 1.05 | 2.95 | 1.15 | 2.85 |
| $6 \times 10^{10}$ | 5 | | 2.47 | 7.53 | 2.90 | 7.10 | 3.12 | 6.88 | 3.26 | 6.74 |
| $1 \times 10^{11}$ | 10 | | 6.12 | 13.88 | 6.65 | 13.35 | 6.94 | 13.06 | 7.12 | 12.88 |
| $3 \times 10^{11}$ | 20 | | 14.0 | 26.0 | 14.7 | 25.3 | 15.1 | 24.9 | 15.3 | 24.7 |
| $6 \times 10^{11}$ | 50 | | 39.4 | 60.6 | 40.3 | 59.7 | 40.8 | 59.2 | 41.1 | 58.9 |
| $1 \times 10^{12}$ | 100 | | 82.3 | 117.7 | 83.4 | 116.6 | 84.0 | 116.0 | 84.4 | 115.6 |
| $3 \times 10^{12}$ | 200 | | 169 | 231 | 171 | 229 | 171 | 229 | 172 | 228 |
| $6 \times 10^{12}$ | 500 | | 435 | 565 | 437 | 563 | 437 | 563 | 438 | 562 |
| $1 \times 10^{13}$ | 1000 | | 884 | 1116 | 885 | 1115 | 887 | 1113 | 887 | 1113 |

*(c) including aggregation of eggs in cysts – 50 eggs per cyst (100,000 simulations)*

| Eggs/ha | Eggs/g | No cores | 50 | | 50 | | 50 | | 50 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Core size (ml) | 11.78 | | 11.78 | | 11.78 | | 11.78 | |
| | | Weight tested (g) | 50 | | 100 | | 200 | | 400 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^9$ | 0.1 | | -0.524 | 0.724 | -0.336 | 0.536 | -0.205 | 0.405 | -0.113 | 0.313 |
| $3 \times 10^9$ | 0.2 | | -0.679 | 1.079 | -0.421 | 0.821 | -0.236 | 0.636 | -0.107 | 0.507 |
| $6 \times 10^9$ | 0.5 | | -0.90 | 1.90 | -0.49 | 1.49 | -0.21 | 1.21 | -0.01 | 1.01 |
| $1 \times 10^{10}$ | 1 | | -0.99 | 2.99 | -0.42 | 2.42 | -0.03 | 2.03 | 0.24 | 1.76 |
| $3 \times 10^{10}$ | 2 | | -0.86 | 4.86 | -0.07 | 4.07 | 0.46 | 3.54 | 0.81 | 3.19 |
| $6 \times 10^{10}$ | 5 | | 0.36 | 9.64 | 1.53 | 8.47 | 2.33 | 7.67 | 2.83 | 7.17 |
| $1 \times 10^{11}$ | 10 | | 3.25 | 16.75 | 4.87 | 15.13 | 5.93 | 14.07 | 6.55 | 13.45 |
| $3 \times 10^{11}$ | 20 | | 10.2 | 29.8 | 12.4 | 27.6 | 13.8 | 26.2 | 14.6 | 25.4 |
| $6 \times 10^{11}$ | 50 | | 33.8 | 66.2 | 37.1 | 62.9 | 39.0 | 61.0 | 40.2 | 59.8 |
| $1 \times 10^{12}$ | 100 | | 75.2 | 124.8 | 79.3 | 120.7 | 82.0 | 118.0 | 83.3 | 116.7 |
| $3 \times 10^{12}$ | 200 | | 161 | 239 | 166 | 234 | 169 | 231 | 171 | 229 |
| $6 \times 10^{12}$ | 500 | | 424 | 576 | 431 | 569 | 435 | 565 | 436 | 564 |
| $1 \times 10^{13}$ | 1000 | | 871 | 1129 | 879 | 1121 | 883 | 1117 | 886 | 1114 |

*(d) including aggregation of eggs in cysts – 200 eggs per cyst (100,000 simulations)*

| Eggs/ha | Eggs/g | No cores | 50 | | 50 | | 50 | | 50 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Core size (ml) | 11.78 | | 11.78 | | 11.78 | | 11.78 | |
| | | Weight tested (g) | 50 | | 100 | | 200 | | 400 | |
| | | | lower | upper | lower | upper | lower | upper | lower | upper |
| $1 \times 10^9$ | 0.1 | | -1.146 | 1.346 | -0.784 | 0.984 | -0.516 | 0.716 | -0.320 | 0.520 |
| $3 \times 10^9$ | 0.2 | | -1.537 | 1.937 | -1.026 | 1.426 | -0.656 | 1.056 | -0.395 | 0.795 |
| $6 \times 10^9$ | 0.5 | | -2.27 | 3.27 | -1.45 | 2.45 | -0.87 | 1.87 | -0.46 | 1.46 |
| $1 \times 10^{10}$ | 1 | | -2.92 | 4.92 | -1.77 | 3.77 | -0.96 | 2.96 | -0.37 | 2.37 |
| $3 \times 10^{10}$ | 2 | | -3.56 | 7.56 | -1.94 | 5.94 | -0.81 | 4.81 | 0.00 | 4.00 |
| $6 \times 10^{10}$ | 5 | | -3.88 | 13.88 | -1.34 | 11.34 | 0.46 | 9.54 | 1.65 | 8.35 |
| $1 \times 10^{11}$ | 10 | | -2.62 | 22.62 | 0.92 | 19.08 | 3.39 | 16.61 | 5.04 | 14.96 |
| $3 \times 10^{11}$ | 20 | | 2.0 | 38.0 | 7.0 | 33.0 | 10.3 | 29.7 | 12.6 | 27.4 |
| $6 \times 10^{11}$ | 50 | | 21.1 | 78.9 | 28.8 | 71.2 | 34.0 | 66.0 | 37.3 | 62.7 |
| $1 \times 10^{12}$ | 100 | | 58.2 | 141.8 | 68.6 | 131.4 | 75.4 | 124.6 | 79.8 | 120.2 |
| $3 \times 10^{12}$ | 200 | | 138 | 262 | 152 | 248 | 161 | 239 | 166 | 234 |
| $6 \times 10^{12}$ | 500 | | 393 | 607 | 413 | 587 | 425 | 575 | 432 | 568 |
| $1 \times 10^{13}$ | 1000 | | 832 | 1168 | 858 | 1142 | 872 | 1128 | 880 | 1120 |

Using a 100 g subsample at 10 eggs per g with 1 egg per cyst, the confidence interval is 7.2 to 12.8. At 10 eggs per cyst this range extends from 6.7 to 13.4. At 50 eggs per cyst the range for 10 eggs per g is 4.9 to 15.1 and at 200 eggs per cyst the expected range is from 0.9 to 19.1 thus demonstrating that **the level of aggregation of eggs within cysts has a major impact on the confidence limits when estimating population levels.** Where a 400 ml subsample is taken with 200 eggs per cyst at 10 eggs per g, the range is reduced to 5.0 to 15.0, markedly improving the estimation. This shows that the larger the subsample size, the smaller the impact of cyst aggregation on the estimation.

***Table 13:*** *Zero-inflated negative binomial estimation of approximate 95% confidence limits based on subsampling. Three target populations of A 5 eggs/g, B 10 eggs/g and C 20 eggs/g. Data are tabulated for four subsampling rates of 100, 200, 400 and 1000g, and for six levels of aggregation of 10, 50, 100, 150, 200 and 400 eggs/cyst. A colour coding system is used: green if the difference between the upper and lower confidence limits is less than the target population, yellow if the difference is greater than the target population by a factor of less than 1.5, and red if the difference is greater than 1.5 x the target population. Where the lower confidence limit is less than zero, a zero value has been recorded.*

**A Target Population = 5 eggs/g**

| Subsample | 100g | | 200g | | 400g | | 1000g | |
|---|---|---|---|---|---|---|---|---|
| Eggs/Cyst | lower CL | upper CL | lower CL | upper CL | lower CL | upper CL | lower CL | upper CL |
| 10 | 2.9 | 7.1 | 3.1 | 6.9 | 3.3 | 6.8 | 3.4 | 6.7 |
| 50 | 1.5 | 8.5 | 2.3 | 7.7 | 2.8 | 7.2 | 3.2 | 6.8 |
| 100 | 0.4 | 9.6 | 1.6 | 8.4 | 2.4 | 7.6 | 3.0 | 7.0 |
| 150 | 0.0 | 10.6 | 1.0 | 9.0 | 2.0 | 8.0 | 2.8 | 7.2 |
| 200 | 0.0 | 11.4 | 0.4 | 9.6 | 1.7 | 8.4 | 2.7 | 7.3 |
| 400 | 0.0 | 13.8 | 0.0 | 11.3 | 0.6 | 9.5 | 2.1 | 7.9 |

**B Target Population = 10 eggs/g**

| Subsample | 100g | | 200g | | 400g | | 1000g | |
|---|---|---|---|---|---|---|---|---|
| Eggs/Cyst | lower CL | upper CL | lower CL | upper CL | lower CL | upper CL | lower CL | upper CL |
| 10 | 6.7 | 13.4 | 7.0 | 13.0 | 7.1 | 12.9 | 7.2 | 12.8 |
| 50 | 4.9 | 15.1 | 5.9 | 14.1 | 6.6 | 13.4 | 7.0 | 13.0 |
| 100 | 3.3 | 16.7 | 4.9 | 15.1 | 6.0 | 14.0 | 6.8 | 13.2 |
| 150 | 2.0 | 18.0 | 4.1 | 15.9 | 5.5 | 14.5 | 6.6 | 13.4 |
| 200 | 0.9 | 19.1 | 3.3 | 16.7 | 5.0 | 15.0 | 6.4 | 13.6 |
| 400 | 0.0 | 22.6 | 1.0 | 19.0 | 3.5 | 16.5 | 5.6 | 14.4 |

**C Target Population = 20 eggs/g**

| Subsample | 100g | | 200g | | 400g | | 1000g | |
|---|---|---|---|---|---|---|---|---|
| Eggs/Cyst | lower CL | upper CL | lower CL | upper CL | lower CL | upper CL | lower CL | upper CL |
| 10 | 14.7 | 25.3 | 15.1 | 24.9 | 15.3 | 24.7 | 15.4 | 24.6 |
| 50 | 12.4 | 27.6 | 13.8 | 26.2 | 14.6 | 25.4 | 15.2 | 24.8 |
| 100 | 10.3 | 29.7 | 12.4 | 27.6 | 13.8 | 26.2 | 14.9 | 25.1 |
| 150 | 8.5 | 31.5 | 11.4 | 28.6 | 13.2 | 26.8 | 14.6 | 25.4 |
| 200 | 7.0 | 33.0 | 10.3 | 29.7 | 12.6 | 27.5 | 14.3 | 25.7 |
| 400 | 2.15 | 37.85 | 7.05 | 33.0 | 10.5 | 29.5 | 13.4 | 26.6 |

Table 13 summarizes the relationship between the confidence interval associated with sampling as the PCN population increases, the aggregation of eggs into cysts increases and the size of the subsample taken increases. We have arbitrarily taken a confidence interval of less than the target population as acceptable (green shading), a confidence interval of between 1 and 1.5 x the target population as marginally acceptable (yellow shading) and a confidence interval of greater than 1.5 x the target population as unacceptable (red shading). Others may disagree with such a definition of acceptability, but we have used it as a baseline for reviewing recommendations on subsampling for the estimation of PCN populations.

Table 13 clearly shows that smaller subsamples can be taken when the target populations are higher and the cysts are less aggregated, i.e. present in a greater number of cysts. For example, with a target population of 20 eggs/g, the confidence interval associated with an aggregation of 200 eggs/cyst and a subsample of 200g would be 10.3 to 29.7 eggs/g. With a target population of 10 eggs/g, the confidence interval associated with a similar level of aggregation and subsampling would be 3.3 to 16.7 eggs/g which we have viewed as only marginally acceptable. With a target population of 5 eggs/g, the confidence interval associated with a similar level of aggregation and subsampling would be 0.4 to 9.6 eggs/g which we view as unacceptable.

The aggregation into eggs/cyst will depend upon initial number of eggs per fresh cyst, typically 300-400, and the annual decline rate. Decline rates are typically higher for *G. rostochiensis* populations than for *G. pallida*. Low levels of eggs/cyst (below 100eggs/cyst) are unlikely to be encountered in fields that have grown susceptible potato cultivars in the last 5 years. Therefore, subsampling to as little as 100g of soil per sample, is only likely to be acceptable under a scenario where the target population is 20 or more eggs/g of soil and the land has been out of potato production for at least 5 years. A subsample of 200g is recommended for shorter rotations. For target populations of 10 eggs/g, a subsample of at least 400g of soil is recommended.

***Table 14:*** *The effect on confidence limits of changing the number of cores based on the zero-inflated negative binomial model with aggregation of eggs into cysts. Here the target population is 10 eggs/g with 100 eggs/cyst assumed, core size of 50 ml and subsample size of 200g. Estimation of approximate 95% confidence limits based on subsampling. A similar colour coding system to Table 12 has been used to indicate the size of the confidence interval in relation to the target population.*

| **Target Population = 10 eggs/g, Subsample size = 200g** | | | | |
|---|---|---|---|---|
| **Eggs/Cyst** | **100** | | **400** | |
| No. of cores | lower CL | upper CL | lower CL | upper CL |
| 25 | 4.3 | 15.7 | 0.8 | 19.2 |
| 50 | 4.9 | 15.1 | 1.0 | 19.0 |
| 100 | 5.3 | 14.7 | 1.1 | 18.9 |
| 200 | 5.4 | 14.6 | 1.2 | 18.8 |
| 400 | 5.5 | 14.5 | 1.2 | 18.8 |

Table 14 indicates the effect of increasing the number of cores used to collect a sample (25 to 400) when the target population is 10 eggs/g and a subsample of 200g is used for laboratory analysis. Two levels of aggregation are illustrated, 100 and 400 eggs/cyst. As with sampling to detect, taking more than 100 cores per ha provides minimal additional benefit, the differences between 50 and 100 cores appear to be less marked. This is in agreement with Figure 16, which presents the results of a similar analysis in graphical format, albeit without any aggregation of eggs within cysts.

# 5. DISCUSSION

Field distributions of PCN in Britain

Been and Schomaker's model for describing the aggregated distribution of cysts within a field is suitable for describing early infestations. The model is based on the introduction of PCN following the planting of seed potato tubers with a very low level of PCN contamination. The PCN population will increase around the original site of introduction on each successive potato crop and will be spread locally by cultivation and from localised movement of the juveniles to surrounding plants. These populations will thus create a focus or hotspot within the field. If there has been more than one initial introduction of PCN, then more than one hotspot can occur. In the Netherlands, three year rotations are generally used which does not allow as much spread with cultivation compared with England and Wales, where the most common rotation used for potatoes was one in five (Minnis *et al*., 2002). Seed potatoes have typically been grown in long rotations – in Scotland a minimum rotation of 1 crop in 6 years is a requirement of the Seed Potato Classification Scheme. The land used for seed production has typically not been used for potatoes other than seed, and the seed planted would itself have been grown on land that had been tested and found free from PCN. Until recently, at least some seed production in Scotland had been able to use virgin land (i.e. no previous potato production). Although seed potato production prior to the UK entry into the EEC in 1973 did not require a statutory PCN test, voluntary testing was widely adopted, at least within the Scottish seed industry. It is therefore reasonable to assume that in seed potato land, distributions of the type described by Been & Schomaker could be encountered with reasonable regularity. However, the frequency with which populations of dead cysts have been encountered in Scottish statutory pre-crop soil testing (Table 2) suggests that many populations in seed potato land may be considerably older and therefore less aggregated than a typical Been & Schomaker population. Such old infestations may be effectively managed by long rotations and/or the cultivation of resistant cultivars.

PCN populations in ware potato land have developed over a long period and most of the land used for potato growing in England and Wales is already infested (Minnis *et al*., 2002). PCN have been known to be present in Britain for over 100 years. Winfield (1965) suggested that up to 84% of fields in Lincolnshire were infested by the 1960s. Seed potato production prior to the 1960s did not require a PCN test, neither has there ever been a requirement for unclassified ('farm-saved') seed to be produced on 'PCN-free' land. Thus it is highly likely that many PCN infestations are well established and have spread widely throughout infested fields. Only since 2010, following the revision of the EU PCN Directive, in have PCN controls similar to those for classified seed been applied to 'farm-saved' seed in Great Britain (and then only to any major extent in Scotland).

It is likely that well established field populations of PCN will be less heterogeneous than the Been and Schomaker model would suggest. The evidence that we have collected supports the view that there is a lot more variation between fields in the nature of their PCN distribution patterns than is suggested by the Been and Schomaker model. We therefore propose that Been and Schomaker model is more suitable for describing the PCN distributions during the early stages of infestation development. More established populations seem to be better described by the zero-inflated negative binomial distribution, a conclusion that we have supported using data from British fields. Thomas

Been (pers. comm.) describes these less heterogeneous distributions as 'full-field' infestations.

For older distributions, we suggest that probability distributions can be used to model the counts of eggs (or cysts). Whilst these models are less mechanistic, they avoid the need to consider the wide range of possibilities of spatial distributions with a more developed infestation. We showed that such models need to accommodate not only heterogeneity in counts but also to allow for a greater number of zero counts than the base model predicts. A suitable model was provided by the zero-inflated negative binomial distribution; which fitted the example data sets reasonably well and allowed calculation of both detection probability and precision of estimation (through the coefficient of variance). An alternative to this model is the zero-inflated Neyman type A distribution; whilst at the time we have not been able to compare the fit of this model, conclusions about sampling strategy are very similar.

## Soil sampling aims

It is clear from our study and subsequent analysis that the motives for sampling are critical for determining how the land should be sampled and how the soil should be treated in the laboratory. For seed potato production and for land believed to be free from PCN, the aim of PCN testing should be detection of infestations. Where there is a known infestation and information is required to underpin management options, quantification or population estimation is required. The optimum approaches to soil testing differ between these two options.

## Detection of PCN

For the detection of PCN the aim is to detect one or more viable cysts. The detection of PCN may be relevant in a number of situations. All seed land within the EU must be tested and found to be free from PCN. PCN is widespread in ware land in many areas of Great Britain, but growers intending to plant ware in land thought to be free from PCN may be better advised to adopt a sampling strategy aimed at detection rather than quantification. If PCN are present, then it is more likely that the field will show an early stage infestation and the model of Been and Schomaker will provide the most appropriate description of the underlying distribution of cysts. As this model describes a highly aggregated distribution, adoption of such a distribution errs on the side of caution when estimating likelihood of detection.

Our analyses show that the major factor when sampling for detection is the amount of soil sampled and analysed in the laboratory. Increasing the core size only slightly increases the probability of detection if the effect of increasing the total volume of soil in the sample is countered by laboratory subsampling. Increasing the number of cores does significantly increase the chance of detection, but there is relatively little to gain by increasing the number of cores beyond 100 per ha. These conclusions are based on sampling in a grid pattern. The amount of time spent sampling a field, and hence the cost of sampling can be reduced by adopting a W-path to cover the field. The grid pattern provides the highest chance of detection although a 4-arm W-pattern can be nearly as effective in detecting PCN, providing that the W-pattern is orientated across the direction of cultivation. If the focus of infestation is elongated in the direction of the W-pattern, then the probability of missing the focus is increased (Figure 15).

It is virtually impossible to detect a single cyst in a field, with the likelihood of detection increasing with the amount of cysts present in the soil and the volume of soil tested. It takes several potato crops for PCN to build up to levels that are detectable using soil sampling rates in commercial use. Even at the EU standard rate of 1500 ml from 1 ha using 100 cores, to have a greater than 90% chance of detecting a single cyst there would need to be 3,800,000 in one hectare.

## Quantification of PCN populations

The EU PCN Directive prohibits seed potato production on infested land and therefore quantitative information is not necessary for soil testing prior to seed production. For land to be used for ware production it is highly valuable to know the level of infestation of PCN in the soil in order to underpin management options, such as cultivar choice, nematicide treatment, extending rotations, etc. The most critical factor for accuracy of the estimation, when not factoring in aggregation of eggs within cysts, is the number of cores that have been taken, with higher numbers of cores giving a greater accuracy in the population estimation. At 200 cores the confidence limits are ±1.50 for 10 eggs/g however this increases to ±2.03 with 100 cores (Table 11).

There is a question that relates to the level of accuracy that is required in relation to population estimation, given that threshold values for treating PCN infestations are highly subjective and given that yield losses are generally perceived to be highly dependent upon soil type, cultivar tolerance, irrigation and nutrient availability. In reality, an accurate diagnosis that includes the quantification of both species of PCN is of considerable value, given that planting with a susceptible cultivar can increase the post-harvest PCN population 50-fold.

## Subsampling

If the sample has been taken for detection purposes, any subsampling will reduce the likelihood of detection and is therefore not recommended. In order to use subsampling for quantification purposes, the soil within the original sample should be thoroughly mixed to ensure an even distribution of PCN. Any aggregation of cysts within the sample or eggs within cysts will increase the width of the confidence limits associated with subsampling. Using a uniform distribution of eggs within the sample and the population in excess of 1 egg/g, taking either a 100 g or 200 g subsample will have little effect on the final outcome. However taking into account that cysts generally contain many eggs, the volume of soil subsampled has an important effect of accuracy, especially at lower population levels.

Our analysis showed little difference between the accuracy of the population estimation, regardless of the level of subsampling that took place down to a minimum of 50 g of soil where there is greater than 1 egg/g providing there is a uniform distribution of eggs within the sample. This would only occur if eggs were not aggregated within cysts, i.e. one egg per cyst. It is exceptionally unlikely that there will be a uniform distribution of eggs within cysts at a level of 1 egg per cyst. Aggregation of eggs within cysts has a significant effect on the confidence limits of population estimates. With a target of 10 eggs per g where a 100g subsample is taken for analysis, the confidence limits are ±2.80 with an aggregation of 1 egg per cyst, increasing to ±9.08 with 200 eggs per cyst.

The impact of the aggregation of eggs within cysts can be mitigated against by increasing the size of the subsample, e.g. the confidence limits are reduced to ±4.96 if the subsample is increased to 400 g at an aggregation of 200 eggs per cyst. The effect of the aggregation of eggs within cysts decreases with increasingly high populations, i.e. when many more cysts will be present within a subsample.

Heterogeneity and age of field infestations

Making best use of the available data on field distributions, the hypothesis that older infestations are more likely to be less heterogeneous than recent infestations appears to be supported by data showing that the heterogeneity index declines with increasing egg density (Figure 7). This may in part be due to the age of the infestation as higher populations of PCN are more likely to be older and therefore to have been more widely distributed across the field by cultivation, creating many more new foci of infestation. This relationship between the level of infestation and the level of heterogeneity therefore seems reasonable, although the conclusion has been reached based on a study of only relatively few fields. It would be beneficial to confirm this observation on a wider set of fields as this relationship provides an important element of our guidance on sampling.

Other Factors

Timing of sampling

The time that the sample is taken is largely dependent upon the reasons for the sample being taken. If the aim of sampling is to indicate what impact the current PCN infestation may have on the next crop, then pre-planting will provide the best information. If the aim is to detect very low populations the optimum time for this would be immediately post-harvest. To assess the effectiveness of management practices then the field should be tested both prior to planting and directly after harvest in order to assess the initial and final population.

In the absence of susceptible groundkeepers, which will continue to host PCN and amplify populations, sampling after harvest makes PCN more likely to be detected as populations should be at their highest. Therefore, for detection purposes, it can be argued that this is the best time to sample. However, for the detection of PCN cysts with live content, it is unlikely that such cysts will either be lost in entirety or will lose viability over the length of a normal potato rotation. Therefore, using most of the diagnostic methods in current use, sampling prior to planting should not significantly reduce the chance of detection. If a highly resistant cultivar has been grown then this in itself is likely to markedly reduce the ability to detect viable PCN.

Decline rates are known to vary according to environmental conditions and there is evidence that microbial populations are able to suppress PCN populations in field soils. Therefore, the best time to assess population levels will be shortly before planting. As too many factors can take effect over the 4-6 year period of a typical rotation period, it is probably too unreliable to extrapolate population levels at planting from a post-harvest test following the previous potato crop.

Sample Depth

Much of the current guidance on soil sampling suggests that cores should be taken to a depth of 15 cm, but there is little evidence to support the value of this. Research in the Netherlands has looked at the distribution of cysts within a vertical plane. Cores of soil were taken to an overall depth of 80 cm and samples analysed separately over the full range of depths (Been and Schomaker, 2013). The conclusion from this research was a uniform distribution of cysts within the top 25 cm of the soil profile, both directly after harvest and after cultivation. Similarly, Boag & Neilson, (1994) concluded that sampling at any depth within the top 20 to 25 cm is suitable. Therefore, a corer taken to a depth of 5 cm should be adequate. Up to 90% of cysts are located within the top 30 cm of soil, so densities will decline at greater depths. However, some cysts can be found at depths as great as 80 cm, making it difficult to target all cysts with a nematicide application.

Spatial scaling.

When sampling a field for PCN the smaller the size of the block or area from which the sample is drawn, the greater will be the discrimination between areas of differing PCN incidence. Whether such finer scale resolution is cost effective or not depends upon the ability/willingness of the grower to make management decisions at this scale of resolution, and the extent of variation that is either expected or found. In one field not used in this study, 1 ha blocks indicated that one side of the field had a higher PCN infestation than the other, but when looking at quarter hectare blocks it was possible to see how the PCN had been spread as a result of a water flow during flooding of the field, which could not have been determined at a lower resolution. If the entire field is to be managed in the same way then it is not as critical to sample at a higher resolution, however the knowledge of the distribution could help with better management of the field and an understanding of the spread within the field. The optimum shape for each block is likely to be elongated in the direction of cultivation, although without more precise geographical information we are unable to expand upon the extent of the elongation.

The size of the area from which a sample is drawn will have a significant effect on the outcome of a soil test, particularly when sampling for detection. We have based the vast majority of the work of this project on the assumption that we are sampling a unit of 1 ha, but fields come in many shapes and sizes and we have been asked whether we can recommend how to effectively sample larger areas, in particular whether savings can be achieved by reducing sampling over larger areas. It is easier to examine this by separating out the two aspects of sampling: sampling for detection and sampling for population estimation.

**Sampling for detection**: This depends very much upon the scenario and the purpose of the sampling. A simple scenario is the detection of a single focus of infestation, with a similar central population density as described by the Been and Schomaker distribution model, present within an entire field, but changing the field size from say 1 ha to 4 ha. In this scenario, the probability of detection will only remain constant if the same sampling rate (cores/ha) is applied across the 1 ha and 4 ha fields. However, if only the same number and size of cores are taken from both field sizes, then the likelihood of detection will be reduced in the 4ha field as only ¼ of the amount of soil will be drawn from the area containing the infestation focus. Another simple scenario is

to consider that the 4 ha unit contains four separate similar foci, one in each hectare block then, with a similar number and size of cores taken per ha, the probability of detection is increased as there are now four chances to hit the focus as opposed to one.

In reality, both of these scenarios are unlikely to reflect reality, both in terms of the PCN distributions and the likely approach to sampling. It is also necessary to consider the purpose of the sampling and the action taken as a result of a positive sample. Any sampling approach in effect sets a threshold for the detection of an infestation. Therefore, if the consequential action after finding PCN is to record the whole field/unit as infested, then it is not practical to establish a sampling protocol that treats fields/units of differing sizes in an equitable manner.

**Sampling for population estimation**: Scaling up to areas of greater than 1 ha when estimating population size is also complicated. A key factor here is the likely variation of PCN population levels across a field. We have shown that at low infestation levels, the distribution of PCN is likely to be more heterogeneous. It is reasonable to assume that high level infestations are older infestations that are more likely to be widespread throughout a field. So there could be an argument to take samples from larger units in fields which are known to have long term or "whole-field" infestations. We also know that PCN populations are influenced markedly by the resistance of a cultivar and by nematicide treatments, as well as by other factors such as tolerance and soil type. Therefore, knowledge of past cultivation practices, where available, should influence sampling decisions, with areas with known differences in previous cultivation practices treated as separate units for sampling purposes. As with sampling for detection (above), the purpose of the sampling should also be considered. What is the level of resolution of the PCN population estimations within a field that is useful to the grower? The smaller an area that is sampled the finer will be the resolution of the distribution of PCN infestations across the field, the importance of this may depend upon the practicalities of implementing different management options on the same scale of resolution.

**Further investigations:** More work on detailed sampling of a wide range of fields is required to gain an accurate insight into the heterogeneity of PCN distributions within fields. This would allow more accurate conclusions to be drawn on the impact of sampling fields using a range of sampling areas, as well as establishing the optimal way to sample areas of different sizes. This is of particular relevance to sampling for population estimation.

Laboratory diagnostics, Cysts, Eggs and PCR

There are several different methods available for detection, speciation and quantification of PCN. For detection; cyst counts and real-time PCR can be used. For speciation; identification using morphological characteristics of the cysts and eggs can be used and for quantification egg counts and cycle threshold (CT) values from real time PCR. Cysts give an indication that PCN are or have been present in the past; the presence of cysts gives no indication of how viable they are. There is no relationship between the number of cysts present in a sample and the number of viable cysts (Table 2).

Eggs only exist in cysts but the proportion of cysts with eggs can vary considerably. As an example, PCN will be introduced into part of a field probably with only a few cysts

(say 100) and without a crop the PCN (egg) population will decline by 20-30% per annum. In the presence of a susceptible crop, the PCN (egg) population will increase about 50-fold, through the production of new cysts, each containing an initial population of approx. 300 eggs. Thus foci will develop around the original sites of introduction. When the population has increased to 5 million cysts/ha, there is a good likelihood that cyst(s) will be detected in a 400 ml soil sample. Real time PCR of DNA extracted from soil float material can detect 1 or 2 eggs in a cyst in a 400 ml sample – or 0.0025 eggs/g. One highly viable cyst could equate to 0.5 eggs/g.

In most cases, it can be hypothesised that by the time the field (ha) has over 1 egg/g, laboratories should be encountering c. 20 viable cysts per 400 ml (cysts with an average content of 30 eggs/cyst = 600 eggs/400 ml (or 600 g)). Average rotations of 6 years (common in Britain) should reduce eggs/cyst to around 30 eggs/cyst, but 3 year rotations (the norm for many parts of the world) would only reduce to about 100 eggs/cyst.

Although we consider detection of PCN in relation to cysts per kg and population estimation in eggs per g, for borderline fields we may be interested in converting between these two scales. This is not easy as the eggs per cyst may vary considerably, as we've shown. For cysts with many live eggs we might have, say, 300 eggs per cyst. So 1 egg per g equates to 3.3 cysts per kg. Whereas a lower level of 10 eggs per cyst would mean that 1 egg per g corresponds to 100 cysts per kg.

So for British situations, once the population is greater than 1 egg/g, then estimating the populations becomes the priority rather than detection. At that point it is eggs rather than cysts that are important as cysts with no eggs are not relevant and viability (eggs/cyst) can vary extensively between the remaining cysts. However, the aggregation of eggs within cysts will have an impact on the effectiveness of subsampling. Once estimation is the goal, the Been & Schomaker model is less suitable and the zero-inflated negative binomial distribution is more appropriate. As the distribution gets older the heterogeneity of the distribution will tend to decrease – unless modified by control practices such as cultivation of cultivars with differing degrees of resistance.

## Speciation

Over the past 50 years, the PCN population in Scotland and in many places elsewhere has altered from being predominantly *G. rostochiensis* to predominantly *G. pallida*, most likely as a result of the extensive cultivation of potato varieties resistant to *G. rostochiensis*. A study of field populations was carried out by Minnis *et al*. (2002) which found PCN to be present in 64% of fields sampled – based on a 200 g subsample from 50 cores taken on a grid pattern from an area of up to 4 ha. Of the 64% of samples that tested positive, 92% were either pure *G. pallida* or a mixed population (with *G. rostochiensis*). It is important to seek speciation as well as detection or quantification from a laboratory. In a few cases the presence of both species may require sampling for both quantification and detection. That is detection of the species that may be present in lower quantities and quantification of the prevalent species.

Standardised Protocol

From the studies carried out, we were able to determine that PCN distributions will vary considerably between fields and the information required from sampling will vary according to whether a population assessment or an assessment of the population level of the two species of PCN is appropriate. Therefore we propose two protocols for sampling fields according to whether the purpose is detection or population estimation.

Table 1 shows that for detection of PCN the standard rate of 100 cores per hectare and a total of 1500 ml of soil as laid out by the European Council PCN Directive (2007/33/EC) provides the greatest probability of detection (compared to other options based on fewer cores/ha). However, even at the standard EU rate, the population within the field would have to exceed 2,000,000 per hectare to achieve at least a 50% chance of detection if there is a single focus of PCN within the field. The likelihood of detection will increase if there are several separate foci. The depth of the cores and the size of the cores is not important, but the larger the sample that is tested, the greater is the probability of detection. Subsampling should not be used, otherwise the probability of detection will be significantly reduced.

For estimating PCN populations in fields with known infestations, we recommend that at least 50 cores per hectare are taken. The sample or subsample should be a minimum of 200 g assuming that the soil sample has been thoroughly mixed prior to subsampling and that there is likely to be a relatively high number of viable cysts of relatively low viability. Where cyst viability is likely to be relatively high, e.g. over 50 eggs/cyst and there are expected to be relatively few viable cysts (e.g. where infestation is relatively recent) laboratory analysis of a 400 g subsample is recommended. As thorough mixing of soil can only be realistically achieved with dry soils, subsampling below 400 g should be avoided where this is not possible.

For both detection and population estimation, existing literature suggests that the depth at which the core is taken is not important, providing that it is within the top 25 cm. Sampling within the top 5 to 10cm often occurs in British fields and is acceptable. Similarly, for both purposes, sampling on a rectangular grid pattern is preferred. However, for a relatively small increase in error, using a W-pattern can reduce the time spent sampling, with a 4 armed W-pattern may be a sensible alternative. The best time to sample for population estimation is close to planting. However, if the aim is the detection of PCN then this can be carried out following the harvest of the previous crop. Spatial sampling recommendations have been based on the distributions found in the Dutch studies of Been & Schomaker. Insufficient spatial data are available from British fields for drawing any conclusions on recommended sampling patterns based on data from British data, particularly from ware potato production.

# 6. CONCLUSIONS

We conclude that there are two purposes for soil testing for PCN – detection of PCN and estimation of the PCN population present. There are two key components of soil testing, sampling in the field and diagnostic work in the laboratory.

**In the case of sampling for detection**, the aim is to maximise the chance of finding a low population of cysts. We assume that this is more likely to be a new infestation and highly aggregated, and we have used the Been & Schomaker model of developing foci to describe the likely distribution of cysts through the field. Modelling sampling strategies onto such distributions, shows us that sampling using 100 cores per ha and taking larger cores to ensure a large soil sample is available for laboratory analysis is the optimum approach. Sampling on a grid pattern is the most effective way, but using a 4-arm 'W' pattern can be nearly as effective, if correctly orientated with respect to the direction of cultivation. Any subsampling in the laboratory will reduce the probability of detection. These conclusions support the method encapsulated within the EU PCN Directive.

**In the case of sampling for estimating PCN populations**, we have examined a range of data on PCN distributions within British ware potato fields. This information indicates that such PCN distributions can be described using a zero-inflated negative binomial distribution. Available data also suggest that such infestations will be relatively long-standing and that a variable proportion of viable and dead cysts will be present in the field. If the reason for sampling is to determine the level of the infestation and allow for management of the infestation, then an egg count per volume or weight of soil should be assessed. The data also suggests that in soils with the highest levels of eggs, distributions become less heterogeneous. Based on this distribution, and adjusting the heterogeneity index according to population levels, we conclude that samples should be drawn at a rate of a minimum of 50 cores/ha (the greater the number of cores taken the higher the level of accuracy of the estimate). If samples are subsampled within the laboratory, a representative subsample of at least 200g should be analysed. A subsample of 400g is recommended if the infestation is expected to be at a low population level, especially if the field has also recently been in potato cultivation and therefore the eggs are likely to be highly aggregated within relatively few cysts. Management decisions should be based on PCN population estimates resulting from the testing of soil samples, combined with an understanding of the level of precision of these estimates. Decisions are usually made by comparing soil test results with threshold population levels. However, thresholds are difficult to define and many factors, including soil type, variety and environmental conditions, will have an effect on the appropriate threshold for any specific production system.

The most appropriate method for soil sampling regardless of the reason is to use a grid, however if the cost of this is prohibitive a W-pattern can be used. The depth to which the soil sample is taken is of little importance as the cysts are evenly distributed within the top 20 cm. The smaller an area the field is divided into the more accurate assessment of the foci and distribution of PCN in the field is. If the sampling is for management purposes and the treatment will be the same regardless of the distribution of PCN within the field then there is no reason to sample areas separately. However, if a grower wishes to manage hotspots and have a better understanding of the distribution of PCN in the field then the higher the resolution the better. Quarter hectare blocks appear to give a more accurate distribution than strips although this is difficult to verify.

Areas for consideration for future work

**Obtaining more data on representative field distributions:** The field data on PCN distributions chosen for this study were not selected using any unbiased criteria, we simply obtained data from as many sources and as many fields as were made available to us. As several fields represent trial sites, it is likely that any bias will be towards more highly infected fields. Therefore, it is uncertain how representative of the British PCN situation the data used for this analysis is. To address this further, data could be obtained from more sites and potentially more field work could be carried out to fill in the gaps. These data would allow further analysis of the relationship between heterogeneity and density of PCN populations (see Figure 8).

Towards the end of this project, further data were received which gave information about PCN distributions in a relatively recently introduced field population. Pre-planting and post-harvest egg counts, cyst counts and PCR values were recorded for 1ha strips and for ¼ ha blocks. This showed a situation where we believe that cysts had likely been introduced from a neighbouring field on higher ground and had been spread by water flow through the field with the highest PCN population resulting at the bottom of the slope where run-off water had accumulated. This is unlikely to be a representative distribution, but it does demonstrate an alternative way in which an infestation can develop and spread. Surveillance over a number of infested fields may reveal further information between field distributions and likely mechanisms of introduction.

**Obtaining geostatistical data from British field distributions:** This review and analysis provides recommendations based on the existing available information. The lack of available intensively mapped, geostatistical information from a range of fields prevents any conclusions being drawn on the precise way to divide fields for sampling, and from assessing the implications of some sampling schemes in common use, e.g. a ring of samples taken from around a GPS located quad bike/tractor. Such intensively mapped data would permit more detailed exploration of models for the spatial distribution of PCN. They would allow more detailed evaluation of the general applicability of the distribution described by the Been & Schomaker models. However, the collection of such detailed information from a representative range of fields is an extensive and expensive undertaking which may or may not provide sufficiently consistent information from which to reach clear conclusions on sampling. Furthermore, we feel it is unlikely that any such conclusions would lead to any significant differences in our recommendations for sampling to detect PCN.

Additional information may enhance the sampling procedure and the estimates coming from it. One possibility is to direct or supplement the tests using related covariates, perhaps using satellites or drones (Wyse-Pester *et al*., 2002; Anon.,1997 Anon., 2015).The use of drones and satellite imagery could be used to provide additional information based on images of the damage caused by PCN to ware potato crops. These could be valuable in mapping the shape of patches with high population levels. The downside of this approach is that low level infestations are unlikely to produce visible symptoms and any damage seen is likely to be causing yield loss and has the potential to leave even more damaging populations of PCN. As potato crops are affected by a wide range of pests and diseases, it should also not be automatically assumed that the damage has been caused by cyst nematodes. Therefore a limited amount of confirmatory soil sampling from affected areas should be carried out to support any such remote sensing approaches.

**Cost/benefit analysis of PCN testing:** A financial analysis of different testing options would complement this project. Factors that should be considered is cost of time to walk different sampling patterns across a field, different laboratories will charge different amounts for testing soil samples.

**Subsampling:** Further work could be done to define a protocol for the effective subsampling of soil samples for PCN detection or quantification. This might involve seeding samples in order to determine how samples can be effectively mixed and subsampled taking account of factors such as soil type and soil moisture. One issue that should be addressed is the potential for PCN cysts to aggregate/separate out as a dry soil sample is mixed. Along with this, samples could be taken from fields known to harbour moderate PCN infestations to determine how within a soil sample the cysts are aggregated and how to ensure uniformity between subsamples. Within this there is also scope to investigate the relationship between eggs per gram, eggs per cyst and cysts per kilogram to determine the risks of subsampling where cysts are highly viable. Obtaining truly representative subsamples is critical to the effectiveness of the sampling guidance outlined here and the development of a standard protocol for laboratories to use for subsampling could be invaluable.

**Estimation of the likely aggregation of eggs within cysts:** As this is a key factor in determining the recommended volume for any laboratory subsampling, a method to estimate the likely numbers of eggs/cyst would be valuable. This is likely to depend upon decline rates and how they likely to vary between factors such as species and soil type.

# 7. REFERENCES


Anon. (1997) Will remote mapping help pinpoint PCN? Potato Review 7: 9-10.

Anon. (2015) New weapon in the fight against PCN? Potato Review 25: 16-17

Been TH, Schomaker CH (2000) Development and evaluation of sampling methods for fields with infestation foci of potato cyst nematodes (*Globodera rostochiensis and G-pallida*). Phytopathology 90, 647-656.

Been TH and Schomaker CH (2006) Distribution patterns and sampling. In: Perry, R.N. and Moens, M. (eds.) Plant Nematology. CABI Publishing, Wallingford.

Been TH, Schomaker CH (2013) Distribution patterns and sampling. In: Perry RN & Moens M (eds.) Plant nematology 2nd Edition. CABI, Wallingford, pp 331-358.

Been T H, Schomaker CH, Molendijk LPG (2007) NemaDecide, a decision support system for the management of potato cyst nematodes. Phytopathology 97, 152-152

Boag B, Neilson R (1994) Nematode aggregation and its effect on sampling strategies. Aspects of Applied Biology 37: 103-111.

Campbell CL, Noe JP (1985) The spatial analysis of soilborne pathogens and root diseases. Annual Review of Phytopathology 23: 129-148.

Cole CS, Howard HW (1962) The effect of growing resistant potatoes on a potato root eelworm population – a micro plot experiment. Annals of Applied Biology 50, 121-127.

Denwood MJ, Stear MJ, Mathews L, Reid SWJ, Toft N, Innocent GT (2008) The distribution of the pathogenic nematode *Nematodirus battus* in lambs is zero-inflated. Parasitology 135, 1225-1235.

Devine KJ, Dunne C, O'Gara F, Jones PW (1999) The influence of in-egg mortality and spontaneous hatching on the decline of *Globodera rostochiensis* during crop rotation in the absence of the host plant in the field. Nematology 1 (6), 637-645.
Diggle PJ (2003) Statistical analysis of spatial point patterns. Arnold, London.
EPPO (2006) Phytosanitary proceedures: Testing of potato varieties to assess resistance to *Globodera rostochiensis* and *Globodera pallida*. EPPO bulletin 36, 419-420.
Evans KA (2015) Potato cysts nematode and risk of multiple generations in response to climate change. Aspects of Applied Biology 130, 131-136.
Evans K and Kerry B (2007). Changing priorities in the management of potato cyst nematodes. Outlooks on Pest Management 18, 6, 265-269.
Evans K, Webster R, Barker A, Halford P, Russell M, Stafford J, Griffin S (2003) Mapping infestations of potato cysts nematodes and the potential for spatially varying application of nematicides. Precision Agriculture 4, 149-162.
Gilligan CA (1988) Analysis of the spatial pattern of soilborne pathogens. In: Kranz J, Rotem J (eds.) Experimental techniques in plant disease epidemiology. Springer, Berlin, pp85-98.
Haydock PPJ, Evans K (1994) Sampling soil for decision-making potato cyst nematode management. Aspects of Applied Biology 37 113-120.
Hockland, S., Back, M., Blok, V., Davie, K., Grove, I., Jones, J. and Pickup, J. (2016). AHDB Potatoes Research Project: Provision of Content for a PCN Management Guide (No: 1240001)
Kaczmarek A, MacKenzie K, Kettle H, Blok VC (2014) Influence of soil temperature on *Globodera rosthochiensis* and *Globodera pallida*. Phytopathologia Mediterranea 53, 3, 396-405.
Kerry B, Barker A, Evans K (2003) Investigation of potato cyst nematode control. Report to Defra for project no HH3111TPO.
Lane A, Trudgill D (1999) Potato cyst nematode: a management guide. MAFF.
Lin CS, Poushinsk G, Mauer M (1979) An examination of five sampling methods under random and clustered disease distributions using simulation. Canadian Journal of Plant Science 59, 121-130.
Madden LV, Hughes G, van den Bosch F (2007) The Study of Plant Disease Epidemics. APS Press, St. Paul, MN.
McSorley R, Parrado JL (1982) Estimating relative error in nematode numbers from single soil samples composed of multiple cores. Journal of Nematology 14, 522-529.
Minnis ST, Haydock PPJ, Ibrahim SK, Gove IG, Evans K and Russell MD (2002) Potato cyst nematodes in England and Wales – occurrence and distribution. Annals of Applied Biology 140, 2, 187-195.
Nicot PC, Rouse DI, Yandell BS (1984) Comparison of statistical methods for studying spatial patterns of soilborne plant pathogens in the field. Phytopathology 74, 1399-1402.
Noe JP Campbell CL (1985) Spatial pattern analysis of plant-parasitic nematodes. Journal of Nematology 17: 86-93.
Oostenbrink M (1950) Het aardappelaaltje ( Heterodera rostochiensis Wollenwebber) een gevaarlijke parasite voor de eenzijdige aardappelcultuur. Verslagen en Mededelingen van de Plantenziektekundige Dienst te Wageningen 115, 230.
Pickup J, Cole Y, Reid A (2015) Implementation of the 2007 PCN directive and its impact on PCN control in Scotland. Aspects of Applied Biology 130, 29-36.
R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.).
Schomaker CH, Been TH (1999) A model for infestation foci of potato cyst nematodes *Globodera rostochiensis* and *G-pallida*. Phytopathology 89,583-590.

Trudgill DL, Phillips MS and Elliot MJ (2014). Dynamics and management of the whole potato cyst nematode *Globodera pallida* in commercial potato crops. Annals of Applied Biology 164, 1, 18-34.
Turner SJ (1993) Soil sampling to detect potato cyst-nematodes (*Globodera* spp.). Annals of Applied Biology 123: 349-357
Turner SJ, Flemming CC (2002) Multiple selection of potato cyst *nematode Globodera pallida* virulence on a range of potato species. 1. Serial selection on a range of potato hybrids. European Journal of Plant Pathology, 108, 461-446.
Van den Berg W, Rossing WAH, Grasman J (2006) Contest and scramble competition and the carry-over effect in *Globodera* spp. in potato-based crop rotations using an extended Ricker model. Journal of Nematology 38, 210-220.
Whitehead A and Turner S (1998) Management and regulatory control strategies for potato cyst nematodes (*Globodera rostochiensis* and *Globodera pallida*). In: Potato Cyst Nematodes. Eds.RJ. Marks and B.Brodie. CAB International.
Winfield AL (1965) Potato root eelworm in Holland, Lincolnshire, NAAS Q. Rev., 67, 110-117.
Wyse-Pester DY, Wiles LJ, Westra P (2002) The potential for mapping nematode distributions for site-specific management. Journal of Nematology 34, 80-87

## ACKNOWLEDGEMENTS

We thank John Keer (Richard Austin Agriculture), Nick Winmill (Agrii), Eric Anderson (Scottish Agronomy), Andy Barker (Barworth) and Emma Keltcher for contributions to the report.