Variability components for macrophyte communities in rivers
Appendix C: River Allen Analysis

Report: SC070051/P2

Integrated catchment science programme

Evidence Directorate

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency.

Published by:
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol, BS32 4UD
Tel: 01454 624400 Fax: 01454 624409
www.environment-agency.gov.uk

ISBN: 978-1-84911-141-6

Further copies of this summary are available from our publications catalogue: http://publications.environment-agency.gov.uk or our National Customer Contact Centre: T: 08708 506506
E: enquiries@environment-agency.gov.uk.

Author(s):
Andrew Davey
David Garrow

Dissemination Status:
Released to all regions
Publicly available

Keywords:
Macrophyte, Ecological Quality Ratio, spatial variability, temporal variability, measurement error, uncertainty, precision

Research Contractor:
WRc plc
Frankland Road
Blagrove
Swindon
SN5 8YF

Environment Agency's Project Manager:
Veronique Adriaenssens, Evidence Directorate

Project Number:
SC070051/P2

Product Code:
SCHO1109BRHM-E-P

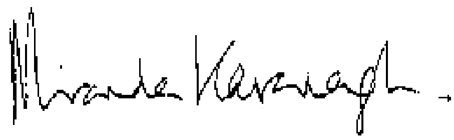
Evidence at the Environment Agency

Evidence underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us, helps us to develop tools and techniques to monitor and manage our environment as efficiently and effectively as possible. It also helps us to understand how the environment is change and to identify what the future pressures may be.

The work of the Environment Agency's Evidence Directorate is a key ingredient in the partnership between research, policy and operations that enables the Environment Agency to protect and restore our environment.

The Research & Innovation programme focuses on four main areas of activity:

- **Setting the agenda**, by informing our evidence-based policies, advisory and regulatory roles;
- **Maintaining scientific credibility**, by ensuring that our programmes and projects are fit for purpose and executed according to international standards;
- **Carrying out research**, either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available to our policy and operations staff.



Miranda Kavanagh

Director of Evidence

Executive summary

To complement a nationwide analysis of the components of variability in riverine macrophyte communities, the Environment Agency commissioned WRc to analyse the results of an intensive macrophyte monitoring programme on the River Allen, a chalk stream in Dorset.

The aim of the study was two-fold:

1. To examine the natural spatial and temporal components of variation in the macrophyte community along a 20-km stretch of the River Allen from 1998 to 2008.
2. To compare by simulation the precision of different sampling schemes, focusing on the number of surveys and the sampling strategy.

The analysis focused on three parameters: total macrophyte cover and cover of the two most abundant taxa, *Ranunculus* and *Scirpus/Sparganium*.

A geostatistical approach was used to model spatial variation in macrophyte communities as a continuous function of distance between sampling points. Estimates were made of the maximum variation between sites (the 'sill'), the distance or lag at which this occurs (the 'range'), and the level of within-site variation (the 'nugget'). Thus, it was possible to measure the total spatial variation and to establish the spatial scales at which this variation occurred. Models were fitted for each year and then compared to determine how the pattern and magnitude of spatial variability changes from year to year.

To compare the precision of different sampling schemes, simulations were run to assess the variability in monitoring results produced using different numbers of surveys and different sampling strategies.

The main findings were:

- Spatial variation in cover values between stretches increases with distance between stretches and gradually plateaus off at a level corresponding to the maximum spatial variance. This pattern was observed in all cases except for total cover in 2007.
- Nearly all the models showed high variability at a spatial scale smaller than that measured (100-m stretches). This could reflect spatial variation within each 100-m sampling unit and/or measurement error (operator variability). This variation often contributed around half of the total spatial variation, indicating that macrophyte communities show high variation at very small spatial scales.
- Most of the spatial variation in *Ranunculus* cover occurs over small spatial scales (less than three km). Therefore, surveying a few localised stretches of a chalk stream will be adequate to estimate the true spatial variation in this taxon. By contrast, *Scirpus/Sparganium* and total cover show increasing spatial variation up to around 15-km and widely-spaced surveys will be required to adequately characterise spatial variability of these parameters.
- Patterns and magnitudes of spatial variation exhibit high levels of inter-annual variation for total cover and *Ranunculus.* By contrast, spatial variation in *Scirpus/Sparganium* is relatively constant from year to year.

Surveys for this taxon can therefore be undertaken at longer intervals than for the other parameters analysed.

- The precision of estimates of mean cover improve with the number of surveys undertaken. The law of diminishing returns applies, however; five to seven surveys will give reasonable precision, and increasing the number of samples further will deliver only minor benefits. The number of surveys required to achieve a given level of precision may be influenced by the length of the water body; all else being equal, smaller water bodies should display less spatial variability, and require fewer surveys.
- Regular spacing of surveys along the water body gives better precision than other sampling strategies. However, there is little difference between this strategy and random or stratified random sampling. Conducting a series of contiguous surveys gives the worst precision for a given sample size. Regular sampling is therefore the most efficient method of estimating ecological status within a water body, particularly when there are large-scale spatial patterns in macrophytes community structure.

Contents

List of tables

List of Figures

1 Introduction

To complement a nationwide analysis of the components of variability in riverine macrophyte[1] communities (Project SC070051), the Environment Agency commissioned WRc to analyse the results of an intensive macrophyte monitoring programme on the River Allen, a chalk stream in Dorset.

The aim of the study was two-fold:

1. To examine the natural spatial and temporal components of variation in the macrophyte community along a 20-km stretch of the River Allen from 1998 to 2008.
2. To compare by simulation the precision of different sampling schemes, focusing on the number of surveys and the sampling strategy.

The analysis focused on three parameters: total macrophyte cover and cover of the two most abundant taxa, *Ranunculus* and *Scirpus/Sparganium*.

The remainder of this report is divided into five sections. Section 2 provides background to the project; Section 3 describes the monitoring data; Section 4 describes the results of the geostatistical modelling of variability; Section 5 presents the results of the simulation studies; and Section 6 draws some general conclusions from these analyses.

[1] Macrophytes are larger plants of freshwater which are easily seen with the naked eye, including all aquatic vascular plants, bryophytes, stoneworts (Characeae) and macro-algal growths.

2 The River Allen Survey

Macrophyte communities along a 20-km length of the River Allen, a chalk stream in Dorset, have been surveyed twice a year (May and July) since 1989, with the exception of 2001 when foot and mouth disease prevented access to the river.

The Environment Agency is not aware of any significant discharges along the river that might affect the macrophyte community. There are some water cress farms but these do not appear to have any major effect and are not expected to affect the results of the analysis.

Land use within the catchment has altered during the course of the survey. Increasing arable farming may have altered drainage patterns but there is no evidence from macroinvertebrate monitoring that this has had an effect on the ecology of the river. Therefore, it is assumed that changes in land use will have no noticeable impact on the results of this analysis.

Limited weed-cutting occurs within the river. This is carried out mainly in June and July so should not affect the cover values recorded in May. Also, the weed-cutting leaves a root shadow that can be recorded by the surveyor. This river management is therefore considered to have no noticeable impact upon the analysis results.

On each survey occasion, the river was divided into 212 contiguous 100-m long stretches of river. The macrophyte community in each stretch was surveyed and recorded. The 212 survey sites span three Water Framework Directive (WFD) water bodies (stretches 1-115 are in GB108043011090, stretches 116-157 are in GB108043015710, and stretches 158-212 are in GB108043015790).

Quantitative estimates of the percentage cover of the dominant instream macrophyte taxa were made from the bank; all measurements were made using a standard protocol, mainly by the same operator. The taxa recorded are listed in the table below.

Table 2.1 List of taxa recorded.

Taxa Recorded
Ranuculus
Scirpus/Sparganium
Callitriche
Apium/Berula
Nuphar
Potamogeton perfoliatus
Oenanthe fluviatilis
Elodea
Rorippa
Potamogeton crispus

3 Data

Data was supplied by the Environment Agency for the period 1998-2008 (no data was supplied for 1989-1997 because this was recorded in a different format) and for May only (the July data is still in a hardcopy format). The data is composed of percentage cover values for each taxon in each of the 212 stretches. During the early surveys, difficulty was experienced in separating *Scirpus* from *Sparganium emersum*. Consequently an additional column of data was created combining the percentage cover of both these species, as it is believed that *Scirpus* is the dominant ribbon weed in the river. Rare taxa were originally recorded as being either present or absent; where a taxon was present, its abundance was subsequently estimated to one per cent, and the total macrophyte cover was computed as the sum of the cover values for the component taxa.

Access to certain sections of the river was not possible (either in specific years or over the whole study period). These sections were excluded from the analysis and are listed in the table below.

Table 3.1 Unavailable sections of the river.

Unavailable Stretches	Years Unavailable
96 to 98	All
144 to 148	All
178 to 192	All
207 to 208	All
36 to 40	2000
41 to 43	2003
41 to 46	2005
41 to 48	2004
193 to 196	2006
210	2003
212	2003
205 to 212	1998
211 to 212	2007
210 to 212	2002, 2004, 2006

The River Allen dataset is exceptional in that it comprises high resolution spatial and temporal measurements of macrophyte communities without the confounding effect of surveys being conducted in different months and by different operators. It is ideally suited for estimating random year-to-year temporal variation and random spatial variation in macrophyte communities.

4 Analysis of variability

4.1 Methodology

The traditional approach to investigating spatial patterns in ecological communities involves splitting up the survey area into progressively smaller units. For example, a river might be divided into three water bodies, each water body might then be divided into a number of reaches of defined length (say, three km), and each reach would then be divided into a number of sites (whose length depends upon the survey method – typically 100 or 500 m). Using a nested ANOVA model, it is then possible to estimate the contribution of each hierarchical level to the overall variance in the data. This was the approach taken in the main macrophyte variability project.

One of the unique features of the River Allen dataset is that surveys are conducted at contiguous 100-m long sites. Dividing up the river into water bodies and reaches within water bodies would mean arbitrarily splitting up what is in reality a continuously varying macrophyte community. A geostatistical approach was therefore used to model spatial variation in macrophyte communities as a continuous function of distance between sampling points. This process involves calculating the variance[2] between pairs of surveys at all possible distances apart (lags), and then plotting the observed average variance against lag. The resulting plot, called a variogram, shows the pattern of spatial dependence between sampling units. Variograms show how average variance between pairs of surveys increases with the distance between the survey sites and typically take the form shown in Figure 4.1.

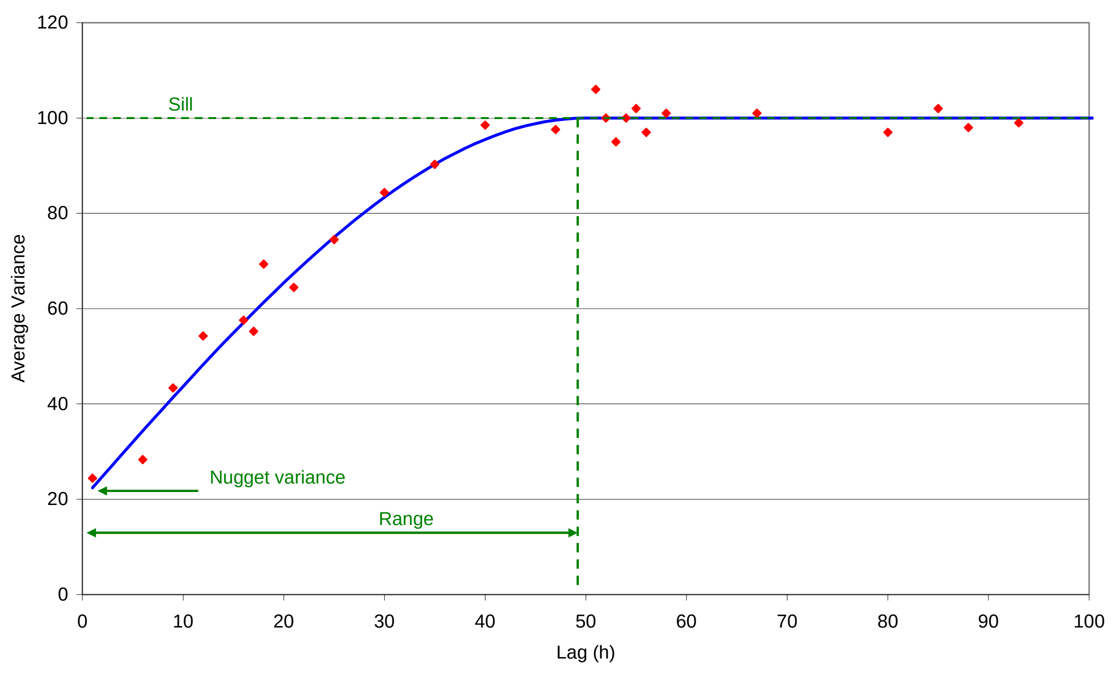


Figure 4.1 Typical variogram.

[2] In statistical literature often termed the 'semi-variance', even though it is strictly a variance.

The variogram contains three useful pieces of information:

1. The lag at which the curve levels off (called the **range**) indicates the distance at which there is no longer any spatial dependence between sites (spatial autocorrelation is zero).
2. The height at which the curve levels off is called the **sill** and it estimates the maximum spatial variance, taking into account variance at all spatial scales under investigation.
3. In theory the variance at the origin (0 lag) should be zero; if it is significantly different from zero for lags very close to zero, then this value is referred to as the **nugget** variance. The nugget variance represents variability at distances smaller than the typical survey spacing, including measurement error.

The survey data was analysed using variograms. A variogram plot for each year of surveying was produced. This was done for the total cover (the combined coverage of all taxa) and for the two most dominant taxa (*Ranunculus* and *Scirpus/Sparganium*). The chart below shows that the surveys are dominated by these two taxa in 2006; a similar pattern is repeated across all years.

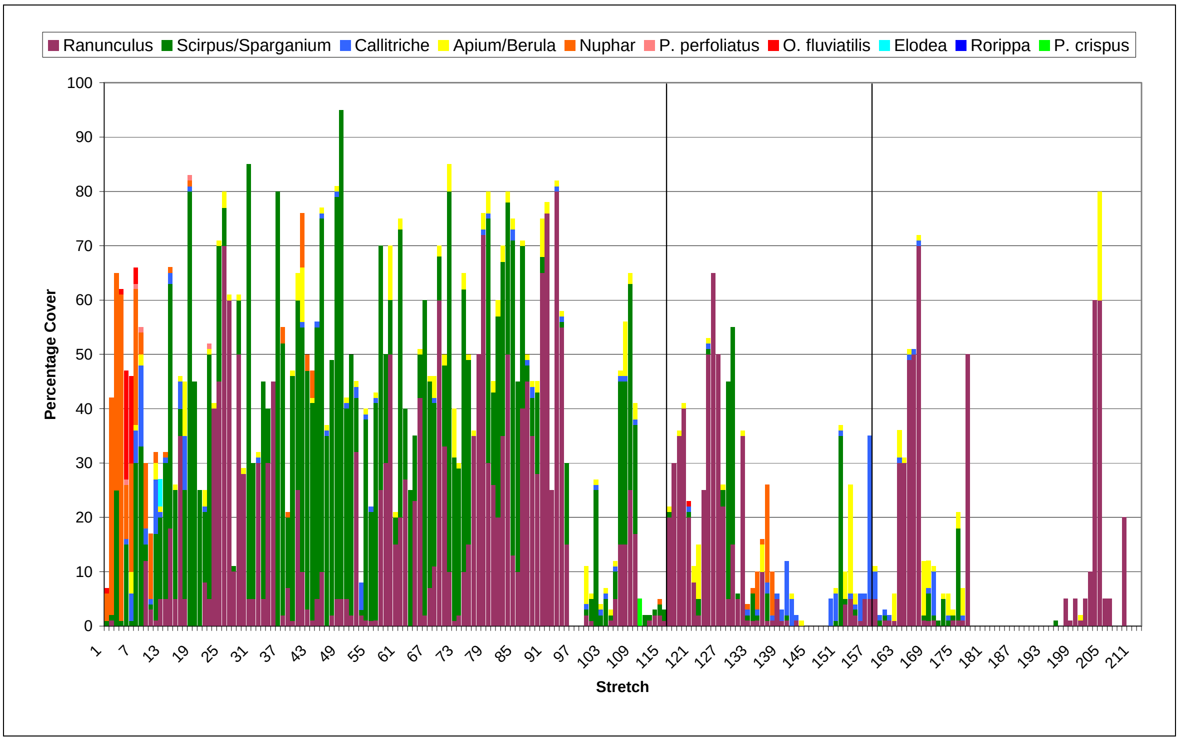


Figure 4.2 Percentage cover values by taxa for 2006.

Models were then fitted to the plots to determine the sill, range and nugget variance for each variogram. It is assumed that the variance values will level off after a certain distance (the range). This levelling off is the sill and models which include a sill are termed transition models. There are three commonly used transition models:

1. Spherical Model

$$\gamma(h) = \begin{cases} c.\left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) + C_0 & \text{if } h \le a \\ c + C_0 & \text{if } h > a \end{cases}$$

Where h is the lag (in number of stretches), $\gamma(h)$ is the average variance for surveys at lag h, a is the range, C_0 is the nugget variance and c is the sill minus the nugget variance. This model reaches the sill at the specified range.

In contrast to the spherical model, the next two models approach the sill asymptotically:

2. Exponential Model

$$\gamma(h) = 1 - \exp\left(-\frac{3h}{a}\right) + C_0$$

This model approaches the sill asymptotically and the range *a* is defined as the lag at which the variogram is at 95 per cent of the sill value. This is termed the practical range.

3. Gaussian Model:

$$\gamma(h) = 1 - \exp\left(-\frac{3h^2}{a^2}\right) + C_0$$

This model also approaches the sill asymptotically and *a* is the practical range where the variogram equals 95 per cent of the sill value.

The Gaussian model has a parabolic nature near the origin, giving an S-shaped curve, and can represent properties that vary smoothly over shorter ranges.

These three models were fitted to each variogram, with the model providing the best fit preferred for each taxon. Figure 4.3 shows the standard pattern that each model takes.

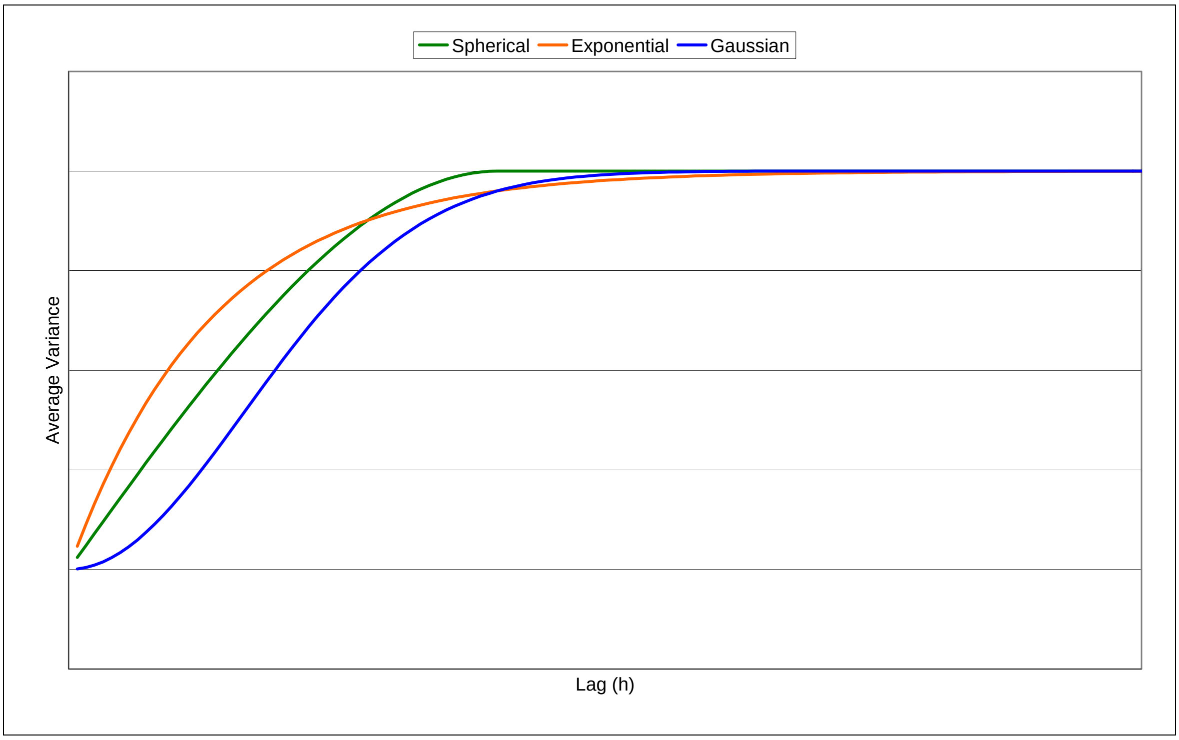


Figure 4.3 Example of variogram models.

4.2 Total cover

Variograms were produced for total cover each year of surveying. These variograms typically fitted the Gaussian model best, as illustrated by the brown line in Figure 4.4. Note that the high scatter and noticeable zig-zag pattern in the data on the right hand side of the graph are an artefact of the data; variances at high lags are based on only a small number of paired surveys, and should not be interpreted as indicating a quasi-cyclic spatial pattern.

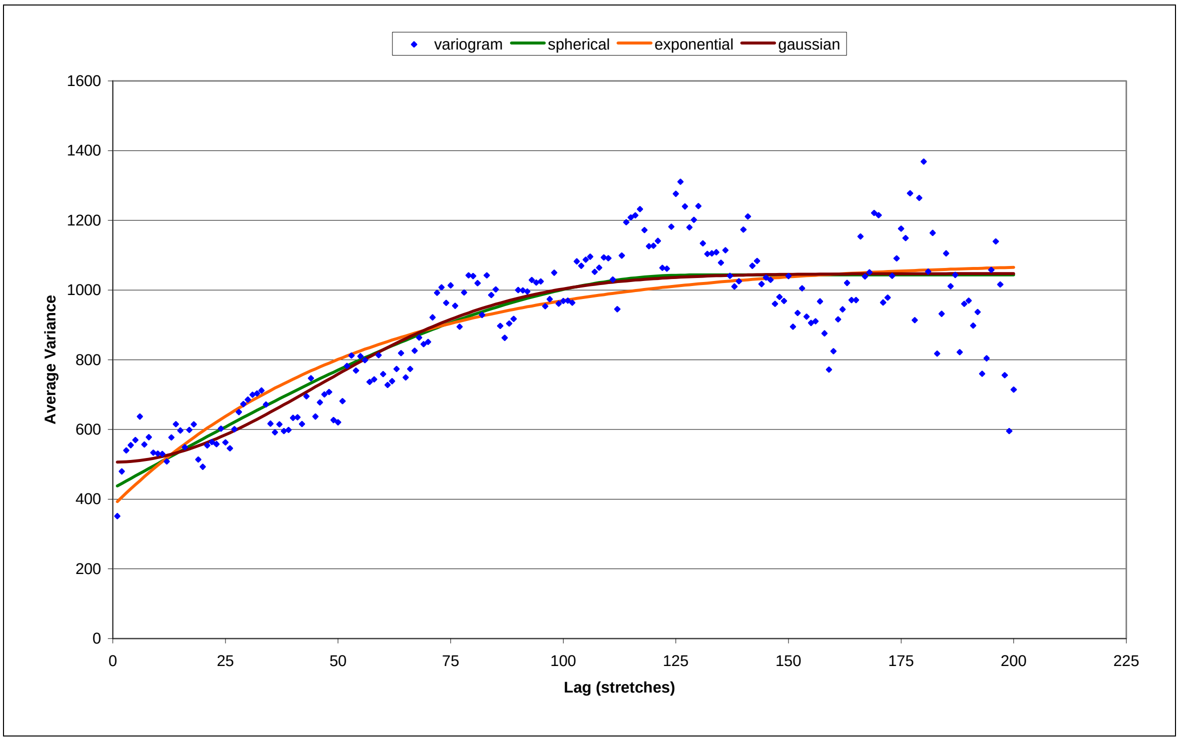


Figure 4.4 Example of typical variogram for total cover – 1999.

The modelled sill, range and nugget values for total cover are shown in Table 4.1.

Table 4.1 Modelled sill, range and nugget values for total cover.

Year	Sill variance	Range (no. of 100-m stretches)	Nugget variance (as % of sill)
1998	786.9	53.7	521.1 (66.2%)
1999	1046.7	109.1	506.0 (48.3%)
2000	1253.9	105.8	644.1 (51.4%)
2002[1]	526.9	140.0	216.9 (41.2%)
2003	977.8	63.9	695.3 (71.1%)
2004	935.9	195.7	501.4 (53.6%)
2005	672.5	192.1	540.6 (80.4%)
2006	848.3	116.6	413.8 (48.8%)
2007[2]	177124.0	14384.8	584.1 (0.3%)
2008	667.9	111.6	507.4 (76.0%)

[1]The model for 2002 was constrained to ensure a realistic range was produced.
[2]A satisfactory model could not be produced for 2007

The fitted Gaussian models for total cover are shown in Figure 4.5. There are three main points to note.

First, all years have a nugget variance that is significantly greater than zero and, in some cases, a large proportion of the total variance indicated by the sill (Table 4.1). This shows that there is a very high level of variation at a spatial scale smaller than the scale of measurement. This could reflect spatial variation within each 100-m sampling unit and/or measurement error (operator variability).

Second, in most years, the variance levelled off at a clearly-defined sill. The average sill value across the years (excluding 2007, for which no satisfactory model could be found) was 857.4, occurring at a typical range of 60 to 140 stretches (6-14 km; Table 4.1). This means that pairs of surveys more than about 10 km apart show no correlation in their results and can be considered to be truly independent replicate measurements of the macrophyte communities in that river or water body. Pairs of surveys less than 10 km apart are correlated to a greater or lesser degree, but even two surveys of adjacent 100-m stretches show high variability because of the high nugget variance.

Third, there is high variation in the magnitude and pattern of spatial variation from year to year. This represents a spatio-temporal variation in macrophyte communities; in other words, the pattern of spatial variation is itself variable over time. This pattern may be driven at least in part by differences in the amount of macrophyte growth or cover from year to year – that is variances tend to be higher in years with prolific macrophyte growth.

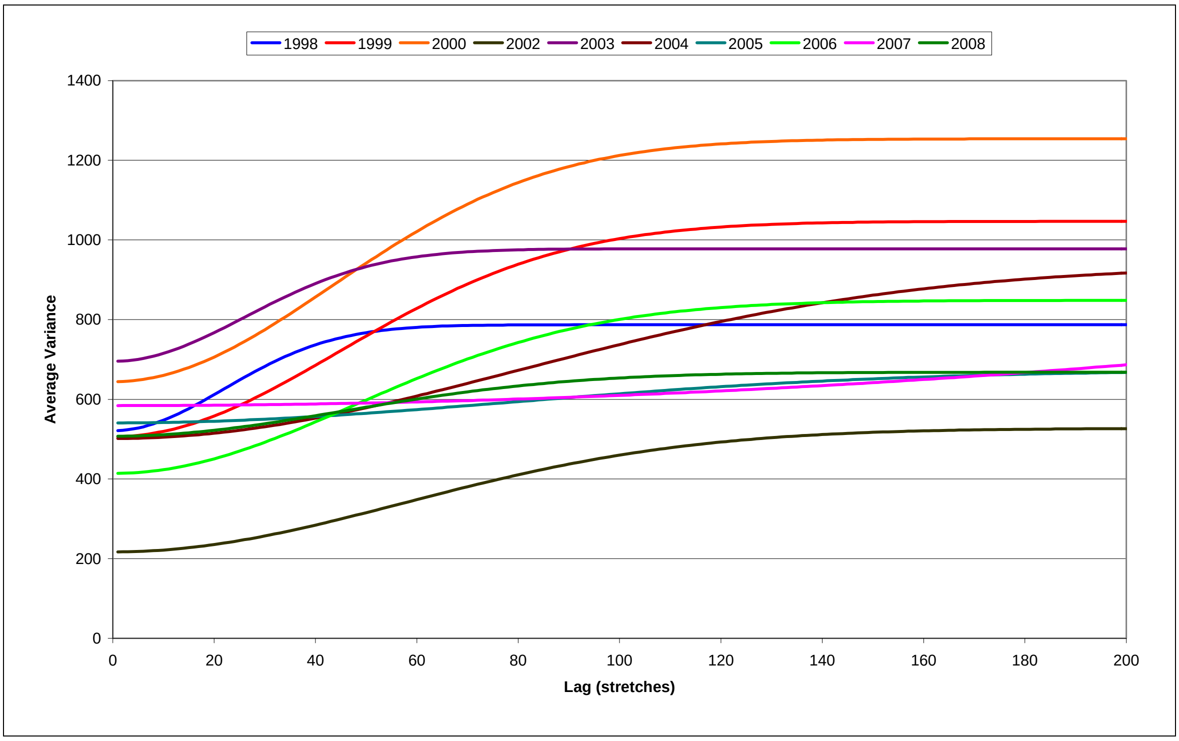


Figure 4.5 Fitted models for total cover.

4.3 *Scirpus/Sparganium* cover

Scirpus/Sparganium was one of the two most commonly occurring taxa. Variograms were produced for each year of surveying for the cover values of this taxon. These variograms typically fitted the spherical model best, as can be seen in Figure 4.6. This figure shows a decreasing variance at high lags (175 stretches and more). However, at this large distance any similarity between cover values is entirely coincidental and so should not be interpreted as a real effect. Also, a lower number of data pairs at large lags means less confidence can be placed in the variances at such distances.

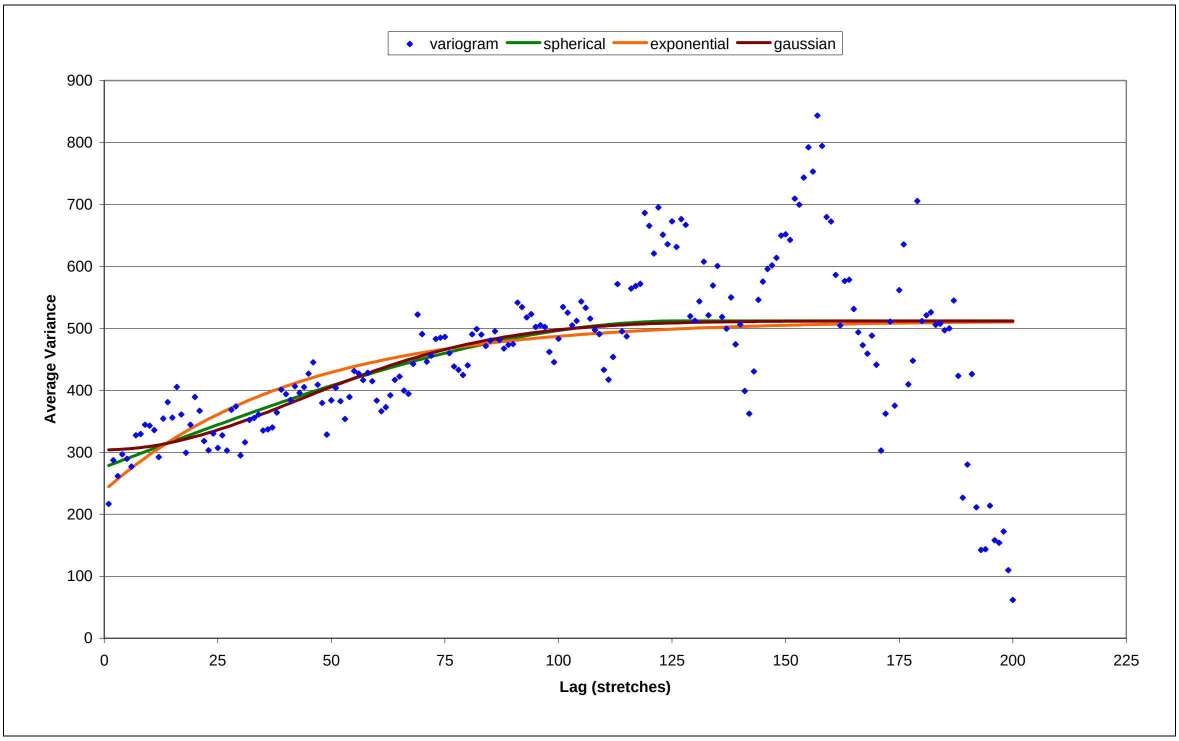


Figure 4.6 Example of typical variogram for *Scirpus/Sparganium* – 2006.

The modelled sill, range and nugget values for *Scirpus/Sparganium* cover are shown in Table 4.2.

Table 4.2 Modelled sill, range and nugget values for *Scirpus/Sparganium* cover.

Year	Sill variance	Range (no. of 100-m stretches)	Nugget variance (as % of sill)
1998	525.0	122.2	285.2 (54.3%)
1999	457.2	153.0	260.4 (56.9%)
2000[1]	576.1	125.0	292.2 (50.7%)
2002	354.0	131.7	221.1 (62.4%)
2003	483.3	164.4	274.4 (56.8%)
2004	681.6	173.7	375.3 (55.1%)
2005	521.6	160.3	334.5 (64.1%)
2006	512.1	127.8	275.9 (53.9%)
2007	498.9	122.6	265.4 (53.2%)
2008	559.7	132.7	298.7 (53.4%)

[1]The model for 2000 was constrained in order to ensure a realistic range was produced.

The fitted spherical models for *Scirpus/Sparganium* cover are shown in Figure 4.7.

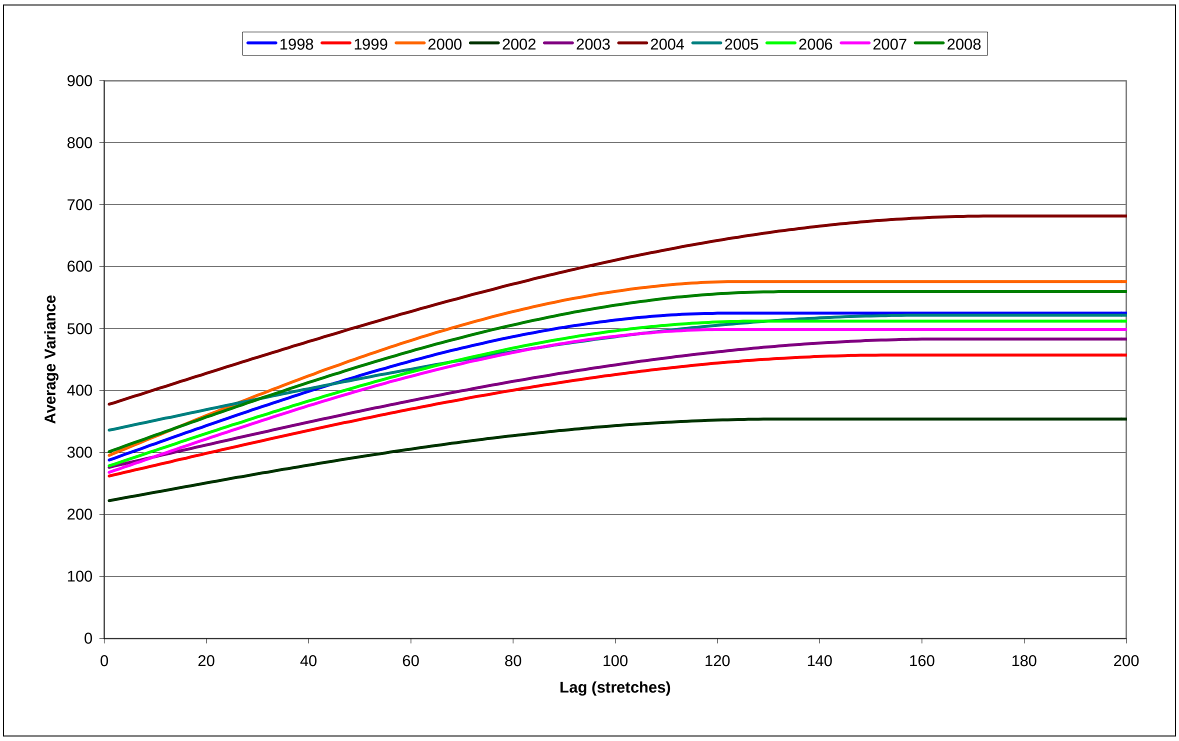


Figure 4.7 Fitted models for *Scirpus/Sparganium* cover.

All years have a significant nugget variance (Figure 4.7). The nugget variance is typically around 50 per cent of the sill variance (Table 4.2), indicating that half of the total spatial variation among stretches is due to variation at a scale of less than100 m.

The average sill value across the years was 517.0, occurring at a typical range of 120 to 170 stretches (12-17 km; Table 4.2). This means that pairs of surveys more than about 15 km apart show no correlation in their results and can be considered to be truly independent replicate measurements of the macrophyte communities in that river or water body. Pairs of surveys less than 15 km apart are correlated to a greater or lesser degree, but even two surveys of adjacent 100-m stretches show high variability because of the high nugget variance.

Compared with total cover, there was relatively little year-to-year variation in the variogram for *Scirpus/Sparganium*, indicating that the abundance and pattern of spatial variation is relatively constant over time for these taxa.

4.4 *Ranunculus* cover

The other commonly occurring taxon was *Ranunculus*. Variograms were produced for each year of surveying for the cover values of this taxon. These variograms typically fitted the spherical model best, as can be seen in Figure 4.8.

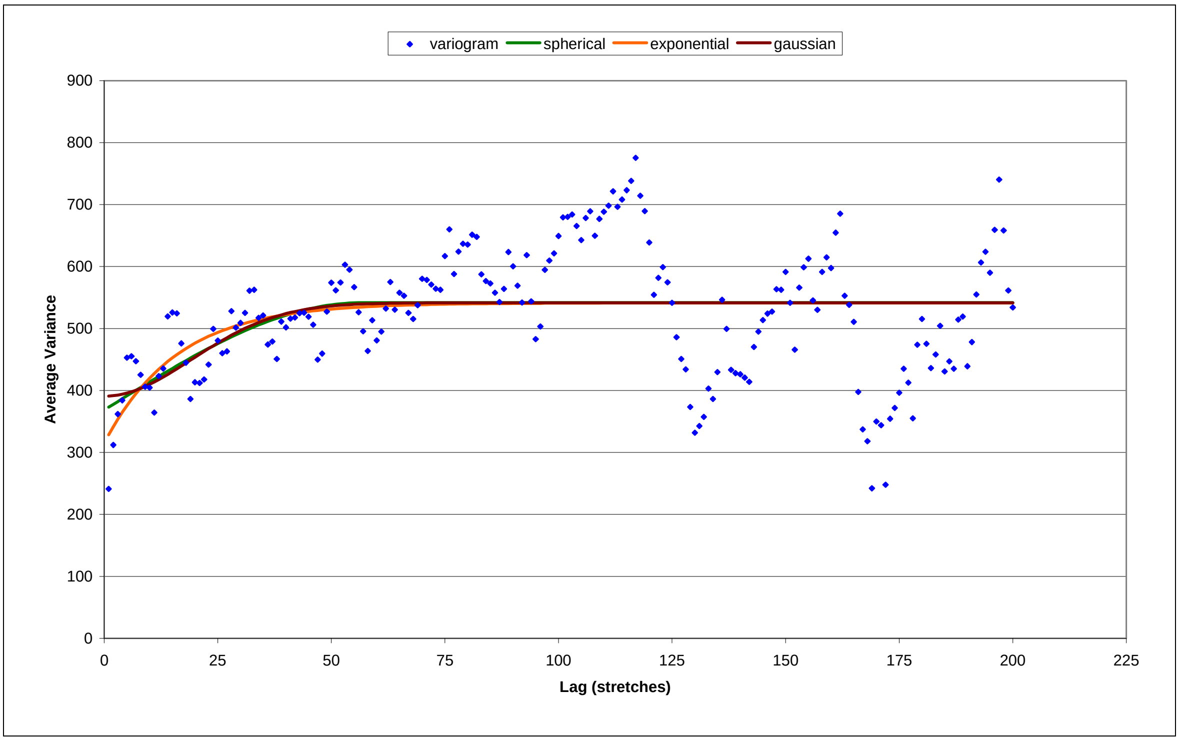


Figure 4.8 Example of typical variogram for *Ranunculus* – 2007.

In certain years the variograms for *Ranunculus* exhibited increasing variances at large lags after a feasible sill value appeared to have been achieved (for example, 2008, as shown in Figure 4.9). This pattern fitted a third-order polynomial model.

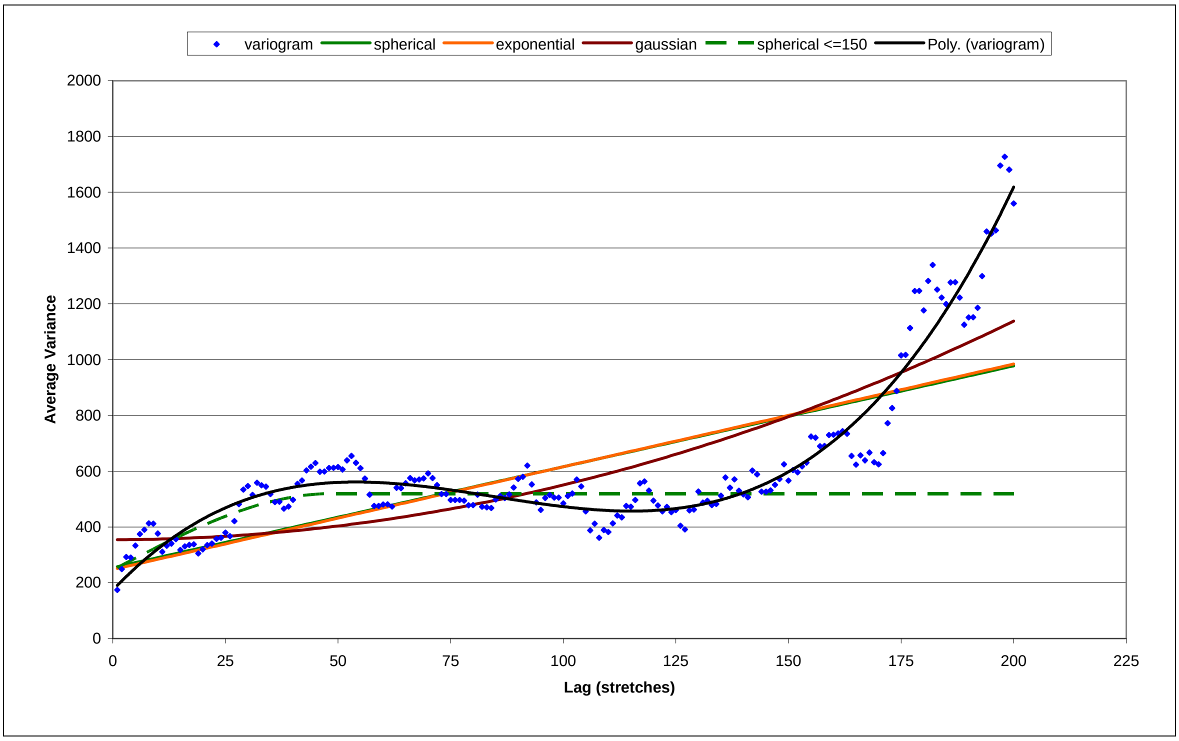


Figure 4.9 Example of polynomial model for *Ranunculus* variogram – 2008.

Although a polynomial model fitted these variograms well, at these large lags any further increase in the variance will be coincidental or the result of specific local conditions and so should not be modelled. A lower number of data pairs at large lags also means that less confidence can be placed in the variances at such distances. Where this issue arose the data was modelled over a shorter lag series to remove any impact of the increase in variance at very large lags. The model fitted using this method is shown by the dashed line in Figure 4.9.

Modelled sill, range and nugget values for *Ranunculus* cover are shown in Table 4.3.

Table 4.3 Modelled sill, range and nugget values for *Ranunculus* cover.

Year	Sill variance	Range (no. of 100-m stretches)	Nugget variance (as % of sill)
1998	554.0	49.8	400.2 (72.2%)
1999	587.3	12.4	362.1 (61.7%)
2000	681.1	6.6	309.4 (45.4%)
2002	26.2	6.6	11.0 (42.1%)
2003	809.1	20.0	543.8 (67.2%)
2004	128.7	2.9	39.3 (30.6%)
2005[1]	336.7	30.4	216.7 (64.4%)
2006	394.5	6.9	142.9 (36.2%)
2007	541.6	56.6	368.3 (68.0%)
2008[1]	519.8	48.7	245.4 (47.2%)

[1]The models for 2005 and 2008 were fitted using a maximum lag of 150 stretches.

The fitted spherical models for *Ranunculus* cover are shown in Figure 4.10.

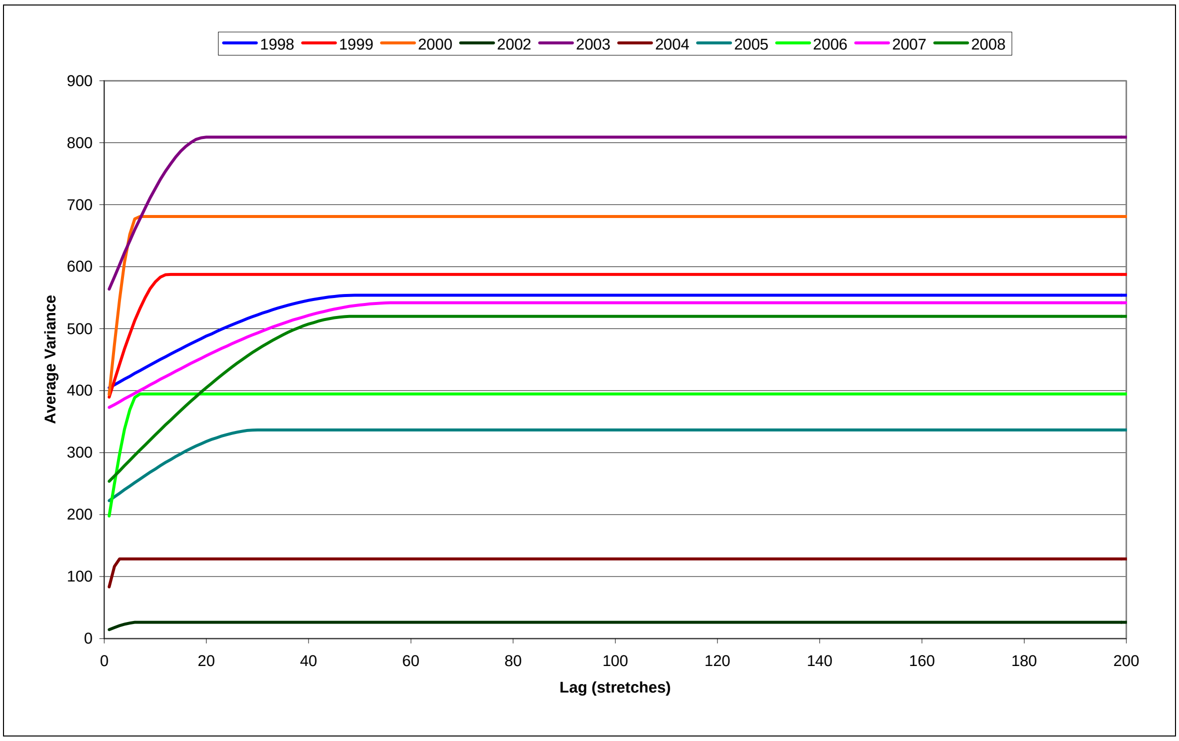


Figure 4.10 Fitted models for *Ranunculus* cover.

With the exception of 2002, all years have a significant nugget variance. This shows that there is a high level of variation at a smaller spatial scale than is measured (within each 100-m stretch), as was also observed for *Scirpus/Sparganium* and total cover values. The nugget variance varies between 30 and 70 per cent of the sill variance (Table 4.3), indicating around half of the total spatial variation among stretches is due to variation at a scale of less than100 m.

The variance values increase steeply from the nugget and typically reach a sill at smaller ranges than for either total cover or *Scirpus/Sparganium*. The average sill value across the years, excluding 2002, is 505.9 and occurs at a typical range of six to 50 stretches (0.6–5 km; Table 4.3). The range values for *Ranunculus* are far lower than for the other response variables measured (total cover and *Scirpus/Sparganium*) showing that there is negligible large-scale variation in *Ranunculus* communities (each three-km reach has an amount of *Ranunculus* very similar to the next three-km reach). This means that pairs of surveys more than about three km apart show no correlation in their results and can be considered to be truly independent replicate measurements of the macrophyte communities in that river or water body. Pairs of surveys less than three km apart are correlated to a greater or lesser degree, but even two surveys of adjacent 100-m stretches can show high variability if the nugget variance is high.

There is a significant variation in the variogram models between the years for *Ranunculus*. In particular, the sill, range and nugget values are far lower in 2002 than in other years. There was far less *Ranunculus* present in this year than normal.

4.5 Variation in models between years

Variograms for *Scirpus/Sparganium* seem to vary least between years. The box-plots of the modelled sills, ranges and nuggets across the years are shown in Figure 4.11 to Figure 4.13.

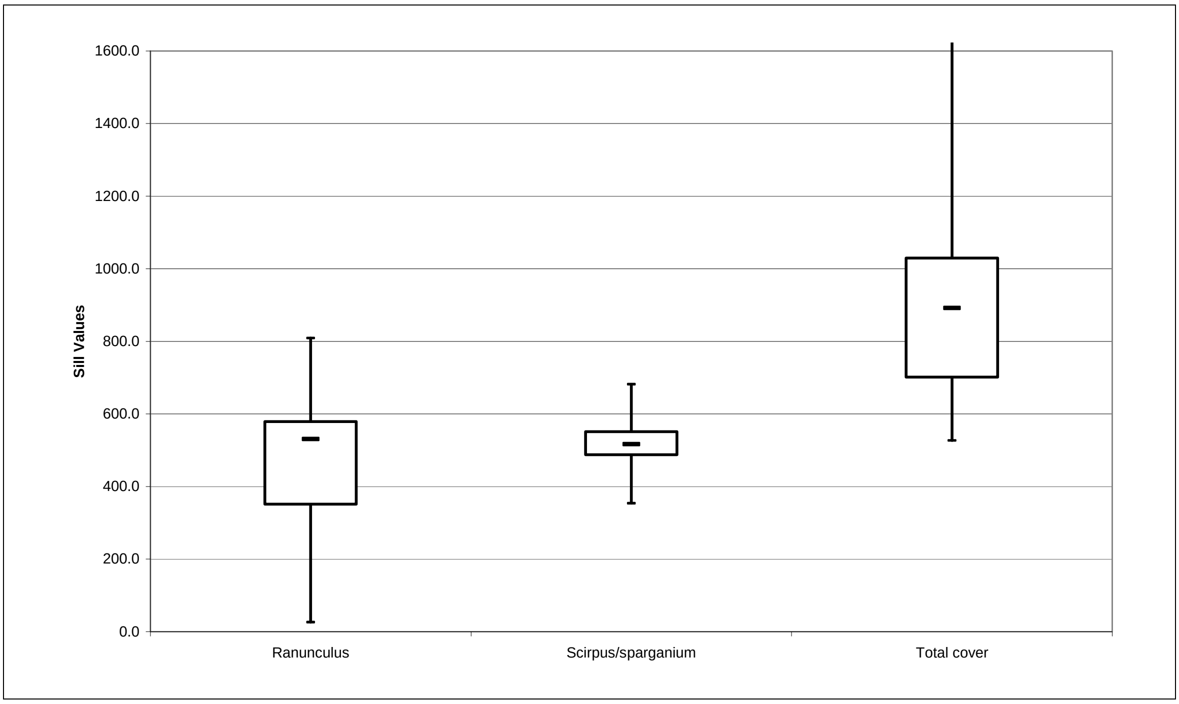


Figure 4.11 Box-plot of modelled sill variances across years.

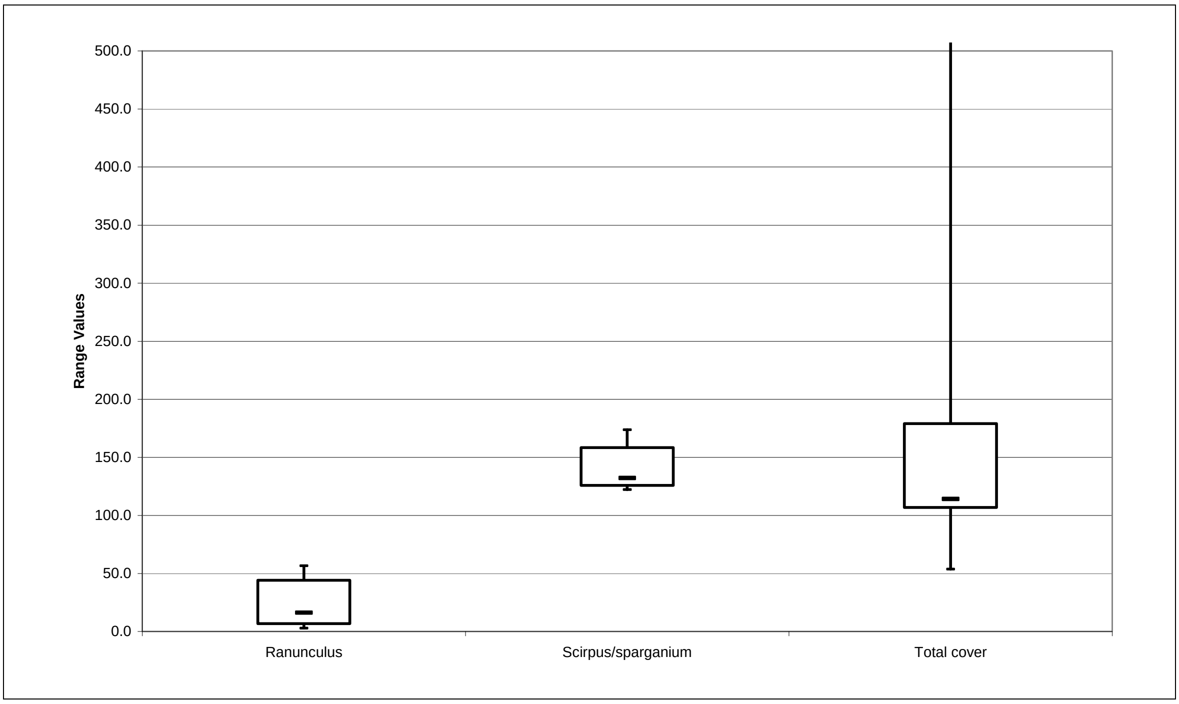


Figure 4.12 Box-plot of modelled range values across years.

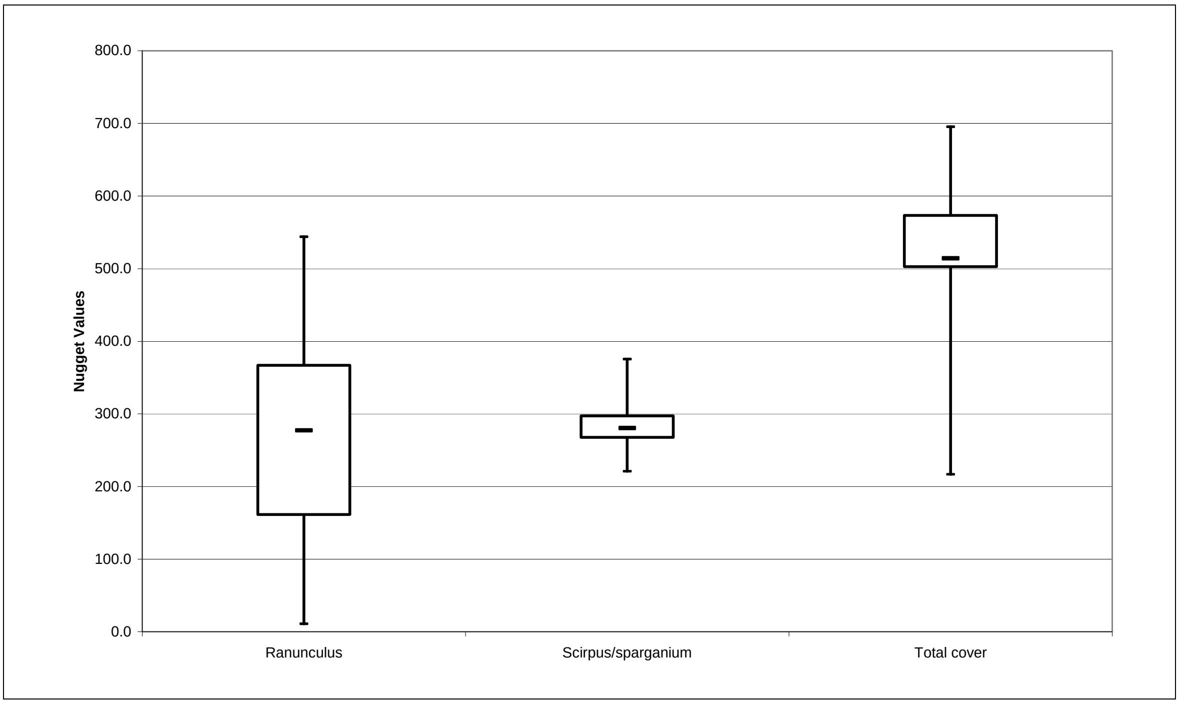


Figure 4.13 Box-plot of modelled nugget variances across years.

The box-plots show that there is least inter-annual variation in the models for *Scirpus/Sparganium*. This suggests that this taxon is less affected by annual changes, such as weather patterns, than *Ranunculus* or the total cover values and may not require as frequent surveying as these taxa.

Total cover varies the most across years for sill and range values but *Ranunculus* exhibits the greatest variation in nugget values. The annual variation in nugget values for *Ranunculus* is greater than the variation in sill and range values. This shows that annual variations affect the small-scale spatial variability of *Ranunculus* more than they affect the larger-scale spatial variability. The variability within a survey stretch will alter more under the impact of annual perturbations than the variability between stretches.

Total cover has a lower annual variation in nugget values than *Ranunculus,* suggesting that the annual variations in one taxon are compensated for by changes in other taxa. The fact that typical nugget and sill variances are higher for total cover than for either of the individual taxa measured is probably due to total cover having higher percentage cover scores than any individual taxon.

5 Comparison of sampling schemes

5.1 Introduction

The high spatial resolution of the River Allen dataset provides an opportunity to investigate the relative efficiency of alternative sampling schemes. Specifically, it is possible to simulate different sampling schemes by selecting sets of 100-m surveys from the dataset, and then to compare the range of results produced using each scheme.

The simulation study sought to answer two main questions:

1. How does the number of 100-m surveys conducted in a water body affect the precision of the results?
2. How does the spatial arrangement of the surveys within the water body affect the precision of the results?

For each question, a number of alternative sampling schemes were each simulated 100 times and the precision of the results produced by the different schemes were compared. The simulations were run using data from the largest of the three water bodies: GB108043011090, which comprises survey stretches 1 to 115 inclusive.

Each simulation was carried out using three separate years worth of data, broadly representing low, medium and high variability years, and for each of the three parameters analysed in Section 4 (Total cover, Ranunculus and Scirpus/Sparganium).

5.2 Number of surveys

How does the number of 100-m surveys conducted in a water body affect the precision of the results?

A simulation exercise was conducted to compare the results that would be produced using different numbers of surveys (*n*, between 1 and 20) located randomly throughout the water body. In each 100 simulations, a set of *n* surveys were selected at random and the cover values recorded in each survey averaged to estimate the mean cover in the water body. For each value of *n*, the mean of the 100 simulation results was determined and plotted, along with the 90 per cent confidence interval as a measure of precision. The results were then compared against the true mean cover calculated from the entire dataset.

The simulation was repeated for each of the three parameters in each of three years.

5.2.1 Total cover results

Figure 4.5 was used to select the three years. The years chosen were not necessarily those with the highest/lowest variability since years that were considered to be ‘extreme’ or that did not fit the appropriate variogram model as well as others were not used. The three years selected to run the total cover simulations were 1998 (medium variability), 2000 (high variability) and 2008 (low variability).

The results for the three years are shown in Figure 5.1 to Figure 5.3. The blue line indicates the mean of the 100 simulations at each value of *n*, whilst the red line indicates the true cover across the whole water body, calculated using all 115 survey results. The black vertical lines indicate 90 per cent confidence intervals, that is they enclose the middle 90 per cent of the results produced by the 100 simulations at each value of *n*.

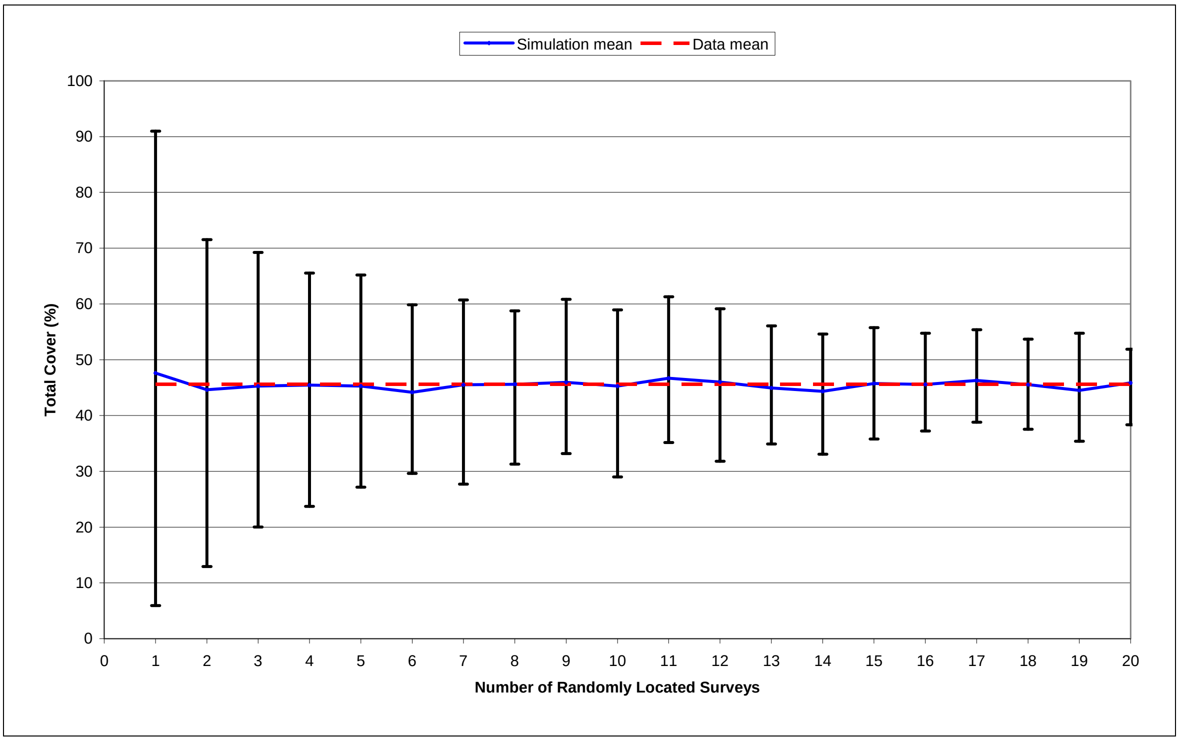


Figure 5.1 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for total cover in 1998.

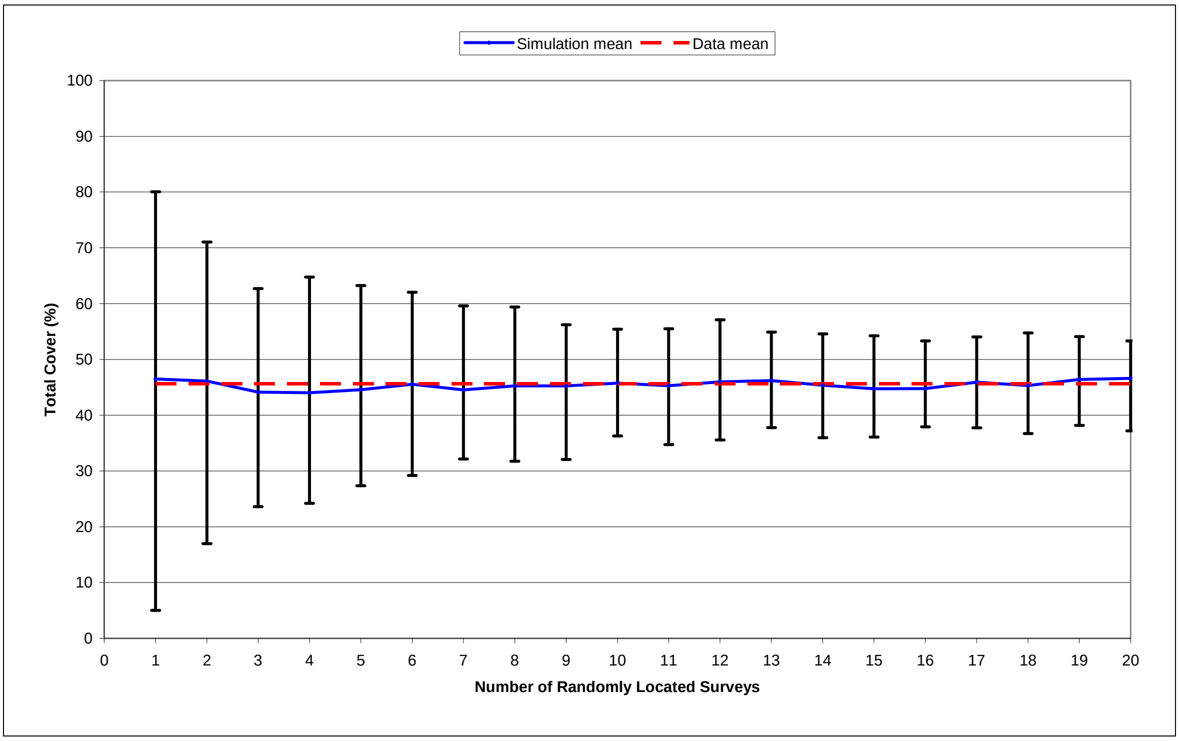


Figure 5.2 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for total cover in 2000.

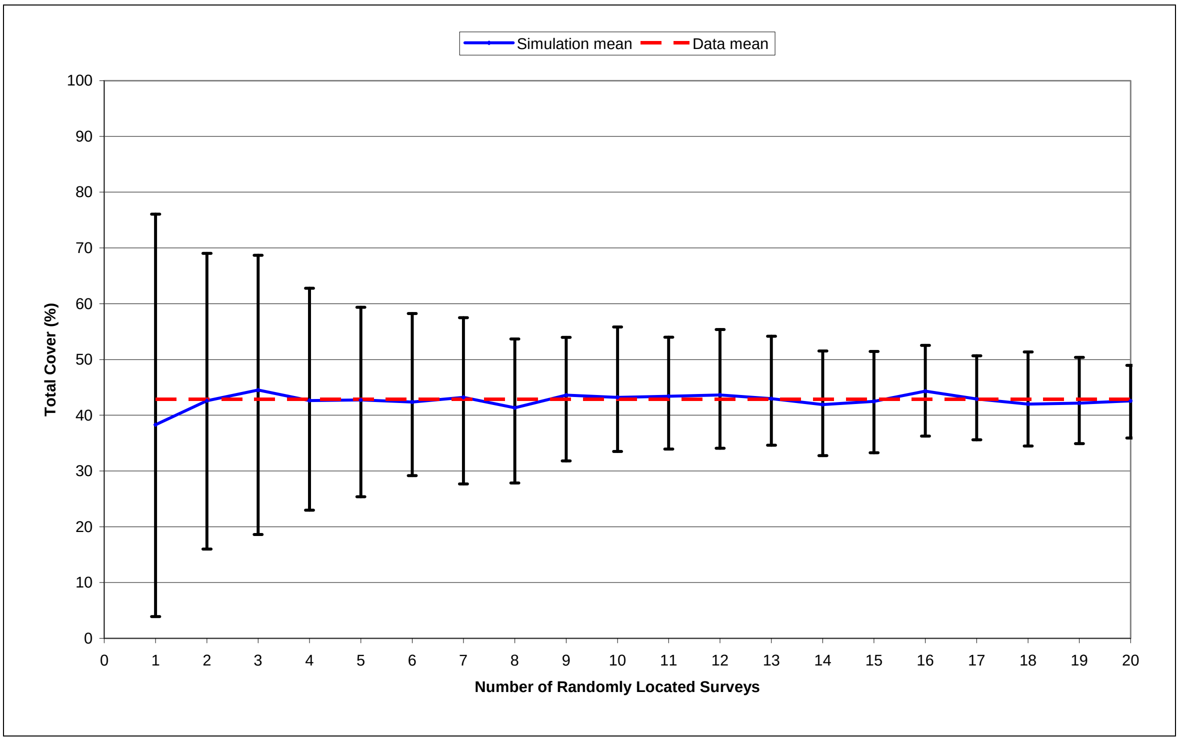


Figure 5.3 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for total cover in 2008.

In all three years, the width of the confidence interval decreases as the number of surveys increases, indicating that the precision of the estimate of mean total cover improves with more sampling effort. However, the rate of improvement decreases, indicating diminishing returns of more sampling.

5.2.2 *Scirpus/Sparganium* cover

Figure 4.7 was used to select the three years. The years chosen were not necessarily those with the highest/lowest variability since years that were considered to be 'extreme' or that did not fit the appropriate variogram model as well as others were not used. The three years selected to run the *Scirpus/Sparganium* cover simulations were 2002 (low variability), 2006 (medium variability) and 2008 (high variability).

The results for the three years are shown in Figure 5.4 to Figure 5.6.

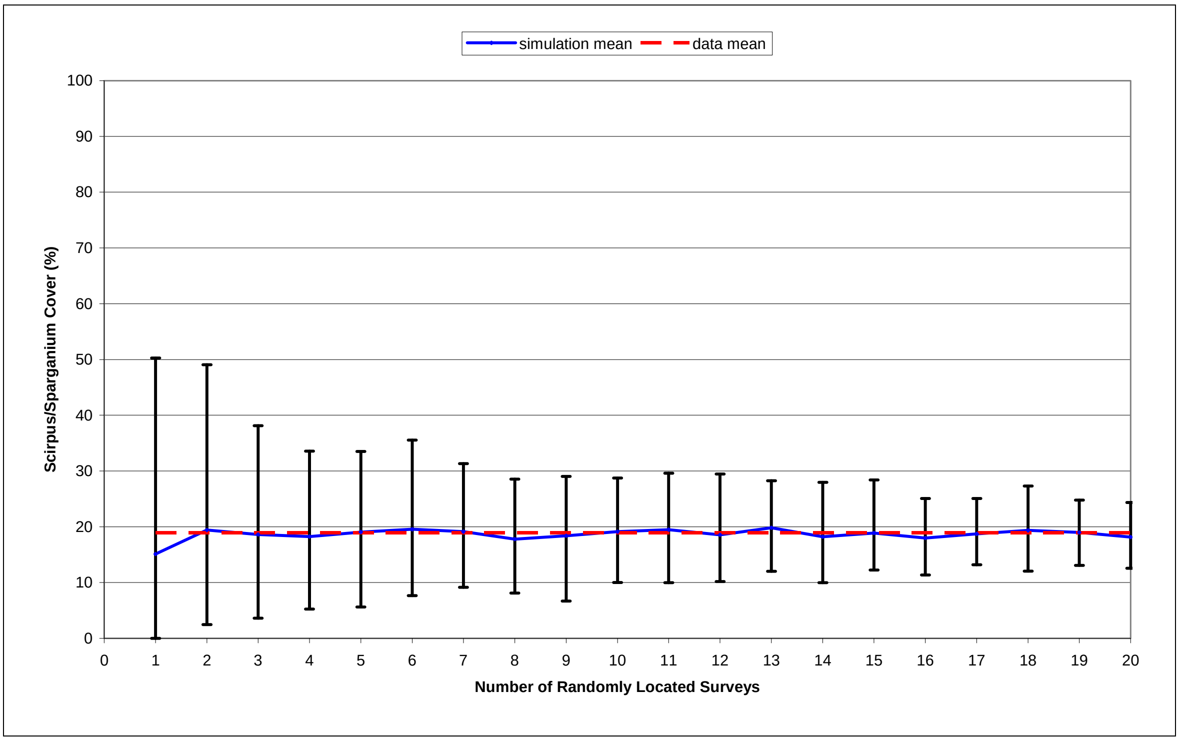


Figure 5.4 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for *Scirpus/Sparganium* cover in 2002.

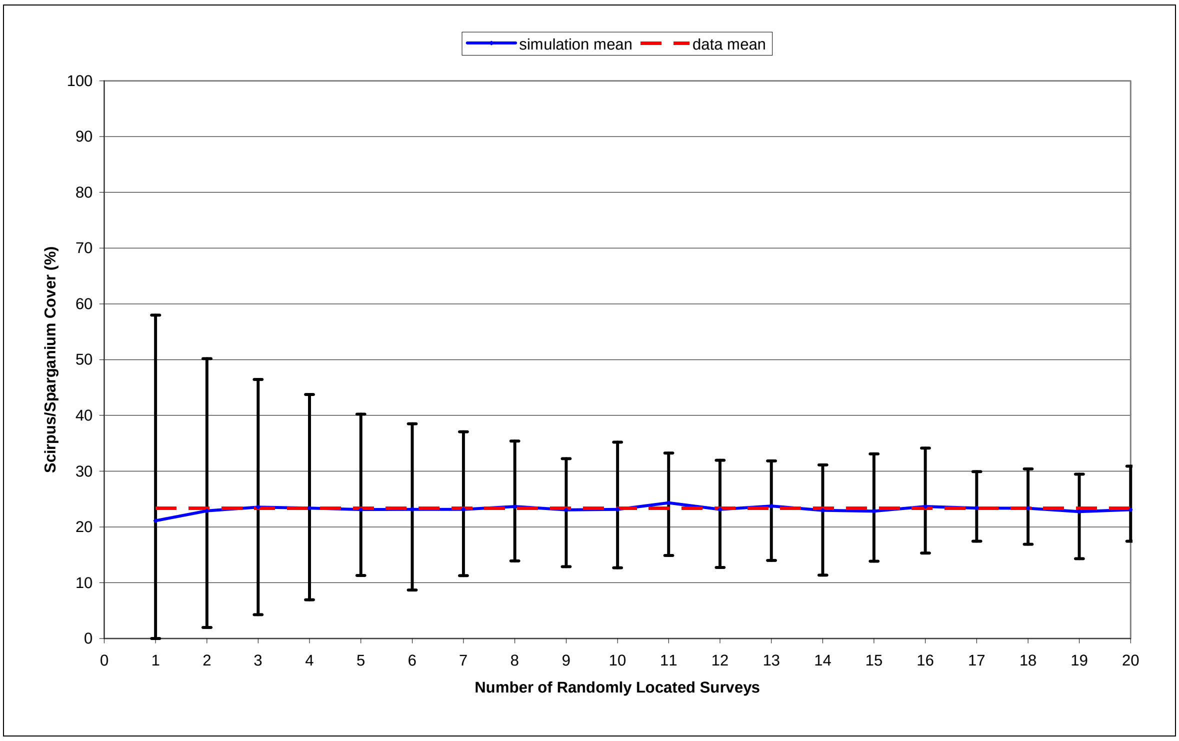


Figure 5.5 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for *Scirpus/Sparganium* cover in 2006.

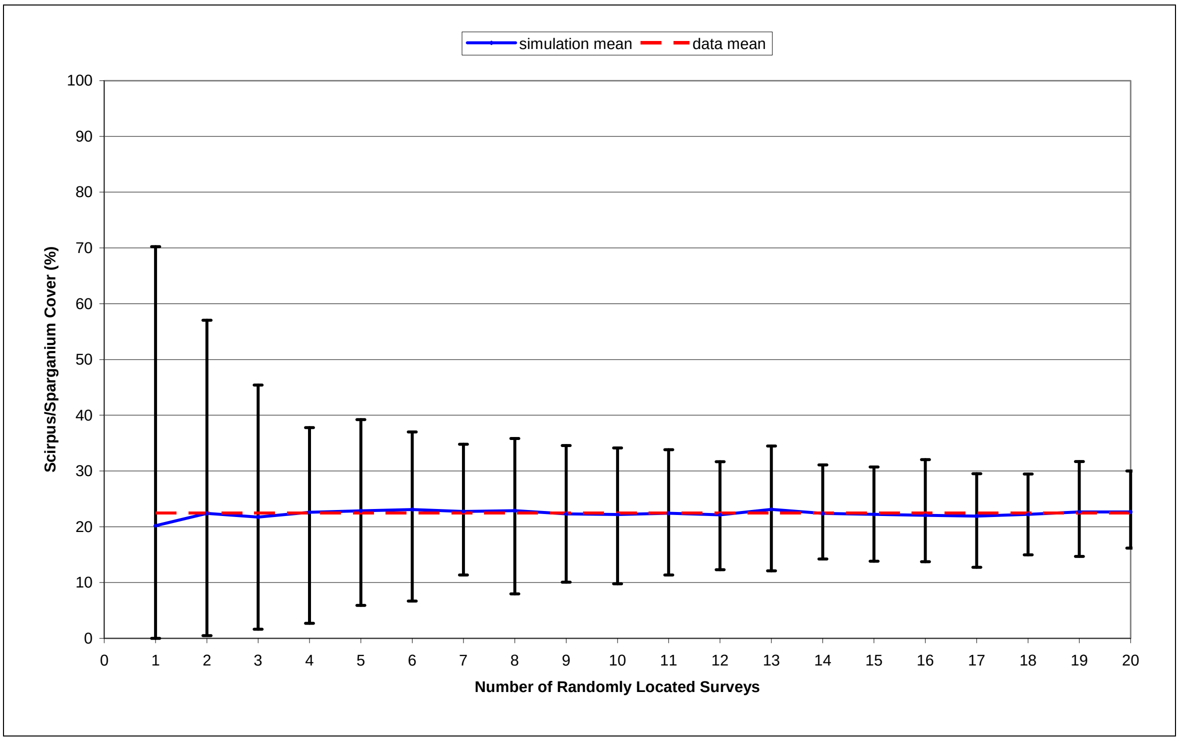


Figure 5.6 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for *Scirpus/Sparganium* cover in 2008.

The size of the confidence interval decreases with an increasing number of surveys. This pattern is evident in all three years, although confidence intervals in 2008 (high variability year) are larger for low numbers of surveys than for the other years analysed.

As with total cover, the more surveys performed, the closer to the actual status of the water body the results will be, although the decreasing rate of improvement shows that returns diminish as the number of surveys increases.

5.2.3 *Ranunculus* cover

Figure 4.10 was used to select the three years. The years chosen were not necessarily those with the highest/lowest variability since years that were considered to be 'extreme' or that did not fit the appropriate variogram model as well as others were not used. The three years selected to run the *Ranunculus* cover simulations were 2003 (high variability), 2005 (low variability) and 2008 (medium variability).

The results for the three years are shown in Figure 5.7 to Figure 5.9.

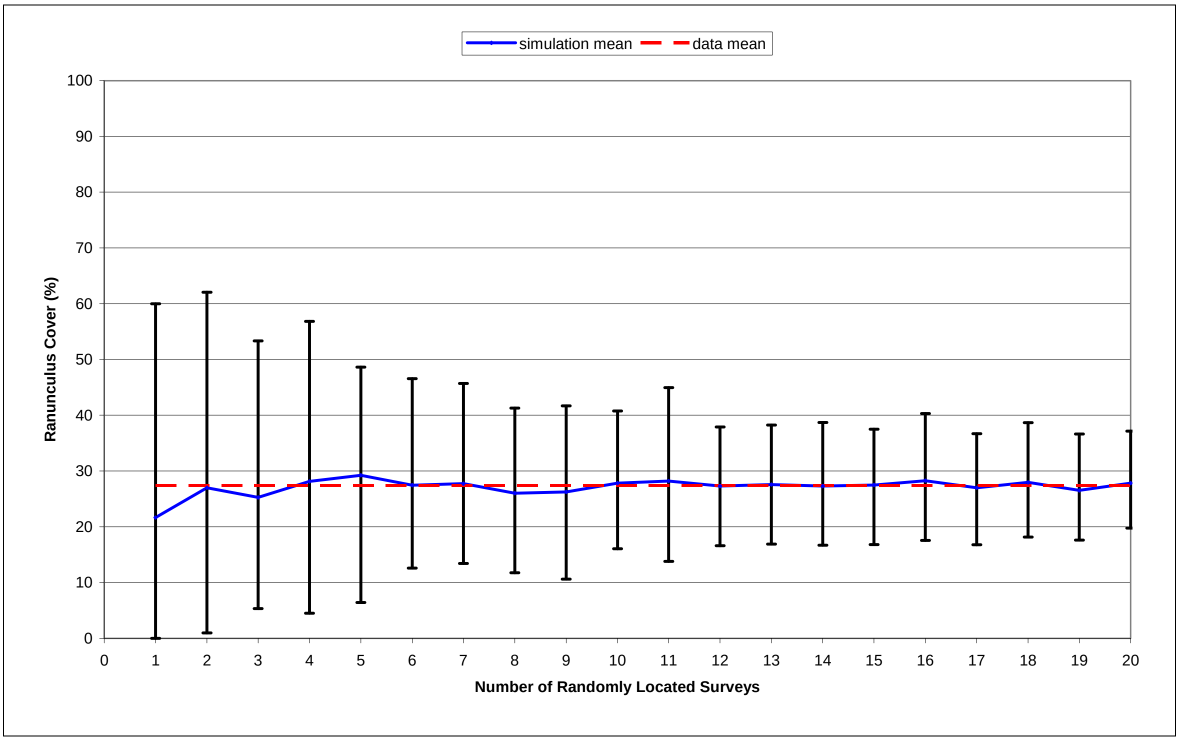


Figure 5.7 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for *Ranunculus* cover in 2003.

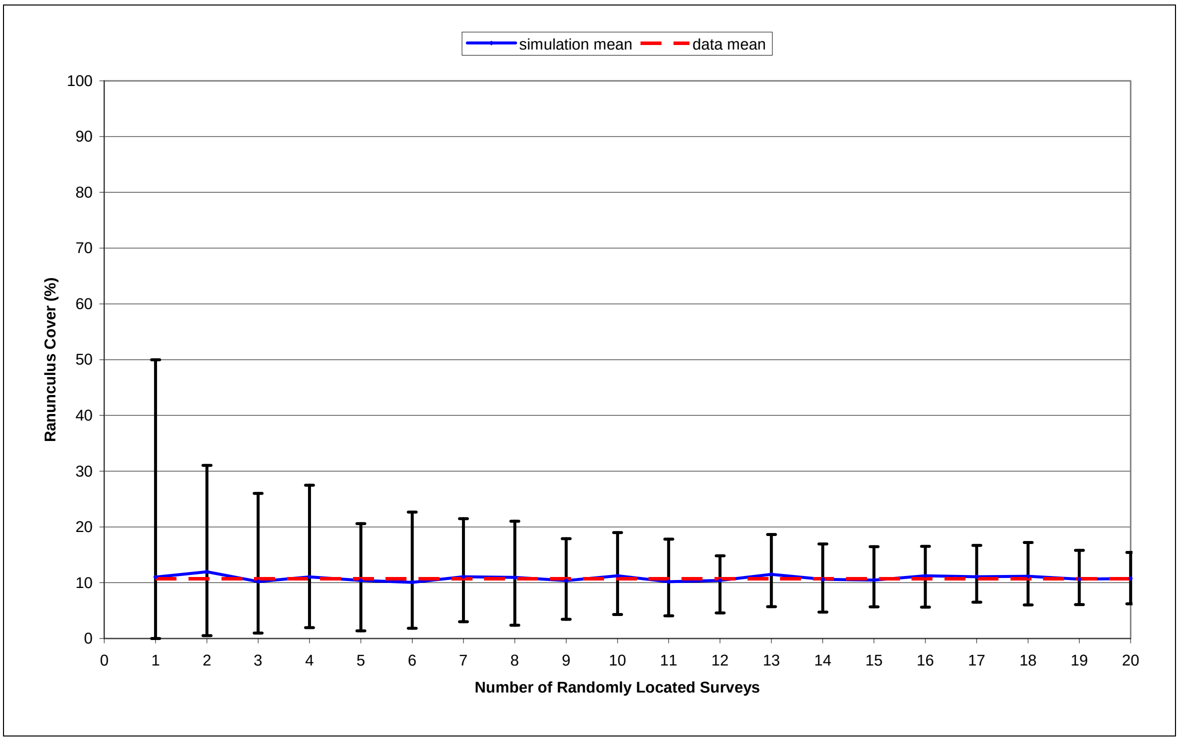


Figure 5.8 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for *Ranunculus* cover in 2005.

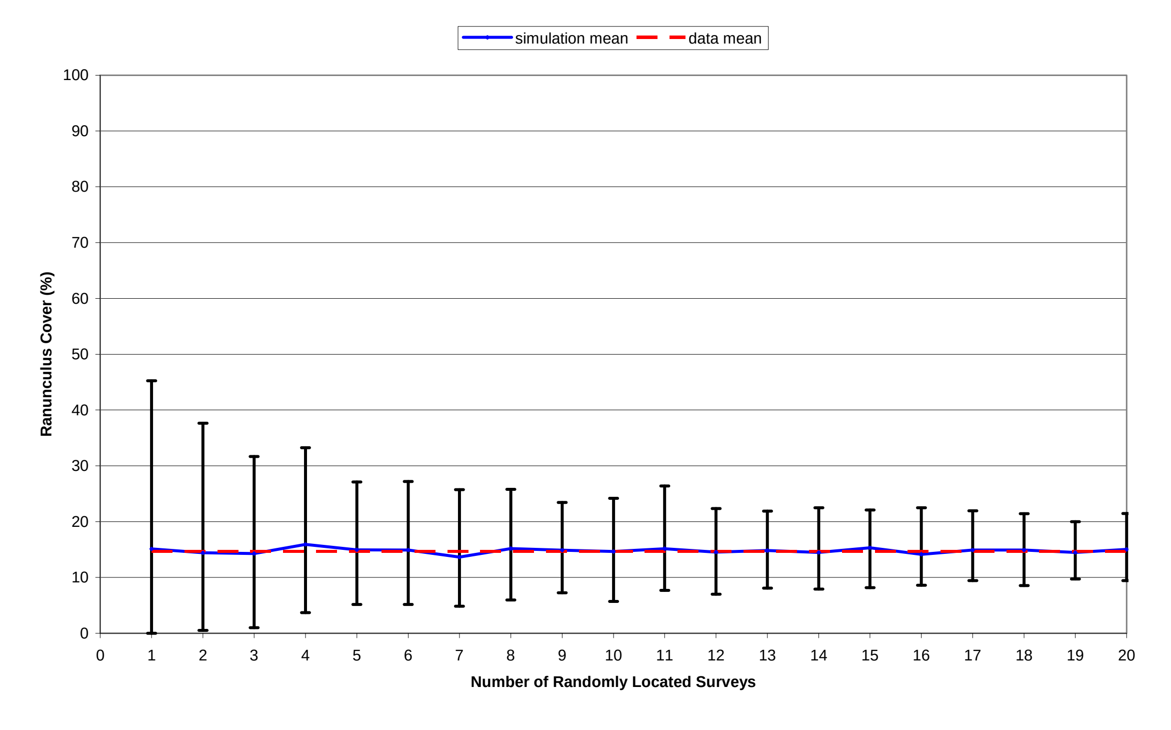


Figure 5.9 Mean and 90% confidence intervals for 100 simulations of different numbers of random surveys for *Ranunculus* cover in 2008.

The size of the confidence interval decreases with an increasing number of surveys performed. This pattern is evident in all three years, although the confidence intervals in 2003 (high variability year) are generally wider than in 2005 and 2008.

As with total cover, the more surveys performed, the closer to the actual status of the water body the results will be. However, performing a greater number of surveys is more expensive and takes more time. The rate of improvement also decreases with an increasing number of surveys, indicating diminishing returns of more sampling. A balance must be struck between cost and confidence in the results to determine the number of surveys to perform.

5.3 Sampling strategy

How does the spatial arrangement of the surveys within a water body affect the precision of the results?

Simulations were run to analyse the impact on the results of different spatial patterns of surveying. Five surveys were chosen from the water body and averaged to give a result for the water body as a whole. The spatial location of these surveys within the water body was varied in four ways:

- **Random**: The five surveys were randomly located throughout the water body.
- **Stratified random**: The water body was divided into five equal areas and a random survey in each area was selected.
- **Regular**: The surveys were located at regular intervals (of 2.3 km) along the water body.

- **Continuous**: The surveys were contiguous (five x 100-m stretches) with a randomly selected starting point.

As for the number simulations, the mean of the 100 average cover values for each of the different survey strategies was determined and plotted, along with the 90 per cent confidence interval, against the mean value calculated from the entire dataset.

As in Section 5.2, the simulation exercise was repeated for each of the three parameters in each of three years.

5.3.1 Total cover

The results of the simulations are shown in Figure 5.10 to Figure 5.12, with the dashed red line showing the actual mean of the water body, the blue dots showing the calculated mean from the simulations and the black lines showing the 90% confidence interval for the simulations.

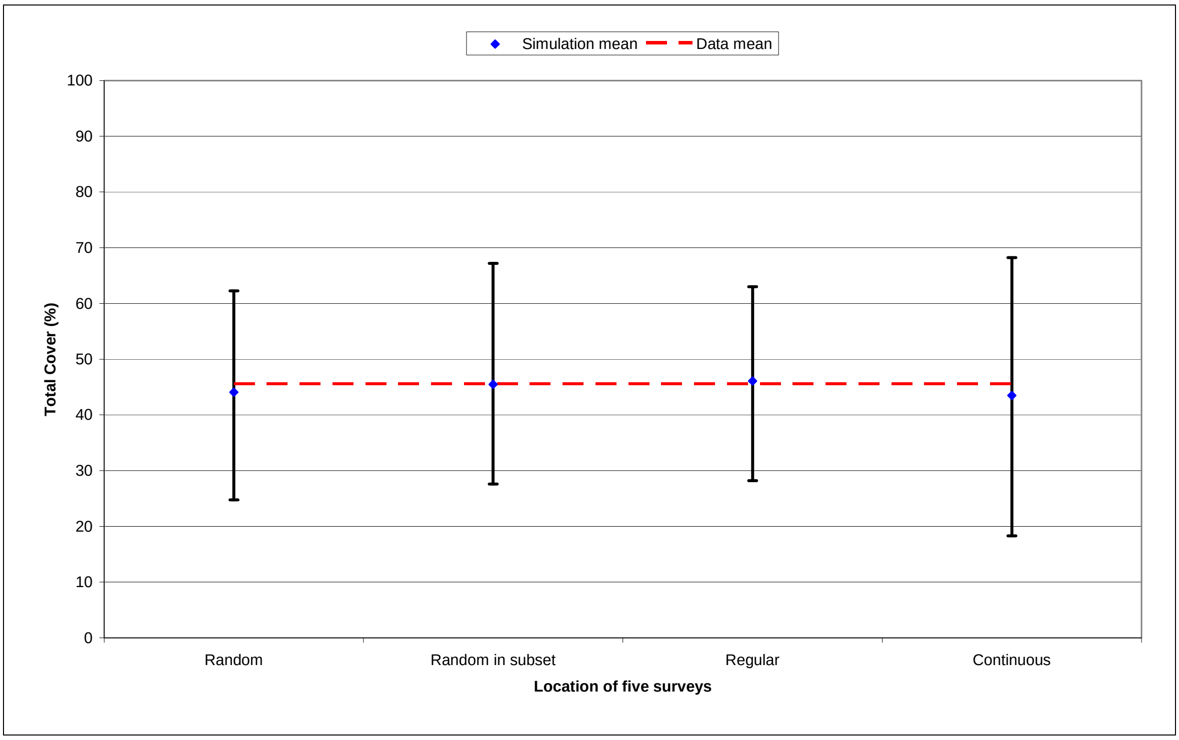


Figure 5.10 Mean and 90% confidence intervals for total cover in 1998 from 100 simulations of each of four sampling strategies.

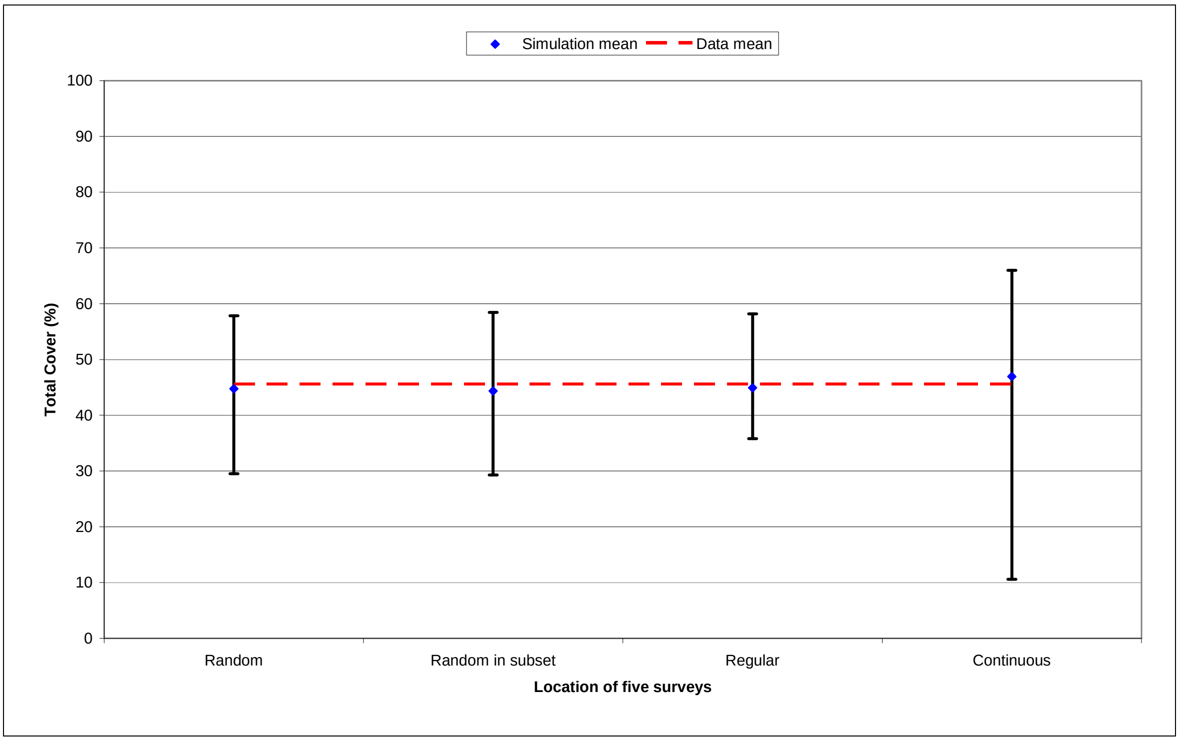


Figure 5.11 Mean and 90% confidence intervals for total cover in 2000 from 100 simulations of each of four sampling strategies.

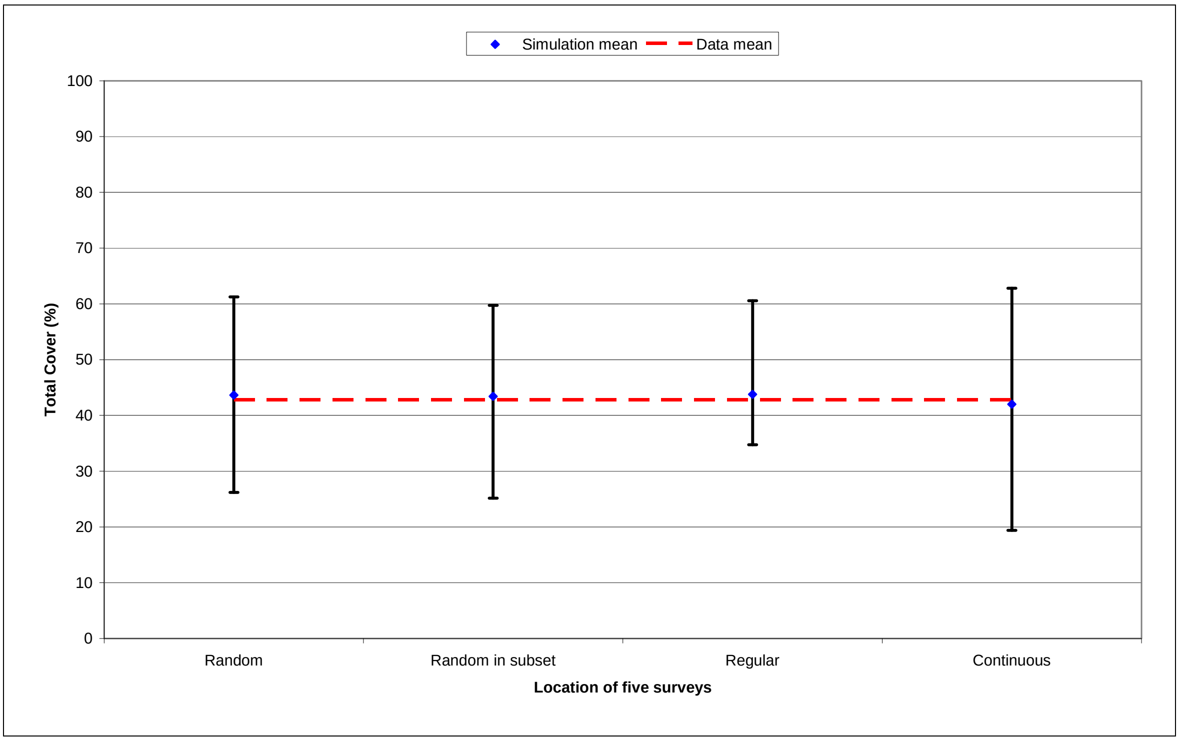


Figure 5.12 Mean and 90% confidence intervals for total cover in 2008 from 100 simulations of each of four sampling strategies.

The size of the confidence interval is typically largest for the continuous strategy. This method of locating survey sites allows less confidence to be placed in the results being representative of the water body. The other three methods have a similar size of confidence interval, with regularly located surveys having the smallest. This method of locating survey sites places the most confidence in the results being representative of the water body. This pattern is observed in all three years but is most pronounced in

2000 (the year with high variability in total cover throughout the river). These differences between sampling strategies are less pronounced than for the other two taxa (below) because total cover shows less spatial variation.

5.3.2 *Scirpus/Sparganium* cover

The results of the simulations are shown in Figure 5.13 to Figure 5.15.

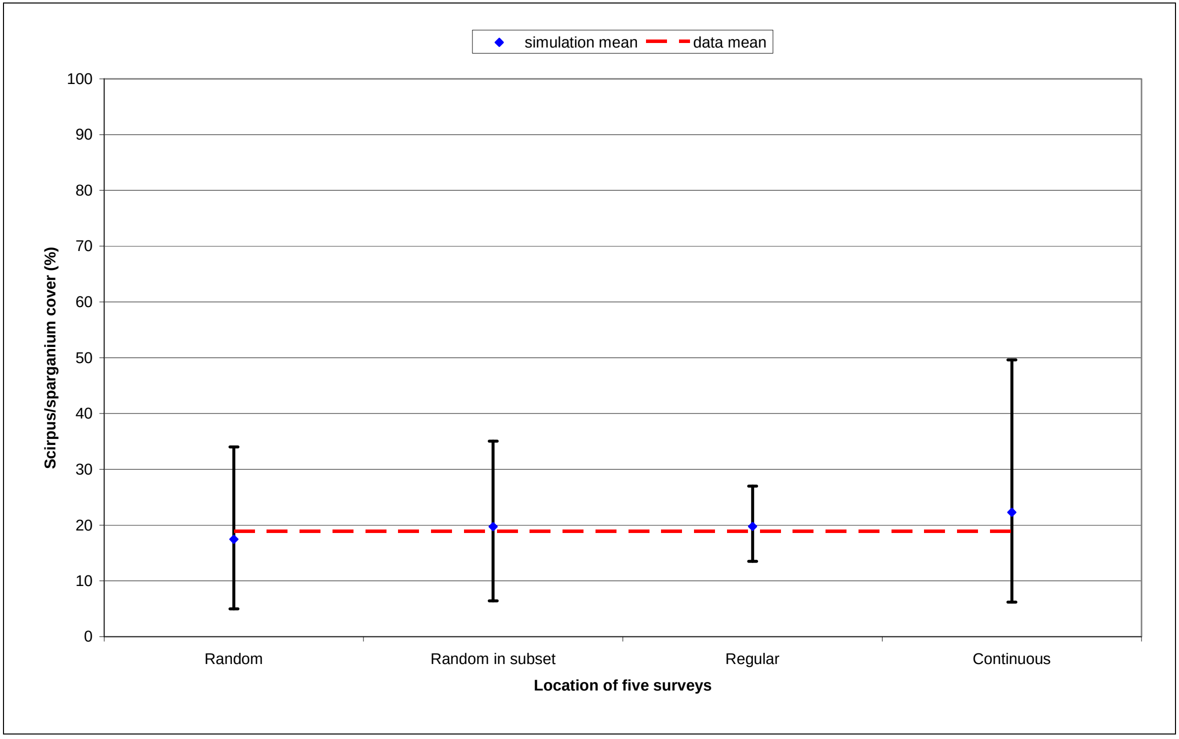


Figure 5.13 Mean and 90% confidence intervals for *Scirpus/Sparganium* cover in 2002 from 100 simulations of each of four sampling strategies.

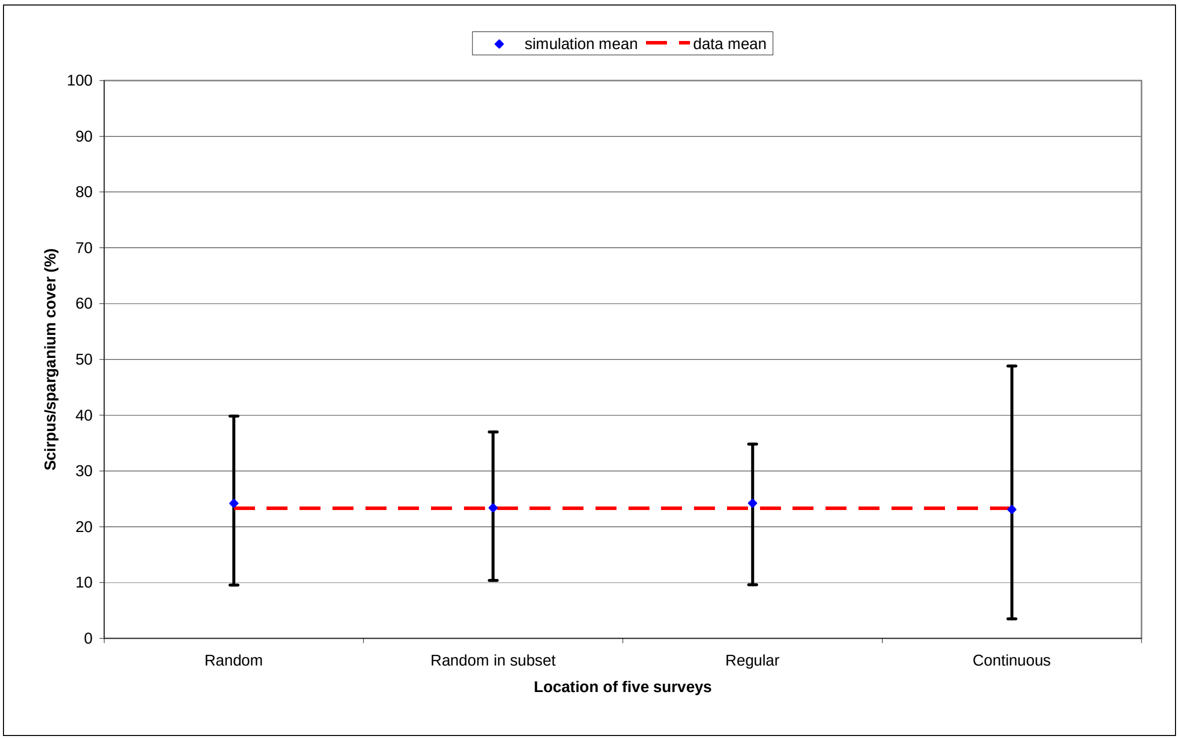


Figure 5.14 Mean and 90% confidence intervals for *Scirpus/Sparganium* cover in 2006 from 100 simulations of each of four sampling strategies.

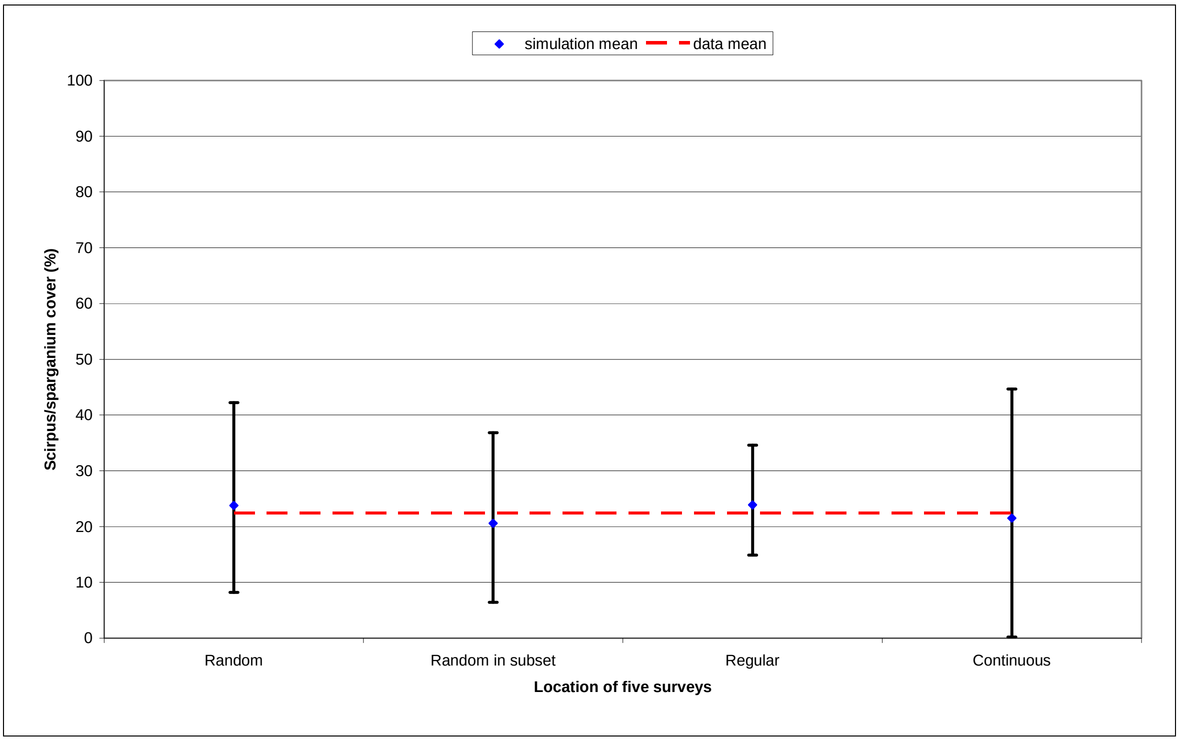


Figure 5.15 Mean and 90% confidence intervals for *Scirpus/Sparganium* cover in 2008 from 100 simulations of each of four sampling strategies.

As with the other taxa, the size of the confidence interval is typically largest for the continuous surveys. This method of locating survey sites allows less confidence to be placed in the results being representative of the water body. The other three methods have a similar size of confidence interval, with regularly located surveys having the smallest. This method of locating survey sites allows the most confidence to be placed in the results being representative of the water body.

This pattern is less noticeable for *Scirpus/Sparganium* than for *Ranunculus* (though more so than for total cover). This reflects the greater range observed for this taxon than for *Ranunculus*. The short range observed for *Ranunculus* (see Section 4.4) is typically less than three km and means that sites located at some distance from each other have no spatial correlation in *Ranunculus* cover values. Therefore, a continuous method of surveying will fail to capture the total variability of the cover values as it is confined to a small area. Regular surveying is best placed to capture as much variation as possible, because survey sites cannot be clustered as is possible with random surveying and, to a lesser extent, random surveying in subsets.

5.3.3 *Ranunculus* cover

The results of the simulations are shown in Figure 5.16 to Figure 5.18.

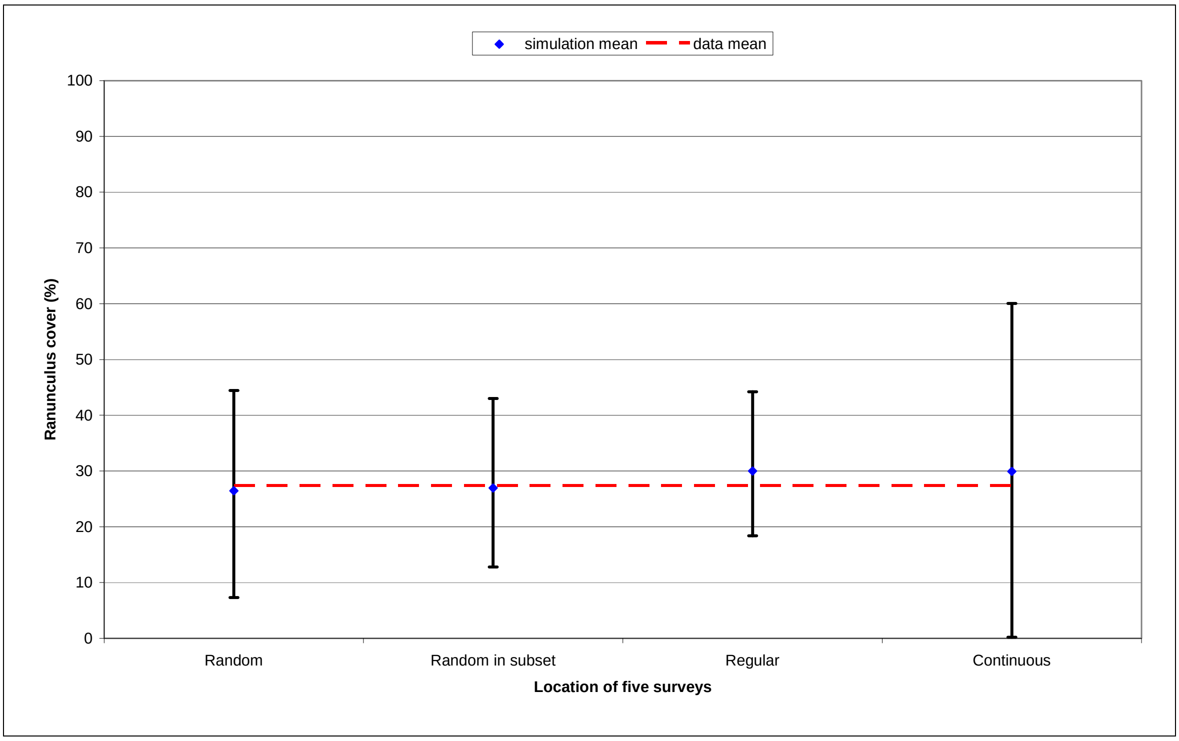


Figure 5.16 Mean and 90% confidence intervals for *Ranunculus* cover in 2003 from 100 simulations of each of four sampling strategies.

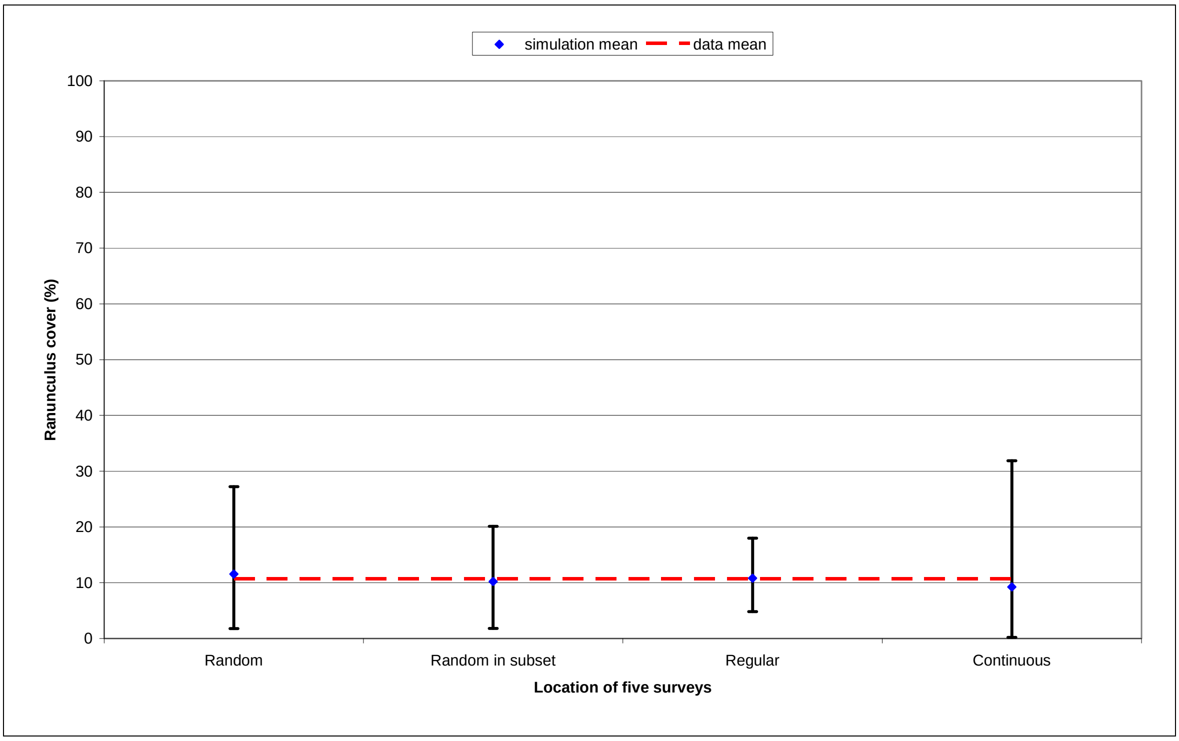


Figure 5.17 Mean and 90% confidence intervals for *Ranunculus* cover in 2005 from 100 simulations of each of four sampling strategies.

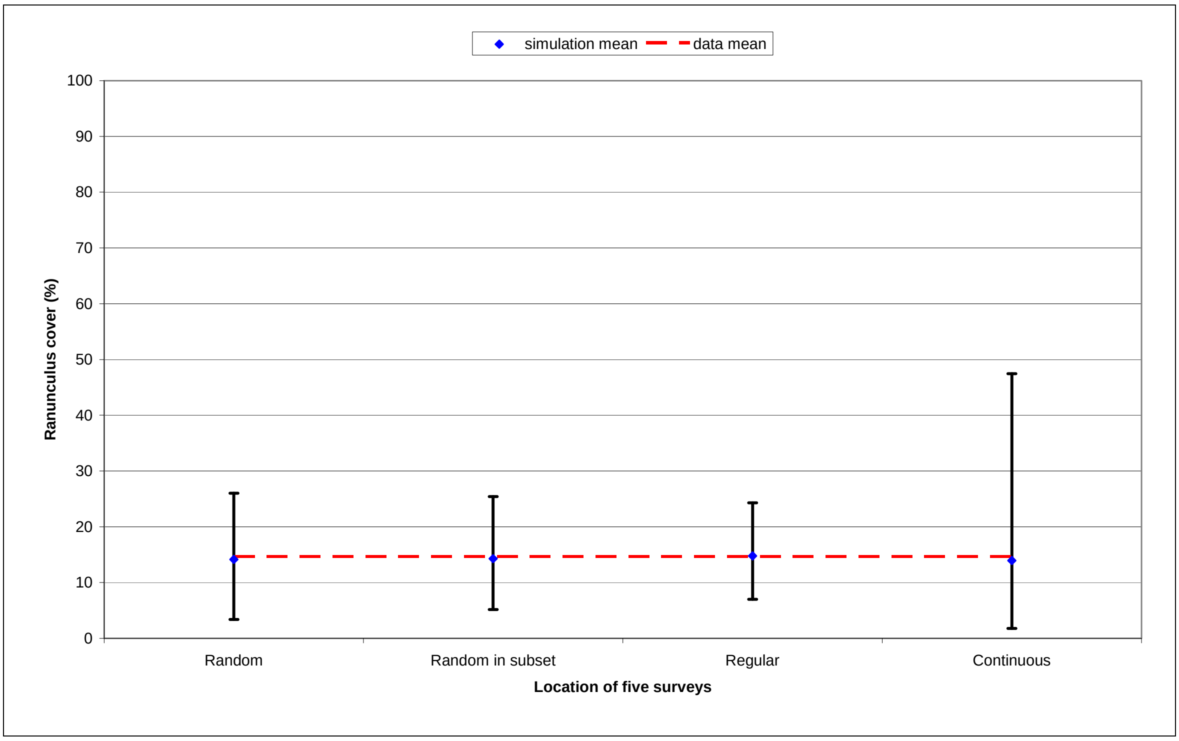


Figure 5.18 Mean and 90% confidence intervals for *Ranunculus* cover in 2008 from 100 simulations of each of four sampling strategies.

As with total cover, the size of the confidence interval is typically largest for the continuous surveys. This method of locating survey sites allows less confidence to be placed in the results being representative of the water body. The other three methods have a similar size of confidence interval, with regularly located surveys having the smallest. This method of locating survey sites allows the most confidence to be placed in the results being representative of the water body.

This pattern is more noticeable for *Ranunculus* than for total cover. This reflects the high level of local spatial variation and short range observed for this taxon (see Section 4.4). This short range, typically less than three km, means that sites located at some distance from each other have no spatial correlation in *Ranunculus* cover values. Therefore, a continuous method of surveying will fail to capture the total variability of the cover values as it is confined to a small area. Regular surveying is best placed to capture as much variation as possible, because survey sites cannot be clustered as is possible with random surveying and, to a lesser extent, random surveying in subsets.

This pattern is clearly observed in all three years.

6 Conclusions

The aim of this study was to analyse the variation in macrophyte communities in the River Allen in Dorset. Although the River Allen dataset has provided interesting new insights in variation in macrophyte communities, it does have two important limitations:

(i) the survey uses a non-standard protocol, which means that the results cannot easily be converted to EQRs for comparison with the national LEAFPACS database used in the nationwide analysis;

(ii) the results are likely to be representative only of conditions in southern English chalk streams.

Nevertheless, the uniquely high spatial and temporal variation of the River Allen macrophyte monitoring programme provides a number of insights into the structure of macrophyte communities, and the influence that this has on monitoring results.

The high resolution of spatial data allowed the average variances between sites at different distances apart (lags) to be determined using variograms. The main conclusions were:

- Spatial variation in cover values between stretches increases with distance between the stretches and gradually plateaus off at a level corresponding to the maximum spatial variance. This pattern was observed for all the variograms, except for total cover in 2007.

- Nearly all the models showed high variability at a spatial scale smaller than that measured (100-m stretches). This could reflect spatial variation within each 100-m sampling unit and/or measurement error (operator variability). This variation often contributed around half of the total spatial variation, indicating that macrophyte communities show high variation at very small spatial scales.

- Most of the spatial variation in *Ranunculus* cover occurs over small spatial scales (less than three km). Therefore, surveying a few localised stretches of a chalk stream will be adequate to estimate the true spatial variation in this taxon. By contrast, *Scirpus/Sparganium* and total cover show increasing spatial variation up to around 15 km and widely-spaced surveys will be required to adequately characterise spatial variability of these parameters.

- Patterns and magnitudes of spatial variation exhibit high levels of inter-annual variation for total cover and *Ranunculus.* By contrast, spatial variation in *Scirpus/Sparganium* is relatively constant from year to year. Surveys for this taxon can therefore be undertaken at longer intervals than for the other parameters analysed.

Simulations were run to estimate and compare the precision of estimates of mean cover achieved using different sampling schemes. The main conclusions were:

- The precision of estimates of mean cover improve with the number of surveys undertaken. The law of diminishing returns applies, however; five to seven surveys will give reasonable precision, and increasing the number of samples will bring only minor benefits. The number of surveys required to achieve a given level of precision may be influenced by the length of the water body; all else being equal, smaller water bodies should display less spatial variability, and require fewer surveys.

- Regular spacing of surveys along the water body gives better precision than other sampling strategies. However, there is little difference between this strategy and either random or stratified random sampling. Conducting a series of contiguous surveys gives the worst precision for a given sample size. Regular sampling is therefore the most efficient method of estimating ecological status within a water body, particularly when there are large-scale spatial patterns in macrophytes community structure.

Would you like to find out more about us, or about your environment?

Then call us on

08708 506 506* (Mon-Fri 8-6)

email

enquiries@environment-agency.gov.uk

or visit our website

www.environment-agency.gov.uk

incident hotline 0800 80 70 60 (24hrs)
floodline 0845 988 1188

* Approximate call costs: 8p plus 6p per minute (standard landline).
Please note charges will vary across telephone providers