The evaluation of woodland status by means of botanical indicator species.

VICKERS, Adrian David.

The Evaluation of Woodland Status by means of Botanical Indicator Species

Adrian David Vickers

A thesis submitted to Sheffield Hallam University for the degree
Doctor of Philosophy

DECEMBER 2001

Collaborating Organisation: South Yorkshire Forest Partnership

ABSTRACT

The aims of this study were to evaluate the use of botanical species as indicators of antiquity and environmental continuity and also to provide greater understanding of the processes responsible for the formation of woodland plant communities. In order to address this, the research was undertaken along four main themes:

1) Plant colonisation rates
2) Plant species lists for woodland sites
3) The impact of surveyor effort and strategy in devising species lists for sites
4) The response of a typical woodland plant to management
5) Plant communities in an area of Scottish pine forest

In particular, this study has focussed on the determination of indicator species. Some of the problems of surveying woodlands have also been raised. These problems include a lack of thorough surveys in secondary woodland habitats, and also the difficulty of comparing woodlands when they have been surveyed for different lengths of time, at different times of the year and different recording methods employed. The rate at which species are recorded during surveys has been studied in detail using three non-linear equations, which can be used to predict the number of species missed for a given survey.

The results of investigating differences between species lists of different types of woodlands have shown that geology and age are the two most important factors affecting species composition of woodland within the study area (mainly South Yorkshire). The best method for determining indicator species appeared to be a simple comparison procedure between ancient and secondary woodland, with species split into two groups depending upon their percentage occurrence in ancient woodland (>90% and 75-90%) after compensating for unequal numbers of woodlands in the two categories. In addition recommendations have been made as to the number of indicator species required to be confident that a site is ancient.

The findings of this study and the conclusions reached will help refine the survey and evaluation procedure for conserving and maintaining the woodland resource.

ACKNOWLEDGEMENTS

The project was funded by a grant from Sheffield Hallam University and South Yorkshire Forest Partnership and supervised by Ian Rotherham, John Rose and Mel Jones of Sheffield Hallam University. Species lists were collected with help from Collin Howes (Doncaster Museum), Jean Glasscock (Sheffield City Ecology Unit) and Rhodri Thomas (Peak National Park). Statistical advice was provided, by Keith Jones (Sheffield Hallam University), throughout the study period.

I also wish to thank Glenn Vickers (Sheffield University) for developing the equations used in Chapter 4, and Keith Kirby (English Nature) who commented on the material included in Chapter 4. Steve Palmer (Centre for Ecology and Hydrology) and Mark Hill (Centre for Ecology and Hydrology) commented on initial drafts of Chapter 6.

CONTENTS

ABBREVIATIONS & CONVENTIONS

Recent secondary woodland: For this study this term has been used to describe woodland that has been planted, or regenerated naturally, since the second world war. This includes mainly natural regeneration on railway cuttings, planting of trees along roadside and some conifer plantations.

Medium aged secondary woodland: For this study this term has been used to describe woodland, that is between 50 and 400 years in age. This type of woodland does not appear on early maps, often has the word plantation in its name and frequently consists of even aged woodland with a mix of broadleaf and coniferous species.

Ancient woodland: For this study this term has been used to describe woodland that originated before the first published Ordnance Survey map of 1600. This type of woodland has frequently been replanted with a mix of broadleaf and coniferous species, but does often contain some areas of woodland which have remained intact such as along water courses and rocky outcrops. Within the study area archaeological features such as charcoal platforms, and white coal, or Q, pits are often abundant within these sites. This type of woodland is also referred to as semi-natural woodland, since it is often highly modified due to different management regimes.

CA: Correspondence analysis – ordination based on just the species scores.
DCA: Detrended correspondence analysis - ordination based on just the species scores with an attempt to remove an arch effect.
CCA: Canonical Correspondence analysis – ordination based on the species and environmental scores.
DCCA: Detrended canonical correspondence analysis - ordination based on the species and environmental scores with an attempt to remove an arch effect.

GLM: General/generalised linear model
GLMM: General/generalised linear mixed model

TWINSPAN: Indicators determined after the binary division of axes produced by correspondence analysis.

INDVAL: A flexible approach to the determination of indicator species giving equal weighting to a species abundance and presence and can also be used with any hierarchical dendogram.

Spatial data analysis

Geostatistical techniques have been used in Chapters, 2,3 and 6 for investigating the abundance of species across geographic space and also species distributions across ordination axes. This analysis has been achieved through the use of semi-variograms and kriging. Semi-variograms are graphs depicting the difference between samples at a number of different lag intervals.

Semi-variance is an auto-correlation statistic defined as

$$\gamma(h) = (1/2m)\Sigma(y_i-y_j)^2$$

where

$\gamma(h)$ = Semi-variance for interval distance class h;
y_I = measured sample value at point i;
y_j = measured sample value at point j; and
m = total number of sample couples for the lag interval h.

Variogram models can be fitted to the semi-variograms, which are often done using three different models - spherical, exponential and gaussian. All of these models fit parameters for the nugget variance (or C_0 – the y intercept of the model), the sill (or C_0+C – the model asymptote) and the range (or A0 – distance over which the spatial dependence is apparent).

An example of a spherical model fitted to a semi-variogram is shown below in Figure 1. Models such as this one can be used to produce smooth 3D surfaces depicting changes across a geographical or ordinational space. This process is called kriging.

The spherical, exponential and gaussian models can be used to define dependent error structures in spatial GLMMs. The same conventions have used in these models i.e nugget, sill and range. Spatial GLMMs have been used in Chapter 2.

For further descriptions of semi-variograms and their applications see Legendre and Legendre (1998), Cressie (1991), Little *et al.*, (1996), SAS Institute (1996a) and SAS Institute (1996b). A SAS syntax program to calculate semi-variance, fit an exponential model and perform kriging is written in Appendix 1.

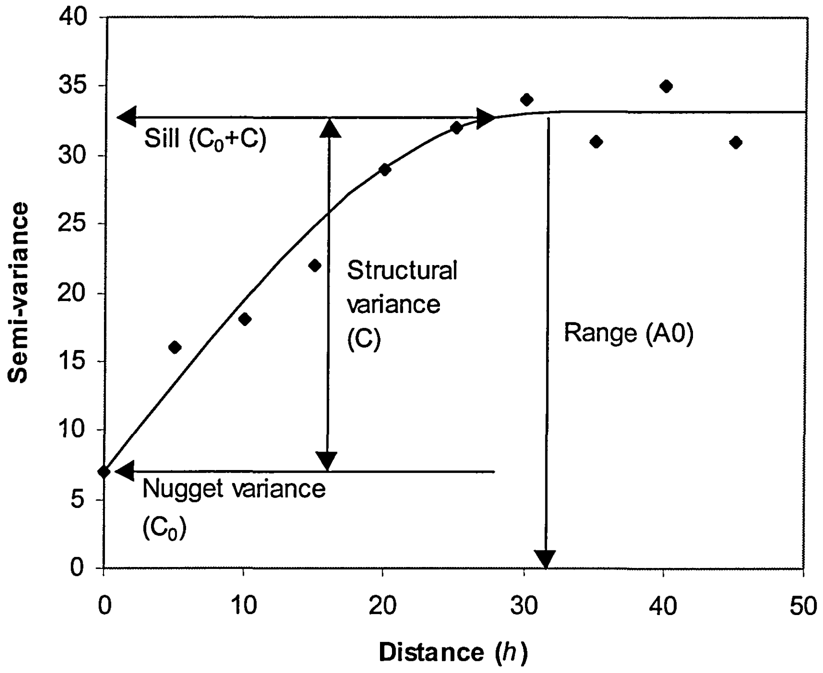


Figure 1. An example of a spherical semi-variogam fitted to semi-variance values.

Statistical significance

Where a statistical value is shown the probability of obtaining that value by chance is shown in the standard way as follows.

*	$0.01 < P < 0.05$
**	$0.001 < P < 0.01$
***	$P < 0.001$

CHAPTER 1

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Aims of the project

The aims of this study are to evaluate the use of botanical species as indicators of antiquity and environmental continuity and also to provide greater understanding of the formation of woodland plant communities. This will enable a more accurate and scientific approach to classification and evaluation of woodland plant communities, and of woodland sites. Recommendations are given for how data should be collected for producing species lists, the statistical analysis required to determine indicator species and applying this information to site evaluation.

This work will inform both surveyors and site managers on the important indicator species for rapid evaluation of sites of interest for the South Yorkshire region. Previous research into botanical indicators of ancient woodland, (Peterken, 1974, 1993; Peterken & Game, 1981, 1984), highlighted the need to form separate lists of species for different geographical regions. Increased understanding of indicator species for woodlands in a region, and in particular the historical factors affecting their distribution and abundance, will help in the effective conservation of woodlands.

The main objectives of this study are: -

1) To investigate colonisation rates and the main environmental gradients that influence species distributions within woodlands.
2) To produce a list of ancient woodland plant species, with a more northerly emphasis than Peterken's list based on woodlands in Lincolnshire.
3) To investigate the length of time it takes to record species during woodland surveys and to provide recommendations as to how woodland survey methods can be improved.
4) To investigate how a species commonly associated with ancient woodland responds to coppicing and seasonal effects (*Hyacinthoides non-scripta*).
5) To investigate the presence of indicator species from data collected from an area of Scots Pine Forest, which can be used as a comparison to the lists produced else where across Europe as well as from part of this study.

1.2 Indicator species

The identification of characteristic or indicator species is an established activity in ecology and biogeography. Studies based on fieldwork, describing sites or habitat types usually mention one or several species characterising each habitat (Dufrene & Legendre, 1997). Out of all the possible species that may exist within a range of sites only a small proportion of them may respond in a way that directly relates to a particular environmental factor. In order to assess the threshold of a particular ecosystem to a pollutant or disturbance, only those species that are likely to change significantly in abundance need to be measured. If only a small number of species in a single taxanomic group need to be recorded for different sites to be compared against each other then the speed at which sites can be evaluated is greatly increased. Examples of the use of biological indicators are very numerous and involve almost every taxon from bacteria, diatoms, fungi, lichens, mosses, vascular plants, insects, molluscs, mammals and birds. Even just a simple listing of those taxa which have been used in the evaluation of ancient woodland, includes carabids (e.g. Terrell-Nield, 1990), molluscs (e.g. Wardhaugh, 1997), flies (e.g. Young & Armstrong, 1994), hoverflies (Stubbs, 1982), butterflies (e.g. Pollard & Yates, 1993), moths (e.g. Waring, 1989), springtails (e.g. Hammond, 1974), lichens (e.g. Rose, 1976), mosses (e.g. Rose, 1992), fungi (e.g. Orton, 1986), vascular plants (e.g. Peterken, 1974), and birds (e.g. Abbate, 1992). The geographical range for these investigations has often been very limited and so their application to other parts of a country such as different regions in the UK should be done with great caution. In particular, the most detailed study of plant species has been Peterken's work in Lincolnshire. The behaviour of these species within different regions, due to aspects of their reproductive or dispersal biology, are rarely fully understood (Spencer, 1990). This makes it very difficult to predict which species are likely to make good indicator species especially where the climate differs from that of the original study area. There may also be the potential for behavioural change if climate vraies significantly through time.

Simberloff (1998) distinguishes between two main uses of indicators:

1 where the presence and fluctuations of the indicators are believed to reflect those of other species in the community;

2 where the presence and fluctuations of the indicator are believed to reflect chemical/physical changes in the environment.

This study is concerned with the second definition i.e. species indicating the antiquity of a site.

1.3 Indicators of antiquity

It is estimated that at the climatic optimum 85-90% of Britain was originally covered in woodland (Spencer and Kirby, 1992). These forests occured between 5000 and 7000 years ago and had developed after a natural succession starting approximately 10 000 years ago when the ice sheets retreated under a milder climate (Peterken, 1993). Remnants of these original woods survive in the modern landscape as ancient, semi-natural woods. Estimates based on the Nature Conservancy Council Inventory of Ancient Woodlands, show that these remnants have been reduced to 318 000 ha, or 1.4% of the land surface (Spencer and Kirby, 1992; Peterken, 1993). The Ordnance Survey First Series 1:25000 maps of South Yorkshire were produced from surveys between 1901 and 1939 (NCC, 1986). It is therefore very difficult to determine the age of woods originating before this date on map evidence alone. Historical documentation has often been lost or destroyed or sites may never have been documented.

Woods in Britain have been actively managed for social and economic purposes over many centuries. A measure of the degree of 'naturalness' is therefore required if we are to understand the ecology of these sites, and if we are to effectively assess them for nature conservation. The effects of these different historical types of management are also very important in terms of their historical value and also in determining how suitable these types of management are for present day woodlands. Indicators of these past management regimes include the archaeological remains of Victorian to medieval, through to prehistoric features, which provide evidence of continuity of woodland cover and/or human activity. Some of these archaeological features such as charcoal platforms only exist in coppice woodland. The degree of naturalness can be measured in terms of species composition (Peterken, 1993; Hermy and Stiepereare, 1981; Koerner *et al.*, 1997).

1.4 Biological indicators

1.4.1 Introduction to biological indicators

All groups of organisms contain members that have very specific requirements in relation to environmental factors such as pH, altitude and temperature range. Species also differ in their adaptation to disturbance (Grime, 1988). Some species are well adapted to a changing environment whilst other species are better adapted to out compete other species given sufficient time of stable conditions. Another strategy is that of being able to tolerate extreme environmental conditions. Shade can be considered as an extreme environmental condition. If a woodland canopy remains dense for long periods of time, then a limited number of species, those adapted to heavy shade may become dominant. Associated with these attributes, both the presence and abundance of these species, can be used to evaluate the 'naturalness' and/or antiquity of these sites. However, there are some complicating factors. Coppicing, as an anthropogenic management system, has had the effect of producing the conditions for many of these woodland indicator species that may be more favourable than under more natural conditions (Peterken, 1993). Coppicing involves the cutting of trees down to just above ground level, which was normally done every 5-25 years (Peterken, 1993). The species encouraged by coppicing are those that have been described as 'shade evaders' (Grime & Hodgson, 1988; Packham & Coin, 1990), i.e. those species which flower early enough to evade the main period of canopy cover. These species include Bluebell (*Hyacinthoides non-scripta*), Wood Anemone (*Anemone nemorosa*), Greater Stitchwort (*Stellaria holostea*), and Yellow Archangel (*Lamiastrum galeobdolon*).

Different taxa may also provide different types of information on site conditions and history. Plants may indicate the continuity of woodland as well as the physical characteristics of a site (such as geology and soil wetness *etc.*). Other taxa such as birds, which can normally move freely between different sites, may be good indicators of woodland structure rather than history or soil *etc.* This can be compared to those invertebrates which are usually more restricted in their movement than bird species, may be good indicators of both woodland structure and continuity. However, invertebrates do include species that are very mobile such as butterflies, moths and hoverflies and others which may be very restricted in their movement such as molluscs and spiders.

1.4.2 Botanical indicators of semi-natural woodland

Several writers have demonstrated that certain species are restricted to, and thus indicative of, ancient woodland. Pigott (1969) concluded that the populations of native *Tilia* in the Peak District (Derbyshire) were relict and indicated woods that had never been cleared. He suggested that certain herb species (e.g. *Convallaria majalis*, *Melica nutans*), which were strongly associated with *Tilia*, were members of a relict community. Peterken (1984) working in Lincolnshire recorded 62 species out of 174 species as having greater than 50% of their localities in ancient woodland, which he considered as those species associated with ancient woodland.

Spencer (1990) lists several species commonly associated with ancient woodland and discussed the reasons why they are so dependent on the older woodland. Most of these species have poor dispersal abilities such Herb Paris (*Paris quadrifolia*) which produces a single berry. Other species such as Wood Anemone have genetic incompatibility mechanisms affecting their reproductive capability and so tend to form vegetative, clonal populations. Species such as Small-leaved Lime (*Tilia cordata*) may be at the edge of their climatic range. Small-leaved lime depends, for succesful seed production, on a long period of hot, sunny, summer weather with the temperatures rising to 20°C for several consecutive days in June (or July and August at the northern reaches of its range) (Pigott & Huntley, 1981).

In the Scottish pine forests, there is no single higher plant or bryophyte that can be regarded as an invariable indicator of antiquity or of continuity (Pitkin *et al.*, 1994). Species stated as being typical of older pine forest include Twinflower (*Linnaea borealis*), Creeping Lady's Tresses (*Goodyera repens*), One-flowered Wintergreen (*Moneses uniflora*) and Lesser Twayblade (*Listera cordata*). Other species associated with ancient pine forest on base-rich sites include *Mercurialis perennis* and *Lonicera periclymenum* (Vickers & Palmer, 2000).

Since the status of individual species may vary across their range or from site to site, evidence from indicator plants should be based on a suite of species rather than a single one. Rackham (1995) suggests that where possible it should be based on their status and behaviour in the same region, not transferred between regions. Rackham argues, that for a regional list of ancient woodland plants, one should begin with those plants that grow in woods known to be ancient, but not in hedges. The flora of recent woods consists mainly of plants that are widespread in hedges and other environments, but the slow colonising species typical of ancient woodland are absent. This could be extended to restrict the survey to post-enclosure hedgerows since ancient woodland indicators are frequently found in older hedgerows, which have previously formed the edge of ancient woodland.

A number of indicator lists have also been derived for woodlands in other countries in Europe. These include Germany (Wulf, 1997), Poland (Dzwonko and Loster, 1992), Denmark (Petersen, 1994) and Belgium (Honnay *et al.*, 1998). Wulf's list for north-western Germany has a number of species in common with Peterken's list for south-eastern England. In particular, *Carex remota*, *Paris quadrifolia*, *Equisetum sylvaticum*, *Lysimachia nemorum*, *Lamiastrum galeobdolon*, *Galium odoratum* and *Melica uniflora* are in common with Peterken's list for species with a strong affinity for ancient woodland. Species in common with Peterken's list of species, with a mild affinity for ancient woodland, are *Mercurialis perennis*, *Viola reichenbachiana*, *Carex sylvatica*, *Veronica montana*, *Ranunculus auricomus.* Additional species, which would have been found to be strongly associated with ancient woodland using Peterken's method, but which were not found to be significant from the Chi-square tests carried by Wulf (1997), are *Scrophularia nodosa* and *Potentilla sterilis*. The only species present in both lists, considered by Peterken (1993) to have a mild or strong association with ancient woodland, but which was not associated with ancient woodland from Wulf's (1997) list is *Stellaria holostea*.

A number of different factors contribute to a species' effectiveness as an ancient woodland indicator. A botanical indicator of ancient woodland would be expected to normally produce very little viable seed, and may rely entirely on vegetative spread (Spencer, 1990). The species will also normally be very shade tolerant or be able to avoid the effects of shade during the main period when the tree canopy is fully in leaf. The former property makes a species a useful indicator of the antiquity of a site providing evidence of very little disturbance. The latter property gives the species the ability to survive in reduced light environment such as underneath a forest canopy. A species which is adapted to be tolerant of low light levels is normally out competed by faster growing species in open environments and so may be almost completely restricted to woodlands. A species with both these properties is likely to make a good indicator species of ancient woodland.

In addition to those factors that contribute to a species being associated with ancient woodland other factors are also important. These additional factors relate to how easy the species are to be found and correctly identified.

Ferris and Humphrey (1999) state that indicators should meet a number of criteria:

- **they need to be easy to assess, even for non-specialists;**
- **they must be repeatable (often using different observer bias);**
- **they must be cost effective, generating reliable data for acceptable costs;**
- **they must be ecologically meaningful, providing data which are easy to interpret.**

1.4.3 Birds as indicators of woodland structure

Bird populations may reflect the structure of the wood as well as being influenced by the surrounding land-use (Fuller, 1992). Historical records of bird species may be used to indicate the former character of woodland. For example, Ecclesall Woods in Sheffield formerly supported Nightjar, Hobby, Tree Pipit, Yellowhammer, Common Whitethroat, Garden and Willow Warbler, which are more typical of open woodlands, and scrub or heath, than of mature high forest (Medforth & Rotherham, 1997).

In a similar way Capercaillie density can be related to the age of the stands of native pine forest in Scotland (Picozzi *et al.*, 1992; Marquiss *et al.*, 1997). Capercaille prefer relatively open forest containing trees with large enough horizontal branches to sit on, and enough space to fly between trees (Picozzi *et al.*, 1992). Crested tits are also relatively sensitive to the structure of the pine forest preferring old pine forest as well as thinned plantations (Summers *et al.*, 1994).

1.4.4 Insects as indicators of woodland continuity and structure

Some insect taxa may fit somewhere between plants and birds. There are groups that can be used to indicate either the continuity of woodland cover on a site, and/or the structure of the contemporary woodland. The popularity of taxanomic groups of invertebrates such as beetles and hoverflies for entomological recorders means that these groups may be useful indicators. This will increase in the future as more species lists are compiled and the data are analysed. Clear differences in the insect assemblages between different stand types according to age and stand density have been found for pine forest communities (Young & Armstrong, 1994).

Spencer (1990) argues that there has been a strong emphasis on the conservation importance of those woodland plant communities which contain many of the species that have come to be associated with ancient woodland, at the expense of woodlands which may appear to be less species rich such as such oak-birch-bracken woods of acidic soils. The latter may be of considerable importance for invertebrates and fungi, but are typically poor in flowering plants. These types of woodlands are likely to fare poorly in any assessment based on purely botanical surveys.

The additional information provided by different insect groups such as hoverflies can be limited. For example some species are associated with particular plants – the hoverfly *Portevinia maculata* is associated with *Allium ursinum* (Stubbs, 1982), so that an additional list of the insect species may not actually be necessary in terms of identifying an area of woodland of conservation value. Some taxa of insect and others such as certain lichens, bryophytes and fungi may give indications of other aspects of an ancient woodland – such as the presence of dead wood *etc.*

1.4.5 Plant colonisation rates

Different authors have investigated the rate of spread of species occurring away from a semi-natural woodland edge. The main geographic areas these types of studies have been carried out are the USA (Matlack, 1994), Sweden (Brunet & Von Oheimb, 1998), Belgium (Bossuyt *et al.*, 1999; Honnay, *et al.*, 1999) and the UK (Vickers *et al.*, 2000). Anecdotal references, to species colonisation distances have also been made by Rackham (1995).

It is often assumed that for a colonising plant species, distributions from a dispersal source will fit a negative exponential curve i.e. $\ln y = a - bx$, where y is a measure of abundance of a plant species occurring at a given distance (x). This equation would normally be fitted to the part of the curve that is distal to the mode (Willson, 1993). Other possible shapes are the negative power function (e.g. Okubo & Levin, 1989) and using log. transformed distances as well as log transformed abundances (e.g. tested by Willson, 1993).

Two authors, Honnay *et al.* (1999) and Bossuyt *et al.* (1999), have used untransformed abundances with log transformed distances. However, this assumes that the species abundance data follow a normal distribution and the distance data are negatively skewed. The problem of using these types of equations is that predicted values become negative at a certain distance away from the source, as well as not fitting the data particularly well before becoming negative. This is probably due to the violations of the assumptions of the model. Bossuyt *et al.* (1999) also made no attempt to restrict the curve fitting routine to the part of the curve that is distal to the mode.

An additional variable for distance squared could be added to model the data for the whole distance studied, assuming that the declines are symmetrical away from the mode in all directions. Otherwise, logistic regression could be used on presence/absence data. A model could then be fitted if the species occurred at a constant abundance until the edge of the forest, where they then declined in abundance. Models based on logistic regression have been fitted by Grashof-Bokdan and Geertsema (1998). They studied a number of species, which included *Lonicera periclymenum* and *Oxalis acetosella*. However, the effect of including distance squared as an additional variable was not tested. It is therefore not possible from their paper to deduce whether any of the species they studied actually peak away from the ancient forest edge. Two of the species studied, *Lonicera periclymenum* and *Hedera helix* show relatively high colonisation rates which is in agreement with Rackham (1980) and also Peterken and Game (1994). They argued that these species are indicative of secondary woodland. Grashof-Bokdan and Geertsema (1998) found that *Oxalis acetosella* had a relatively high colonisation rate in older target patches of forest habitats, apparently being able to spread vegetatively at a high rate under low light levels. An unexpectedly high colonisation rate in young patches of secondary woodland was found for *Viola riviniana,* which had almost exclusively colonised young patches of secondary woodlands (Grashof-Bokdan & Geertsema, 1998).

The exponential decline away from the dispersal source or a symmetrical distance with a peak away from the dispersal source are not the only distributions that have been observed. Other studies have shown that the mode may be closer to the tail of the distribution than to the origin. This is the case for a number of ballistically dispersed seed shadows (e.g. Stamp & Lucas, 1983).

Modern regression techniques are now available for situations in which different distributions can be assumed. For example, generalised linear models can be used where a

Poisson or mixed Poisson distribution is assumed together with a log link function. The link function links the linear model to the mean of *y*. This type of model still gives the negative exponential decline away from a source, but may be able to handle the data better than simply log transforming abundance since it removes the possibility of negative predicted counts

The reason for the negative exponential curve is explained, by Brunet and Von Oheimb (1998), to be a result of the establishment of isolated individuals and followed by the infilling of the plant species between these pioneers. The probability of chance events that influence the colonisation patterns of species might be expected to decline exponentially away from the parent plants. This would give an exponential decline of a species' abundance away from the forest edge. The chance of a species becoming established will also decline with distance from the forest edge into adjacent open landscape. This is because the canopy and other environmental requirements, that influence competition with other species, may not occur beyond a certain distance from the forest edge. Matlack (1994) argues that some species are so particularly well suited to the forest edge to be considered as 'edge specialists'. These species can be identified as those which have greater mean rates of migration than interior species. These species are generally less competitive than interior species but are also tolerant of the deep shade, and cannot withstand the competition from plant species growing outside of the shade produced by the woodland canopy. Species in the UK which fit this description and have been shown to peak away from the forest edge include *Circea lutetiana, Rubus fruticosus* and *Pteridium aquilinum* (Vickers and Rotherham, 2000).

In addition, to comparing the slopes of regression equations, other methods have also been used where the mean, median or maximum distances of a species have been calculated away from the forest edge (Honnay *et al.*, 1999; Bossuyt *et al.*, 1999; Grashof-Bokdan & Geertsema, 1998; Brunet and Von Oheimb, 1998; Matlack, 1994). These values have been used to calculate colonisation rates in metres per year or metres per century for a number of common woodland species. Brunet and Von Oheimb (1998) recorded migration rates that ranged from no observed colonisation to 1m year^{-1} for a range of woodland vascular plant species. This is in close agreement with Bossuyt *et al.*, who attested to rates of between < 0.05m to 1.15m year^{-1}. The colonisation rates published by Honnay *et al.* (1999) and, Grashof-Bokdam and Geertsema (1998) also fall within this range. Matlack

(1994) observed slightly higher colonisation rates, of vascular plant species, from northern Delaware and southern Pennsylvania that ranged from 0.08 to 2.5m year^{-1}.

1.4.6 Seed Dispersal

A number of different methods of dispersal have been described for plant species. These include wind-dispersed (anemochores), animal-dispersed (zoochores), ant-dispersed (myrmecochores), auto dispersal (autochores), water-dispersed (hydrochores) and bird-dispersed (ornithochores) and no specific dispersal agent (barochores). Zoochores can be further split into adhesive (epizoochores) and ingested (endozoochores) (Brunet & Von Oheimb, 1988; Matlack 1994; Grashof-Bokdan & Geertsema, 1998).

Differences between the vegetation patterns of myrmecochores and ornithocochorous species along contiguous ancient-recent forest stands can be explained at least in part by dispersal mechanisms (Bossuyt *et al.* 1999). In temperate deciduous forest, myrmecochorous species have a lower dispersal capacity than ornithochorous species (Brunet & Von Oheimb, 1988; Matlack 1994; Grashof-Bokdan & Geertsema, 1998). A general rule has been suggested by Bossuyt *et al.* (1999) and independently by Matlack (1994) who both place some of the different mechanisms in the same order of colonisation rates - ingested zoochores > adhesive zoochores > anemochores ≥ myrmecochores ≥ barochores. Assuming this order to be correct, those species that have been described as ancient woodland indicators would be expected to belong to the latter two categories. Out of the species described by Peterken as being strongly associated with ancient woodland, seven are myremcochores, two are barcochores, three are ingested zoochores, and one is an anemochore. Hermy *et al.* (1999) carried out a review of published data on 132 plant species that have been described as indicators of ancient woodland. The data they collected is reproduced in Figure 1.1. A mix of different dispersal strategies can again be seen for those species that have been referred to as ancient woodland indicators. It would therefore appear not to be reliable to use plant species dispersal mode alone, to predict whether a species is likely to be a good ancient woodland indicator. Other factors need to be taken into account and these should include the probability of the use of other different dispersal mechanims. For example, *Carex strigosa* and *Carex remota* have been described as hydrochores, but they are frequently found in damp flushed areas of woodland with no running water implying that they may rely on another mechanism. *Anemone nemorosa* is frequently referred to as a myrmecochore, but in broadleaf woodland in the UK, it probably relies almost entirely on vegetative spread (Spencer, 1990).

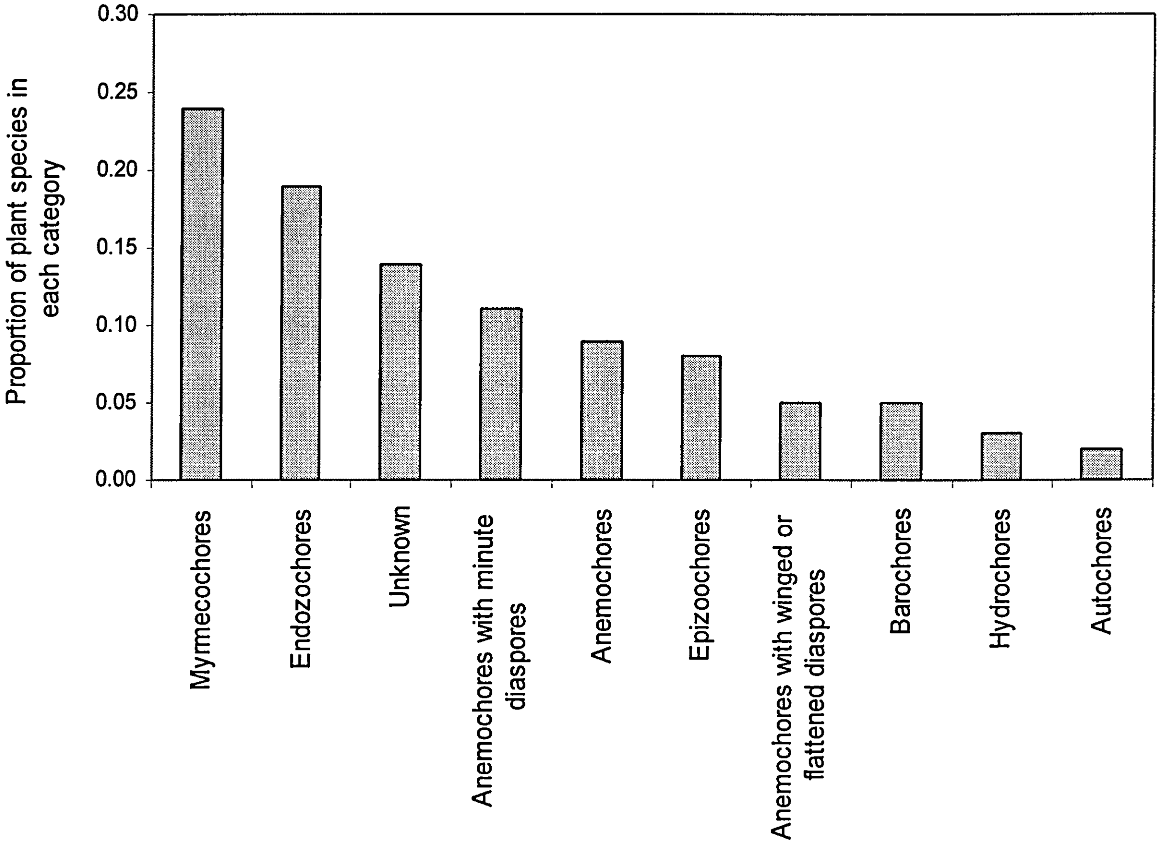


Figure 1.1. Proportion of plants that have been described as ancient woodland indicators in each dispersal category from data published by Hermy *et al.* (1999), (n=132).

The number of seeds or berries produced by a plant should also be taken into consideration (for example, as already noted, *Paris quadrifolia* only produces a single berry). Those species that are considered as anemochores may produce only small numbers of seeds, and these may not be as well adapted as other species, such as *Chamerion angustifolium* for example, for travelling great distances. Wind speeds are expected to be much less within woodlands than in open environments, especially at ground level. Without this extra amount of information to produce an index of the effectiveness of a plant's dispersal strategy, investigations on the rates of spread of species contained in these broad categories are somewhat restricted. It is unlikely that this will produce further insight as to whether a species can be regarded as an ancient woodland indicator or not. Hermy *et al.* (1999) argue that there is no single mechanism, which explains completely the low colonisation ability of ancient forest plant species. According to Erikson and Ehrlen (1992) it is not only the availability of seeds but also the availability of microsites (small-scale sites suitable for germination and survival of seedlings). Out of nine woodland plant species they studied, they concluded that three species were seed limited and six were limited by a combination of seed and microsite availability.

Instead of looking at the dispersal mechanism, a simpler and probably more effective method, is to measure the colonisation rates for a number of species growing under different environmental conditions.

1.4.7 Shade tolerance / avoidance

Shade tolerance may be considered as one of the strategies against stress proposed by Grime (1988) in his C-S-R (Competition, Stress, Ruderal) classification of particular species. Packham and Cohn (1990) used this method to look at a number of woodland communities. They list a number of stress tolerators including *Sanicula europaea*, *Oxalis acetosella* and *Viola riviniana* and also a number of species considered as shade-tolerant competitors such as *Hedera helix* and *Mercurialis perennis*. Packham and Cohn (1990) also argue that many woodland herbs require freedom from grazing rather than only substantial shade.

Some species are able to show a fairly high degree of phenotypic plasticity. For example, *Lamiastrum galeobdolon* produces more flowers and seeds from fewer and shorter stolons in years with a dry, sunny spring in contrast to its performance in wet, springs with fewer hours of sunshine (Salisbury, 1942). Further investigations of this species carried out by Packham and Willis (1982) have shown that under a dense canopy *Lamiastrum galeobdolon* produces large, deep green, relatively thin, fragile shade leaves, whilst plants exposed to high light flux along hedgerows have smaller, thicker, yellowish-green leaves of low specific leaf area (SLA). This degree of phenotypic plasticity is not common to all species. For example, *Oxalis acetosella* shows much less difference between plants grown under different conditions (Packham and Willis, 1977).

1.4.8 Life histories and demography of shade-tolerant forest herbs

Species considered to be ancient woodland indicators are often thought to be long-lived, although very few published data are available to support this. Indeed annuals are relatively rare in forest herb communities. A study from the USA found that only 5.6 % of forest herbs were annuals (Struik, 1965). Bierzychudek (1982) lists a number of plant species from various references where the age of individual species or colonies has been estimated. This list includes *Allium ursinum* which has an expected life span of 8-10 years, *Anemone nemorosa* which is thought to live to at least 9-13 years, *Narcissus pseudonarcissus* which has been recorded as living for up to 12 years in shaded environments and 18 years in less shaded environments and *Teucrium scorodonia* for which colonies have been recorded as being 50-100 years old. However, this evidence appears to neglect data on size or extent of a clonal patch which might give evidence that clones may be able to perist for graeat lengths of time. Only half of the species described by Bierzychudek (1982) were reported as relying on vegetative reproduction as a primary means of population growth. Bierzychudek (1982) argues that even for those species that replace themselves primarily vegetatively, seeds are necessary for the establishment of new populations, and nearly all woodland plant species flower and produce seed regularly. However, Bierzychedek (1982) does state that the degree of compatibility, can vary widely among different populations of the same species. He suggests that incompatibility may be exhibited by about half the woodland herb species. Even for common species, such as *Hyacinthoides non-scripta*, there has been debate in the literature as to whether this species relies on vegetative spread (division of bulbs) (Grabham & Packham, 1983) or from seeds (Corbet, 1998) with the most recent published work simply saying that it relies on both mechanisms (Merryweather & Fitter, 1998).

1.4.9 The importance of other factors affecting species composition of woodland

A large amount of published work has been concerned with environmental gradients within forest environments. Some of this work has relied on the use of multivariate analysis such as TWINSPAN (Hill, 1976) and Correspondence Analysis (e.g. Brunet, 1993; Diekman *et al.*, 1999; Dzwonko, 1993; Hermy & Stieperaere, 1981; Bradfield & Scagel, 1984). A number of problems exist with some of these papers. For example, Diekman *et al.* (1999) did not test the gradient produced from DCA against any environmental factors. Instead they try to deduce the gradients from a two-way table produced using TWINSPAN. The other authors mentioned, except Dzwonko (1993), have used the interset correlations of environmental variables with the ordination axes. This has often produced a high proportion of significant variables on the first axes. Some of these variables are probably highly correlated and some of the sites that are close to each other may be spatially correlated. This will result in inflated statistical values (e.g. *t*-values and *F*-values) along with over high values for the degrees of freedom. It is therefore difficult to deduce which are the most important variables influencing the species composition of the ground vegetation from these papers. Brunet (1993) found high correlations for soil pH, soil organic matter, and also high correlations with the abundance of beech, ash and oak in the canopy. Dzwonko and Gawronski (1994) found significant effects for light intensity, distance to ancient woodland and significant correlations with Ellenberg indicator values for light, moisture and nitrogen within four different woodlands. Bradfield and Scagel (1984) appear to have found large effects for altitude, soil moisture content, soil nutrient content, soil depth and soil pH for data collected from spruce-fir forests in British Columbia, although canopy cover was not included in their analysis.

These papers suggest that the factors affecting species distributions in woodlands are very complex. Alhough some of these authors appear to have identified the main factors involved, none of them has looked for interactions **between** the main factors.

1.4.10 Historical factors affecting species distributions

The plant species composition of woodland has often been highly modified by a sequence of different past management regimes. These include sites being used for woodland pasture and/or charcoal and white coal production. In some cases woods have been used for large-scale timber production. The origins of much of the woodland across northern Europe go back to the end of the last ice age (11 000 BC). In the UK, at about 8 500 BC, tree species such as pine and hazel spread across the UK followed by oak and alder and then followed by lime and elm, then by ash, beech, hornbeam and field maple (Rackham, 1995). Destruction of the so-called ‘wildwood’ started approximately 4000 years ago. Rackham (1995) guesses that half of England had ceased to be wildwood by around 500 BC. Woodland clearance has continued up to the present day when we are left with less than 10 % woodland (Evans, 1991). Across much of the temperate area of the world there are fragments of isolated woodlands. Most British woodlands from the medieval period to the eve of the First World War, were carefully managed coppice-with-standards woods with stock-proof fences (Jones, 1984). By the late nineteenth century, coppicing was in decline within much of the UK, and woodlands were being converted to mixed plantation by singling the coppice and planting in the gaps where timber trees had been removed. Conifers such as Larch and Scots Pine were introduced to produce a harvestable crop in 30-50 years, and hornbeams, chestnuts and beeches have been planted for ornamental timber (Jones and Talbot, 1995).

Charcoal production in the UK has been documented since in the twelfth century (Jones, 1995). In order to produce the charcoal, the wood had to be burnt with a reduced amount of oxygen. In order to achieve this the stack of cords (timber of four feet in length) were covered with bracken, turf and soil (Jones, 1993). The amount of turf used during the centuries of charcoal production or the effect this had on the present-day ancient woodland is unknown, although recent research has begun to address this (Rotherham & Doram, 1990). The removal of turf is believed to produce thinner soils with possibly higher acidity as they would be more vulnerable to the leaching of bases. Subsequent conversion to high forest management, which produces a dense shade, may have had a further dramatic change on the woodland floor. The effects of these former management regimes will of course further complicate any analysis of indicator species. In woodland converted to coniferous plantation, deep shade, especially from March to May, and the accumulation of coniferous litter and humus have been shown to cause large changes in the field layer (Pigott, 1990). After 25-30 years following the replanting of a coniferous crop Pigott (1990) recorded a reduction of *Hyacinthoides non-scripta* from being the vernal dominant, to about 1 % cover. Flowering had also ceased. *Lamiastrum galeobdolon* was similarly reduced and *Anemone nemorosa* was absent. The ancient woodland of today would therefore be expected to be far more homogenous and have a much lower plant species diversity than the original wildwood.

1.4.11 Indicators of former management

Little work has been done on the long-term impacts of woodland management such as coppicing, high forest, charcoal burning or use as wood pasture, on site fauna and flora. However Koerner *et al.* (1997) (working in France), showed that areas with previous land-use such as pasture, cropland and gardens still show marked differences in their vegetation composition a century later. Former gardens and croplands were characterised by a high frequency of nitrophilic species such as *Galeopsis tetrahit*, *Geum urbanum*, *Impatiens parviflora*, *Mycelis muralis*, *Rubus idaeus*, *Silene dioica*, *Stachys sylvatica* and *Urtica* *dioica*. They found that old forests and pastures were characterised by acidophilic, or low nitrogen demanding species such as *Vaccinium myrtillus*, *Luzula luzuloides* and *Leucobryum glaucum*. They concluded that past agricultural land-use strongly modified spatial patterns of soil fertility, organic matter characteristics and contemporary plant species distributions. These highly modified patterns of distribution are still clearly visible a century later.

Management as coppice and as high forest may have had long-lasting effects. Coppicing is likely to increase the number of plant species in the soil-based seed bank, allowing both shade-tolerant and ruderal species to exist on the same site even if rarely apparent as mature plants at the same moment in time. High forest is likely to result in species-poor plant communities with shade-tolerant species tolerant of shade surviving and often becoming dominant (Evans & Barkham, 1992).

It is not clear what the long-term implications of charcoal burning are. Charcoal burning involved stripping the ground of turf to use to cover the hearths. This may have caused severe damage to the soil. Once the vegetation and topsoil have been removed, the soil would be more prone to leaching and soil compaction. This would create thin compacted soils of high acidity and low nutrient content, which in some cases may be able to prevent any plant growth under the canopy. Charcoal and whitecoal features may be good evidence of ancient woodland status. Furthermore, the impact of this usage is represented by clear changes in vegetation. Additionally, woodlands managed in this way may well be impoverished in terms of typical ancient woodland indicators.

1.4.12 Indicators of the lack of grazing and over grazing

The effects of grazing by large mammals, wild or domestic, can have both long-term and immediate effects on the ground flora of a woodland. Low levels or absence of grazing can lead to an abundance of competitive species such as *Pteridium aquilinum and Rubus fruticosus*, where there is sufficient light getting through the canopy (Ash & Barkham, 1976; Kirby, 1997). Extensive species-poor mats of *Deschampsia flexuosa* may also occur in acidic sites which have been over-grazed (Mitchell & Kirby, 1990). Regeneration of trees and shrub may be very limited due to the highly competitive nature of plants, such *as Rubus fruticosus*, which benefit from a lack of grazing. Linhart and Whelan (1980) recorded negligible oak regeneration, in a Welsh site. They considered this to be due to the shading effect of dense bramble. Grazing may create niches for seedling establishment and the reduction in the height of competing field layer vegetation (Miles & Kinnaird, 1979).

A high level of grazing can lead to bryophyte or bracken (in the absence of cattle) dominated vegetation, with ground vegetation low in stature (20cm) and no tree regeneration greater than 20cm (Mitchell & Kirby, 1990). The more palatable, grazing sensitive species such as *Lonicera periclymenum, Rubus fruticosus, Luzula sylvatica, Vaccinium myrtillus* may be confined to inaccessible areas. In Yarncliffe Wood near Sheffield, for example, birch only regenerated successfully, immediately after the establishment of a sheep exclosure (Pigott, 1983).

Excessive grazing can often result in the total loss of the shrub layer and the formation of extensive bare patches and soil erosion. The invasion of ruderal species such as *Rumex* spp and *Poa annua* commonly occurs under these conditions.

Although the effects of grazing have been well-documented in the literature, the long-term implications of individual grazing regimes and the recovery time involved are unclear. If the grazing levels within a woodland have been extreme (either high or low) for a very long time, species extinctions may have resulted. Some species may have survived in inaccessible parts of a wood. However, due to the very poor dispersal mechanisms of plant species typical ancient woodland it may require centuries to recolonise sites. The species composition and structure may therefore serve as reliable indicators of past grazing regimes.

1.5 Obtaining lists of indicator species

1.5.1 Methods used to determine ancient woodland indicator species

Various survey methods have been used to collect data, which can then be interrogated for potential indicator species. The simplest technique used, has been to compare species lists for known ancient and known recent sites. Examples of this method include Peterken's work in Lincolnshire (Peterken, 1993; Peterken & Game, 1984; Peterken, 1974), Wulf's (1997) work in northwestern Germany and Hermy and Stieperaere's (1981) work on riverine woodlands in Belgium and also that of Honnay *et al.*'s (1998) work in western Belgium. Wulf, Hermy and Stieperaere, and Honnay *et al.* used Chi-square and Fisher's exact tests to determine those species significantly associated with ancient woodland.

Peterken used data from 273 recent woods and 83 ancient woods and he used an arbitrary value of 75% for the occurrence of species locations in ancient woodland, these species he called species with a strong affinity for ancient woodland. Those species with 50-75% locations in ancient woodland he called species with a mild affinity for ancient woodland. No statistical tests were used. However, due to the size of data set used by Peterken almost all of the species he lists are significant except for only the very infrequent species. The authors from the rest of Europe have relied upon Chi-square and Fisher's exact tests. The use of such statistical tests may be misleading since they only indicate an association between a species occurrence and ancient woodland instead of providing information on how confident a surveyor should be that a site is ancient if one or more of these species is found.

It is interesting to note that the methods that have been used so far have only adhered to the last criterion on the list devised by Ferris and Humphrey (1999) i.e. they must be ecologically meaningful. None of the indicator lists published so far takes into account how easy the species are to assess or how repeatable the survey data may be.

1.5.2 Other methods available for determining indicator species

One of the most commonly used methods for identifying indicator species is by means of analysis such as through the use of TWINSPAN (Hill, 1976). However, this method has not as yet been used in the determination of ancient woodland indicator species, except for in hedgerows (Helliwell, 1975). TWINSPAN derives indicator species for classifications obtained by splitting sites along correspondence analysis (CA) or detrended correspondence analysis (DCA) axes. COINSPAN (Carleton *et al.*, 1996) is an extension of TWINSPAN which is also able to bisect a primary canonical correspondence analysis (CCA) axis. Advantages of using these multivariate techniques include that they are able to cope reasonably well with both presence/absence data as well as abundance data. Different indicator species may also be determined for ancient and recent environments or communitites in different geographic areas from the same data set. Computer progams such as CANOCO (ter Braak, 1985) can also be used to compute the ordination axes after certain unwanted variables have been removed from the ordination. This makes these methods even more flexible.

However, TWINSPAN and COINSPAN do suffer from a number of problems. The main disadvantage of these methods is that axes are divided at the value of zero along the axis with no consideration as to what clusters of samples may exist within the ordination. The other main problem is that the same criteria for dividing axes is used on both sides of the dendogram. Dufrene and Legendre (1997) present an alternative method to TWINSPAN and COINSPAN, which derives indicator species from any hierarchical or non-hierarchical classification of samples, which they called IndVal.

1.5.3 Problems with using lists of species

Woodlands are often a complex mix of different habitat-types making a systematic survey method difficult to implement. Peterken (1974) argues that complete species lists of all groups are virtually impossible to obtain even for a single site. This is his main rationale for the use of indicator species. However, in order to derive a list of indicator species ideally all the species occurring at different sites should be known. As Peterken has stated this is never the situation. Some species, such as *Anemone nemorosa*, appear above ground early in the spring and then have completely died back by mid-summer while other species, such as *Epipactis helleborine*, only appear above ground in mid-summer and disappear by late-summer or August. Differences in assessing a site for a species list, have also been found between observers (Kirby *et al.*, 1986). For data analysis purposes it is also useful to have information on the abundance of species. However, this causes even more complications for analysis, such as additional differences between observers and the prefered time of year to use to derive the abundance assessments. Observers have been found to differ in their consistency with which they tended to under- or over-estimate cover in relation to both species and to quadrat size; this consistency was not correlated with experience (Sykes *et al.*, 1983). Sykes *et al.* (1983) found that variability was highest with fine-leaved species and bryophyes and lowest when estimating broad-leaved species.

From the literature available on colonisation rates, the abundance of slow colonising species should be capable of being used to provide information on site stability and longevity. This information may still provide a useful way of identifying sites, which have had minimal disturbance. However, the method used for recording the abundance of species often relies on the DAFOR rating, i.e. recording species as rare, occasional, frequent, abundant, dominant, with the prefix of local which can be used with any of these (Kent &

Coker, 1996). Although at first this may appear a useful way of recording the abundance of species, there appears to have been no work done to try and calibrate these categories in terms of percentage coverage, or how repeatable these values are both between the same observer and between different observers and which species may be under or over estimated. Further problems have a risen from observers recording different species *i.e.* deliberately not recording very common species such as *Plantago major*, which often grow along woodland paths and rides. A further complication factor is when a species that appears to be restricted to ancient woodland may in fact occur in a wide variety of undisturbed environments rather than be a true indicator of ancient woodland. This may be particularly common with wet woodland habitats where there has been a successional change from, for example, a lowand bog to a woodland. In these circumstances the wood may appear to be ancient and contain a large number of rare species, but has not existed as woodland for a very long period of time. In order to address these types of problems, Peterken (1974) suggests that one should limit the consideration to 'woodland species', defined as (a) those species which can bear the shade of a closed woodland canopy, (b) those which create the canopy, and (c) others which in some way require woodland conditions.

CHAPTER 2

2. VEGETATION SUCCESSION AND COLONISATION RATES AT THE FOREST EDGE

2.1 Introduction

A number of recent authors have attempted to quantify the rates of spread of selected woodland vascular plants from ancient to recent woodland (Bossuyt, Hermy & Deckers, 1999; Brunet & Oheib, 1998; Matlack, 1994; and Honnay *et al.*, 1999). Brunet and Von Oheimb (1998) observed migration rates based on a maximum cover in recent woods varied between species from 0.00 to 1.00 m year^{-1}, with a median migration rate of 0.30 m year^{-1}. They concluded that the movement of species into a new community that could produce communities that are comparable to the adjacent stands proceeded at a rate of 0.3-0.5 m year^{-1}.

These types of studies have important implications for placing a nature conservation value on existing ancient woodland. In particular this concerns the implications of the length of time required for these ecosystems to form under natural conditions. The findings may also be used to help inform reintroduction programmes, which try to recreate woodland communities more typical of ancient sites in a short space of time.

Lists of ancient woodland plant species have been found to differ in their association with ancient woodland according to geographical location (Wulf, 1997). Since these indicators are often used as the first stage in assessing woodland for nature conservation, it is very important to understand the processes underlying their occurrence and abundance. The colonisation rates reported often differ considerably between research papers. For example, for *Mercurialis perennis,* Rackham (1980) mentions colonisation rates of *c.* 100 m, and Honnay *et al.* (1999) of 22 m per century. Honnay *et al.* (1999) attributed these differences mainly to variations in the level of phosphate in the soil. Similar differences between the two authors have been reported for other common woodland species such as *Anemone nemorosa* and *Hyacinthoides non-scripta*.

The objectives of this research were to compare colonisation rates for a number of different woodland plant species, and to assess the extent to which environmental factors may influence these rates of spread. A sampling strategy was devised to assess colonisation under natural conditions. However, there are a number of complicating factors. The main problem is that of the possible site related effects. It is difficult to separate effects such as the historical management of the ancient woodland or recent woodland, along with identifying effects of altitude, precipitation *etc.* A consequence of the methodology employed, was that only the environmental factors that varied at the quadrat level, could be used in the analysis. Random effects were used to try and compensate for possible site effects. The assumption behind the use of random effects is that they do not represent all the variation known to occur. Random effects normally represent an unknown site effect caused by randomly selecting a number of different sites for the investigation. The approach undertaken was based on the assumption that the observed colonisation patterns can be extrapolated from the present environmental conditions.

The second part of this study was concerned with the plant communities that occur along the ecotone from ancient to recent woodland. A number of studies point to differences between ancient and recent sites (Peterken, 1977; Whitney & Foster, 1988; Matlack (1994b); Dzwonko, (1993); Bossuyt *et al.*, (1999). A method was required which would be able to group similar quadrats together, as well as determining possible indicator species. The TWINSPAN procedure developed by Hill (1979) was developed with this aim. Sites are classified using a divisive hierarchical algorithm. They are first divided into two subsets according to their sign on the first axis of a correspondence analysis ordination (CA or DCA), each subset is then divided in two smaller subsets by repeating the same procedure (Dufrene & Legendre, 1997). However, there is a number of problems with this method. TWINSPAN assumes the existence of a strong gradient dominating the data structure, and so it may fail to identify secondary gradients or other types of structure in the data set (Belbin and McDonald, 1993). Belbin and McDonald (1993) have also criticised the way in which the cutting points are determined as they do not select for large gaps in the data. Dufrene and Legenre (1997) raise a number of other points. One is that rare and widespread species are compared on the same footing, although they do not respond to the same site clustering levels. This means that there must be several species typical of a high-level structure to make the analysis recognise them as forming a homogeneous group, distinct from other species groups(s) typical of lower level structures(s).

An alternative and very flexible method, has been proposed by Dufrene and Legendre (1997), which can make use of any ordination method derived from species presence/absence or abundance data. The sample scores produced from an ordination method are clustered to produce a hierarchical dendrogram. This can then be used in a subsequent analysis to determine species indicators at each level of the hierarchical structure. In their example, Dufrene and Legendre use principal coordinates analysis based on a similarity matrix computed using the Bray-Curtis coefficient. A non-hierarchical clustering method (*k* means) was then used to produce the dendrogram. However, their method does lend itself well to an ordination derived from direct gradient analysis such as canonical correspondence analysis (CCA). An advantage of using CCA is that it preserves Chi-square distances. Unlike the Bray-Curtis coefficient (a non-metric coefficient) Chi-square distances are metric and have the same properties as Euclidean distances (Legendre & Legendre, 1998). This means that the sample scores can be used directly in a large number of different clustering methods. The data used by Dufrene and Legendre (1997) came from very different plant communities which meant that *k* means clustering would be expected to find these groups in the data. However, the data presented in this paper come from transects. Hierarchical clustering is a much more robust method when the data have been collected from a hierarchical design and where continuous gradients exist within the data. The sample scores that CANOCO calculates, which are a linear combination of the environmental variables, may be very useful for producing the hierarchical dendrogram. In this way, the user has some influence over where a partition should be formed. This may be particularly useful when trying to split continuous gradients.

2.2 Materials and Methods

2.2.1 Study Site

Sites were selected where ancient woodland existed next to an area of recent woodland or scrub. Five sites were found where ancient woodland was adjacent to recent woodland of only 30-40 years old. The five sites selected were situated on either the Coal Measures Series or Magnesian Limestone. Graves Park, Little Matlock Wood and Smithy Wood are situated on the Coal Measures Series and Old Spring Wood and Norwood are located on the Magnesian Limestone. A brief description of each site with a summary plus NVC communities is given in Table 2.1.

Graves Park (GP) contained two areas on a very thin loamy soil; one of which had been planted with lime and sycamore and the other mainly planted with beech, although it also contained a high proportion of regenerating ash.

Little Matlock Wood (LMW) contained a very thin, sandy soil supporting a semi-natural oak woodland with the ecotone changing to hawthorn and elder scrub or *Pteridium aquilinum* dominated acid grassland.

Old Spring Wood (OSW) consisted of a very deep soil of sand and clay lenses. For this site, the recent area had been planted with mainly *Fagus sylvatica* and some *Pinus sylvestris* which are now mature trees. The sand rich areas were dominated by *Hyacinthoides non-scripta* and *Anemone nemorosa* and the Clay rich areas were dominated by *Allium ursinum*.

Norwood (NW) was located at the interface between the Coal Measures Series and the Magnesian Limestone. This wood contained a deep soil which supported a mixed broadleaf woodland of ash, oak and sycamore in the semi-natural woodland progressing to a hawthorn and elder scrub for the recent area. This site also contained a poorly drained silty area, part of an old water-course.

Smithy Wood (SW) consisted of a relatively deep soil which supported a mixed broadleaf woodland of ash and oak with the ecotone changing to a damp neutral grassland often dominated with bracken and bramble.

Table 2.1. Description of the five sites studied. The initials of each site are given.

Site name	Grid Ref. SK	Geology	Past landuse of ecotone	Trees planted	NVC in woodland	NVC along ecotone
GP	354 825	Coal Measures	Parkland	beech, sycamore, lime	W8	W8
LMW	305 893	Coal Measures	Hay meadow	hawthorn	W10	W10, W21, W2
NW	478 812	Corallian	Hay meadow	hawthorn	W21	W21, W6
OSW	537 814	Corallian	Arable	beech, pine, sycamore	W8f, W8b	W8f, W8b
SW	302 773	Coal Measures	Hay meadow	-	W10, W8	W10, W8, W24, M

2.2.2 Data collection

Five transects were laid out at each site starting at the ancient woodland boundary and in the direction away from the ancient woodland. One metre square quadrats were used with a distance between them of 2 m, making transects of 10 quadrats and 28 m in length (250 quadrats in total). The percentage covers of all vascular plant species were recorded. The number of ancient woodland indicator species in each quadrat was also recorded (according to the Peak Park list) (Peak National Park unpublished). Canopy cover was estimated by eye using the method of Vickers and Palmer (2000). A spirit level was used to record the slope for each transect and a compass was used to record aspect. A digital pH meter was used to measure soil pH. Soil samples were taken, weighed and then dried at 102°C until constant mass was achieved and the soil moisture content calculated. They were then heated to 450°C until no further loss in mass and the soil organic content calaculated. Soil depth was measured using a soil auger. At three of the sites the soil depths were too great to be measured, and so were estimated from nearby exposures of the soil profiles. The soil texture was determined using the hand-texture method (Waugh, 1990).

The patches of recent woodlands were often small, and so relatively short transects were used. It would have been desirable to extend each transect back into the ancient woodland. However, there were a number of problems with this. At two of the sites, Smithy Wood and Graves Park, the areas of ancient woodland were very thin (only a few metres in width). Old Spring Wood contained extensive areas of homogeneous plant communities often consisting entirely of a single species. The woodland that had been in existence since 1600 contained a dominant covers of *Allium ursinum* on the clay and *Anemone nemorosa* on the sand lenses with *Hyacinthoides non-scripta* dominating the sand rich areas that had been planted with a mix of tree species approximately forty years ago. This would have been difficult to model using Gaussian response curves, although a logistic model would have been appropriate. Norwood and Little Matlock Wood, were therefore the only two woods where extending the transects back into the ancient woodland may have been useful. However, in order to keep the experimental design as balanced as possible transects were started at the ancient woodland boundary.

Some information was available on the historical management of the sites studied. However, because of the small number of sites it was not possible to separate the effects of this from those of geographical location such as longitude, latitude and altitude. Estimated values for any unexplained error for each site were therefore calculated for each wood and these discussed in terms of possible historical and environmental effects.

2.3 Data analysis

2.3.1 Generalised Linear Models

Generalised linear models were used in the analysis of the data to investigating species declines from the ancient woodland boundary. The GENMOD procedure was used in SAS (SAS Institute, 1996) with a log link function and Poisson error term specified.

An over-dispersion parameter was included in the model to compensate for the possibility that the data did not come from a single Poisson distribution. If the data do not belong to a single Poisson distribution then then test statistics become inflated. If an over-dispersion parameter is included this compensates for this lack of fit by reducing the test statistics appropriately. The greater this lack of fit the greater the over-dispersion parameter. If a significant difference occurred for the deviance statistic between linear and quadratic equations (Chi-square test with 1 DF) then the quadratic equation was used and Type-I F tests reported.

2.3.2 Ordinations

Ordinations produced by canonical correspondence analysis (CCA) were computed using CANOCO (ter Braak & Smilauer, 1998). The species cover scores were log transformed and rare species were down weighted. This is a very common method since log transforming the data effectively increases the weight of rare species (Legendre & Legendre, 1998). If there are a lot of rare species in the data then these species may only be providing information on peculiar environments of where each is found, and are therefore not as useful as the more common species for finding the main environmental gradients within the data. The environmental effects included in the analysis were distance from the ancient woodland, estimated canopy cover (arcsine square root transformed), slope (log transformed), soil depth, soil pH, easterliness (sine of the radians of a compass), northerliness (cosine of the radians of a compass), soil moisture content and soil organic content. Soil texture types were also included as dummy variables (sand, silt, clay and loam).

The sample scores were of greatest interest in this analysis, so the option 'sample scores are weighted mean species scores' was used in CANOCO. This gives sample scores that are approximate chi-square distances on the ordination axes (Legendre & Legendre, 1998). CCA was preferred over CA as it provides a more elegant extension to CA (Palmer, 1993) and provides a more powerful method in terms of relating species to environmental gradients (ter Braak & Prentice, 1988). CCA is less prone to the arch effect (Palmer, 1993) which is a result of rendering non-linear phenomenon in Euclidean space, in particular as two-dimensional plots (Legendre & Legendre, 1998; Kent & Coker, 1996). It can also help to separate out samples according to the environmental variables included in the analysis, when they might otherwise occur very close together on the ordination diagram. However, with strong unimodal responses, CA (and CCA with many environmental variables) tends to show an arch effect on the ordination diagram (ter Braak & Smilauer, 1998).

2.3.3 Trend surface analysis: Regression models

A number of techniques can be used to look at spatial patterns. These can be applied to analysis of the abundance of species across ordinations of sample scores. One of he most common methods is to use a polynomial regression equation to model a quadratic surface model.

$$\hat{y} = b_0 + b_1 X + b_2 Y + b_3 X^2 + b_4 XY + b_5 Y^2 \qquad (2.1)$$

The additional parameters X^3, X^2Y, XY^2 and Y^3 can be included to extend the equation to a cubic surface model. Legendre and Legendre (1998) recommend the backward elimination of non-significant terms in the model, until all terms in the polynomial equation are significant. This method is also appropriate for generalised linear models i.e. binomial distribution with logistic link function for presence/absence data and poisson error term with a log-link function for counts or abundance data. An over-dispersion parameter can also be included for both models so that the assumption that the data come from a single binomial or Poisson distribution is not required.

2.3.4 Kriging

Another popular method for mapping species distributions involves semi-variograms and kriging to produce smooth 3D-terrain maps. A number of different models (spherical, exponential, linear, linear to sill and Gaussian) can be used to fit the semi-variogram to the data.

Semi-variograms were plotted for each species' change in semi-variance with separation distance across the ordination of the first two axes derived from the sample scores, which are linear combinations of the environmental variables. Spherical, exponential or Gaussian equations were used in fitting the semi-variograms to the data in this study. This was conducted on Geostatistics for the Environmental Sciences – GSWIN (Gamma Design Software, 1998). The reduced sum of squares was used to evaluate the models used. The model with the smallest value was the one used.

A common problem with semi-variograms is that at greater lag distances there are fewer points to measure the difference from. This can lead to inflated semi-variance, particularly at large lag distances. The semi-variograms were plotted and the range was reduced as necessary to improve the fit. The interval distances used were also shortened when the range was reduced.

Point kriging was used with the default settings in GSWIN (a search radius of 3.38 and number of neighbours equal to 16). Point kriging uses the exact values for each point that has been given spatial coordinates. The alternative to this is block kriging, which uses the mean value for groups of points close together. The grid values and the values calculated by kriging were imported into MapInfo (MapInfo, 1996). A 3D-terrain map could then be drawn, by using rectangular interpolation, in Vertical Mapper (Northwood Geoscience, 1996). The 3D map could then be used as the background to CCA charts drawn in Excel 97 (Microsoft Corporation, 1997).

Both these types of trend surface analysis (polynomial regression and kriging) are available in CanoDraw 3.1. (Smilauer, 1992). CanoDraw uses a generalised linear model procedure to determine the polynomial regression, but does not use an over-dispersion parameter. In CanoDraw, there is also no option available to shorten the length of the range or to specify different models for semi-variance analysis prior to kriging.

2.3.5 General Linear Mixed Models

The data for the number of ancient woodland indicator species, % cover of *Hyacinthoides non-scripta* and the CCA axes were analysed using general linear mixed models which were conducted using PROC MIXED in SAS. The data for these variables contained relatively few noughts and so a more thorough analysis could be performed. If as large number of noughts is found then a more appropriate method is often to assume a Poisson error distribution using a generalised linear or generalised linear mixed model. However, neither of these two models can include dependent error structures. The data collected from the transects include quadrats which have been placed fairly close together. It is therefore probably more desirable to assume normal distributions (after a log or squareroot transformation) and use a general linear mixed model to model the spatial dependency.

The analysis was carried out in a series of steps. Distance was first treated as a categorial variable. Estimated values could then be calculated for distance interval along the transect. A test could then be done to assertain whether the estimated values changed siginificantly along the transects. The presence of significant, linear and quadratic, effects were determined using tests written in the contrast statements. Once this had been achieved, a model could be fitted using distance as a continuous variable, with a quadratic term included if the test for this had been found to be significant in the previous model. Polynomials such as using quadratic effects allows the program to model non-linear effects in a fairly simple way.

Random effects were included for the categorical variables – 'wood' and also for 'transect nested in wood'. Time series error structures were fitted for quadrats along the same transect. The time series error structures were tested for goodness of fit so that an appropriate error structure could be decided upon. This was carried out through the use of Akaikes Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) (Little *et al.*, 1996; Verbeke, 1997). These are essentially log likelihood values penalised for the number of parameters estimated. SBC imposes a heavier penalty than AIC. The covariance structure, with values of the criteria, closest to zero are considered most desirable (Little *et al.*, 1996). First-order auto-regressive AR(1), first-order auto-regressive moving average ARMA(1,1) and compound symmetry CS structures were tested for each of the models as well as random linear and quadratic slopes models. The AR(1) structure specifies that the covariance between two measurements *w* lag intervals apart is $\sigma^2\rho^w$. The parameter σ^2 stands for the variance of an observation and ρ^w for the correlation between adjacent observations on the same subject (Littell *et al.*, 1996). The CS structure specifies that all subjects (or quadrats along each transect) are equally correlated with each other. This is the same as specifying random effects for each transect, without a repeated statement. However, in the analysis used in this study the variable 'wood' was treated as a grouping variable, in the SAS syntax program (SAS Institute, 1994), so that different variances could be calculated for each wood. This meant that for the CS structure, five additional parameters were used, compared to a model with a random variable for transect nested in wood and wood used as grouping variable. The ARMA(1,1) structure is similar to the AR(1) structure, but includes an additional variable for each group of observations which models a moving average component.

The sites used in this study varied a great deal in the sizes of plant communities that existed along the transects. Old Spring Wood contained extensive homogeneous communities in contrast to both Graves Park and Little Matlock Wood. These contained much smaller patches of different species. It therefore seemed sensible to use the site variable 'wood' as a grouping variable for all the models used. This allowed the different variances and covariances to be calculated separately for each wood.

Satterthwaite approximate degrees of freedom (SAS Institute, 1994) were used instead of the default degrees of freedom for all the mixed models. This option is often required for any model with multiple error terms (Littell *et al.*, 1996). This is because the random effects may only explain part of the error and may also interfere with the fixed effects in the model. The degrees of freedom are therefore not necessarily whole numbers. The option is available in SAS to use Satterthwaite approximated degrees of freedom.

Multiple comparison tests (Tukey-Kramer) (SAS Institute, 1994) were used to test for differences between the adjusted mean values for each wood when the variable 'wood' was treated as a fixed effect. BLUPs (best linear unbiased predictors) were used to estimate differences when 'wood' was treated as a random variable. Multiple contrast statements were also used to test for differences between the regression coefficients for the relation with distance for each of the woods using t-tests.

The backward elimination of non-significant terms in the model was used to find a simple equation, to relate the number of indicators to distance and site effects. This was repeated for the logarithm of percentage cover values of selected species and the number of ancient woodland indicator species.

The procedure used to analyse the sample scores produced by CANOCO was as similar as possible to that used by the CANOCO program, except for the addition of the random site effects and time series error structures. The sample scores that were derived from just the species data and the environmental data, were all standardised by subtracting the mean and dividing by their standard deviation to give a mean of zero and variance of one as in CANOCO. The weights produced by CANOCO were also applied to the sample scores. These weights are derived from the sum of the abundance values for each quadrat and compensate for the down weighting of rare species. Random effects were used for wood and transect nested in wood and different time series error structures were fitted as in the above analysis. The differences between the estimated random values for each wood were tested by estimating the differences between best linear unbiased predictors (BLUPs) using the PROC MIXED syntax language. 'Wood' could have been treated as a fixed effect However, a large number of variables was recorded and the addition of site variables on an ordination diagrams could have made them very confusing. Sites were not selected to include all possible variation. They were selected on the occurrence of ancient woodland next to newly colonising sites. The sites have been selected from a large and unknown distribution of possible sites and therefore come under the definition of random effects.

For a site variable to be treated as a random effect, it is preferable to have a large number of different sites, probably greater than 20 (Keith Jones pers. comm.). This is because, in order to perform a test on whether there is an overall difference between the covariance parameters, a chi-square distribution is calculated from the site parameters. A large number of values are required to estimate this distribution accurately. The tests for the random effects for 'wood' presented in the results are therefore unreliable. The comparison tests between the BLUPs, however should be more reliable.

The advantage of treating 'wood' as a random variable, is that any model produced is more applicable to a larger geographic area. If 'wood' is treated a fixed effect, then all the variation of interest is assumed to occur, between only the sites sampled. This would assume that the species studied will always behave in the same way no matter what their geographic location. This is likely to be an unrealistic assumption.

2.3.6 Cluster analysis

As mentioned above, in Section 2.3.2, the sample scores determined in the analysis are approximate chi-square distances and so are appropriate for clustering. Chi-square distances behave as Euclidean distances. A coefficient is said to be Euclidean if the distances are fully embeddable in an Euclidean space; principal coordinate analysis of such a distance matrix does not produce negative eigenvalues (Legendre & Legendre, 1998). The calculation of Euclidean distances assumes the variables are uncorrelated with one another (Everitt, 1993). This criteria is also met by using the scores produced from an ordination method.

Several authors have suggested that clustering should be carried out simultaneously with ordination on a set of objects (Sneath & Sokal, 1973; Field *et al.*, 1982; and Legendre & Legendre, 1998). Clustering the sample scores from canonical correspondence analysis is particularly attractive as this represents a type of constrained clustering. A large number of different methods is available for cluster analysis. Single linkage (nearest neighbour) is stated as having the greatest appeal by Jardine and Sibson (1971). However, this method has a tendency to cluster together, at relatively low levels, individuals linked by a series of intermediates. This property known as 'chaining' often makes this method unsuitable for investigating ecological questions (Legendre & Legendre, 1998). The following clustering methods were applied to the first eight axes produced by canonical correspondence analysis – Single linkage, Average linkage, Ward's minimum variance, Flexible clustering (beta = -0.25 and beta = -0.5) and k-means clustering.

2.3.7 TWINSPAN

TWINSPAN (Hill, 1973) was used with the correction for the modified convergence criteria of Oksanen and Minchin (1997). An extension to TWINSPAN called COINSPAN (Carleton *et al.*, 1996) has also been developed, but this is not widely available as yet. COINSPAN bisects canonical correspondence analysis axes unlike TWINSPAN, which bisects correspondence axes. The default options were used in TWINSPAN i.e. cut off levels of 0, 2, 5, 10, and 20% to construct pseudospecies, as proposed by Hill (1979).

2.3.8 INDVAL

The method developed by Dufrene and Legendre (1997) using the program INDVAL was used as an alternative method to TWINSPAN to determine indicator species. Indicator species can be defined as the most characteristic species of each group, found mostly in a single group of a typology and present in the majority of the sites belonging to that group (Dufrene & Legendre, 1997). INDVAL uses a symmetrical index for the indicator values so that a species presence and abundance are given equal weighting. The index used is given below.

$$A_{ij} = \mathrm{Nindividuals}_{ij}/\mathrm{Nindividuals}_{i.} \tag{2.2}$$

$$B_{ij} = \mathrm{Nsites}_{ij}/\mathrm{Nsites}_{.j} \tag{2.3}$$

$$\mathrm{INDVAL}_{ij} = A_{ij} \times B_{ij} \times 100. \tag{2.4}$$

Where $\mathrm{Nindividuals}_{ij}$ is the mean number of individuals of species i across sites of group j, while $\mathrm{Nindividuals}_{i.}$ is the mean number of individuals of species i over all groups and INDVAL_{ij} is the indicator value of species i in cluster j. A_{ij} gives an index of abundance and B_{ij} gives an index of presence/absence. The computer program INDVAL was applied to the dendogram produced from the cluster analysis of the sample scores derived from the canonical correspondence analysis.

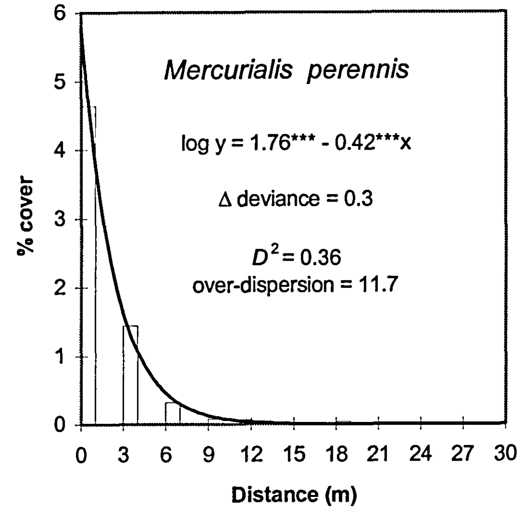
Contrary to TWINSPAN, INDVAL calculates an index value that is independent of the other species relative abundances, and there is no need for pseudospecies (Dufrene & Legendre 1997). A random reallocation procedure of sites among site groups is used to test for significance of INDVAL_i. In addition, if taxa show very similar specificity and fidelity trends, but differ in abundance, their INDVAL score remains the same, hence making comparisons across taxa robust to differences in abundance (McGeoch & Chown, 1998).

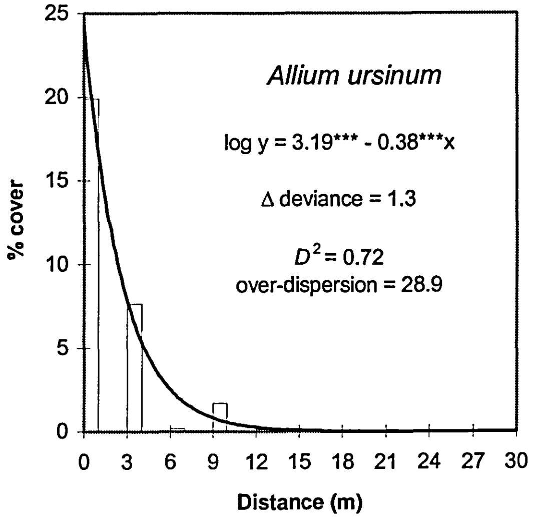
2.4 Results

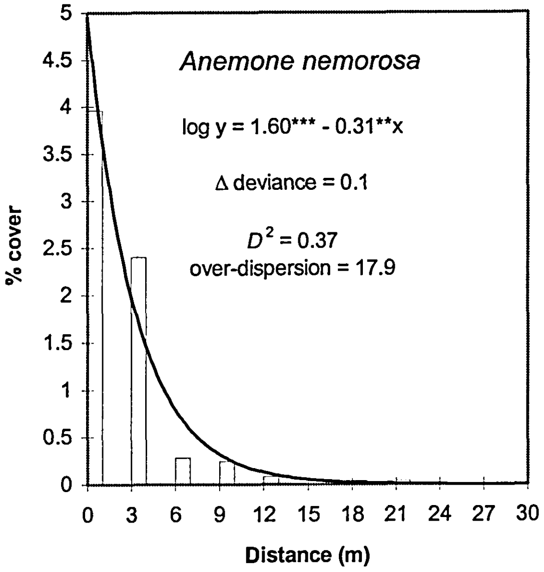
2.4.1 Colonisation rates

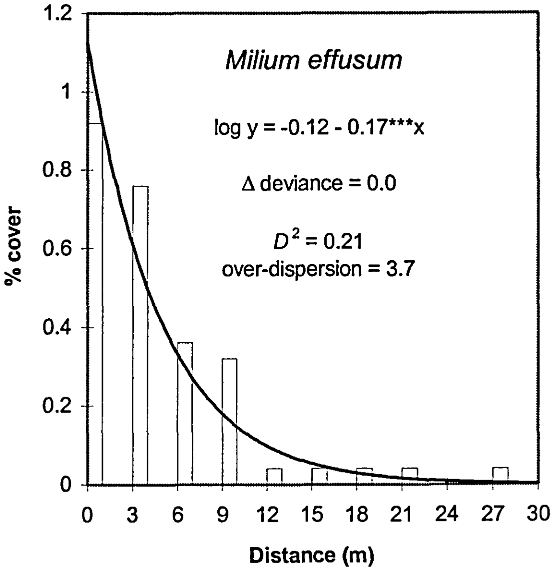
A number of species showed clear declines from the ancient woodland boundary (Figure 2.1). The fitting of models using a Poisson distribution and log-link function appeared to fit the data very well. Five of the species described by Peterken (1993) as having an affinity for ancient woodland, *Allium ursinum*, *Mercurialis perennis*, *Anemone nemorosa* and *Lamiastrum galeobdolon,* all show a clear decline from the ancient woodland boundary. *Circaea lutetiana*, *Lonicera periclymenum*, *Oxalis acetosella* and *Rubus fruticosus* all show clear quadratic relationships. *Pteridium aquilinum* and *Anthriscus sylvestris* both show a clear increase away from the ancient woodland boundary. *Hyacinthoides non-scripta* (not shown since this species will be dealt with in more detail later on) appeared as a very uniform distribution when plotted with distance with no tests proving to be significant. In order to test the response of *Hyacinthoides non-scripta* to distance from the ancient woodland a general linear mixed model was used to test for differences in slopes between the different woodlands as well as compensating for a dependent error structure.

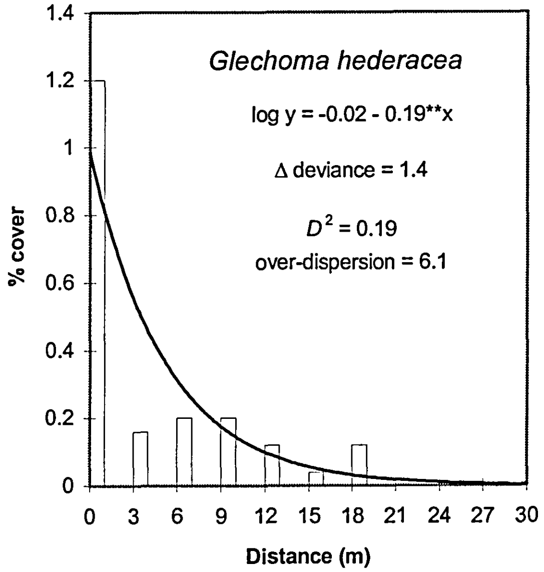
Distance was first treated as a categorical variable. The four different error structures, mentioned in the methods section, were fitted and the model fitting values produced by PROC MIXED used to evaluate the error structures. When a dependent error structure was specified, the models would not converge with the random statement included for transect nested in wood. This random statement for transect nested in wood was therefore omitted. The unstructured error structure would not converge for any of the models used. With no error structure specified, but with the random statement included the value for -2 Restricted Log Likelihood equaled 648.9 and If a chi-square distribution is assumed then the difference between these values and the models with an AR(1) is significant at $P < 0.001$ with 9 degrees of freedom and ARMA(1,1) with 14 degrees of freedom is significant at $P < 0.001$ ant the CS significant at $P < 0.001$ with 4 degrees of freedom. The AR(1) error structure can be seen to give the best fit in Table 2.2 since the values given for AIC and SBC are closest to zero.

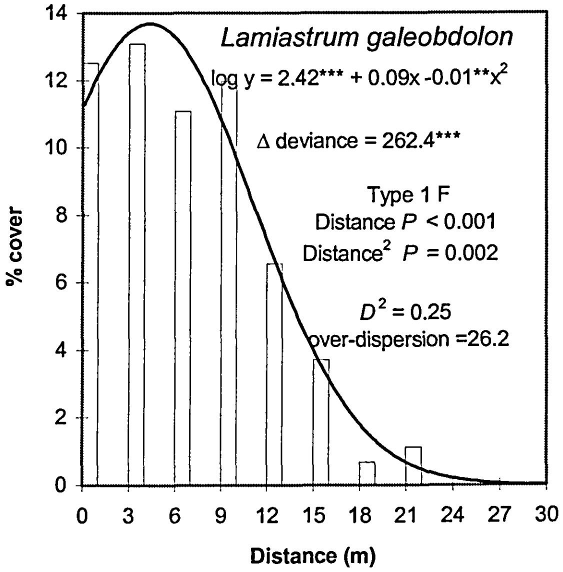

Mercurialis perennis
log y = 1.76*** - 0.42***x
Δ deviance = 0.3
D² = 0.36
over-dispersion = 11.7
% cover
0
1
2
3
4
5
6
0 3 6 9 12 15 18 21 24 27 30
Distance (m)


Allium ursinum
log y = 3.19*** - 0.38***x
Δ deviance = 1.3
D² = 0.72
over-dispersion = 28.9
% cover
0
5
10
15
20
25
0 3 6 9 12 15 18 21 24 27 30
Distance (m)


Anemone nemorosa
log y = 1.60*** - 0.31**x
Δ deviance = 0.1
D² = 0.37
over-dispersion = 17.9
% cover
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0 3 6 9 12 15 18 21 24 27 30
Distance (m)


Milium effusum
log y = -0.12 - 0.17***x
Δ deviance = 0.0
D² = 0.21
over-dispersion = 3.7
% cover
0
0.2
0.4
0.6
0.8
1
1.2
0 3 6 9 12 15 18 21 24 27 30
Distance (m)


Glechoma hederacea
log y = -0.02 - 0.19**x
Δ deviance = 1.4
D² = 0.19
over-dispersion = 6.1
% cover
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0 3 6 9 12 15 18 21 24 27 30
Distance (m)


Lamiastrum galeobdolon
log y = 2.42*** + 0.09x -0.01**x²
Δ deviance = 262.4***
Type 1 F
Distance P < 0.001
Distance² P = 0.002
D² = 0.25
over-dispersion =26.2
% cover
0
2
4
6
8
10
12
14
0 3 6 9 12 15 18 21 24 27 30
Distance (m)

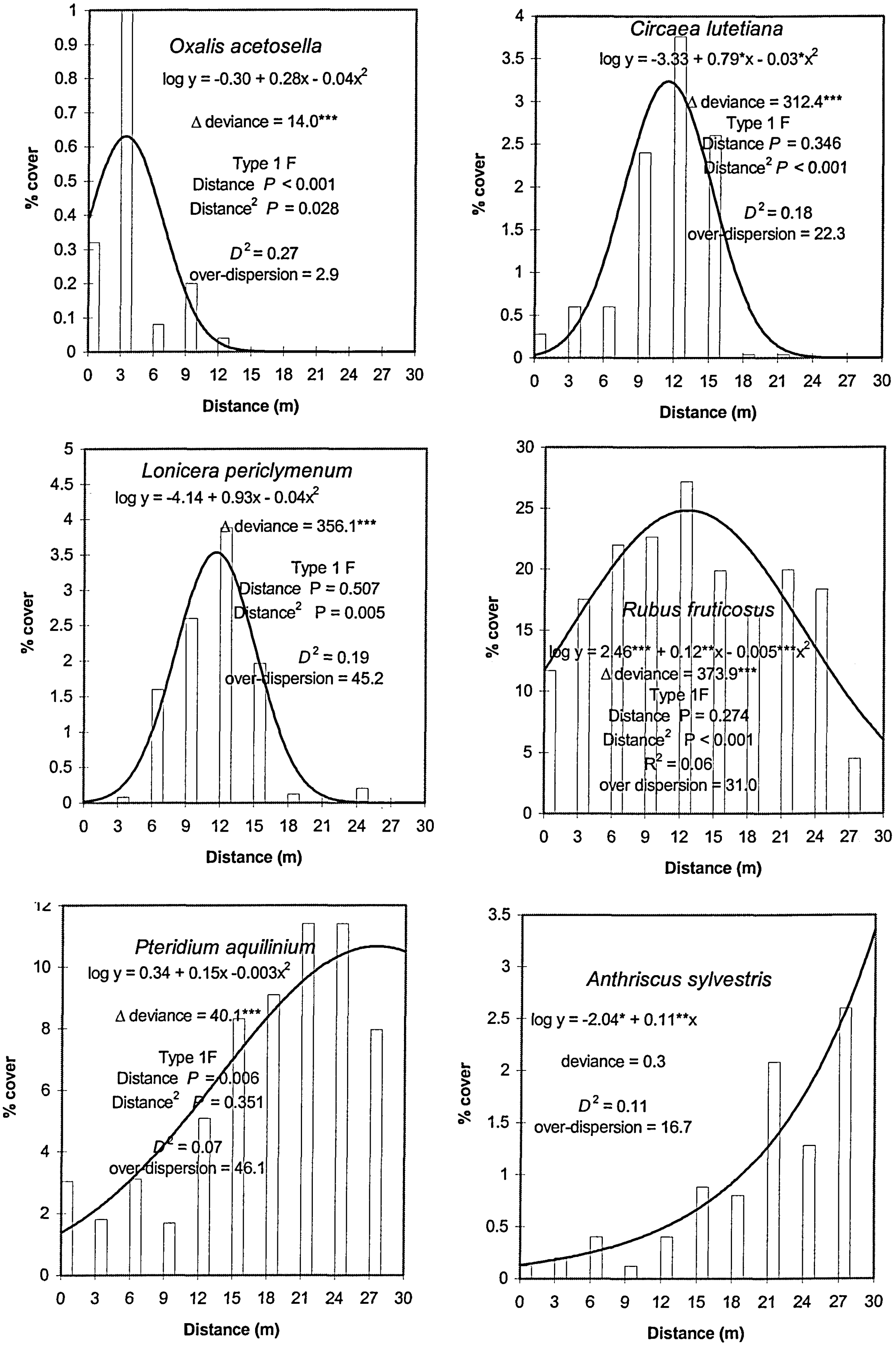


Figure 2.1. Poisson regression models and mean % covers of selected species that occurred along the transects. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. D^2 = proportion of total deviance explained.

Table 2.2. Determining an appropriate covariance structure for % cover of *Hyacinthoides non-scripta* (log) with Wood, Distance and Wood × Distance as the independent variables. Distance was treated as a categorical variable. Values closest to zero are in **bold**.

Description	AR(1)	ARMA(1,1)	CS
Restricted Log Likelihood	-218.0	**-215.6**	-239.9
Akaike's Information Criterion	**-228.0**	-230.6	-249.9
Schwarz's Bayesian Criterion	**-244.5**	-255.4	-266.3
-2 Restricted Log Likelihood	436.0	**431.3**	479.7

The tests of the fixed effects are given in Table 2.3. This table indicates that there is a significant difference between woods and the response of *Hyacinthoides non-scipta* between the different woods.

Table 2.3. Tests of fixed effects with *Hyacinthoides non-scipta* as the dependent variable and distance the interaction term for distance and wood treated as independent variables. Distance was treated as a categorical variable.

Source	NDF	DDF	Type III F
Wood	4	14.5	29.91 ***
Distance	9	85.1	1.36
Distance × Wood	36	53.4	2.04 **

Contrast statements were used in PROC MIXED to test for the presence of linear and quadratic responses of *Hyacinthoides non-scripta* with distance from the ancient woodland as an overall effect with distance (Table 2.4) and with separate tests for each wood (Table 2.5). In order to carry out these tests the model statement used to produce Table 2.4 included effects for wood, distance and the interaction term wood × distance. In order to produce the results in Table 2.4 the model statement included effects for wood and distance nested in wood. *Hyacinthoides non-scripta* was found to show a significant quadratic relation with distance for Old Spring Wood and a decline from the ancient woodland boundary in Smithy Wood.

Table 2.4. Results of contrast statements used to test for overall responses of *Hyacinthoides non-scripta* with distance from the ancient woodland.

Source	NDF	DDF	F
Linear	1	62.7	0.08
Quadratic	1	106	2.82

Table 2.5. Results of contrast statements used to test the response of *Hyacinthoides non-scripta* with distance from the ancient woodland at each site using the nested model i.e. distance nested in wood as one of the fixed effects in the model statement

Source	NDF	DDF	F
linear GP	1	21.2	1.48
quadratic GP	1	27.5	0.30
linear LMW	1	16.1	0.32
quadratic LMW	1	26	1.29
linear NW	1	22.5	0.02
quadratic NW	1	39.9	0.00
linear OSW	1	17.6	0.97
quadratic OSW	1	38.4	16.24 ***
linear SW	1	14	4.54 *
quadratic SW	1	24.1	0.92

A regression model treating distance as a continuous variable was also fitted to the data. The tests of the random effects are given in Table 2.6. Transect nested in wood was included as a random variable and the AR(1) error structure, grouped by wood, was used in the repeated statement. Transect can be seen to have a very small effect. However, much of the differences between transects may occur within the correlated error structures. The variances and correlations for the AR(1) structure can be seen to differ considerably between the different sites. Old Spring Wood and Norwood can be seen to contain very highly correlated data which can be compared to Graves Park which does not show a significant correlated error structure.

The solution for the fixed effects is given in Table 2.7. The only linear and quadratic terms that are significant are those for Old Spring Wood. Type III F tests are given in Table 2.8. All variables can be seen to be significant. Adjusted mean values and back transformed values are given in Table 2.9. Multiple comparison tests between adjusted means for all the sites are given in Table 2.10. Old Spring wood contained a significantly higher % cover of *Hyacinthoides non-scripta* than Graves Park, Little Matlock Wood and Norwood. Smithy Wood contained a significantly higher % cover *of Hyacinthoides non-scripta* than Graves Park and Little Matlock Wood. Multiple *t*-tests using the estimate statement in PROC MIXED were used test for differences between the slopes for each wood. The results are given in Table 2.11. The only significant differences found were those comparisons, which included Old Spring Wood.

Table 2.6. Covariance parameter estimates with *Hyacinthoides non-scripta* as the dependent variable and distance treated as a continuous variable.

Covariance parameter	Wood	Estimate
Transect (nested in wood)		0.00
Variance	GP	0.01 ***
AR(1)	GP	-0.05
Variance	LMW	0.63 ***
AR(1)	LMW	0.38 **
Variance	NW	2.37 *
AR(1)	NW	0.78 ***
Variance	OSW	1.57 **
AR(1)	OSW	0.73 ***
Variance	SW	1.72 ***
AR(1)	SW	0.45 **

Table 2.7. Solution for fixed effects. Testing unequal slopes with *Hyacinthoides non-scripta* as the dependent variable and distance treated as a continuous variable.

Effect	WOOD	DF	Estimate
Intercept		21.6	3.005 ***
Wood	GP	21.8	-3.003 ***
Wood	LMW	35.7	-3.119 ***
Wood	NW	30.3	-2.232 *
Wood	OSW	35.0	-2.125 *
Wood	SW	.	0.000
Distance		29.0	-0.136
Distance × Wood	GP	29.2	0.134
Distance × Wood	LMW	47.7	0.209
Distance × Wood	NW	68.7	0.167
Distance × Wood	OSW	62.2	0.538 ***
Distance × Wood	SW	.	0.000
Distance × Distance		30.6	0.003
Distance × Distance × Wood	GP	30.8	-0.003
Distance × Distance × Wood	LMW	50.3	-0.005
Distance × Distance × Wood	NW	71.8	-0.003
Distance × Distance × Wood	OSW	65.3	-0.015 ***
Distance × Distance × Wood	SW	.	0.000

Table 2.8. Type III F tests for the fixed effects. Testing unequal slopes with *Hyacinthoides non-scripta* as the dependent variable and distance treated as a continuous variable.

Source	NDF	DDF	Type III F
Wood	4	12.0	7.33 **
Distance	1	64.9	5.01 *
Distance × Distance	1	133	4.90 *
Distance × Wood	4	43.5	7.23 ***
Distance × Distance × Wood	4	77.0	6.15 ***

Table 2.9. Adjusted means for each wood. Testing unequal slopes with *Hyacinthoides non-scripta* as the dependent variable and distance treated as a continuous variable.

Wood	Adjusted means	Std error	Back transformed adjusted means
GP	0.01	0.01	0.0
LMW	0.27	0.16	0.3
NW	1.13	0.49	2.1
OSW	3.10	0.37	21.2
SW	1.87	0.28	5.5

Table 2.10. Multiple comparison tests for woods (Probability test – Tukey-Kramer) with *Hyacinthoides non-scripta* as the dependent variable and distance treated as a continuous variable.

Wood	_Wood	DF	Difference
GP	LMW	10.9	-0.25
GP	NW	5.4	-1.12
GP	OSW	5.1	-3.09 ***
GP	SW	8.2	-1.86 ***
LMW	NW	6.5	-0.86
LMW	OSW	7.0	-2.83 ***
LMW	SW	13.2	-1.60 ***
NW	OSW	9.9	-1.97 *
NW	SW	8.8	-0.74
OSW	SW	10.4	1.23

Table 2.11. Multiple t-tests comparing the slopes produced by PROC MIXED for *Hyacinthoides non-scripta*.

Parameter		DF	Estimate	t
Distance	b1(GP) - b2(LMW)	31.1	-0.075	-1.32
Distance	b1(GP) - b3(NW)	43.2	-0.032	-0.35
Distance	b1(GP) - b4(OSW)	39.5	-0.404	-5.01 ***
Distance	b1(GP) - b5(SW)	29.2	0.134	1.42
Distance	b2(LMW) - b3(NW)	67.5	0.042	0.39
Distance	b2(LMW) - b4(OSW)	66.3	-0.329	-3.35 **
Distance	b2(LMW) - b5(SW)	47.7	0.209	1.90
Distance	b3(NW) – b4(OSW)	81.2	-0.371	-3.03 **
Distance	b3(NW) – b5(SW)	68.7	0.167	1.26
Distance	b4(OSW) - b5(SW)	62.2	0.538	4.34 ***
Distance × Distance	b1(GP) - b2(LMW)	32.4	0.002	1.29
Distance × Distance	b1(GP) - b3(NW)	45.8	0.000	0.13
Distance × Distance	b1(GP) - b4(OSW)	42.7	0.013	4.72 ***
Distance × Distance	b1(GP) - b5(SW)	30.8	-0.003	-0.84
Distance × Distance	b2(LMW) - b3(NW)	72.1	-0.002	-0.58
Distance × Distance	b2(LMW) - b4(OSW)	71.6	0.010	3.10 **
Distance × Distance	b2(LMW) - b5(SW)	50.3	-0.005	-1.39
Distance × Distance	b3(NW) – b4(OSW)	86.8	0.012	3.02 **
Distance × Distance	b3(NW) – b5(SW)	71.8	-0.003	-0.70
Distance × Distance	b4(OSW) - b5(SW)	65.3	-0.015	-3.69 ***

The list of species described by Peterken (1993) as having a strong or mild affinity for ancient woodland, were analysed in the same way as the previous section on *Hyacinthoides non-scipta*. The log of the number of indicator species found in each quadrat, were used in the analysis. The same approach was used as the % cover of *Hyacinthoides non-scripta*. Distance was first treated as a categorical variable and the different error structures evaluated using the model fitting statistics produced by PROC MIXED. The unstructured error structure would again not converge. With no covariance structure specified the value for -2 Restricted Log Likelihood equaled 200.9, for the number of ancient woodland indicator species. If a chi-square distribution is assumed then the difference between these values and the models with an AR(1) is significant at $P < 0.001$ with 10 degrees of freedom and ARMA(1,1) with 15 degrees of freedom is significant at $P < 0.01$ ant the CS significant at P < 0.01 with 5 degrees of freedom. These models all included a random variable for transect nested in wood.

Table 2.12. Model fitting statistics produced by PROC MIXED. Number of ancient woodland indicator species (log) with Wood and Distance and Wood × Distance as the independent variables. Distance was treated as a categorical variable.Values closet to zero are in **bold.**

Description	AR(1)	ARMA(1,1)	CS
Akaike's Information Criterion	**-95.5**	-96.9	-101.5
Schwarz's Bayesian Criterion	**-113.7**	-123.3	-118.0
-2 Restricted Log Likelihood	169.0	**161.8**	183.0

The AR(1) structure gives values for AIC and SBC closer to zero and so this structure was used in fitting the models (Table 2.12). The Type III F Tests for the fixed effects is given in Table 2.13. Linear and quadratic tests were conducted using contrast statements in PROC MIXED as described previously. The models with the AR(1) and ARMA(1,1) error structures produced very similar results with the same effects being significant. However, the model with the CS error structure and the model without a repeated statement did give significant interaction effects for the Distance × Wood effect, $P = 0.03$ and $P = 0.01$ respectively.

The results of the contrast statements using the crossed model and the nested model are given in Table 2.14 and Table 2.15 respectively. A strong linear relationship can be seen to occur. All the woods show strong linear relationships with two of the sites also giving significant quadratic relationships. However, on examination of the adjusted means for the number of woodland indicators with distance in each wood, the two woods showing quadratic relationships, show no indicator species after about halfway along the transect giving an inverted regression curve.

Table 2.13. Type III F tests for the fixed effects. Unequal slopes model with the log of the number of woodland indicators as the dependent variable.

Source	NDF	DDF	Type III F
Wood	4	15.2	4.82 **
Distance	9	121	11.35 ***
Distance × Wood	36	49	1.54

Table 2.14. Results of contrast statements to test for the overall relationship with the number of woodland indicators and distance from the ancient woodland using a crossed model.

Source	NDF	DDF	F
Distance linear	1	62.7	93.38 ***
Distance quadratic	1	113	1.91

Table 2.15. Results of contrast statements used to test the response of the number of woodland indicator species with distance from the ancient woodland at each site using a nested model.

Source	NDF	DDF	F
linear GP	1	14.0	11.20 **
quadratic GP	1	22.1	7.20 *
linear LMW	1	15.9	24.37 ***
quadratic LMW	1	25.1	0.97
linear NW	1	17.0	44.75 ***
quadratic NW	1	24.6	5.74 *
linear OSW	1	15.3	21.40 ***
quadratic OSW	1	28.3	2.28
linear SW	1	14.5	10.36 **
quadratic SW	1	25.5	0.00

Since the overal quadratic effect shown in Table 2.14 is not significant, a new model was fitted to the data treating distance as a continuous variable with only the linear term included. A common slope model was fitted to the data i.e. only two independent variables - wood and distance with no interaction term. The covariance parameters and associated significance values are given in Table 2.16 and the regression equation given in Table 2.17. The tests for the fixed effects are given in Table 2.18. The residuals produced from this model were analysed using the standard procedure of plotting them against the predicted values and producing a histogram of the values. PROC UNIVARIATE was used to produce the histogram and conduct a test of deviance from a normal distribution using a Shapiro-Wilk statistic (SAS Institute, 1994; Zar, 1996). The model produced excellent model fitting diagnostics. A random scatter of points was observed for the plot of the residuals and predicted values and a symmetrical histogram was produced for the residuals. The Shapiro-Wilk test gave a value for W of 0.98, which was not significantly different from a normal distribution. This indicates that if any interaction between wood and distance does exist it is very small.

Table 2.16. Covariance parameter estimates with the number of woodland indicators as the dependent variable and distance as a continuous variable.

Covariance Parameter	Wood	Estimate
Transect (nested in wood)		0.00
Variance	GP	0.09 ***
AR(1)	GP	0.44 **
Variance	LMW	0.11 ***
AR(1)	LMW	0.31 *
Variance	NW	0.08 ***
AR(1)	NW	0.30 *
Variance	OSW	0.14 ***
AR(1)	OSW	0.53 ***
Variance	SW	0.18 ***
AR(1)	SW	0.38 **

Table 2.17. Solution for fixed effects with the number of woodland indicators as the dependent variable and distance as a continuous variable.

Effect	WOOD	DF	Estimate
Intercept		17.5	0.848 ***
Wood	GP	18.8	-0.234 *
Wood	LMW	21.9	-0.084
Wood	NW	20.0	-0.110
Wood	OSW	20.8	0.111
Wood	SW	.	0.000
Distance		86.6	-0.029 ***

Table 2.18. Type III F tests for the fixed effects with the number of woodland indicators as the dependent variable and distance as a continuous variable.

Source	NDF	DDF	Type III F
Wood	4	17.8	2.93 *
Distance	1	86.6	91.24 ***

Table 2.19. Adjusted means for the number of woodland indicators at each wood. A common slope model with the number of woodland indicators as the dependent variable and distance treated as a continuous variable.

Wood	Adjusted means	Std error	Back transformed adjusted means
GP	0.20	0.06	0.23
LMW	0.35	0.06	0.42
NW	0.33	0.05	0.39
OSW	0.55	0.09	0.73
SW	0.44	0.08	0.55

A multiple comparison test was conducted on the adjusted mean values for the differences between each of the woods to find where any significant differences occurred (Table 20). The only significant difference was found to occur between Graves Park (GP) and Old Spring Wood (OSW).

Table 2.20. Multiple comparison tests between the adjusted means for the number of woodland indicators at each wood given in Table 18 using a Tukey-Kramer test.

Wood	_Wood	DF	Difference
GP	LMW	19.1	-0.15
GP	NW	17.3	-0.12
GP	OSW	17.0	-0.34 *
GP	SW	18.8	-0.23
LMW	NW	24.6	0.03
LMW	OSW	18.0	-0.20
LMW	SW	21.9	-0.08
NW	OSW	16.1	-0.22
NW	SW	20.0	-0.11
OSW	SW	20.8	0.11

2.4.2 Differences between the sites surveyed

All the variables recorded showed significant differences between the sites using a One-Way ANOVA (Table 2.21). In particular, soil depth shows considerable difference between the sites. Other large differences can be seen for the soil texture categories and the tree species planted in the newly wooded sites.

Table 2.21. Means and standard errors for the environmental variables recorded at quadrats in each. F values are for 1 Way ANOVA tests. Canopy cover refers to the arcsine square root transformation of this variable. The variables for trees from oak to elder refer to the square root of the number of trees + 3/8 in 100m^2 quadrats.

Variable	GP		LMW		NW		OSW		SW		ANOVA		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	NDF	DDF	F
Canopy cover	**1.43**	0.08	**1.01**	0.47	**1.41**	0.20	**1.37**	0.08	**0.87**	0.36	4	245	41.4 *
Log slope	**2.49**	0.47	**2.25**	0.86	**2.62**	0.85	**1.40**	0.62	**2.90**	0.49	4	245	35.3 *
Easterliness	**0.49**	0.38	**-0.10**	0.53	**-0.48**	0.80	**-0.37**	0.58	**0.66**	0.14	4	245	46.1 *
Northerliness	**0.01**	0.80	**0.65**	0.54	**-0.14**	0.35	**0.52**	0.53	**-0.68**	0.29	4	245	50.5 *
PH	**5.45**	0.74	**4.40**	0.23	**4.71**	0.88	**4.52**	1.28	**4.28**	0.27	4	245	17.5 *
Soil Depth	**36.00**	4.95	**32.62**	16.18	**150.00**	0.00	**300.00**	0.00	**175.00**	0.00	4	245	10755.0 *
Water	**0.32**	0.04	**0.36**	0.07	**0.28**	0.06	**0.23**	0.06	**0.29**	0.07	4	245	31.7 *
Organic matter	**0.06**	0.01	**0.08**	0.02	**0.06**	0.01	**0.04**	0.01	**0.07**	0.01	4	245	37.3 *
Sand	**0.00**	0.00	**1.00**	0.00	**0.80**	0.40	**0.76**	0.43	**0.60**	0.50	4	245	61.1 *
Silt	**0.00**	0.00	**0.00**	0.00	**0.12**	0.32	**0.00**	0.00	**0.00**	0.00	4	245	6.7 *
Clay	**0.00**	0.00	**0.00**	0.00	**0.00**	0.00	**0.24**	0.43	**0.00**	0.00	4	245	15.5 *
Loam	**1.00**	0.00	**0.00**	0.00	**0.08**	0.27	**0.00**	0.00	**0.40**	0.50	4	245	142.3 *
Oak	**0.61**	0.00	**0.84**	0.37	**1.25**	0.92	**0.61**	0.00	**0.75**	0.31	4	245	16.1 *
Sycamore	**1.89**	0.60	**0.63**	0.11	**1.11**	0.69	**1.05**	0.51	**0.61**	0.00	4	245	60.7 *
Beech	**0.89**	0.41	**0.63**	0.11	**0.61**	0.00	**2.42**	0.69	**0.61**	0.00	4	245	233.0 *
Ash	**0.63**	0.11	**0.61**	0.00	**0.77**	0.25	**0.61**	0.00	**0.99**	0.55	4	245	17.7 *
Birch	**0.61**	0.00	**0.61**	0.00	**0.61**	0.00	**1.28**	0.69	**0.61**	0.00	4	245	47.4 *
Pine	**0.61**	0.00	**0.61**	0.00	**0.61**	0.00	**1.19**	0.77	**0.61**	0.00	4	245	28.4 *
Hawthorn	**0.61**	0.00	**1.42**	0.60	**1.77**	0.90	**0.61**	0.00	**0.69**	0.19	4	245	60.5 *
Elder	**0.74**	0.35	**1.10**	0.59	**0.83**	0.38	**0.61**	0.00	**0.80**	0.42	4	245	10.2 *

2.4.3 Calculating colonisation rates

Average and maximum colonisation distances from the ancient woodland boundary were calculated for the main indicator species i.e. those which exhibited a clear decline from the ancient woodland boundary. Three methods were used. The first used presence/absence data and followed the procedure used by Honnay *et al.* (1999) (Table 2.22). The second method used values weighted by the % cover as given in equation below (Table 2.23). The third method used log % cover value and is also shown in Table 2.23.

$$\text{Weighted mean distance} = \frac{\text{distance}_i \ \text{x} \ \%\,\text{cover}_i}{\sum \%\,\text{cover}_i} \qquad (2.5)$$

There is a reasonable agreement between the three methods used. The species showing the slowest rates of spread from Table 2.22 and Table 2.23 are *Allium ursinum*, *Mercurialis perennis* and *Oxalis acetosella*. *Lonicera periclymenum* can be seen to have the quickest rate of spread for all methods used. The main disagreement is for *Anemone nemorosa,* which can be seen to vary from 7 - 28 m/century depending upon which method is used.

Table 2.22. Average and maximal colonisation distances and rates from the ancient woodland boundary for selected woodland speices using the method described by Honnay et al. (1999). d_{max}: Maximal colonisation distance; d_{av}:Average colonisation distance; $d_{max\ i}$: Maximal colonisation distance in transect i; $d_{av\ i}$: Average colonisation distance in transect i; t_i: Time since site was abandoned; n: Number of transects in which species occurred.

Woodland Species	no transects	Maximal distance d_{max} (m)	Average distance d_{av} (m)	Max. col. rate Per transect m/century 1/n $\Sigma d_{max\ i}/t_i$ (m/century)	Av. col. rate per transect m/century 1/n $\Sigma d_{av\ i}/t_i$ (m/century
Allium ursinum	6	22	6	25	12
Anemone nemorosa	6	22	9	36	28
Circea lutetiana	9	22	9	32	25
Glechoma hederacea	4	19	10	29	24
Lamiastrum galeobdolon	16	22	8	29	19
Lonicera periclymenum	5	25	13	49	45
Mercurialis perennis	6	10	5	19	15
Milium effusum	10	28	8	31	26
Oxalis acetosella	8	13	6	13	12

Table 2.23. Calculation of distances from the ancient woodland boundary and rates using values weighted by the % cover and the log of the % cover for each quadrat.

Woodland Species	no. quadrats	Average distance Weighted by % cover (m)	Average distance Weighted by ln % cover (m)	Average distance weighted by % cover (m/century)	Average dista weighted by ln cover (m/centu
Allium ursinum	16	2	3	5	10
Anemone nemorosa	12	3	6	7	17
Circaea lutetiana	16	11	10	33	28
Glechoma hederacea	9	4	8	12	22
Lamiastrum galeobdolon	60	7	8	19	22
Lonicera periclymenum	10	12	12	33	35
Mercurialis perennis	11	5	3	15	9
Milium effusum	19	2	7	5	19
Oxalis acetosella	12	3	4	9	12

2.4.4 The estimated canopy cover

A nonlinear model was fitted to the estimated % canopy cover and the tree density so that the estimated values could be calibrated (Figure 2.2). A negative exponential model was used with parameters included for the intercept, the maximum % cover and the slope of the curve. The area used for counting the tree density was fairly small (10 × 10 m) so that trees outside this area often added to the canopy cover so the value of zero was not assumed for no trees present within the specified area surrounding the quadrat. It should be noted that the model should only be used for the same specified area if repeated for comparison reasons.

The arcsine square root of the proportion canopy was used to help normalize the estimated canopy values and the square root of the tree density plus three eighths was used to normalize the tree density values. The addition of the three eighths in the latter transformation helps stabilize the variance (Zar, 1996).

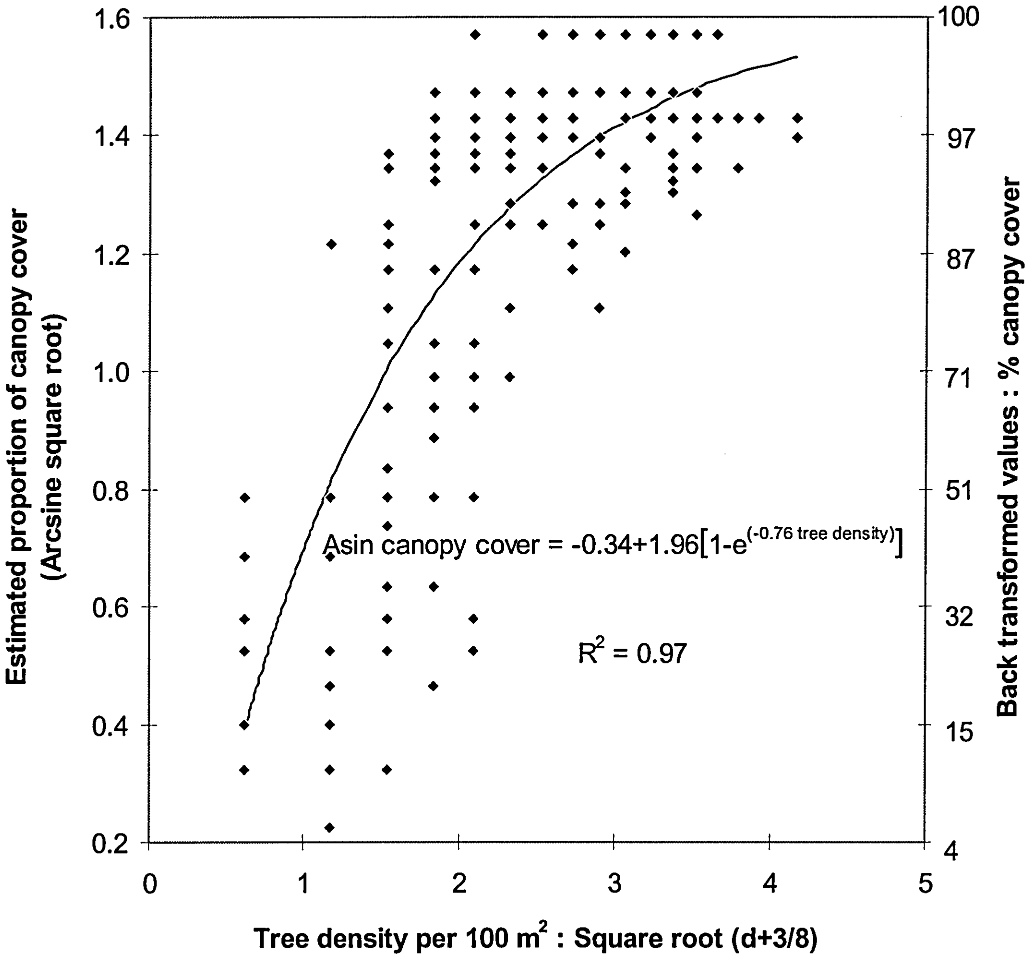


Figure 2.2. The relationship between the values for the % canopy cover estimated by eye and the tree density surrounding the quadrats. A negative exponential curve has been fitted to the data. (Asin = arcsine square root)

2.4.5 Results from the ordinations

A number of different combinations of environmental variables together with interactions was tested. This was as well as using correspondence analysis and detrended correspondence analysis (by segments) and detrended canonical correspondence analysis (by polynomials).

If interactions were included in the analysis, then the ordinations became uninterpretable due to the large number of significant interactions. This resulted in arrows incorporating the same variable pointing in different directions close to other apparently important variables. In this situation a species could not be said to be associated with a particular environmental variable. The addition of variables for the proportion of different tree species in the canopy, also resulted in too many arrows on the ordination diagrams for the ordinations to be easily interpretable. The soil types were finally simplified by dropping the variable for sandy loam and adding this variable to the variable for sand. This was due to sand on its own being strongly correlated with soil depth, which was also highly correlated with the clay-rich soils.

The ordinations produced by CA gave a very close agreement to those produced by CCA. There is some indication of an arch effect using CA and CCA when plotting the first two axes. CA appeared to be more sensitive to outliers where as CCA showed clearer clusters of sample scores on the ordination. When detrending was used with 2-4 polynomials for CCA very few differences occurred. Very little difference was also observed by using DCA. The ordinations produced from CA and DCA both produced solutions showing the first two axes to be mainly correlated with canopy cover and soil depth and the third and forth axes separating the quadrats on the poorly drained silty soils from the rest.

Descriptions of the first four axes produced by some of the different methods available in CANOCO are given in Table 2.24. The eigenvalues for the first two axes are high. Ter Braak (1988) suggests that eigenvalues greater than 0.4 should show strong unimodality of species across the CCA axes. CCA and DCCA have higher correlation coefficients than CA and DCA but they do have smaller eigenvalues. CCA can be seen to have higher or equal eigenvalues and correlation coefficients than DCCA for all but the third axis.

Table 2.24. Comparison of the results of ordinations by correspondence analysis (CA), detrended correspondence analysis (DCA), canonical correspondence analysis (CCA), and detrended canonical correspondence analysis (DCCA) Descriptions of the first four axes produced by CANOCO. DCA detrending by segments, DCCA detrending by second order polynomials. Eigenvalues and species-environment correlation coefficents are given for the first four axes. Total inertia (sum of all eigenvalues) = 5.66.

	Axis			
	1	2	3	4
		Eigenvalues		
CA	0.59	0.55	0.44	0.36
DCA	0.59	0.47	0.24	0.20
CCA	0.47	0.44	0.18	0.16
DCCA	0.47	0.34	0.16	0.04
		Correlation coefficients		
CA	0.87	0.89	0.53	0.58
DCA	0.87	0.89	0.68	0.49
CCA	0.91	0.89	0.73	0.76
DCCA	0.91	0.86	0.79	0.55

Three different time series structures were fitted to the data - first order auto-regressive AR(1), first order moving average ARMA(1,1) and compound symmetry CS. Wood was used as a grouping variable so that different estimates for each structure could be calculated separately for each wood. The different structures were evaluated in terms of their goodness of fit using Akaikes Information Criterion and Schwarz's Bayesian Criterion (Table 2.25). A larger table is given in Appendix 1B giving the model fitting information for a number of different random effects including models with random slopes for each transect. All the first order auto-regressive error structure have values that are closer to zero for Schwarz's Bayesian Criterion than the other two error structures. Akaikes Information Criterion gives lower values for the AR(1) structure for axis 3 and equal to the ARMA(1,1) for axis 1. The ARMA(1,1) receives values that are closer to zero for axes 2 and 4 from Schwarz's Bayesian Criterion. The compound symmetry structure can therefore be considered to be inappropriate for these data, however there is little to choose between the AR(1) and ARMA(1,1) structures. All three models gave very similar estimates for the environmental variables. Only the model incorporating the AR(1) structure is presented. The covariance parameter estimates are given in Table 2.26 and the regression equation for the environmental variables is given in Table 2.27.

Table 2.25. Comparison of three different error structures - First order autoregressive error structure - AR(1), First order moving average structure - ARMA(1,1) and Compound symmetry -CS on the goodness of fit statistics produced by PROC MIXED (Akaikes Information Criterion - AIC and Schwarz's Bayesian Criterion – SBC).

CCA Axis	AR(1)		ARMA(1,1)		CS	
	AIC	SBC	AIC	SBC	AIC	SBC
Axis 1	**-127.5**	**-148.3**	**-127.5**	-157.0	-139.0	-159.8
Axis 2	-101.6	**-122.5**	**-97.2**	-126.7	-133.8	-151.2
Axis 3	**-190.6**	**-211.5**	-193.4	-223.0	-202.2	-223.1
Axis 4	-224.2	**-245.0**	**-221.1**	-250.6	-250.3	-253.8

Table 2.26. Covariance parameter estimates produced the GLMM analyses of the first four axes produced by canonical correspondence analysis.

Covariance Parameter	Subject	Axis 1	Axis 2	Axis 3	Axis 4
Wood		0.07	0.11	0.55	0.28
Transect (nested in wood)		0.01	0.00	0.02	0.00
Variance	GP	0.74 **	0.98 **	3.08 ***	4.35 **
AR(1)	GP	0.38 *	0.61 ***	0.40 *	0.65 ***
Variance	LMW	3.98 **	0.73 ***	0.80 **	1.75 **
AR(1)	LMW	0.56 ***	0.04	0.42 *	0.43 *
Variance	NW	1.42 ***	2.82 **	4.63	6.35 **
AR(1)	NW	0.44 **	0.73 ***	0.75 ***	0.45 *
Variance	OSW	0.32 **	1.58 *	3.37 *	2.75 **
AR(1)	OSW	0.58 ***	0.77 ***	0.55 *	0.64 ***
Variance	SW	2.95 **	1.17 ***	1.29 ***	1.99 ***
AR(1)	SW	0.52 ***	0.57 ***	0.20	0.25

Table 2.27. Regression coefficients estimated by weighted GLMMs of standardised variables with random effects for transect nested in wood and wood. Different first order auto-regressive error structures were fitted for each wood. * P < 0.5, ** P < 0.01, *** P < 0.001. Estimated values greater than 0.1 are in bold type.

Variable	Axis 1		Axis 2		Axis 3		Axis 4	
	DF	Estimate	DF	Estimate	DF	Estimate	DF	Estimate
Distance	32.5	**0.12** **	50.3	**-0.10** **	14.3	0.09	39.0	**0.18** **
Canopy cover	19.9	**-0.58** ***	24.8	0.06	7.5	**0.21** **	11.7	-0.04
log slope	82.2	0.04	72.3	-0.02	21.7	0.08 *	46.8	-0.04
Easterliness	55.2	-0.04	38.2	-0.06	7.5	0.01	16.8	-0.03
Northerliness	54.7	-0.06 *	10.8	0.02	1.4	0.05	8.6	-0.07
pH	74.2	**-0.11** **	96.1	0.07	62.2	0.04	116	**-0.13**
Soil Depth	0.1	0.05	0.1	**0.77**	0.01	-0.05	0.1	**0.16**
Water	97.5	0.04	95.9	**0.11** **	35.2	0.01	69.5	-0.04
Organic matter	95.9	0.04	59.1	-0.06	9.8	-0.01	25.8	0.04
Sand	7.0	0.02	8.4	0.05	6.9	-0.04	16.1	0.01
Silt	13.0	-0.02	29.9	-0.06	21.3	0.05	9.3	**0.35** ***
Clay	13.2	0.04	33.4	-0.02	26.7	**-0.25** **	51.7	-0.06
Loam	.	0.00	.	0.00	.	0.00	.	0.00

The first axis appears to mainly represent the combined effects of canopy cover and distance from the ancient woodland. The second axis appears to be highly correlated with soil depth. However, since the soil depth is confounded with the random site effect and interfered with by the auto-regressive error structures only 0.1 degrees of freedom have been allocated to this variable. The third and fourth axes appear to mainly represent variation in the soil texture. The third axis separates out the samples on the clay-rich soils and the fourth axis separates out the samples on the poorly-drained silty soils. If wood is treated as a fixed effect the results are very similar. The main difference is that soil depth is found to be significant on the second axis (regression coefficent = 0.80, degrees of freedom = 22.8 and $P = 0.014$).

Table 2.28. Multiple comparison tests for the differences between the random site variables in the GLMM using BLUPs (Best Linear Unbiased Predictors). t values greater than 2.0 are in **bold type**.

Comparison	Axis 1		Axis 2		Axis 3		Axis 4	
	DF	T	DF	t	DF	t	DF	t
GP - LMW	5.2	**-2.35**	4.2	**-5.77** **	3.7	**-8.91** ***	8.1	**4.93** ***
GP - NW	0.2	-1.35	0.7	-1.05	< 0.1	**-3.92**	0.7	0.12
GP - OSW	0.1	-0.55	0.2	-1.19	< 0.1	-1.35	0.2	0.92
GP - SW	0.2	**-2.66**	0.2	-1.71	< 0.1	**-2.41**	0.3	1.25
LMW - NW	0.9	1.03	0.2	**2.80**	< 0.1	1.24	0.2	**-3.65**
LMW - OSW	0.1	0.90	0.1	1.18	< 0.1	1.50	0.1	-1.22
LMW - SW	0.9	-0.48	0.1	1.88	< 0.1	**2.38**	0.1	**-2.09**
NW - OSW	0.1	0.38	0.5	-0.74	< 0.1	1.26	0.3	1.28
NW – SW	22.2	-1.93	7.0	-0.87	0.7	**2.08**	3.2	1.74
OSW - SW	0.2	-1.92	1.2	0.16	< 0.1	-0.11	0.6	-0.23

The differences between the random site effects for each wood were investigated further in order to get some idea as to what site effects may be important. Comparison tests using BLUPs (Best Linear Unbiased Predictors) for each wood were used to test for differences and are shown in Table 2.18. The main difference between the sites are for Graves Park (GP) and Little Matlock Wood (LMW). Both these woods contained very thin soil. However, Graves Park was thought to contain soil of a relatively high nutrient status. *Urtica dioica* and *Geum urbanum* were very common in gaps in the canopy in the site at Graves Park whereas *Holcus lanatus* and *Pteridium aquilinum* were the main species to occur in gaps at Little Matlock Wood. Although other large *t*-values are shown in Table 2.28, no other comparaisons were found to be significantly different.

Figure 2.3 shows the species scores across the first two axes produced by canonical correspondence analysis. The species normally considered as ancient woodland indicators, such as *Allium ursinum* and *Lamiastrum galeobdolon*. *Melica uniflora*, *Milium effusum* and *Mercurialis perennis* can be seen to be placed on the left hand edge of the ordination in the direction of increasing canopy cover and close to the ancient woodland boundary. Species that appear to be associated with the richer deeper soils include *Hyacinthoides non-scripta*, *Allium ursinum*, *Scrophularia nodosa* and *Ribes uva-crispa*. The species in the bottom right corner of the ordination are those species mainly associated with the thin soil consisting of a high proportion of organic matter and a sandyloam texture. These species include *Holcus lanatus*, *Luzula campestre*, *Pteridium aquilinum*, *Deschampsia flexuosa* and *Quercus* seedlings. The plants species associated with the thin loamy soil occur in the bottom left corner of the ordination. These species include *Geum urbanum*, *Bromus ramosus Rumex obtusifolius*, *Rumex sanguineus* and *Anthriscus sylvestris*.

Figure 2.4 shows the species scores across the second and third axes. The third axis appears to represent a gradient from the clay rich soils to the sand rich soils. Species in the direction of the clay rich soils include *Ribes uva-crispa* and *Allium ursinum.* Species in the direction of the sandy soils include *Milium effusum*, *Stellaria holostea*, *Galium saxatile* and *Luzula sylvatica*. The quadrats on the pooly drained silty soils are clearly separated on the fourth axis. The main species associated with this soil type are *Cardamine flexuosa, Angelica sylvestris*, *Epilobium hirsutum*, *Heracleum sphondylium* and *Urtica dioica*.

The correlation matrix for the environmental variables is given in Table 2.29. Many of the correlations can be seen to be significant. In particular, distance and canopy cover, soil moisture and soil organic content, and soil depth with moisture and organic content are all significant. Since, these are some of the variables that were found to strongly associated with the CCA axes, caution should be taken in the interpretation of this part of the analysis.

Table 2.29. Weighted correlation matrix of the independent variables used in the CCA analysis. Significant correlation are in **bold type** (P < 0.01).

	Distance	Canopy cover	log slope	East	North	pH	Soil Depth	Water	Organic matter	Sand	Silt	Clay	Loam
Distance	1.00												
Canopy cover	**-0.41**	1.00											
log slope	-0.13	**0.34**	1.00										
Easterliness	-0.02	-0.04	**-0.56**	1.00									
Northerliness	**0.25**	**-0.47**	**-0.28**	0.16	1.00								
PH	-0.08	0.32	-0.01	0.06	**-0.31**	1.00							
Soil Depth	-0.01	0.04	-0.13	**-0.21**	0.07	**-0.24**	1.00						
Water	-0.20	-0.04	0.04	-0.05	-0.01	0.10	**-0.53**	1.00					
Organic matter	-0.06	**-0.30**	-0.12	0.11	0.14	-0.14	**-0.39**	**0.59**	1.00				
Sand	0.08	**-0.24**	**0.37**	**-0.35**	**0.21**	**-0.55**	0.15	**-0.22**	0.13	1.00			
Silt	-0.08	0.11	0.06	-0.14	0.01	**0.33**	**0.33**	**-0.20**	**-0.27**	**-0.21**	1.00		
Clay	0.12	0.02	0.06	**-0.27**	**-0.26**	**0.25**	0.04	**0.18**	0.03	**-0.23**	-0.03	1.00	
Loam	-0.10	**0.20**	**-0.43**	**0.51**	-0.12	**0.36**	**-0.29**	**0.23**	-0.05	**-0.88**	-0.12	-0.12	1.00

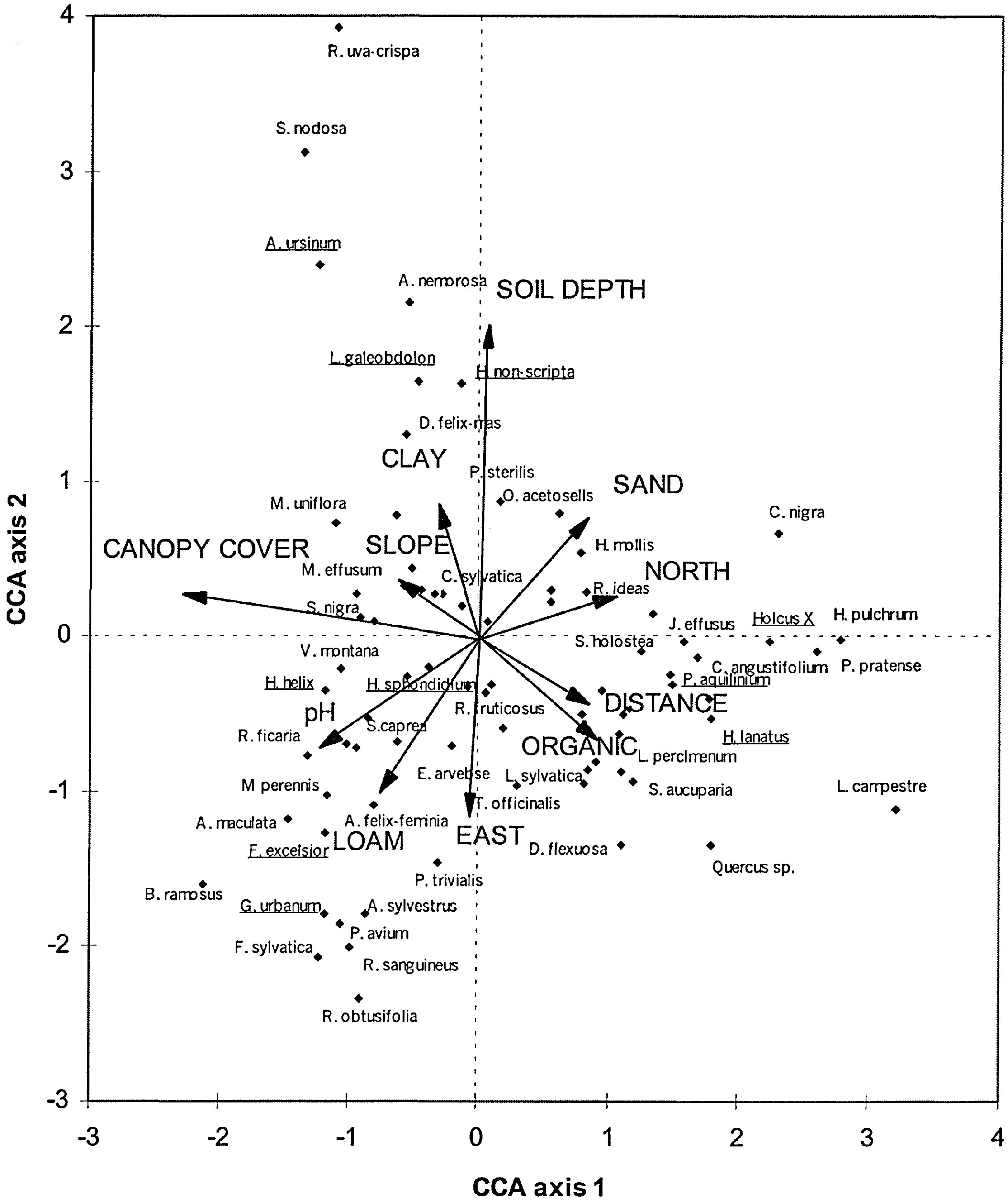


Figure 2.3. Ordination of species scores on the first two axes produced by canonical correspondence analysis. Species with greater than 30% of their variance explained by the first 4 axes are underlined. *Holcus* x, refers to the hybrid between *Holcus lanatus* and *Holcus mollis*. Species names are given in full in Table 2.35.

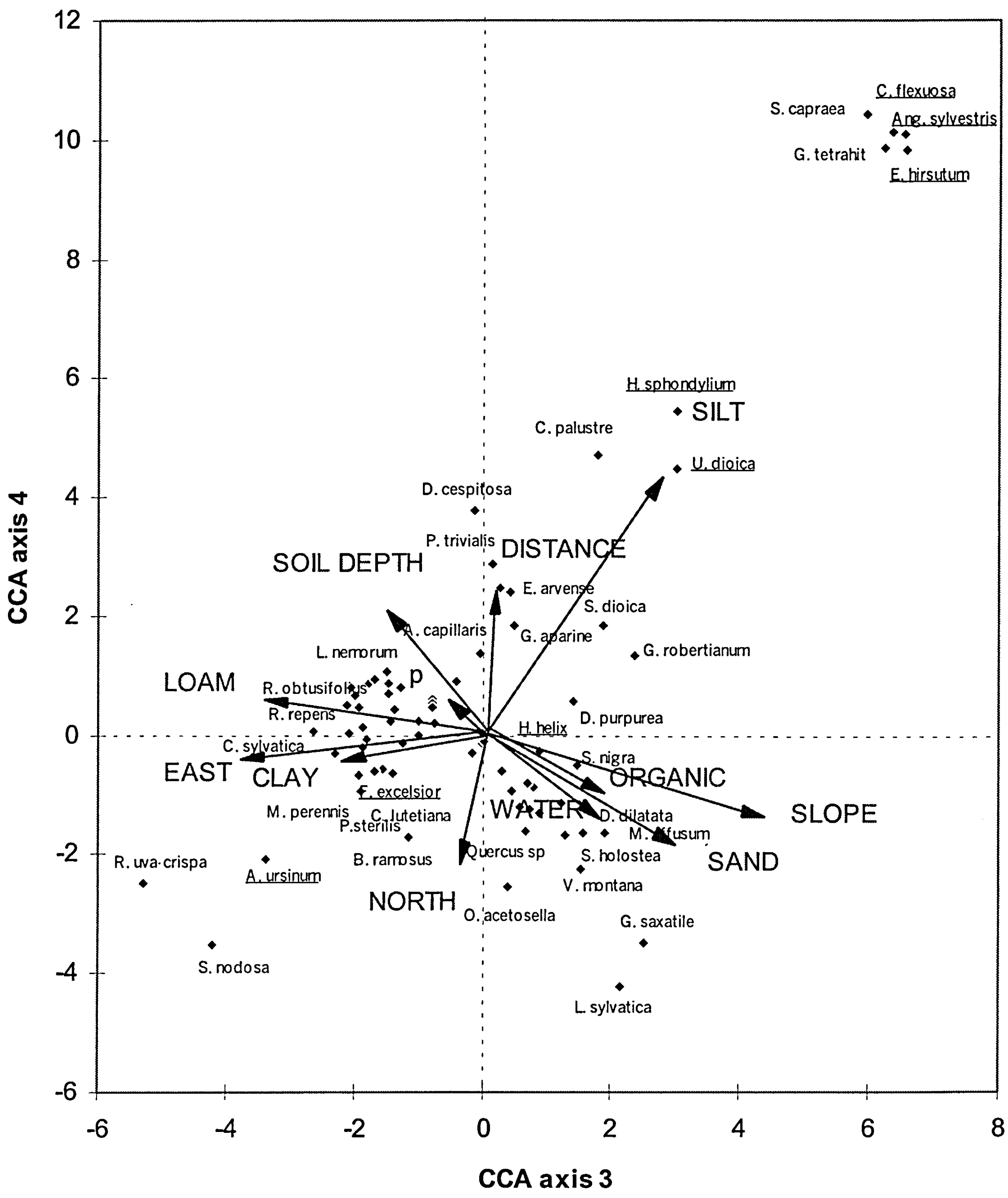


Figure 2.4. Ordination of species scores on the third and forth axes produced by canonical correspondence analysis. Species with greater than 30% of their variance explained by the first 4 axes are underlined. Species names are given in full in Table 2.35.

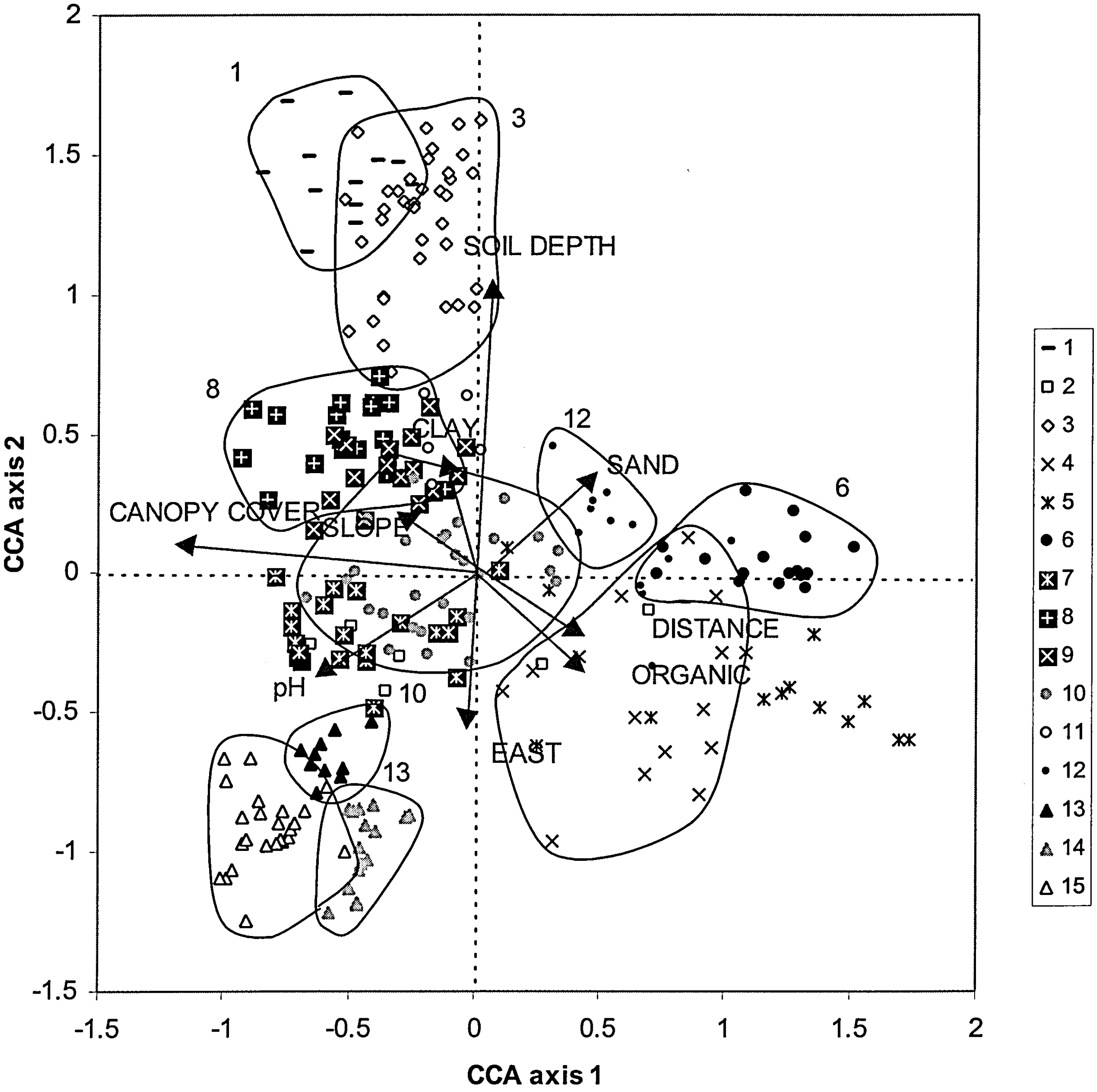


Figure 2.5. Ordination of sample scores on the first two axes produced by canonical correspondence analysis. The numbers refer to plant communities shown by the dendogram in Figure 2.10.

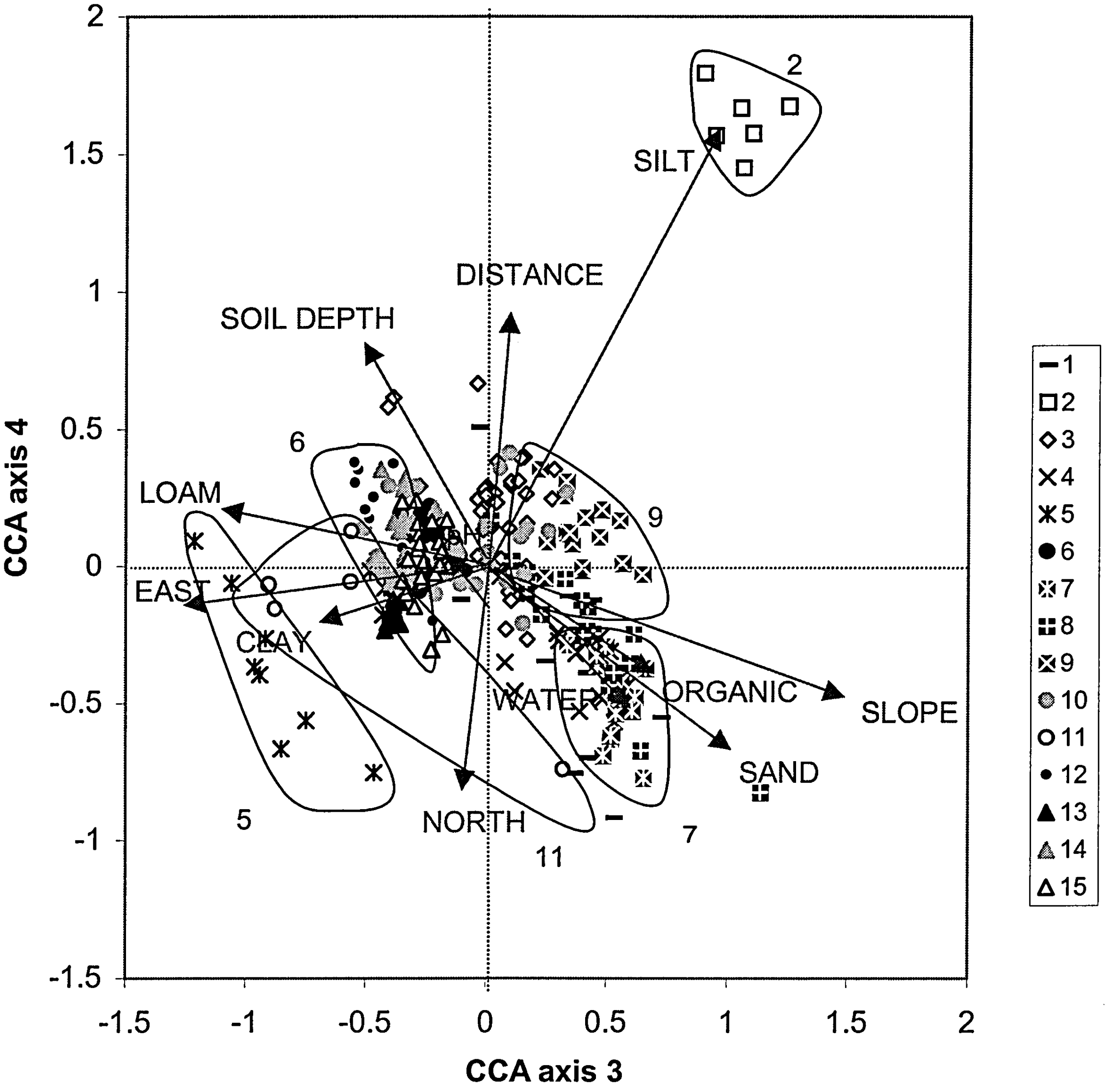


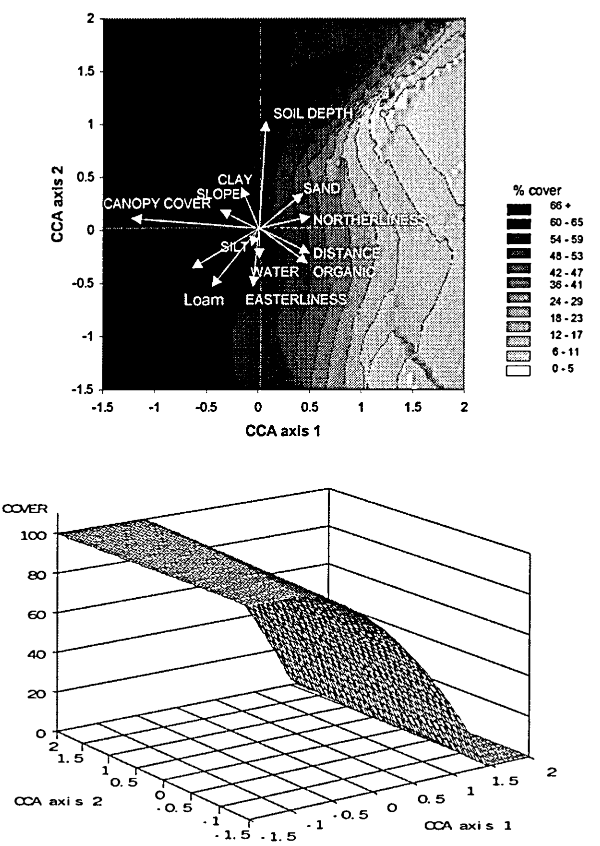
Figure 2.6. Ordination of sample scores on the third and fourth axes produced by canonical correspondence analysis. The numbers refer to different plant communities shown by the dendrogram in Figure 2.9.

2.4.6 Spatial patterns across the CCA ordination

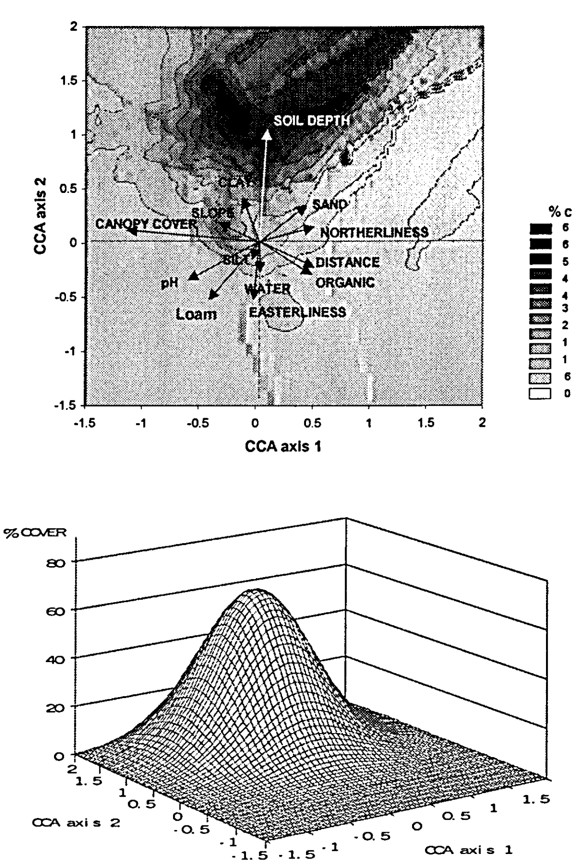
Semi-variograms were plotted for the more abundant species. If the semi-variance showed a clear increase from lag zero, then kriging was applied to the data using the semi-variogram. The exception to this rule was *Lamiastrum galeobdolon,* which showed a high degree of variability along the semi-variogram. An anisotropic model was used for this species, which greatly improved the fit. Some of the distribution maps produced, showed only a narrow single peak, such as *Circea lutetiana* and *Mercurialis perennis*. This reflected only a small number of quadrats with these species present. Ten species that showed clear patterns across the ordination are presented in Figure 2.8. The quadratic surface generalised linear models are also shown. Two of the environmental variables, canopy cover and soil depth are also shown.

Canopy cover can be seen to decrease from left to right across the ordination. The top left corner of the ordination can be seen to contain a high percentage cover of *Allium ursinum;* generally occurring in quadrats on clay-rich soil. *Lamiastrum galeobdolon* also appears to have an affinity for the top left corner of the ordination, but is absent where *Allium ursinum* reaches its optimum. *Hyacinthoides non-scripta* can be seen to have an affinity for the deeper soils. *Pteridium aquilinum* can be seen to occur away from the ancient woodland boundary, and may also be associated with soil with a high organic matter content. *Rubus fruticosus* can be seen to have its optimum in the centre of the ordination. *Crataegus monogyna* and *Sambucus nigra* can be seen to occur on the left hand-side of the ordination. The distributions of the *Holcus* family are also shown in Figure 2.7. *Holcus lanatus* appears to have a similar distribution to *Pteridium aquilinum,* being restricted to the bottom right of the ordination in the direction of distance from the ancient woodland boundary and soil with high organic content. *Holcus mollis,* appears to occur much more centrally in the ordination diagram, probably preferring a higher proportion of canopy cover. The distribution of the hybrid between these two grass species appears to have a distribution located above *Holcus lanatus* and to the right of *Holcus mollis*. The hybrid would appear to prefer an open canopy like *Holcus lanatus* but appears to show a preference for deeper and possibly more nutrient rich soil types than *Holcus lanatus*.

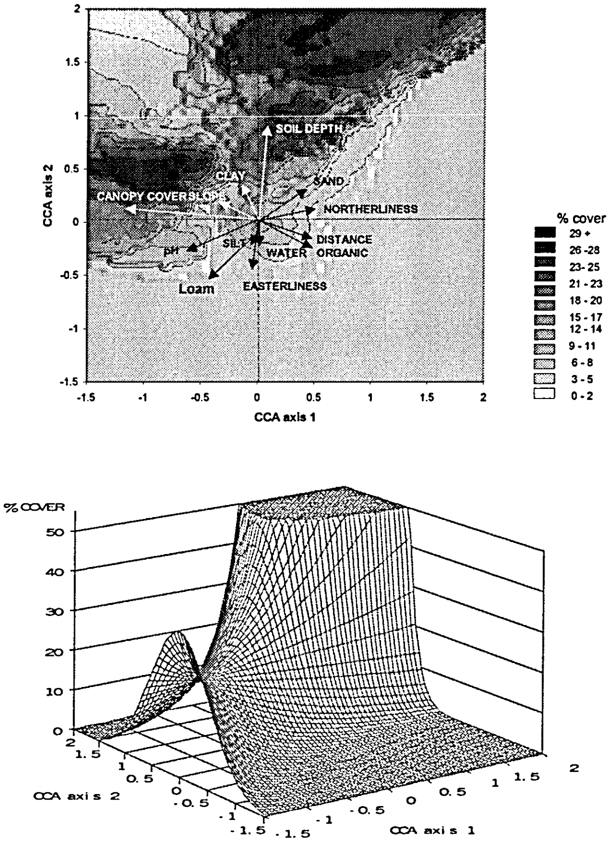
Canopy cover



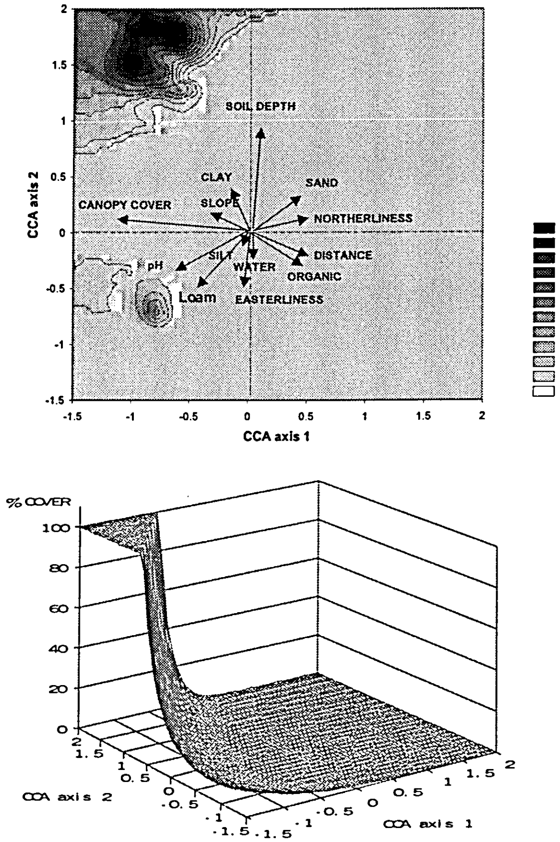
Hyacinthoides non-scripta

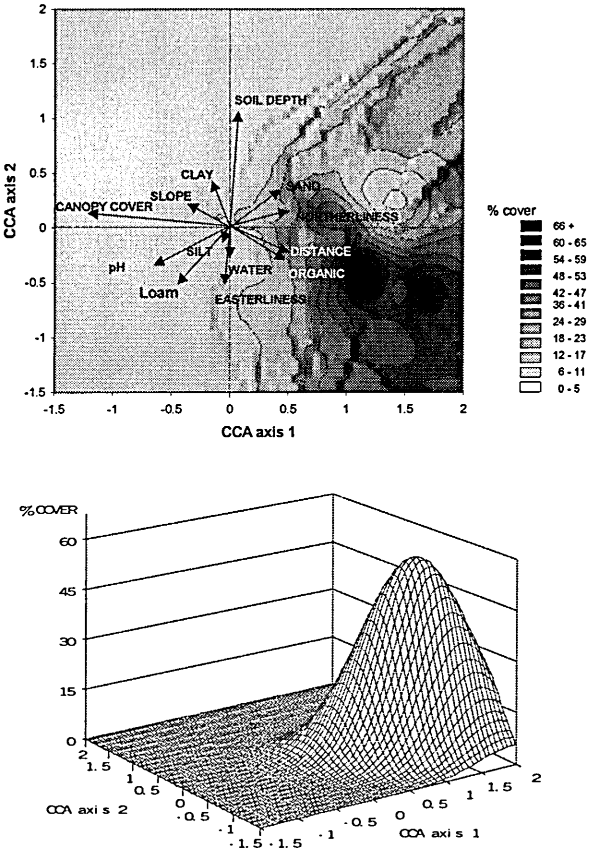


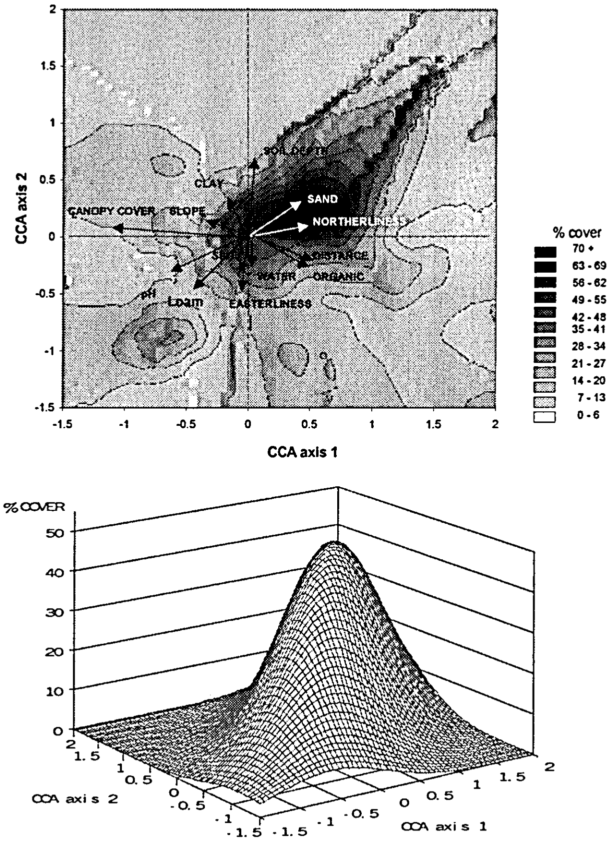
Lamiastrum galeobdolon

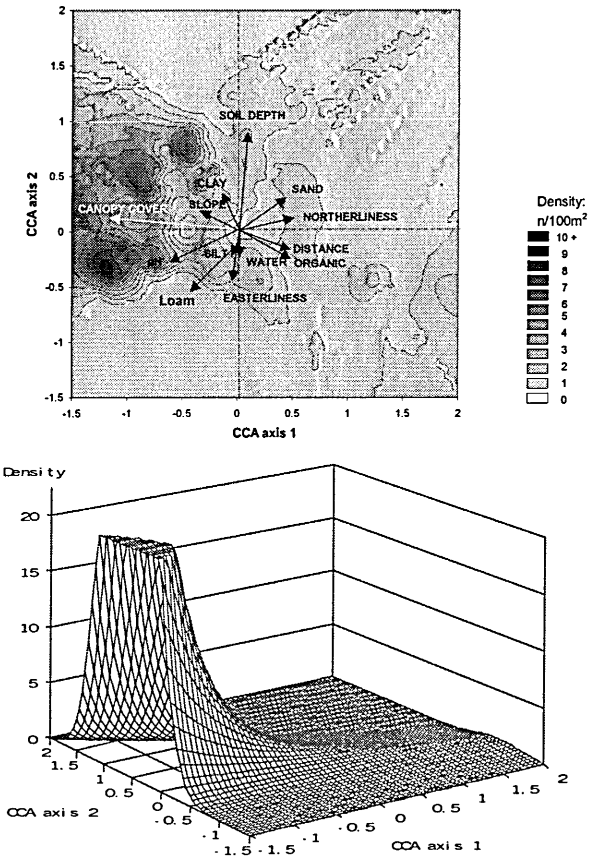


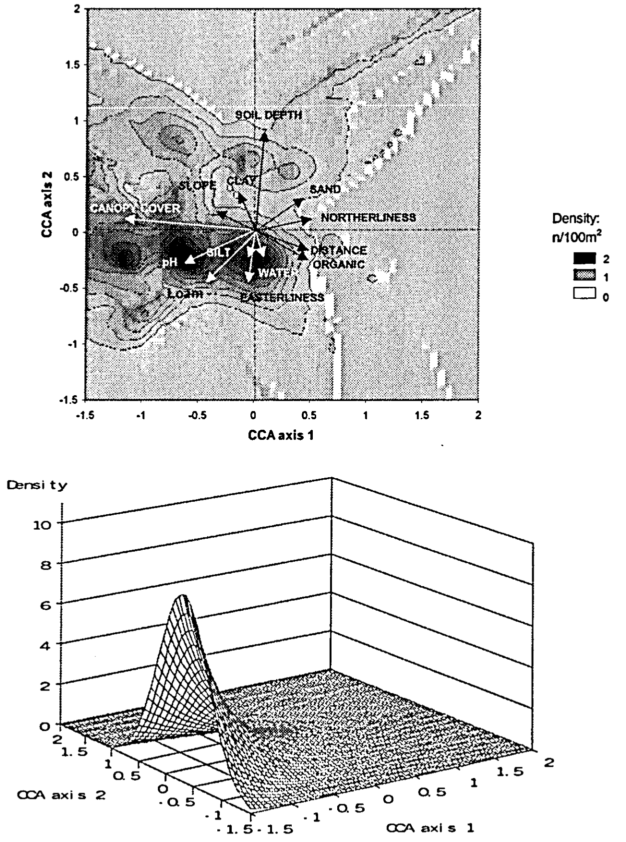
Allium ursinum




Pteridium aquilinum


Rubus fruticosus


Crataegus monogyna


Sambucus nigra

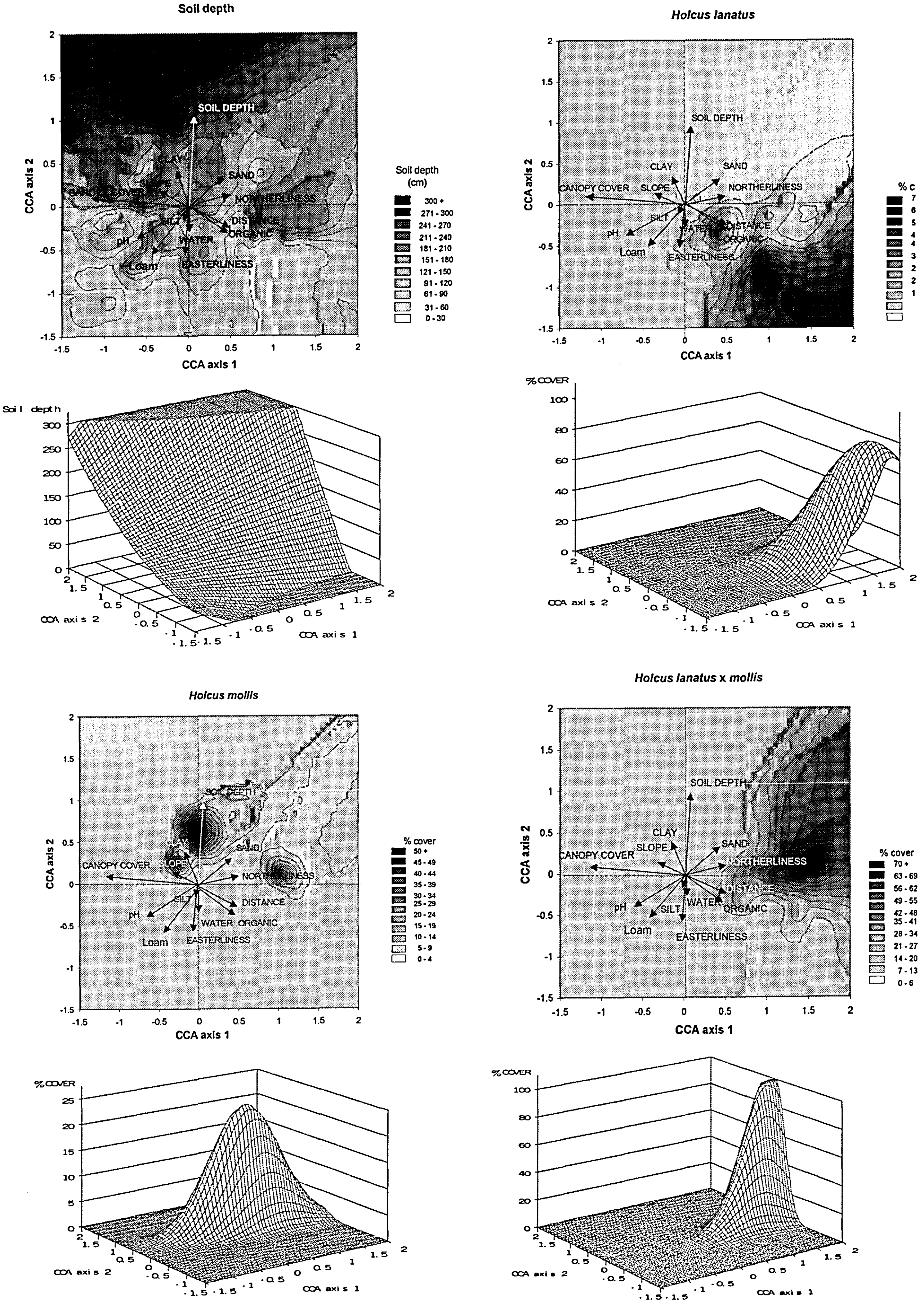


Figure 2.8. Spatial patterns of selected species across the CCA ordination presented by the use of kriging.

Table 2.30. Parameter estimates for semi-variograms used for kriging in order to produce the 3D terrain maps for the abundance of each species across the first two axes produced by canonical correspondence analysis.

Variable	Model		Nugget	Sill	Range	Minor range	R^2
Canopy cover	Isotropic	Gaussian	64.0	2238.0	3.64		0.98
Soil depth	Isotropic	Exponential	1040.0	5230.0	0.93		0.93
Allium ursinum	Isotropic	Spherical	19.5	157.9	0.33		0.67
Holcus hybrid	Isotropic	Exponential	53.8	278.7	3.57		0.95
Holcus lanatus	Isotropic	Spherical	27.0	364.9	3.86		0.96
Holcus mollis	Isotropic	Spherical	49.0	161.7	0.49		0.92
Hyacinthoides non-scripta	Isotropic	Spherical	306.0	906.0	2.97		0.83
Lamiastrum galeobdolon	Anisotropic	Spherical	173.3	557.6	7.76	28.8	0.30
Pteridium aquilinum	Isotropic	Spherical	60.0	137.9	0.74		0.71
Rubus fruticosus	Isotropic	Spherical	263.0	805.7	1.28		0.84
Crataegus monogyna	Isotropic	Spherical	0.68	5.5	0.71		0.97
Sambucus nigra	Isotropic	Spherical	0.65	1.4	0.92		0.94

Table 2.31. Quadratic surface regression models for the response of each species with the first two axes produced from canonical correspondence analysis. The equations for canopy cover and soil depth were estimated using linear regression, all other equations were estimated using generalised linear models assuming a poisson distribution and with a log-link function and over-dispersion parameter.

Dependent variable		Intercept	Axis 1	Axis 2	Axis 1 × 2	Axis 1^2	Axis 2^2	R^2/D^2
Canopy cover	Full	84.24 ***	-34.54 ***	1.91 *	-0.52	-14.80 ***	-0.16	0.95
	Stepwise	84.17 ***	-34.37 ***	2.02 ***		-14.79 ***		0.95
Soil depth	Full	122.42 ***	16.47 **	128.10 ***	45.18 ***	0.41	13.23 **	0.83
	Stepwise	122.59 ***	16.66 ***	127.96 ***	45.09 ***		13.29 **	0.83
Allium ursinum	Full	-2.27 *	-3.44 *	0.07	-1.20	0.25	1.14 **	
	Stepwise	-2.22 **	-3.35 ***		-1.28 *		1.15 ***	0.63
Holcus lanatus	Full	0.99 ***	2.89 ***	-1.71 *	-2.00 **	-1.61 ***	-2.06 *	0.59
Holcus mollis	Full	0.32 ***	2.77 **	10.42 ***	-2.74 **	-1.43 **	-11.22 ***	0.45
Holcus hybrid	Full	-0.50 *	5.98 ***	-4.20 ***	3.84 ***	-1.79 ***	-9.45 ***	0.52
Hyacinthoides	Full	2.36 ***	1.04 *	3.02 ***	-0.79	-1.86 ***	-1.38 ***	0.53
non-scripta	Stepwise	2.29 ***	0.69 *	3.08 ***		-1.4 ***	-1.313 ***	0.53
Lamiastrum	Full	1.05 **	-1.71 **	3.92 ***	2.58 **	-0.13	-1.77 ***	0.34
galeobdolon	Stepwise	1.02 ***	-1.70 **	3.98 ***	2.64 **		-1.79 ***	0.34
Pteridium	Full	1.24 ***	4.59 ***	0.90	-3.60 ***	-2.47 ***	-3.34 ***	0.71
aquilinum	Stepwise	1.34 ***	4.15 ***		-2.79 ***	-2.15 ***	-3.41 ***	0.71
Rubu fruticosus	Full	3.62 ***	0.99 ***	-0.44 *	1.53 ***	-1.06 ***	-1.20 ***	0.38
Crataegus monogyna	Full	0.62 ***	-1.54 ***	-0.06	-2.92 ***	-0.12	-3.60 ***	0.40
density	Stepwise	0.59 ***	-1.55 ***		-2.76 ***		-3.60 ***	0.37
Sambucus nigra	Full	-0.40 *	-1.66 ***	0.66	2.08 *	-0.30	-2.63 ***	0.36
density	Stepwise	-0.46 **	-1.60 ***		1.02 *		-2.52 ***	0.33

2.4.7 Further analysis of *Rubus fruticosus*

The response of *Rubus fruticosus* to the environmental variables recorded was investigated further in order to determine whether canopy cover on its own was sufficient to limit the abundance of this species. A mixed model was used, as in the previous analysis, with the same time series error structures and random effects. The AR(1) error structure was again found to give the best fit in terms of the AIC and SBC values (Table 2.32). The estimates for the covariance parameters are given in Table 2.33 and the solution for the fixed effects is given in Table 2.34. Table 2.34 also includes the results of removing the least significant effect and repeating the analysis until all effects were found to be significant at the 5% level.

Table 2.32. Comparison of three different error structures - First order autoregressive error structure - AR(1), First order moving average structure - ARMA(1,1) and Compound symmetry -CS on the goodness of fit statistics produced by SAS (Akaikes Information Criterion - AIC and Schwarz's Bayesian Criterion – SBC).

	AR(1)		ARMA(1,1)		CS	
	AIC	SBC	AIC	SBC	AIC	SBC
Rubus fruticosus	-384.7	**-405.4**	**-384.4**	-413.7	-400.6	-421.3

Table 2.33. Covariance Parameter Estimates (REML) for model with all variables.

Covariance Parameter	Wood	Estimate
Wood		0.00
Transect (nested in wood)		0.00
Variance	GP	2.35 ***
AR(1)	GP	0.53 ***
Variance	LMW	1.76 **
AR(1)	LMW	0.53 ***
Variance	NW	1.70 ***
AR(1)	NW	0.49 ***
Variance	OSW	0.39 **
AR(1)	OSW	0.48 **
Variance	SW	1.21 **
AR(1)	SW	0.51 **

Table 2.34 shows that canopy cover, distance from the ancient woodland, slope and the proportion of *Crataegus monogyna* all appear to have significant impacts on *Rubus fruticosus*. Quadratic relationships can be seen for canopy cover, soil depth and distance from the ancient woodland boundary. The predicted response of *Rubus fruticosus,* according to the backward stepwise model, with canopy cover and the proportion of

Crataegus monogyna in the canopy model is shown in Figure 2.8. The presence of *Crataegus monogyna* appears to greatly reduce the cover of *Rubus fruticosus*.

Table 2.34. Solution for fixed effects with the log transformed % cover values of *Rubus fruticosus* as the dependent variable.

		Full model			Backward stepwise model	
Variable	DF	Estimate	Type I F	DF	Estimate	Type I F
Intercept	124	-2.22		41.2	0.98	
Canopy cover	75.7	4.96 **	11.5 **	56.3	4.40 *	9.6 **
Canopy cover^2	105	-2.88 **	14.2 ***	73.3	-2.69 **	11.6 **
Distance	71.3	0.11 **	0.0	33.4	0.077 *	0.7
Distance^2	78.9	-0.0036 **	4.5 *	38.6	-0.0026 *	1.7
Soil depth	45.4	-0.01	29.4 ***	19.6	0.012 *	30.1 ***
Soil depth^2	51.4	0.00	24.9 ***	23.0	-5.8×10^{-4} ***	20.8 ***
pH	57.3	0.62	1.7			
pH^2	50.2	-0.05	0.0			
Water	80.9	-0.13	0.1			
Organic matter	121	1.06	0.4			
Easterliness	76.9	-0.20	1.0			
Northerliness	64.8	0.23	0.0			
log slope	110	0.29 **	14.3 ***	108	0.30 **	13.2 ***
Proportion hawthorn	62.6	-1.08 *	5.3 *	41.4	-1.01 *	5.6 *
Proportion elder	75.4	0.79	2.8			
Sand	54.6	-5.16				
Silt	22.7	-0.49				
Clay	55.9	-4.73				
Sandy loam	55.2	4.97				
loam		0.00				

	NDF	DDF	Type III F	Type I F
Soil Texture	4	43.1	1.1	1.1

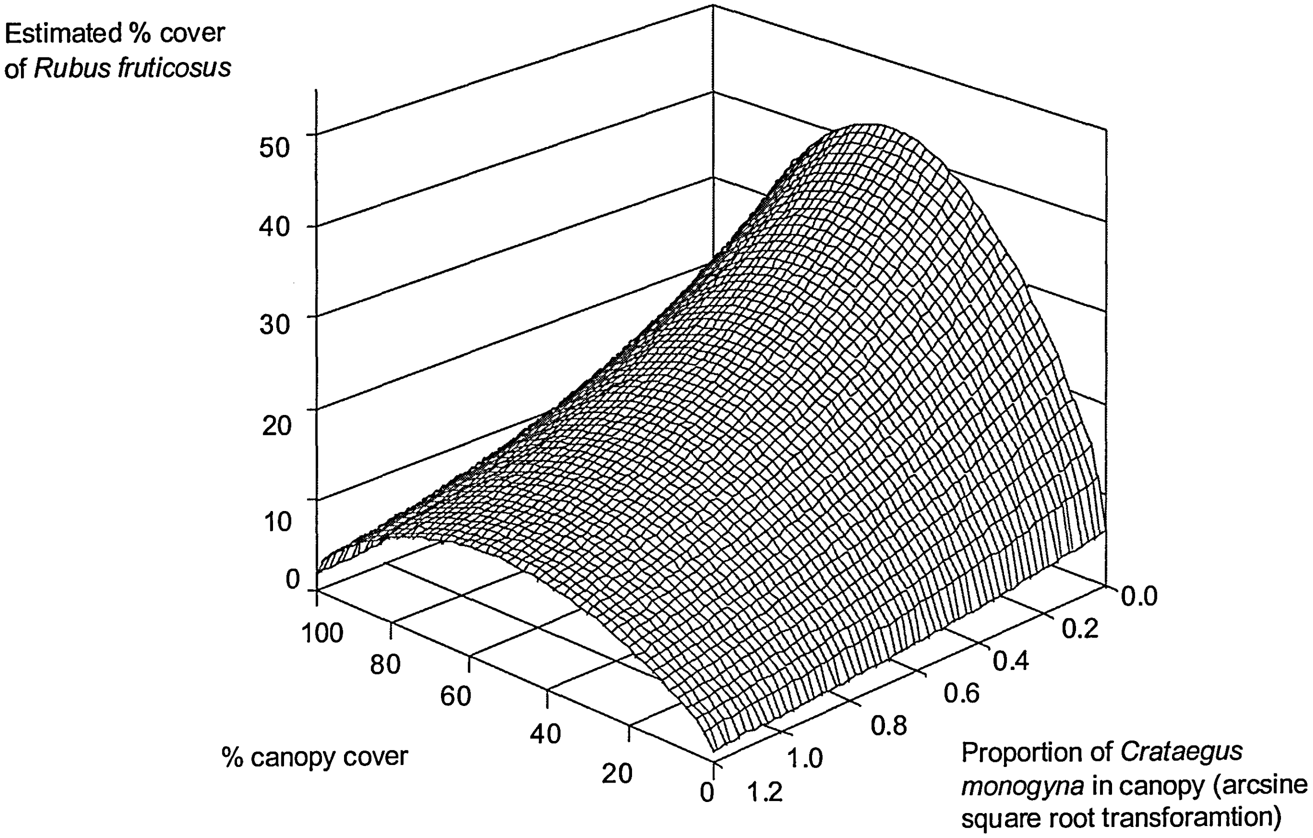


Figure 2.9. Predicted response surface for *Rubus fruticosus* using the regression equation from Table 2.34 after the elimination of non-significant terms.

2.4.8 Further analysis of the number of ancient woodland indicator species

The same analysis as that done on *Rubus fruticosus* was repeated on the log transformed data for the number of woodland indicator species. The AR(1) structure can be seen to give the best fit to the data as shown by Table 2.35. The compound symmetry structure can be seen to give a much poorer fit than the AR(1) and AR(1,1) error structures. The covariance parameter estimates are shown in Table 2.36. Old Spring Wood can be seen to have a much higher AR(1) parameter estimate than the other woods.

The solution for the fixed effects is given in Table 2.37. Distance from the ancient woodland can be seen to have greatest effect on the number of woodland indicator species. Even the type-I F test for canopy cover entered first is not significant. There is evidence for a quadratic relation with distance given by the type-I F tests, but this is for an inverse quadratic curve. This may simply be due to the problem of zeros being present for a large proportion of the transects after 15m. This problem could have been overcome by assuming a Poisson distribution but then the time series error structure could not have been fitted. There is also some evidence for the proportion of *Crataegus monogyna* reducing the number of indicator species as well as extreme soil texture types i.e. high proportions of clay or silt. However, once all non-significant terms have been removed using the backward stepwise method only distance remains significant.

Table 2.35. Comparison of three different error structures for the number of woodland indicator species- First order auto-regressive error structure - AR(1), First order moving average structure - ARMA(1,1) and Compound symmetry -CS on the goodness of fit statistics produced by SAS (Akaikes Information Criterion - AIC and Schwarz's Bayesian Criterion – SBC).

	AR(1)		ARMA(1,1)		CS	
	AIC	SBC	AIC	SBC	AIC	SBC
Woodland species	**-117.0**	**-137.6**	-119.0	-148.3	-400.6	-421.3

Table 2.36. Covariance Parameter Estimates (REML) for model with all variables with the number of woodland indicators as the dependent variable.

Covariance Parameter	Wood	Estimate
Wood		0.06
Transect (nested in wood)		0.00
Variance	GP	0.08 ***
AR(1)	GP	0.42 **
Variance	LMW	0.11 ***
AR(1)	LMW	0.27
Variance	NW	0.07 ***
AR(1)	NW	0.32 *
Variance	OSW	0.19 **
AR(1)	OSW	0.63 ***
Variance	SW	0.16 ***
AR(1)	SW	0.28 *

Table 2.37. Solution for fixed effects with the log transformed number of woodland indicator species per square metre as the dependent variable.

	Full model			Backward stepwise model		
Variable	DF	Estimate	Type I F	DF	Estimate	F
Intercept	36.2	-0.48		6.4	0.78 ***	
Canopy cover	55.6	-0.12	4.0			
Canopy cover^2	62.5	0.05	7.3			
Distance	105	-0.042 ***	50.5 ***	81.1	-0.030 ***	90.4 ***
Distance^2	114	6×10^{-4}	6.9 *			
Soil depth	18.6	-0.003	0.6			
Soil depth^2	3.4	3×10^{-5}	0.0			
PH	122	-0.07	1.1			
pH^2	120	-0.01	0.3			
Water	80.9	0.91	3.5			
Organic matter	172	-0.49	0.2			
Easterliness	15.7	-0.02	2.5			
Northerliness	49.4	0.03	1.5			
log slope	102	0.01	0.2			
Proportion hawthorn	42.5	-0.15	4.4 *			
Proportion elder	31.3	-0.17	1.1			
Sand	1.1	-1.35	2.4			
Silt	22.8	-0.43 *	3.9			
Clay	1.1	-1.61	5.4 *			
Sandy loam	1.0	1.12	0.7			
Loam		0.00				

	NDF	DDF	Type III F	Type I F
Soil Texture	4	0	2.5	2.5

2.4.9 Cluster analysis

The dendrograms produced by using Ward's method and Flexible clustering gave very similar and easily interpretable results. The Flexible method with $\beta = -0.5$ appeared able to split the end clusters into near and far from the ancient woodland boundary, and so it was the dendogram produced from this method that was used for further analysis (Figure 2.10). Prior to these partitions the Ward's clustering algorithm had formed virtually identical clusters to the Flexible method. The pattern of chaining described in the data analysis section was found to occur with single linkage and average linkage methods. The application of *k*-means clustering to the data gave different results according to the starting values used, and no clear structure could be obtained by specifying different numbers of clusters to be produced.

The primary divisions can be seen to be due to the soil texture. The quadrats on the thin loam, deep clay and sands and the poorly drained soils have all formed separate groups early on in the dendogram. The remaining data comprises the quadrats on the sandy loam which can then be further separated according to canopy cover, soil depth and whether they are from the start or end of a transect. This clustering method has therefore appeared to be able to recreate the experimental design i.e. different woods on different soil types with transects nested in each site. The NVC types of each cluster are also shown in Figure 2.10. Since, the NVC methodology mainly relies on species abundances, climate and the species of trees forming the canopy it may not be that surprising to find that quadrats of the same NVC type are not necessarily closes together on the dendogram. However, another interpretation could be that the dendogram canot be correctly displayed in 2 dimensions. In order to test this the dendogram has been superimposed onto the first 2 CCA axes. Figure 2.10 shows that it is possible to draw polygons onto the ordination, which link up the clusters according to the broad NVC types.

2.4.10 Indicator species analysis

The main difference between the two methods used is the method of clustering, making comparisons between the power of the methods used harder to compare. A large number of 1s appear in the TWINSPAN diagram (Figure 2.11) indicating that a species has a frequency >0%. This is not very useful for prediction; it simply indicates that a species is present in all sites of a group.

The first division produced by the TWINSPAN dendogram separates the loam, silty loam and some sandy loam sites from the rest. A similar separation can be seen for the primary division produced from the Flexible clustering method of the first eight CCA axes. However, only the sites on the thin loam have been separated by the Flexible clustering method. As stated in the introduction, TWINSPAN does not look for groups in the data, which can lead to site groups that are more heterogeneous than others. The second division produced by TWINSPAN appears to separate the sites on the deep clay and sand from the sandy loam and to a lesser extent the sites on thin loam with a relatively open canopy from the sites with a higher proportion of canopy cover. The third division appears to mainly separate the deep sand from the deep clay. Both methods give *Hyacinthoides non-scripta* as an indicator of deep sand and *Allium ursinum* as an indicator of the quadrats on the deep clay soil. The fourth division separates the quadrats on the poorly drained silty soil from the quadrats on the sandy loam.

The dendogram produced from the flexible clustering method is much more easily interpretable than that produced by TWINSPAN. A large number of species are often significant for the divisions in the hierarchical typology. An arbitrary value of 15% for the INDVAL index was chosen. This supposes that a characteristic species is present in at least 30% of one site group and that its relative abundance in that group reaches at least 30%. Figure 2.13 presents on the dendogram of sites, all the species that have an index value greater than 15%. Many species are typical of both high and low hierarchy levels. For example, *Rubus fruticosus* and *Geum urbanum* have higher index values at the start of the dendogram, indicating that they occur in a broad range of habitat-types. Other species only occur at the ends of the dendogram, such as *Mercurialis perennis* and *Anemone nemorosa,* indicating a very narrow range of habitat types. Most of the species described as being associated with or having a mild or strong affinity for ancient woodland by Peterken (1993) can be found towards the ends of the dendogram. Of these, *Anemone nemorosa* and *Mercurialis perennis* show a very restricted distribution across the dendogram. *Allium ursinum*, *Hyacinthoides non-scripta*, *Veronica montana* and *Milium effusum* show a more general distribution and *Lamiastrum galeobdolon* appears in several locations on the dendogram.

A two-way indicator table produced from the flexible clustering method and INDVAL is show in Table 2.35. The structure of this table is shown in Figure 2.13. The table contains a list of all the species found in the quadrats in the division of the hierarchical structure where they had their optimum. This table also includes the sum of abundance followed by the sum of occurrence for each species, in each partition of the dendogram. Species that have an indicator value >55% are symmetrical indicators i.e. both their presence and abundance can be used as indicators. The other species must be considered accidental or anecdotal and are asymmetrical indicators (Dufrene & Legendre, 1997).

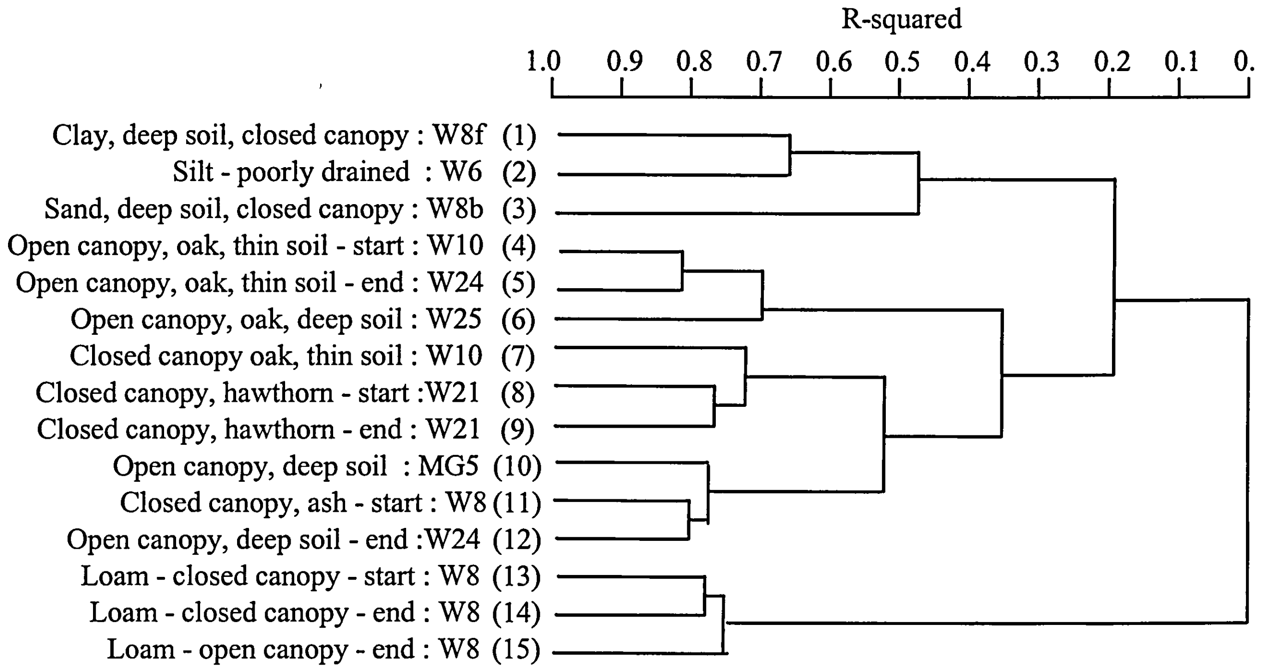


Figure 2.10. Hierarchical dendogram produced from the flexible clustering method. A desciption of each cluste given followed by NVC community.

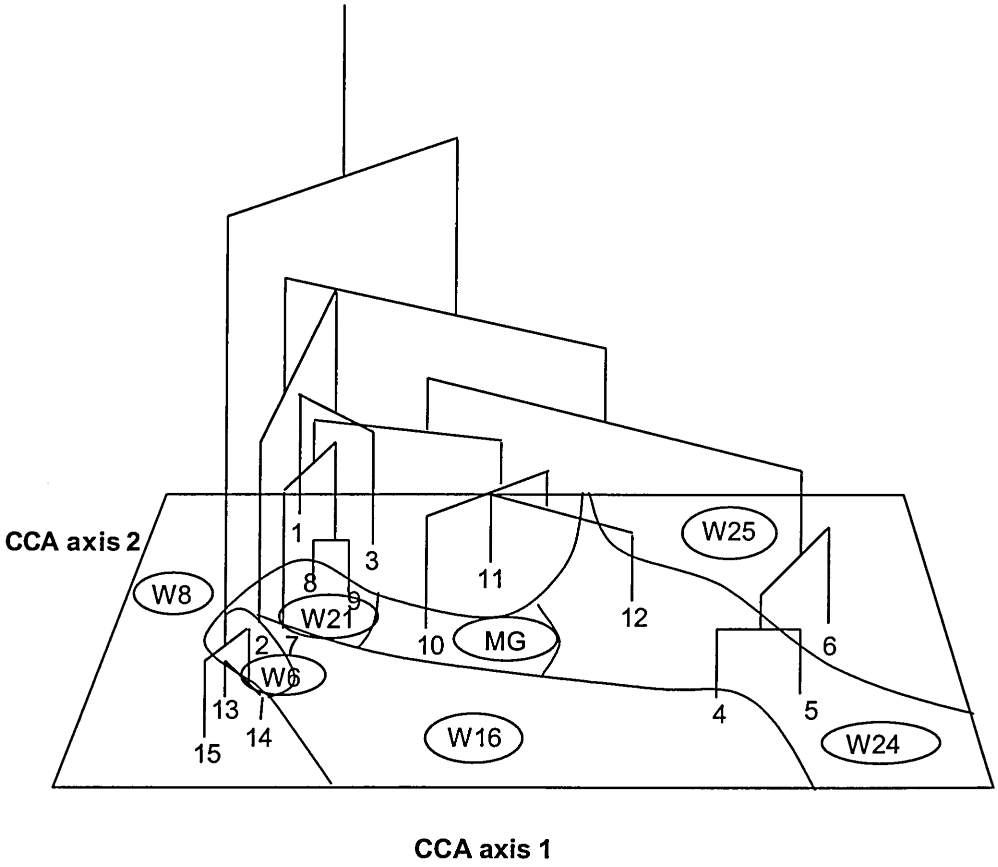


Figure 2.11. The dendogram of Figure 2.5 showing the orientation of the clusters when superimposed onto the CCA ordination of the first two axes. The ends of the dendogram are located in the most central point of the samples for each cluster. The NVC communities are represented on the ordination as polygons.

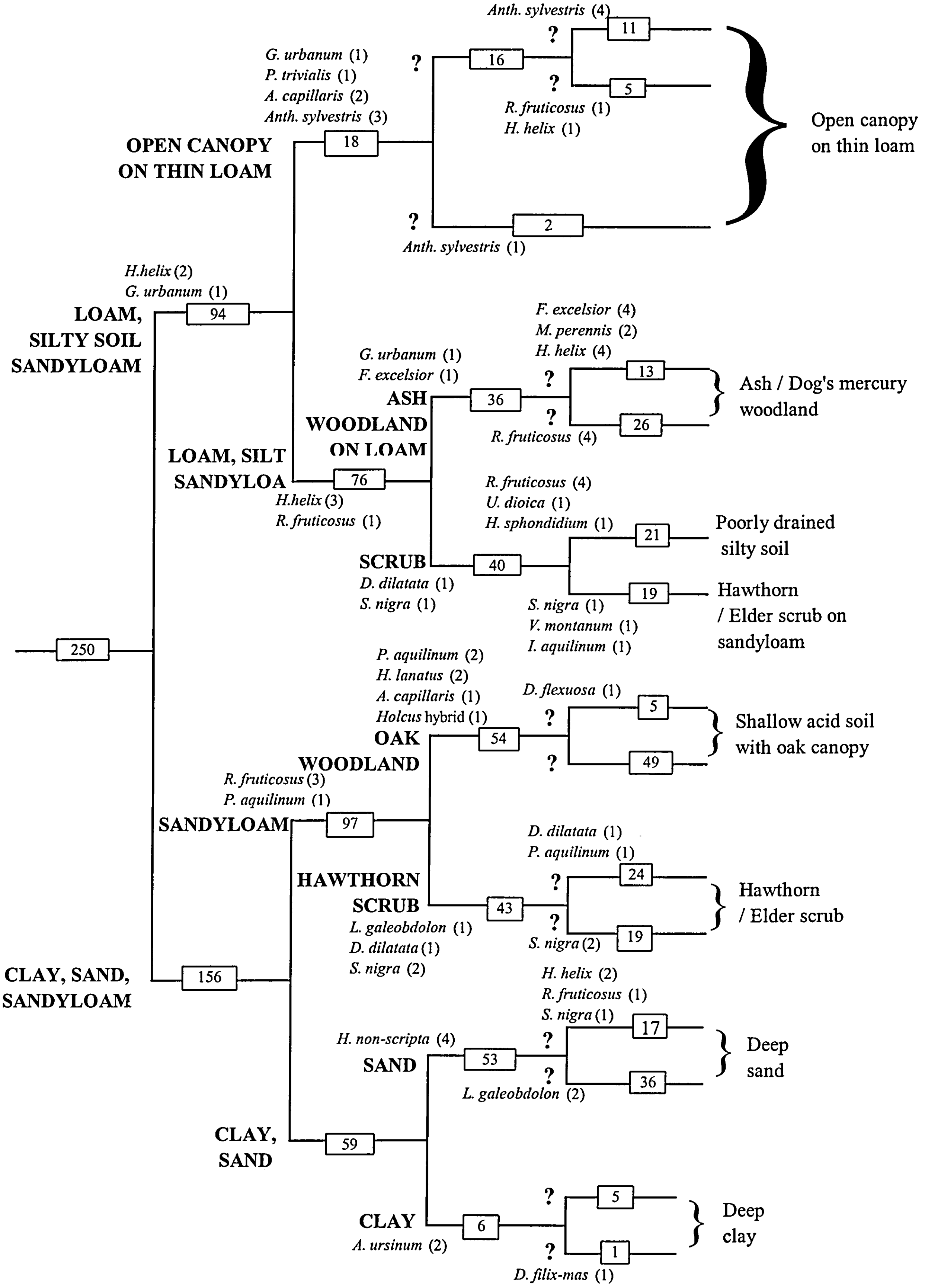


Figure 2.12. Dendogram representing the TWINSPAN classification of the plant % cover data. The indicator species relative abundance levels are expressed on an ordinal scale in parentheses (1, 0-2%; 2, 2-5%; 3, 5-10%; 4, 10-25%; and 5, 20-100%). Numbers in boxes refer to the number of quadrats in each division.

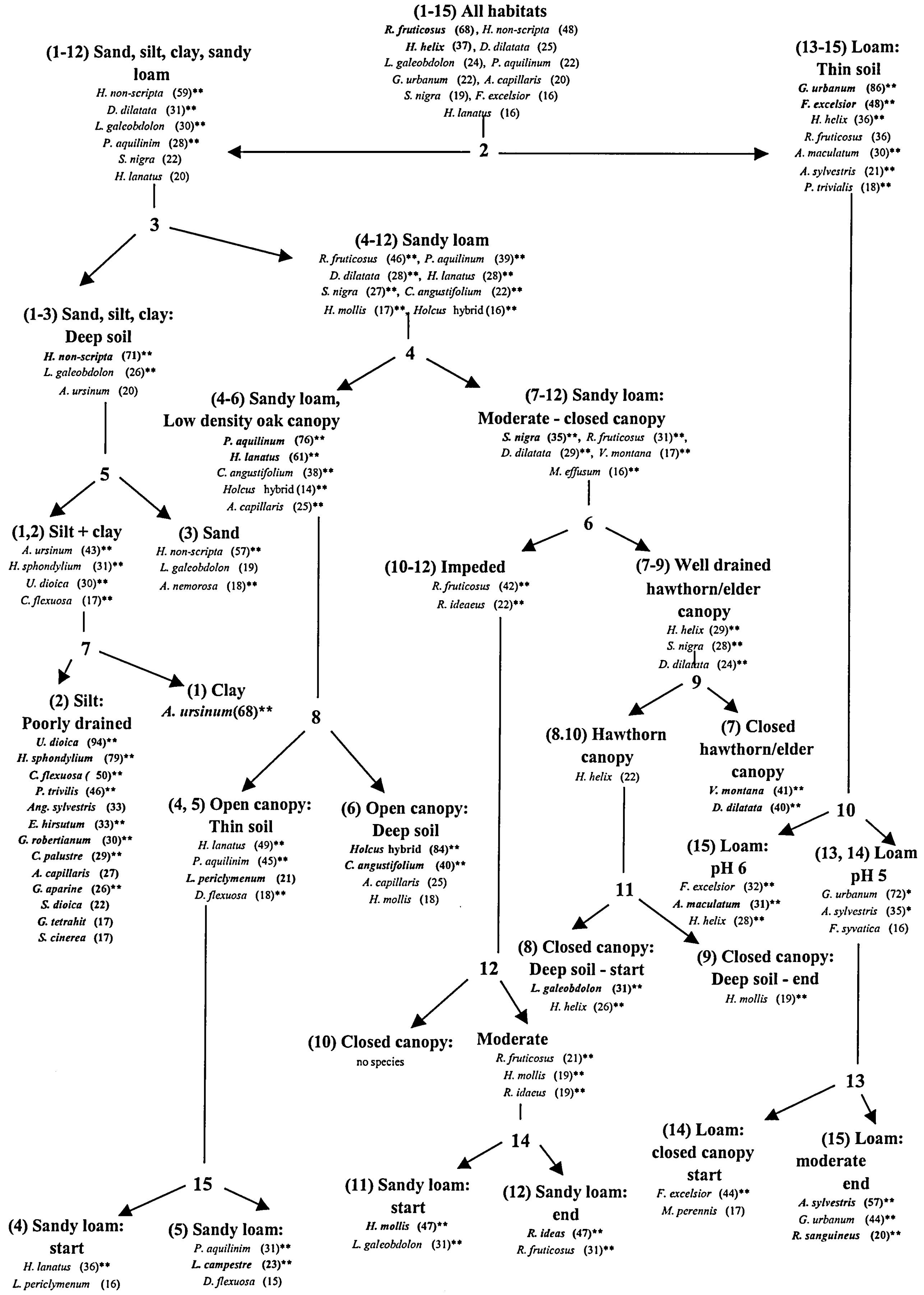


Figure 2.13. Site clusters obtained with the flexible hierarchical method. The associated indicator values are in parentheses. All species with an indicator value >15% are presented for each site cluster where they are found. The maximum is indicated in **bold**.

Table 2.38. Two-way indicator table showing the species indicator power for the site clustering hierarchy. The co headings correspond to those of Figure10. Numbers in bold type represent the main data set structure. ** $P < 0.01$.

Species	IndVal (%)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Clay, deep soil, closed canopy (1)															
Allium ursinum	68 **	**528/9**		105/4										105/2	2/1
Ribes uva-crispa	8	**2/1**													
Scrophularia nodosa	8	**1/1**													
Silt+clay (1,2)															
Dryopteris filix-mas	13 **	**7/2**	**1/1**	1/1							10/4				
Silt, poorly drained (2)															
Urtica dioica	94 **		**290/6**		5/1	2/1	5/1		100/3	6/3	13/3	2/1			25/1
Heracleum sphondylium	79 **	1/1	**80/5**						3/1		19/4		7/1		
Cardamine flexuosa	50 **		**6/3**												
Poa trivalis	46 **		**46/4**				3/3		1/1		25/3		1/1	1/1	141/9
Angelica sylvestris	33 **		**3/2**												
Epilobium hirsutum	33 **		**15/2**												
Geranium robertianum	30 **	2/1	**5/3**					5/2	12/4		4/2				
Cirsium palustre	29 **		**11/2**				5/2				6/2				
Agrostis capillaris	27		**37/5**		2/2	117/7	155/11		5/3	6/4	110/5		74/5	6/2	113/6
Galium aparine	26 **	1/1	**40/2**				38/5		26/3	3/2	9/5		8/5		3/2
Silene dioica	22 **		**10/2**	1/1					22/7		15/3				3/3
Galeopsis tetrahit	17		**1/1**												
Salix cineria	17		**1/1**												
Deschampsia caespitosa	10	1/1	**1/1**								1/1		1/1		
Sand+silt+clay (1,2,3)															1/1
Hyacinthoides non-scripta	71 **	**363/11**		**1688/38**	20/1	52/5	50/11	2/1	90/8	304/15	499/15	120/5	115/9		
Sand, closed canopy (3)															
Anemone nemorosa	18 **	3/1		**168/7**			1/1				4/2	1/1			
Sand+silt+clay+sandy loam (1-12)															
No species indicator of this group															
Moderate canopy cover, thin soil -start (4)															
Epilobium montanum	8				**4/3**	1/1	1/1	2/2			1/1				1/1
Stellaria holostea	8				**45/2**			7/1				3/1	1/1		
Open canopy, thin soil (4,5)															
Lonicera. periclymenum	21			3/2	**166/4**	**85/2**		2/1					5/1		
Deschampsia flexuosa	18 **				**3/3**	**80/2**									
Sorbus aucuparia	11				**2/2**	**2/2**					2/2				
Luzula sylvatica	7				**10/1**	**1/1**									
Quercus spp.	7				**1/1**	**1/1**									
Oxalis acetosella	7	3/2		1/1	**12/2**	**15/1**		1/1	1/1		1/1	1/1	6/2		
Open canopy, thin soil - end (5)															
Luzula campestre	23 **					**3/3**									
Galium saxatile	7					**7/1**		2/1							
Rosa arvensis	6					**20/1**					17/3				
Open canopy (4,5,6)															
Pteridium aquilinum	76 **				**421/12**	**527/11**	**371/13**	15/4	5/1		70/6	27/2	136/7		
Holcus lanatus	61 **				**480/11**	**254/8**	**132/9**	20/1	1/1		37/3		43/7		
Open canopy, deep soil (6)															
Holcus hybrid	84 **					60/1	**927/15**				35/3		190/4		
Epilobium angudtifolium	40 **		1/1		12/4	32/8	**167/9**	1/1			27/2	40/2	51/6		
Juncus effusus	12						**4/3**				3/2		3/1		
Centaurea nigra	6						**5/1**								
Hypericum pulchrum	6						**3/1**								
Poa pratense	6						**5/1**								
Sandy loam (4-12)															
No species indicator of this group															
Closed hawthorn and elder canopy - (7)															
Veronica montana	41 **						1/1	**1/10**			25/5	8/2			
Digitalis dilatata	40 **		4/2	18/11	21/6	7/2	1/1	**26/16**	61/5	2/2	58/13	4/3	2/2		
Ilex aquifolium	10			2/2	2/2			**6/5**			2/1				1/1
Ligustrum ovalifolium	5							**1/1**							
Closed hawthorn and elder canopy (7,8,9)															
No species indicator of this group															

Species	IndVal (%)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Closed hawthorn canopy - start (8)															
Lamiastrum galeobdolon	31 **	96/5	35/1	512/16	3/2	25/1	1/1 1	11/3	**463/13**		89/9	150/5	33/4		
Closed hawthorn canopy - (8,9)															
No species indicator of this group															
Closed hawthorn canopy - end (9)															
Digitalis purpurea	6									**1/1**					
Closed canopy – Sandy loam (7-12)															
Sambucus nigra	35 **		15/1		3/3	1/1		**79/13**	**58/8**	**63/9**	**117/9**	**2/1**			
Milium effusum	16 **				1/1			**11/4**	**13/7**	**2/2**	**36/4**				
Closed canopy – start (10)															
Viola riviniana	13	25/1		1/1				1/1			**60/8**		1/1		
Glechoma hederacea	8										**19/7**				2/1
Equisetum arvense	7										**2/2**				
Carex sylvatica	4										**1/1**				
Melica uniflora	3			1/1							**3/1**				
Impeded drainage (10,11,12)															
Lysimachia nemorum	7										**30/2**		**7/1**		
Impeded drainage, closed canopy - start (11)															
Holcus mollis	47 **						95/8			205/6	13/4	**190/4**	90/3		
Potenilla sterilis	14											**1/1**	1/1		
Impeded drainage, moderate canopy cover (11,12)															
No species indicator of this group															
Impeded darainage, little canopy cover- end (12)															
Rubus idaeus	32 **						70/1				76/7		**101/7**		
Dactylis glomerata	15 **					1/1	2/2						**4/3**		1/1
Ranunculus repens	13						1/1				14/4	3/1	**61/4**		110/2
Stachys sylvatica	12							1/1	3/1	1/1			**4/3**		1/1
All habitats (1-15)															
Rubus fruticosus	68	**14/2**	**91/6**	**24/4**	**407/14**	**229/12**	**352/14**	**59/15**	**237/13**	**119/11**	**1156/29**	**95/4**	**704/12**	**166/8**	**215/7**
Hedera helix	37	**25/1**	**67/4**	**153/6**		**6/3**	**2**	**29/18**	**564/15**	**230/7**	**303/10**			**42/2**	**52/3**
Crataegus monogyna	7			**2/1**	**1/1**	**1/1**	**3/2**	**3/3**	**1/1**	**1/1**	**3/3**		**5/1**		
Loam, closed canopy – start (13)															
Mercurialis perennis	17			1/1					1/1		15/2			**120/2**	4/2
Loam, pH=5 (13, 14)															
Prunus avium	8													**1/1**	**10/1**
Loam, moderate canopy cover – end (14)															
Anthriscus sylvestris	57 **										31/6				**189/10**
Rumex sanguineus	20 **													2/1	**13/4**
Fagus sylvatica	16													1/1	**17/3**
Rumex obtusifolius	6														**2/1**
Taraxicum officinale	5				1/1	1/1					2/2		1/1		**2/2**
Athyrium filix-femina	5										5/1				**10/1**
Thin Loam (13,14,15)															
Geum urbanum	87 **		2/1						3/1		145/7			**258/10**	**492/16**
Fraxinus excelsior	48 **						1/1				10/9	21/2	3/3	**95/8**	**3/3**
Ranunculus ficaria	17 **	9/2									2/1			**17/2**	
Acer pseudoplatanus	10			3/3	1/1			4/4			5/1			**1/1**	**3/3**
Loam, moderate canopy cover, pH=6 (15)															
Arum maculatum	31 **	11/2		2/2											12/2
Bromus ramosus	8														
Circaea lutetiana	8					1/1	1/1	3/3			41/5	4/1	3/1	55/1	
Number of quadrats		12	6	38	15	13	16	21	17	16	29	5	12	10	16

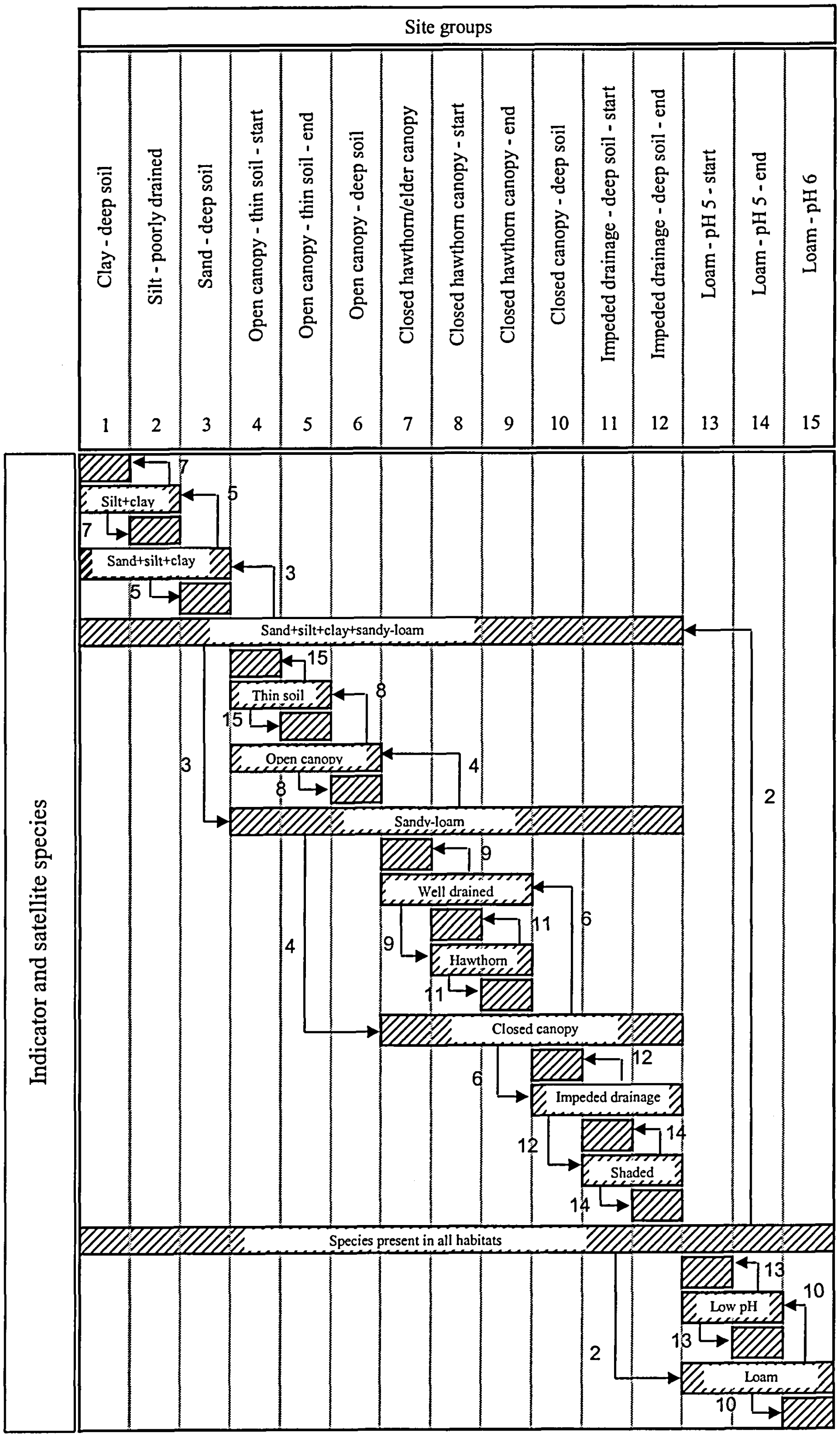


Figure 2.14. The structure of the two-way table presented in Table 2.38.

2.5 Discussion

2.5.1 Colonisation rates

Similar declines of woodland species, from an ancient wood / secondary wood boundary, have been reported by other authors. Bossuyt *et al.* (1999) and Brunet and Von Oheimb (1998) present several graphs depicting declines of species from ancient woodland boundaries at sites in Belgium. However, the regression equations they calculated cannot be compared with this study. This should simply be because the abundance data were collected using different sized sampling units. However, both the above papers have log transformed the distance from the ancient woodland and left the abundance data untransformed. The abundance data are likely to be a mixture of Poisson and normal distributions and probably strongly negatively skewed and so would have required a square root or log transformation. The distance data will have had a uniform distribution and so should therefore have not been transformed. There was also no attempt to model the quadratic relationships, which can be clearly seen from the graphs they have produced. However, despite these problems the graphs appear very similar for both the above papers and also those produced in this study. In particular, *Lamiastrum galeobolon* shows almost identical distributions as those presented in this study i.e. having its highest cover away from the ancient woodland boundary. This gives further evidence that this species is a poor competitor compared to other woodland species such as *Anemone nemorosa*, *Mercurialis perennis* and *Allium ursinum*, but it is able to spread at a faster rate.

The factors that determine the rates of spread of plant species such as *Hyacinthoides non-scripta* are also of interest. This species seems to be one of the first to colonise sites with deep sandy soils under a closed canopy. Under these conditions *Hyacinthoides non-scripta* would appear to be a very poor indicator species of ancient woodland. Dzwonko and Gawronski (1994) also reported a rapid spread of several woodland species on rich brown soils.

The equations used to calculate the colonisation rates appeared to show the greatest differences for *Anemone nemorosa* and to a lesser extent *Allium ursinum*. These two species frequently form continuous mats of vegetation, probably due to poor colonisation abilities and a strong competitive ability. Interpretation of any evaluation of these two species as potential indicators is therefore particularly sensitive to whether it is their abundance or presence that is used in calculating their colonisation rates. If the abundance is used, then the calculations may be for the time required to form dominant covers of these species. This has different implications from information on the time required for only a few individuals to be found amongst other vegetation. It may require a great length of time for these two species to become established in certain types of vegetation.

2.5.2 Succession and competition

Evidence for the competitive displacement of species can be seen with *Allium ursinum* displacing *Lamiastrum galeobdolon* and *Rubus fruticosus* displacing *Pteridium aquilinum*. *Rubus fruticosus* in turn appears to be unable to tolerate a high level of shade under a closed canopy, especially one with multiple layers formed by species such as *Crataegus monogyna* growing underneath the main canopy. Competition between species does make predicting the outcome of a succession more difficult when only a small number of sites have been studied. It would be very useful to set up competition experiments using a simple Latin square design with two or more species growing together under a number of different conditions e.g. a range of different light and nutrient levels. A more traditional means of analysis such as 2-way ANOVA could then be used to analyse the results.

2.5.3 Historical effects

Little Mattock Wood and Smithy Wood are the two sites where the newly colonising areas were situated on well-established abandoned pasture. The best linear unbiased predictors, for the comparison tests, which involved these two woods with other three sites, were relatively large. Some of this variation may be explained by the nutrient status of the soils but this may also be related to the historical management of these sites. The historical management will probably have exaggerated any difference in the nutrient status i.e. sites which were used for arable land were probably selected because they already contained deep rich soils whereas land used for pasture would have included the thinner, more nutrient poor soils. The land used for arable would probably have had nutrients added, whereas the soil used for pasture may have decreased in nutrient status overtime as the cut hay is removed from the site as well as leaching taking place. The site located within the historic parkland (Graves Park) was on soil of moderate nutrient status (the soil being a loam with areas rich in clay).

The planting of different tree species may also be an important factor. If no planting occurs it is likely that *Crataegus monogyna* and *Sambucus nigra* may form the first canopy cover. These species may be very persistent in the newly formed woodland. The findings of this study suggest that presence of an additional layer of canopy cover, especially from the former species can significantly reduce the cover of *Rubus fruticosus*. If the site is planted with tree species such as *Quercus* sp., *Fraxinus excelsior*, *Acer pseudoplatanus* or *Fagus sylvatica*, the species that often form an under-storey such as *Crataegus monogyna* and *Sambucus nigra* are not able to establish. However, if a site is planted with a high density of trees, a dense even-aged woodland is formed. With little light reaching the forest floor, this may also only contain a small amount of *Rubus fruticosus*. In terms of site management and conservation, the problem with such an even-aged stand is that the ground vegetation is likely to be dominated by just a single species such as *Hyacinthoides non-scripta* or *Anemone nemorosa*.

2.5.4 Implications for the creation of new woodlands

The results from the CCA analysis give an insight into the factors, which determine how a woodland community develops. *Pteridium aquilinum* appears to be restricted to sites wih relatively litlle shade, perhaps as a direct result from competition from *Rubus fruticosus*. The latter appears to be limited by a dense complete canopy cover. Both these species are often considered problem species for woodland managers. However, both may help facilitate the movement of other woodland species into a site, before a closed canopy develops. *Hyacinthoides non-scripta* was frequently found in small patches under *Pteridium aquilinum*. Likewise *Circaea lutetiana* and *Mercurialis perennis* were found growing in small patches under *Rubus fruticosus*.

If the control of species such as *Rubus fruticosus* and *Pteridium aquilinum* is desired, when creating a new woodland, then the creation of an under-storey is the most important factor. For this to be achieved the planting of species such as *Corylus avellana* and *Crataegus monogyna* should take place first or with the planting of the main tree species. If groups of these under storey species are planted then a more diverse woodland ground flora is likely to develope with no one species becoming dominant.

2.5.5 The role of the INDVAL procedure in plant community ecology

The INDVAL procedure, developed by Dufrene and Legendre (1997), appears to be able to cope equally well with plant percentage cover data as it did with the numbers of carabids in different environments. The method presented in this paper does differ slightly since the experimental design consisted of sampling across ecotones instead of in distinct habitat types. This meant that the clustering method used by Dufrene and Legendre (1997) i.e. *k*-means clustering, was not appropriate for this data set. The experimental design did follow a hierarchical design, so the assumption of a hierarchical structure could be imposed on this data set. The hierarchical clustering methods tried were all able to reproduce the hierarchical design of the experiment. The use of the CCA samples scores that are linear combinations of the environmental effects appeared to greatly ease the interpretability of the dendograms produced. This is in agreement with Palmer (1993) who found that the scores that are linear combinations of environmental effects produced very little noise compared to the weighted average scores or those scores produced from correspondence analysis. The use of constrained ordinations together with the INDVAL procedure therefore appears to be a very useful extension to the method proposed by Dufrene and Legendre (1997).

2.5.6 GLMMs and CCA

Random effects cannot be analysed directly using CANOCO. The inclusion of the variable for 'wood' as a fixed effect or random depends on the assumptions desired. If 'wood' is treated as a fixed effect then it is assumed that this variable includes all the variation of interest. In a local context this could be said to be true, since the woods chosen occur on the Coal Measures and Magnesian Limestone Series. So for example, species would be expected to behave in a similar way to factors such as soil depth, across the local region. However, if 'wood' is treated as a random effect then it is assumed that the variance explained is only part of the total variance from a larger population. The inclusion of wood as a random effect should therefore be more appropriate for applying the results to a wider geographical area, in which other factors not recorded may also be important such as altitude, precipitation or different soil types or topography which were not included in the analysis presented.

It is also difficult to analyse time series or spatially correlated data in CANOCO. Two methods exist. The first is the use of the cyclic shift option when performing the Monte Carlo F tests (ter Braak & Smilauer, 1998). The method used by CANOCO bends the time series into a circle, so that the start and ends meet. This means that all the data can be used and not just the data that is linked in both directions. The cyclic shift only corrupts the autocorrelation structure of each time series at the beginning and end of each series. For line transects, the dependence structure is not unidirectional as in time series. Usually, a point is related to its neighbours in both directions (ter Braak & Smilauer, 1998). The second method is to include the geographic coordinates of the samples as covariables. Linear, quadratic and cubic surface models can be included as covariables. Meot *et al.* (1998) suggest that the terms can be selected using 'forward selection of explanatory variables' and then these used as covariables in order to remove the effects of space. Variance partitioning can alo be used to separate out the variation derived from the spatial component and the environmental variables (see Anderson & Gribble, 1998; Meot *et al.*, 1998; Legendre & Legendre, 1998 and Chapter 3 of this thesis). However, in this project the variable for the geographic location i.e. distance along each transect, was one of the main variables of interest. The use of a random slopes model or a time series error structure was therefore preferred, although this cannot be done directly in CANOCO. CCA uses weighted linear regression of matrix Q bar (r×c) of correspondence analysis on the explanatory variables (matrix X). The predicted values from a separate analysis, such as GLMM, can therefore not be used to replace the sample scores which are linear combinations of the environmental variables, since the algorithm used by CANOCO preserves a specific distance measure. This means at present it is not easily to derive predicted sample scores which have resulted from analysis with random effects or which include autocorrelation functions. Instead CANOCO has been used to derive the constrained sample scores and general linear models have been used to produce more conservative estimates for the significance values of the fixed effects.

CHAPTER 3

3. BOTANICAL INDICATORS OF ANCIENT WOODLAND OF TARGET AREAS WITHIN SOUTH YORKSHIRE

3.1 Introduction

3.1.1 Ancient woodland indicators

Surprising little work has been carried out on the determination of botanical indicators of ancient woodland since Peterken's paper of 1974. In this Peterken limited the list of species under consideration to those species, that can tolerate the shade of a closed woodland canopy, that create the canopy or those that in some way require woodland conditions. Comparisons between the species present in ancient woodland and those species occurring in secondary woodland resulted in lists of species with varying degrees of being restricted to the ancient woodland. Peterken (1993) further refined this method to produce lists of species he considered to have a strong or mild affinity for ancient woodland. Those species, which had greater than 75 % of their occurrence in ancient woodland were listed as those having a strong affinity for ancient woodland and those species with 50 – 75 % of their occurrence in ancient woodland were listed as those species with a mild affinity for ancient woodland. However, the data he used are unbalanced, i.e. this list contains 89 ancient woodlands and 273 recent woodlands. A number of different authors (Honnay *et al.*, 1998 and Wulf, 1997) have collected similar data but have used chi-square tests to determine whether a species is significantly associated with ancient woodland. A third method of listing the species thought to be good indicators, based on the experience of a number of surveyors, with no collection or analysis of data has been used in the south of England (Rose, 1999) and also for the Peak District (Peak Park, unpublished).

Peterken's (1993) method suffers from the lack of a robust statistical test and the unbalanced nature of the data. However, the percentage values Peterken (1993) calculated can be converted to give values for an equal number of ancient and recent woods by using the equation below.

$$\begin{array}{c}\text{Proportional}\\ \text{occurrence in}\\ \text{ancient woodland}\end{array} = \frac{\text{occurrence in ancient woodland}}{\begin{array}{c}\text{occurrence in}\\ \text{ancient woodland}\end{array} + \left[\frac{\text{Number of ancient sites}}{\text{Number of recent sites}} \times \begin{array}{c}\text{occurrence in}\\ \text{recent woodland}\end{array}\right]} \tag{3.1}$$

This would change a value of 50% occurrence (mild affinity) in ancient woodland to 75% and change a value of 75% (strong affinity) to 90% occurrence in ancient woodland. Using this equation to convert these values means that lists produced from other data sets can be directly compared with Peterken's data.

As noted above chi-square tests have proved to be a popular method for determining indicator species. However these tests may be misleading. A species that is found to be significantly associated with ancient woodland may also be found to be fairly common in recent woodland. For example, from Honnay *et al.*'s (1998) list of species, *Teucrium scorodonia* occurred in forty-six out of fifty-eight ancient woods and thirteen out of forty-six recent woods and was found to be significantly associated with ancient woodland ($P < 0.001$). However, if the above equation is used so that these data can be compared with Peterken's (1993) list then it has a value of 74% of its occurrence in ancient woodland. This species would then not have made Peterken's list for species with a mild affinity for ancient woodland. Honnay *et al.* (1998) suggest that from their list of indicator species twenty-five, or more species are needed from a particular site to be confident that it is ancient. However, if we take Peterken's (1993) list of species with a strong affinity for ancient woodland and assume that this refers to greater than 90% occurrence in ancient woodland, then if one indicator species is found from this list then there is a 10% (or less) chance the site is recent, and if two species are found there is a 1% or less chance that the site is recent and if three then a 0.1% or less chance that the site is recent. If Peterken's (1993) list of species, with a strong affinity for ancient woodland, is used then only two or three indicator species may be needed to be confident that a site is ancient. A species present from Peterken's (1993) list behaves as an indicator in its own right where as it is suggested that a number of species from Honnay *et al.*'s (1998) list would have to found for them to behave as an indicator.

The method used by both Rose (1999) and by the Peak National Park ecologists, of speculating which species may be an indicator species of course provides no quantitative information to the surveyor as to how confident he/she should be that a site is ancient based on the number of indicator species found. However, this method is still useful for identifying the most species-rich sites and also for selecting sites for conservation. Using a shortlist of important species, making comparisons quick and easy between sites, significantly aids site evaluation and hence conservation.

3.1.2 TWINSPAN and INDVAL

A number of other methods for determining indicator species have been used on other data sets. These include the use of TWINSPAN and INDVAL (See section 2.3.8 for an explanation of these procedures).

3.1.3 Problems with spatial data

This study is not only concerned with those species that may be considered ancient woodland indicators and the methods likely to give the most reliable and easy to use species list. It is also concerned with, when different lists should be used for different regions.

If different regional lists are to be formed, then the major environmental factors influencing woodland plant distributions and occurrence need to be found. This will normally involve an ordination method using species composition data and then testing these axes produced, against known environmental factors. However, other factors that have not been recorded may be important. These unknown factors as well as the known factors would be expected to operate within a limited area, giving rise to spatially correlated data. There are a number of different methods that can be used with this type of data. The first is test for the presence of auto-correlation using a correlogram with a value of correlation determined for specified lag distances such as Moran's I or Geary's c (Cliff & Ord, 1981). A model can then be fitted using the known environmental data using multiple regression. The residuals can then be subjected to the same analysis using Moran's I or Geary's c to check whether any further auto-correlation exists with in the data. If the residuals remain spatially correlated then the results from the multiple regression should be treated with caution. Some computer programs also allow spatial structures to be assumed when fitting the regression equations, such as PROC MIXED in SAS (SAS Institute, 1994).

Another method popular with ecologists is variance partitioning. Here the proportion of variance explained by the environmental variables can be compared with the variance explained by spatial and/or temporal variables.

3.1.4 Moran's *I*

Before computing spatial auto-correlation coefficients, a matrix of geographic distances must be calculated. These are calculated for a number of different distance classes. A compromise has to be made between the resolution of the correlogram and the power of the test. Legendre and Legendre (1998) recommend the use of Sturge's rule (given below) to decide on the number of classes in histograms.

$$\text{Number of classes} = 1 + 3.3\log_{10}(m) \qquad (3.2)$$

where m is the number of distances in the upper or (lower) triangular distance matrix: the number of distances classes is rounded to the nearest integer.

The second problem with Morans'I is how to determine whether a coefficient for a given distance class is significant. Cliff and Ord (1981) describe how to compute confidence intervals and test the significance of spatial auto-correlation coefficients.

$$Stat_a = z_a\sqrt{Var(Stat)} - (n-1)^{-1} \qquad (3.3)$$

Where *Stat* is a confidence interval at significance level α.

An alternative method is to use a Monte Carlo permutation test such as the Excel 97 add-on, Rook's Case (Sawada, 2000). However, both these methods suffer from the same problems. The main concern is that the interest in the results of tests of significance decreases as the periods (or lags) get longer. This problem can be resolved by resorting to a progressive Bonferonni correction (Legendre and Legendre, 1998). In this method, the first periodogram is tested against the α significance level e.g. 0.05; the second and subsequent coefficients are tested against the Bonferroni-corrected level $\alpha' = \alpha/k$ (eg. at the second lag $\alpha' = 0.05/2 = 0.025$).

3.1.5 Variance Partitioning

It has recently become very popular in CCA analysis to calculate the proportion of variance explained by different sets of explanatory variables using a technique called variance partitioning. So, for example, spatial data can be entered as a large number of polynomials, usually the nine terms of a cubic trend surface model and then forward selection of explanatory variables selected and a Monte Carlo permutation test used to select a small number of significant terms (Anderson and Gribble, 1998; Bocard *et al.*, 1992; Bocard and Legendre, 1994; Meot *et al.*, 1998; Legendre and Legendre, 1998).

The use of polynomials especially using quadratic and cubic equations is an established method for modelling non-linear relationships. However, caution does need to be used when applying polynomials to ordination axes. The methods used by programs such as CANOCO assume unimodal responses of species to the environmental variable, so that, a single term for an effect actually models a quadratic response of a species to that variable. If a quadratic term is added, then a community would be assumed to change along an environmental gradient and then go back to something very similar as it started to at the opposite end of the gradient. It is difficult to imagine the circumstances that would lead to this. The only likely scenario would be sampling from patchy environments with latitude and/or longitude being used as environmental variables.

Palmer (1999) raises some further problems concerned with variance partitioning. He demonstrated that with simulated data for a perfect linear environmental gradient, the variance explained on the first axis, for his data, was only 0.243 as a proportion instead of 1.000. Many different types of ordination suffer from the arch effect so giving the appearance that large polynomials are needed to accurately model a community's response to an environmental gradient. Palmer (1999) demonstrated that if a quadratic effect is included to his perfect linear community then the variable erroneously appears to be important, and causes an arch effect. Palmer argues that this quadratic effect is a nuisance variable, since it does not explain species composition, instead, it clearly explains the arch effect. Palmer also demonstrated that it does not stop there, if cubic and quartic terms are added they also appear to explain "significant" additional variation. Okland (1999) has also independently discovered and analysed some of these problems. Okland (1999) demonstrates that any lack-of-fit of data to the response model implicit in any eigenvector ordination method results in polynomial distortion of ordination axes. It is also not possible to compare the total variation explained between data sets. There is currently no method for quantifying the relative contributions of gradient structure, polynomial distortion and random variation to total inertia, and the amount of "polynomial distortion variation" varies from data set to data set (Okland, 1999). Okland (1999) concludes that variance partitioning may still remain a useful tool for ecological analysis but that a shift of focus is needed from the total inertia and the 'unexplained variation' to relative amounts of variation explained. This means that instead of just quoting the percentage of explained variation, which may typically be very low as a result of problems associated with non-linear relationships, the relative difference between the spatial variation and the variation accounted for by a statistical model is of greater interest. If the variation explained by the spatial component is equal to or considerably larger then that accounted for by the environmental variables, then it is likely that another environmental factor or factors have been omitted from the analysis. In such cases, constraining the sample scores with the environmental factors may not be recommended, since at least one important variable is likely to be missing, which could give misleading results.

3.1.6 Modelling spatial structures within a GLM

Further analysis may be carried out on the ordination axes using separate programs. For example SAS can be used to model the environmental data to the ordination axes with a number of different spatial structures included in the model. This type of analysis avoids the problem of partialling out the spatial effects from the ordination. The spatial structures available in SAS include spherical, exponential and Gaussian. However, problems can occur when trying to fit these structures. The first is that the procedure may not converge, even when using the different options available such as increasing the number of iterations and using the Fisher scoring algorithm instead of the Newton-Raphson algorithm (Littell *et al.*, 1996). The second problem is that it is very difficult to find out whether the parameters used in the spatial structure are sensible values. Semi-variance can become inflated at large distances, since the reliability of these values is very suspect because they are computed from too few values. There is no option in SAS to restrict the fitting of a semi-variogram to a maximum distance.

The alternative methods available are to include spatial effects in the fixed or random effects. If they are included as fixed effects then the assumption used would be that all the variation of interest is included in the latitude and longitude. Since we know that species may have a northwest to southeast preference in their distribution then it may also be reasonable to constrain the ordination by these effects. If the spatial effects are included as random effects then it assumed that they do not take into account all the variation of interest and can only be used to give pseudo-tests for the other variables. If the spatial terms are treated as fixed effects then the only interest maybe in whether there is a significant change with latitude, longitude and the interaction between the two. If the spatial terms are treated as random effects then all the terms for a quadratic or cubic surface model can be used. A backward stepwise procedure can be used to remove terms in the model which are not significant by using the change in the value for –2 Restricted Log Likelihood with one degree of freedom for each time a term is removed. If removing a term causes a significant increase in the –2RLL (looked up in a chi-square table) then no more terms should be removed (Verbeke, 1997).

3.2 Materials and Methods

3.2.1 Study area

The study area consists of selected sites across South Yorkshire, extending into North Derbyshire and the southern edge of West Yorkshire. Approximately 6% of the area of South Yorkshire is covered by woodland compared with the national average of 9% (NCC, 1986). About 47 % of South Yorkshire's woodland is thought to be ancient (NCC, 1986). Altitude, rainfall, and geology all appear to play an important role in determining the types of woodland in the study area. A small number of woodlands and plantations exists on the gritstone at a relatively high altitude to the west of the County. These ancient woodlands are typically birch – sessile oak woodlands, with a vegetation cover including *Holcus mollis* and *Deschampsia flexuosa*. In the middle of the County lies the Coal Measures at a lower altitude, again supporting mainly birch sessile oak woodlands, often with a high percentage cover of *Hyacinthoides non-scripta*. A narrow band of Magnesian Limestone and Permian Marl runs North – South towards the east of the county. This geology type consists of beds of Magnesian Limestone separated by bands of clay and sand of the Permian Marl. This geology types supports mainly ash and elm woodland with species such as *Allium ursinum*, *Mercurialis perennis* and *Sanicula europaea* in abundance. Further east there is very low lying ground situated on the Bunter Sandstone. Many of these woods have developed on sites, which have been used for the extraction of sand and are typically birch woodland and scrub (NCC, 1986).

A map showing the locations of the sites that have been included in the analysis is shown in Figure 3.1.

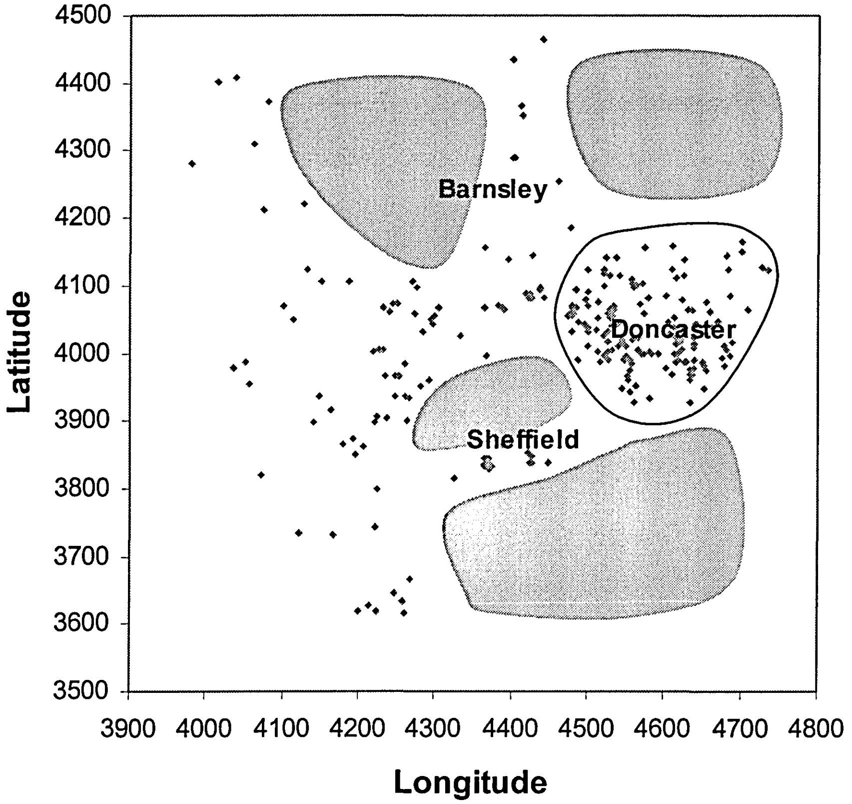


Figure 3.1. Map of the study area showing the biogeographic coverage of sites included in the analysis. The area marked around Doncaster has been particularly well surveyed. The shaded areas have received very little systematic survey work suitable for this analysis.

3.2.2 Data collection

Species lists were collated from a number of different Biological Record Centres and other sources. The survey effort varies considerably for different sites throughout the surveyarea For example the sites that were thought to be of the greatest nature conservation status or close to urban areas have often been more thoroughly surveyed. Most regions have had most of their sites thoroughly surveyed, but there was little consistency as to how species have been recorded. This is particularly with respect to whether species had been recorded as present or absent, or whether an abundance scale was used such as DAFOR. There is also variation in whether a site had been split up into smaller compartments, the time spent surveying, the time of year, and whether hybrids had been recorded separately. Only Doncaster Metropolitan Borough Council had systematically surveyed all the sites that they considered to be Sites of Scientific Interest to the same standard. An emphasis has therefore been put on the statistical analysis of indicator species of the Doncaster region and the distributions of some of the commoner woodland species across this region. Sample sites also differed according to whether they had been surveyed only once or a number of times. For the purposes of this study, a sample of the woods surveyed were selected that had been approximately surveyed to a standard level. Since, most of the data available had been recorded using the DAFOR system for species abundance and contained sites which have had received only a single survey with the date recorded, only records that met these criteria were included in the analysis. This facilitates certain statistical compliance and is therefore valid. This may differ fundamentally from some other equally (but differently) valid assessment of woodland species but which may not be amenable to statistical interrogation. Through using data of a common standard, it meant that a number of different methods of analysis could be tried on the same data. Some new sites were surveyed during the summer of 1999 and the spring of 2000, for the purposes of this study.

The origin of each wood was ascertained with the help of the information in the Inventories of Ancient Woodland produced by the then Nature Conservancy Council (NCC, 1986). This information provided evidence as to whether a wood of greater than 2 ha in area should be considered as ancient woodland. The surveyor's reports often included a description of the site which would normally include whether the site was semi-natural, broadleaf/conifer plantation or developing scrub. Other sites had documented historical records as to when the site was planted or abandoned. Complicated sites such as areas of developing scrub, which contained old hedgerows, were not included. The sites were split into three categories – ancient woodland (greater than 400 years old), secondary woodland (pre-Second World War – mainly mature mixed broadleaf and coniferous plantations) and recent (post-Second World War – less than 50 years old, consisting mainly of newly planted woodlands, developing scub and developing birch woodland). This method may be contrasted to that of Honnay *et al.* (1999) who only used isolated woods in their analysis and considered woods of greater than 215 years as ancient.

3.4. Data analysis

3.4.1 Partial CCA

Canonical correspondence analysis was used to produce a constrained ordination. The area (log transformed) of each site and the survey day from the first of January were entered as covariables, so removing their influence from the ordination. The subsequent ordination was then constrained by altitude, longitude, latitude, longitude × latitude, the age category (recent, secondary, ancient), the altitude, the geology (Millstone Grit, Coal Measures, Magnesian Limestone, Permian Marl and Bunter Sandstone), the dafor rating of introduced tree species (beech, sycamore, sweet chestnut, pine and larch), the adjacent land use (urban, and whether a recent or secondary habitat was bordered by any ancient woodland or hedgerow). The species data contained 585 species and 285 sites. In order to ease the handling of the data the number of species used was reduced to 255. INDVAL has a limit of 300 species and Excel was the main method used for entering and handling the data, which has a limit of 256 columns. The species included in the analysis, were from those that Peterken (1973) considered as having a mild or strong association with ancient woodland, the species that Rose (1999) considered to be indicators of ancient woodland in the south of England, those species considered by Wulf (1997) to be indicators of ancient woodland in north west Germany and the species listed by Honnay *et al.* (1998) as being associated with ancient woodland in Belgium. The remaining species were those with the greatest sum of their abundance scores, within the data set, which mainly consisted of common woodland and some common grassland species. The DAFOR ratings were entered to give a value of 1 for rare, 2 for locally occasional, 3 for occasional, 4 for locally frequent, 5 for frequent, 6 for locally abundant, 7 for abundant, 8 for locally dominant, and 9 for dominant. Since this type of information is ordinal no transformation was used. Since a large number of rare non-woodland species had already been omitted, no down weighting of rare species was used.

A correlation matrix was produced using Spearman rank correlation coefficients. This was carried out using SAS.

3.4.2 Constrained cluster analysis

Since, CANOCO produces sample scores that are chi-square distances the distances have euclidean properties so can be used in a wide range of different clustering procedures. Two methods of cluster analysis were used – Ward's method and Flexible clustering (beta = -0.5 and –0.25). The first four, six and eight axes were entered into the cluster procedure to see which gave the clearest separations.

3.4.3 Variance Partitioning

The method of variance partitioning described by Anderson and Gribble (1998) was used on the species and environmental variables, which was conducted on CANOCO. The data were divided up into environmental factors, the spatial variables and the temporal variable. The matrix of spatial variables was calculated by including all terms for a cubic trend surface regression, with x = longitude (centred) and y = latitude (centred). So that the nine terms included were $x, y, x^2, xy, y^2, x^3, x^2y, xy^2$, and y^3. These nine terms were then included in a procedure of 'forward selection' using CANOCO. The environmental variables were altitude, the age categories, the geology categories and variables for an urban boundary and if a recent or secondary site was adjacent to an ancient site and also the log of the area. The date (number of days since January 1) was entered as the temporal matrix.

The total trace, or sum of all eigenvalues, obtained by an unconstrained correspondence analysis of the species data provides a measure of the total variation in the species data. The sum of canonical eigenvalues obtained by any CCA is a proportion of the total trace obtained by CA on that data (Bocard *et al.*, 1992; Anderson and Gribble, 1998).

A series of steps, involving constrained and/or partial CCA, were done using CANOCO (Table 3.1). For an explanation of how the explained variance is calculated for each component see Anderson and Gribble (1998).

Table 3.1. Steps in the analysis to perform variance partitioning using CANOCO (taken from Anderson and Gribble, 1998).

Step	Description
[1]	CCA of species matrix, constrained by the environmental matrix
[2]	CCA of species matrix, constrained by the spatial matrix
[3]	CCA of species matrix, constrained by the temporal matrix
[4]	CCA of species matrix, constrained by the environmental matrix, with spatial variables treated as covariables
[5]	CCA of species matrix, constrained by the environmental matrix, with temporal variables treated as covariables
[6]	CCA of species matrix, constrained by the environmental matrix, with spatial + temporal variables treated as covariables
[7]	CCA of species matrix, constrained by the spatial matrix, with environmental variables treated as covariables
[8]	CCA of species matrix, constrained by the spatial matrix, with temporal variables treated as covariables
[9]	CCA of species matrix, constrained by the spatial matrix, with environmental + temporal variables treated as covariables
[10]	CCA of species matrix, constrained by the temporal matrix, with environmental variables treated as covariables
[11]	CCA of species matrix, constrained by the temporal matrix, with spatial variables treated as covariables
[12]	CCA of species matrix, constrained by the temporal matrix, with environmental + spatial variables treated as covariables

3.4.4 Moran's *I*

The degree of spatial auto-correlation was determined for each of the first four axes produced by the partial CCA analysis. This was conducted on GSWIN (Gamma Design Software, 1999). This program uses Sturge's rule to determine the number of classes and gives maximum distance recommended for calculating an auto-correlation coefficient. Once these values had been recorded then Rook's Case (an add-on for Excel 97) was used. This program calculates confidence intervals and z values for each auto-correlation coefficient. Rook's Case can also be used to calculate the estimated probability by using a Monte Carlo permutation test. Once the probabilities had been determined for each lag interval a progressive Bonferonni correction was applied for identifying which coefficients were significant on correlograms, with α equal to 0.05.

The above procedure was repeated for the residuals after a weighted multiple regression analysis had been carried out on the weighted average scores using the same environmental variables used in CANOCO. This was conducted on SAS. All the variables were first standardised by centring them and dividing by their standard deviation. The weights calculated by CANOCO were used to weight all the sample scores.

3.4.5 Spatial GLMMs

General Linear Models were used to fit the ordination axes to the environmental data using models with a spatial covariance structure specified and also a separate analysis using random slope models for cubic and quadratic surfaces. The PROC MIXED procedure in SAS was used for fitting these models. All variables were first centred and standardised by their standard deviation and the weights calculated by CANOCO used to weight the sample scores.

Three different spatial structures were applied to the data, these were spherical, exponential and Gaussian. The Fisher scoring method was used as recommended by Littell *et al.* (1996). If the analysis had not converged by 50 iterations then the program was rerun with the last parameter estimates given as starting parameters for the sill, range and nugget effects. The model which gave the value for the –2 Restricted Maximum Likelihood (REML) that was closest to zero i.e. contained the smallest error was considered to be the model giving the best fit.

A separate analysis was also done but instead of specifying a spatial structure random slopes were applied. The nine terms of a cubic trend surface were entered as random effects. The effect that was found to be least significant was removed from the model and the change in the –2 REML used as a significance test (chi-square test with one degree of freedom). If the removal of a parameter caused a significant increase in the –2 REML then that parameter was kept and no further terms were removed.

3.4.6 Indicator Analysis

A number of different methods were applied to the data. The first consisted of following the method used by Dufrene and Legendre (1997) once a hierarchical dendogram had been decided upon. The program INDVAL was used to determine where species occurred on the dendogram, give them an indicator value and perform significance tests.

The second set of analyses was again based on the method using INDVAL. Simplified dendograms were entered, i.e. giving all sites and then simply dividing them up into recent and ancient. The program INDVAL was then used to determine which species were associated with the ancient woodland. The species data contained the DAFOR rating for the abundance of each species using the values from zero to nine. The species found to be associated with the ancient woodland were then divided up into categories according to their percentage occurrence in ancient woodland. The equation given in the introduction was used to correct for the unbalanced design in the data. Species were divided up into categories for 90-100% occurrence in ancient woodland, 75-90%, 60-75% and 50-60%. Chi-square tests were also applied to the data as a comparison to the INDVAL significance test. If any of the expected values were less than five, then Fisher's exact tests were used. This was conducted on SPSS (SPSS Inc., 1990). The data were also split up according to geological type and then the same procedure applied to the data to find the ancient woodland indicator species for the three main geology types i.e. Coal Measures, Bunter Sandstone and also Magnesian Limestone and Permian Marl combined.

3.5 Results

3.5.1 Partial CCA

The results of the canonical correspondence analysis are shown in Tables 3.2 and 3.3 after the effects of time since January 1 and the log of the area have been removed by partialling them out. The eigenvalues and correlation coefficients are given in Table 3.2 and the *t*-values of the regression coefficients produced by CANOCO are given in Table 3.3. The eigenvalues can be seen to be fairly low, which is probably due to the large number of species that have been included in the analysis, since only a small number of which would be expected to be sufficiently abundant to show clear unimodal trends across the CCA axes. A further problem, which is discussed in section 3.5.8, is that a large number of samples are located in the centre of the ordination.

The first axis appears to be mainly associated with the effects for altitude and the geology i.e. the limestone woods are negatively significant on the first axis compared to the woods on sandstone which are positively associated with the first axis. The abundance of sycamore and sweet chestnut were also found to be significant on this axis. The second axis appears to mainly explain the variation between recent, ancient and secondary woods as well as the effects of neighbouring ancient woodland by the recent and secondary woods. The third and fourth axes appear to mainly explain the variation attributed to longitude and latitude and the interaction between the two. The multivariate F test for the first four axes appears to indicate that the main sources of variation can be attributed to altitude, the age category of the sites and the geology.

A correlation matrix of Spearman rank correlation coefficients is given in Table 3.4. A large number of the independent variables can be seen to significantly correlated with each other. In particular the variable for Millstone Grit can be seen to significantly correlated with high altitude and longitude (west) and the variable for Bunter Sandstone can be seen to be highly correlated with low altitude and longitude (east). The time of year that the sites were surveyed, can be seen to be negatively correlated with the recent woods and positively correlated with the secondary woods. This may be a because the recent sites had not been classified as woodlands prior to the survey, and so a more appropriate time for surveying was considered to be later in the summer when more grassland species would be in flower. This means that there may be a bias towards not finding early spring flowering species such as *Hyacinthoides non-scripta* and *Anemone nemorosa* in the recent sites.

Table 3.2. Eigenvalues and correlation coefficients for the first four CCA axes produced by CANOCO. Total inertia (sum of all eigenvalues) = 4.889, sum of canonical eigenvalues = 0.884.

	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Eigenvalues	0.27	0.14	0.08	0.05
Correlation coefficients	0.91	0.83	0.79	0.60

The ordinations of the sample scores produced by the partial CCA analysis are shown in Figure 3.2 and Figure 3.3. Figure 3.2 shows a clear separation of sites on the Millstone Grit to the right of the ordination along axis 1 and sites on the Magnesian Limestone and Permian Marl on the left of the ordination. The recent sites can be seen to be separated along axis 2 and occur at the top of the ordination. The ordination produced by the third and fourth axes (Figure 3.3) is less clear. The sites on The Bunter Sandstone are separated out on axis 3. The fourth axis appears to separate the sites on the Gritstone, that are in West Yorkshire, from those that are to the west of Sheffield, in South Yorkshire.

The ordination of the species scores on the first two axes is shown in Figure 3.4. The species towards the ends of the axes have been displayed showing some of the possible indicator species for these different woodland types. For example, *Blechnum spicant*, and *Calluna vulgaris* can be seen to occur in the direction of the Gritstone woods, *Brachypodium sylvaticum* and *Mercurialis perennis* can be seen to occur in the direction of the Magnesian Limestone and *Chamerion angustifolium* and *Achillea millefolium* can be seen to occur in the direction of the recent sites.

Table 3.3. *t*-values of regression coefficients for standardised variables. F – Monte Carlo permutation for the first four axes under a full model.

Effect	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4	F
	t	t	t	t	
Altitude	4.01 ***	0.84	2.43 **	2.96 **	12.23 ***
Longitude	-1.56	0.58	-5.39 ***	5.41 ***	1.84 **
Latitude	-1.30	0.44	-5.37 ***	5.42 ***	2.21 **
Longitude × Latitude	1.24	-0.37	5.56 ***	-5.17 ***	2.20 **
Recent	-1.57	9.60 ***	5.04 ***	-1.61	6.95 ***
Ancient	0.79	-3.63 ***	-1.42	-0.52	1.53 *
Secondary	0.00	0.00	0.00	0.00	
Urban	-2.30 *	0.66	-0.10	3.05 **	1.61 **
Recent × Ancient Boundary	2.04	-2.68 **	-0.30	-1.49	1.61 **
Secondary × Ancient Boundary	-1.12	-2.25 *	-0.08	1.00	1.13
Bunter Sandstone	2.89 **	1.68	-2.36 *	-1.79	1.68 *
Permian Marl	0.17	0.76	-0.07	-0.37	2.05 **
Magnesian Limestone	-1.14	0.06	0.44	-0.45	6.60 ***
Coal Measures	1.82	2.41 *	-2.35 *	-0.84	1.61 *
Millstone Grit	3.31 **	2.02	0.45	0.20	2.79 **
Carboniferous Limestone	-2.72 **	2.23 *	2.16 *	0.28	4.00 ***
Sycamore (DAFOR)	-3.48 ***	-3.22 **	2.12 *	-3.96 ***	1.91 **
Beech (DAFOR)	-1.03	-3.13 **	-2.78 **	0.13	1.34 *
Sweet Chestnut (DAFOR)	3.57 ***	-0.37	0.38	-2.75 **	1.23
Larch (DAFOR)	-0.97	-0.98	1.08	-0.67	1.28
Pine/Spruce (DAFOR)	1.25	2.85 **	-1.39	4.60 ***	1.74 *

Table 3.4. Spearman rank correlation matrix of the explanatory variables andovariables used in theCCA analysis.

	Altitude	Longitude	Latitude	Recent	Ancient	Secondary	Ancient boundary	Urban boundary	Bunter Sandstone	Permian Marl	Magnesian Limestone	Coal Measures	Millstone Grit	Carboniferous Limestone	Sycamore	Beech	Sweet Chestnut	Larch	Pine	Time
Longitude	-0.93 ***																			
Latitude	-0.40 ***	0.25 ***																		
Recent	-0.11	0.09	-0.11																	
Ancient	0.41 ***	-0.42 ***	-0.04	-0.61 ***																
Secondary	-0.40 ***	0.42 ***	0.17 **	-0.26 ***	-0.61 ***															
Ancient boundary	0.32 ***	-0.30 ***	-0.32 ***	0.19 ***	-0.03	-0.15 *														
Urban	-0.14 *	0.12 *	0.04	0.06	-0.15 *	0.12 *	-0.10													
Bunter Sandstone	-0.71 ***	0.74 ***	0.17 **	0.01	-0.30 ***	0.35 ***	-0.35 ***	0.16 **												
Permian Marl	-0.24 ***	0.20 ***	0.00	0.02	-0.18 **	0.19 ***	-0.01	0.03	-0.16 **											
Magnesian Limestone	-0.05	0.12 *	0.28 ***	-0.10	0.10	-0.03	-0.06	-0.09	-0.28 ***	-0.11										
Coal Measures	0.35 ***	-0.44 ***	-0.15 *	0.23 ***	0.03	-0.26 ***	0.32 ***	0.02	-0.41 ***	-0.20 ***	-0.33 ***									
Millstone Grit	0.57 ***	-0.55 ***	-0.22 ***	-0.20 ***	0.30 ***	-0.17 **	0.03	-0.14 *	-0.23 ***	-0.11	-0.19 ***	-0.28 ***								
Carboniferous Limestone	0.19 ***	-0.20 ***	-0.24 ***	-0.07	0.12 *	-0.07	0.11	-0.07	-0.08	-0.04	-0.07	-0.10	-0.06							
Sycamore	0.16 **	-0.15 *	0.11	-0.32 ***	0.27 ***	-0.01	-0.05	0.03	-0.27 ***	0.01	0.32 ***	0.03	-0.05	0.01						
Beech	0.36 ***	-0.35 ***	0.05	-0.34 ***	0.40 ***	-0.15 *	0.02	0.00	-0.25 ***	-0.16 **	0.12	0.09	0.17 **	0.05	0.40 ***					
Sweetchestnut	-0.11	0.15 *	-0.01	-0.21 ***	0.09	0.10	-0.02	0.08	0.22 ***	-0.13 *	-0.04	-0.07	-0.02	-0.07	0.12 *	0.27 ***				
Larch	0.25 ***	-0.26 ***	-0.04	-0.24 ***	0.33 ***	-0.16 **	0.04	-0.12 *	-0.10	-0.15 *	0.00	-0.04	0.24 ***	0.19 **	0.13 *	0.36 ***	0.15 *			
Pine	0.18 **	-0.17 **	-0.01	-0.28 ***	0.26 ***	-0.03	-0.07	-0.10	0.06	-0.16 **	-0.11	-0.09	0.25 ***	0.13 *	0.12 *	0.41 ***	0.30 ***	0.51 ***		
Time	-0.40 ***	0.43 ***	0.20 ***	-0.22 ***	-0.08	0.32 ***	-0.24 ***	0.11	0.39 ***	0.33 ***	-0.06	-0.50 ***	0.07	-0.09	-0.09	-0.18 **	0.14 *	-0.07	0.07	
Area	0.02	-0.06	0.32 ***	-0.38 ***	0.39 ***	-0.09	-0.17 **	0.02	0.01	-0.04	0.08	-0.19 ***	0.19 ***	0.07	0.28 ***	0.29 ***	0.23 ***	0.34 ***	0.37 ***	0.13 *

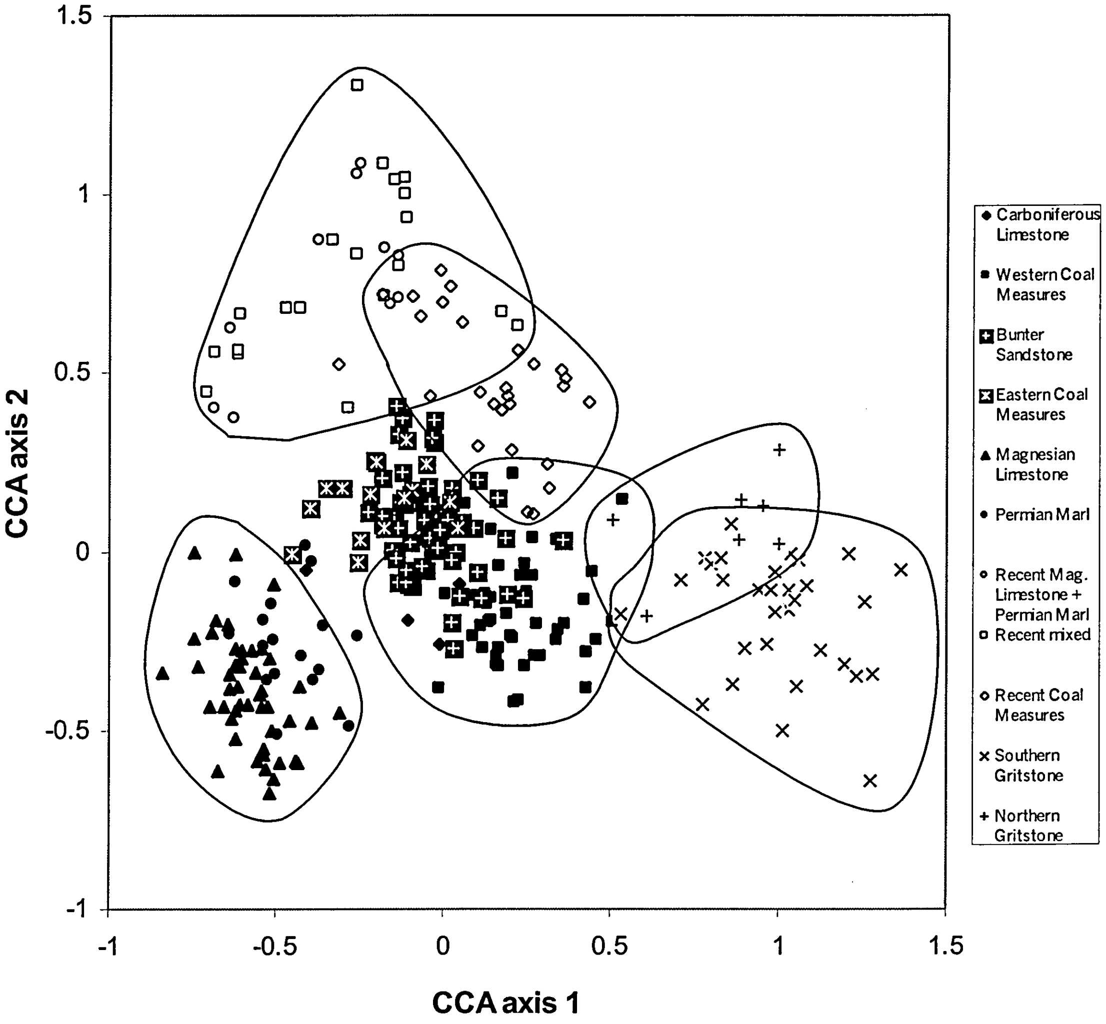


Figure 3.2. Ordination of sample scores on the first axes produced by canonical correspondence analysis. The polygons and symbols refer to the groups obtained by the Ward's clustering method shown in Figure 3.6.

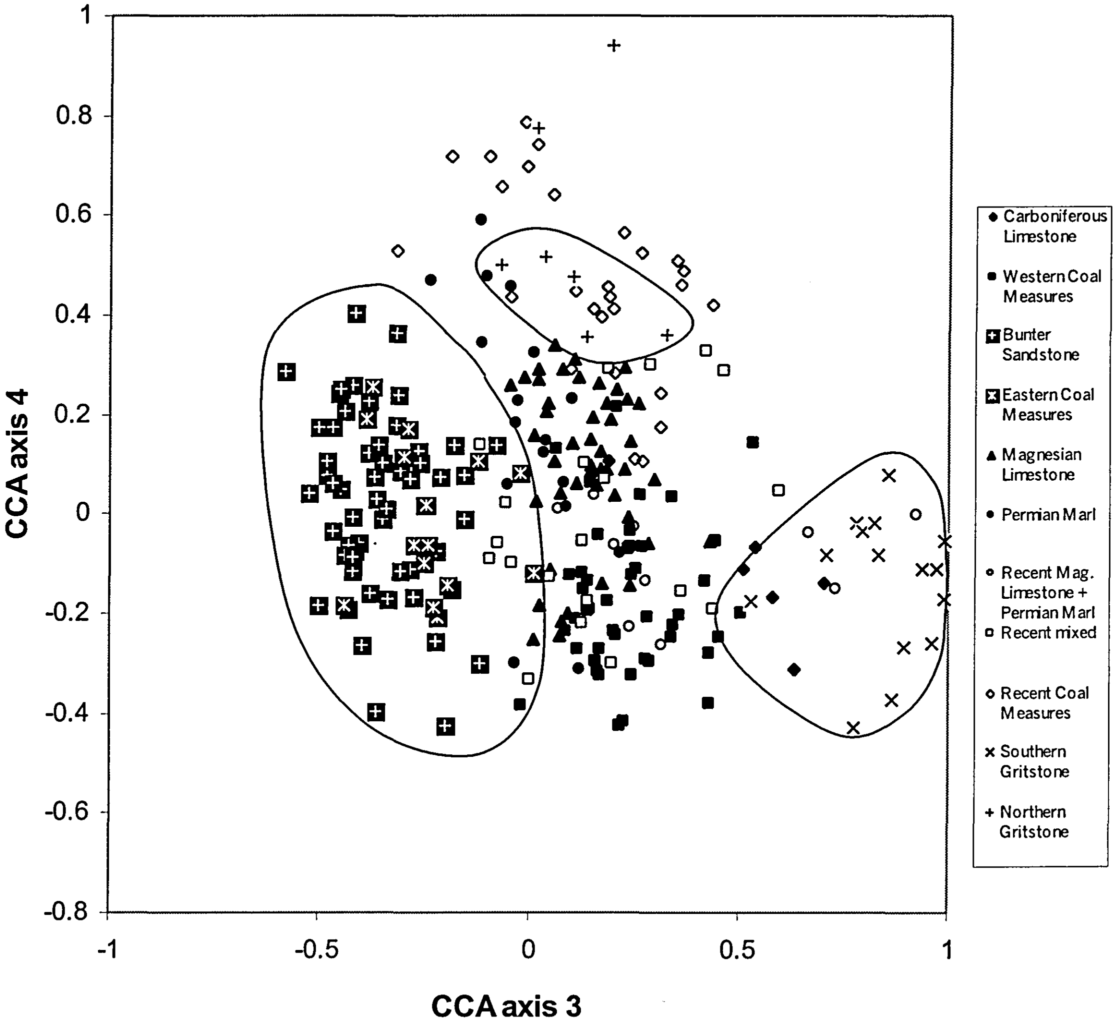


Figure 3.3. Ordination of the sample scores produced on the third and forth axes produced by canonical correspondence analysis. The symbols refer to the groups obtained with the Ward's clustering method. Polygons have been drawn round the sites on the Bunter Sandstone, the northern Millstone Grit and the southern Millstone Grit.

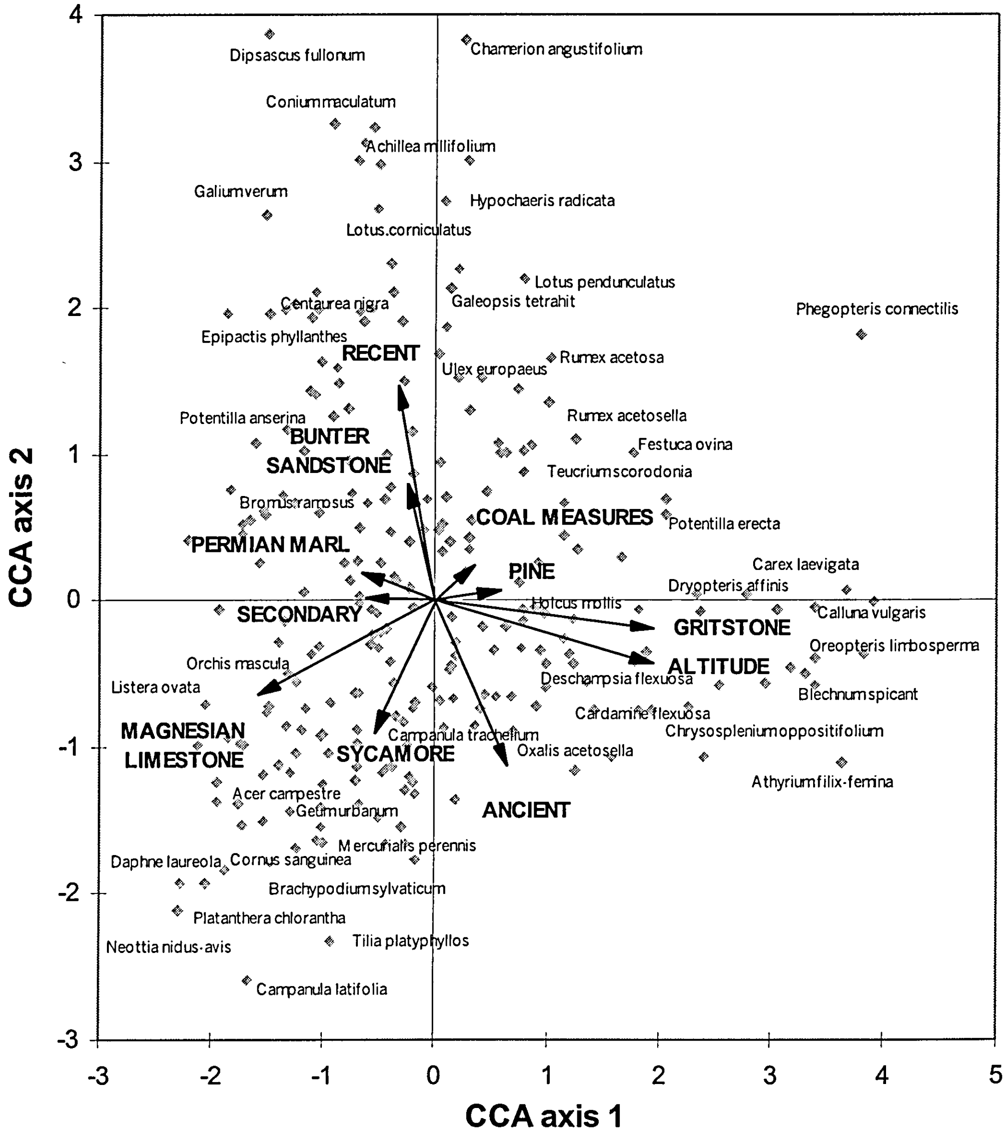


Figure 3.4. Ordination of the species scores on the first two axes produced from canonical correspondence analysis.

3.5.2 Spatial auto-correlation

The model used in CANOCO was fitted using multiple regression, on the sample scores that are weighted averages of the species values, in SAS. All the variables were first standardised by centring then on their mean and then dividing by their standard deviation. The same weights were also applied to the data as those calculated by CANOCO. Moran's *I* was used to examine the weighted average scores produced by CANOCO for any spatial auto-correlation and also on the residuals produced by the weighted multiple regression equation. The resulting correlograms are shown in Figure 3.5. A high degree of spatial auto-correlation can be seen to occur in the scores produced by CANOCO. However, after the mode has been fitted then very little auto-correlation can be seen to remain in the data. This suggests that the model has fit reasonably well i.e. the variables used in the analysis explain most of the variation that is contributing to this spatial auto-correlation.

Two further tests were applied to the data. The first is variance partitioning, which can be used to calculate the proportion of the variance that may be attributed to a number of different sources. The second method involved the fitting of general linear models with a spatial covariance structure specified.

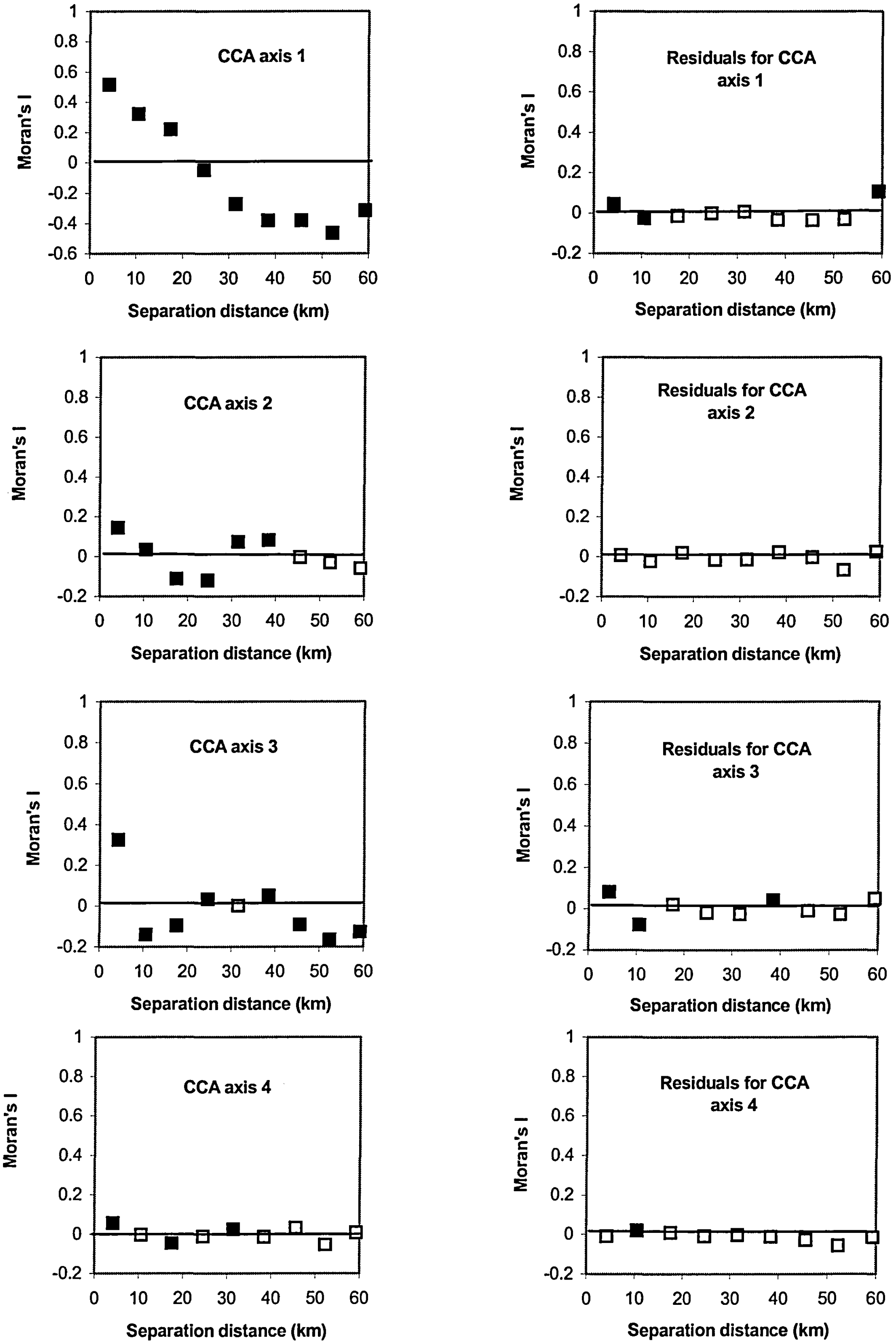


Figure 3.5. Spatial auto-correlogram for the CCA sample scores that are weighted averages of the species matri and the residuals produced after the same model has been fitted on the data using multiple regression. Dark symbols correlation statistics that are significant after progressive Bonferroni correction ($\alpha = 0.05$).

3.5.3 Variance Partitioning

The variables used in the partial CCA analysis of the previous section were separated into three different groups. The first contained the environmental data minus the effects for longitude and latitude and the interaction between the two, but did include the log of the area of the sites. The second contained the spatial variables, which consisted of the cubic surface model for longitude and latitude. The third contained the temporal variable i.e. the number of days since January 1. The spatial variables were first entered as the only environmental variables present and the forward selection option used in CANOCO. Monte Carlo permutations were used to test for the significance of the nine terms used in the cubic response model. The results are given in Table 3.5. All effects can be seen to be significant. The main sources of variation appear to be for the polynomials for longitude.

Table 3.5. Forward selection of spatial variables, Monte Carlo F under a full model. Longitude and Latitude were first centred on their means to reduce collinearity in the spatial data.

Spatial variable	F
Longitude	11.52 ***
Longitude^2	5.06 ***
Longitude^3	4.43 ***
Latitude^2	3.16 **
Longitude × Latitude	2.31 **
Latitude	3.04 **
$\text{Longitude} \times \text{Latitude}^2$	2.48 **
Latitude^3	2.55 **
$\text{Longitude}^2 \times \text{Latitude}$	2.04 *

The sum of canonical eigenvalues and corresponding percentage of explained variation for each of the CCAs indicated by steps [1]-[12] of Table 3.1 are shown in Table 3.6. This table also shows the probabilities of significance from Monte Carlo permutation tests for each CCA with 1000 permutations done for each test.

The calculations of Anderson and Gribble (1998) were repeated so that the explained variation of the three components could be ascertained as well as the overlap of variation occurring between the different components. The calculations are shown in Table 3.7. A graphical representation of the explained variation is shown, with the aid of rectangles with their areas corresponding to the components of variation, in Figure 3.6.

Table 3.6. Summary of results of constrained and partial canonical correspondence analysis. See Table 3.1 for an explanation of the steps used in the analysis.

Step in Analysis	Value in Calculations	Sum of canonical Eigenvalues	Explained Variation (%)	P
[1]	Ω_E	0.953	19.49	< 0.001
[2]	Ω_S	0.593	12.13	< 0.001
[3]	Ω_T	0.087	1.78	< 0.001
[4]		0.559	11.43	< 0.001
[5]		0.914	18.70	< 0.001
[6]	E	0.608	12.44	< 0.001
[7]		0.256	5.24	< 0.001
[8]		0.556	11.37	< 0.001
[9]	S	0.251	5.13	< 0.001
[10]		0.048	0.98	< 0.001
[11]		0.051	1.04	< 0.001
[12]	T	0.042	0.86	< 0.001

Table 3.7. Summary of calculations for the partitioning of variation with spatial, temporal and environmental variables.

Component	Calculation	Explained variation (%)
Environmental = E	[6]	12.44
Spatial = S	[9]	5.13
Temporal = T	[12]	0.86
SE	$SE_T - STE$ = [1] – [4] – STE	7.40
TE	$TE_S - STE$ = [1] – [5] – STE	0.14
ST	$ST_E - STE$ = [2] – [8] – STE	0.10
STE	[9] + ([2] – [7]) + ([2] – [8]) – [2], or [6] + ([1] – [4]) + ([1] – [5]) – [1], or [12] +([3] – [10]) + ([3] – [11]) – [3]	0.65
Total explained	[1] + [7] + [12], or [2] + [4] + [12], or [3] + [5] + [9], or $S + T + E + SE + TE + ST + STE$	25.59
Unexplained	100% – Ω	74.41
Total variation	Total trace of CA = 4.889	100.00

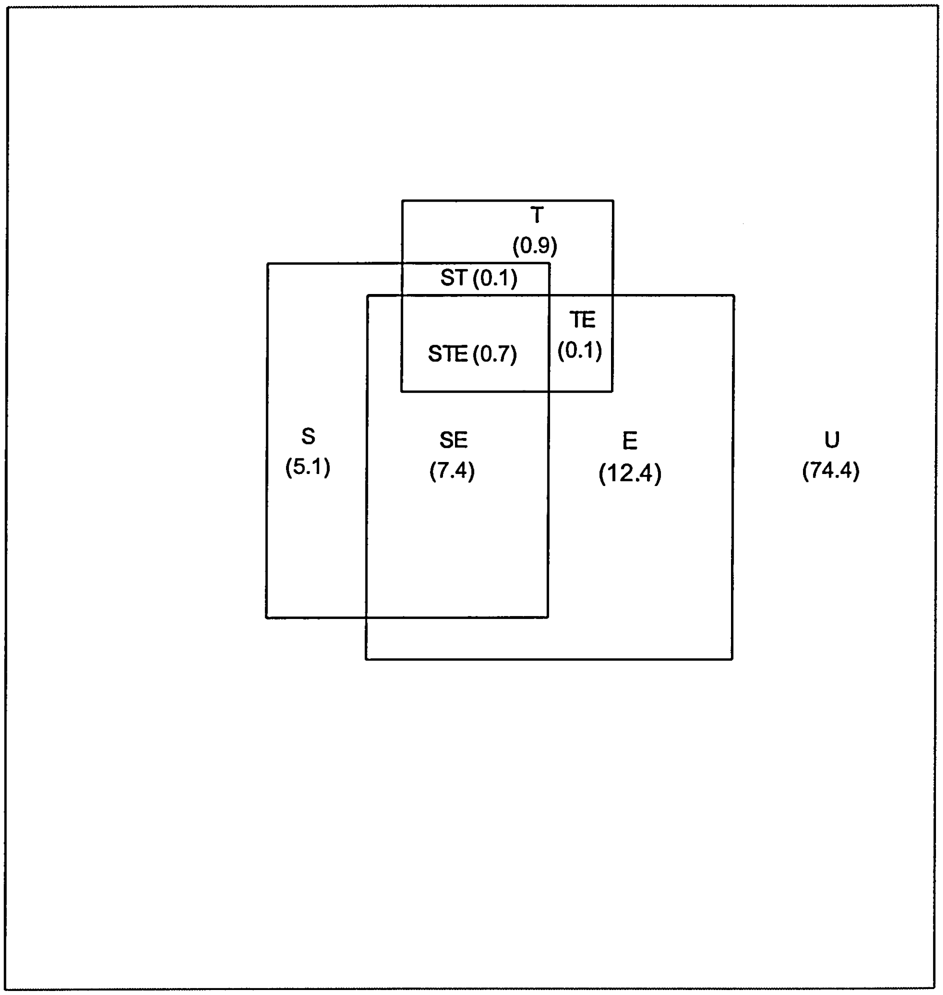


Figure 3.6. Approximate graphical representation of the partitioning of variation for the CCA analysis of Table 3.6 following the method of Anderson and Gribble (1998). The components in the diagram are spatial (S), environmental (E) and the combined effects for time and area (T). The area outside these components and their interactions is the unexplained variation (U).

3.5.4 Spatial GLMs

General linear models were fitted to the original partial CCA analysis i.e. of Tables 3.2 and 3.3, but with spatial structures specified. The values for –2 × the restricted maximum likelihood (REML) estimates were used to compare the fit of the different models. A value for the -2REML estimate which is closest to nought contains the least error and should give the best fit. The -2REML estimates are shown in Table 3.8 for GLMs with spatial, exponential, Gaussian structures and also when no structure is specified. The models with exponential and Gaussian structures appear to give the best fit.

Since these models all include a nugget effect (an intercept which is not necessarily nought) they contain three covariance parameters whereas the independent errors model has one. A chi-square test with two degrees of freedom can be used to give a test of significance for the improvement of fitting a model with a spatial structure specified. The results of these tests are shown in Table 3.9. Only the exponential structure can be seen to improve the fit of all the models significantly.

Table 3.8. -2 REML Log Likelihood estimates for each of the models fitted with different spatial structures specified, when relating the CCA axes to the environmental variables. The models with the smallest amount of error i.e. values closest to zero are shown in **bold type**.

Spatial structure	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Spherical	432.72	633.43	642.76	940.59
Exponential	435.63	**621.41**	600.26	**912.19**
Gaussian	**431.62**	622.55	**592.82**	935.63
No structure specified	460.90	633.80	642.76	940.59

Table 3.9. Chi-square tests between the -2 REML Log Likelihood estimates with spatial structures specified and when no spatial structure is specified.

Structure	DF	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Spherical	2	28.18 ***	0.37	0.00	0.00
Exponential	2	25.27 ***	12.39 **	42.50 ***	28.40 ***
Gaussian	2	29.28 ***	11.25 **	49.94 ***	4.96

The *t*-values for the estimates produced by the GLMs with exponential structures are shown in Table 3.10. These values can be compared with the *t*-values produced by CANOCO. There appears to be a problem with models fitted with spatial structures for axes 2 and 3 since the nugget value is larger than the value for the sill, suggesting that the semi-variograms for these models have been fitted upside down to that expected. The results shown in Table 3.10 for these two axes should therefore be treated with caution. However, there appears to be very little difference between the estimates produced from the two different methods of analysis. In particular, there also appears to be little difference between the *t*-values produced by the spatial GLMs and CANOCO for the first axis. The same variables that were found to be significant by CANOCO are still significant with the spatial GLM. Only the degree of significance has changed i.e. the variables are not as significant with the spatial GLM. This is what would normally be expected to happen since the spatial GLM tries to compensate for the possibility of other factors giving the observed result, so normally gives more conservative estimates than a model with no structure specified. The results produced for the fourth axis also appear to be reliable. The major difference is that the effects for the geology types, except Bunter Sandstone, are no longer significant. The effects for longitude, latitude and their interaction remain highly significant.

Differences between the *t*-values produced by the different spatial structures were also examined. The results produced were very consistent. The Gaussian models also produced nugget values that were higher than the sill values for axes 2 and 3.

Table 3.10. *t*-values for the fixed effects in the GLMs with exponential spatial structures.

Source	DF	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
		t	t	t	t
Effect					
Altitude	264	2.19 *	0.83	2.6 **	2.67 **
Longitude	264	-0.83	1.65	-4.22 ***	5.35 ***
Latitude	264	-0.73	1.59	-4.21 ***	5.34 ***
Longitude × Latitude	264	0.61	-1.40	4.29 ***	-5.07 ***
Recent	264	-1.53	9.52 ***	5.60 ***	-2.73 **
Ancient	264	0.46	-1.98 *	-0.77	-2.05 *
Secondary	.	.	.	.	.
Urban	264	-2.07 *	1.21	-0.41	2.81 **
Recent × Ancient Boundary	264	2.17 *	-3.29 **	-0.54	-0.54
Secondary × Ancient Boundary	264	-1.31	-1.66	-0.09	-0.31
Bunter Sandstone	264	1.97 *	0.86	-2.05 *	-2.04 *
Permian Marl	264	1.03	-0.11	-1.00	-0.50
Magnesian Limestone	264	-0.35	-0.41	-0.57	-0.72
Coal Measures	264	1.64	2.08 *	-1.96	-0.73
Millstone Grit	264	2.78 **	2.27 *	0.57	-0.70
Carboniferous Limestone	264	-2.05 *	3.25 **	2.14 *	0.53
Sycamore (DAFOR)	264	-3.02 **	-2.04 *	1.95	-4.92 ***
Beech (DAFOR)	264	-1.69	-2.17 *	-2.81 **	-0.71
Sweet Chestnut (DAFOR)	264	2.39 *	0.45	1.39	-1.84
Larch (DAFOR)	264	-0.91	0.66	0.96	-1.83
Pine/Spruce (DAFOR)	264	1.37	3.98 ***	-1.61	3.82 ***
Covariance parameter estimates					
Sill		26.92 **	21.25 **	35.95 **	2494.02
Range (km)		18.20 *	0.71	11.28 *	210.68
Nugget		20.30 ***	34.54 ***	37.50 ***	109.58 ***

3.5.5 Cluster analysis

The sample scores produced from the original partial CCA analysis, of Tables 3.2 and 3.3, that were linear combinations of environmental variables, were subjected to cluster analysis. The option available in CANOCO to preserve the chi-square distance between sample scores was used so that they could be treated as euclidean co-ordinates. The first eight axes were used in the cluster analysis. This number of axes appeared to give the least amount of ambiguity as to why a group had been formed in the resulting dendogram. The dendogram produced by Ward's clustering method is shown in Figure 8. Fisher's flexible method for clustering was also applied to the data. The major difference found was that the Fisher's flexible method first separated the recent sites on all the different geology types together with the ancient sites on the Millstone Grit from the rest. The second division then formed a separate group for the sites on the Millstone Grit. There would therefore appear to be some confusion as to where the sites situated on the Millstone Grit should be located on the dendogram, suggesting that they are very different to the other sites. Ward's method was preferred since it separated these sites out on the first partition.

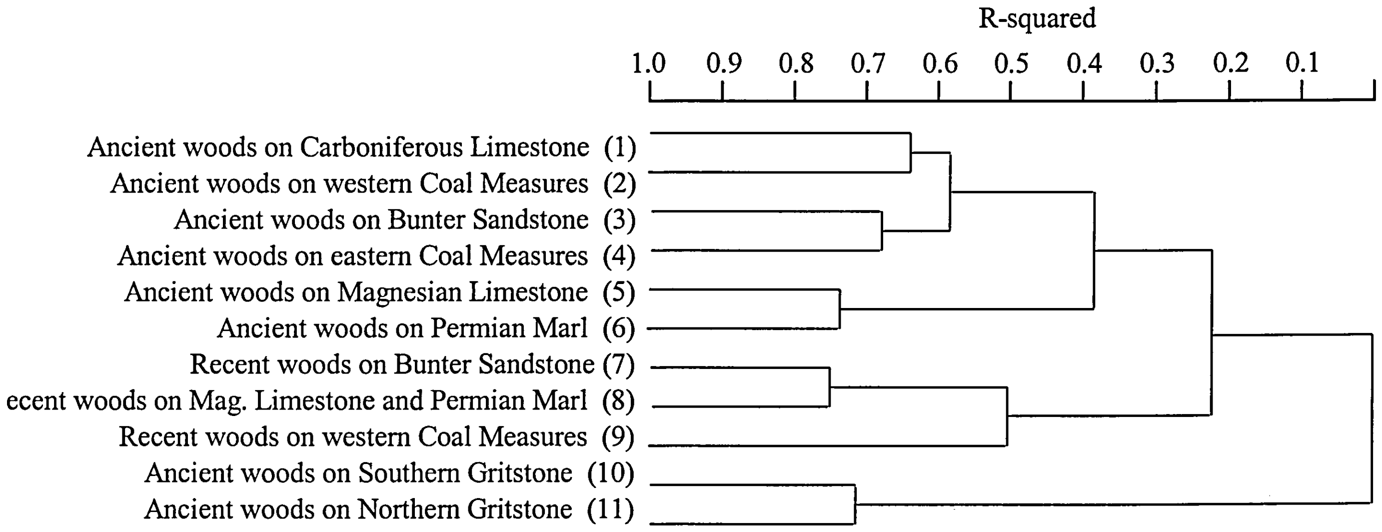


Figure 3.7. Dendogram produced using Wards clustering method on the first eight CCA sample scores that were linear combinations of the environmental variables. Ancient refers to both secondary and ancient woodland.

A confusion table, comparing the *a priori* habitat typology with the groups produced by the clustering procedure, is shown in Table 3.11. The main point of interest is that in each case for the different geology types where recent, secondary and ancient woods are present secondary sites are always grouped with the ancient sites. This suggests that the recent sites may be very different e.g. contain young trees which have not yet formed a closed canopy or contain a considerable amount of grassland or other open community with maybe hawthorn scrub forming the main canopy. The groups produced containing the ancient woodland may therefore be considered to contain mature woodland rather than just ancient woodland. The cluster analysis can be also be seen to appear to have some problems in separating recent sites on the Bunter Sandstone and eastern Coal Measures and with recent sites on the Permian Marl and Magnesian Limestone. The soil types can differ considerably within these geology types, especially on the Magnesian Limestone and Permian Marl. This may be a complicating factor.

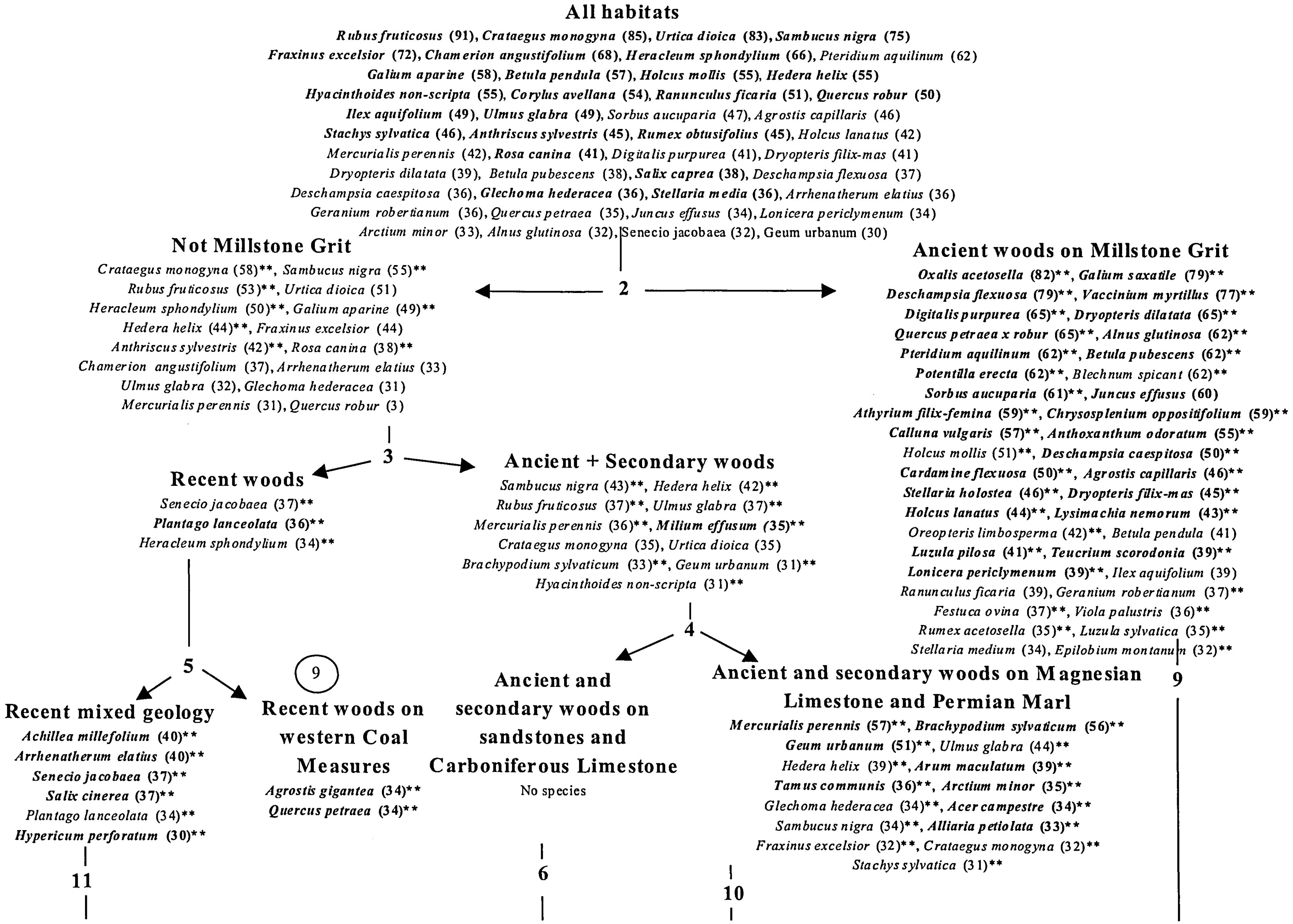
Table 3.11. Confusion table comparing the a *priori* habitat typology to the typology produced from Ward's clustering method (Figure 3.7).

	Habitat (Cluster number)											
	1	2	3	4	5	6	7	8	9	10	11	S
Ancient woods on Carboniferous limestone	**6**											
Ancient and secondary woods on western Coal Measures		**45**	4	2								5
Ancient and secondary woods on Bunter Sandstone			**47**	4								5
Ancient and secondary woods on eastern Coal Measures			1	**7**								
Ancient and secondary woods on Magnesian Limestone				1	**37**	1						3
Ancient and secondary woods on Permian Marl						**16**						1
Recent woods on Bunter Sandstone and eastern Coal Measures							**7**	10				1
Recent woods on Permian Marl and Magnesian Limestone					10		3	**8**				2
Recent woods on western Coal Measures		2	5	1				2	**28**			3
Ancient woods on southwestern Millstone Grit										**31**		3
Ancient woods on northern Millstone Grit											**7**	
Sum	6	47	57	15	47	17	10	20	28	31	7	2

3.5.6 Indicator analysis

The hierarchical structure shown in Figure 3.7 was entered into the program INDVAL, together with the species abundance data and species names. This dendogram is redrawn in Figure 3.9 with the main species, that are associated with each partition, in the dendogram shown. An arbitrary threshold value of 30% was chosen for showing species in each cluster. Species that are significantly ($P < 0.01$) associated with a particular cluster are indicated on the diagram. Legendre and Dufrene displayed a similar diagram but only showed the location of significant species and also only showed the location of species until they had reached their maximum. The diagram of Figure 3.9 includes all species above the chosen threshold value, and may therefore be more useful in terms of describing the community at each partition in the dendogram instead of being concerned solely with indicator species. High indicator values can be seen for some of the species of the ancient woods on the Millstone Grit. Species, which appear to be associated with this group, include *Oxalis acetosella*, *Galium saxatile*, *Deschampsia flexuosa and Vaccinium myrtillus*, which all have their maximum indicator value in this group. The species displayed in the opposing group have lower indicator values revealing the greater heterogeneity of sites in this group i.e. contain a number of different geology types as well as recent and mature woodlands. Further down the table species that are particularly associated with the recent sites are *Senecio jacobaea*, *Plantago lanceolata* and *Heracleum sphondylium*. These species all require open environments with little canopy cover. Species that are associated with mature woodland include *Sambucus nigra*, *Hedera helix*, *Rubus fruticosus*, *Mercurialis perennis* and *Milium effusum*. All these species are normally found in well-shaded habitats. Species that are associated with the mature woods on the Magnesian Limestone and Permian Marl include *Mercurialis perennis*, *Brachypodium sylvaticum*, *Geum urbanum*, *Ulmus glabra*, *Hedera helix* and *Arum maculatum*. The other main group of interest shown on this diagram is that for the ancient woods on the Carboniferous Limestone. Species associated with this group include *Mycelis muralis*, *Hypericum hirsutum*, *Filipendula ulmaria*, *Vicia sepium*, *Potentilla sterilis* and *Rubus uva-crispa*. Many species were found to be significantly associated with different groups below the threshold value that was chosen. Species that are rare will have low indicator values, despite often being some of the species that are often considered to be the best indicators. This can be seen from Table 3.12. For example, the group labelled 'Ancient and secondary woods on Magnesian Limestone' contains five rare species that were only found in this group. These species *are Daphne laureola*, *Neottia nidus-avis*, *Iris foetidissima*, *Ophrys insectifera* and *Platanthera chlorantha*. All of these species are listed by Rose (1999), as species he considers to be ancient woodland indicators in the south of England.

er maximum m ka or va e are m o



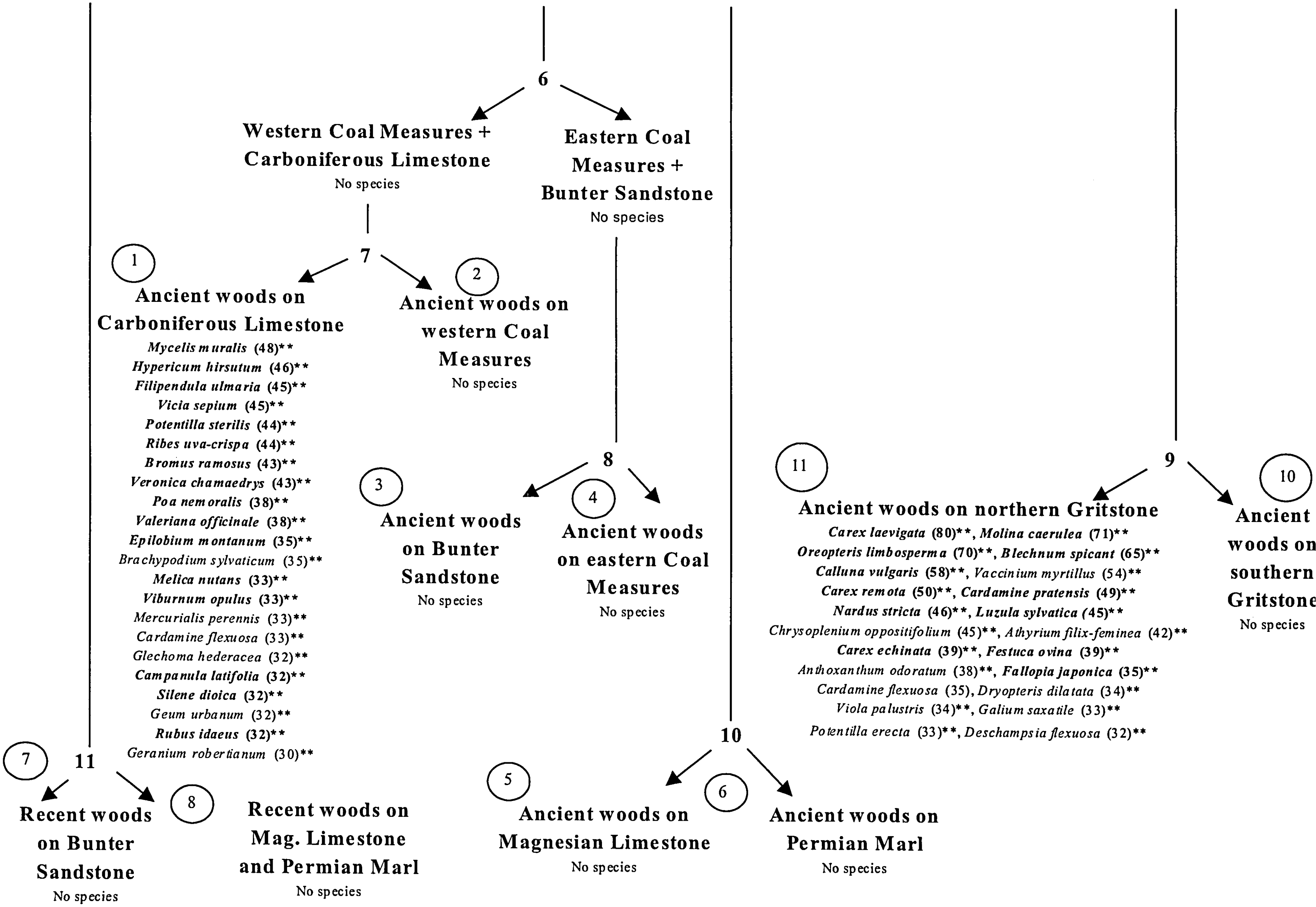

6
Western Coal Measures + Carboniferous Limestone
No species
Eastern Coal Measures + Bunter Sandstone
No species
7
1
Ancient woods on Carboniferous Limestone
Mycelis muralis (48)**
Hypericum hirsutum (46)**
Filipendula ulmaria (45)**
Vicia sepium (45)**
Potentilla sterilis (44)**
Ribes uva-crispa (44)**
Bromus ramosus (43)**
Veronica chamaedrys (43)**
Poa nemoralis (38)**
Valeriana officinale (38)**
Epilobium montanum (35)**
Brachypodium sylvaticum (35)**
Melica nutans (33)**
Viburnum opulus (33)**
Mercurialis perennis (33)**
Cardamine flexuosa (33)**
Glechoma hederacea (32)**
Campanula latifolia (32)**
Silene dioica (32)**
Geum urbanum (32)**
Rubus idaeus (32)**
Geranium robertianum (30)**
2
Ancient woods on western Coal Measures
No species
8
3
Ancient woods on Bunter Sandstone
No species
4
Ancient woods on eastern Coal Measures
No species
9
11
Ancient woods on northern Gritstone
Carex laevigata (80)**, Molina caerulea (71)**
Oreopteris limbosperma (70)**, Blechnum spicant (65)**
Calluna vulgaris (58)**, Vaccinium myrtillus (54)**
Carex remota (50)**, Cardamine pratensis (49)**
Nardus stricta (46)**, Luzula sylvatica (45)**
Chrysoplenium oppositifolium (45)**, Athyrium filix-feminea (42)**
Carex echinata (39)**, Festuca ovina (39)**
Anthoxanthum odoratum (38)**, Fallopia japonica (35)**
Cardamine flexuosa (35), Dryopteris dilatata (34)**
Viola palustris (34)**, Galium saxatile (33)**
Potentilla erecta (33)**, Deschampsia flexuosa (32)**
10
Ancient woods on southern Gritstone
No species
11
7
Recent woods on Bunter Sandstone
No species
8
Recent woods on Mag. Limestone and Permian Marl
No species
10
5
Ancient woods on Magnesian Limestone
No species
6
Ancient woods on Permian Marl
No species

Table 3.12. Two-way indicator table showing the species indicator power for the site clustering hierarchy. The first value given before the slash refers to the sum of the abundance values and the value after the slash represents the number of sites in that group that the species is present. Bold numbers represent the main data set structure (shown in Figure 3.10).

Species	IndVal (%)	1	2	3	4	5	6	7	8	9	10	11
Ancient woods on Carboniferous Limestone (1)												
Mycelis muralis	48.4 **	**5./4**	1./1			9./4	3./1				4./2	
Hypericum hirsutum	46.0 **	**12./4**			9./3	41./12		3./1		1./1		
Filipendula ulmaria	44.8 **	**19./6**	26./13	18./6	13./3	28./8	9./3	12./2	16./5	12./3	22./8	15./5
Vicia sepium	44.6 **	**13./5**	12./7		9./3	2./2	1./1	4./2		26./10	11./7	4./2
Potentilla sterilis	44.0 **	**6./4**	2./2		3./1	7./2		1./1			5./3	6./2
Ribes uva-crispa	43.8 **	**6./5**	3./3	3./1	2./2	35./12				4./1	1./1	2./2
Bromus ramosa	43.0 **	**20./5**	41./13	12./4	12./4	78./21	26./8	3./1	1./1		3./2	3./1
Veronica chamaedrys	42.7 **	**12./5**	5./3	3./1	17./4	27./9			18./6	3./1	12./6	3./1
Poa nemoralis	38.0 **	**7./3**	11./5	3./1		6./2						
Valeriana officinalis	37.6 **	**9./4**	15./8		3./1	4./2		4./2			14./9	9./5
Epilobium montanum	35.0 **	**13./6**	67./25	1./1	6./2	10./4		3./1		38./12	23./15	11./5
Melica nutans	33.3 **	**6./2**										
Viburnum opulus	32.7 **	**8./4**	13./10	17./8	4./2	38./14	4./2	3./1		1./1		1./1
Campanula trachelium	32.3 **	**6./2**							1./1			
Silene dioica	32.3 **	**17./5**	56./17	23./7	46./11	63./16	9./3	3./1	8./3	6./2	14./4	10./4
Rubus idaeus	31.7 **	**16./5**	35./14	4./2	25./6	73./19	8./3	3./1	4./2	21./6	27./11	9./3
Adoxa moschatellina	29.6 **	**7./2**	2./2								3./2	1./1
Fragaria vesca	29.0 **	**9./3**	1./1	7./3	8./2	12./4		7./1	6./1		3./1	6./1
Ribes rubrum	28.8 **	**3./3**	5./5		3./3	13./5	1./1					
Poa trivialis	28.7 **	**14./5**	37./11		18./5	43./13		8./2		38./9	28./11	14./4
Succisa pratensis	28.3 **	**6./3**	4./2			9./2		3./1	6./1	6./2	1./1	
Angelica sylvestris	27.8 **	**12./5**	49./20	14./5	14./5	12./5	14./4	15./3	4./2	22./8	17./8	11./5
Galium verum	27.1 **	**5./3**			4./1	3./1		7./2	11./4			
Geum rivale	25.9 **	**6./2**	5./3			3./1						5./3
Festuca rubra agg.	25.7 **	**15./5**	12./6	19./4	26./5	10./3		31./6	30./7	42./10	23./7	16./4
Alopecurus pratensis	25.1	**10./3**	7./7	3./1	31./8			3./1	5./1	14./5		10./3
Scrophularia auriculata	24.9 **	**6./4**	12./8	9./4	8./3	37./14	4./2			8./2	8./6	2./2
Polypodium vulgare agg.	22.8 **	**3./2**								5./1	1./1	1./1
Hypericum perforatum	22.6 **	**4./2**	5./3		5./3	1./1		3./1			1./1	
Myosotis sylvatica	22.3 **	**6./2**	7./5			7./3					6./4	3./1
Primula veris	22.1 **	**5./3**			11./4	15./6		9./3	11./3			
Prunus padus	21.6 **	**2./2**	1./1			3./1		1./1			2./2	1./1
Primula vulgaris	21.5	**3./2**			1./1	15./5					1./1	
Melica uniflora	19.6	**9./3**	45./19	15./4	4./1	50./12	7./2				4./3	4./2
Carex digitata	16.7	**1./1**										
Dryopteris aemula	16.7	**1./1**										
Ribes alpinum	16.7	**1./1**										
Sedum telephium	16.7	**4./1**										
Polystichum aculeatum	15.8	**3./1**									1./1	
Crataegus laevigata	15.7	**2./1**	1./1									
Sanicula europaea	15.6	**6./2**	4./2	4./1	7./3	42./12		1./1			6./3	2./1
Ribes nigrum	15.4 **	**1./1**		1./1								
Rubus saxatilis	15.1	**3./1**									1./1	1./1
Carex sylvatica	14.9	**7./2**	9./6	4./1	10./4	34./10	8./2			1./1	7./4	7./2
Torilis japonica	14.8	**7./2**	2./2	4./1	6./2	30./8	2./1	3./1	13./4	7./1		
Convallaria majalis	13.8	**6./1**	3./1			3./1		3./1				
Malus sylvestris	13.2	**4./2**	4./4	16./8	8./4	5./4	3./1	3./1	4./2	3./1	4./4	1./1
Epipactis helleborine	13.0	**3./1**	6./6	1./1								
Equisetum telmateia	12.8	**4./1**	1./1			5./2						4./1
Phalaris arundinacea	9.1	**3./2**	2./2	15./5	1./1	3./1	11./3	7./2	12./3		8./1	
Orchis mascula	6.6	**1./1**	1./1			15./3						
Ancient and seconary woods on Coal Measures and Carboniferous limestone west (1 + 2)												
Lamiastrum galeobdolon	23.5 **	**3./2**	**99./25**		13./3	58./14	3./1 1	4./1	33./9			
Sorbus aria agg.	13.5 **	**1./1**	**23./13**	4./2	5./3	4./2	5./2	2./2	1./1			
Ancient and secondary woods on western Coal Measures (2)												
Anemone nemorosa	17.1	1./1	**62./20**	2./1	20./5	51./13				6./2	6./2	12./3
Carex pendula	2.1		**1./1**									
Ancient and secondary woods on Coal Measures, Bunter Sandstone and Carboniferous limestone (1 + 2 + 3 + 4)												
No indicator species												
Ancient and secondary woods on Bunter Sandstone (3)												
No indicator species												
Woods on Bunter Sandstone and eastern Coal Measures (3 + 4)												
Ceratocapnos claviculata	17.1 **	1./1	3./1	**67./13**	**16./4**						8./2	6./2

Species	IndVal (%)	1	2	3	4	5	6	7	8	9	10	11
Ancient and secondary woods on Eastern Coal Measures (4)												
Lamium album	25.3 **	2./ 1	3./ 1	28./10	**30./ 9**	24./ 7	11./ 4	10./ 3	24./ 8	5./ 3		
Moehringia trinervia	22.4 **	2./ 2	6./ 4	14./ 4	**26./ 7**	48./14	11./ 3		1./ 1		8./ 6	
Populus tremula	16.2			16./ 4	**16./ 4**	20./ 7			3./ 1			
Luzula campestris	15.1		7./ 4		**11./ 5**	4./ 1		3./ 1		5./ 1	8./ 4	7./ 2
Circaea lutetiana	14.5	2./ 1	26./ 7	15./ 3	**22./ 6**	44./11	15./ 4			2./ 1	14./ 7	3./ 1
Galium cruciata	10.5		1./ 1		**11./ 3**	22./ 6			9./ 2			
Hypericum tetrapterum	9.2			5./ 1	**5./ 3**	8./ 2	3./ 1		4./ 1			
Sorbus torminalis	7.8				**2./ 2**	6./ 2						
Viola reichenbacjiana	6.9				**4./ 2**	3./ 1			6./ 1			
Dryopteris carthusiana	6.7				**1./ 1**							
Polygonatum multiflorum	6.7				**3./ 1**							
Ranunculus auricomus	6.7				**1./ 1**							
Primula veris x *Primula vulgaris*	6.3				**4./ 1**	1./ 1						
Calamagrostis canescens	5.4				**4./ 1**	4./ 1						
Pimpinella major	3.9				**5./ 1**	7./ 2		3./ 1			1./ 1	
Narcissus pseudonarcissus	2.5		1./ 1		**3./ 1**	7./ 2				2./ 2	5./ 1	
Aquilegia vulgaris	2.4				**1./ 1**	4./ 2			1./ 1		1./ 1	
All ancient and secondary woods (1 + 2 + 3 + 4 + 5 + 6)												
Milium effusum	35.0 **	**3./ 1**	**118./31**	**52./16**	**29./10**	**86./18**	**6./ 2**		1./ 1		7./ 3	3./ 1
Tilia cordata	18.0 **	**1./ 1**	**14./ 6**	**29./10**	**12./ 5**	**39./11**	**12./ 2**			1./ 1		
Carpinus betulus	5.3		**2./ 2**	**9./ 5**		**9./ 3**						
Tilia cordata x *platyphyllos (T.* x *vulgaris)*	5.3		**7./ 3**	**11./ 3**	**2./ 1**	**3./ 3**						
Ancient and secondary woods on Magnesian Limestone (5)												
Daphne laureola	6.4					**8./ 3**						
Listera ovata	4.4				1./ 1	**7./ 3**						
Neottia nidus-avis	4.3					**2./ 2**						
Tilia platyphyllos	2.6		1./ 1			**4./ 2**					1./ 1	
Iris foetidissima	2.1					**2./ 1**						
Ophrys insectifera	2.1					**1./ 1**						
Platanthera chlorantha	2.1					**1./ 1**						
Ancient and secondary woods on Magnesian Limestone and Permain Marl (5 + 6)												
Mercurialis perennis	57.2 **	29./ 6	91./24	15./ 5	48./11	**285./45**	**60./11**	3./ 1	10./ 2	11./ 3	24./ 8	13./ 3
Brachypodium sylvaticum	55.7 **	19./ 5	18./ 8	16./ 5	30./ 8	**165./38**	**33./ 8**	3./ 1			3./ 1	4./ 2
Geum urbanum	51.1 **	19./ 5	33./11	17./ 4	37./ 9	**156./37**	**35./10**	5./ 2	3./ 1		6./ 5	6./ 2
Arum maculatum	38.5 **	13./ 4	19./10	11./ 5	17./ 7	**103./33**	**19./ 5**	3./ 1	8./ 2	6./ 2	6./ 3	4./ 2
Tamus communis	36.3 **	2./ 2	3./ 3	10./ 4	10./ 4	**65./28**	**13./ 5**	11./ 3	7./ 3			
Arctium minus	34.8 **	9./ 4	14./ 6	43./17	31./ 9	**119./33**	**16./ 6**	11./ 5	21./ 7	3./ 1	6./ 4	2./ 2
Acer campestre	34.2 **	5./ 2	10./10	37./11	19./ 9	**88./27**	**23./ 7**	4./ 2	10./ 4	1./ 1	5./ 1	
Alliaria petiolata	32.8 **	4./ 2	6./ 4	27./ 8	23./ 7	**106./25**	**18./ 6**	1./ 1	24./ 6	1./ 1		
Ligustrum vulgare	29.6 **	5./ 2	3./ 3	15./ 5	10./ 4	**61./19**	**16./ 6**	1./ 1	6./ 2			
Rumex sanguineus	26.3 **	1./ 1	41./16	17./ 4	38./ 9	**89./26**	**13./ 4**	3./ 1	9./ 3		1./ 1	9./ 3
Lapsana communis	25.8 **	2./ 2	21./11	6./ 4	13./ 5	**54./22**	**19./ 7**	6./ 2	9./ 4	5./ 3	4./ 2	3./ 1
Cornus sanguinea	25.6 **	6./ 3		6./ 2	8./ 2	**63./19**	**3./ 2**	1./ 1	5./ 1		1./ 1	
Euonymus europaeus	21.4 **	1./ 1	1./ 1		4./ 2	**36./14**	**2./ 2**		3./ 1			
Viola odorata	18.5 **			7./ 2	7./ 2	**57./15**	**1./ 1**	7./ 2	4./ 1	1./ 1		
Ulmus procera	15.7 **		1./ 1	19./ 5	12./ 3	**46./12**	**4./ 2**		3./ 1			
Campanula latifolia	14.8 **	1./ 1			5./ 3	**38./10**	**1./ 1**				1./ 1	1./ 1
Symphoricarpos albus	13.1 **		8./ 3	15./ 4	20./ 4	**47./10**	**13./ 3**		3./ 1	1./ 1		4./ 2
Rubus caesius	8.1	1./ 1		3./ 1		**18./ 5**	**9./ 3**	4./ 1	3./ 1		3./ 1	
Potentilla anserina	6.1			7./ 2	6./ 2	**21./ 6**	**3./ 1**	3./ 1	8./ 2			
Agropyron caninum	5.3				6./ 1	**17./ 4**						
Ancient and secondary woods on Permian Marl (6)												
Clematis vitalba	4.9				1./ 1	9./ 2	**8./ 2**	6./ 1	6./ 1			
Carex acutiformis	4.2			4./ 1			**3./ 1**					
All woods not on Millstone Grit (1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9)												
Prunus spinosa	28.6	**6./ 2**	**23./10**	**55./15**	**39./11**	**59./15**	**9./ 4**	**8./ 2**	**7./ 3**	**53./14**	1./ 1	2./ 2
Solanum dulcamara	18.5		**34./14**	**17./ 8**	**15./ 5**	**17./ 4**	**7./ 3**	**6./ 2**	**20./ 9**	**9./ 3**		1./ 1
Salix fragilis	17.9	**1./ 1**	**37./14**	**27./10**	**12./ 3**	**29./ 7**	**17./ 5**	**13./ 3**	**16./ 5**			2./ 2
Bryonia dioica	11.7			**10./ 6**	**6./ 4**	**27./11**	**3./ 1**	**5./ 3**	**8./ 4**			
Recent woods on Bunter Sandstone and eastern Coal Measures (7)												
Vicia cracca	21.8 **	2./ 1	8./ 4	20./ 6	7./ 2	19./ 7	6./ 1	**19./ 5**	5./ 3		4./ 1	
Hypericum montanum	16.1 **		9./ 3					**8./ 2**				
Origanum vulgare	13.1	6./ 1			6./ 2	7./ 2		**12./ 3**				
Bromus erectus	10.5		8./ 2		8./ 1	15./ 4	4./ 1	**16./ 2**	4./ 1			
Dactylorhiza fuchsii	9.7	3./ 1	3./ 1		4./ 2	6./ 2		**9./ 2**				
Rosa arvensis	8.9		6./ 4	3./ 1	3./ 1	5./ 1	2./ 1	**6./ 2**				1./ 1
Epipactis phyllanthes	6.0				1./ 1			**1./ 1**				
Lathyrus montana	5.7		1./ 1					**3./ 1**		3./ 1	3./ 2	
Calamagrostis epigejos	5.5					12./ 3		**4./ 1**		2./ 1		
Stachys officinalis	4.8		1./ 1			6./ 2		**3./ 1**		4./ 1	1./ 1	

ecies	IndVal (%)	1	2	3	4	5	6	7	8	9	10	11
cent woods on Permian Narl and mixed geology types (7 + 8)												
hillea millefolium	39.9 **	5./3	7./4	21./6	12./3	3./1	7./2	21./5	50./13	8./5	5./2	10./3
henatherum elatius	39.7 **	21./5	24./9	103./23	28./7	70./21	16./5	58./10	54./13	12./4	1./1	10./4
necio jacobaea	38.6 **	7./4	23./13	40./14	27./9	16./7	3./2	30./8	53./14	28./11	4./4	12./4
lix cinerea	36.5 **	5./3	15./7	73./18	17./5	28./9	7./3	29./6	62./14	11./3	12./5	14./5
ypericum pulchrum	30.1 **	4./2	11./3	27./9	6./2	27./10	13./4	19./4	46./10	5./1		
entaurea nigra	29.7 **	7./4	15./7	14./3	19./6	34./10	8./2	35./7	39./8	15./6	2./2	
temisia vulgaris	28.1 **		7./5	41./16	11./4	22./10	24./7	28./7	23./8	3./1		1./1
ontodon autumnalis	25.5 **			28./8	5./1	9./3	4./1	15./3	37./9	10./2	3./3	3./1
thyrus pratensis	23.3 **	7./3	7./5	10./3	6./2	3./1		15./4	21./6	5./2	1./1	1./1
ymbrium officinale	22.3 **		1./1	15./6	7./2	17./5	3./1	4./2	26./8			
pha latifolia	20.6 **			28./7	6./3	7./2	3./1	16./3	28./6	6./1		1./1
onium maculatum	20.6 **			23./7	7./2	4./2	11./4	8./2	30./7	3./1		
onvolvulus arvensis	20.3 **			6./2	3./1			3./1	19./6	1./1		
tus corniculatus	20.1 **	6./2	8./4	19./5	16./4	4./2	6./1	19./4	28./7	14./3	2./2	7./3
tentilla reptans	20.1 **	1./1	9./5	23./6	16./4	26./6	6./1	14./3	26./8	2./1	5./2	
uisetum arvense	17.7 **	3./2	3./3	16./5	6./2	13./5	10./3	12./3	18./7	1./1	3./2	7./3
ipsacus fullonum	16.0 **			4./1		7./3		4./2	13./4			
lix viminalis	14.8 **	1./1		18./5		10./2	6./2	6./2	15./5			
rophularia nodosa	11.6 **			3./1	3./1	9./3	3./1	7./2	9./3			
s pseudacorus	8.5		3./3	11./4	5./2	11./4	1./1	4./2	9./3			3./1
ecent woods on Magnesian Limestone and Permian Marl (8)												
pericum humifusum	10.0								6./2			
rdelymus europaeus	6.0		2./2	3./1		5./2			6./2			
ecent woodland (7 + 8 + 9)												
antago lanceolata	35.9 **	5./3	21./9	39./12	32./9	40./11	8./2	32./7	58./14	35./11	3./3	12./4
pochaeris radicata	19.6 **		22./9	13./5	12./4			21./5	25./6	23./6	1./1	12./4
cent woods on western Coal Measures (9)												
rostis gigantea	34.1		6./2	3./1						42./10		
ercus petraea	33.8		89./16	49./13	13./2	16./4	4./2		13./3	67./18	8./2	
uisetum sylvaticum	10.8	4./1	21./9		3./1					22./5	6./3	5./3
l habitats												
bus fruticosus agg.	91.2	23./6	239./44	258./52	83./15	251./45	79./15	53./10	79./18	95./26	101./22	34./7
ataegus monogyna	84.9	17./5	142./43	175./45	71./15	232./45	65./15	45./9	87./17	95./21	49./20	21./7
ica dioica	83.2	22./6	188./43	171./46	68./14	185./39	66./15	36./7	68./17	79./20	65./23	24./7
mbucus nigra	74.7	16./6	109./37	166./47	65./15	187./43	53./13	16./4	48./15	45./15	33./12	15./6
axinus excelsior	71.6	21./6	107./34	95./33	52./14	194./41	63./14	27./6	60./15	72./16	55./20	14./5
amerion angustifolium	68.1	17./4	161./40	164./40	47./11	91./24	14./4	35./9	54./12	75./20	61./24	20./6
racleum sphondylium	65.6	14./5	150./42	75./25	46./13	134./35	16./7	36./9	33./11	92./22	24./13	9./5
lium aparine	57.9	18./4	131./38	76./20	70./14	170./33	28./7	18./5	39./10	79./20	18./11	7./3
tula pendula	56.8	13./5	116./34	168./40	41./8	44./13	13./3	16./4	54./11	54./16	68./21	34./7
lcus mollis	55.1	15./3	259./43	219./40	61./10	37./8	10./2	9./3	22./4	70./13	136./24	45./7
dera helix	54.7	20./5	131./32	144./29	40./10	224./38	77./12	2./2	31./6	32./9	23./10	10./3
acinthoides non-scripta	54.7	21./5	182./38	129./29	62./12	143./33	7./2	1./1	9./3	35./12	65./15	21./6
rylus avellana	53.7	19./6	112./36	72./23	38./11	139./34	17./5	7./3	11./3	29./8	44./18	20./6
nunculus ficaria	50.5	14./5	136./36	43./12	40./10	71./18	14./3	14./3	32./9	73./21	58./20	22./7
ercus robur	50.2	10./4	58./15	150./34	49./10	113./32	32./9	18./6	41./11	23./7	26./8	29./7
x aquifolium	48.8	3./3	128./38	55./26	24./8	44./18	16./6	3./1	6./2	22./11	49./22	12./4
mus glabra	48.8	14./4	83./27	60./19	33./11	150./40	58./13	4./2	10./2	10./4	28./14	7./3
chys sylvatica	45.6	13./5	80./25	26./11	45./12	137./35	30./9	9./4	2./2	15./5	40./18	12./4
thriscus sylvestris	44.9	10./5	81./27	39./13	49./12	132./31	26./7	15./4	19./6	71./17	5./3	7./3
mex obtusifolius	44.6	5./2	102./32	58./18	31./10	54./18	10./4	12./3	25./9	58./16	19./10	17./5
sa canina agg.	41.4	10./6	13./5	79./23	35./10	107./31	25./9	23./5	60./15	16./7	8./6	3./1
lix caprea	37.5	9./5	45./19	52./15	26./9	44./15	5./3	8./2	34./10	32./10	25./13	14./6
echoma hederacea	36.1	21./6	10./5	75./18	55./11	158./33	31./9	9./2	46./10	3./1	8./4	8./4
llaria media	36.1	5./3	39./17	74./20	45./12	43./12	8./2	3./3	44./9	5./2	44./19	10./4
ntago major	29.5	2./2	39./17	35./10	25./8	58./19	16./5	17./5	34./10	6./2	2./2	10./4
nunculus repens	28.1	16./5	94./24	6./1	30./8	85./17				38./13	21./7	19./5
silago farfara	26.7	6./3	18./12	27./8	17./7	22./6	9./4	27./8	22./8	13./6	12./9	11./5
la riviniana	26.0	5./3	38./18	26./7	29./7	86./21	9./2	3./1	4./2	11./4	9./6	10./3
tuca gigantea	25.3	6./4	81./25	5./1	18./5	71./21	9./3	1./1	1./1	18./6	5./3	4./2
ilobium hirsutum	24.9	14./4	44./19	27./9	25./7	40./12	12./4	10./2	25./6	22./6		6./2
sa arvensis x *canina*	22.1	9./4	69./24	18./6	16./4	48./12	7./2	3./1	14./4	7./3	5./3	
nus avium	20.4	1./1	35./20	6./4	10./4	59./17	6./4	8./2	1./1	3./1	2./1	7./3
us baccata	17.9	5./3	6./5	35./14	14./6	45./16	9./2		3./1		7./4	
ium ursinum	16.1	9./3	39./12		20./5	69./13				11./3	18./7	18./3
x europaeus	15.4		11./6	36./12	9./5	4./2		8./1	15./6	19./5	9./5	2./2
lystegia sepium	11.6	3./2	19./10	18./7	7./3	4./2	5./3		6./1	13./3	1./1	3./1
atiens glandulifera	7.4	1./1	18./6	7./2	1./1	14./4	13./4	1./1	4./1			6./1
podium podagraria	7.0		5./2	12./3	1./1	27./7		3./1	4./1	7./4		1./1
tha aquatica	4.2		1./1	5./2	1./1	11./3	3./1	5./1	3./1		5./2	
cient woods on southwestern Millstone GritGritstone (10)												
llaria neglecta	3.2										5./1	

Species	IndVal (%)	1	2	3	4	5	6	7	8	9	10	11
Ancient woods on Millstone Grit (10 + 11)												
Oxalis acetosella	81.9 **	12./4	87./26	10./3	18./6	14./4	4./1			7./3	**121./29**	**28./7**
Galium saxatile	79.4 **	5./3	32./13	26./6	11./4			4./1		28./7	**98./28**	**30./6**
Deschampsia flexuosa	79.1 **	9./2	176./39	65./15	27./5	7./2			9./2	15./3	**158./30**	**47./7**
Vaccinium myrtillus	76.9 **		20./7								**81./24**	**36./6**
Digitalis purpurea	70.6 **	4./2	71./24	121./31	30./9	16./6	4./2		10./3	8./4	**90./29**	**29./7**
Dryopteris dilatata	65.4 **	15./5	66./21	58./18	32./8	40./14	16./4	4./2	3./1	20./6	**93./25**	**43./7**
Quercus petraea x *robur*	64.6 **	14./4	129./24	76./20	34./6	31./8	1./1		16./4		**125./26**	**27./6**
Alnus glutinosa	62.5 **	6./2	55./25	29./11	11./4	22./8	7./3	8./2	1./1	19./4	**70./24**	**22./6**
Pteridium aquilinum	62.3 **	15./4	170./36	225./39	78./14	108./25	21./6	4./1	38./6	45./9	**163./29**	**45./7**
Betula pubescens	62.1 **	11./3	57./15	128./29	37./7	17./6	21./6	9./3	21./6	4./2	**95./25**	**37./7**
Potentilla erecta	61.8 **		7./5	6./2	7./2	7./2		3./1	12./4	12./3	**57./21**	**21./5**
Sorbus aucuparia	61.4 **	9./5	70./25	88./34	37./10	32./15	4./2	1./1	2./2	20./8	**79./25**	**30./7**
Juncus effusus	59.8 **	4./1	63./24	57./18	10./5	8./3	10./3	15./3	29./8	6./3	**72./26**	**25./4**
Athyrium filix-femina	59.2 **	5./2	49./19	7./2	10./3	1./1		1./1			**50./19**	**22./7**
Chrysosplenium oppositifolium	58.8 **	8./3	43./15		3./1	1./1				5./2	**53./18**	**25./7**
Anthoxanthum odoratum	55.3 **	11./4	8./6	8./2	23./6	4./1		3./1		11./4	**49./17**	**25./7**
Deschampsia caespitosa	50.4 **	19./5	96./27	31./9	17./4	37./10	8./2	12./3	15./5	32./10	**58./21**	**31./7**
Cardamine flexuosa	49.8 **	13./5	68./24	1./1	7./2	11./4				3./1	**38./17**	**21./7**
Agrostis capillaris	46.3 **	14./5	101./24	120./26	24./6	25./7	14./3	28./6	71./13	58./12	**75./22**	**33./7**
Stellaria holostea	46.0 **	12./5	75./22	29./9	20./7	33./8	3./1			5./2	**72./18**	**14./5**
Dryopteris filix-mas	45.3 **	13./5	81./24	61./19	28./8	66./21	19./5		6./2	14./6	**63./20**	**24./6**
Holcus lanatus	44.1 **	9./3	96./21	59./18	27./7	26./9	3./1	30./7	56./12	70./14	**54./21**	**32./7**
Lysimachia nemorum	43.1 **	6./3	16./8		12./4	16./5					**32./14**	**16./5**
Luzula pilosa	41.1 **	3./2	18./10								**31./14**	**5./3**
Teucrium scorodonia	39.2 **	9./4	43./16	79./19	27./9	2./1		7./2	21./5	7./3	**48./16**	**15./6**
Lonicera periclymenum	39.0 **	6./4	76./24	66./20	18./5	38./13	11./3	9./2	1./1	5./2	**58./18**	**15./4**
Geranium robertianum	37.1 **	17./6	73./22	22./6	13./4	112./28	28./6	4./2	7./2	7./2	**44./18**	**18./6**
Viola palustris	36.4 **				4./1						**32./10**	**13./4**
Rumex acetosella	35.4 **	4./2	13./7	30./7	11./3			3./1	10./3	10./2	**34./13**	**13./4**
Rumex acetosa	32.2 **	4./3	12./8	19./6	8./4	8./3		15./4	16./5	34./12	**28./12**	**18./5**
Dryopteris affinis	30.5 **	1./1	1./1		4./2						**25./12**	
Ajuga reptans	26.2 **	4./3	1./1		13./3	18./6		1./1	1./1	4./1	**15./9**	**6./4**
Veronica montana	25.5 **	6./4	30./14		9./3	16./6				1./1	**19./10**	**9./3**
Galeopsis tetrahit agg.	19.4	4./2	21./9	16./6	12./3	6./2	9./3	6./2	4./2	9./3	**26./10**	**1./1**
Prunella vulgaris	16.2	2./2	6./4	18./6	16./4	20./6	4./1	8./2	21./6	3./1	**20./8**	**4./2**
Stellaria graminea	14.7 **	3./1	1./1	4./2	4./1			9./3	3./1	2./2	**12./5**	**4./2**
Campanula rotundifolia	8.5	2./2		9./3	4./1			8./2	6./2	3./1	**6./4**	**3./1**
Gymnocarpium dryopteris	2.0					2./1					**1./1**	
Ancient woods on northern Millstone Grit (11)												
Carex laevigata	80.1 **		2./2								3./3	**14./6**
Molinia caerulea	71.0 **			4./1	2./1						23./8	**32./6**
Oreopteris limbosperma	69.9 **										20./10	**20./6**
Blechnum spicant	65.3 **		12./7								51./17	**25./7**
Calluna vulgaris	57.7 **		21./7	4./2	8./3				2./2		43./17	**36./6**
Carex remota	50.0 **	1./1	33./16	5./1	6./2	3./1				4./2	9./5	**18./6**
Cardamine pratensis	48.8 **	1./1	5./3		3./1						8./7	**11./5**
Nardus stricta	45.7 **				1./1					3./1	20./7	**23./4**
Luzula sylvatica	45.2 **		37./14			1./1					24./10	**19./5**
Carex echinata	39.2 **										5./2	**12./3**
Festuca ovina agg.	38.9 **	4./2	24./7	3./1	3./1			7./2	17./4	5./1	40./11	**31./5**
Fallopia japonica	35.4 **		3./1			17./3			1./1			**12./3**
Carex pallescens	28.6 **											**4./2**
Rhododendron ponticum	27.9 **	6./1	23./8	26./6	15./5	3./1	4./1		8./1		35./7	**20./5**
Lotus pedunculatus	27.0 **		1./1		3./1	4./2		9./2		3./1	2./1	**9./3**
Cynosurus cristatus	25.8 **	2./1	2./2	3./1	9./2	4./1		5./2	4./1		1./1	**15./3**
Caltha palustris	23.9 **	1./1	3./1		6./2	2./2		3./1			1./1	**7./3**
Salix aurita	22.7 **		1./1		3./1							**6./2**
Conopodium majus	18.1	3./2	6./3	4./1	7./3	22./7				6./2	1./1	**9./3**
Cardamine amara	17.7		7./3							2./1	4./1	**4./2**
Phegopteris connectilis	14.3											**1./1**
Cytisus scoparius	12.0		11./4	31./10	6./4		3./1	11./2	9./3	6./3	3./2	**6./3**
Phyllitis scolopendrium	11.1		1./1		1./1	5./2						**4./1**
Petasites hybridus	10.8	4./1	1./1	9./3	1./1	7./2	1./1		4./1			**5./2**
Stellaria nemorum	10.4			3./1								**1./1**
Galium odoratum	8.2	5./1	19./8			33./7					1./1	**5./2**
Melampyrum pratense	7.7		16./4	4./1	3./1							**5./1**
Lathraea squamaria	6.1										6./1	**1./1**
Number of sites		6	47	57	15	47	17	10	20	28	31	7

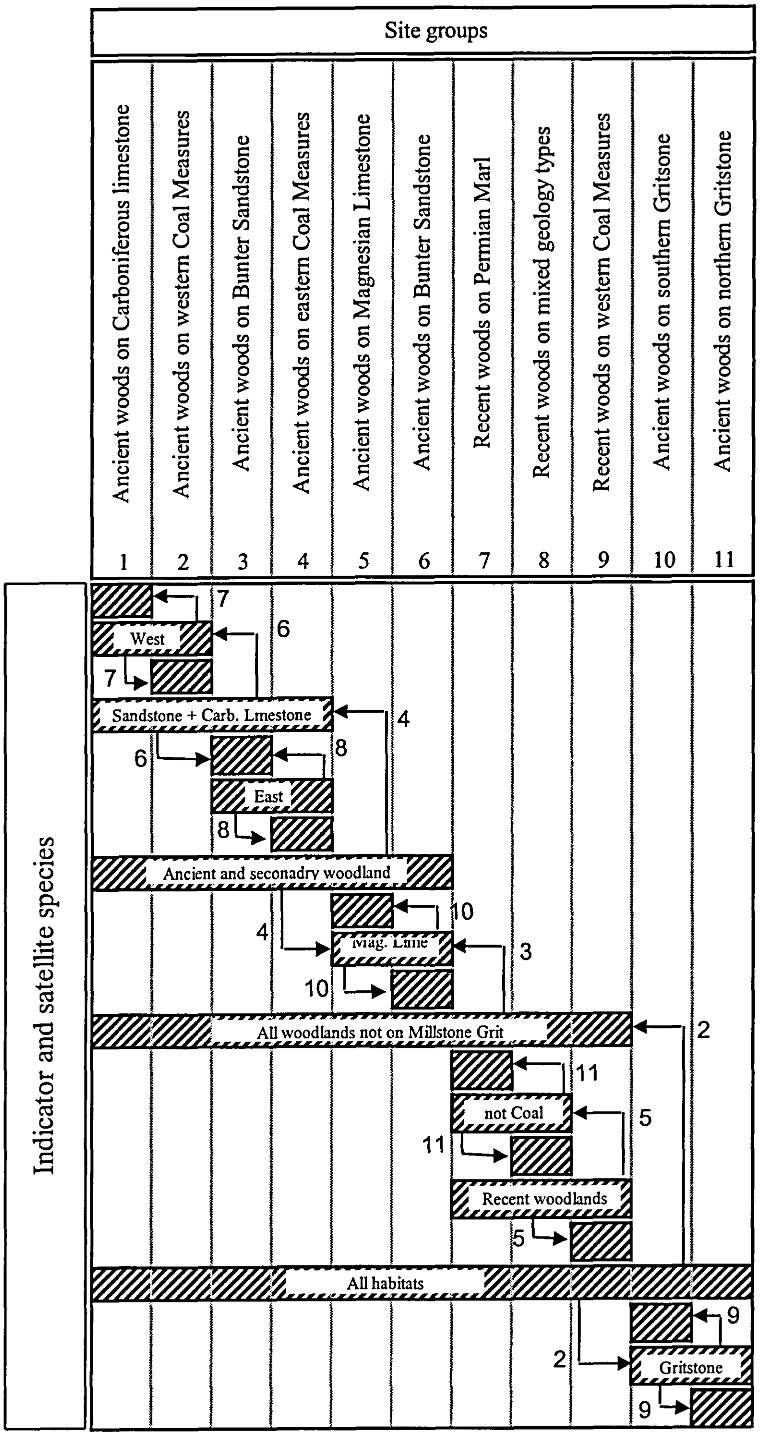


Figure 3.9. The structure of the two-way table of Table 3.12 showing the order in which groups are formed according to the dendogram of Figure 3.7.

3.5.7 Removal of the Carboniferous Limestone and Gritstone woodlands

The analysis reported in previous section was repeated on the same data set, but this time the woods on the Millstone Grit and Carboniferous Limestone were removed from the data set since there were few recent sites on these geology sites, which had resulted in unbalanced experimental design. The constraining factors were also changed. Since the main factors from the previous analysis appeared to be the geology, age, longitude, latitude and altitude these factors were kept. This time, age was entered as either ancient or recent, with recent including all the secondary sites. Permian Marl and Magnesian Limestone were pooled together since they were difficult to separate from each other. Interactions were added for the three geology types, and both the age of the sites and the interaction term for longitude × latitude were kept in. The same two covariables were also left in the analysis i.e. the date the sites were surveyed and the log. of the area of the sites surveyed. The eigenvalues and correlation coefficients are shown in Table 3.13. Most of these values can be seen to be lower than in the previous analysis, except for the correlation coefficient of the fourth axis. The *t*-values for the regression coefficients are shown for each axis and also Monte Carlo permutation tests reported, for the first four axes under a full model, in Table 3.14. Ordinations for the sample scores for the first two axes are shown in Figure 3.10 and the sample scores for the third and fourth axes are shown in Figure 3.11. The species scores and environmental biplot scores for the first two CCA axes are shown in Figure 3.12.

Table 3.13. Eigenvalues and correlation coefficients after the sites on the Millstone Grit and Carboniferous Limestone have been omitted. Total inertia = 4.543, sum of canonical eigenvalues = 0.433.

	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Eigenvalues	0.16	0.12	0.06	0.04
Correlation coefficients	0.84	0.77	0.75	0.72

Table 3.14. *t*-values of regression coefficients for standardised variables. *F* – Monte Carlo permutation for the first four axes under a full model.

Variable	CCA axis 1	CCA axis 2	CCA axis3	CCA axis 4	F
	t	t	t	t	
Altitude	0.92	-2.06 *	1.15	3.67 ***	1.84 ***
Longitude	-2.98 **	1.43	6.98 ***	1.30	3.78 ***
Latitude	-3.40 ***	2.72 **	-5.83 ***	4.16 ***	2.28 ***
Longitude × Latitude	-1.70	2.11 *	-10.07 ***	-5.64 ***	2.27 ***
Ancient	0.64	-9.30 ***	6.27 ***	4.92 ***	5.50 ***
Bunter Sandstone	2.58 *	-3.07 **	1.78	6.06 ***	1.64 **
Permian Marl/Mag. Limestone	-6.10 ***	-4.61 ***	-0.96	8.27 ***	8.59 ***
Coal Measures	0.00	0.00	0.00	0.00	
Bunter Sandstone × Ancient	-0.46	3.33 **	-4.24 ***	-4.27 ***	1.53 **
Mag. Limestone × Ancient	-0.71	2.03 *	-1.28	-9.21 ***	
Coal Measures × Ancient	0.00	0.00	0.00	0.00	1.97 ***

A clear separation is evident from Figure 3.10 and Table 3.13 indicating that the first axis explains most of the variation due to the difference between the sites on the Magnesian Limestone/Permian Marl and the sites on the sandstone (Bunter Sandstone and Coal Measures). The second axis shows a very clear separation with the ancient sites at the bottom of the ordination and the recent sites at the top of the ordination. However, there is no gap in the groups shown along the second axis. This suggests that this is a continuous gradient including heavily disturbed or managed ancient sites, as well as some species-rich secondary sites which may be ancient in part or close to an existing ancient woodland. Any separations on the third and fourth axes are much less evident. The only definite groupings appear to be for the recent versus the ancient sites on the Bunter Sandstone.

The ordination of the species scores for the first two axes is now much clearer in terms of the species associated with the ancient woodland on the main geology types. For example, *Luzula sylvatica*, *Luzula pilosa*, *Carex pendula*, *Blechnum spicant* and *Adoxa moschatellina* are in the direction of the ancient sites on the sandstones. This can be compared to species such as *Equisetum telmateia*, *Tilia platyphyllos*, *Daphne laureola* and *Melica uniflora*, which are in the direction of ancient sites on the Magnesian Limestone and Permian Marl. Species in the direction of the ancient sites on the Bunter Sandstone can also be seen including *Teucrium scorodonia*, *Ceratocapnos claviculata* and *Viola palustris*. Some of the species associated with the recent woods appear to be *Rumex acetosella*, *Cytisus scoparius*, *Galium verum* and *Agrostis gigantea*.

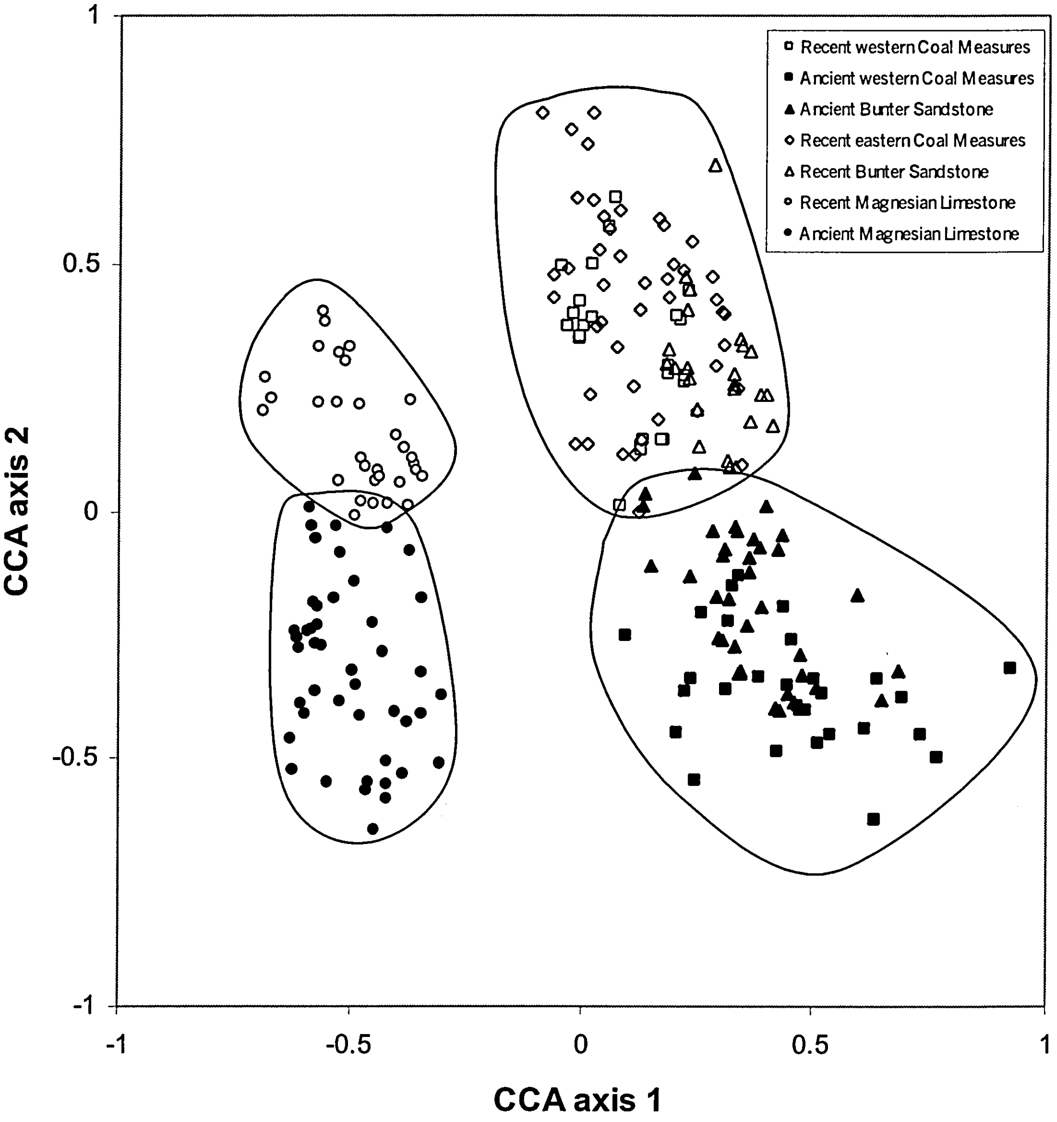


Figure 3.10. Ordination of sample scores on the first axes produced by canonical correspondence analysis. The polygons and symbols refer to the groups obtained by the Ward's clustering method shown in Figure 3.12. Magnesian Limestone refers to both Magnesian Limestone and Permian Marl.

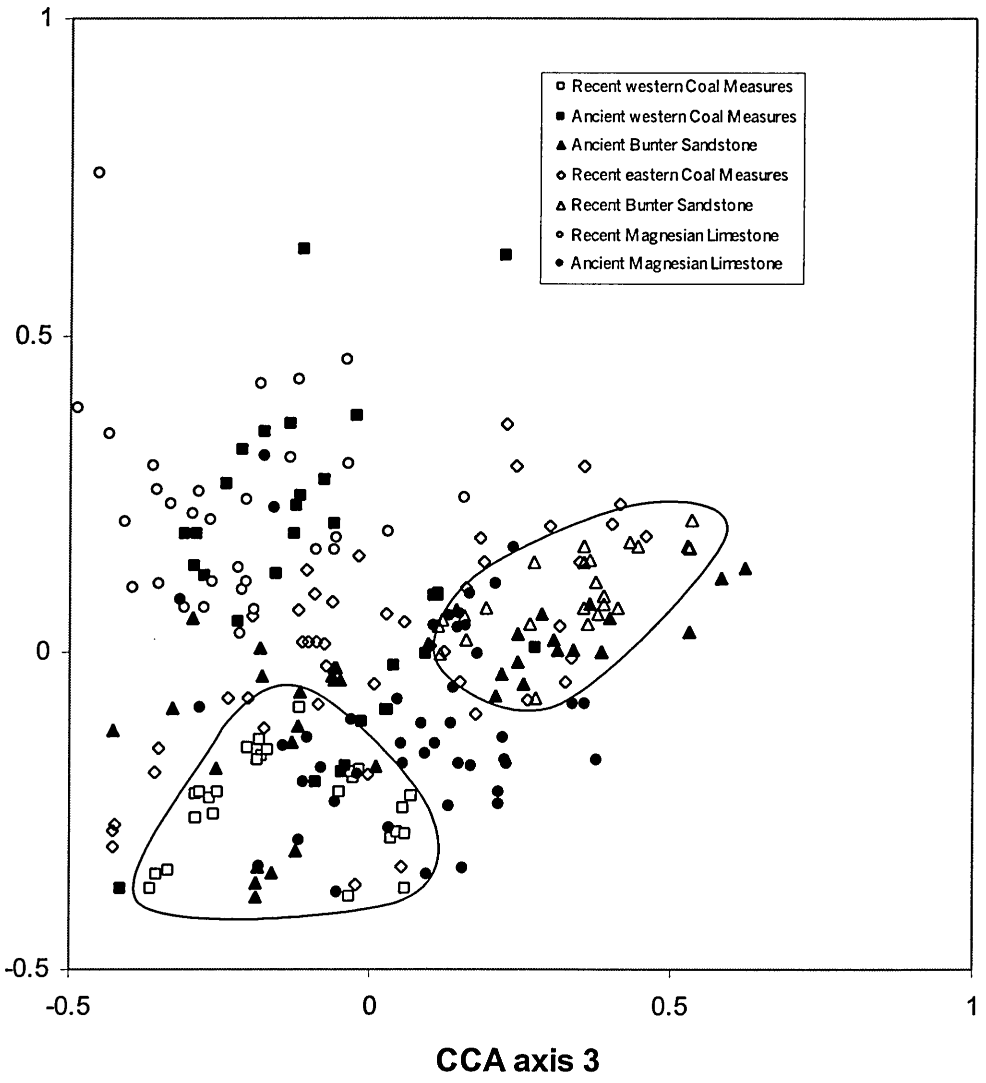


Figure 3.11. Ordination of sample scores on the third and fourth axes produced by canonical correspondence analysis. The polygons and symbols refer to the groups obtained by the Ward's clustering method shown in Figure 3.16. Magnesian Limestone refers to both Magnesian Limestone and Permian Marl. The clearest separation can be seen to occur between the recent sites on the western Coal Measures Series and the recent sites on the Bunter Sandstone.

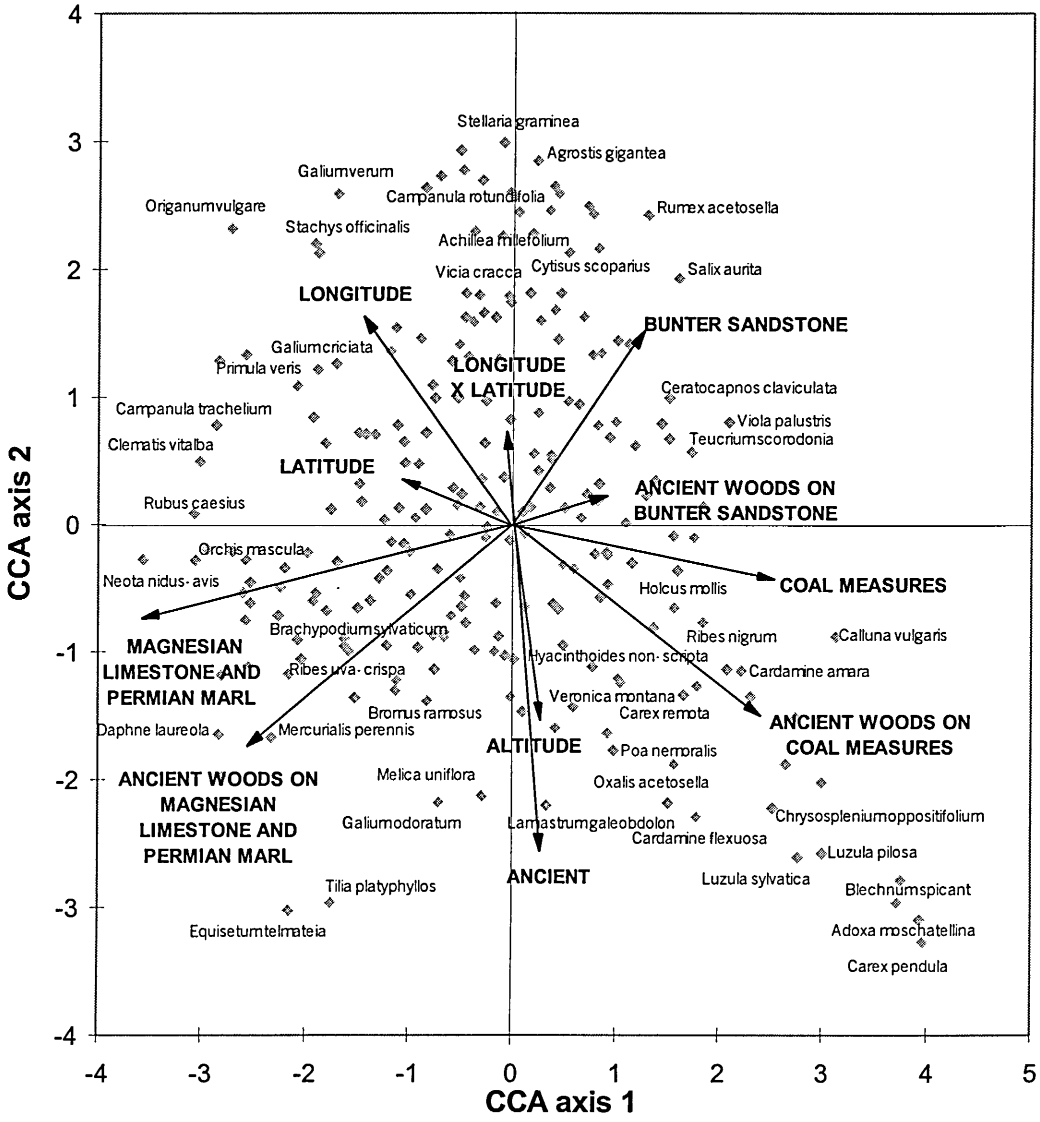


Figure 3.12. Ordination of the species scores on the first two axes produced by canonical correspondence analysis.

3.5.8 Species distributions across the ordinations

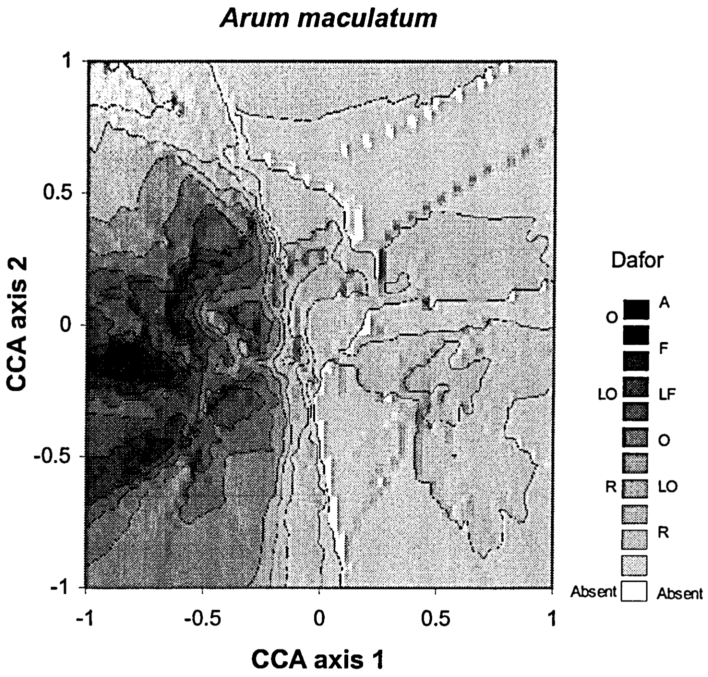
In order to look at the trends of species across the ordination more thoroughly kriging was used. This had also been tried for the first analysis with the ordination shown in Figure 3.2. However, this ordination suffers from the fact that there are a large number of sites at the centre of the ordination and any attempt at kriging results in inestimable values in the middle of the ordination diagram. This could have been overcome by averaging some of the values at the centre of the ordination. However, since the second analysis has produced such a clear spread of samples kriging was only applied to the ordination shown in Figure 3.10.

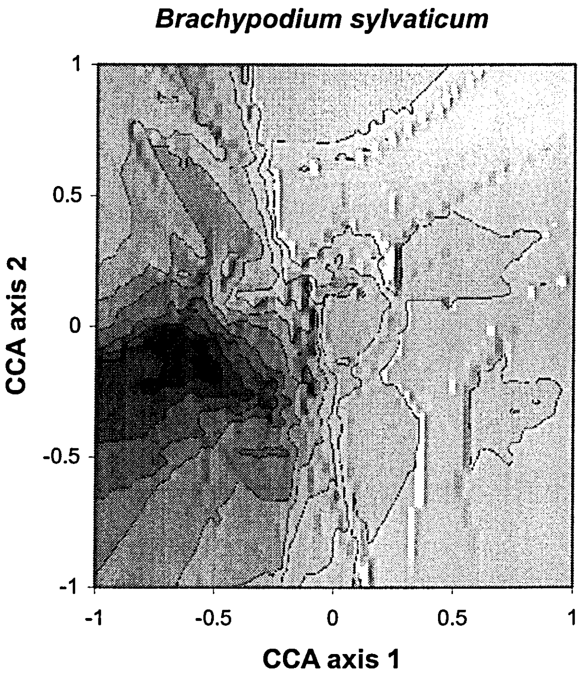
Only those species, which showed clear semi-variograms, which started with low intercepts (nugget) and increased towards a maximum value (the sill), were included in the analysis. The descriptions of the semi-variograms for the species selected are shown in Table 3.15. The range refers to the distance after which there is no further difference between samples for that species. The distributions of species produced by kriging are shown in Figure 3.13.

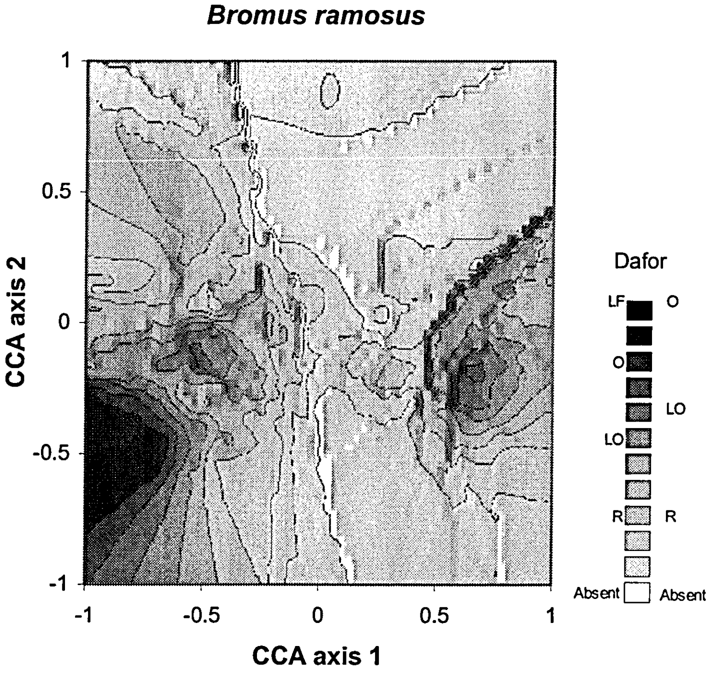
Table 3.15. Descriptions of the semi-variograms used for the kriging displayed in Figure 3.14.

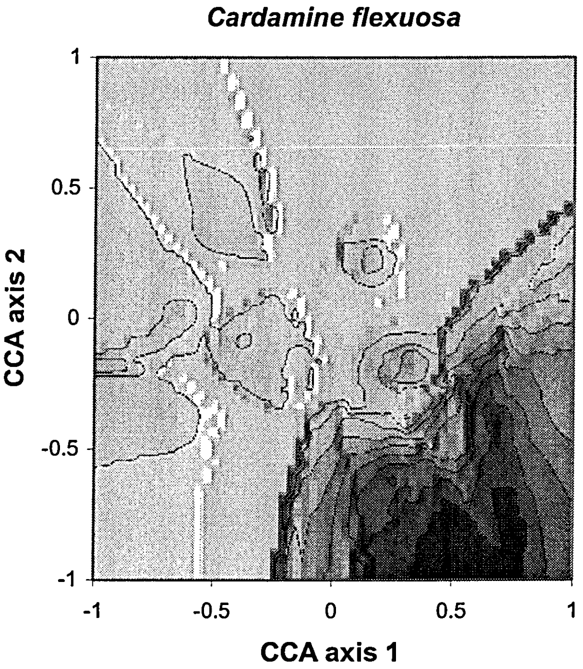
Species	Model	Nugget	sill	Range	R^2
Arum maculatum	Gaussian	1.30	3.10	1.46	0.93
Brachypodium sylvaticum	Gaussian	1.77	5.43	1.23	0.94
Bromus ramosa	Spherical	1.27	4.44	3.00	0.94
Cardamine flexuosa	Exponential	0.56	3.14	8.89	0.80
Corylus avellana	Exponential	3.06	6.35	4.28	0.83
Deschampsia flexuosa	Spherical	3.12	8.64	2.45	0.91
Geum urbanum	Gaussian	1.95	7.04	1.79	0.94
Holcus mollis	Spherical	5.06	16.55	2.63	0.92
Hyacinthoides non-scripta	Exponential	3.06	6.12	0.32	0.96
Lamiastrum galeobdolon	Exponential	1.88	5.77	6.73	0.81
Mercurialis perennis	Gaussian	2.96	14.64	1.49	0.94
Milium effusum	Exponential	2.53	9.07	6.08	0.71
Oxalis acetosella	Exponential	1.29	3.61	6.74	0.89
Pteridium aquilinum	Exponential	2.78	8.66	0.25	0.81
Ranunculus ficaria	Exponential	2.08	4.17	0.87	0.76
Stachys sylvatica	Exponential	2.15	7.31	6.57	0.87

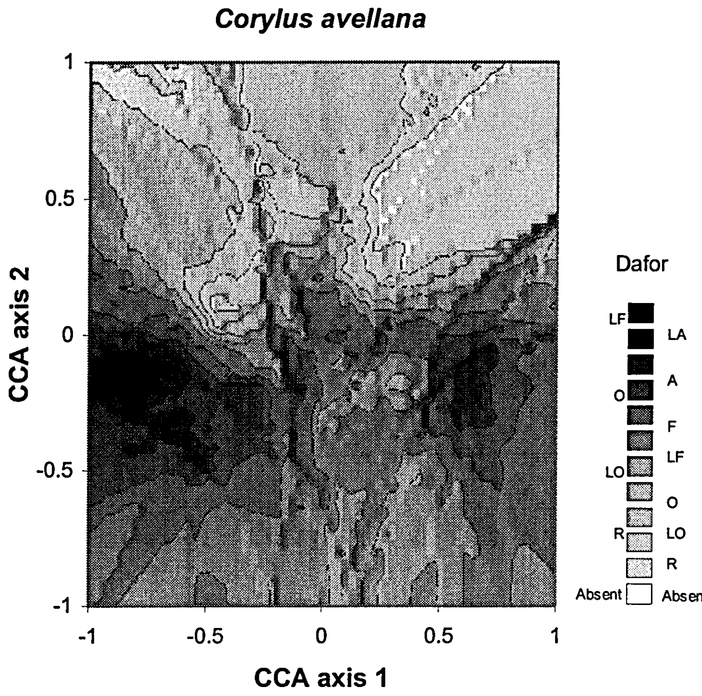
A number of species can be seen to occur mainly on the Magnesian Limestone and Permian Marl; these are *Arum maculatum, Brachypodium sylvaticum, Geum urbanum, Mercurialis perennis,* and *Stachys sylvatica.* Other species can be to occur mainly on the sites on the sandstones; these are *Holcus mollis* and *Pteridium aquilinum.* The species *Bromus ramosus* and possibly *Mercurialis perennis* appear to be most abundant in the ancient sites on the Magnesian Limestone and Permian Marl. *Cardamine flexuosa, Deschampsia flexuosa*, and *Oxalis acetosella* appear to have their greatest abundance in the ancient sites on the sandstones. *Corylus avellana*, *Lamiastrum galeobdolon* and *Milium effusum* appear to be general indicators of ancient woodland, with little preference for a particular geology type. It is not clear what might be influencing the distributions of *Hyacinthoides non-scripta* and *Ranunculus ficaria*, although both species appear to be generally in greater abundance in the ancient sites.

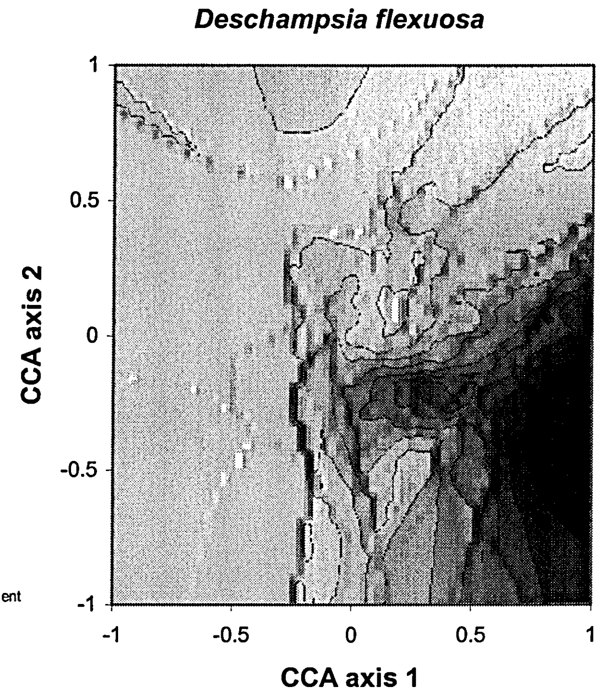

Arum maculatum

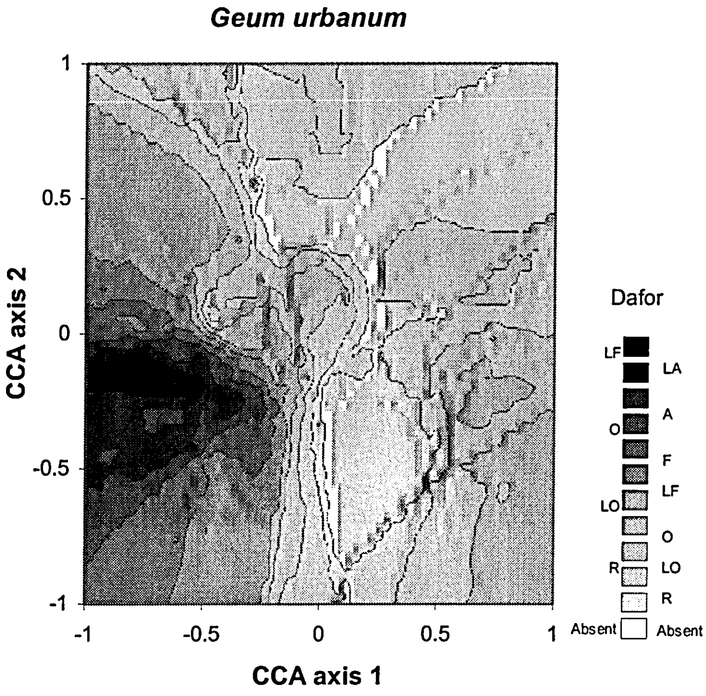

Brachypodium sylvaticum

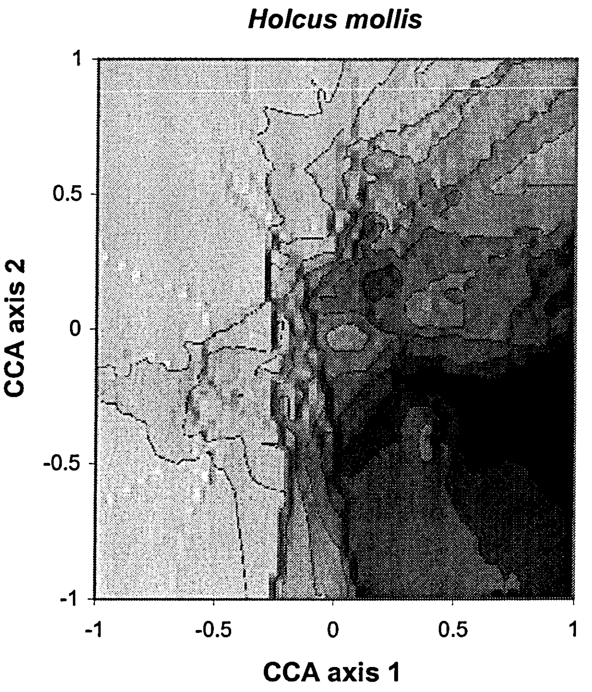

Bromus ramosus

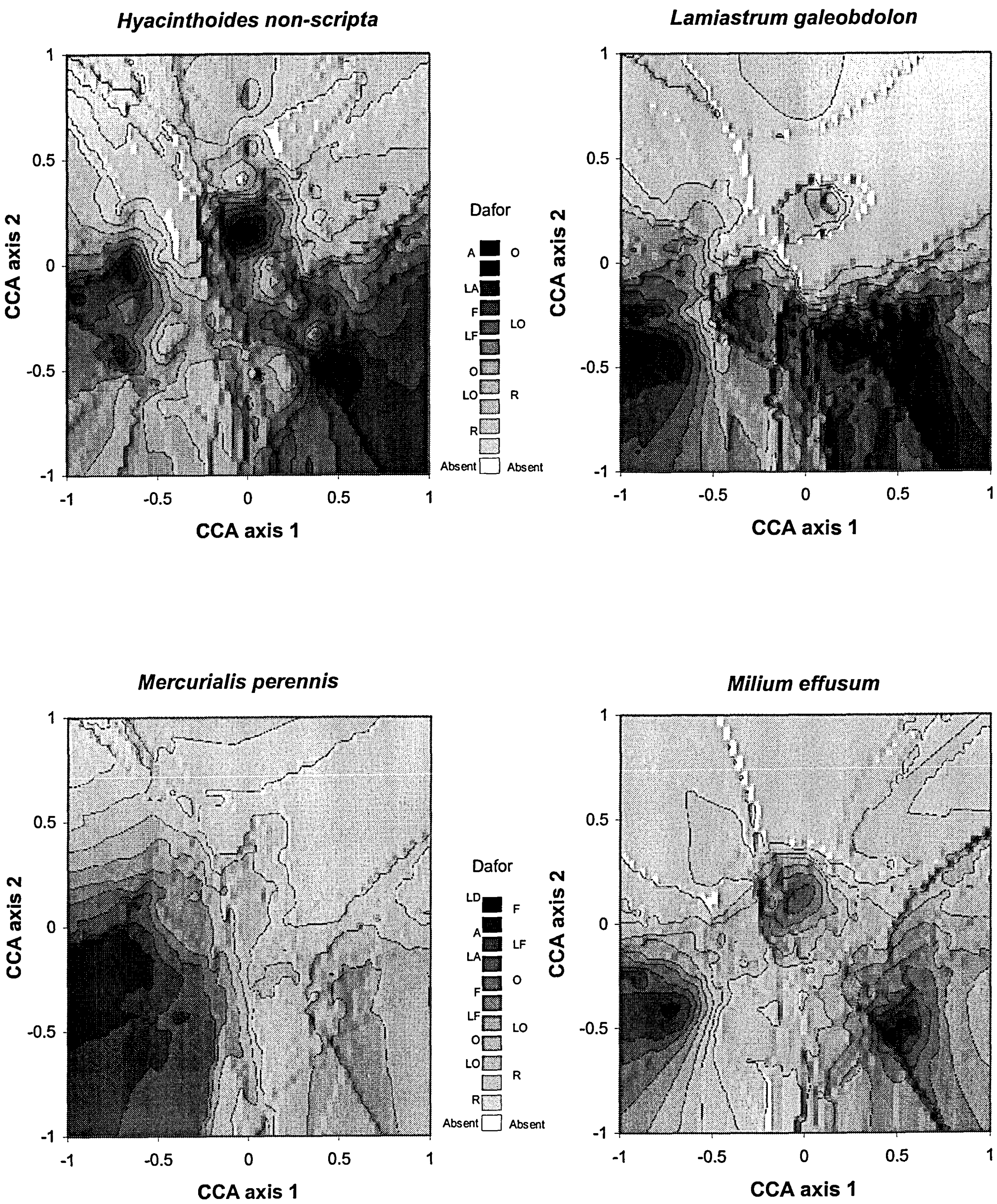

Cardamine flexuosa


Corylus avellana
CCA axis 2
CCA axis 1
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1
Dafor
LF
LA
A
O
F
LF
LO
O
R
LO
R
Absent
Absent


Deschampsia flexuosa
CCA axis 2
CCA axis 1
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1


Geum urbanum
CCA axis 2
CCA axis 1
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1
Dafor
LF
LA
A
O
F
LF
LO
O
R
LO
R
Absent
Absent


Holcus mollis
CCA axis 2
CCA axis 1
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1


Hyacinthoides non-scripta
Lamiastrum galeobdolon
Mercurialis perennis
Milium effusum
CCA axis 1
CCA axis 2
-1
-0.5
0
0.5
1
Dafor
A
LA
F
LF
O
LO
R
Absent
O
LO
R
Absent
LD
A
LA
F
LF
O
LO
R
Absent
F
LF
O
LO
R
Absent

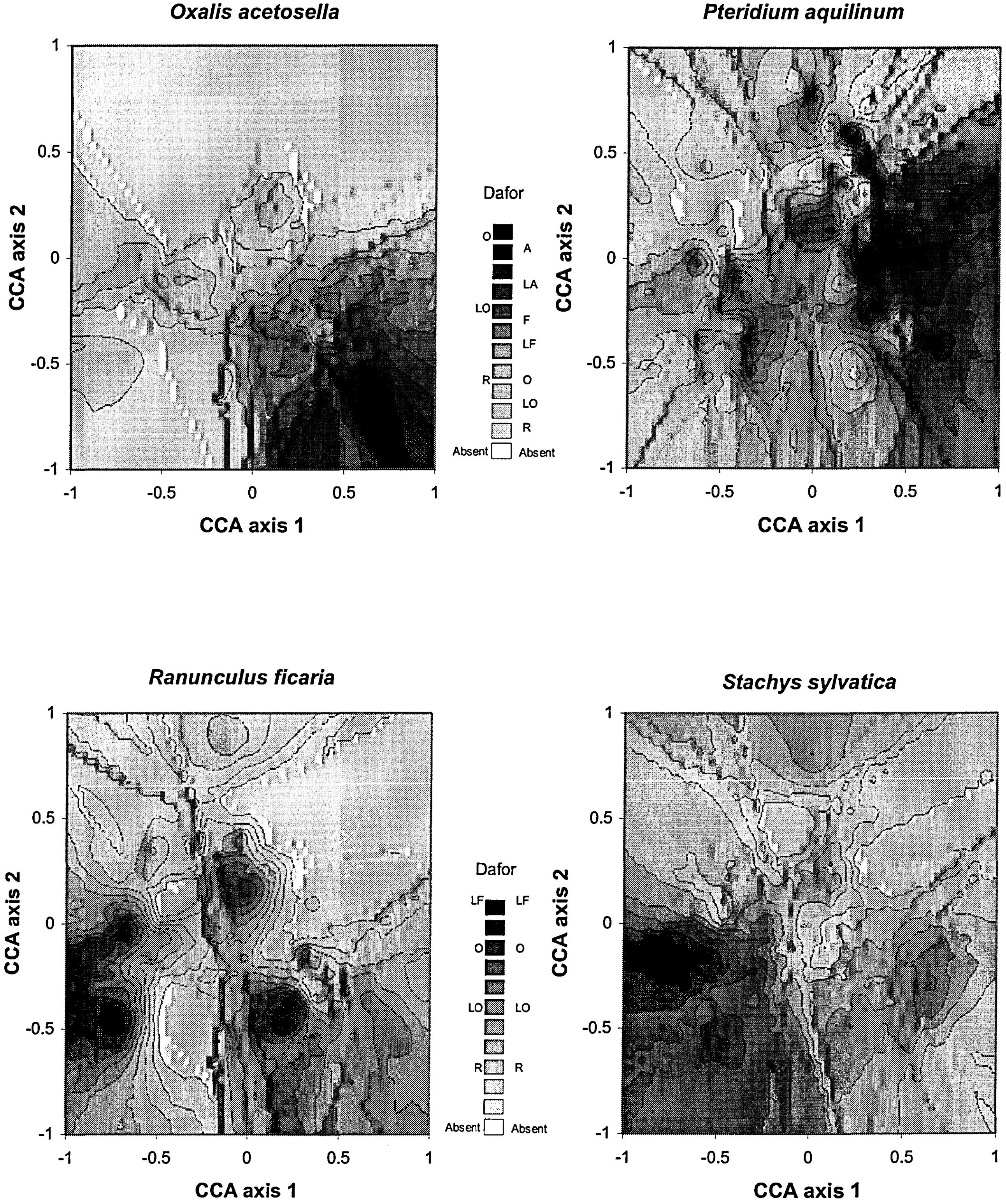


Figure 3.13. Spatial patterns of species across the first two axes produced by partial CCA analysis presented by the use of kriging. Descriptions of the semi-variograms used are given in Table 3.15.

3.5.9 Spatial auto-correlation

Spatial correlograms using Moran's *I*, were produced in the same way as in the original analysis with all the data. All the variables were first standardised as before and the weights calculated by CANOCO used in the multiple regression equations. The correlograms produced are displayed in Figure 3.14. The axes produced by the partial CCA analysis can all be seen to contain significant spatial auto-correlation which is not completely removed by models with the environmental variables listed in Table 3.14. This means that other factors not studied may explain this unaccounted for spatial-autocorrelation. If other variables not recorded are important then this spatial-autocorrelation may at occur at a range of different scales. The correlograms for the ordination axes show that woods of distances up to 10 km are significantly correlated. Once the models have been fitted the residual variation left appears to be correlated at a much smaller distances, less than 5 km. This auto-correlation occurring between woods very close to each other is much more likely to be associated with local soil variation rather than variables which differ at much greater distances such as precipitation.

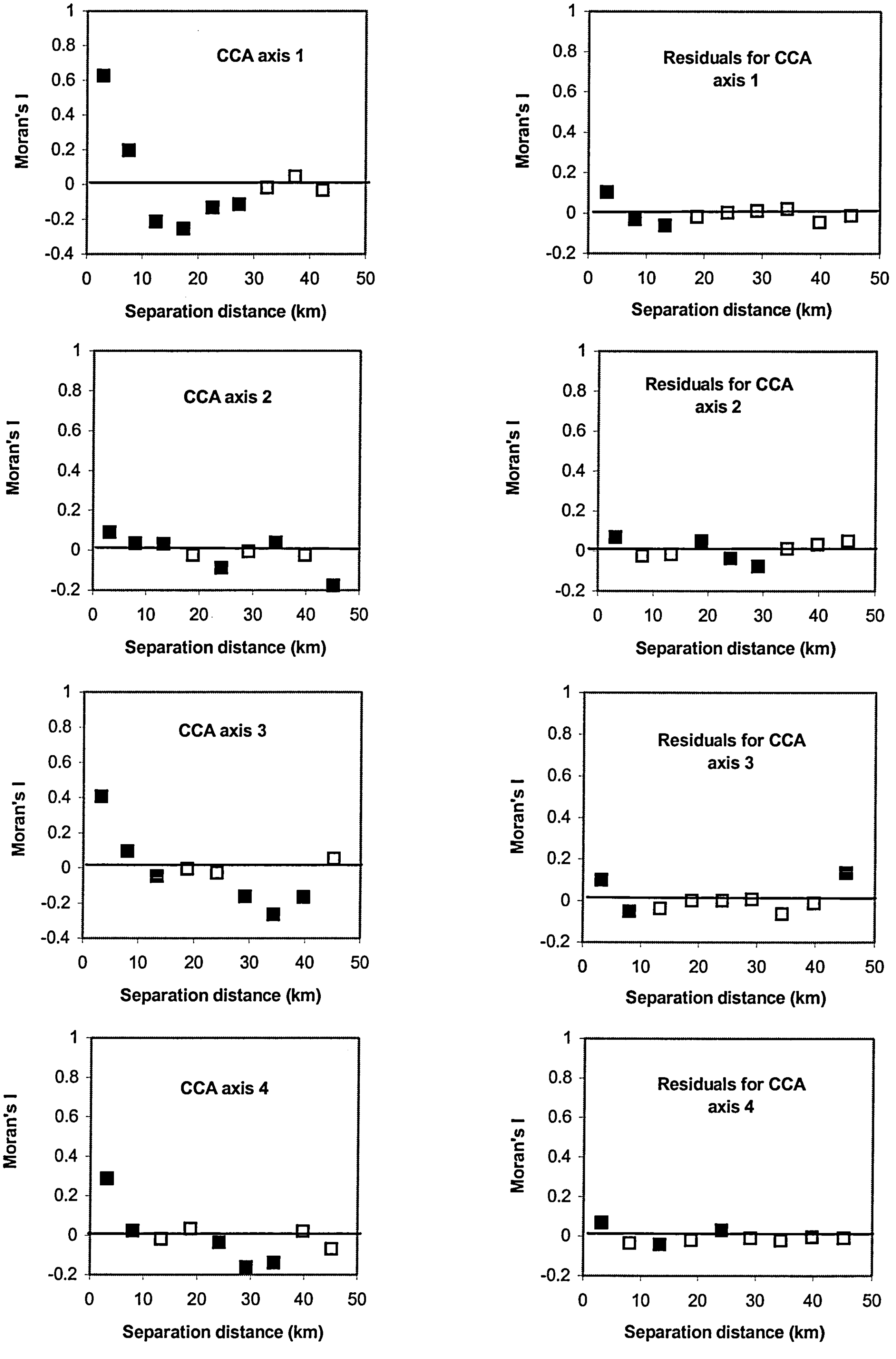


Figure 3.14. Spatial auto-correlogram for the CCA sample scores that are weighted averages of the species matrix and the residuals produced after the same model has been fitted on the data using multiple regression. Dark symbols: correlation statistics that are significant after progressive Bonferroni correction ($\alpha = 0.05$).

3.5.10 Variance partitioning

The technique for partitioning the variation described by Anderson and Gribble (1998) was again applied to the data to give an estimate as to what proportion of the variation in the data could be attributed to spatial, environmental and temporal factors. Monte Carlo permutation tests were again used to select the significant terms of a cubic surface regression model. The results of this test are shown in Table 3.16. Longitude still remains the variable that explains the largest amount of variation in the data and all the terms entered still remain significant. The longitude is likely to represent the main variation attributed to geology, precipitation and altitude. The geology varies from the Coal Measures Series in the East to Magnesian Limestone and Permian Marl in the middle followed by Bunter Sandstone in the West. Since this essentially represents a band of geology (Magnesian Limestone) capable of producing alkali soils, sandwiched between two types of sandstone which are normally associated with acidic soils this would be expected to produce an inverse curve if modelled using polynomial regression.

Table 3.16. Forward selection of spatial variables, Monte Carlo *F* under a full model. Longitude and Latitude were first centred on their means to reduce collinearity in the spatial data.

Spatial variable	F
Longitude	6.83 ***
Longitude^2	5.22 ***
Latitude	4.14 ***
Latitude^2	3.04 ***
Latitude^3	2.72 ***
Longitude^3	2.29 ***
$\text{Longitude} \times \text{Latitude}^2$	2.13 ***
$\text{Longitude}^2 \times \text{Latitude}$	2.01 *
Longitude × Latitude	1.49 **

The sum of canonical eigenvalues and corresponding percentage of explained variation for each of the CCAs indicated by steps [1]-[12] of Table 3.1 are shown in Table 3.17. The calculations used to find the variation explained by each component and the degree of overlap between them were repeated and presented in Table 3.18. A graphical representation is shown in Figure 3.15. The main difference between the original analysis and this one, after the removal of the Carboniferous Limestone and Gritstone woods, is that the latter analysis contains a smaller component that can be explained by the environmental variables. This emphasises the difference between the Gritsone woods and the other types of woodland, since this difference appears to explain much of the variation in the original data set.

Table 3.17. Summary of the results of constrained and partial canonical correspondence analysis.

Step in analysis	Value in Calculations	Sum of canonical eigenvalues	Explained Variation (%)	P
[1]	Ω_E	0.524	11.53	< 0.001
[2]	Ω_S	0.535	11.78	< 0.001
[3]	Ω_T	0.102	2.25	< 0.001
[4]		0.299	6.58	< 0.001
[5]		0.479	10.54	< 0.001
[6]	E	0.279	6.14	< 0.001
[7]		0.310	6.82	< 0.001
[8]		0.491	10.81	< 0.001
[9]	S	0.309	6.80	< 0.001
[10]		0.057	1.25	< 0.001
[11]		0.057	1.25	< 0.001
[12]	T	0.056	1.23	< 0.001

Table 3.18. Summary of calculations for the partitioning of variation with spatial, temporal and environmental variables.

Component	Calculation	Explained Variation (%)
Environmental = E	[6]	6.14
Spatial = S	[9]	6.80
Temporal = T	[12]	1.23
SE	$SE_T - STE = [1] - [4] - STE$	4.01
TE	$TE_S - STE = [1] - [5] - STE$	0.04
ST	$ST_E - STE = [2] - [8] - STE$	0.02
STE	[9] + ([2] – [7]) + ([2] – [8] – [2], or [6] + ([1] – [4]) + ([1] – [5]) – [1], or [12] +([3] – [10]) + ([3] – [11]) – [3]	0.95
Total explained	[1] + [7] + [12], or [2] + [4] + [12], or [3] + [5] + [9], or $S + T + E + SE + TE + ST + STE$	19.59
Unexplained	100% – Ω	80.41
Total variation	Total trace of CA = 4.543	100.00

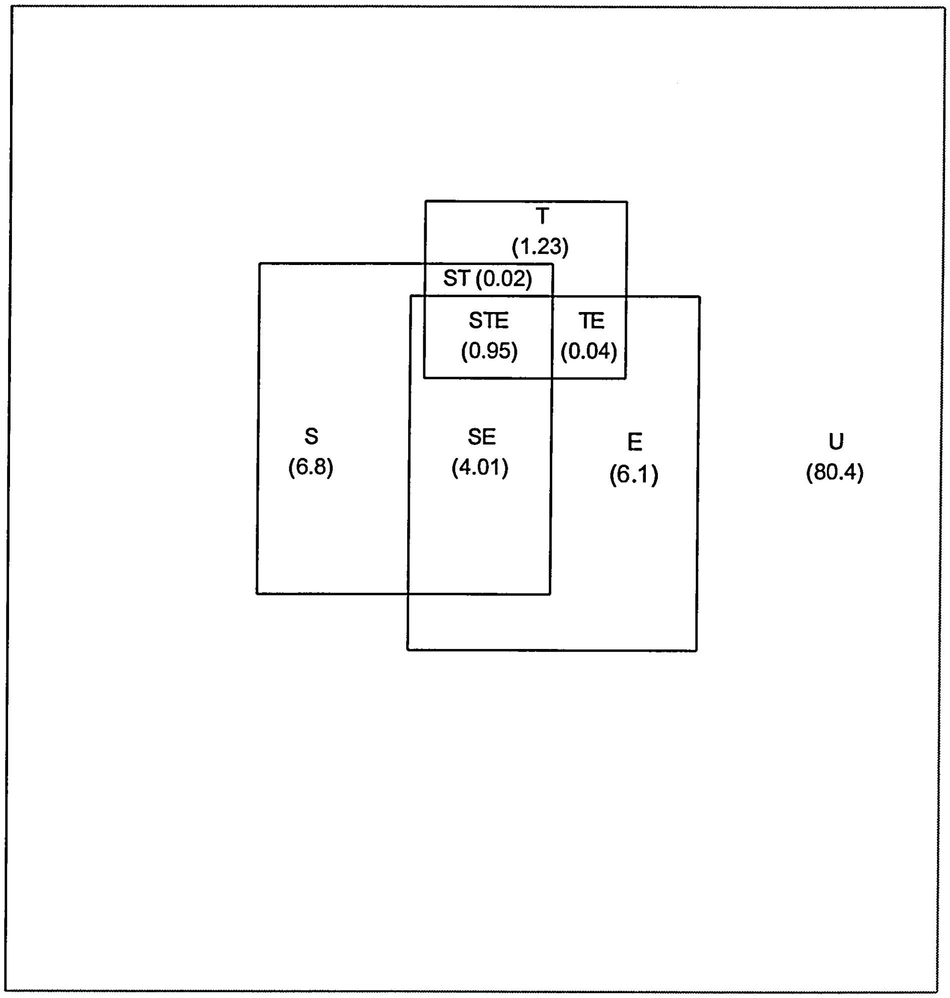


Figure 3.15. Approximate graphical representation of the partitioning of variation for the CCA analysis displayed in Table 3.18, following the method of Anderson and Gribble (1998). The components in the diagram are spatial (S), environmental (E) and the combined effects for time and area (T). The area outside these components and their interactions is the unexplained variation (U).

3.5.11 Spatial GLMs

General linear models with spatial structures were fitted to the partial CCA axes for the sample scores that were weighted averages of the species values. This again provides a further check using a more conservative method to test the significance of the results produced by CANOCO. The same three spatial structures, spherical, exponential and Gaussian, were fitted to the GLMs. However, instead of simply using the dummy variables that were used in CANOCO, categorical variables were used for age (ancient or recent) and for the three geology types. The values for the -2REML were used to select the best fitting structure. These values are shown in Table 3.19. The two structures that appear to give the best results are spherical and exponential. Tests for the improvement of the fit of the models are shown in Table 3.20. All three models with spatial structures fitted, give a significantly better fit than when no structure is specified. The results of the GLMs with exponential spatial structures are given in Table 3.21. It is worth noting that the GLMs with the spherical structures gave virtually identical results. The only differences between the GLMs with the spherical structures, to that of the ones with the exponential structures, was that the former gave longitude on the second axis a probability value of less than 0.001 and longitude and latitude on the forth axis were given probability values of $0.01 < P < 0.05$. The GLMs with the Gaussian structure did show some further minor differences with longitude not being significant for either the third or fourth axes.

Table 3.19. -2 REML Log Likelihood estimates for each of the models fitted when relating the CCA axes to the environmental variables. Values closest to zero are in **bold type.**

Spatial structure	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Spherical	423.7	**518.1**	**550.2**	565.7
Exponential	**422.9**	518.7	550.3	**564.1**
Gaussian	424.9	527.0	551.0	575.2
No structure specified	467.6	570.8	592.3	605.3

Table 3.20. Chi-square tests between the -2 REML Log Likelihood estimates with spatial structures specified and when spatial structure is specified

Structure	DF	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Spherical	2	43.9 ***	52.6 ***	42.1 ***	39.6 ***
Exponential	2	44.7 ***	52.1 ***	42.0 ***	41.2 ***
Gaussian	2	42.7 ***	43.8 ***	41.3 ***	30.1 ***

Some quite large differences can be seen between the *t*-values produced by CANOCO, given in Table 3.14, and those produced by the spatial GLMs of Table 3.21. A much clearer pattern can be seen from the spatial GLM of Table 3.21, with the main significant variables being geology, age and the age and geology interaction. Altitude and latitude now appear to explain far less of the variation on any of the axes. The importance of determining separate lists of indicator species is clearly evident from this table due to the very high *F*-values produced for age, geology and the interactions between the two. The overall fit of these spatial GLMs also appears to be much better than that done with the original data. Sensible values can be seen for the covariance parameter estimates given in Table 3.21. The nugget is always less than the sill and the ranges show a good agreement with correlograms displayed in Figure 3.14. Only the estimate for the range of the second axis does not appear to correspond to that indicated by the correlogram.

Table 3.21. Type-III F tests relating the first four axes to the environmental data using categorical variables for age (ancient and recent) and the four different geology types and including exponential spatial structures. The estimated vales for the sill, range and nugget are also shown.

Source	NDF	DDF	CCA axis 1 F	CC axis 2 F	CC axis 3 F	CC axis 4 F
Effect						
Altitude	1	231	0.0	2.0	0.5	0.5
Longitude	1	231	1.0	10.8 **	10.6 **	1.5
Latitude	1	231	2.3	0.8	0.1	0.4
Longitude × Latitude	1	231	1.8	0.5	10.1 **	0.1
Age	1	231	2.2	119.1 ***	30.0 ***	11.7 ***
Geology	2	231	20.8 ***	13.1 ***	2.3	5.2 **
Age × Geology	2	231	1.3	2.1	3.8 *	19.7 ***
Covariance Parameter Estimates						
Sill			55.0 *	174.9	150.9	242.8
Range (km)			19.0	79.0	33.3	30.6
Nugget			27.4 ***	46.0 ***	48.0 ***	44.4 ***

3.5.12 Cluster analysis

The first eight CCA axes, that were linear combinations of the environmental variables, were used for the cluster analysis as before. The clustering methods used were flexible clustering and Ward's clustering method. Both methods produced the same primary division i.e. separating the sites on the Magnesian Limestone/Permian Marl from the sites in the sandstones. The Ward's clustering method then separated the sites on the eastern Coal Measures and Bunter sandstone from the sites on the western Coal Measures with the following separations mainly concerned with the age of the sites. The second division produced by the Flexible clustering method separated the ancient sites from the recent sites on the sandstones and then further divisions were mainly concerned with the geology. The two clustering methods produced almost identical end clusters with the main differences occuring at the second and third divisions. Since the Ward's clustering method split the sites up according to their geology at a stage earlier and then grouped them according to age it was the preferred method for this particular study. The dendogram produced by the Ward's clustering method is shown in Figure 3.16. A confusion table is shown in Table 3.22 showing the '*a priori*' habitat classifications with those produced by the Ward's clustering method. The major problem the clustering algorithm appears to have had is when grouping sites on the eastern and western Coal Measures. This is perhaps not surprising since this gradient represents a continuous gradient of altitude and longitude. This makes making the decision very arbitrary as to how the sites should be separated. For this study the sites that have been classified as western are those sites with a longitude of less than 4440 and an altitude above 100 m.

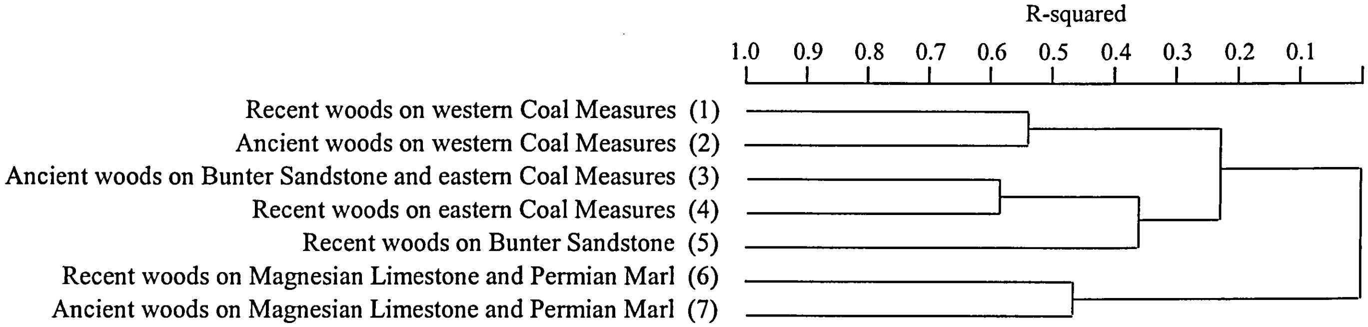


Figure 3.16. Dendogram produced by Wards clustering method on the data set with the sites on the Carboniferous Limestone and Millstone Grit removed from the data.

Table 3.22. Confusion table for the cluster analysis after the removal of sites on the Millstone Grit and Carboniferous Limestone.

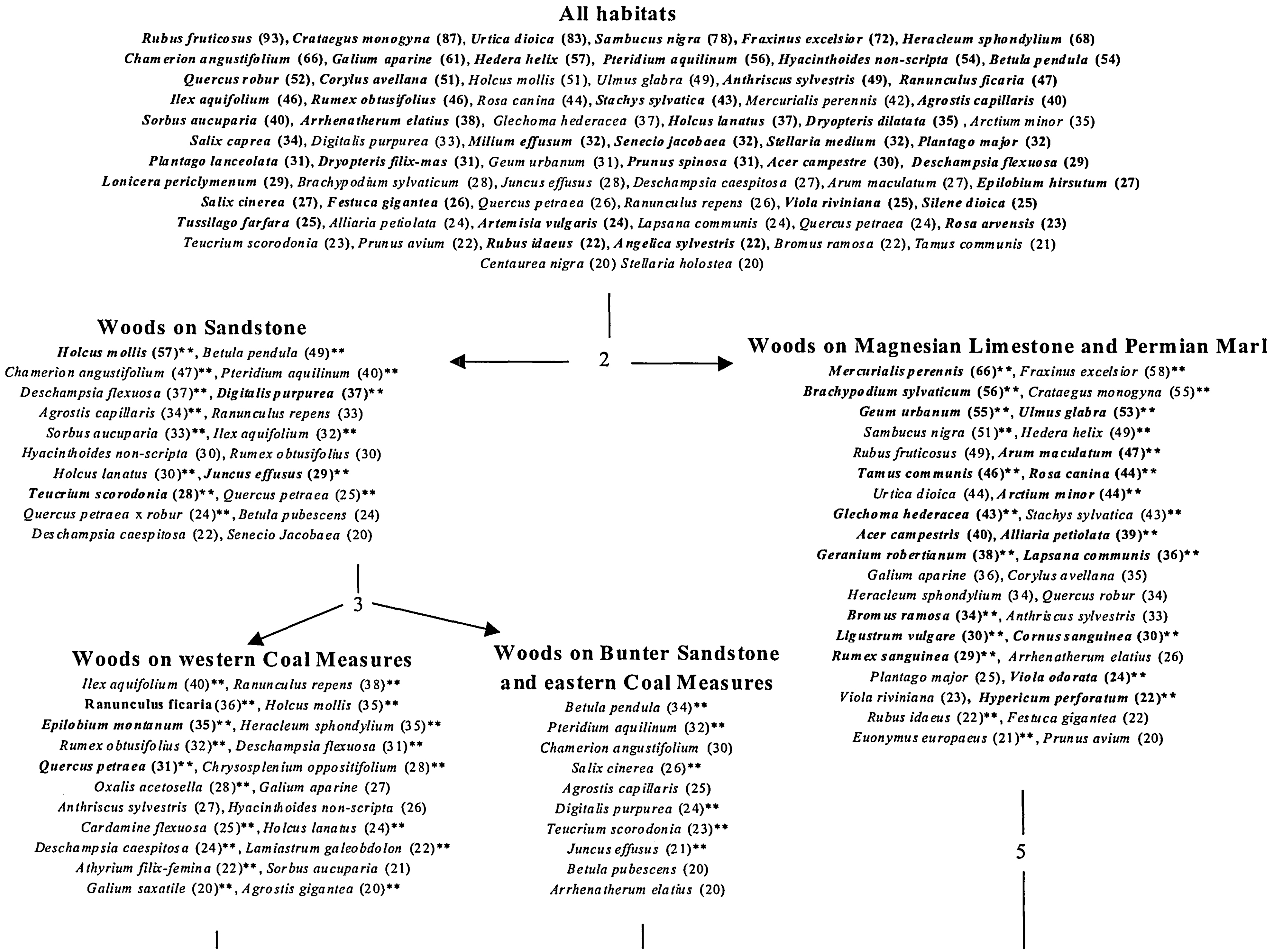
Habitats (*a priori*)	Habitat (cluster number)							
	1	2	3	4	5	6	7	S
Recent woods on western Coal Measures	**21**	1		1				
Ancient woods on western Coal Measures		**30**	15	1				
Ancient woods on Bunter Sandstone and eastern Coal Measures	6		**23**	12				
Recent woods on eastern Coal Measures				**8**				
Recent woods on Bunter Sandstone				23	**24**			
Recent woods on Magnesian Limestone and Permian Marl						**31**		
Ancient woods on Magnesian Limestone and Permian Marl							**45**	
Sum	27	31	38	45	24	31	45	

3.5.13 Indicator analysis

The species associated with each partition of the cluster analysis, of Figure 3.16, are shown in Figure 3.17. Since this dendogram contains fewer partitions than the previous analysis with all the data included, a smaller threshold value of 20 % was chosen. The groups, which did not contain any species making this criteria, have their highest four species shown. The indicator values for the sites on the Coal Measures can be seen to be considerably higher than the previous analysis indicating less variation within these partitions. As mentioned previously the partitions shown in this present analysis separate sites according to whether they are ancient or not instead of combining the secondary sites with the ancient sites. In the original analysis with all the data included the cluster analysis and INDVAL procedure are struggling to find indicator species for the sites on the Coal Measure Series. The species that are particularly associated with sites on the different sandstones can now be clearly seen from Figure 3.17. These species include *Holcus mollis*, *Pteridium aquilinum*, *Deschampsia flexuosa*, and *Agrostis capillaris*. The next partition separates the sites on the western Coal Measures from those on the eastern Coal Measures and Bunter Sandstone. Species which are associated with the western Coal Measures include *Ranunculus repens*, *Ranunculus ficaria*, *Holcus mollis* and *Deschampsia flexuosa*. Those species that are associated with the Bunter Sandstone and eastern Coal Measures, include *Pteridium aquilinum*, *Digitalis purpurea* and *Teucrium scorodonia*. The species that appear to be indicators of the ancient sites on the western Coal Measures include *Deschampsia flexuosa*, *Athyrium filix-femina*, *Cardamine flexuosa* and *Oxalis acetosella*. The species that are significantly associated with the ancient sites on the eastern Coal Measures and Bunter Sandstone include *Pteridium aquilinum*, *Teucrium scorodonia* and

Ceratocapnos claviculata. The species that are indicators of the ancient sites on the Magnesian Limestone and Permian Marl appear to have changed little from the original analysis. *Mercurialis perennis*, *Geum urbanum*, *Brachypodium sylvaticum* and *Tamus communis* all remain good indicators of this group. However, the number of species which have their maximum indicator value in this group is greatly reduced. These species are now *Prunus avium*, *Euonymus europaea*, *Campanula latifolia* and *Hypericum humifusum*. These species may therefore be the best indicators of this group as they are not common to the sites on the Magnesian Limestone and Permian Marl in general. A complete list of the species entered into the analysis is shown in Table 3.23 together with the sum of their abundance and presence values and indicator values. The structure of this table is presented in Figure 3.18. This table shows more clearly which of the rare species are good indicators of particular communities which may be very restricted to a habitat-type but occur at too low an abundance or frequency to be meet the threshold used in Figure3.18. For example, *Carex remota*, *Lysimachia nemorum*, *Epipactis helleborine* and *Myosotis sylvatica* can all be seen to have their maximum indicator value in the ancient sites on the western Coal Measures. A complete list of the species that have their maximum indicator value on the Magnesian Limestone and Permian Marl can also be seen. As well as the species listed in Figure 3.17, other species associated with the ancient sites can be seen to include *Ribes uva-crispa*, *Viburnum opulus*, *Carex sylvatica*, *Allium ursinum* and *Melica uniflora*.

Fig. 3.17. Site clusters obtained with the Ward's clustering method after the removal of the sites on the Carboniferous Limestone an Millstone Grit (as shown in Fig. 3.16), but with associated indicator values in parentheses. All species with indicator values of >30 % ar mentioned for each site cluster they are found. Species that are at their maximum are in bold.



All habitats

***Rubus fruticosus* (93), *Crataegus monogyna* (87), *Urtica dioica* (83), *Sambucus nigra* (78), *Fraxinus excelsior* (72), *Heracleum sphondylium* (68)**
***Chamerion angustifolium* (66), *Galium aparine* (61), *Hedera helix* (57), *Pteridium aquilinum* (56), *Hyacinthoides non-scripta* (54), *Betula pendula* (54)**
***Quercus robur* (52), *Corylus avellana* (51),** *Holcus mollis* (51), *Ulmus glabra* (49), ***Anthriscus sylvestris* (49), *Ranunculus ficaria* (47)**
***Ilex aquifolium* (46), *Rumex obtusifolius* (46),** *Rosa canina* (44), ***Stachys sylvatica* (43),** *Mercurialis perennis* (42), ***Agrostis capillaris* (40)**
***Sorbus aucuparia* (40), *Arrhenatherum elatius* (38),** *Glechoma hederacea* (37), ***Holcus lanatus* (37), *Dryopteris dilatata* (35)**, *Arctium minor* (35)
***Salix caprea* (34),** *Digitalis purpurea* (33), ***Milium effusum* (32), *Senecio jacobaea* (32), *Stellaria medium* (32), *Plantago major* (32)**
***Plantago lanceolata* (31), *Dryopteris filix-mas* (31),** *Geum urbanum* (31), ***Prunus spinosa* (31), *Acer campestre* (30), *Deschampsia flexuosa* (29)**
***Lonicera periclymenum* (29),** *Brachypodium sylvaticum* (28), *Juncus effusus* (28), *Deschampsia caespitosa* (27), *Arum maculatum* (27), ***Epilobium hirsutum* (27)**
***Salix cinerea* (27), *Festuca gigantea* (26),** *Quercus petraea* (26), *Ranunculus repens* (26), ***Viola riviniana* (25), *Silene dioica* (25)**
***Tussilago farfara* (25),** *Alliaria petiolata* (24), ***Artemisia vulgaris* (24),** *Lapsana communis* (24), *Quercus petraea* (24), ***Rosa arvensis* (23)**
Teucrium scorodonia (23), *Prunus avium* (22), ***Rubus idaeus* (22), *Angelica sylvestris* (22),** *Bromus ramosa* (22), *Tamus communis* (21)
Centaurea nigra (20) *Stellaria holostea* (20)

2

Woods on Sandstone

***Holcus mollis* (57)**,** *Betula pendula* (49)**
Chamerion angustifolium (47)**, *Pteridium aquilinum* (40)**
Deschampsia flexuosa (37)**, ***Digitalis purpurea* (37)****
Agrostis capillaris (34)**, *Ranunculus repens* (33)
Sorbus aucuparia (33)**, *Ilex aquifolium* (32)**
Hyacinthoides non-scripta (30), *Rumex obtusifolius* (30)
Holcus lanatus (30)**, ***Juncus effusus* (29)****
***Teucrium scorodonia* (28)**,** *Quercus petraea* (25)**
Quercus petraea x *robur* (24)**, *Betula pubescens* (24)
Deschampsia caespitosa (22), *Senecio Jacobaea* (20)

Woods on Magnesian Limestone and Permian Marl

***Mercurialis perennis* (66)**,** *Fraxinus excelsior* (58)**
***Brachypodium sylvaticum* (56)**,** *Crataegus monogyna* (55)**
Geum urbanum* (55)**, *Ulmus glabra* (53)*
Sambucus nigra (51)**, *Hedera helix* (49)**
Rubus fruticosus (49), ***Arum maculatum* (47)****
Tamus communis* (46)**, *Rosa canina* (44)*
Urtica dioica (44), ***Arctium minor* (44)****
***Glechoma hederacea* (43)**,** *Stachys sylvatica* (43)**
Acer campestris* (40), *Alliaria petiolata* (39)*
Geranium robertianum* (38)**, *Lapsana communis* (36)*
Galium aparine (36), *Corylus avellana* (35)
Heracleum sphondylium (34), *Quercus robur* (34)
***Bromus ramosa* (34)**,** *Anthriscus sylvestris* (33)
Ligustrum vulgare* (30)**, *Cornus sanguinea* (30)*
***Rumex sanguinea* (29)**,** *Arrhenatherum elatius* (26)
Plantago major (25), ***Viola odorata* (24)****
Viola riviniana (23), ***Hypericum perforatum* (22)****
Rubus idaeus (22)**, *Festuca gigantea* (22)
Euonymus europaeus (21)**, *Prunus avium* (20)

3

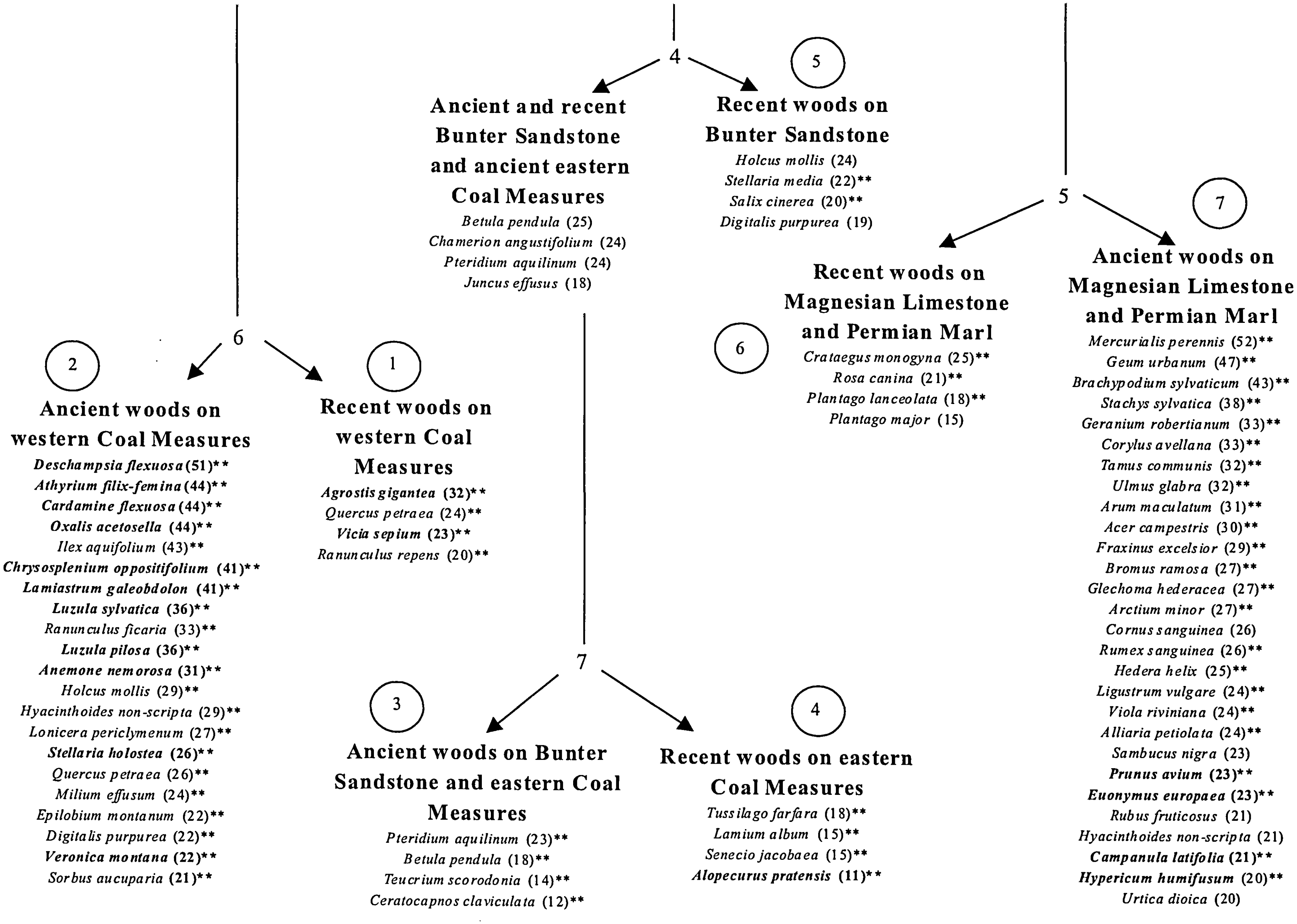
Woods on western Coal Measures

Ilex aquifolium (40)**, *Ranunculus repens* (38)**
Ranunculus ficaria (36),** *Holcus mollis* (35)**
***Epilobium montanum* (35)**,** *Heracleum sphondylium* (35)**
Rumex obtusifolius (32)**, *Deschampsia flexuosa* (31)**
***Quercus petraea* (31)**,** *Chrysosplenium oppositifolium* (28)**
Oxalis acetosella (28)**, *Galium aparine* (27)
Anthriscus sylvestris (27), *Hyacinthoides non-scripta* (26)
Cardamine flexuosa (25)**, *Holcus lanatus* (24)**
Deschampsia caespitosa (24)**, *Lamiastrum galeobdolon* (22)**
Athyrium filix-femina (22)**, *Sorbus aucuparia* (21)
Galium saxatile (20)**, *Agrostis gigantea* (20)**

Woods on Bunter Sandstone and eastern Coal Measures

Betula pendula (34)**
Pteridium aquilinum (32)**
Chamerion angustifolium (30)
Salix cinerea (26)**
Agrostis capillaris (25)
Digitalis purpurea (24)**
Teucrium scorodonia (23)**
Juncus effusus (21)**
Betula pubescens (20)
Arrhenatherum elatius (20)

5



② **Ancient woods on western Coal Measures**

Deschampsia flexuosa **(51)****
Athyrium filix-femina **(44)****
Cardamine flexuosa **(44)****
Oxalis acetosella **(44)****
Ilex aquifolium (43)**
Chrysosplenium oppositifolium **(41)****
Lamiastrum galeobdolon **(41)****
Luzula sylvatica **(36)****
Ranunculus ficaria (33)**
Luzula pilosa **(36)****
Anemone nemorosa **(31)****
Holcus mollis (29)**
Hyacinthoides non-scripta (29)**
Lonicera periclymenum (27)**
Stellaria holostea **(26)****
Quercus petraea (26)**
Milium effusum (24)**
Epilobium montanum (22)**
Digitalis purpurea (22)**
Veronica montana **(22)****
Sorbus aucuparia (21)**

① **Recent woods on western Coal Measures**

Agrostis gigantea **(32)****
Quercus petraea (24)**
Vicia sepium **(23)****
Ranunculus repens (20)**

Ancient and recent Bunter Sandstone and ancient eastern Coal Measures

Betula pendula (25)
Chamerion angustifolium (24)
Pteridium aquilinum (24)
Juncus effusus (18)

③ **Ancient woods on Bunter Sandstone and eastern Coal Measures**

Pteridium aquilinum (23)**
Betula pendula (18)**
Teucrium scorodonia (14)**
Ceratocapnos claviculata (12)**

④ **Recent woods on eastern Coal Measures**

Tussilago farfara (18)**
Lamium album (15)**
Senecio jacobaea (15)**
Alopecurus pratensis **(11)****

⑤ **Recent woods on Bunter Sandstone**

Holcus mollis (24)
Stellaria media (22)**
Salix cinerea (20)**
Digitalis purpurea (19)

⑥ **Recent woods on Magnesian Limestone and Permian Marl**

Crataegus monogyna (25)**
Rosa canina (21)**
Plantago lanceolata (18)**
Plantago major (15)

⑦ **Ancient woods on Magnesian Limestone and Permian Marl**

Mercurialis perennis (52)**
Geum urbanum (47)**
Brachypodium sylvaticum (43)**
Stachys sylvatica (38)**
Geranium robertianum (33)**
Corylus avellana (33)**
Tamus communis (32)**
Ulmus glabra (32)**
Arum maculatum (31)**
Acer campestris (30)**
Fraxinus excelsior (29)**
Bromus ramosa (27)**
Glechoma hederacea (27)**
Arctium minor (27)**
Cornus sanguinea (26)
Rumex sanguinea (26)**
Hedera helix (25)**
Ligustrum vulgare (24)**
Viola riviniana (24)**
Alliaria petiolata (24)**
Sambucus nigra (23)
Prunus avium **(23)****
Euonymus europaea **(23)****
Rubus fruticosus (21)
Hyacinthoides non-scripta (21)
Campanula latifolia **(21)****
Hypericum humifusum **(20)****
Urtica dioica (20)

Table 3.23. Two-way indicator table showing the species indicator power for the site clustering hierarchy. The first valu given before the slash refers to the sum of the abundance values and the value after the slash represents the number of site in that group that the species is present. Bold numbers represent the main data set structure (shown in Figure 18).

Species	INDVAL (%)	1	2	3	4	5	6	7
Recent woods on western Coal Measures (1)								
Agrostis gigantea	32.3 **	**42./10**	6./ 2		3./ 1			
Vicia sepium	23.3 **	**26./10**	7./ 5	5./ 2	9./ 3		2./ 2	5./ 3
Woods on Coal Measures (1+2)								
Ranunculus ficaria	35.7 **	**34./12**	**93./23**	1./ 1	36./ 9		12./ 2	77./16
Epilobium montanum	35.4 **	**35./11**	**43./17**	24./ 8	10./ 4		3./ 1	10./ 4
Quercus petraea x *robur* (*Q.* x *rosacea*)	30.5 **	**67./18**	**59./10**	47./11	27./ 6	26./ 6	8./ 2	17./ 5
Galium saxatile	20.2 **	**27./ 6**	**28./11**	19./ 5	21./ 7	6./ 2		
Equisetum sylvaticum	19.8 **	**22./ 5**	**15./ 8**	6./ 1	3./ 1			
Cardamine amara	3.6	**2./ 1**	**3./ 2**	4./ 1				
Ancient woods on western Coal Measures (2)								
Deschampsia flexuosa	50.7 **	12./ 2	**128./27**	92./20	42./10	18./ 5		7./ 2
Athyrium filix-femina	43.7 **		**41./17**	15./ 4	11./ 4			1./ 1
Cardamine flexuosa	43.6 **	3./ 1	**53./19**	15./ 5	5./ 2		3./ 1	11./ 4
Oxalis acetosella	43.5 **	7./ 3	**75./20**	22./ 9	15./ 5			21./ 6
Chrysosplenium oppositifolium	41.2 **	5./ 2	**43./15**		3./ 1			1./ 1
Lamiastrum galeobdolon	38.8 **	4./ 1	**84./21**	15./ 4	10./ 2		8./ 2	59./15
Luzula sylvatica	36.3 **		**33./12**	4./ 2				1./ 1
Luzula pilosa	32.3 **		**18./10**					
Anemone nemorosa	31.3 **	6./ 2	**60./18**	2./ 2	21./ 5			52./14
Stellaria holostea	26.2 **	2./ 1	**62./15**	31./13	31./10			39./10
Veronica montana	21.5 **	1./ 1	**27./11**	3./ 3	6./ 2			19./ 7
Blechnum spicant	18.7 **		**11./ 6**	1./ 1				
Vaccinium myrtillus	18.2 **		**17./ 6**	3./ 1				
Carex remota	15.6 **	1./ 1	**15./10**	18./ 6	11./ 3			6./ 2
Lysimachia nemorum	12.7 **		**16./ 8**		9./ 3			19./ 6
Valeriana officinalis	12.2 **		**13./ 6**	2./ 2	3./ 1		1./ 1	7./ 3
Sorbus aria agg.	10.1		**12./ 8**	11./ 5	5./ 3	5./ 2	2./ 2	6./ 2
Epipactis helleborine	10.1 **		**4./ 4**	2./ 2	1./ 1			
Calluna vulgaris	10.0		**17./ 5**	5./ 3	9./ 4	4./ 2		
Festuca ovina agg.	9.4		**24./ 7**		8./ 2	12./ 3	15./ 4	
Poa nemoralis	9.3		**11./ 5**			3./ 1		6./ 2
Myosotis sylvatica	8.9 **		**7./ 5**				3./ 1	4./ 2
Melampyrum pratense	8.1		**15./ 3**	5./ 2	3./ 1			
Cardamine pratensis	6.9		**5./ 3**					3./ 1
Carex laevigata	6.5		**2./ 2**					
Hypericum tetrapterum	4.6		**5./ 3**		5./ 3		3./ 1	1./ 1
Rosa arvensis	4.4		**6./ 4**		9./ 3		5./ 2	5./ 1
Geum rivale	4.0		**4./ 2**	1./ 1				3./ 1
Crataegus laevigata	3.2		**1./ 1**					
Adoxa moschatellina	2.4		**1./ 1**	1./ 1				
Woods on sandstones (1+2+3+4+5)								
Holcus mollis	56.9 **	**74./14**	**168./28**	**186./30**	**109./21**	**97./18**	5./ 2	48./10
Digitalis purpurea	35.0 **	**2./ 2**	**60./21**	**74./19**	**57./16**	**44./12**	1./ 1	22./ 8
Juncus effusus	28.9 **	**2./ 2**	**43./16**	**56./18**	**55./18**	**21./ 6**	4./ 2	17./ 5
Teucrium scorodonia	28.4 **	**4./ 2**	**31./12**	**63./17**	**49./15**	**30./ 6**	4./ 1	5./ 2

Species	IndVal (%)	1	2	3	4	5	6	7
Ancient woods on Bunter Sandstone and eastern Coal Measures (3)								
Carex pendula	2.6			**1./ 1**				
Ribes nigrum	2.6			**1./ 1**				
Stellaria nemorum	2.6			**3./ 1**				
Woods on Bunter Sandstone and eastern Caol Measures (3+4)								
Galeopsis tetrahit agg.	14.1 **	9./ 3	6./ 2	**27./11**	**26./ 9**			15./ 5
Dryopteris affinis	3.6			**1./ 1**	**4./ 2**			
Salix aurita	2.4			**1./ 1**	**3./ 1**			
Recent woods on Bunter Sandstone and eastern Coal Measures (4)								
Alopecurus pratensis	13.3 **	13./ 4	3./ 3	4./ 4	**40./11**			3./ 1
Vicia cracca	11.9		1./ 1	10./ 4	**41./12**		13./ 4	19./ 7
Phalaris arundinacea	11.1 **		2./ 2	3./ 1	**31./ 9**		8./ 2	7./ 3
Luzula campestris	7.2		7./ 4		**18./ 6**			5./ 2
Lotus pedunculatus	6.5		1./ 1		**15./ 4**			4./ 2
Cynosurus cristatus	4.9		2./ 2		**17./ 4**	3./ 1	1./ 1	4./ 1
Molinia caerulea	4.4				**6./ 2**			
Nardus stricta	4.4				**4./ 2**			
Stellaria graminea	4.2	2./ 2	1./ 1	1./ 1	**13./ 4**		6./ 2	
Dryopteris carthusiana	2.2				**1./ 1**			
Polypodium vulgare agg.	2.2				**5./ 1**			
Viola palustris	2.2				**4./ 1**			
Woods on eastern sandstones (3+4+5)								
Cytisus scoparius	16.3 **	5./ 2	4./ 1	**25./10**	**24./ 8**	**16./ 5**		3./ 1
Ceratocapnos claviculata	14.9 **		3./ 1	**44./ 9**	**30./ 6**	**9./ 2**		
Conium maculatum	13.7 **			**13./ 4**	**40./11**	**18./ 4**	7./ 3	8./ 3
Typha latifolia	12.1 **	6./ 1		**16./ 4**	**47./12**	**12./ 2**	9./ 3	4./ 1
Recent woods on Bunter Sandstone (5)								
Tilia cordata x *platyphyllos* (*T.* x *vulgaris*)	17.1 **	5./ 2	3./ 2	8./ 4	10./ 5	**27./ 8**	10./ 3	44./11
Rumex acetosa	14.8 **	3./ 1	13./ 7	3./ 1	30./ 7	**25./ 6**	3./ 1	
Rhododendron ponticum	13.3 **		8./ 4	15./ 4	21./ 7	**28./ 5**		7./ 2
Leontodon autumnalis	11.4	6./ 1		16./ 5	37./ 8	**22./ 6**	18./ 4	9./ 3
Carpinus betulus	11.2		2./ 2	1./ 1		**8./ 4**	4./ 2	5./ 1
Hordelymus europaeus	10.0 **		1./ 1	1./ 1		**9./ 3**		5./ 2
Salix viminalis	8.9			7./ 2	18./ 6	**14./ 4**	2./ 1	14./ 3
Hypericum hirsutum	8.3 **					**6./ 2**		
Tilia cordata	5.3		1./ 1	6./ 2	5./ 2	**8./ 2**		3./ 3
Viola reichenbachiana	3.1				1./ 1	**6./ 1**		6./ 2
All habitats								
Rubus fruticosus agg.	93.4	**95./25**	**162./30**	**189./35**	**210./42**	**97./22**	**145./29**	**39./42**
Crataegus monogyna	87.1	**88./20**	**92./29**	**118./32**	**178./38**	**77./19**	**147./31**	**12./41**
Urtica dioica	83.4	**85./21**	**116./28**	**142./35**	**154./35**	**72./18**	**107./25**	**85./39**
Sambucus nigra	78.4	**45./15**	**80./26**	**106./30**	**127./33**	**60./19**	**101./26**	**70./40**
Fraxinus excelsior	71.8	**67./15**	**75./25**	**63./22**	**113./31**	**36./13**	**109./27**	**07./40**
Heracleum sphondylium	68.1	**91./21**	**97./27**	**77./25**	**108./30**	**25./ 9**	**68./21**	**16./31**

Species	IndVal (%)	1	2	3	4	5	6	7
Chamerion angustifolium	66.4	71./19	96./25	149./33	127./31	59./15	57./15	82./22
Galium aparine	61.0	79./19	91./24	56./19	124./28	31./ 9	71./17	59./31
Hedera helix	57.3	29./ 8	87./23	107./21	69./18	62./12	113./20	14./36
Pteridium aquilinum	56.4	45./ 9	94./23	197./32	137./25	78./14	20./ 6	18./27
Hyacinthoides non-scripta	53.9	34./11	138./26	112./27	97./22	28./ 6	18./ 7	41./31
Betula pendula	53.5	48./15	73./21	130./32	119./28	66./13	34./10	36./10
Quercus robur	51.5	19./ 6	55./14	65./14	121./28	53./14	52./18	19./30
Corylus avellana	51.0	29./ 8	69./24	96./30	50./14	8./ 3	26./ 8	47./36
Anthriscus sylvestris	48.6	75./17	63./21	22./10	78./21	18./ 6	67./14	09./28
Ranunculus repens	46.5	73./21	87./23	61./17	87./22	21./ 6	27./ 6	67./17
Ilex aquifolium	45.6	23./12	108./30	47./19	33./13	23./10	16./ 8	48./18
Rumex obtusifolius	45.6	55./15	73./23	44./15	74./23	23./ 6	27./10	54./18
Stachys sylvatica	42.7	12./ 4	50./16	47./15	60./20	1./ 1	33./11	41./36
Agrostis capillaris	40.3	53./11	51./13	99./21	109./21	58./13	41./ 9	30./ 9
Sorbus aucuparia	40.3	20./ 8	59./20	60./23	51./16	23./10	6./ 4	35./16
Arrhenatherum elatius	38.2	12./ 4	15./ 6	48./13	111./22	48./11	67./16	64./20
Holcus lanatus	36.9	71./14	68./14	43./13	96./23	28./ 7	29./ 7	32./11
Dryopteris filix-mas	35.3	14./ 6	60./17	51./16	39./12	21./ 7	16./ 5	74./22
Salix caprea	34.4	24./ 8	31./13	43./14	55./18	12./ 4	46./12	35./14
Milium effusum	32.4		74./19	74./22	46./15	3./ 1	14./ 4	81./17
Senecio jacobaea	32.4	20./ 9	15./ 9	29./13	83./22	18./ 5	37./13	18./ 7
Stellaria media	32.0	5./ 2	31./13	32./13	61./16	54./12	35./ 9	43./12
Plantago major	31.5	6./ 2	19./ 9	32./12	57./18	17./ 4	51./14	48./17
Plantago lanceolata	31.1	30./10	17./ 7	17./ 6	87./22	15./ 5	63./15	36./10
Betula pubescens	30.7	4./ 2	53./13	70./18	64./13	53./11	21./ 8	29./ 9
Dryopteris dilatata	30.7	20./ 6	56./17	38./13	50./14	13./ 4	14./ 4	48./16
Prunus spinosa	30.7	51./14	21./ 9	22./ 6	77./22	8./ 2	6./ 3	68./18
Deschampsia caespitosa	29.1	26./ 9	70./18	40./13	58./15	3./ 1	7./ 2	44./12
Lonicera periclymenum	29.1	5./ 2	65./21	48./14	37./11	16./ 5	11./ 3	42./14
Epilobium hirsutum	27.0	19./ 5	21./12	41./12	55./16	4./ 1	28./ 7	37./12
Salix cinerea	27.0	3./ 1	7./ 3	42./13	85./20	46./11	27./ 7	32./10
Festuca gigantea	26.1	18./ 6	49./14	32./11	20./ 5		18./ 7	67./20
Quercus petraea	26.1		95./18	82./15	35./ 8	38./12	11./ 3	26./ 7
Viola riviniana	25.7	11./ 4	27./13	17./ 6	33./10	15./ 4	9./ 3	94./22
Silene dioica	24.9	1./ 1	30./10	38./11	56./13	11./ 4	11./ 3	67./18
Tussilago farfara	24.5	10./ 5	15./ 9	12./ 6	69./21	3./ 1	24./10	22./ 7
Alnus glutinosa	24.1	19./ 4	31./17	48./16	22./ 9		7./ 3	25./ 9
Artemisia vulgaris	24.1	3./ 1	1./ 1	16./10	61./18	20./ 7	36./11	22./10
Rosa arvensis x *canina (R.* x *verticillacantha)*	23.2	11./ 4	39./14	35./12	25./ 7	3./ 1	21./ 5	48./13
Rubus idaeus	22.0	21./ 6	26./10	13./ 6	23./ 6	3./ 1	26./ 8	61./16
Angelica sylvestris	21.6	16./ 6	24./10	34./14	40./12		7./ 2	23./ 8
Centaurea nigra	20.3	7./ 4	5./ 3	10./ 4	69./16		46./10	42./12
Solanum dulcamara	19.9	6./ 2	20./10	26./ 9	33./12	4./ 2	10./ 6	26./ 7
Salix fragilis	19.5		14./ 8	32./10	45./13	7./ 2	19./ 6	34./ 8
Lamium album	18.7	2./ 2	3./ 1	5./ 2	58./19	19./ 6	24./ 8	24./ 7
Taxus baccata	18.3		6./ 5	17./ 7	22./ 9	12./ 4	18./ 5	37./14
Filipendula ulmaria	17.8	4./ 2	19./ 8	13./ 7	54./13		4./ 2	40./11
Rumex acetosella	17.4	26./10	11./ 7	13./ 5	40./11	6./ 2	7./ 3	9./ 4
Festuca rubra agg.	17.0	32./ 8	12./ 6	4./ 1	67./13	7./ 2	33./ 7	15./ 4

Species	IndVal (%)	1	2	3	4	5	6	7
Achillea millefolium	16.2	**8./ 5**	**6./ 3**	**4./ 2**	**41./10**	**22./ 6**	**42./11**	**6./ 2**
Ulex europaeus	15.4	**19./ 5**	**6./ 4**	**26./ 8**	**17./ 8**	**20./ 7**	**9./ 2**	**5./ 3**
Hypochaeris radicata	14.5	**15./ 4**	**20./ 8**	**6./ 3**	**45./13**	**7./ 2**	**23./ 5**	
Lotus corniculatus	12.5	**6./ 1**	**2./ 2**	**13./ 4**	**47./11**	**16./ 4**	**26./ 6**	**4./ 2**
Prunella vulgaris	12.5	**3./ 1**	**2./ 2**	**5./ 3**	**25./ 6**	**18./ 6**	**16./ 4**	**27./ 8**
Calystegia sepium	12.0	**13./ 3**	**15./ 7**	**5./ 4**	**23./ 7**	**7./ 3**	**8./ 4**	**1./ 1**
Equisetum arvense	12.0		**2./ 2**	**10./ 4**	**30./10**	**6./ 2**	**21./ 7**	**10./ 4**
Malus sylvestris	10.4	**3./ 1**	**3./ 3**	**9./ 5**	**12./ 6**	**7./ 3**	**5./ 3**	**7./ 4**
Lathyrus pratensis	9.5	**5./ 2**	**2./ 2**	**5./ 3**	**34./ 9**		**15./ 5**	**6./ 2**
Anthoxanthum odoratum	8.3	**6./ 3**	**7./ 5**	**1./ 1**	**29./ 7**	**4./ 1**	**3./ 1**	**7./ 2**
Iris pseudacorus	7.9		**3./ 3**	**3./ 1**	**20./ 8**	**3./ 1**	**5./ 3**	**10./ 3**
Potentilla erecta	7.9	**7./ 2**	**6./ 4**	**1./ 1**	**20./ 6**	**10./ 3**	**7./ 2**	**3./ 1**
Scrophularia auriculata	4.6			**3./ 1**	**13./ 4**	**3./ 1**	**3./ 1**	**12./ 4**
Caltha palustris	2.5		**3./ 1**		**6./ 2**		**1./ 1**	**4./ 2**
Recent woods on Magnesian Limestone and Permian Marl (5)								
Clematis vitalba	9.6						**20./ 4**	10./ 3
Galium verum	9.5 **				8./ 3		**14./ 4**	3./ 1
Agropyron caninum	7.4				6./ 1		**14./ 3**	3./ 1
Convolvulus arvensis	6.7				11./ 5	7./ 2	**14./ 4**	
Dipsacus fullonum	6.5				9./ 4	6./ 1	**12./ 4**	1./ 1
Origanum vulgare	6.1				6./ 2		**12./ 3**	7./ 2
Campanula rotundifolia	5.5			3./ 1	10./ 3	4./ 2	**13./ 3**	
ypericum montanum	4.2		6./ 2	3./ 1			**8./ 2**	
Campanula trachelium	3.2						**1./ 1**	
Carex acutiformis	2.2				4./ 1		**3./ 1**	
athyrus montana	1.9	3./ 1		1./ 1			**3./ 1**	
'oods on Permian Marl and Magnesian Limestone (5+6)								
ercurialis perennis	65.6 **	11./ 3	76./19	24./ 8	46./11	3./ 1	**90./16**	**73./44**
rachypodium sylvaticum	55.9 **		12./ 6	14./ 4	32./10		**59./13**	**48./35**
eum urbanum	54.9 **		26./ 8	11./ 4	46./11	3./ 1	**43./13**	**57./37**
lmus glabra	53.3 **	9./ 3	56./18	49./17	41./15	30./ 9	**68./18**	**55./38**
rum maculatum	47.3 **	6./ 2	15./ 8	5./ 3	21./ 9	3./ 1	**43./13**	**93./29**
amus communis	46.2 **		2./ 2	4./ 2	17./ 7		**30./12**	**66./27**
osa canina agg.	44.4 **	16./ 7	3./ 1	43./14	93./25	28./ 8	**75./21**	**00./29**
rctium minus	44.3 **	3./ 1	4./ 2	27./11	50./16	25./ 9	**42./15**	**07./30**
lechoma hederacea	42.8 **		3./ 3	51./12	89./20	40./ 8	**57./14**	**47./32**
cer campestre	39.7 **	1./ 1	7./ 7	25./ 9	29./13	7./ 3	**32./10**	**91./28**
lliaria petiolata	38.9 **		5./ 3	4./ 2	35./11	22./ 6	**41./12**	**99./24**
eranium robertianum	38.4 **	7./ 2	56./17	26./ 7	19./ 5	4./ 2	**30./ 9**	**24./30**
apsana communis	36.0 **	5./ 3	7./ 5	16./ 8	14./ 6	4./ 2	**28./14**	**59./20**
romus ramosa	33.5 **		21./ 7	26./ 8	15./ 5		**28./ 8**	**83./24**
igustrum vulgare	30.2 **		1./ 1	9./ 4	8./ 4	13./ 4	**17./ 6**	**64./21**
ornus sanguinea	29.6 **			6./ 2	5./ 1		**18./ 6**	**57./18**
umex sanguineus	29.3 **		22./11	19./ 5	58./14	3./ 1	**21./ 8**	**87./24**
iola odorata	24.4 **	1./ 1		7./ 2	4./ 1		**31./ 9**	**41./11**
ypericum pulchrum	21.7 **			25./ 8	21./ 6	25./ 6	**49./11**	**34./12**

Species	IndVal (%)	1	2	3	4	5	6	7
Bryonia dioica	16.8 **			7./ 3	12./ 8	2./ 2	**12./ 6**	**26./10**
Sanicula europaea	16.5 **		4./ 2		8./ 3		**11./ 4**	**35./10**
Moehringia trinervia	16.4 **		2./ 2	14./ 5	30./ 8	1./ 1	**15./ 5**	**44./12**
Scrophularia nodosa	16.0 **	3./ 1	8./ 4	5./ 5	19./ 5	1./ 1	**7./ 3**	**35./14**
Primula veris	15.2 **				10./ 3		**21./ 7**	**15./ 6**
Ulmus procera	15.1 **			8./ 3	16./ 3	10./ 3	**17./ 5**	**34./10**
Symphoricarpos albus	14.6 **	7./ 2	1./ 1	1./ 1	20./ 5	10./ 3	**18./ 4**	**50./10**
Circaea lutetiana	14.4 **	2./ 1	16./ 4	13./ 4	31./ 7		**11./ 3**	**51./13**
Potentilla reptans	14.4	2./ 1	9./ 5	4./ 1	28./ 6	18./ 5	**32./ 9**	**29./ 7**
Veronica chamaedrys	14.3 **		4./ 2	1./ 1	19./ 5	4./ 2	**15./ 5**	**30./ 9**
Rubus caesius	12.6 **				3./ 1		**11./ 3**	**23./ 7**
Sisymbrium officinale	11.8			1./ 1	21./ 8	9./ 4	**25./ 7**	**17./ 5**
Torilis japonica	10.7			2./ 2	27./ 7		**22./ 6**	**16./ 5**
Impatiens glandulifera	9.0		12./ 4	13./ 4	4./ 1		**10./ 5**	**19./ 5**
Bromus erectus	8.9 **			8./ 2	8./ 1		**20./ 3**	**19./ 5**
egopodium podagraria	8.5	6./ 3	5./ 2	3./ 1	13./ 4		**10./ 2**	**22./ 7**
Fragaria vesca	7.9			4./ 2	7./ 3		**16./ 3**	**14./ 4**
Galium cruciata	6.9		1./ 1		17./ 4		**6./ 2**	**19./ 5**
ypericum perforatum	5.2				9./ 3		**4./ 1**	**12./ 4**
Calamagrostis canescens	5.0	2./ 1					**4./ 1**	**12./ 3**
etasites hybridus	4.9		1./ 1	1./ 1	8./ 2		**5./ 2**	**8./ 3**
uccisa pratensis	4.2	2./ 1	4./ 2		4./ 1		**9./ 2**	**9./ 2**
actylorhiza fuchsii	4.2		3./ 1		5./ 2		**5./ 1**	**9./ 3**
quilegia vulgaris	3.6					1./ 1	**1./ 1**	**4./ 2**
allopia japonica	3.6		3./ 1		1./ 1		**9./ 2**	**8./ 1**
impinella major	3.2				5./ 1		**3./ 1**	**7./ 2**
tachys officinalis	3.1		1./ 1		4./ 1		**3./ 1**	**6./ 2**
otentilla sterilis	3.1		2./ 2		3./ 1		**1./ 1**	**7./ 2**
pipactis phyllanthes	2.6						**1./ 1**	**1./ 1**
runus padus	2.4		1./ 1				**1./ 1**	**3./ 1**
onvallaria majalis	2.1		3./ 1				**3./ 1**	**3./ 1**
ncient woods on Permian Marl and Magnesian Limestone (6)								
runus avium	22.8 **	6./ 2	25./14	11./ 7	19./ 7	1./ 1	6./ 3	**60./19**
uonymus europaeus	22.6 **			1./ 1	4./ 2		7./ 3	**34./14**
ampanula latifolia	20.8 **				4./ 2		3./ 1	**37./11**
ypericum humifusum	20.4 **				7./ 3		6./ 2	**41./12**
ibes uva-crispa	19.0 **	4./ 1	1./ 1	5./ 3	1./ 1		4./ 1	**32./12**
iburnum opulus	18.2 **	1./ 1	8./ 5	20./11	6./ 4		5./ 3	**40./14**
arex sylvatica	18.2 **	1./ 1	7./ 4	2./ 2	11./ 4		4./ 1	**41./12**
llium ursinum	16.0 **	11./ 3	38./11	1./ 1	16./ 4		5./ 1	**68./13**
elica uniflora	15.8 **		25./11	29./11	10./ 2		3./ 1	**54./13**
opulus tremula	11.0			9./ 2	15./ 5		3./ 1	**28./ 8**
alium odoratum	11.0 **	1./ 1	14./ 5	4./ 2				**33./ 7**
onopodium majus	10.8	6./ 2	6./ 3		10./ 3			**23./ 8**
rimula vulgaris	10.7 **				1./ 1			**15./ 5**
otentilla anserina	10.3 **			2./ 1	12./ 3	4./ 1	3./ 1	**27./ 8**
juga reptans	10.1	4./ 1	1./ 1		11./ 3		1./ 1	**21./ 7**
ibes rubrum	9.1		5./ 5		2./ 2		1./ 1	**14./ 6**

Species	INDVAL (%)	1	2	3	4	5	6	7
Mycelis muralis	7.4		1./ 1				1./ 1	**11./ 4**
Mentha aquatica	7.3		1./ 1	2./ 1	11./ 3			**15./ 5**
Daphne laureola	6.7							**8./ 3**
Orchis mascula	6.4			1./ 1				**15./ 3**
Listera ovata	6.2				1./ 1			**7./ 3**
Phyllitis scolopendrium	5.9		1./ 1					**6./ 3**
Neottia nidus-avis	4.4							**2./ 2**
Primula veris x *Primula vulgaris*	4.4							**5./ 2**
Equisetum telmateia	4.0			1./ 1				**5./ 2**
Sorbus torminalis	3.8				2./ 2			**6./ 2**
Tilia platyphyllos	3.7		1./ 1					**4./ 2**
Narcissus pseudonarcissus	2.9	1./ 1	1./ 1		4./ 2			**7./ 2**
Gymnocarpium dryopteris	2.2							**2./ 1**
Iris foetidissima	2.2							**2./ 1**
Ophrys insectifera	2.2							**1./ 1**
Platanthera chlorantha	2.2							**1./ 1**
Polygonatum multiflorum	2.2							**3./ 1**
Ranunculus auricomus	2.2							**1./ 1**
Calamagrostis epigejos	1.4				4./ 1			**4./ 1**
Number of sites		27	31	38	45	24	31	45

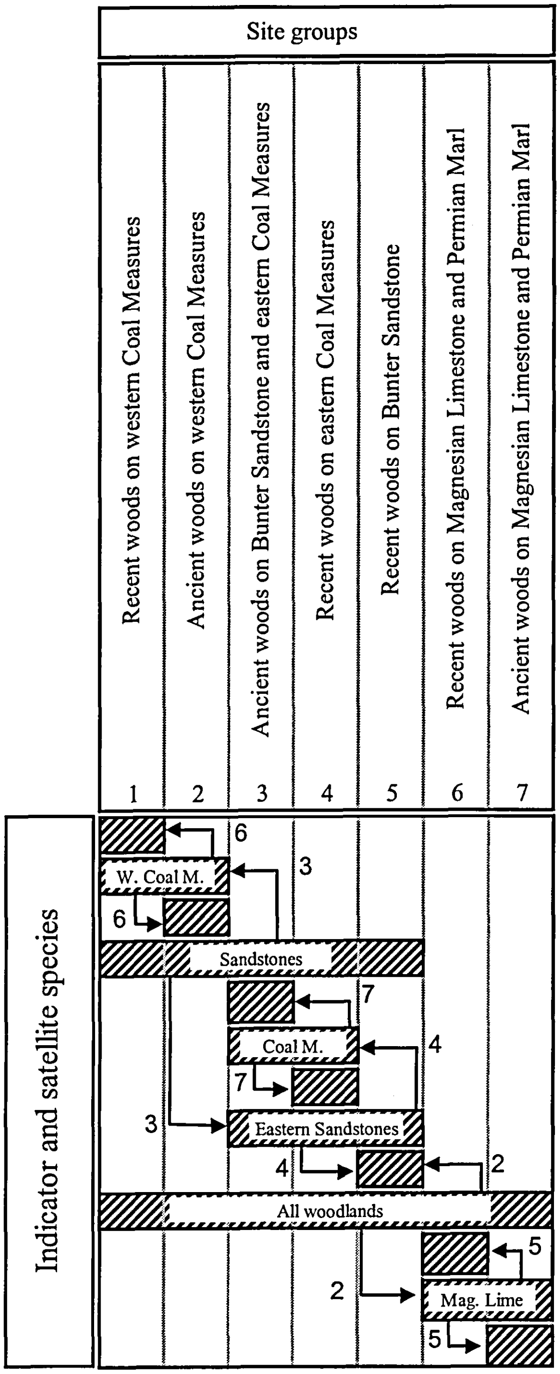


Figure 3.18. The structure of the two-way table of Table 3.22 showing the order in which groups are formed according to the dendogram of Figure 3.16.

3.5.14 Peterken's method

A method similar to that of Peterken (1993) has been used to give lists of species, which are associated with ancient woodland. First all the data were analysed with only the sites on the Millstone Grit omitted since these sites appeared to be very different in terms of their species composition to the rest of the sites as demonstrated previously. The program, INDVAL, was used with all the remaining sites followed by a division for ancient woods and all other sites (secondary and recent) given as the hierarchical structure. Since a statistical test generally has more power with the greater the number of occurrences in the data, the rarer species that may be completely restricted to ancient woodland, and hence the most reliable species, may not appear to be very good indicators when relying on a statistical method alone. Species have therefore been split up into separate tables depending upon their percentage occurrence in ancient woodland. These percentages have been corrected for the unbalanced nature of the data using equation 3.1 and so the first two tables are directly comparable with that of Peterken (1993). These tables are shown in Table 3.24, 3.25 and 3.26 for > 90 %, 75-90% and 60-75% of their occurrence in ancient woodland respectively. Chi-square and Fisher's exact tests have also been carried out on the presence of each species making the tables also comparable with Honnay *et al.*'s (1998) list for western Belgium and Wulf's (1997) list for northwestern Germany. Species that were too rare to be significant with any test and which were found exclusively in ancient woodland are given in Table 3.27. One of the main differences between the method using INDVAL and that using the chi-square test is that INDVAL produced a list of species that are common to both groups. This is shown by Table 3.28 which lists the species that INDVAL deems to be generalists, but which chi-square tests deem to be significantly associated with ancient woodland. However, when a species has been labelled as an indicator of a particular group then INDVAL can be seen to give very similar values to that produced by the chi-square tests. The only difference in terms of a species being significant or not, is for *Orchis mascula* which just makes the less than 5% level using the permutation test in INDVAL but just misses it using the Fisher's exact test.

From Table 3.24 (the equivalent of Peterken's (1993) list for species with a strong affinity for ancient woodland) there appears to be reasonable agreement between his list and the ones reported in this study. Twelve out of the twenty-four species of Table 3.24 are mentioned in Peterken's list for species with a strong affinity for ancient woodland, and a further four of the species are mentioned in Peterken's list for species with a mild affinity for ancient woodland. Only four of the species in this list do not appear in the literature as indicators of ancient woodland. Three more species that Peterken (1993) lists as species with a strong affinity for ancient woodland indicator species appear in Table 3.25 as species with a mild affinity for ancient woodland. Five more of these species are in Table 3.27 for the rare species restricted to ancient woodland. The species that Peterken (1993) lists as those species that have a mild affinity for ancient woodland also have reasonable agreement with the lists presented in this study. Eight out of the twenty-three species listed by Peterken (1993), as having a mild affinity for ancient woodland, occur in the equivalent list of Table 3.25. Four more species can be found listed in Table 3.24 and two more can be found listed in Table 3.27. There are a number of species that have been reported to be indicators of ancient woodland which do not appear in these tables. These are *Hypericum tetrapterum* (seven occurrences in ancient woodland, one in recent), *Ceratocapnos claviculata* (fourteeen occurrences in ancient woodland, five in recent) *Prunus padus* (four occurrences in ancient woodland, one in recent), *Narcissus pseudonarcissus* (five occurrences in ancient woodland, two in recent) and *Hypericum pulchrum* (eight occurrences in ancient woodland, two in recent).

The table showing the rare species restricted to ancient woodland (Table 3.27) appears to include a large number of species which have been previously reported as indicators of ancient woodland. A large number of infrequent species that are not normally associated with woodland were removed from the data set in order to meet the requirements of INDVAL. This also had the affect of reducing the risk of non-woodland species being found to be associated with ancient woodland purely by chance.

The program INDVAL appears to find very few indicator species with an occurrence of less than 75% in ancient woodland. Only seven species fall into this category (Table 3.26). All of these species appear to be common woodland species, only *Hyacinthoides non-scripta* appears in the literature as a species that has been reported as being associated with ancient woodland. The point at which Peterken has stopped looking for indicator species therefore appears to be a sensible place to stop, this is especially important if a different test to INDVAL is being used, such as chi-square tests and Fisher's exact tests. If Chi-square tests are used beyond this threshold then a further 33 species are found to be significantly associated with ancient woodland. However, on examination of this list it includes a number of species that would be expected to colonise sites quickly but may require some shade. These species include *Anthriscus sylvestris*, *Epilobium hirsutum*, *Epilobium montanum, Galium aparine, Juncus effusus, Poa trivialis, Rubus idaeus, Rumex obtusifolius and Stellaria media.* Unfortunately this is the method that has been used by Honnay *et al.* (1998). The method used by Honnay *et al.* (1998) may lead to list of indicator species which may range from species that are completely restricted to ancient woodland to species with a mild affintity for ancient woodland which will behave as poor indicator species. Honnay *et al.* (1998) suggests that greater than twenty-five of the species he lists are required to be confident that a site is ancient. The data from the present study were used to test how many of the species with greater than 90% of their occurrence in ancient woodland are required to be similarly confident that a site is ancient. As mentioned previously we would expect this number to be low based on probability e.g. if two species are found then we should be 99% confident that a site is ancient. However, different numbers of these species may be required for effective analysis on different geology types.

Table 3.24. Species with greater than 90% of their occurrence in ancient woodland. Species indicated with P_1 are also included in the list of Peterken (1993) for species with a strong affinity for ancient woodland and P_2 for a mild affinity with ancient woodland. Species indicated with H are also included in the list of Honnay *et al.* (1998) for Western Belgium. Species indicated with W are also included in the list of Wulf (1997) for northwestern Germany. Species indicated with R are also included in the list of Rose (1999) for southern England); species indicated with HS are included in the list produced by Hermy and Stieperaere (1981) for riverine woodlands in Belgium. Species included in the list produced by the Peak Park (Unpublished) for the Dark Peak are indicated with PP. '(F)' in the chi-square column indicates that Fisher's exact test was used to calculate the probability. For the occurrence in ancient and recent woodland the first value is the sum of the abundance values and the second value refers to the number of sites a species is present.

Separate tests: Species significantly associated with ancient woodland on C (Coal Measures), M (Magnesian Limestone and Permian Marl) and B (Bunter Sandstone).

Species with greater than 90 % Occurrence in ancient woodland (Species with a strong affinity for ancient woodland)	Separate Tests	INDVAL (%)	Occurrence in ancient woodland (n=132)	Occurrence in recent Woodland (n=115)	Chi-square
Anemone nemorosa (P_1,HS,W,R,PP)	C+M	27.7 ***	134 / 39	8 / 3	31.6 ***
Athyrium filix-femina	C	18.5 **	69 / 26	4 / 2	19.7 ***
Blechnum spicant (H,R)	C	5.3 *	12 / 7		6.3 * (F)
Campanula latifolia (P_2,R,PP)	M	9.8 **	42 / 14	3 / 1	10.2 ***
Cardamine flexuosa	C	24.8 ***	97 / 35	6 / 2	29.6 ***
Carex remota (P_1,W,R,PP)	C	14.5 ***	48 / 21	4 / 2	14.6 ***
Carex sylvatica (P_2,H,R,PP)	C+M	16.8 ***	68 / 24	5 / 2	17.6 ***
Epipactis helleborine (R,PP)		6.1 **	10 / 8		7.2 ** (F)
Galium odoratum (P_1,W,R,PP)	M	11.1 ***	56 / 15	1 / 1	11.2 ***
Geum rivale (P_2,R,PP)		4.6 *	14 / 6		5.4 * (F)
Lamiastrum galeobdolon (P_1,HS,R,PP)	C+M	29.6 ***	168 / 43	15 / 4	33.8 ***
Luzula pilosa (P_1,H,R,PP)	C	9.1 **	21 / 12		11.0 ***
Luzula sylvatica (P_1,H,R,PP)	C	11.4 ***	38 / 15		13.9 ***
Lysimachia nemorum (P_1,H,W,R,PP)	C+M	15.2 **	50 / 20		19.0 ***
Melampyrum pratense (P_1,H,R,PP)		4.6 *	23 / 6		5.4 * (F)
Melica uniflora (P_1,H,W,R,PP)	C+B+M	29.5 ***	127 / 40	3 / 1	38.5 ***
Orchis mascula (P_1,R,PP)		3.8 *	17 / 5		4.5 NS (F)
Oxalis acetosella (P_1,H,R,PP)	C+B+M	31.6 ***	145 / 44	7 / 3	37.7 ***
Primula vulgaris (P_1,H,R,PP)		6.1 **	19 / 8		7.2 ** (F)
Ribes rubrum (R)	C	11.6 **	24 / 16	1 / 1	12.1 ***
Ribes uva-crispa	M	13.8 **	45 / 22	8 / 2	15.6 ***
Stellaria holostea (PP)	C+B+M	35.0 ***	165 / 50	12 / 4	42.6 ***
Vaccinium myrtillus (H,R)	C	5.3 *	20 / 7		6.3 * (F)
Veronica montana (P_2,H,W,R,PP)	C+M	20.1 ***	61 / 27	1 / 1	29.0 ***

Table 3.25. Species with 75-90 % of their occurrence in ancient woodland. See Table 3.24 for an explanation of the abbreviations used.

Species with 75-90 % of their Occurrence in ancient woodland (Species with a mild affinity for ancient Woodland)	Separate tests	INDVAL (%)	Occurrence in ancient woodland (n=132)	Occurrence in recent Woodland (n=115)	Chi-square
Ajuga reptans		8.3 *	36 / 13	6 / 3	5.3 *
Allium ursinum (P_2,H,W,R,PP)	M	20.1 ***	129 / 31	19 / 5	18.1 ***
Alnus glutinosa	B	25.6 ***	118 / 47	40 / 13	19.7 ***
Brachypodium sylvaticum (H,PP)	C+B	32.1 ***	218 / 57	66 / 16	25.3 ***
Bromus ramosus (R,PP)	C+B	29.0 ***	161 / 47	32 / 10	25.1 ***
Chrysosplenium oppositifolium (P_2,H,R,PP)	C	13.7 ***	55 / 20	5 / 2	13.6 ***
Circaea lutetiana (H,W,PP)	C+B	17.8 ***	108 / 28	18 / 5	15.1 ***
Conopodium majus (P_2,R,PP)	M	10.4 **	42 / 16	6 / 2	9.8 **
Cornus sanguinea (H)		14.2 **	74 / 24	18 / 6	9.7 **
Corylus avellana (HS)	B+M	64.7 ***	367 / 106	77 / 23	89.6 ***
Deschampsia flexuosa (H)		33.8 ***	256 / 55	52 / 13	28.4 ***
Euonymus europaeus (R)	M	9.8 *	37 / 17	10 / 4	7.0 **
Festuca gigantea (R)	C	31.6 ***	170 / 53	40 / 14	24.3 ***
Geranium robertianum	C	37.1 ***	230 / 62	53 / 16	31.1 ***
Geum urbanum (PP)	C	36.7 ***	245 / 62	60 / 17	29.3 ***
Hypericum hirsutum (P_2,H)	M	10.7 **	56 / 17	10 / 4	7.0 **
Lonicera periclymenum (H,PP)	C+B	35.5 ***	185 / 60	45 / 14	32.9 ***
Mercurialis perennis (P_2,H,W,PP)	C+M	50.2 ***	438 / 86	114 / 22	52.9 ***
Milium effusum (P_1,H,R,PP)	C+B+M	47.3 ***	270 / 69	25 / 10	53.6 ***
Moehringia trinervia (R)	B	16.0 **	87 / 27	21 / 8	9.2 **
Mycelis muralis		6.4 *	17 / 9	1 / 1	5.6 * (F)
Myosotis sylvatica (P_2,R,PP)		5.7 *	17 / 9	3 / 1	5.6 * (F)
Poa nemoralis (H,R)		6.6 *	24 / 10	3 / 1	6.5 *
Populus tremula (R)		7.9 *	45 / 13	10 / 3	5.3 *
Potentilla sterilis (P_1,H,R,PP)		6.4 *	18 / 9	1 / 1	5.6 * (F)
Prunus avium (R,PP)	M	27.6 ***	107 / 45	22 / 9	24.8 ***
Rosa arvensis (H,R,PP)	C	26.5 ***	147 / 47	44 / 13	19.7 ***
Sanicula europaea (H,W,R,PP)		10.4 **	53 / 17	11 / 4	7.0 **
Scrophularia nodosa (P_1)	B+M	18.5 ***	68 / 31	16 / 6	16.1 ***
Silene dioica	C+M	30.7 ***	185 / 52	46 / 13	25.0 ***
Sorbus aria agg. (R)	C+B+M	10.0 *	32 / 18	10 / 5	6.3 *
Valeriana officinalis (P_2)		9.9 **	31 / 15	4 / 2	8.9 **
Viburnum opulus (R,PP)	C+M	23.8 ***	78 / 36	10 / 6	13.1 ***
Viola riviniana (P_2,H)	C+M	31.3 ***	172 / 52	39 / 13	25.0 ***

Table 3.26. Species with 60-75 % of their occurrence in ancient woodland.

Species with 60 - 75 % of their Occurrence in ancient woodland (Common woodland species)	Separate Tests	INDVAL (%)	Occurrence in ancient woodland (n=132)	Occurrence in recent Woodland (n=115)	Chi-square
Dryopteris dilatata		33.6 ***	194 / 60	60 / 19	23.6 ***
Dryopteris filix-mas	B+M	39.2 ***	226 / 68	62 / 22	27.8 ***
Hyacinthoides non-scripta (H,R,PP)	B+M	57.8 ***	468 / 99	121 / 36	47.3 ***
Quercus petraea		29.3 ***	244 / 49	57 / 18	13.2 ***
Ranunculus ficaria	M	30.7 ***	216 / 52	53 / 16	20.0 ***
Rumex sanguineus	C	28.3 ***	166 / 49	45 / 15	18.6 ***
Stachys sylvatica		46.4 ***	282 / 80	75 / 28	32.4 ***

Table 3.27. Species infrequently recorded that were found exclusively in ancient woodland.

Species	INDVAL (%)	Ancient (n=132)	Recent (n=115)
Adoxa moschatellina (P_2,R,PP)	3.0	9 / 4	
Calamagrostis canescens (P_1,R)	1.5	8 / 2	
Cardamine pratensis	3.8	9 / 5	
Carex digitata (PP)	0.8	1 / 1	
Carex laevigata (P_1,R,PP)	1.5	2 / 2	
Carex pendula (P_1,H,R)	0.8	1 / 1	
Crataegus laevigata (R)	1.5	3 / 2	
Daphne laureola (R,PP)	2.3	8 / 3	
Dryopteris aemula (R)	0.8	1 / 1	
Dryopteris carthusiana (R,PP)	0.8	1 / 1	
Equisetum telmateia (H,PP)	3.0	10 / 4	
Gymnocarpium dryopteris	0.8	2 / 1	
Iris foetidissima (R)	0.8	2 / 1	
Listera ovata	3.0	8 / 4	
Melica nutans (PP)	1.5	6 / 2	
Neottia nidus-avis (R)	1.5	2 / 2	
Ophrys insectifera (R)	0.8	1 / 1	
Phyllitis scolopendrium (R,PP)	3.0	7 / 4	
Platanthera chlorantha (P_1,W,R)	0.8	1 / 1	
Polygonatum multiflorum (H,R,PP)	0.8	3 / 1	
Polystichum aculeatum (R,PP)	0.8	3 / 1	
Primula veris x *Primula vulgaris*	1.5	5 / 2	
Ranunculus auricomus (P_2,W,R,PP)	0.8	1 / 1	
Ribes alpinum	0.8	1 / 1	
Ribes nigrum (R)	1.5	2 / 2	
Rubus saxatilis (PP)	0.8	3 / 1	
Salix aurita	1.5	4 / 2	
Sedum telephium (R)	0.8	4 / 1	
Sorbus torminalis (P_1,R,PP)	3.0	8 / 4	
Stellaria nemorum (H,PP)	0.8	3 / 1	
Tilia platyphyllos (PP)	2.3	5 / 3	
Viola palustris (R)	0.8	4 / 1	

Table 3.28 Species which INDVAL indicates are generalists but which the chi-square test suggests are significantly associated with Ancient woodland. All species in the table have a corrected percentage occurrence of between 50-75%.

Species	Separate Tests	INDVAL	Occurrence in ancient woodland (n = 132)	Occurrence in recent woodland (n = 115)	Chi-square
Acer campestre (H,R,PP)	B	general	145 / 53	52 / 20	15.3 ***
Angelica sylvestris		general	108 / 42	48 / 15	12.2 ***
Anthriscus sylvasetris		general	241 / 73	201 / 49	4.0 *
Arum maculatum (H)		general	142 /50	57 / 19	13.9 ***
Deschampsia caespitosa (H)		general	191 / 52	76 / 23	10.9 **
Digitalis purpurea	C	general	191 /58	73 / 23	16.0 ***
Epilobium hirsutum		general	132 / 45	87 / 24	5.3 *
Epilobium montanum (H)		general	96 / 37	42 / 14	9.4 **
Filipendula ulmaria		general	102 / 35	52 / 14	7.9 **
Fraxinus excelsior		general	407 / 104	284 / 75	5.7 *
Galeopsis tetrahit agg.		general	64 / 23	23 / 9	5.0 *
Galium aparine		general	381 / 89	248 / 62	4.7 *
Galium saxatile (H)		general	68 / 24	38 / 10	4.7 *
Glechoma hederacea		general	268 / 63	140 / 32	10.3 **
Hedera helix		general	462 / 94	239 / 49	20.6 ***
Holcus mollis (R)		general	466 / 79	236 / 47	8.9 **
Ilex aquifolium (R)	B	general	228 / 78	73 / 35	20.3 ***
Juncus effusus		general	135 / 46	67 / 22	7.6 **
Lapsana communis		general	94 / 39	41 / 21	4.3 *
Ligustrum vulgare		general	86 / 31	31 / 11	8.4 **
Poa trivialis		general	103 / 32	56 / 13	6.9 **
Pteridium aquilinum (H)		general	481 / 97	223 / 43	32.6 ***
Ranunculus repens		general	275 / 73	162 / 44	7.2 **
Rubus idaeus		general	136 / 42	53 / 16	11.0 ***
Rumex obtusifolius		general	211 / 69	144 / 43	5.5 *
Sambucus nigra		general	412 / 112	293 / 83	5.9 *
Sorbus aucuparia (R)	M	general	195 / 73	68 / 29	22.9 ***
Stellaria media		general	153 / 51	113 / 29	5.1 *
Tamus communis (H,R)	C+B	general	85 / 38	36 / 14	10.2 **
Taxus baccata		general	78 / 34	39 / 13	8.3 **
Teucrium scorodonia (H)	C+B	general	133 / 43	62 / 16	11.8 ***
Ulmus glabra (R)		general	302 / 86	120 / 36	28.2 ***
Urtica dioica		general	508 / 118	375 / 89	6.5 **

3.5.15 The number of species required to be confident that a site is ancient

A test was carried out using a generalised linear model with logistic link function and binomial error term on the data. This included categorical effects for geology and age and also included the interaction for geology and age. This analysis estimates the probability that a site is ancient based on the geology, the number of ancient woodland indicators (the number of species listed in Table 3.24) and the interaction between the number of species and geology. The estimates produced from this analysis are given in Table 3.29 and they have been used to calculate the probabilities displayed in Figure 3.19. The data for the geology types for Magnesian Limestone and Permian Marl have been pooled together and have been simply called 'Mag.Limestone' in the following tables. The tests for the main effects and interaction between them are given in Table 3.30 and separate tests for unequal slopes are given in Table 3.31. The only statistically significant differences that were found between the regression coefficients (steepness of slope) were between the Bunter Sandstone and Coal Measures Series and the Bunter Sandstone and Magnesian Limestone. This indicates that significantly fewer ancient woodland species are required to be confident that a site is ancient on the Bunter Sandstone compared to sites on the Coal Measures Series or on Magnesian Limestone. From Figure 3.19 it appears that maybe only two or three of the species listed in Table 3.24 are required to be confident that a site on the Bunter Sandstone is ancient. Since the slopes for Coal Measures and Magnesian Limestone are not significantly different, then greater than five species may be required to be confident that a site is ancient on these two geology types.

Table 3.29. Estimates produced from the generalised linear model with logistic link function and binomial error distribution.

Effect	DF	Estimate
Intercept	236	-1.08 **
Bunter Sandstone	236	-0.91
Coal Measures	236	-0.28
Mag. Limestone	.	0.00
Indicators	236	1.72 ***
Indicators × Bunter Sandstone	236	2.11 *
Indicators × Coal Measures	236	-0.65
Indicators × Mag. Limestone	.	0.00

Table 3.30. Type III F tests for the main effects and interactions used in the GLM.

Effect	NDF	DDF	Type III F
Geology	2	236	1.49
Indicators	1	236	51.18 ***
Indicators × geology	2	236	6.05 **

Table 3.31. Testing for unequal slopes of the estimates produced by the GLM.

Unequal slopes tests	DF	Estimate	t	
Bunter Sandstone – Coal Measures	236	2.76	3.34	***
Bunter Sandstone – Mag. Limestone	236	2.11	2.35	*
Coal Measures – Mag. Limestone	236	-0.65	-1.38	

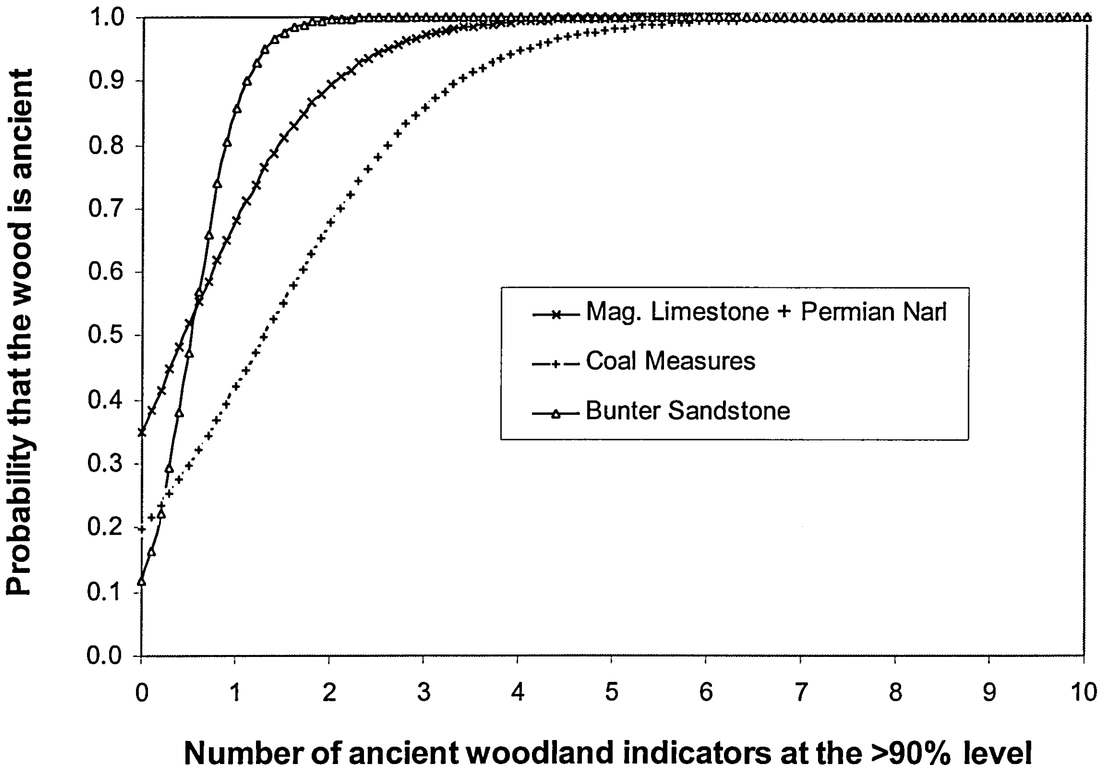


Figure 3.19. The number of species from Table 3.24 that are required to be confident that a site is ancient.

3.5.16 Indicators on the different geology types

The procedure used in the previous section for the complete data set minus the Gritstone woods was repeated separately for the sites on the Coal Measure Series, the Bunter Sandstone and the sites on the Magnesian Limstone plus Permian Marl. The INDVAL program was used to determine the species that are indicators of the ancient sites and determine their significance. The same cut-off values (>90% and 75-90 %) were used after correction for the unbalanced data. Only species that were found to be significant are shown. A table giving the rare species that are restricted to the ancient woodland is also given for each geology type.

3.5.17 Coal Measures Series

The results produced by INDVAL using the equivalent cut-off values to Peterken (1993) are shown in Table 3.32 for species with greater than 90% of their occurrence in ancient woodland and Table 3.33 for species with 75-90% of their occurrence in ancient woodland. The presence of *Impatiens glandulifera* is the most surprising species on the list in Table 3.32. Since, this species is not native and probably colonises sites such as those where garden rubbish has been dumped it should not be considered as an indicator of woodland continuity. It could perhaps be associated with ancient rather than secondary woodland sites and this might be for a number of reasons. This might include recreational use and deliberate release by people into ancient woodlands and perhaps because ancient woodlands are often wetter and this is a requirement for successful establishment. The presence of *Geum urbanum* and *Tamus communis* are also surprising. These two species are regarded by Peterken (1993) as fast-colonising woodland species. However, locally they may be good indicators within the South Yorkshire region. There are also a few surprising species listed in Table 3.33, for species with a mild affinity for ancient woodland. These species include *Calluna vulgaris*, *Deschampsia flexuosa*, *Digitalis purpurea*, *Geranium robertianum* and *Silene dioica,* which are not normally considered as ancient woodland species. *Geranium robertianum* and *Silene dioica* are other species that Peterken (1993) considers to be fast-colonising woodland species. *Deschampsia flexuosa*, *Digitalis purpurea* and *Calluna vulgaris* are species that Peterken regards as shade-bearing species that occur in both ancient woodland, heathland and unimproved grassland. However, it should be noted that *Impatiens glandulifera, Calluna vulgaris* and *Digitalis purpurea*, were not found to be indicator species when all the data were combined. Caution is needed in regarding these species as indicators since there does appear to be a large amount of spatial variation which has not been accounted for by the variables entered into the analysis. It is possible that species such as *Impatiens glandulifera* are an artefact of an unbalanced experimental design. This could arise if most of the woodlands close to urban areas are ancient. This is likely since the ancient sites do have some protection compared to recent sites which are more likely to be developed on. This could therefore give rise to some garden plants having a positive correlation with ancient woodland.

Table 3.32. Species with a strong affinity for ancient woodland on the Coal Measures Series. Species in brackets were not considered to be indicators species in Tables 3.24 or 3.25 and so should be treated with caution.

Species with 90-100% of their Occurrence in ancient woodland	INDVAL (%)	Ancient n=58	Recent n=37
Blechnum spicant	12.1 *	12 / 7	
Bromus ramosus	25.2 **	50 / 16	3 / 1
Cardamine flexuosa	42.2 ***	75 / 26	3 / 1
Geum urbanum	25.7 **	56 / 17	5 / 1
(*Impatiens glandulifera*)	15.5 *	26 / 9	
Luzula pilosa	17.2 *	18 / 10	
Luzula sylvatica	24.1 **	37 / 14	
Lysimachia nemorum	19.0 **	27 / 11	
Melica uniflora	34.5 ***	49 / 20	
Ribes rubrum	12.1 *	7 / 7	
Salix fragilis	26.9 **	51 / 18	5 / 1
Sorbus aria agg.	25.9 ***	27 / 15	
(*Tamus communis*)	15.5 *	18 / 9	
Vaccinium myrtillus	12.1 *	20 / 7	
Veronica montana	28.2 **	39 / 17	1 / 1

Table 3.33. Species with a mild affinity for ancient woodland on the Coal Measures Series. Species in brackets were not considered to be indicators species in Tables 3.24 or 3.25 and so should be treated with caution.

Species with 75-90% of their Occurrence in ancient woodland	INDVAL (%)	Ancient n=58	Recent n=37
Anemone nemorosa	35.2 ***	75 / 23	6 / 2
Athyrium filix-femina	32.6 **	56 / 21	4 / 2
Brachypodium sylvaticum	18.6 *	38 / 13	5 / 1
(*Calluna vulgaris*)	14.6 *	25 / 9	1 / 1
Carex remota	26.7 **	39 / 18	4 / 2
Carex sylvatica	15.9 *	19 / 10	1 / 1
Chrysosplenium oppositifolium	23.6 **	46 / 16	5 / 2
Circaea lutetiana	17.5 **	38 / 11	2 / 1
Deschampsia flexuosa	63.1 ***	211 / 45	31 / 7
(*Digitalis purpurea*)	48.9 ***	106 / 33	11 / 5
Geranium robertianum	36.3 ***	83 / 25	10 / 3
Lamiastrum galeobdolon	42.3 ***	109 / 27	7 / 2
Lonicera periclymenum	43.7 ***	90 / 28	6 / 3
Mercurialis perennis	43.4 ***	121 / 32	21 / 5
Milium effusum	66.5 ***	149 / 41	6 / 3
Oxalis acetosella	50.0 ***	105 / 32	7 / 3
(*Rumex sanguineus*)	34.7 ***	67 / 22	4 / 2
Silene dioica	38.8 **	90 / 26	9 / 3
Stellaria holostea	42.5 ***	92 / 28	8 / 3
Viburnum opulus	20.8 **	20 / 13	1 / 1

Table 3.33. Species infrequently recorded on the Coal Measures Series that were found exclusively in ancient woodland. Species in bold are those that were found to be indicators in Tables 3.24 or 3.25.

Species	INDVAL (%)	Ancient n=58	Recent n=37
Adoxa moschatellina	3.5	2 / 2	
Campanula latifolia	5.2	5 / 3	
Cardamine pratensis	6.9	8 / 4	
Carex laevigata	3.5	2 / 2	
Carpinus betulus	3.5	2 / 2	
Ceratocapnos claviculata	6.9	14 / 4	
Cornus sanguinea	6.9	14 / 4	
Dryopteris carthusiana	3.5	2 / 2	
Epipactis helleborine	10.3	6 / 6	
Fragaria vesca	5.2	9 / 3	
Geum rivale	5.2	5 / 3	
Hordelymus europaeus	3.5	2 / 2	
Hypericum montanum	5.2	9 / 3	
Hypericum tetrapterum	5.2	5 / 3	
Ligustrum vulgare	10.3	12 / 6	
Melampyrum pratense	6.9	16 / 4	
Myosotis sylvatica	8.6	7 / 5	
Petasites hybridus	3.5	2 / 2	
Phalaris arundinacea	5.2	3 / 3	
Phyllitis scolopendrium	3.5	2 / 2	
Poa nemoralis	8.6	11 / 5	
Populus tremula	5.2	13 / 3	
Potentilla anserina	5.2	5 / 3	
Sanicula europaea	8.6	11 / 5	
Teucrium scorodonia	5.2	7 / 3	
Ulmus procera	5.2	7 / 3	
Viola reichenbachiana	3.5	4 / 2	

The species that occurred exclusively in ancient woodland on the Coal Measures Series are listed in Table 3.33. This table contains many of the species that have often been regarded as being very good indicators of ancient woodland. *Adoxa moschatellina*, *Campanula latifolia*, *Carex laevigata*, *Ceratocapnos claviculata*, *Fragaria vesca*, *Geum rivale*, *Hypericum tetrapterum*, *Melampyrum pratense*, *Myosotis sylvatica* and *Viola reichenbachiana a*re all listed by Peterken (1993) as either species with a strong or mild affinity for ancient woodland. However, there are a few species that appear to be in this table by chance, which include *Cardamine pratensis* and *Potentilla anserina*. it should be remembered that this table is biased towards rare woodland species that are restricted to ancient woodland, since the rare species in the data set are mainly those which have been previously noted as ancient woodland species.

3.5.18 Bunter Sandstone

The species that have greater than 90 % of their occurrence in ancient woodland, on the Bunter Sandstone, are listed in Table 3.34 and those with 75-90 % of their occurrence in ancient woodland in Table 3.35, after compensating for the number of recent and ancient sites. Only five species make the table for species with a strong affinity for ancient woodland. A number of species that would normally be considered to be common woodland species occur in these two tables, suggesting that many species have problems colonising these sites. A fairly large number of rare species can also be seen to be very restricted to the ancient woodland as shown in Table 3.36. This list includes species such as *Cardamine flexuosa*, *Athyrium filix-femina* and *Ranunculus ficaria*, which are normally found in damp shady environments. The area of Bunter Sandstone studied occurs to the east of Doncaster in an area that typically receives very little rainfall (less than 600mm per year) and also contains extensive areas of free-draining sandy soils. It is therefore likely that soil moisture content and soil acidity are limiting the colonisation of many species that are typical of both secondary and ancient woodland.

Table 3.34. Species with a strong affinity for ancient woodland on the Bunter Sandstone. Species in brackets were not considered to be indicators species in Tables 3.24 or 3.25 and so should be treated with caution.

Species with 90-100% of their occurrence in ancient woodland	INDVAL (%)	Ancient n = 24	Recent n = 47
Melica uniflora	16.7 *	15 / 4	
Mercurialis perennis	19.4 *	20 / 5	3 / 1
Oxalis acetosella	12.5 *	10 / 3	
Stellaria holostea	30.8 **	25 / 8	4 / 1
(*Tamus communis*)	20.8 **	13 / 5	

Table 3.35. Species with a mild affinity for ancient woodland on the Bunter Sandstone. Species in brackets were not considered to be indicators species in Tables 3.24 or 3.25 and so should be treated with caution.

Species with 75-90% of their occurrence in ancient woodland	INDVAL (%)	Ancient n = 24	Recent n = 47
(*Acer campestre*)	23.4 *	25 / 7	12 / 6
Alnus glutinosa	29.1 **	21 / 8	6 / 4
Brachypodium sylvaticum	19.8 **	19 / 5	2 / 2
(*Ceratocapnos claviculata*)	28.7 **	45 / 9	27 / 5
Circaea lutetiana	14.8 *	20 / 4	5 / 1
Corylus avellana	66.4 **	63 / 18	16 / 5
(*Dryopteris affinis*)	48.3 **	49 / 14	20 / 7
(*Dryopteris dilatata*)	33.2 **	39 / 11	29 / 10
(*Galium saxatile*)	17.1 *	23 / 5	10 / 3
(*Hyacinthoides non-scripta*)	50.1 **	77 / 16	50 / 12
(*Ilex aquifolium*)	39.9 *	27 / 15	30 / 13
Lonicera periclymenum	44.7 **	47 / 14	28 / 8
Milium effusum	43.1 **	40 / 11	5 / 3
Moehringia trinervia	18.1 *	17 / 5	5 / 2
(*Scrophularia auriculata*)	15.9 *	10 / 4	1 / 1
Sorbus aucuparia	52.6 **	49 / 18	41 / 16
(*Teucrium scorodonia*)	33.6 *	49 / 12	47 / 10
(*Vicia cracca*)	21.6 **	13 / 6	4 / 2

Table 3.36. Species infrequently recorded on the Bunter Sandstone that were only found in ancient woodland. Species in bold are also listed in Tables 3.24 or 3.25 as those species that were found to be indicators when all the data are pooled together.

Species	INDVAL (%)	Ancient n = 24	Recent n = 47
Agropyron caninum	4.2	6 / 1	
Athyrium filix-femina	8.3	7 / 2	
Calamagrostis canescens	4.2	4 / 1	
Cardamine flexuosa	4.2	1 / 1	
Carex acutiformis	4.2	4 / 1	
Carex remota	4.2	5 / 1	
Carex sylvatica	4.2	4 / 1	
Conopodium majus	4.2	4 / 1	
Epipactis helleborine	4.2	1 / 1	
Euonymus europaeus	4.2	1 / 1	
Hypericum tetrapterum	4.2	5 / 1	
Lysimachia nemorum	4.2	1 / 1	
Melampyrum pratense	8.3	7 / 2	
Nardus stricta	4.2	1 / 1	
Poa trivialis	4.2	4 / 1	
Ranunculus ficaria	8.3	9 / 2	
Ribes nigrum	4.2	1 / 1	
Ribes rubrum	4.2	1 / 1	
Ribes uva-crispa	4.2	3 / 1	
Rubus caesius	4.2	3 / 1	
Salix aurita	4.2	3 / 1	
Sanicula europaea	4.2	4 / 1	
Sorbus torminalis	4.2	1 / 1	
Stellaria nemorum	4.2	3 / 1	
Viola reichenbachiana	8.3	7 / 2	

3.5.19 Magnesian Limestone and Permian Marl

The species that have greater than 90 % of their occurrence in ancient woodland, on the Magnesian Limestone and Permian Marl, are listed in Table 3.37 and those with 75-90 % of their occurrence in ancient woodland in Table 3.38 after compensating for the number of recent and ancient sites. Out of the nine species listed in Table 3.37 only *Stellaria holostea* is not mentioned by Peterken (1993) as either having a strong or mild affinity for ancient woodland. The good agreement between Peterken's list and the one presented here is perhaps not surprising considering the similarity of the calcareous soils of central Lincolnshire with those on the Magnesian Limestone. However, there is far less agreement with the species listed in Table 3.38 with those mentioned by Peterken (1993). A large number of woodland species, which are not normally considered to be ancient woodland indicators, can be seen in Table 3.38. These species include *Deschampsia caespitosa*, *Digitalis purpurea*, *Dryopteris filix-mas*, *Prunus spinosa*, *Ranunculus ficaria*, *Silene dioica* and *Sorbus aucuparia*.

The rare species that were only found in ancient woodland on the Magnesian Limestone and Permian Marl are listed in Table 3.39. This list includes a number of rare species, which have been referred to as indicators of ancient woodland such as *Carex remota, Chrysosplenium oppositifolium*, *Daphne laureola*, *Equisetum telmateia*, *Luzula sylvatica*, *Platanthera chlorantha*, *Tilia platyphyllos* and *Viola reichenbachiana*. However, this list also includes a number of fairly common species of unimproved grassland such as *Alopecurus pratensis*, *Lotus pedunculatus* and *Luzula campestris*, suggesting that not all of these species make reliable indicator species in woodlands.

Table 3.37. Species with a strong affinity for ancient woodland on the Magnesian Limestone and Permian Marl. Species in brackets were not considered to be indicators species in Tables 3.24 or 3.25 and so should be treated with caution.

Species with 90-100% of their occurrence in ancient woodland	INDVAL (%)	Ancient n = 45	Recent n = 31
Allium ursinum	26.1 **	68 / 13	5 / 1
Anemone nemorosa	31.1 ***	52 / 14	
Conopodium majus	17.8 *	23 / 8	
Galium odoratum	15.6 *	33 / 7	
Lysimachia nemorum	13.3 *	19 / 6	
Melica uniflora	26.7 **	54 / 13	3 / 1
Oxalis acetosella	13.3 *	21 / 6	
Stellaria holostea	22.2 *	39 / 10	
Veronica montana	15.6 *	19 / 7	

Table 3.38. Species with a mild affinity for ancient woodland on the Magnesian Limestone and Permian Marl. Species in brackets were not considered to be indicators species in Tables 3.24 or 3.25 and so should be treated with caution.

Species with 75-90% of their occurrence in ancient woodland	INDVAL (%)	Ancient n=45	Recent n=31
Campanula latifolia	21.9 *	37 / 11	3 / 1
Carex sylvatica	23.4 *	41 / 12	4 / 1
Corylus avellana	63.7 ***	147 / 36	26 / 8
(*Deschampsia caespitosa*)	21.7 *	44 / 12	7 / 2
(*Digitalis purpurea*)	16.7 *	22 / 8	1 / 1
(*Dryopteris filix-mas*)	37.2 **	74 / 22	16 / 5
Euonymus europaeus	24.0 *	34 / 14	7 / 3
(*Filipendula ulmaria*)	21.4 *	40 / 11	4 / 2
(*Hyacinthoides non-scripta*)	58.1 ***	141 / 31	18 / 7
Hypericum hirsutum	22.0 *	41 / 12	6 / 2
Lamiastrum galeobdolon	27.9 *	59 / 15	8 / 2
Milium effusum	30.2 *	81 / 17	14 / 4
Prunus avium	36.9 ***	60 / 19	6 / 3
(*Prunus spinosa*)	35.5 **	68 / 18	6 / 3
(*Pteridium aquilinum*)	48.2 ***	118 / 27	20 / 6
(*Ranunculus ficaria*)	29.0 *	77 / 16	12 / 2
Ribes uva-crispa	22.6 *	32 / 12	4 / 1
Scrophularia nodosa	24.1 *	35 / 14	7 / 3
Silene dioica	32.3 *	67 / 18	11 / 3
(*Sorbus aucuparia*)	28.5 *	35 / 16	6 / 4
Viburnum opulus	26.3 *	40 / 14	5 / 3
Viola riviniana	42.9 ***	94 / 22	9 / 3

Table 3.39. Species infrequently recorded on the Magnesian Limestone and Permian Marl that were only found in ancient woodland. Species in bold are those that were also listed in Table 3.24 and 3.25.

Species	IndVal (%)	Recent n=45	Ancient n=31
Alopecurus pratensis	2.2	3 / 1	
Athyrium filix-femina	2.2	1 / 1	
Calamagrostis epigejos	2.2	4 / 1	
Cardamine pratensis	2.2	3 / 1	
Carex remota	4.4	6 / 2	
Chrysosplenium oppositifolium	2.2	1 / 1	
Cytisus scoparius	2.2	3 / 1	
Daphne laureola	6.7	8 / 3	
Deschampsia flexuosa	4.4	7 / 2	
Equisetum telmateia	4.4	5 / 2	
Galeopsis tetrahit agg.	11.1	15 / 5	
Geum rivale	2.2	3 / 1	
Gymnocarpium dryopteris	2.2	2 / 1	
Hordelymus europaeus	4.4	5 / 2	
Iris foetidissima	2.2	2 / 1	
Listera ovata	6.7	7 / 3	
Lotus pedunculatus	4.4	4 / 2	
Luzula campestris	4.4	5 / 2	
Luzula sylvatica	2.2	1 / 1	
Mentha aquatica	11.1	15 / 5	
Narcissus pseudonarcissus	4.4	7 / 2	
Neottia nidus-avis	4.4	2 / 2	
Ophrys insectifera	2.2	1 / 1	
Orchis mascula	6.7	15 / 3	
Phyllitis scolopendrium	6.7	6 / 3	
Platanthera chlorantha	2.2	1 / 1	
Poa nemoralis	4.4	6 / 2	
Polygonatum multiflorum	2.2	3 / 1	
Primula veris x *Primula vulgaris*	4.4	5 / 2	
Primula vulgaris	11.1	15 / 5	
Ranunculus auricomus	2.2	1 / 1	
Rhododendron ponticum	4.4	7 / 2	
Sorbus torminalis	4.4	6 / 2	
Tilia cordata	6.7	3 / 3	
Tilia platyphyllos	4.4	4 / 2	
Viola reichenbachiana	4.4	6 / 2	

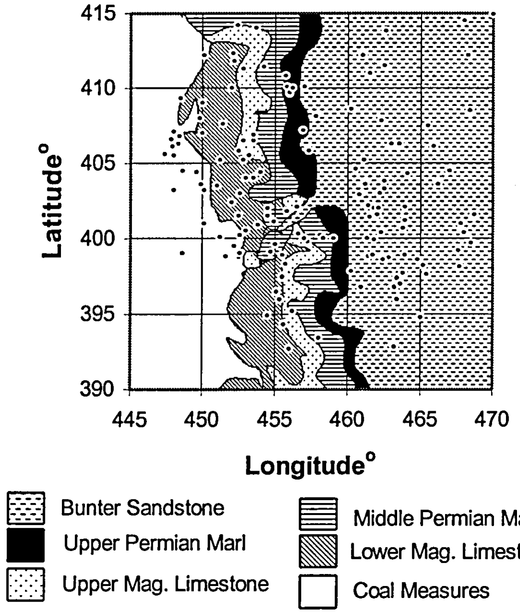
3.5.20 Species distributions in the Doncaster area

The use of semi-variograms and kriging were used to investigate the species distribution in the Doncaster region. There was not sufficient coverage for the whole of the South Yorkshire region to produce reliable distribution maps by the use of kriging. However, a good fit was found with the data from the Doncaster region. The descriptions of semi-variograms for those species, which showed clear increases in their semi-variance with separation distance, are shown in Table 3.40. Altitude is also included in this table. This table shows that a number of the more common woodland species produced semi-variograms with a high R^2 value. Maps of the species distributions and altitude are shown in Figure 3.19, as well as a map showing the site locations and geology. A number of species can be seen to have a greater abundance on the Coal Measures and Bunter Sandstone. These species are: *Holcus mollis*, *Hyacinthoides non-scripta*, *Pteridium aquilinum*, *Milium effusum*, *Ceratocapnos claviculata*, *Sorbus aucuparia*, *Agrostis capillaris*, *Betula* sp., and *Teucrium scorodonia*. A number of species appear to be virtually absent from the Bunter Sandstone. These species are *Bromus ramosus*, *Brachypodium sylvaticum*, *Lamiastrum galeobdolon*, *Mercurialis perennis* and *Arum maculatum*. *Fraxinus excelsior* and *Mercurialis perennis* can be seen to have their greatest abundance on the Magnesian Limestone and Permian Marl.

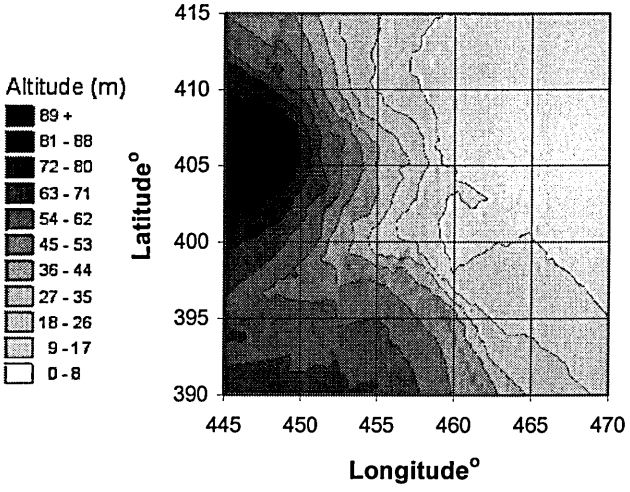
Table 3.40. Desciptions of the semi-variograms used for showing the distributions of species across the Doncaster region by the use of kriging.

Variable	Model	Nugget	Sill	Range (km)	R^2
Altitude	Gaussian	261.0	2994.0	125.4	0.96
Agrostis capillaris	Exponential	4.04	13.09	133.7	0.77
Arum maculatum	Exponential	1.83	3.67	52.6	0.77
Betula sp.	Spherical	3.50	8.84	24.1	0.91
Brachypodium sylvaticum	Spherical	2.10	1.09	17.4	0.90
Bromus ramosa	Exponential	1.75	3.50	90.9	0.67
Ceratocapnos claviculata	Spherical	1.48	2.95	10.8	0.82
Fraxinus excelsior	Exponential	3.39	7.16	51.1	0.87
Holcus mollis	Spherical	3.51	9.84	17.8	0.98
Hyacinthoides non-scripta	Exponential	2.53	6.45	41.1	0.97
Lamiastrum galeobdolon	Exponential	1.44	3.29	88.4	0.71
Mercurialis perennis	Spherical	2.61	13.12	20.9	0.98
Milium effusum	Exponential	0.30	3.89	7.6	0.89
Pteridium aquilinum	Exponential	4.75	9.51	15.5	0.77
Silene dioica	Exponential	1.53	3.34	68.4	0.82
Sorbus aucuparia	Exponential	1.35	2.71	26.3	0.96
Teucrium scorodonia	Exponential	2.00	4.00	50.4	0.76

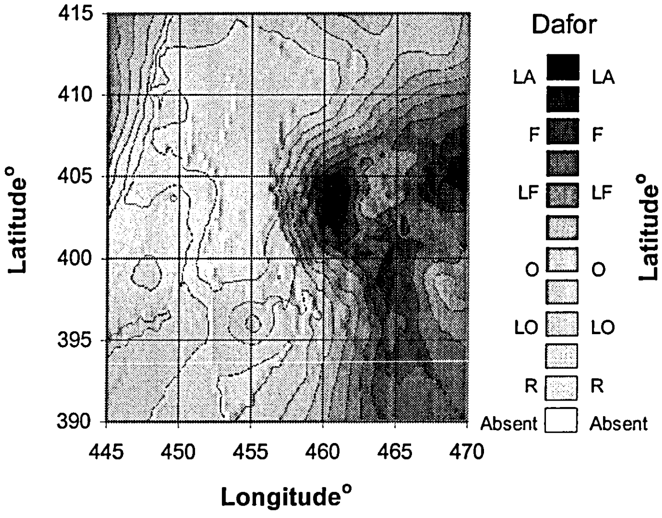
Geology + sample locations



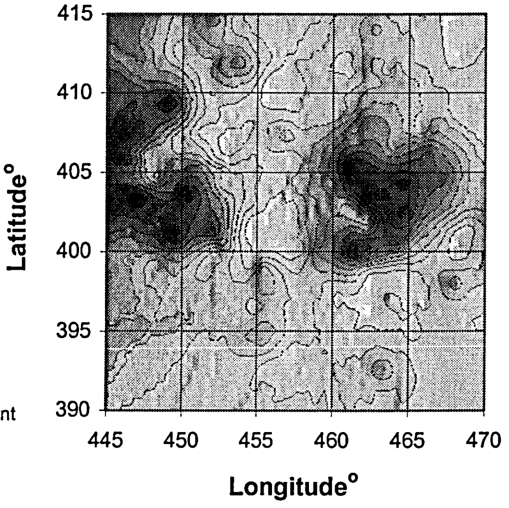
Altitude



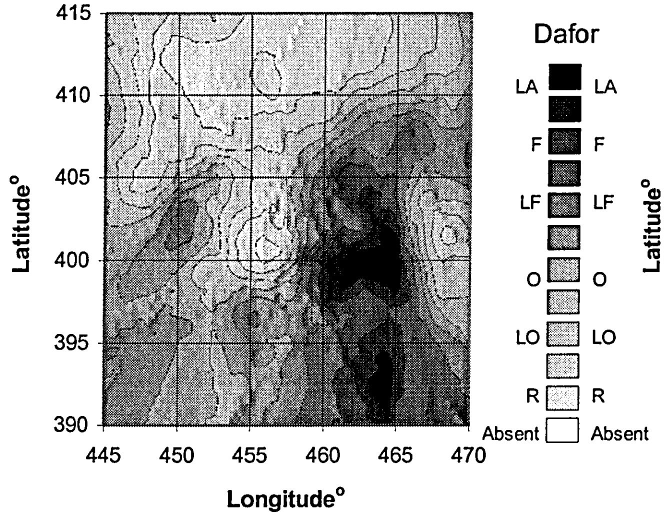
Holcus mollis



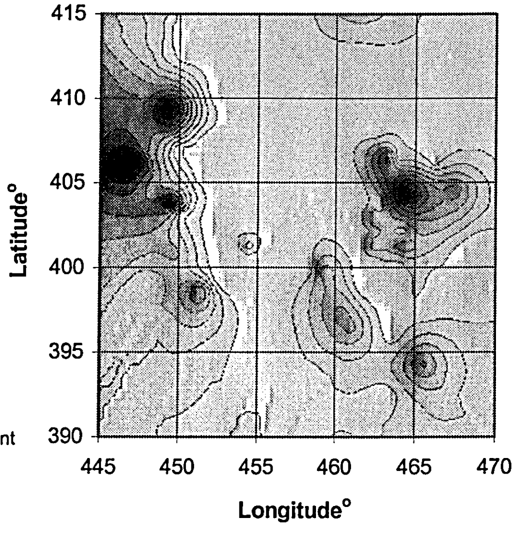
Hyacinthoides non-scripta

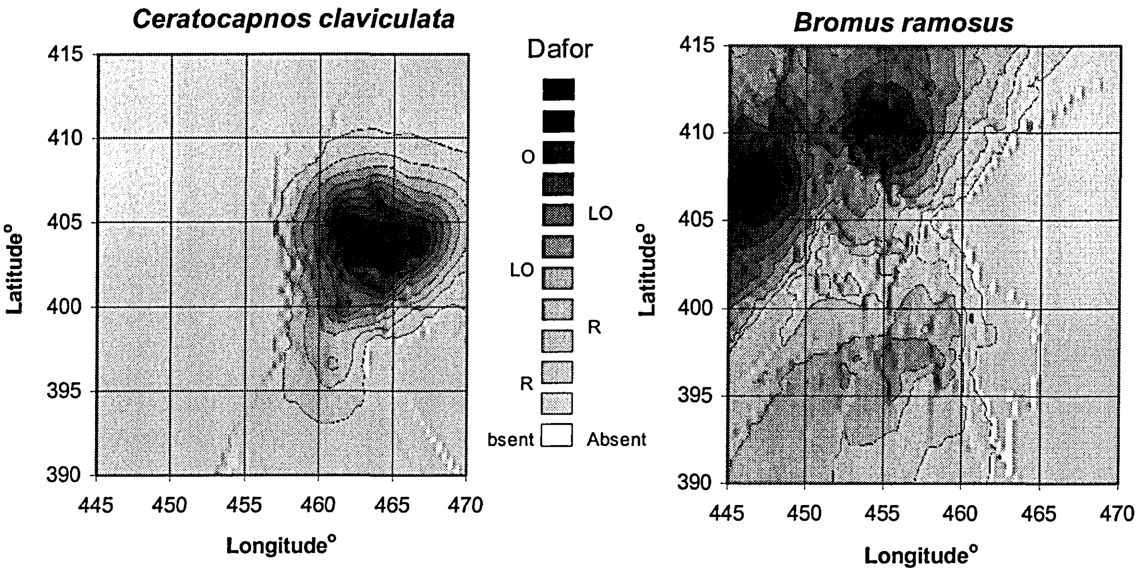


Pteridium aquilinum

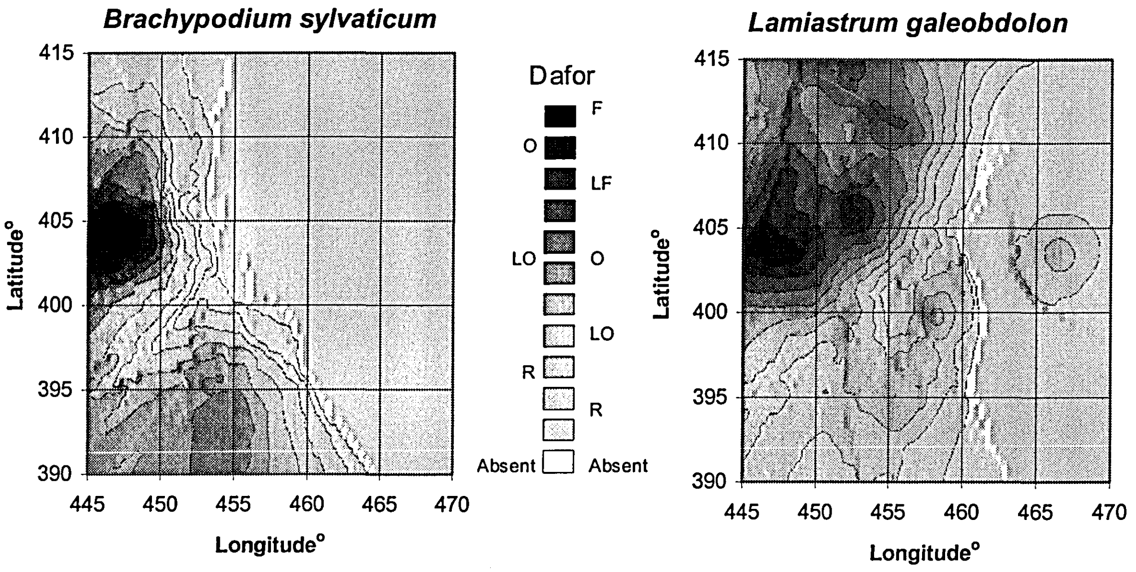


Milium effusum

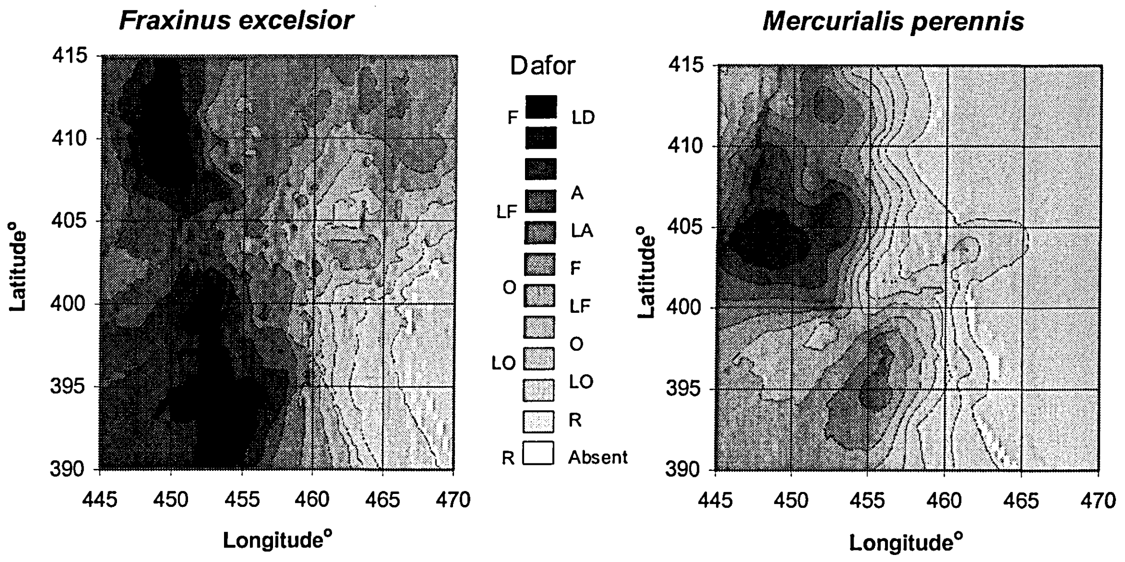



Ceratocapnos claviculata

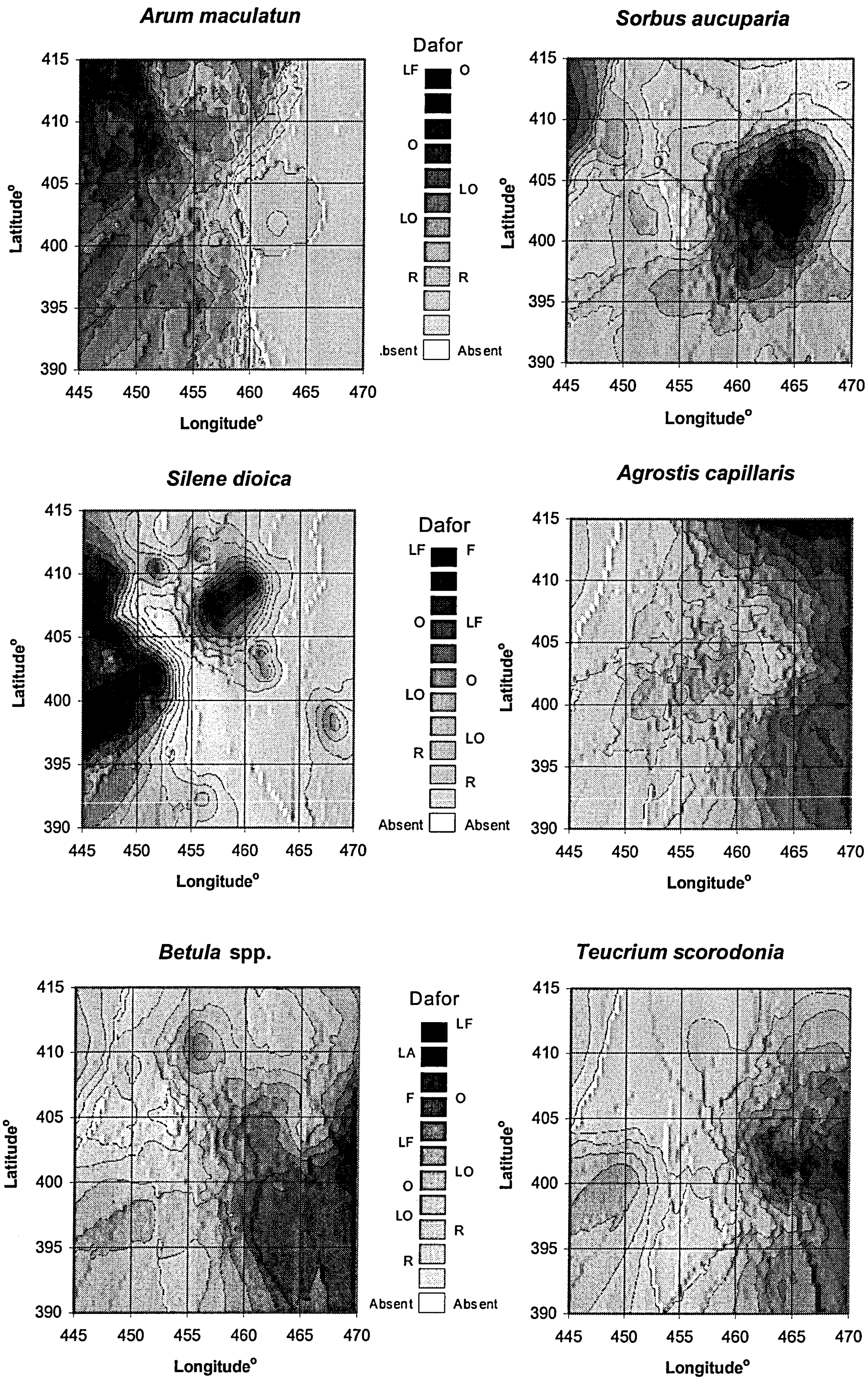
Bromus ramosus


Brachypodium sylvaticum

Lamiastrum galeobdolon


Fraxinus excelsior

Mercurialis perennis


Arum maculatun

Sorbus aucuparia

Silene dioica

Agrostis capillaris

Betula spp.

Teucrium scorodonia

Figure 3.19 Maps, showing the geology, site locations, altitude and the distribution some common woodland species. Th maps for altitude and species distributions have been produced by kriging.

3.6 Discussion

A number of different methods for determining indicator species has been evaluated. This consisted of analysing the data using ordinations, variance partitioning, cluster analysis and GLMs to determine the importance of various factors influencing species composition. Once this had been achieved two main types of indicator species analysis were carried out. These were, the INDVAL indicator procedure for the determination of indicator species at each partition in a dendogram produced by a hierarchical cluster analysis and chi-square type tests and permutation tests using INDVAL on similar data to that analysed by Peterken (1993).

The first analysis involving the ordinations, cluster analysis and GLMs showed that spatial patterns were a major contributing factor to differences in the species composition of woodlands. In particular it has been shown that the effects of age are significantly different depending upon the geology of a site. The application of the INDVAL procedure to these data showed a number of clear species associations with the different habitat-types. However, this method suffered from a number of drawbacks. The first being that it is difficult to fit many of the species onto a dendogram which results in a number of groups containing no species at a given threshold value. The second problem is that a species with a high indicator value, may still be fairly common in the opposing habitat-type, as displayed in a dendogram. This can be compared to rarer species that may be completely restricted to a certain environment being given a much lower indicator value. In terms of producing a table summarising all the data, the two-way table devised by Dufrene and Legendre appears to be a very useful way of displaying the data. However, despite these drawbacks the method developed by Dufrene and Legendre and the method used in this study will still be expected to be far more useful that TWINSPAN or COINSPAN. On the original analysis, with all the sites included, TWINSPAN would have probably separated the sites on the Millstone Grit from the rest. At a later division it will then probably have further separated the Gritstone woods according to their geographical location. In doing this it has to separate the other partitions in the dendogram according to the same criteria which may have produced nonsense groups within the dendogram. TWINSPAN also suffers from the problems of how to assign pseudo-species, since predefined cut off values must be used in order to split abundance data. Even if all these problems are removed, then there is still the problem of reliability i.e. when if an indicator species may still be common in recent or secondary woodland. COINSPAN would have been expected to work better on the ordinations presented, especially in the analysis with two of the geology types removed from the data set, since this analysis formed two clear groups along the first axis and then separated the two groups formed according to the second axis. However, some misclassifications would have likely since this method still divides the CCA scores into two groups, one for those with positive values and another for those with negative values. Despite the big gap down the middle of this ordination it is slightly slanted, making it difficult to get every sample in the correct half of the ordination. So despite not having used either of these two indicator analysis programs, it is felt that no further information would be gained by their use beyond that already gained.

Peterken's method of calculating the percentage of each species occurrence in ancient woodland can be seen to lack any significance test. However, it has been shown that splitting species up according to their percentage occurrence in ancient woodlands is a very useful way of categorising the species. This information cannot be derived from either Honnay *et al.* (1998) or Wulf's (1997) method of using chi-square tests or of speculating as to which species may be indicators such as Rose (1999). Even if a species is found to be significantly associated with ancient woodland it does not necessarily mean that it will make a good indicator species. Simply carrying out chi-square tests may produce a large list of species with some species that are common to both recent and ancient woodlands. These species should not be considered as reliable indicators. The main advantage would appear to be that the INDVAL method will place species in a general group so avoiding some of the risk in performing statistical tests on species that are common to both groups. There actually appears to be little advantage of using the INDVAL index over chi-square tests, so long as species have split up into tables of their percentage occurrence prior to such tests.

It has often been suggested that finding a suite of species that are considered to be associated with antiquity is the best way of identifying ancient woodland. This could be taken a step further using the method described previously. In this case a value of 0.1 is given for each species that is listed in Table 3.24, and a value of 0.25 for each species listed in Table 3.24. If all these values are multiplied together this will give an approximate probability that the site is not ancient.

This study has emphasised the need to have both secondary and ancient woodland data if comparisons are to be made. Since the permutation tests appear to show very little difference from the chi-square tests, it can be assumed that the species abundance information is of little additional use. It may be more useful to collect data for sites that have been surveyed at least twice and possibly not include any abundance information. The importance of including at least the same number of recent sites must also be emphasised. If the data set contains considerably more ancient sites then there is a much greater risk of saying that a species may be considered as an indicator species when it may not be assoicated with ancient woodland. If the data set contains substantially more recent sites then there is the risk of saying a species is not an indicator when it should it may be very closely associated with ancient woodland. From a statistical point of view the latter situation is preferred. This is because the INDVAL procedure and Chi-square tests can be considered as multiple tests. If one species is tested for its association with ancient woodland then the test is valid. If for example, one hundred species are tested for their association with ancient woodland then it would be expected that purely by chance, one of those species will be found to associated with ancient woodland. If it is not feasible to re-survey a large number of recent sites the other option is to include in the analysis considerably more recent woodlands than ancient ones. This is what Peterken did.

The present study has produced a number of tables with different species listed as indicators of ancient woodland. It is suggested that, Tables 3.24 and to a lesser extent Table 3.25, are the most useful lists of species for this purpose. These two tables appear to contain no spurious species. Some species, more typical of other environments such as unimproved grassland do appear when separate lists are produced for the different geology types. Instead of using different lists of species, it may be preferred to use different numbers of indicator species for the different geology types, i.e. two or three for the Bunter Sandstone and five or six for sites on the Magnesian Limestone, Permian Marl and Coal Measures Series. However, there still remains the question of what to do with the rare species reported in Table 3.27. Since this also appears to contain just species that would normally be considered to be ancient woodland indicators, this list should still be very useful. If the characteristics of an indicator species as described by Ferris and Humphrey (1999) are adhered to i.e. species need to be easy to assess, meaningful and cost effective, then rare species would not meet these criteria. However, since they are rare they maybe useful in evaluating site conservation status. This list is of course not complete since a number of rare woodland species are absent from this list such as *Paris quadrifolia*. This is an example of a rare plant of ancient woodland which did occur here historically, but is now largely extinct and where it occurs is rarely found.

Reasons for the different lists of indicator species with geology type

The differences between the lists of indicator species are likely to be associated with the main factors. The sites which are close to urban areas such as many of the sites on the Coal Measures Series or where the drainage is very good such as on the Bunter Sandstone may contain soils with a very low moisture content. The species that would be expected to make good indicators under these conditions will therefore be those that are associated with poor drainage. In a similar way both these geology types tend to produce fairly acidic soils resulting in slower rates of spread for species that are better adapted to soils of high pH. These hypotheses appear to supported from the evidence provided in the results. The species found to have a strong affinity for ancient woodland on the Coal Measures include species often associated with wet habitats such as *Blechnum spicant, Cardamine flexuosa*, *Lysimachia nemorum* and *Salix fragilis*. However, the same does not appear to follow for the Bunter Sandstone. For this geology type *Mercurialis perennis* and *Tamus communis* appear in the list for species with a strong affinity for ancient woodland. These two species are often associated with neutral to alkaline soils. It would therefore appear that the main limiting factor for the colonisation of woodland species, on the Coal Measure Series is the soil moisture content and on the Bunter Sandstone the soil pH. The list of species for those with a strong affinity for ancient woodland on the Magnesian Limestone and Permian Marl contain species that prefer both damp alkali-neutral soils such as *Allium ursinum* and dry soils such as *Anemone nemorosa* and *Melica uniflora*. This geology type appears to have very little in the way of limiting factors that would slow the colonisation rates of woodland species. This probably explains the presence of the relatively high numbers of species that have been refered to as ancient woodland indicator species in woods on this type of geology.

CHAPTER 4

4. ESTIMATING SPECIES TOTALS AND THE TIME REQUIRED TO RECORD SPECIES IN ANCIENT WOODLAND

4.1 Introduction

Woodlands may be surveyed for a variety of reasons - scientific or historical research, management planning, or conservation assessment. Such surveys may be of species, communities, indicator species, historic features and a diversity of other physical factors. Approaches to the surveys range from quick, 'look-see' evaluations, to timed assessments, and meticulous, comprehensive surveys. Woodland surveys and methods are discussed in Peterken (1993), and by Usher (ed.) (1986). Some indication of the reliability of minimal surveys in predicting the species richness and ecological quality of a site such as a woodland, may be of considerable benefit. Furthermore, evaluation of the time requirement for thorough and effective surveys of higher plants, will assist researchers, consultants and agencies in assessing the time required for site work.

Kirby *et al.* (1986) and Sykes *et al.* (1983), considered aspects of timed surveys, seasonal variation, and recorder variability. These discussions raised key points of interest in terms of the reliability of ecological surveys and in their evaluation. Kirby *et al.* found significant differences between the number of species recorded by different observers and also significant differences between different observers according to different seasons. Sykes *et al.* found significant differences in cover estimates between observers.

Increasingly driven by commercial consultancy, and constrained by the budgeting restrictions of both local government and agencies, ecological surveyors are often pressed for 'quick fix' evaluations of potentially complex environments such as woodlands. Botanical species lists are frequently used in evaluation, but how comparable are they?

Comparison of the results of such surveys, against good base-line data, quickly shows cause for concern.

Ecologists are under pressure to quantify the cost of damaging effects on the environment such as deforestation. Numerous methods have been used to estimate the number of species, or species richness in a community. The most common methods are the number of observed species, extrapolation of species-area curves, assuming the log-normal distribution and non-parametric estimators (Palmer, 1990). Extrapolation of species-area curves normally involves assuming that the number of species is proportional to the logarithm of the area or that the logarithm of the number of species is proportional to the logarithm of the area. The total area of the community being surveyed can then be used with the regression equation produced to estimate the number of species.

The total area of the community being surveyed can then be used with the regression equation produced to estimate the number of species. Different models that have been used to predict species richness from accumulation curves has been tested by Soberon and Llorente (1993). They tested three models. The first assumes that adding a new species decreases proportionally to survey effort and eventually reaches zero.

$$y = \frac{a}{b}\left(1 - e^{-bt}\right) \tag{4.1}$$

Where y represents the expected value of the number of species at time t. The parameter α represents the list increase rate at the beginning of the collection and the asymptote is given by α/b.

The second model assumes that increasing the size of the collection decreases the probability of adding a new species in a non-linear way.

$$y = \frac{1}{z}\ln\left(1 + zat\right) \tag{4.2}$$

However the probability of finding a new species never reaches zero.

The third model is the Clench equation. This is also referred to as the Michaelis-Menton equation used by Clench (1979). The asymptote is given by α/b.

$$y = \frac{at}{1+bt} \qquad (4.3)$$

This model assumes that the probability of adding a new species will improve up to a ceiling with increasing survey effort. Soberon and Llorente (1993) argue that this assumption may be met if the surveyor gains experience with the site, taxa, and methods during the survey.

The current study is concerned with time surveys where there is no definitive length of time required to record all species. The abundance of species is also not known so that the integration of the log-normal distribution cannot be used. Non-parametric estimators such as the jackknife and bootstrap also cannot be used on this type of data since the number of new species was recorded for each time interval and these types of estimates only give values for species richness and do not estimate the total number of species.

The assumption that all species are equally detectable has been discussed by Bunge and Fitzpatrick (1993). Differences in detectability can be attributed to the fact that the number of individuals will vary among species as well as the size and appearance of the species will vary. A number of different methods are described by Bunge and Fitzpatrick (1993), for models with different class (species) sizes. Departure from equal representation where equal class sizes are assumed results in an underestimate of the total number of classes (Bunge & Fitzpatrick, 1993; Lewontin & Prout, 1956).

Species-area relationship models have also been developed to compensate for different ranges and occupancy of species (Ney-Nifle & Mangel, 1999). However, data would be needed for all species in each time unit, in order to be able to use the models they devised.

This study presents an alternative method for estimating the total number of species when only new species have been recorded for each time unit, through the use of non-linear models

4.2 Materials and methods

4.2.1 Study area

Woodlands on three different geology types were selected. These included six woods on each of Coal Measures Series, the Millstone Grit Series and the Corallian Series. All the woods were subjected to timed surveys of higher plant species. All woods included in this chapter are considered to be ancient woodland i.e. have existed continually as some form of woodland since before 1600 AD (Watkins, 1990; Rackham, 1995; Peterken and Game, 1984; Spencer and Kirby, 1992).

The Coal Measures Series woods are located within the Sheffield City boundary, at the southeastern tip of the Pennines, in central England. The soils are mainly acidic and free draining but they do contain some mineral enriched flushes associated with shale outcrops.

The woods on the Gritstone Series are located to the west of Sheffield. They have soils which are mostly acidic, and topography ranging from steep sloping, with freely-drained soils, to flat with impeded drainage and occurred at higher altitude than either the Coal Measures Series or Corallian Series.

Sites on the Corallian Measures were either located on slopes with very thin alkaline soils, or were situated by rivers on flood plains with deep alluvial soils. The woods with thin alkaline soils typically contained large areas dominated by *Mercurialis perennis*. The woods located on flood plains had deeper, more eutrophic soils, with PHs neutral to slightly acid. They had an abundant associated cover of *Urtica dioica* and *Heracleum sphondylium*.

The names and locations of the woods are given in Table 4.1.

Table 4.1. The locations and areas of the sites surveyed on the three different rock types.

Site	Grid Ref.	Area (ha)
Coal Measures Series		
Hang Bank Wood	48 364843	3.0
Buck Wood	48 370844	7.1
Ashes Wood	48 360838	7.7
Leeshall Wood	48 372835	12.2
Rollestone Wood	48 366835	15.4
Ecclesall Woods	48 323825	130.0
Millstone Grit Series		
Holes Wood	48 239908	2.5
Pears House Clough	48 237905	4.5
Owler Carr	48 225898	5.5
Raynor Clough	48 428940	6.6
Blackbrook Wood	48 244971	15.7
Agden Wood	48 250935	19.8
Corallian Series		
Chafer Wood *	44 840845	3.1
Bridge Wood*	44 732623	4.5
Willowbridge Wood*	44 724625	13.4
Little Park Wood	44 803870	27.8
Settrington Wood*	44 850690	64.0
Haugh Wood	44 820860	100.5

4.2.2 Data collection

The woodlands were chosen to include (a) two groups with very comprehensive botanical species lists, and (b) one group with only partially complete lists. The analysis was applied differently to (a) and (b). All the woods on the Coal Measures Series and Millstone Grit Series had all been thoroughly surveyed as part of an on-going programme of work. The woods on the Corallian Series had been partially surveyed. The botanical surveys undertaken were of all vascular plants including trees and shrubs, by two surveyors working together. Both surveyors were moderately experienced woodland ecologists and competent field botanists.

The surveyors walked through the woods aiming to effectively cover as much of the site as possible. Species were recorded by five-minute intervals. The number of species found in the first five minutes was recorded and then the number of new species recorded for each subsequent five-minute interval. The time spent recording ranged from fifty minutes for the smallest wood (2.5 ha) to 140-160 minutes for larger woods (60-130 ha). The surveys were carried out between July and August 1994.

Data analysis 4.3

4.3.1 Deriving equations for the species accumulation curves

Three methods were used to predict the total number of species present from the species accumulation curves. The first involved plotting the number of new species against time and then fitting a best fit exponential decay to the data. The equation

$$y = N(1 - e^{-at}) \qquad (4.4)$$

was used to fit an exponential decline for the number of species recorded with time and estimate the total number of species for each wood. This is the same as the first equation given in the introduction (4.1). Where N is the predicted total number of species and α is a constant. This equation assumes that the probability of finding a new species declines exponentially but also has the assumption that all species are equally abundant **(Model 1)**. The latter is known not to be the case (Krebs, 1985). Three theories are widely accepted, the geometric decline, the log normal distribution and the 'broken-stick' distribution. Using information on abundance in order to predict population size from accumulation curves is difficult (see Goudie and Goldie, 1981). This is particularly true when the survey is for a pre-assigned time (rather than for a pre-assigned number of species). The following method cannot claim to be theoretically well-founded but did provide good results.

A new method was used which assumes an exponential (or geometric) decline in the abundance of species and also that species are found in the order of their abundance **(Model 2)**. Now if the relative abundance of species i is $e^{-\alpha}$ and the total number of species is N then the rate of finding new species (when n have been found) is proportional to:

$$e^{-\alpha n} - e^{-\alpha n}. \qquad (4.5)$$

When N is large (so that the discrete process can be replaced by a continuous one) it is found that the accumulation curve is:

$$y = \frac{1}{\alpha} \ln\left[e^{\alpha N} - \left(e^{\alpha n} - 1\right)e^{-Rt}\right]. \tag{4.6}$$

Here *n* is the number of species found after time *t* and *R* is a rate constant. (Fast walkers/workers have a large value of *R* but it does also depend upon the wood etc). Thus the observed accumulation curves have to be fitted to a three-parameter family (rather than a two-parameter family as for the first method). The fitting was achieved by a least-squares method and enables us to find *N*.

The third model used was the Clench Equation, which was described in the introduction (4.3).

All of the models assume a homogeneous environment.

The woods on the Coal Measures and Millstone Grit Series were used to test the reliability of these two models. The model giving the best fit for the total number of species predicted, was then used to assess the proportion of species recorded by the surveyor for the woods on the Corallian Series.

Finally, the equation relating the number of species to the accumulation curves was rearranged to give the predicted length of time required to record a given proportion of the species present.

4.3.2 Analysis of the species composition data

Principal component analysis, based on a covariance matrix, conducted on CANOCO (ter Braak, 1988), was used as a means of identifying any difficulties in estimating the total number of species according to their species composition. Presence/absence data of suggested 'ancient woodland indicators', was used together with some environmental data to identify the main environmental factors influencing the variation in the species data. The list of indicator species used was that produced as a provisional list by the Peak Park Joint Planning Board (Peak National Park unpublished).

4.4 Results

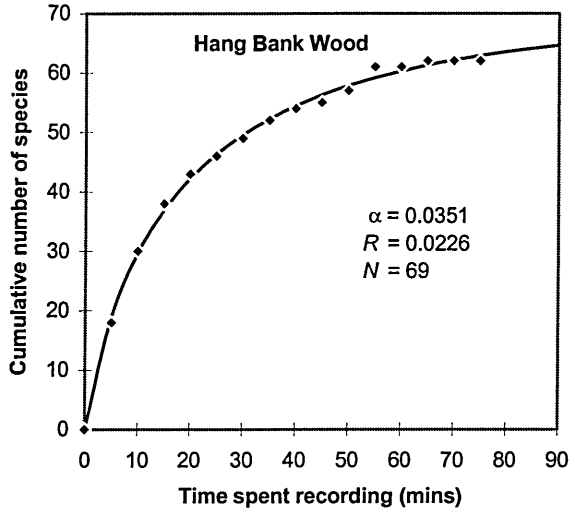
4.4.1 Estimating the total number of species

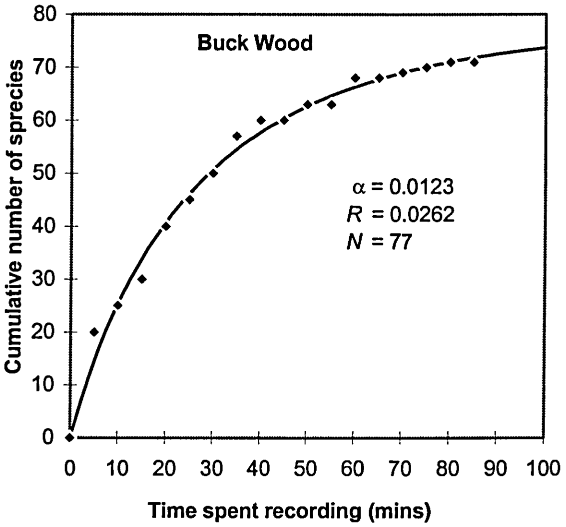
Two of the cumulative species curves for each of the three rock types are shown in Figure 4.1. The figures also show the best-fit curves for the new method. Some of the woods, in particular Settrington Wood, Blackbrook Wood and Agden Wood contain large areas of species-poor plant communities. These sites all show a series of steps for the cumulative curve. However, the total numbers of species predicted for these woods are very close to the actual number recorded.

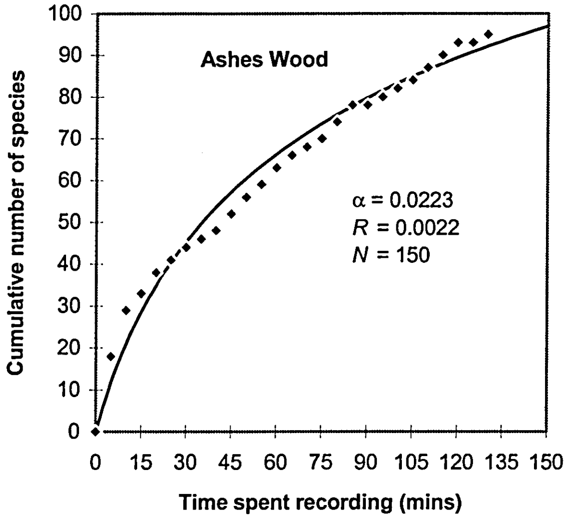
The number of species predicted by the single exponential decline for new species, and the model for a decline in species abundance, both can be significantly related to the total number of species recorded. Of the two, the model for a decline in species abundance gives a closer fit to the actual number of species recorded (Table 4.3) and gives a value for b in the regression equation much closer to 1. This means that the reliability of the prediction remains fairly constant. Simply adding 5.6 to the predicted value, gives the best estimate for the total number of species. The two regression coefficients were found to be significantly different, $P = 0.006$ ($t = -3.04$, $v = 20$).

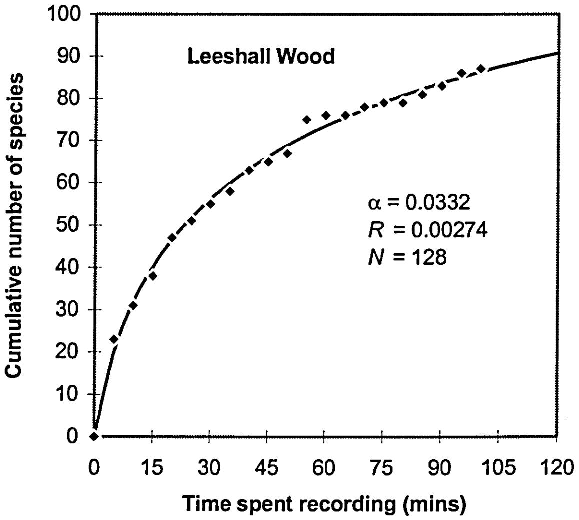
Table 4.2. Predicted number of species by the three models compared with the actual number recorded, not the timed surveys. The four woods which were not thoroughly surveyed on the Corallian Series are marked *.

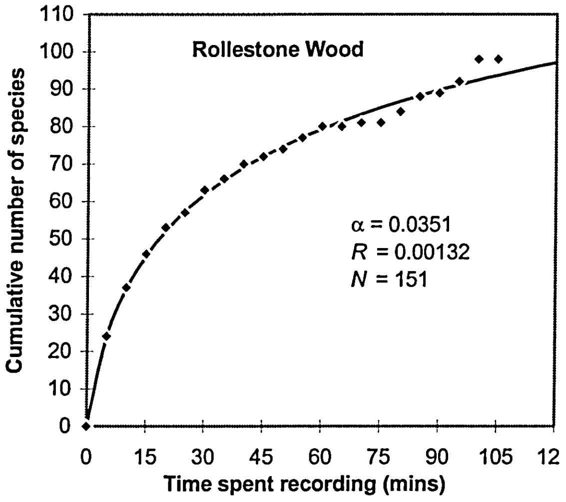
Site	Actual number of species recorded	Total predicted by single exponential	Total predicted by new model	Total Predicted by the Clench Equation	Proportion of the total number of species predicted by the new model actually recorded
Coal Measures Series					
Hang Bank Wood	85	76	69	76	1.23
Buck Wood	109	79	78	96	1.40
Ashes Wood	138	136	150	133	0.92
Leeshall Wood	143	97	128	108	1.12
Rollestone Wood	165	105	151	112	1.09
Ecclesall Woods	176	105	164	118	1.07
Millstone Grit Series					
Holes Wood	61	63	67	63	0.91
Pears House Clough	79	63	54	67	1.46
Owler Carr	97	88	88	97	1.10
Raynor Clough	94	95	111	86	0.85
Blackbrook Wood	102	93	90	69	1.13
Agden Wood	75	62	66	70	1.14
Corallian Series					
Chafer Wood *	99	75	126	91	0.79
Bridge Wood*	96	109	115	119	0.83
Willowbridge Wood*	62	81	113	84	0.56
Little Park Wood	141	138	149	157	0.95
Settrington Wood*	90	97	117	84	0.77
Haugh Wood	159	106	123	185	1.29

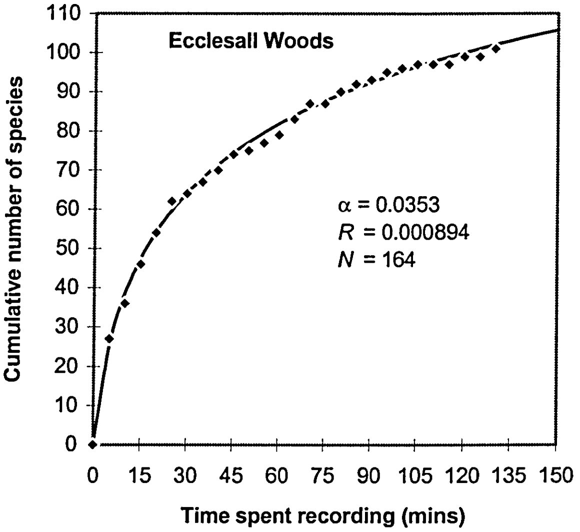

Hang Bank Wood
α = 0.0351
R = 0.0226
N = 69
Cumulative number of species
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
70
80
90
Time spent recording (mins)

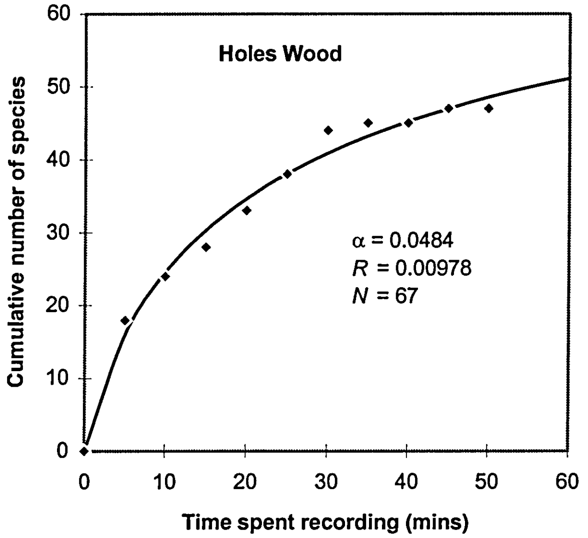

Buck Wood
α = 0.0123
R = 0.0262
N = 77
Cumulative number of species
0
10
20
30
40
50
60
70
80
0
10
20
30
40
50
60
70
80
90
100
Time spent recording (mins)

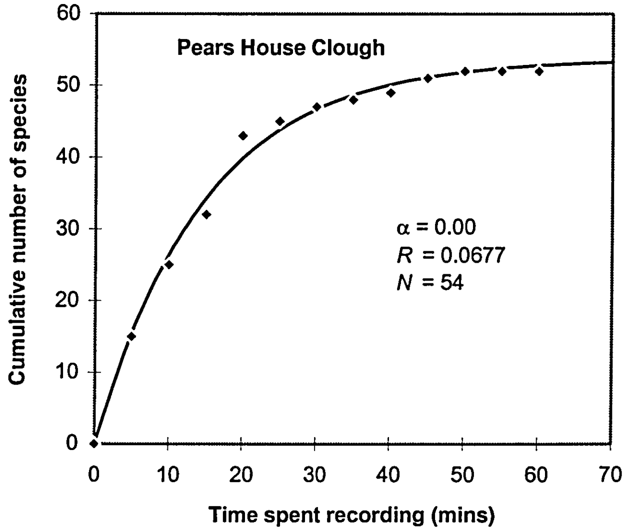

Ashes Wood
α = 0.0223
R = 0.0022
N = 150
Cumulative number of species
0
10
20
30
40
50
60
70
80
90
100
0
15
30
45
60
75
90
105
120
135
150
Time spent recording (mins)

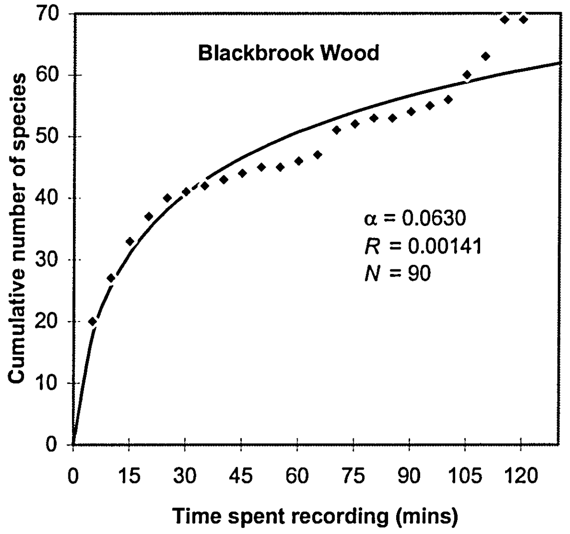

Leeshall Wood
α = 0.0332
R = 0.00274
N = 128
Cumulative number of species
0
10
20
30
40
50
60
70
80
90
100
0
15
30
45
60
75
90
105
120
Time spent recording (mins)

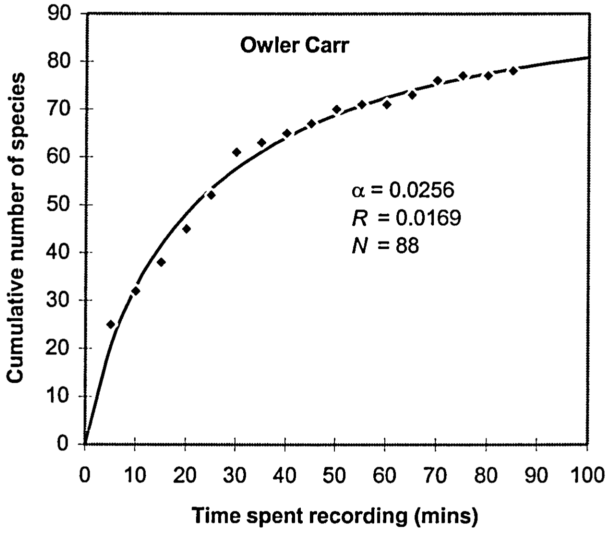

Rollestone Wood
α = 0.0351
R = 0.00132
N = 151
Cumulative number of species
0
10
20
30
40
50
60
70
80
90
100
110
0
15
30
45
60
75
90
105
12
Time spent recording (mins)

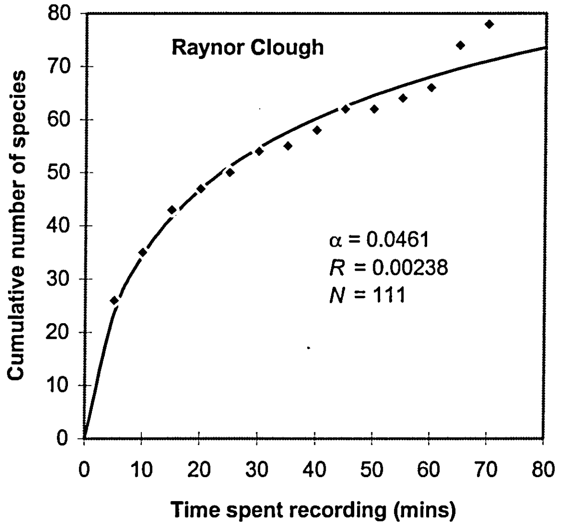

Ecclesall Woods
α = 0.0353
R = 0.000894
N = 164
Cumulative number of species
0
10
20
30
40
50
60
70
80
90
100
110
0
15
30
45
60
75
90
105
120
135
150
Time spent recording (mins)

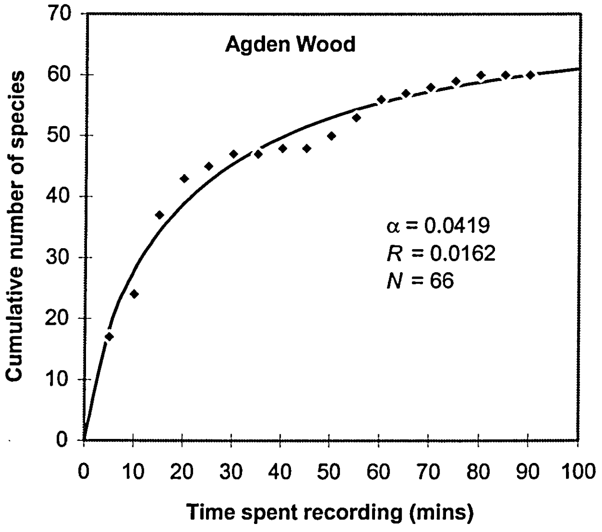

Holes Wood
α = 0.0484
R = 0.00978
N = 67
Cumulative number of species
0
10
20
30
40
50
60
Time spent recording (mins)
0
10
20
30
40
50
60


Pears House Clough
α = 0.00
R = 0.0677
N = 54
Cumulative number of species
0
10
20
30
40
50
60
Time spent recording (mins)
0
10
20
30
40
50
60
70


Blackbrook Wood
α = 0.0630
R = 0.00141
N = 90
Cumulative number of species
0
10
20
30
40
50
60
70
Time spent recording (mins)
0
15
30
45
60
75
90
105
120


Owler Carr
α = 0.0256
R = 0.0169
N = 88
Cumulative number of species
0
10
20
30
40
50
60
70
80
90
Time spent recording (mins)
0
10
20
30
40
50
60
70
80
90
100


Raynor Clough
α = 0.0461
R = 0.00238
N = 111
Cumulative number of species
0
10
20
30
40
50
60
70
80
Time spent recording (mins)
0
10
20
30
40
50
60
70
80


Agden Wood
α = 0.0419
R = 0.0162
N = 66
Cumulative number of species
0
10
20
30
40
50
60
70
Time spent recording (mins)
0
10
20
30
40
50
60
70
80
90
100

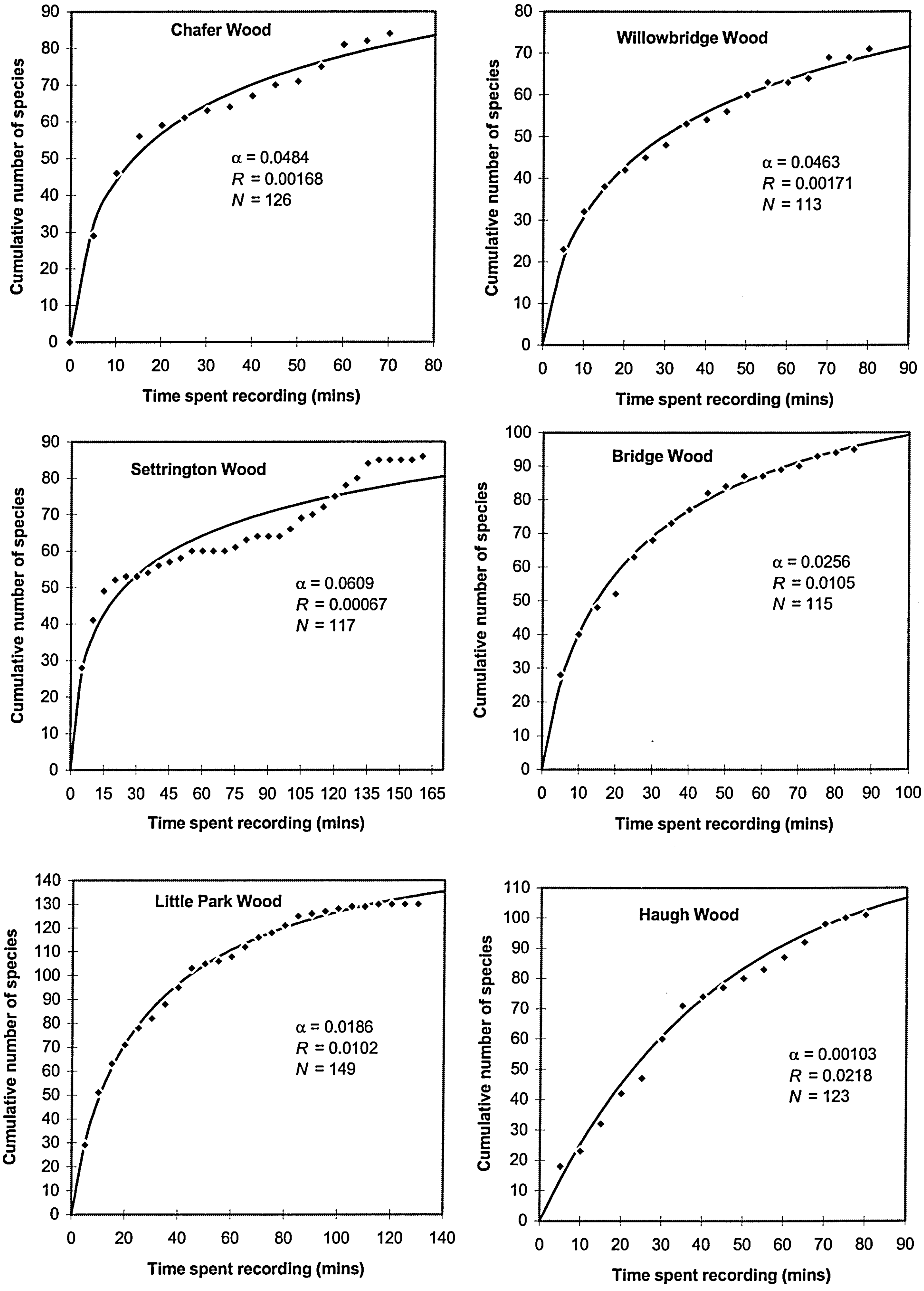


Figure 4.1. The relationship between survey time and the number of vascular plants recorded for walked surveys. The lines are drawn from the equation assuming species are found in order of abundance and that their abundance declines exponentially.

4.4.2. Comparing the different models

Simple regression and correlation procedures were used to calculate regression equations and correlation coefficients for the actual number of species recorded and the predicted number of species (Table 4.3). All the models gave reasonable results with high correlation coefficients. The model appears to give slighter better results since the slope paramter is closest to 1.0. A value of 1.0 would mean that any error is constant no matter on the actual total number of species. A value less than 1.0 means that the predicted numbers are underestimated to a greater extent with increasing number of species.

Table 4.3 Simple regression equations with the actual number of species recorded as the independent variable the predicted values from the three models as the dependent variables.

Model	α	b	F		R^2
Single Exponential	25.46	0.427	51.7	***	0.84
New Model	-5.19	0.965	63.6	***	0.86
Clench Equation	31.70	0.540	29.8	***	0.75

The results of the regression procedures were tested to see whether the slope parameters were significantly different to each other. Analyis of covariance was used in order to this. The results of these tests are given in Table 4.3 and 4.4. Table 4.3 shows the F tests. The predicted values can be seen to be highly correated with the actual numbers. The models do not appear to give significantly overall higher or lower estimates than each other. However, the interaction effect was found to be significant. The slope parameters were tested against each other as shown in Table 4.5

Table 4.4 Analysis of covariance for the actual number of species recorded against the predicted values produced by the three equations.

	NDF	DDF	F	
Expected	1	30	133.6	***
Model	2	30	3.1	
Expected × model	2	30	8.65	**

Table 4.5 Testing the differences in the slope paramters produced by the three different models.

	DF	Estimate	*t*
New - Single	30	0.54	3.94 ***
New - Clench	30	-0.43	3.12 **
Clench - Single	30	-0.11	0.83

4.4.3 Assessment of the number of plant species recorded on the Corallian Series

The model for an exponential decline in species abundance was used to assess the timed surveys carried out on the Corallian Series (Table 4.2). All the sites are considered ancient, but there was considerable variation in the conditions of the contemporary woodlands, from 'semi-natural' in character, to rather disturbed and secondary. Willowbridge Wood and Bridge Wood are located on flood-plains with evenly spaced, even-aged trees on nutrient-enriched soils. These woods were therefore not considered as being of a 'semi-natural' character. Settrington Wood contained a similar ground flora to these woods and Chafer wood also had the characteristics of 'secondary' woodland. Little Park Wood consists of a shrub layer and field layer characteristic of ancient woodland. Haugh Wood contained a variety of different vegetation types, and the accumulation curve data were restricted to the semi-natural woodland areas. The time predicted to record 95 % of the species was also very small, at less than two hours. This site was subsequently omitted from the analysis for the predicted lengths of time required to find a given percentage of the species present. The survey of the woods on the Corallian Series was primarily involved in accurately surveying ancient woodland and so less time was spent recording species in the four woods that contained woodland that was more secondary in character. These four woods were found to have significantly smaller proportions of species predicted by the new model actually recorded than the other woods, (Mann-Whitney U test, $z = -2.97$, $P = 0.004$).

4.4.4 Evaluation of the predicted species totals using ordinations

The evaluation of the estimated totals was conducted using Principal Component Analysis. Extra data on woods on the geological types were available and so these were added to help show the variability of the woods on the three different rock types. The ordination diagram for the sample scores is shown in Figure 4.2. The correlation coefficients for the sample scores and environmental variable are shown in Table 4.5.

Table 4.5. Pearson correlation coefficients between the sample scores of the first two axes of the Principal Component Analysis and the environmental data. The probability values are for 28 degrees of freedom (n=33) and for a two tailed test. The Eigenvalues are also given.

	Axis 1		Axis 2	
	Eigenvalue = 0.254		Eigenvalue = 0.117	
Variable	*r*	*P*	*r*	*P*
Millstone Grit Series	-0.64	< 0.001	-0.08	0.66
Coal Measures Series	0.02	0.90	0.74	< 0.001
Corallian Series	0.56	0.001	0.64	< 0.001
Area: ln	0.45	0.01	0.13	0.49
Mean Altitude	-0.51	0.002	0.14	0.44

Geology, size of wood and altitude appear to influence species composition and species richness of the sites, at least in terms of suggested ancient woodland indicators. The two woods classed as ‘semi-natural’ on the Corallian Series, Haugh Wood and Little Park Wood, are located to the right of the ordination diagram. The two woods described as the poorest in terms of ancient woodland quality on the Corallian Series, Willowbridge Wood and Bridge Wood, appear to the left of the ordination and appear to be similar in their species composition to the woods on the Millstone Grit Series. Settrington Wood and Chafer Wood, both ‘ancient’ but with mainly secondary woodland communities and only small areas of a semi-natural character, are located in the centre of the ordination diagram.

Sites with soils of low pH are distributed to the left of the diagram and those of higher pH appear to the right of the ordination diagram. There is a clear gradient from the Gritstone Series to the Coal Measures Series, with the wood located on the border of the two rock types, Raynor Clough, in the middle of the ordination diagram.

The woods towards the bottom of the ordination appear to have been underestimated in terms of their total number of species recorded. The woods in this part of the diagram were those which contained extensive areas of poorly drained ground, in particular Agden Wood, Owler Carr, Settrington Wood and Willowbridge Wood.

Woods containing flushes are more likely to be of a more heterogeneous character and so it is probably the assumption of a homogenous environment, which is where the model is weakest.

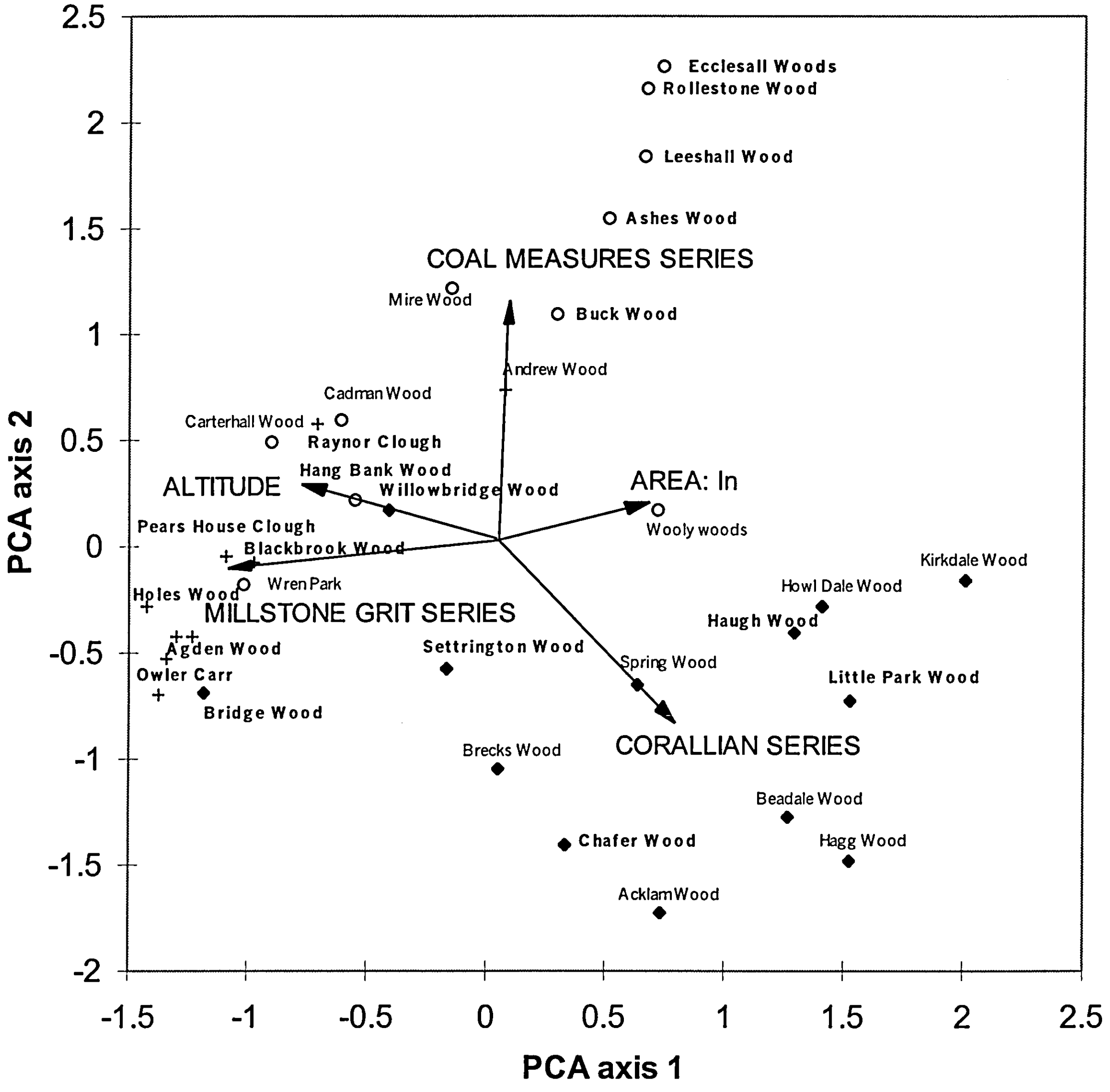


Figure 4.2. Ordination produced by principal component analysis of the woods on the different rock types - Coal Measures Series (o), Millstone Grit Series (+), Corallian Series (♦). The sites are distributed according to the presence of ancient woodland indicators. The environmental axes have then been superimposed on the ordination diagram. This therefore represents an example of indirect gradient analysis. The lengths of the arrows are proportional in length to the significance of the relation between the scores produced by PCA. and the environmental variables. *r* species-environment: axis 1, 0.818; axis 2, 0.814; 37.2% variance accounted for by the first two axes. The woods which were used for the timed surveys are printed in bold.

4.4.5 Time well spent? Justifying the time spent on surveys

The equation for the exponential decline in species abundance was used to estimate the proportion of species expected to be found for a given length of time surveying and a given total number of species. The equation can be re-arranged to give:

$$\text{Predicted proportion of species recorded} = \frac{1}{aN}\ln\left[e^{aN} - \left(\frac{e^{aN}-1}{e^{t-\frac{1}{R}}}\right)\right] \qquad (4.7)$$

A mean value for α was used (0.033). R and N, were found to be correlated with each other. However, R was negatively skewed. A regression equation using the log of R was therefore used (F = 14.8, DF=15, P=0.002).

$$\ln R = -2.52 - 0.0285N \qquad (4.8)$$

The exponential of this predicted value was used to estimate the value of R in the former equation above. Predicted values could then be calculated for a grid of values for time spent surveying and the total number of species at a site (Figure 4.3).

From the above equations the time required to record 95% of a total of 200 species is approximately 80 hours, 99% of species is approximately 170 hours and the time required to record 100% of the species is approximately 500 hours. If we assume that the rate constant 'R' is not related to the number of species then the predicted length of time required to record 100% of 200 species is 5 hours.

However, the total number of predicted species, area and R, were all found to be correlated with each other. This would mean that for a small area with a small number of species all or almost all the species can be seen on entering a site. However, for a large site a much smaller proportion of the total number of species can be seen on entering the site, the rest of the species may take a great deal of effort in terms of survey time to find. A regression equation was used to relate the predicted total number of species (log) to the area (log). The equation is:

$$\text{Ln}N = 4.26 + 0.1604\ln\text{Area(ha)} \qquad (4.9)$$

The regression equation has an R^2 value of 0.26 and probability value of $P = 0.036$. From this equation the predicted number of species in a wood of one hectare equals 70, for 10 ha, the predicted number is 100 species and for a 100 ha wood the predicted number of species is 150. If these values are applied to Figure 4.3, it would be appear to be virtually impossible to record all the species present in a wood greater than approximately100 ha in size.

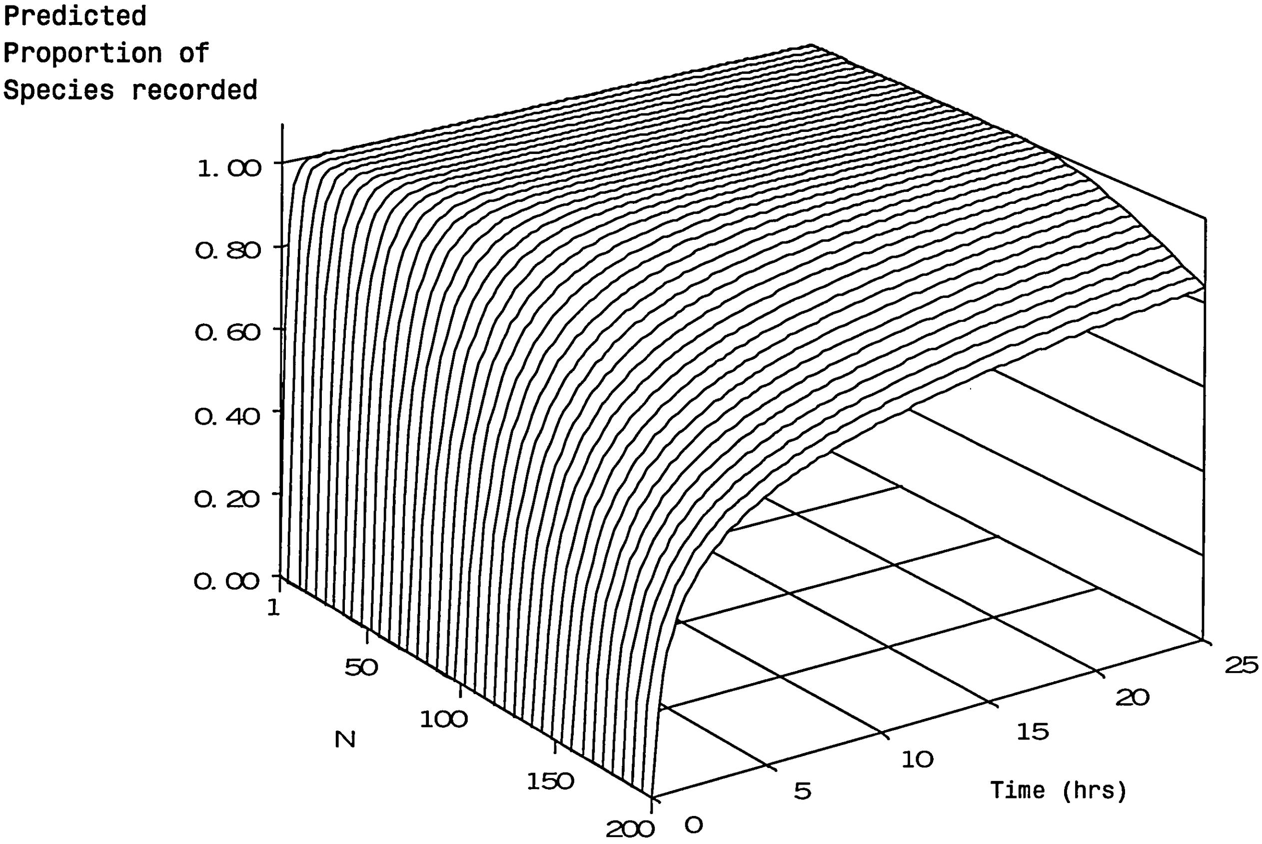


Figure 4.3. The predicted proportion of species expected to be found according to the number of species in a site and the time spent recording.

4.4.6 Testing the assumption of an exponential decline in species abundance: Rank-abundance curves

A number of different models exist for the rank-abundance diagrams. These are the 'geometric', 'broken stick', 'log-normal' and 'log series'. The least equitable distribution of these is the geometric series (Begon *et al.*, 1990). Since an exponential (geometric) decline in species abundance has been assumed in the new model shown in equation 4.3, this can be tested on data collected on the % cover data of Chapters 2 and 6. The data in Chapter 2 are for just vascular plant species in one metre square quadrats and that used in Chapter 6 are for all vascular plants and mosses in 100 m^2 quadrats. The rank-abundance diagrams for these two sets of data are shown in Figure 4.4.

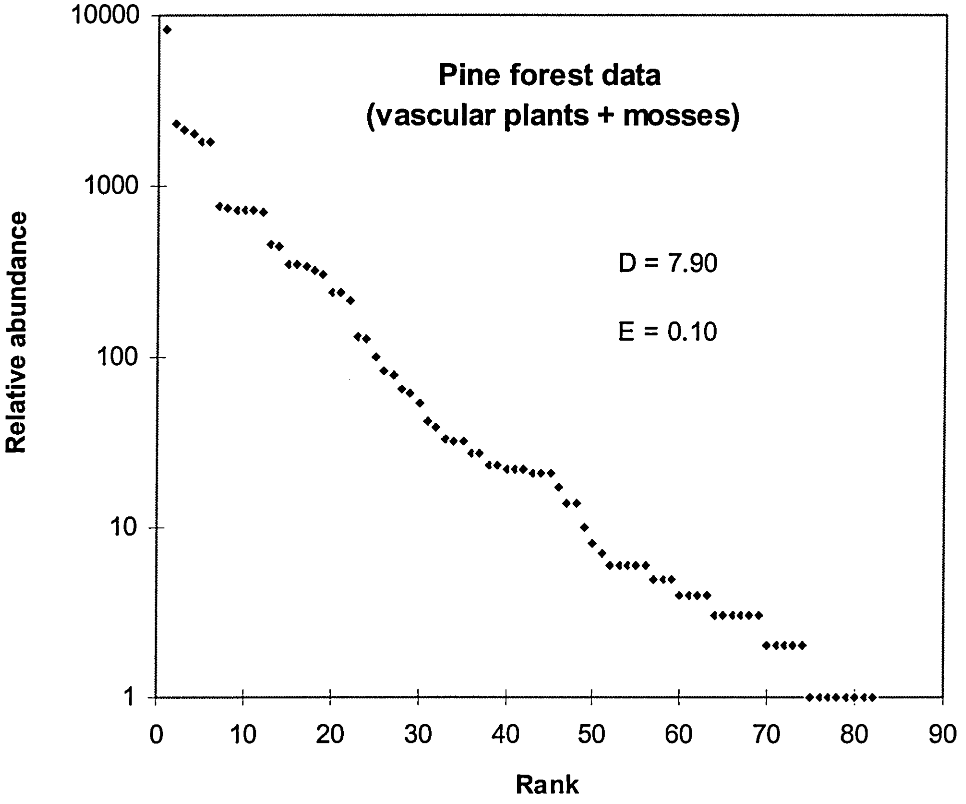
The graphs drawn appear to be most similar to the geometric series and log series models. Communities dominated by a single species or small number of species will have low diversity and equitability indices (Begon *et al.*, 1990). These have been calculated using the equations below.

$$\text{Simpson's index } D = \frac{1}{\sum_{i=1}^{S} P_i^2} \qquad (4.10)$$

$$\text{equitability } E = \frac{D}{D_{\max}} = \frac{1}{\sum_{i=1}^{S} P_i^2} \text{X} \frac{1}{S} \qquad (4.11)$$

The rank-abundance diagram, for the colonisation data, is mainly dominated by *Rubus fruticosus* and *Hyacinthoides non-scripta.* The pine forest data is mainly dominated by *Calluna vulgaris* and the moss, *Hyalocomium splendens.* It is not surprising therefore that both these habitat-types have low diversity and equitability indices.

The two data sets shown in Figure 4.4 appear to support the assumptions used in the new model of Equation 4.3 *i.e.* .two very different woodland data sets both showing exponential declines in the abunadance of species.



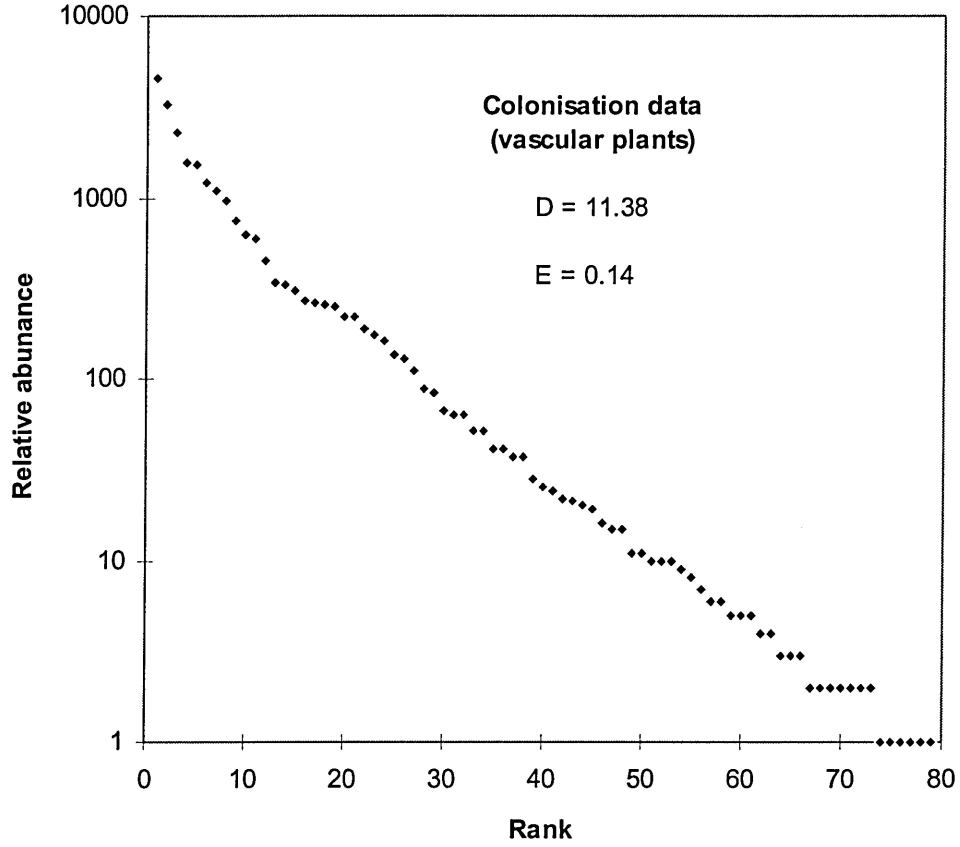


Figure 4.4. Rank-abundance diagrams for the colonisation data (Chapter 2) and pine forest data (Chapter 6). Simpson's diversity index D and Equitability values (Simpson's E) are shown.

4.5 Discussion

Exponential declines in species abundance or appearance were assumed in the model, and the improved performance of Model 2 over Model 1 and the Clench Eqiuation supports this assumption. The predicted and actual numbers of higher plant species recorded with time, were significantly closer with Model 2. A single exponential equation alone, clearly underestimates the total number of species. The main justification for the assumption that species are found in order of their abundance has to be that the results are good, although exponential declines in the abundance of species are commonly found, and the assumption of finding species in order of abundance is the simplest model to use.

Moreno and Halffter (2000) argue that the linear dependence model (Model 1) predicts the lower limit and the Clench model the upper limit, and the true species accumulation curve should lie between them. The present study clearly contradicts this assumtpion, suggesting that the Clench model may also underestimate species richness. If this is the case then it is worth considering the assumptions behind the Clench equation. Soberon and Llorente (1993) state that the biological meaning of the clench equation is that the probabilty of adding new species will improve (up to a ceiling) as more time is spent in the field. They go on to claim that as one accumulates experience with the site, taxa, and methods, the chance of adding new species will improve. It is felt that this argument is very weak compared to the reason that the single expontial model understimates species richness is because species occur in different quantities.

However, there are still some obvious complicating factors. One is the clear link between species diversity and environmental variation within a site. This variation will have complex relationships with woodland 'types' (geology, topography, micro-climate *etc.*) and woodland history. Furthermore, it is likely that there will be a link between woodland size and the probability of internal site diversity i.e. streams, flushes, pedological variation *etc.* These relationships clearly require further, more rigorous consideration. For the woods that were thoroughly surveyed the predicted totals were often lower than the actual recorded number of species. This is likely to be the result of the assumption of a homogenous environment in the model.

An additional complication, is that the survey doesn't record 'occurrence' as such, but the success of the surveyor in 'finding' the various plant species. This success is assumed (and appears) to decline exponentially. Ease of finding or 'appearance' will be influenced by size, colour, shape, morphological state/phenological changes, surveyor experience, and even site topography and micro-topography. If the surveyor knows what species to look for in terms of conservation interest and what types of communities they may be found in, then it is more likely that these species will be found towards the beginning of a survey rather than at the end, even if very rare. This is contrary to the assumption used in the model.

Again, some of the recording curves show smooth, exponential trends, but with sudden pulses. It is presumed that these pulses are when the surveyor has entered or even crossed, a new plant community. Indeed the discrepancy between actual and predicted totals in some of the cases may relate to a surveyor missing small but species-rich communities. This would adversely affect the model.

With three or more parameters, a good fit will be produced for the accumulation data. Even if a wood had an infinite number of species, a good fit can be produced which by simple inspection looks close to the actual number recorded. It is therefore surprising that the predicted numbers of Models 2 are similar to the actual number of species found. However, certain trends are apparent.

Equations can be fitted to cumulative species data, and they can help predict the length of time required for survey work. In this way, the technique may aid cost/benefit analysis and so be used to justify and refine survey techniques. This may be useful in the preparation of contract tenders for consultancy and research work, and also perhaps for highlighting the short-comings of less intensive surveys. For example, if the aim of a survey is to estimate the total number of species then the results presented in this paper would indicate that only a small amount of data may be needed to give reasonable estimates. However, if the aim of a survey is record all the species, then the findings presented in this study suggest that a considerable amount of time and effort is required.

Other aspects of site survey are also relevant to efficiency and effectiveness. The value of surveying at different time of year is well-known (Kirby *et al.*, 1986). Ideally, a site should be visited at least three times in spring, early and late summer. Furthermore, continuous monitoring of a site may also produce dividends. Ecclesall Woods had 176 species recorded in the 1987 survey, but the total number of species now (1997) recorded is 236. This total includes considerable 'background noise' of urban exotics, but this is not the whole story.

Whilst the present model is felt to be useful, further refinements are clearly needed. One approach would be to record all the species in every time interval, and to analyse them using the computer programme EstiMateS (Colwell, 1995). The same approach could be used to improve the estimates produced by the model presented in this paper with an exponential decline in species abundance.

CHAPTER 5

5. THE RESPONSE OF BLUEBELL (*HYACINTHOIDES NON-SCRIPTA*) TO SEASONAL DIFFERENCES BETWEEN YEARS AND WOODLAND MANAGEMENT

5.1 Introduction

The response of *Hyacinthoides non-scripta* to coppicing has been well documented. Ford and Newbould (1977) recorded a peak in the biomass of *Hyacinthoides non-scripta* at two years after coppicing and again at nine years after coppicing. The intervening period may have been the result of dominance by *Pteridium aquilinum* and *Rubus fruticosus*. However, site effects unaccounted for because of the experimental design may also have influenced the results found.

Hyacinthoides non-scripta is a shade evader rather than a species tolerant of shading, such as *Lamiastrum galeobdolon* and *Oxalis acetosella* (Grabham & Packham, 1983; Packham & Willis, 1976). The greater part of the growth of *Hyacinthoides non-scripta* occurs before the canopy expands. The plant is not found where the mean light intensity between April and June falls below a tenth of full daylight, and hence it is generally absent from coniferous woodland and beech woods (Blackman & Rutter, 1954). *Hyacinthoides non-scripta* is also sensitive to soil moisture. If too wet *Anemone nemorosa* has a competitive advantage to such an extent it may exclude *Hyacinthoides non-scripta* (Shirreffs, 1985). *Hyacinthoides non-scripta* is regarded as a poor competitor having adapted a strategy mainly against the stressful conditions of the shaded woodland floor and also possesses some characteristics of ruderal species (Grime, 1988). *Hyacinthoides non-scripta* is able to survive and produce large bulbs under *Pteridium aquilinum*. The bulbs avoid the *Pteridium aquilinum* rhizomes by moving further down the soil profile (Grabham & Packham, 1983). *Hyacinthoides non-scripta*, appears to have little answer to the vigorous growth of *Rubus fruticosus*. However, conditions for *Hyacinthoides non-scripta* do improve once the canopy returns to its former cover or grazing occurs from sheep or deer (Kirby, 1997).

Hyacinthoides non-scripta is more successful on light soils as opposed to heavy ones (Knight, 1964). This is most likely to be the result of impeded drainage as well as difficulties posing the bulbs in moving down the soil profile.

Hyacinthoides non-scripta is a perennial and exhibits both sexual and vegetative reproduction (Corbet, 1998; Grabham & Packham, 1983; Merryweather & Fitter, 1998). Raceme architecture and overall resource supply are determined by the performance of the bulb months before flowering begins (Merryweather & Fitter, 1998).

This study aims to examine in detail the response of *Hyacinthoides non-scripta* to coppicing as well as be able to compensate for seasonal differences between different years and also take into account the influence of different plots in an experimental design. The experiment was set up with limited resources primarily to investigate the feasibility of reintroducing coppicing into urban-fringe woodland within the Sheffield City area. *Hyacinthoides non-scripta* was the main species of interest due to its popularity with local residents. South Yorkshire Forest had also expressed an interest in the project in terms of predicting when seed should be collected from existing *Hyacinthoides non-scripta* populations in order to introduce this species to new sites. In order to find answers to these problems an attempt at answering the question raised by Corbet (1998) is required – 'Is seed number in an insect pollinated plant limited by resources or pollinators?' In terms of this project, the effects of year to year climatic differences need to be separated from the effects of coppicing and plot effects such as soil wetness. Once these factors can be accounted for then predictions can be made as to both the response of *Hyacinthoides non-scripta* to woodland management as well as seasonal differences with the possible link to climate change.

Much of this study has been already been published (Vickers and Rotherham, 2000).

5.2 Materials and methods

5.2.1 Study site

The study site chosen was Ecclesall Woods, which is situated five miles west of Sheffield City centre (SK 325 825) and is approximately 100 ha in size. The wood has a very detailed and well documented past historical record. Evidence of former coppice management such as charcoal hearths, whitecoal kilns, wood banks, ditches, drains and boundary features, are present in abundance.

The site contains mainly mature broadleaf and mixed conifer and broadleaf woodland, with some areas dominated by *Acer pseudoplatanus*, *Castanea sativa*, *Fagus sylvatica* or *Quercus petraea* x *robur.* Three adjacent plots of woodland of approximately 50 × 50m have been coppiced in an area mainly dominated *by Acer pseudoplatanus*, with an under-storey containing some *Corylus avellana* and *Ilex aquifolium* with the ground flora dominated by *Hyacinthoides non-scripta.*

The ground vegetation in the coppiced plots at present is very patchy with some areas dominated by *Hyacinthoides non-scripta*, other areas consisting mainly of ferns – *Dryopteris filix-mas*, *Dryopteris dilatata* or *Pteridium aquilinum*, while other areas consist mainly of *Holcus mollis*, *Rubus fruticosus* or *Stellaria holostea.* A flushed area also existed in plots 1 and plot 2, which contained abundant *Ranunculus repens.*

5.2.2 Data collection

Three experimental plots were set up in 1994 each of approximately 50 × 50 m in area (0.25 ha). These areas lie side by side and were each coppiced in consecutive years. Data were collected on the density of *Hyacinthoides non-scripta* stems, the height of the stems and the number of fertile and infertile capsules on each stem. The experimental design is summarised in Table 5.1. Plots were coppiced in the winter of each year and data collected in the following spring (June) of each year.

Table 5.1. The experimental design used. Numbers refer to the 1st, 2nd, 3rd, 4th and 5th years after coppicing. (- no data recorded)

Year	Plot 1	Plot 2	Plot 3	Plot 4
1995	2	1	uncoppiced	-
1996	3	2	1	-
1997	4	3	2	Uncoppiced
1998	5	4	3	Uncoppiced

A total of ten 1 m^2 random quadrats were placed within patches of *Hyacinthoides non-scripta* within each plot in each year. The number of *Hyacinthoides non-scripta* stems were counted in each quadrat. All the stems in a quadrat were picked and then ten selected at random. The lengths of the stems were measured and the number of fertile and infertile capsules recorded.

5.3 Data Analysis

The design of the experiment for the recording of *Hyacinthoides non-scripta* densities, heights of stems and number of flowers can be likened to an incomplete, non-random and unbalanced Latin square. The plot, plot by year interaction and plot by time since coppiced interaction were treated as random effects because the plots had been randomly re-sampled each year. Time since coppiced and year, were treated as fixed effects. The number of fertile and infertile capsules per stem and the number of stems per quadrat were treated as having the classical assumptions about counts data i.e. a poisson distribution. Anscombe's variation of the arcsine square root transformation (Zar, 1996) was used to normalise the data for proportion of fertile capsules. This is given below.

$$P' = \arcsin\sqrt{\frac{x+\frac{3}{8}}{n+\frac{3}{4}}} \qquad (5.1)$$

Where x = number of fertile capsules
n = number of capsules

The density of stems and the number of capsules per stem were analysed using GLMMIX, a SAS macro for the analysis of generalised linear mixed models (Little *et al.*, 1996). A poisson error term was assumed and a log-link function used as well as an over-dispersion parameter. The over-dispersion parameter compensates for the counts data having a mixed poisson distribution. The data for height of stems and proportion of fertile flowers were analysed using PROC MIXED (SAS Institute, 1996). Correlograms and partial correlograms were used to detect for any auto-correlation in the residuals produced by the GLMMs.

The estimated values produced by the GLMMs were entered as dependent variables and data for the weather, for the winter and spring of each year, entered as independent variables. A combined forward backward stepwise regression procedure was used to find the combination of weather variables that best explained the predicted values for each year.

5.4 Results

5.4.1 Effects of year and time since coppiced

The estimates for the random effects of the GLMMs are given in Table 5.2, and the type III F tests shown in Table 5.3. The adjusted means and standard errors are shown in Figure 5.1. None of the random effects were found to be significant. For the fixed effects, 'year' can be seen to be have a much greater effect on the variables recorded than time since coppiced.

Table 5.2. Covariance parameter estimates for the random variables produced by the GLMMs.

Dependent variable	Random effect	Estimate	Z	*P*
Density of stems	Plot	0.022	0.45	0.65
	Plot × Year	0.000	.	.
	Plot × Time	0.067	0.91	0.36
Height of stems	Plot	0.000	.	.
	Plot × Year	0.000	.	.
	Plot × Time	0.034	1.48	0.14
Number of capsules	Plot	0.008	0.68	0.49
	Plot × Year	.	.	.
	Plot × Time	0.092	1.00	0.32
Proportion of fertile capsules	Plot	0.004	0.80	0.42
	Plot × Year	0.001	0.71	0.48
	Plot × Time	0.000	.	.

Table 5.3. Type III F tests for the fixed effects produced by the GLMMs. 'Year' refers to the actual year (1995-1998) and 'Time' refers to time (years) since coppiced. (NDF Numerator Degrees of Freedom, DDF Denominator Degrees of Freedom)

Dependent variable	Fixed effect	NDF	DDF	Type III F	*P*
Density of	Year	3	4.40	3.46	0.120
Stems	Time	5	3.15	2.63	0.220
Height of	Year	3	6.89	31.24	< 0.001
Stems	Time	5	4.95	2.13	0.214
Number of	Year	3	4.78	14.29	0.008
Capsules	Time	5	3.39	1.49	0.379
Proportion of	Year	3	2.88	20.87	0.018
Fertile capsules	Time	5	2.83	2.37	0.263

The estimates produced from the GLMMs for each year show almost identical patterns with 1995 producing the lowest estimated values and 1997 the highest. The estimated values for the years 1996 and 1997 have intermediate values. The exception to this is the density of stems which has estimated values for 1996, 1997, and 1998 that are all very similar. The estimated values for the density of stems, height of stems and the number of capsules per stem all show very similar trends according to the time since coppiced. All show a drop in the first year after coppicing. The proportion of fertile capsules shows a very different response. This variable peaks at the second and third years after coppicing. However, it should be remembered that the effect of coppicing was not found to be significant. This can probably be attributed to the fact that *Hyacinthoides non-scrita* avoids shade rather than tolerates it. The effect of a canopy may only suppress competitors instead of having a direct effect on this species.

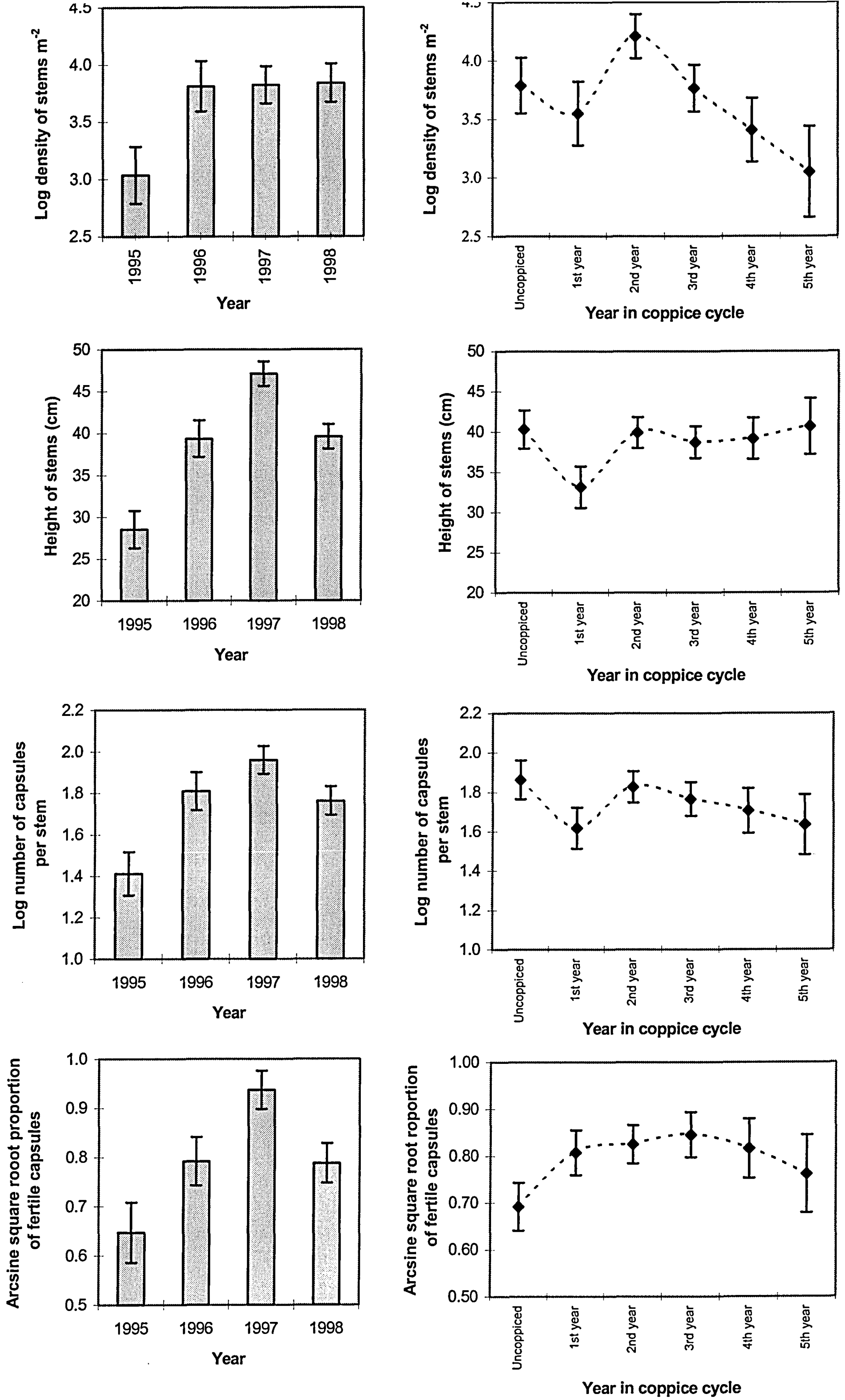


Figure 5.1. Adjusted mean values and standard errors for the density of stems, height of stems number of capsules per stem, and the proportion of fertile capsules per stem produced by the GLMMs.

Multiple comparison tests using Tukey-Kramer tests on the adjusted means to test for significant differences between them (Table 5.4).

Table 5.4. Contrasts between adjusted mean values for the year and time since coppiced effects. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Probabilities are for Tukey-Kramer adjusted P values.

Effect	Contrast	Density of stems		Height of Stems		Number of capsules per stem		Proportion of Fertile capsules	
		DF	*t*	DF	*t*	DF	*t*	DF	*t*
Year	1995 - 1996	3.24	-2.89	4.92	-3.80 *	3.59	-4.12 *	4.71	-3.81
Year	1995 - 1997	3.82	-2.79	4.99	-6.50 **	4.71	-4.93 *	4.20	-5.66 *
Year	1995 - 1998	4.99	-2.63	5.37	-3.80 *	5.89	-2.94	3.71	-2.18
Year	1996 - 1997	3.17	-0.05	5.04	-2.84	3.78	-1.56	4.56	-3.78
Year	1996 - 1998	4.40	0.10	5.44	-0.09	5.28	0.46	5.07	0.07
Year	1997 - 1998	30.88	-0.11	397	7.82 ***	81.08	4.09 *	3.54	4.59
Time	0 - 1	3.89	0.69	4.91	2.05	3.76	1.98	4.33	-2.41
Time	0 - 2	3.49	-1.45	4.90	0.13	4.60	0.31	4.30	-2.56
Time	0 - 3	3.62	-0.07	4.84	0.51	4.54	0.76	2.90	-2.34
Time	0 - 4	3.70	1.04	4.84	0.33	4.14	1.01	2.52	-1.48
Time	0 - 5	4.48	1.60	5.03	-0.08	4.75	1.21	2.87	-0.67
Time	1 - 2	3.54	-2.23	4.94	-2.21	3.70	-1.95	4.38	-0.42
Time	1 - 3	3.83	-0.65	4.94	-1.63	4.42	-1.17	5.15	-0.68
Time	1 - 4	4.41	0.34	4.94	-1.50	4.94	-0.55	4.65	-0.11
Time	1 - 5	5.05	0.97	5.11	-1.61	5.49	0.08	4.72	0.47
Time	2 - 3	2.74	1.75	4.99	0.43	3.25	0.67	4.02	-0.53
Time	2 - 4	3.63	2.49	4.99	0.22	4.30	0.99	5.05	0.18
Time	2 - 5	5.03	2.68	5.24	-0.18	5.36	1.20	5.14	0.82
Time	3 - 4	3.22	1.24	4.94	-0.15	3.38	0.56	4.01	0.71
Time	3 - 5	4.54	1.81	5.11	-0.51	4.41	0.91	4.86	1.33
Time	4 - 5	3.98	0.90	5.07	-0.38	3.44	0.52	3.22	1.01

The only pair-wise comparisons found to be significant involved the effect for year. The main differences can be seen to occur between 1995 and 1997 and between 1997 and 1998. The height of the stems showed the greatest differences.

5.4.2 Testing for auto-correlation in the GLMMS

The correlation between observations of lag k apart can be calculated by the equation

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \qquad (5.2)$$

The standard error is equal to 1/√N. So that if this is produced for a random time series then 95% of the coefficients will lie between ±2/√N (Diggle, 1992).

The data presented in this study represent a short time series but with many replicates. If the number of years is used for N then ±2/√N equals 1. Since a large number of samples has been taken the mean of the number of samples was used to give approximate confidence intervals (Keith Jones pers.comm.). The correlograms and partial correlograms calculated from the means of the residuals produced by the GLMMs are shown in Figure 5.2. Partial correlograms are constructed by successively fitting auto-regressive processes of order 1,2..., and, at each stage, defining the partial auto-correlation coefficient, α_k to be the estimate of the final auto-regressive coefficent (Diggle, 1992). The values were calculated using SAS. If there is a sharp cut-off in the correlogram then it indicates the presence of a moving average process, but if not and instead there is a sharp cut-off in the partial correlogram, then it indicates the presence of an auto-regressive process (Diggle, 1992). The correlograms and partial correlograms displayed in Figure 5.2 show some evidence that the density of stems are positively auto-correlated, but this is not significant. There appears to be little difference between the two types of correlograms, so that no conclusions can be made as to the types of process that may be occurring i.e. whether or not a moving average model may fit this data better than a first order aut-regressive model.

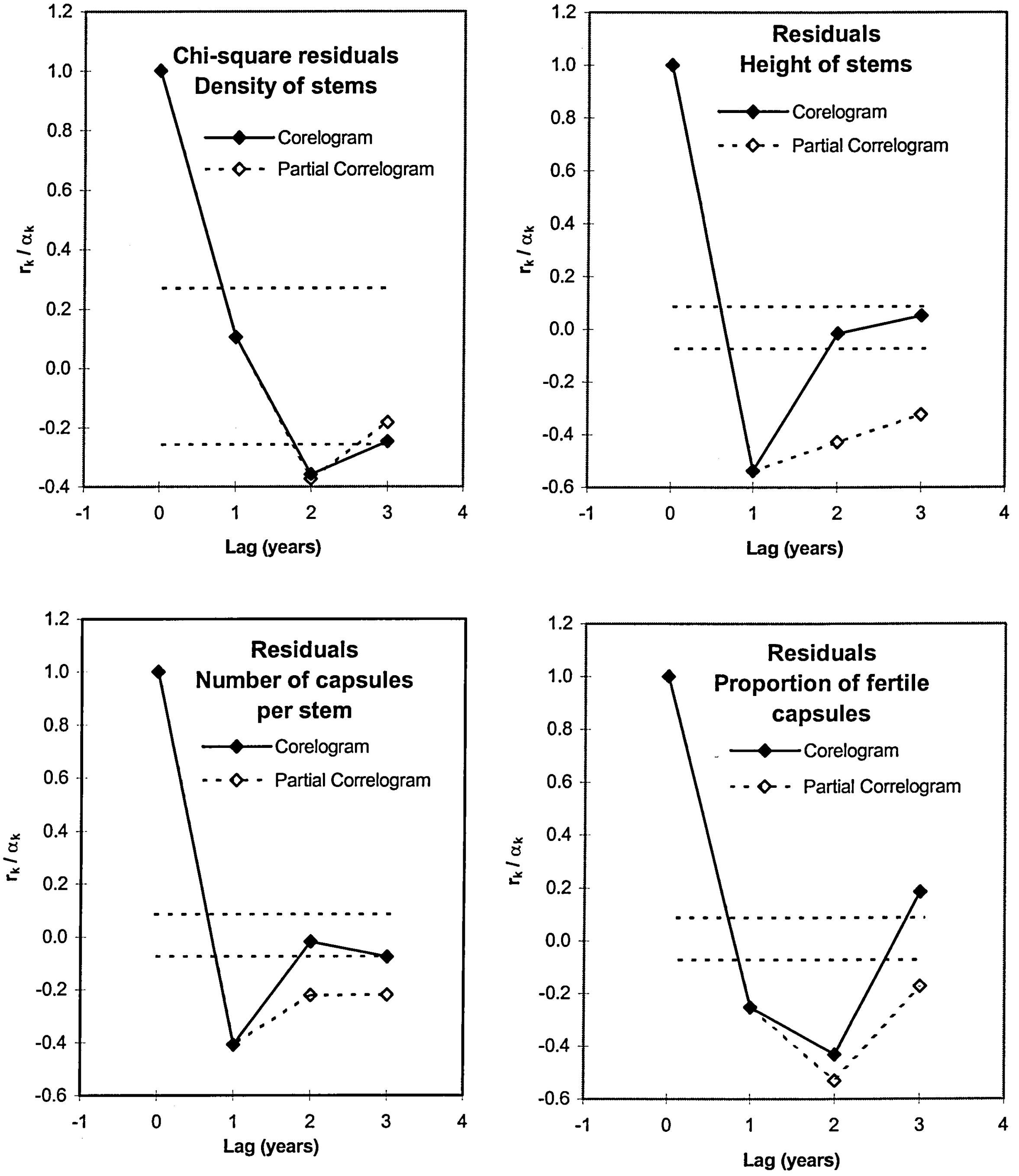


Figure 5.2. Correlograms of the residuals produced by the GLMMs. r_k refers to the correlograms and α_k to the partial correlogram.

5.4.3 Seasonal differences between years and the estimates produced from the GLMMs

Weather data are shown in Table 5.5 for the months of January to May for the four years of the study period. According to Grabham and Packham (1983) January, February and March represents the main period of fresh weight increase, and April to mid June represents the main flowering period for bluebells. The analysis of the weather data has therefore been split into winter (January, February and March) and spring (April and May) corresponding to these two times of year. June was not included since most of the data were collected at the beginning of this month. A correlation matrix of the weather variables is shown in Table 5.6.

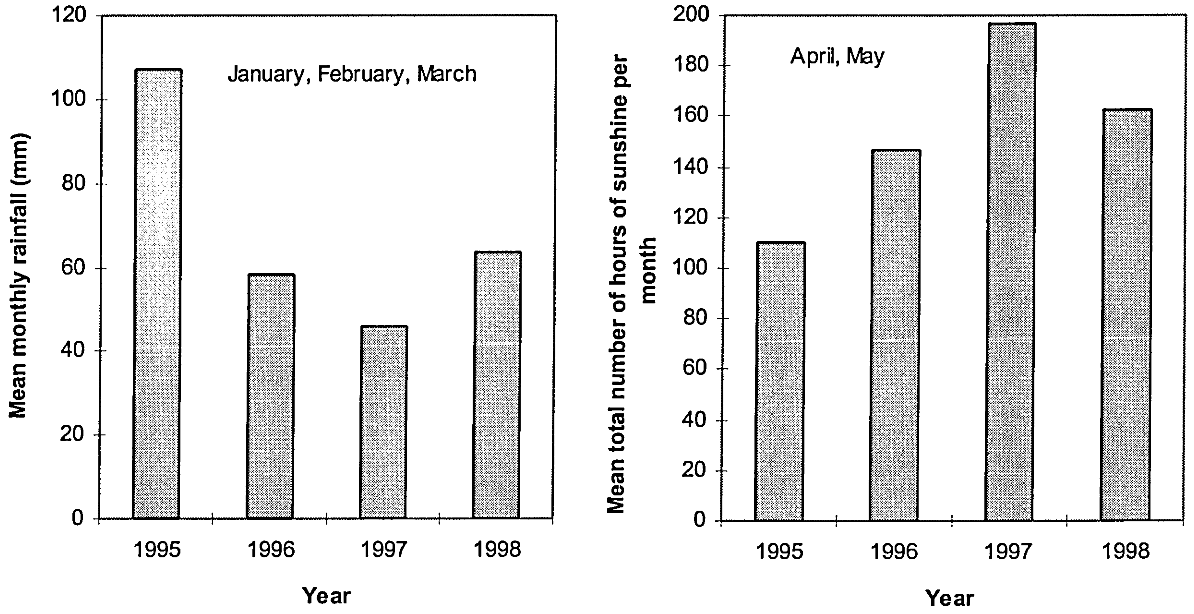


Figure 5.3. Weather data for mean monthly rainfall during the winter and number of hours of sunshine during the spring.

The graphs displayed in Figure 5.3 show some very similar trends to those displayed in Figure 5.1. The winter with the least amount of rainfall was in 1997, and the highest in 1995. This corresponds with the estimates produced by the GLMMs. These showed the height of the stems, the number of capsules per stem and the proportion of fertile capsules per stem to all peak in 1997 and have their minimum in 1995. A similar pattern can also be seen for the number of hours of sunshine per month during the spring of each year. The correlation matrix of Table 5.6 shows some variables to be highly correlated with each other. Some correlations are what would be expected, such as winter snowfall, winter frost and winter temperature, and also spring snowfall, spring frost, and spring temperature. However, there are some unexpected correlations between spring sunshine and winter rainfall, winter sunshine and spring snowfall, and winter sunshine and spring frost. This does make it more difficult to separate out the effects of winter rainfall and spring sunshine in any analysis.

The results of the stepwise regression analysis on the estimates produced by the GLMMs fitted to the weather data, for the two time periods, are given in Table 5.7. The only variable picked out by the stepwise regression procedure for the density of stems was winter rainfall, although this was not significant at the $P < 0.05$ level. The height of the stems can be seen to be significantly related to winter rainfall. The number of capsules can be seen be significantly related to the winter rainfall and also possibly spring rainfall. The proportion of fertile capsules can be seen to be significantly related to the spring hours of sunshine. The number of days with snow lying on the ground was also picked out by the stepwise procedure but was not significant at the 5 % level, for the proportion of fertile capsules.

Table 5.5. Weather data provided by Weston Park Museum.

Year	Month	Mean temp °C	Rainfall (mm)	Hours of sunshine	Days with snow	No. of days with air frost
1995	January	4.1	157.7	35.8	6	10
	February	6.2	99.2	63.0	0	0
	March	5.4	63.9	172.7	4	7
	April	9.1	24.6	17.7	0	1
	May	11.8	48.1	203.7	0	0
1996	January	3.6	67.5	5.8	9	9
	February	2.5	75.1	54.3	13	16
	March	3.8	52.9	38.6	3	5
	April	8.9	37.7	116.0	1	4
	May	9.2	39.1	178.0	0	0
1997	January	2.7	8.5	51.4	10	15
	February	6.7	106.5	81.6	0	1
	March	8.6	22.0	152.1	0	1
	April	9.1	28.2	142.0	0	1
	May	11.5	77.5	252.3	0	0
1998	January	4.8	88.5	46.3	0	5
	February	8.2	5.4	106.7	0	4
	March	8.0	97.4	91.9	0	2
	April	7.7	117.9	125.7	0	2
	May	12.8	26.7	199.3	0	0

Table 5.6. Correlation matrix of weather variables: Winter refers to January, February, March; Spring refers to April, May. ***P* < 0.05 bold type, P < 0.01 bold and underlined**.

	Winter temp	Winter rain	Winter Sun	Winter snow	Winter Frost	Spring temp	Spring rain	Spring Sun	Spring snow
Winter rain	-0.17								
Winter sun	0.81	0.11							
Winter snow	**-0.98**	-0.01	-0.79						
Winter frost	**-0.98**	-0.05	-0.85	**0.99**					
Spring temp	0.82	0.24	**0.98**	-0.84	-0.90				
Spring rain	0.82	-0.51	0.34	-0.77	-0.71	0.35			
Spring sun	0.36	**-0.96**	0.18	-0.17	-0.17	0.03	0.55		
Spring snow	-0.89	-0.13	**-0.98**	0.89	0.94	**-0.99**	-0.46	-0.14	
Spring frost	-0.70	-0.20	**-0.99**	0.69	0.76	**-0.96**	-0.18	-0.09	0.94

Table 5.7. Combined forward backward stepwise regression for the estimates produced by the GLMMs and the weather data.

Dependent variable	*Independent* Variable	Estimate	*P*	Partial R^2
Density of Stems	Intercept	4.632	0.003	
log (n+1)	Winter rain	-0.014	0.061	0.882
Height of stems	Intercept	59.164	0.001	
Stems : cm	Winter rain	-0.292	0.010	0.980
Number of Capsules	Intercept	2.483	0.007	
log (n+1)	Winter rain	-0.009	0.014	0.989
	Spring rain	-0.002	0.121	0.011
Proportion of fertile	Intercept	0.103	0.136	
Capsules	Spring sun	0.004	0.020	0.936
arcsine square root	Winter snow	0.011	0.080	0.063

5.5 Discussion

An attempt has been made to quantify the response of *Hyacinthoides non-scripta* to weather variables and woodland management. The main effect shown by this study has been that of differences between years rather than that produced by the coppicing. The main effect of the coppicing appears to be a peak in the second year after coppicing for the density of stems and the proportion of fertile capsules, and a drop in the first year after coppicing for the height of stems and the number of capsules. The decrease in the height of stems and the number of capsules may be in part related to soil moisture conditions. The water table goes up after a site has been coppiced due to the reduced amount of transpiration from trees (Mitchell *et al.*, 1992). This would also be in agreement with the reduced height of stems and a drop in the number of capsules during the year with the wettest winter. This suggests that the growth of *Hyacinthoides non-scripta* is severely limited by the soil moisture content, which may explain the competitive advantage of *Anemone nemorosa* described by Shirreffs (1985) when the two species are growing together in wet conditions.

The decline in the density of stems after the second year in coppicing was probably the result of *Rubus fruticosus* and *Pteridium aquilinum* becoming dominant. The site at present is now dominated by these two species although *Hyacinthoides non-scripta* is becoming established again under the shade of the coppice regrowth. The increase in the proportion of fertile capsules during the second and third years after coppicing is probably due to the increased amount of sunlight reaching the coppice before the site is dominated by *Rubus fruticosus* and *Pteridium aquilinum*, allowing the pollination of the flowers by bees (Corbet, 1998). The increase in the proportion of fertile capsules in 1999 can be attributed to the high number of hours of sunshine during the spring. The winter snowfall may also be important. Snow can help prevent fungal infections of a number of hibernating insects such as bees.

Speculations can also be made as to how these results affect the colonisation rate of *Hyacinthoides non-scripta*. The height of stems and the amount of seed produced will be directly related to the rate of spread of this species and hence to recolonisation within a particular woodland. The predicted effects of global warming are for warmer wetter winters and cooler wetter springs and summers. If this is true then the growth and hence the competitiveness of *Hyacinthoides non-scripta* is likely to be reduced in the future.

The collection of seed has become a common practice for introducing *Hyacinthoides non-scripta* to newly created woodlands, especially those close to urban areas. This study should help improve the efficiency of seed collection. Seed should be collected after a cold, dry winter followed by a sunny spring. The optimum time in the coppice cycle would appear to be in the second spring after coppicing, when both the density of stems and the proportion of fertile stems are high.

CHAPTER 6

6. THE INFLUENCE OF CANOPY COVER AND OTHER FACTORS UPON VEGETATION COMPOSITION WITHIN GLEN TANAR NATIONAL NATURE RESERVE

6.1 Introduction

This study is based on work on pine regeneration initially reported by Vickers and Palmer (2000). Whereas that paper focused mainly on factors influencing pine regeneration, the present study is concerned with factors affecting vegetation composition and the identification of different plant communities and indicator species.

These data provide a useful comparison to the research interrogated in Chapters 2 and 3. The pine forests of Scotland represent the extreme end of a gradient from woodlands of warmer drier lowland climates in the south east to woodlands of cooler and wetter climates in the north and west. In particular, if species can be seen to behaving in a similar way in the data from Scotland compared with information collected elsewhere in the UK, it is suggested that these lists of potential indicator species maybe of use throughout the north and west of the UK. However, if the species found to be associated with semi-natural pine forest are found to be considerably different to those found else where in the UK, then it is likely that the lists should be used with great caution. In particular, this would suggest that further research would be required to determine lists for different geographic areas.

6.1.2 Characteristic species of Caledonian forest

The native Caledonian pinewoods of Scotland, dominated by Scots pine (*Pinus sylvestris* L.), constitute a small remnant of a much larger forest, possibly covering 1.5 million ha. The total area of the pinewoods today is now estimated to have been reduced to about one percent of this value, approximately 16 000 ha (Forestry Commission, 1994). The decline of the forest has mainly been the result of hundreds of years of deforestation by people and a changed in climatic conditions, such as increased precipitation (Dickson, 1992; Bennett, 1994).

Open stands of Scots pine are considered to have great aesthetic appeal, yet other stand types may exist of equal age but with a much higher tree density. High density stands probably result from fire or windthrow caused by gales (Steven and Carlisle, 1959). The density of trees can be seen to alter the ground vegetation composition to a large extent. Ground vegetation below open stands is usually dominated by heather (*Calluna vulgaris*) with mosses underneath the heather. The ground layer below dense stands is often dominated by bilberry (*Vaccinium myrtillus*) and wavy hair-grass (*Deschampsia flexuosa*). The native pine forest would have probably contained a mosaic of these different stand types, interspersed with birch (*Betula* spp.) and aspen (*Populus tremula*) (Steven and Carlisle, 1959; Rodwell and Cooper, 1995), and separated by rivers or bogs, which would have restricted the spread of fire (Vickers and Palmer, 2000).

The vascular plants of the Scottish pinewoods have often been referred to as a species poor assemblage with few characteristic species (Peterken, 1993; and Pitkin *et al.*, 1994). Out of sixty-nine plant species, from seventy-seven samples listed in the National Vegetation Classification (NVC) (Rodwell, 1991), only thirty-three are vascular plants (Pitkin *et al.*, 1994). The remaining species are mosses. The NVC table for the W18 category contains only five sub-communities.

The species most commonly referred to as being associated with Scots pine forest are *Goodyera repens*, *Linnaea borealis*, *Moneses uniflora,* the moss *Ptilium crista-castrensis* and the liverwort *Anastrophyllum hellerianum*. However, none of these occur exclusively with Scots pine (Pitkin *et al.*, 1994). A number of fungi and lichens have also been found to be associated with Caledonian pine (Orton, 1986). Several insect groups (*Hemiptera*, *Coleoptera* and *Diptera*) also appear to be good indicators of old pine forest stands (Young and Armstrong, 1994).

6.2 Materials and methods

6.2.1 Study site

Glen Tanar NNR is located approximately 10 km south of Aboyne in Aberdeenshire (National Grid Reference NO 470950). The main forest area of the reserve covers approximately 860 ha. The forest consists almost entirely of Scots pine. The stands of pine generally fall into three age categories - less than fifty years old, approximately 150 years old and approximately 250 years old (Steven and Carlisle, 1959). All sites studied had regenerated naturally. Juniper (*Juniperus communis*) grows throughout the forest, but is more abundant in the open semi-natural forest areas (Steven and Carlisle, 1959). The predominant soil type are podzols, mostly freely drained and with a shallow layer of raw humus. Some areas have much thicker peat deposits. Other types of soils include forest brown earths, peaty gleys and non-differentiated fluvial material (Vickers & Palmer 2000)

6.2.2 Data collection

The data collected was that used by Vickers and Palmer (2000). The field study was conducted during May and June 1996. Twenty-five plots of 50 m x 50 m were established in such a way as to cover as much of the variation as possible in mature tree density, soil wetness, slope, aspect and different types of historical disturbance within the site. In each plot the altitude, aspect (sixteen compass points), slope, nutrient status of the soil. Degrees of slope were derived from a geographical information system (GIS) model of the estate (P.J. Bacon, A. Webb, unpublished data). The precise locations of the plots were chosen on the ground with the aim of selecting areas that had an even mature tree density as well as being surrounded by a similar mature tree density.

In each plot five quadrats of 10 m x 10 m were set up to determine the densities of pine saplings within three height categories (< 1 m, 1-2 m and > 2 m). Two of the quadrats were placed at random, one selected to be in a gap, one selected under higher canopy cover and the fifth was placed in a ground vegetation type not represented by any of the first four quadrats. If any one of these conditions could not be met, then a quadrat was placed at random. For each quadrat the percentage canopy cover and soil wetness (poorly drained or not) were recorded. The canopy cover was estimated as the percentage of the total area of sky visible through the canopy above the plot hidden by the canopy. The percentage cover of all vascular plants and mosses was estimated by eye within each quadrat.

6.3 Data Analysis

Species scores were transformed by taking the log i.e. log (% cover +1). Canopy cover was arcsine squareroot transformed. The sine and cosine of the aspect were used to give values for easterliness and northerliness respectively and slope was log transformed. Canonical correspondence analysis (CCA) was used as a means of producing a direct gradient analysis, which was conducted using CANOCO (ter Braak & Smilaur, 1998).

The significance of the environmental data with the CCA axes (derived from the weighted averages of the species scores) were tested using a general linear mixed model. All variables were standardised by their standard deviation to give values with a mean zero and variance of one. Plot was included as a random effect. Satterthwaite approximated degrees of freedom were used to compensate for the unbalanced nature of the design of the experiment. This analysis was conducted using PROC MIXED in SAS (SAS Institute, 1996).

The sample scores which, were linear combinations of the environmental variables were used to produce the ordination plots. The first eight of these axes were used to produce a constrained ordination. The option 'focus on samples' was used in CANOCO so that the chi-square distances between sample scores were preserved. A number of methods were used: Average linkage, *k*-means, Ward's method and Flexible clustering ($\beta = -0.25$, -0.5), with the aim of finding the most easily interpretable solution.

The presence of indicator species at each level of the dendogram produced was determined using the computer package INDVAL (Dufrene & Legendre, 1997). The number of clusters used was determined by the maximum number of clusters that could be explained from a knowledge of the site and obvious differences in the species and environmental data.

6.4 Results

6.4.1 Ordinations

The eigenvalues, % species-environment data explained and correlations are shown in Table 6.1. The eigenvalues are fairly low, suggesting that species distributions may not follow clear unimodal patterns across the ordination. The main factors appear to be canopy cover and soil wetness on the first two axes and fire and scarification on the third and fourth axes. The results of the GLMMs on the first four CCA axes are presented in Table 6.2, and show a number of similarities with the correaltion coefficients produced by CANOCO. Canopy cover can be seen to be significantly associated with the first axis and soil wetness to the second axis. Altitude, slope and fire can be seen to be significantly associated with the third axis and slope and scarification are significant on the fourth axis. The random plot effect can be seen to significant on all the axes.

Table 6.1. Descriptions of the first four CCA axes produced by CANOCO and correlation coefficients of the environmental variables with the CCA axes.

	CCA axis 1	CCA axis 2	CCA axis 3	CCA axis 4
Eigenvalue	0.31	0.20	0.11	0.07
% variance explained	38.2	25.0	13.2	8.6
r species environment	0.82	0.76	0.55	0.56
Factor				
Canopy cover	-0.67 **	0.32 **	-0.08	-0.03
Soil wetness	0.49 **	0.44 **	-0.26 *	0.04
Altitude	-0.02	0.23	0.09	-0.17
Easterliness	-0.12	-0.24 *	0.09	0.08
Northerliness	0.26 *	0.28 *	-0.12	0.01
Slope	-0.10	-0.33 **	-0.11	0.03
Fire	0.38 **	0.10	0.35 **	-0.28 *
Windthrow	0.02	-0.12	-0.04	-0.08
Scarification	-0.03	-0.18	-0.10	0.49 **

Table 6.2. The solution for the fixed effects in the GLMMs. All variables were standardised to give mean zero and variance of one. Canopy cover was first arcsine square root transformed and the slope was log transformed. Northerliness and easterliness were calculated by using the radians of a compass. Fire, windthrow and scarified were all entered as dummy variables. $^*P < 0.05$, $^{**}P < 0.01$, $P^{***} < 0.001$

	CCA axis 1		CCA axis 2		CCA axis 3		CCA axis 4	
Effect	DF	Estimate	DF	Estimate	DF	Estimate	DF	Estimate
Canopy cover	111	**0.48** ***	109	0.08	113	0.04	116	0.03
Soil wetness	20.0	-0.23	18.7	**0.51** ***	16.6	-0.03	19.2	0.14
Altitude	16.4	0.18	15.1	-0.12	14.8	**0.38** *	16.3	0.29
Easterliness	15.8	0.03	14.5	-0.01	14.6	0.05	15.9	-0.06
Northerliness	16.1	-0.14	14.9	0.16	14.7	0.16	16.1	-0.32
Slope	17.1	-0.08	15.8	0.14	15.3	**-0.42** *	17.0	**-0.52** *
Fire	20.0	-0.19	18.6	-0.28	16.5	**0.48** *	19.1	0.01
Windthrow	16.4	-0.04	15.1	-0.14	14.9	0.10	16.4	-0.04
Scarified	17.7	-0.09	16.4	-0.26	15.5	0.31	17.4	**0.49** *
Plot (random)		0.35 **		0.35 **		0.56 **		0.68 **

Figure 6.1 shows the species associated with the wetter environments in the top left corner and those associated with the denser canopy cover on the right of the diagram. Species associated with the wetter sites include *Drosera rotundifolia*, *Narthecium ossifragum* and a number of *Sphagnum* species. Species associated with the denser shade include *Deschampsia flexuosa*, *Vaccinium myrtillus* and *Lycopodium annotinum*. Figure 6.2 shows some of the species associated with the recent fire towards the right of the ordination and those species associated with undisturbed habitat at the extreme left. The arrows for scarification and soil wetness are in the same direction making it difficult to see which species are associated with either environmental factor. Species associated with the burnt sites include *Chamerion angustifolium* and *Stellaria media*. Species associated with the scarified sites include *Potentilla erecta* and *Digitalis purpurea,* although this is not obvious from the ordination diagram. Species in the opposite direction to the arrows for the disturbance categories, include *Hypericum pulchrum*, *Mercurialis perennis*, and *Potentilla sterilis*.

The corresponding ordinations for the sample scores are shown in Figure 6.3 and Figure 6.4. These two figures show the results of grouping the scores using cluster analysis, which will be discussed later.

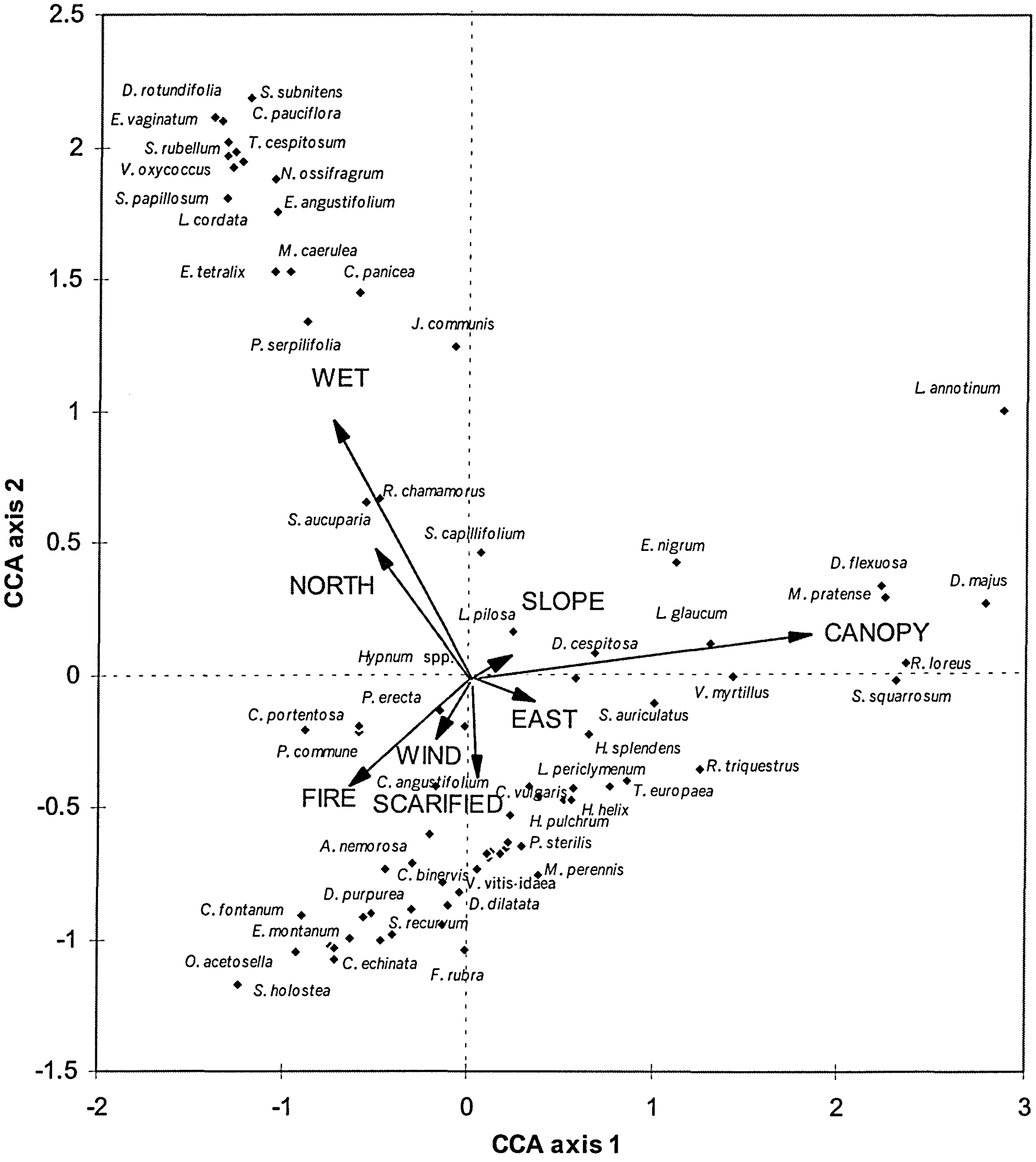


Figure 6.1. Ordination of the species scores on the first two axes produced by canonical correspondence analysis. Species names are given in full in Table 6.5.

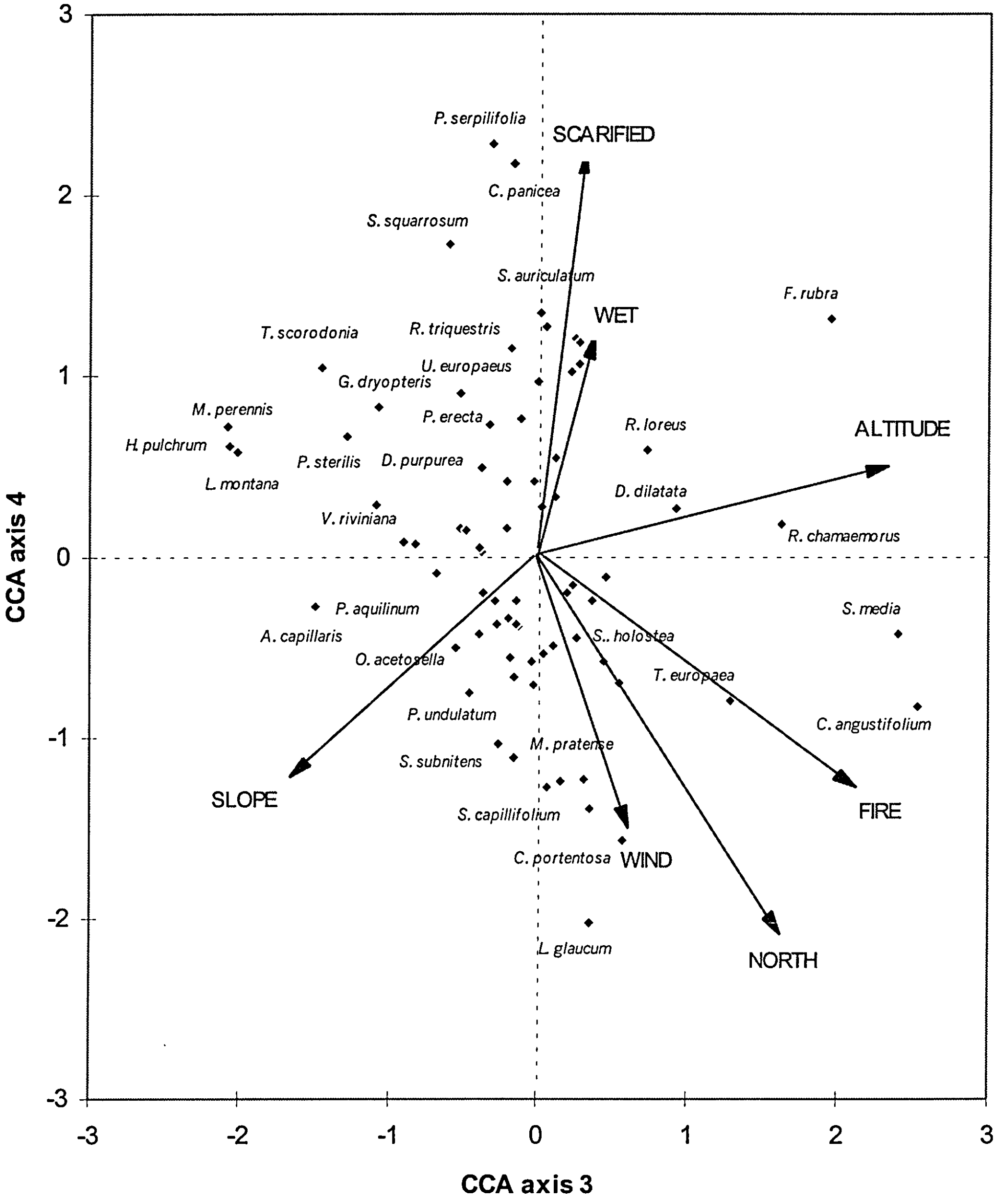


Figure 6.2. Ordination of the species scores on the second and third axes produced by canonical correspondence analysis.

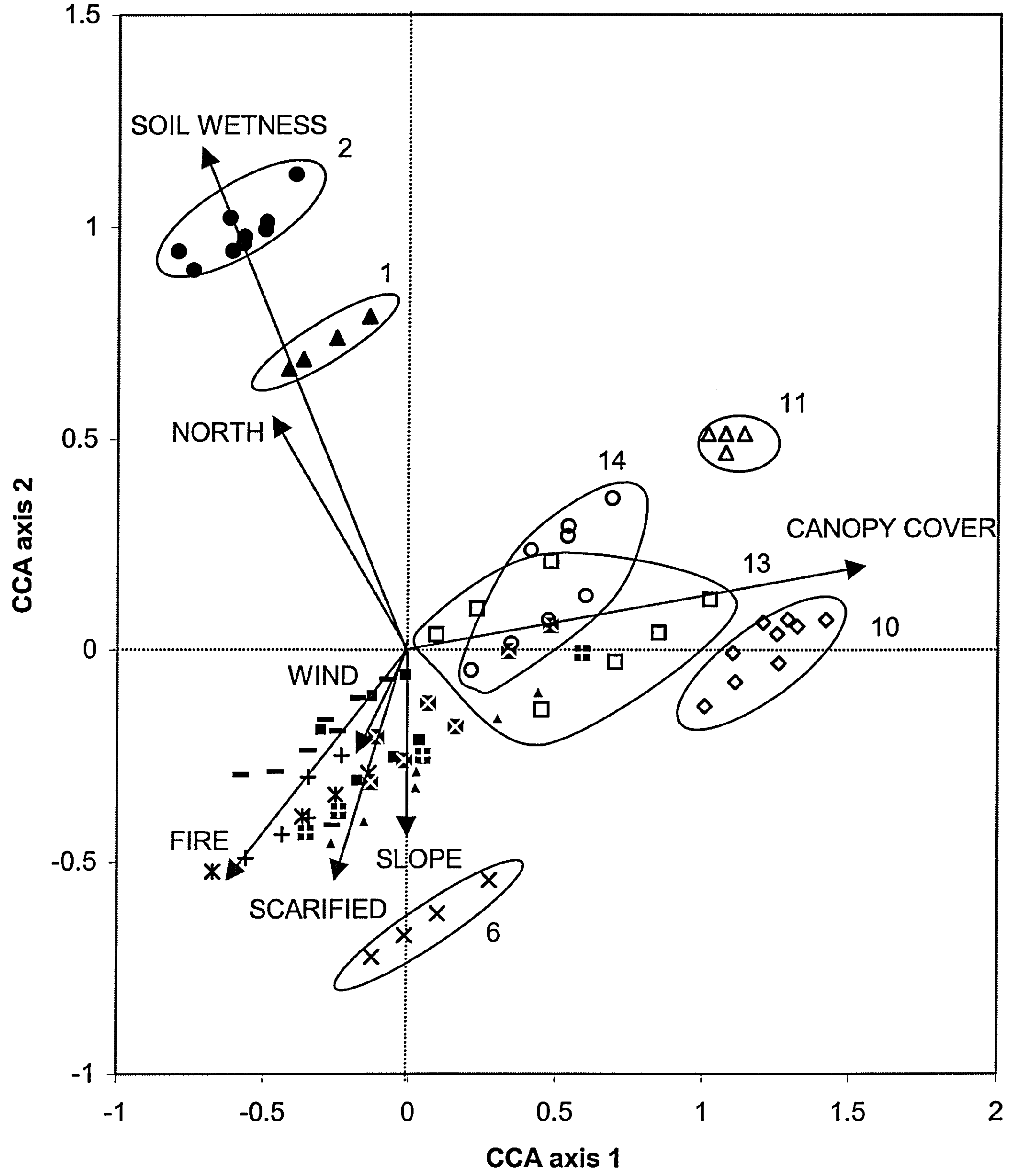


Figure 6.3. Sample scores of the first two axes produced by CCA that are a linear combination of the environmental variables. Numbers refer to the clusters given in Figure 6.5.

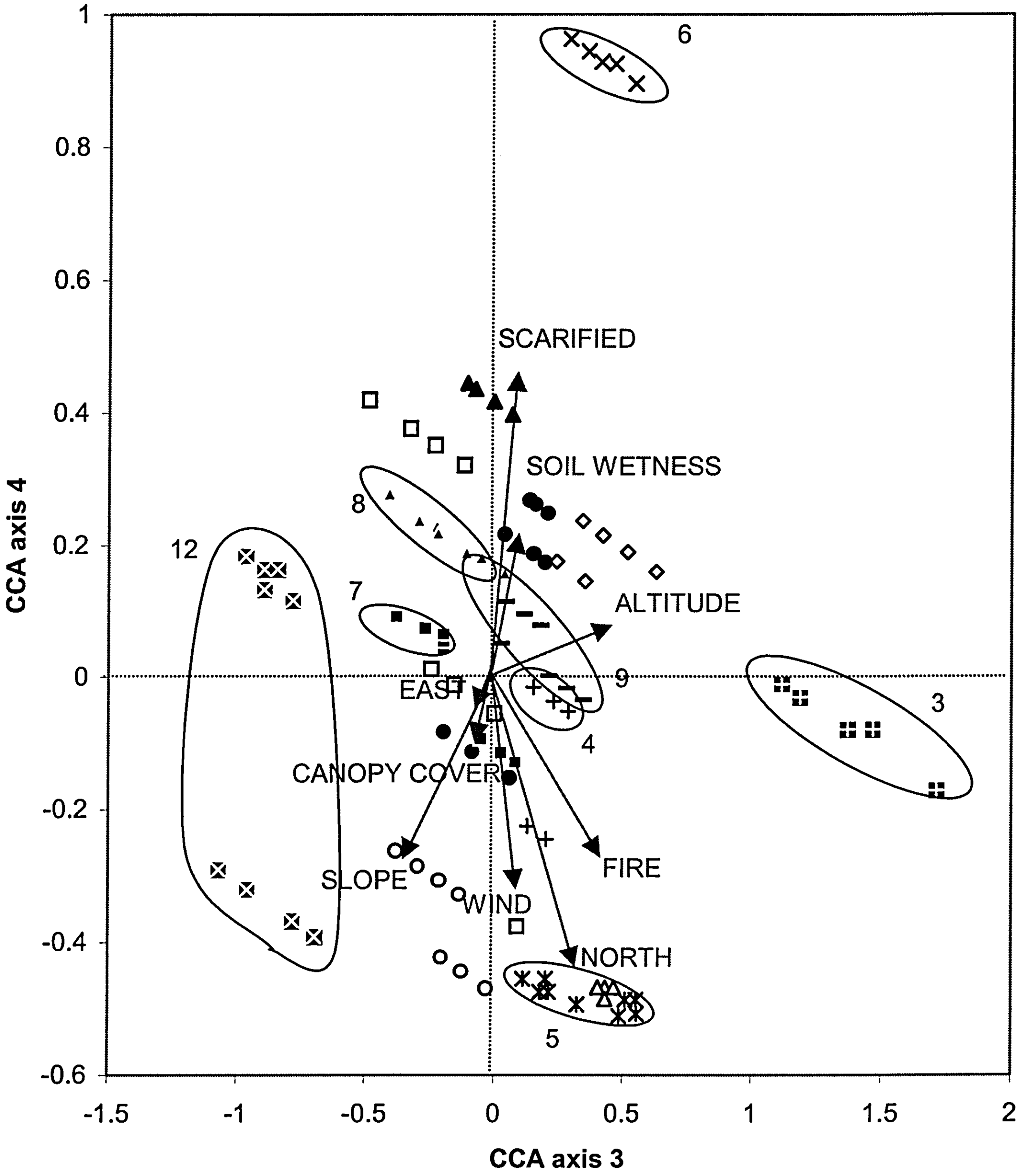


Figure 6.4. Sample scores of the second and third axes produced by CCA that are a linear combination of environmental variables. Numbers refer to the clusters given in Figure 6.5.

6.4.2 Regeneration across the ordination axes

The trends of the regenerating pine classes across the ordinations were investigated in the earlier paper by Vickers and Palmer (2000). However, using recently developed techniques the trends across the ordination can be more easily visually presented by the use of kriging and tested for significance using trend surface regression models. These diagrams are presented in Figure 6.5. Presence absence data was used to produce the semi-variograms (Table 6.3) and then perform the kriging. Backward stepwise logistic regression (Table 6.4) was used to test for significance. The estimates produced from the backward stepwise method have been used to produce the quadratic trend surface models shown in Figure 6.5.

Table 6.2. Parameter estimates for semi-variograms used for kriging which were used to produce the 3D terrain maps for the presence/absence for the three categories of pine regeneration and for the presence of regenerating *Sorbus aucuparia* in the quadrats.

Variable Presence/absence data	Model		Nugget	Sill	Range	Minor range	R^2
Pinus sylvestris < 1m	Isotropic	Spherical	0.09	0.25	0.98		0.87
Pinus sylvestris 1-2 m	Anisotropic	Spherical	0.17	0.52	5.82	13.82	0.31
Pinus sylvestris >2 m	Anisotropic	Spherical	0.16	0.52	11.99	31.81	0.24
Sorbus aucuparia	Isotropic	Exponential	0.04	0.16	0.63		0.87

Table 6.3. Quadratic trend surface models for the probability of recording the regeneration from the three categories of *Pinus sylvestris* and the probability of recording regenerating *Sorbus aucuparia*. Backward stepwise logistic regression was used to produce the estimates.

Dependent Variable	Model	Intercept	Axis 1	Axis 2	Axis 1^2	Axis 1 × 2	Axis 2^2
Pinus sylvestris	Full	1.04 **	-1.31 *	-0.56	-5.86 **	-2.37	0.61
<1 m	Stepwise	1.26 **	-1.29 *		-6.11 **	-2.02 *	
Pinus sylvestris	Full	-0.93 *	-1.63 **	-0.59	-1.50	-1.82	-1.30
1-2 m	Stepwise	-0.90 ***	-2.13 ***	-1.13 **			
Pinus sylvestris	Full	-1.18 **	-1.75 **	-0.74	-0.93	3.19	1.10
>2 m	Stepwise	-1.20 ***	-1.95 ***	-1.38 ***			
Sorbus aucuparia	Full	-0.45	2.09 **	0.15	-2.54 *	-7.23	-7.85 *
	Stepwise	-0.58	2.01 **		-2.61 *		-9.22 *

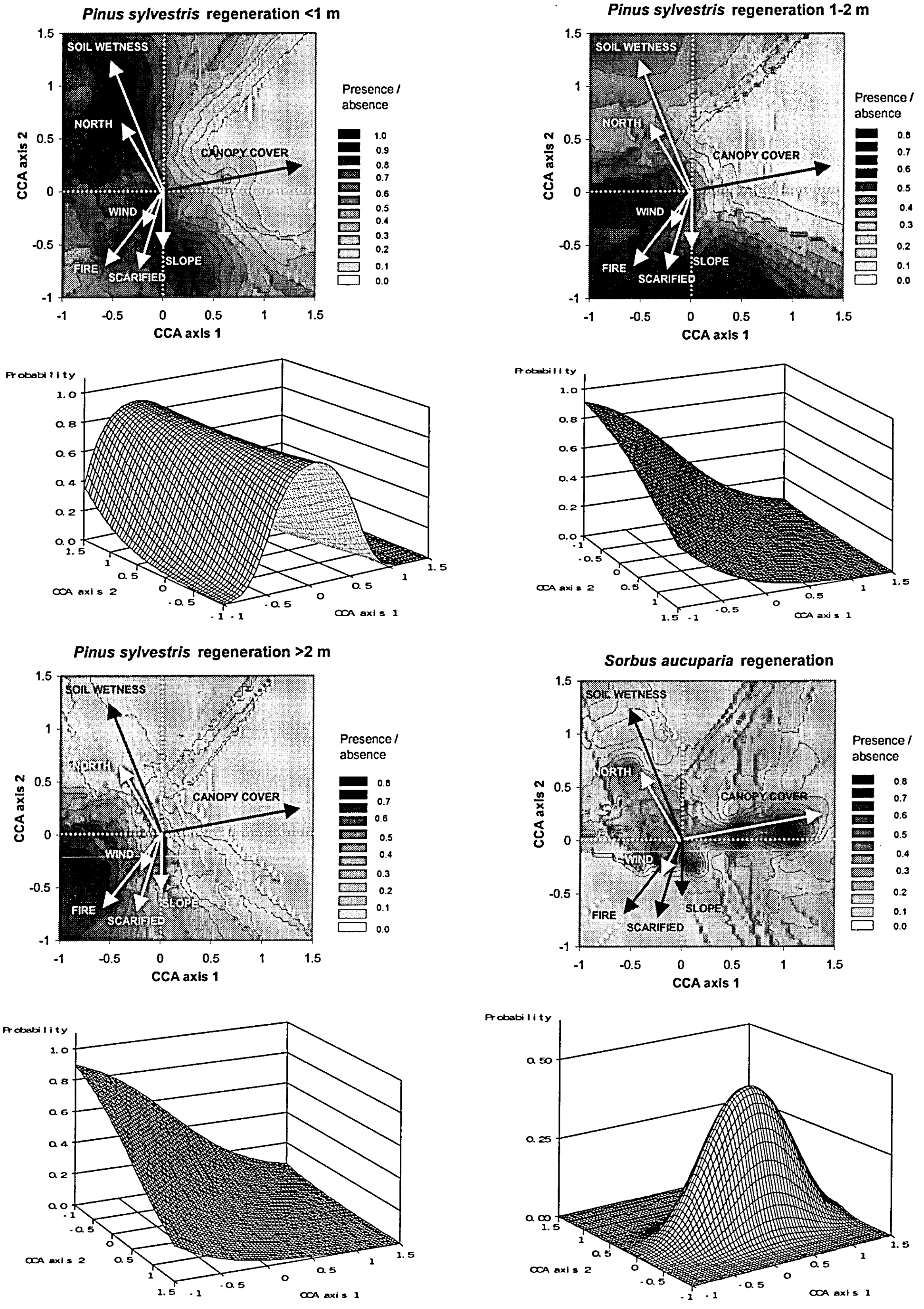


Figure 6.5. Trend surface maps produced by kriging for the presence/absence data for regeneration in the quadrats a also, underneath each graph, a trend surface map produced from a stepwise quadratic logistic regression equation. Axi has been reversed for graphs showing the quadratic logistic regression surfaces for pine of 1-2 m and >2 m.

6.4.3 Clustering

The first eight CCA axes, that were linear combinations of the environmental variables, were used in the cluster analysis. This number of axes appeared to give the dendogram which was most interpretable. Axes 5-8 still contained significant estimates produced by additional GLMMs. Single linkage, average linkage, complete linkage, flexible clustering (β = -0.25 and β = -0.5) and Ward's clustering method were used on squared euclidean distances. Single linkage suffered from chaining (Everitt, 1993) and average linkage and complete linkage appeared to give to greater importance to two plots that were very different to the rest i.e. the recently disturbed sites making it difficult to interpret. Flexible clustering and Ward's method gave very similar results, with Ward's method giving a slightly more easily interpretable dendogram. The dendogram produced by Ward's method is shown in Figure 6.6.

6.4.4 Indicator species

The structure of the dendogram produced by Ward's clustering method was entered into the program, INDVAL, together with the species cover score, to obtain the indicator values. The density of pine saplings, were added to the species data. The classical INDVAL index described by Dufrene and Legendre (1997) was used. The main two-way table giving the maximum of each species in each partition is shown in Table 6.4. The structure of the two-way table according to divisions of different habitat types is shown in Figure 6.7. The species characteristic of each partition are shown in Figure 6.8. An arbitrarily cut-off value of 20% was chosen. All species are shown whether they are significant or not (** $P < 0.01$). The NVC tables, keys and the computer program MATCH were used to classify the various partitions according to NVC types. Wet sites are characterised by *Eriophorum* species, *Erica tetralix* etc., which can be contrasted to the dry sites, which are characterised by *Vaccinium* and *Hypnum* species *etc.* The wet sites could be divided into two different NVC types (M19a – *Calluna vulgaris-Eriophorum vaginatum* blanket mire *Erica tetralix* sub-community and M15a *Trichophorum cespitosum-Erica tetralix* wet heath *Carex panicea* sub-community). However, all these sites did contain a few very old pine trees as well as a scattering of dwarfed pine saplings, which are not characteristic of these NVC communities. Species associated with the recently disturbed fire are *Festuca rubra*, *Chamerion angustifolium*, *Carex nigra*, *Trientalis europaea*, and *Dryopteris dilatata*.

Species associated with the heathy habitats that have developed 40-80 years after a major fire or gale are *Hypnum* species (probably *H. cupressiforme*), pine regeneration, *Cladonia portentosa* and *Calluna vulgaris*. Species characteristic of scarified sites are *Erica cinerea* and pine regeneration of less than 1 m and 1 – 2 m in height. Table 6.4 also shows that two ruderal species, *Ulex europaeus* and *Digitalis purpurea*, have their maximum in the scarified sites although they are not significant. Species associated with undisturbed pine forest are *Vaccinium myrtillus*, *Deschampsia flexuosa*, *Dicranum majus*, *Hylocomium splendens*, *Plagiothecium undulatum* and *Sorbus aucuparia* saplings. This habitat type could then be further divided according to the density of canopy cover. The denser forest was characterised by *Deschampsia flexuosa*, *Dicranum majus*, *Rhytidiadelphus loreus*, *Vaccinium myrtillus*, *Melampyrum pratense* and *Empetrum nigrum*. This species list is fairly characteristic of W18b (*Pinus sylvestris-Hylocomium splendens* woodland *Vaccinium myrtillus – Vaccinium vitis-idaea* sub-community). However, this could be further divided according to whether it had a semi-natural character with very old trees (>200) or whether it was more characteristic of a plantation style forest. The semi-natural high density forest contained *Empetrum nigrum*, *Lycopodium annotinum*, *Dicranum majus Juniperus communis*, *Deschampsia flexuosa* and *Vaccinium myrtillus* as indicator species. The plantation style woodland contained *Rhytidiadelphus loreus* as an indicator species. The medium to low density forest could also be split according to whether it was of a semi-natural character or plantation style forest. Species found to be significantly associated with the semi-natural pine forest include *Hypericum pulchrum*, *Mercurialis perennis*, *Pteridium aquilinum*, *Potentilla sterilis* and *Viola riviniana.* These species are generally typical of sites with a pH, perhaps indicating that the soil type might have been a brown earth.

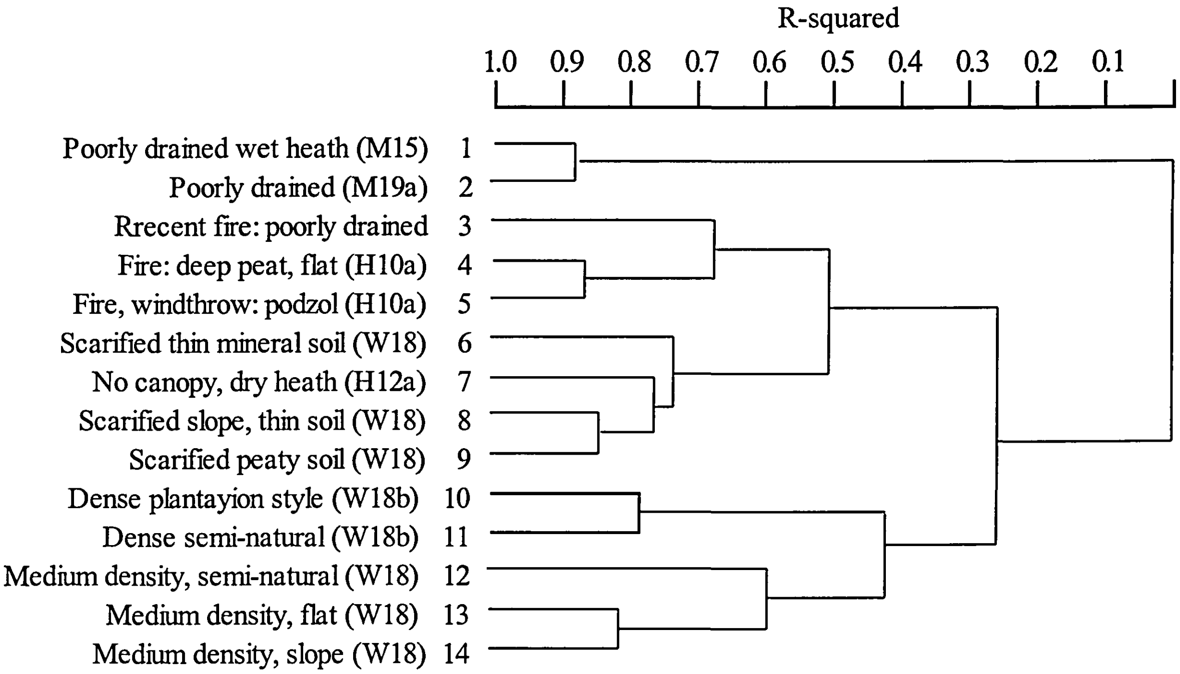


Figure 6.6. Hierarchical dendogram produced from Ward's method. A description of each cluster is followe by its NVC community.

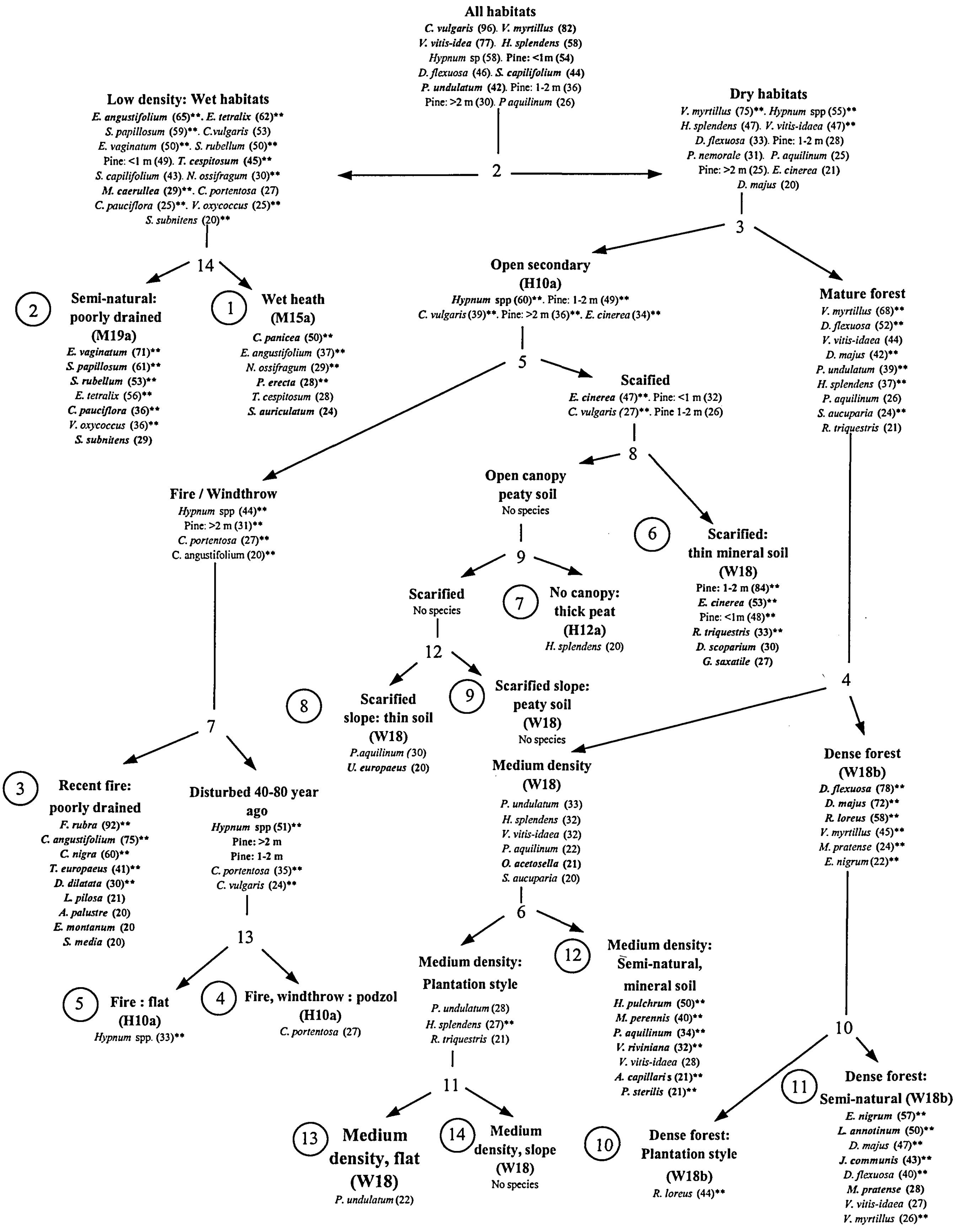


Figure 6.7. Site clusters produced from Ward's method. The associated indicator values are in parentheses. All species with an indicator value >20% are presented for each cluster where they are found. The maximum is indicated in **bold**.

Table 6.5. Two-way indicator table showing the species indicator power for the site clustering hierarchy. The column headings correspond to those in Figure 4.8. Numbers in bold type represent the main data set structure

Species	Indval	Habitats													
	%	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Poorly drained, wet heath (1)															
Carex panicea	50 **	3/3													
Potentilla erecta	28 **	4/4			4/3				2/2	1/1			2/2		
Sphagnum auriculatum	24	3/2								2/2					
Polygala serpyllifolia	17	1/1													
Wet habitats (1,2)															
Eriophorum angustifolium	65 **	15/4	26/9		1/1										
Erica tetralix	62 **	5/3	48/10					5/4	3/2	1/1					3/2
Trichophorum cespitosum	45 **	13/4	42/5												
Narthecium ossifragum	30 **	3/3	5/3												
Molinia caerulea	29 **	6/2	26/4					7/3							
Semi-natural: Poorly drained (2)															
Eriophorum vaginatum	71 **		240/10												
Sphagnum papillosum	61 **	3/2	91/10		2/2				2/1				1/1	1/1	
Sphagnum callifolium var. *rubellum*	52 **	2/2	68/8						1/1						
Carex pauciflora	36 **		6/5												
Vaccinium oxycoccus	36 **		10/5												
Sphagnum subnitens	29		12/4												
Listera cordata	14		2/2												
Drosera rotundifolia	7		1/1												
All habitats (1-16)															
Calluna vulgaris	96	444/6	1050/14	7/5	996/10	887/10	210/5	921/10	892/10	886/10	200/5	29/4	685/10	692/11	575/10
Vaccinium myrtillus	82	42/2	45/4	16/4	17/9	114/10	75/5	53/8	76/10	95/10	395/8	240/4	202/8	381/11	355/10
Vaccinium vitis-idaea	77	10/1	52/7	3/3	12/8	43/10	35/4	12/6	15/8	30/10	27/5	55/4	81/9	41/11	52/10
Hylocomium splendens	58	30/3	111/4		23/2	35/3	75/5	235/5	510/9	165/10	250/5	22/2	250/5	430/10	235/10
Pinus sylvestris : <1 m	54	22/4	135/11	3/2	77/9	12/4	156/4	46/8	26/7	94/7			17/6	4/2	9/4
Sphagnum callifolium var. *callifolium*	44	67/6	266/7		125/4	145/10		113/6	6/3	48/6		46/3	1/1	112/4	330/5
Plagiothecium undulatum	42	4/2	6/4		4/4	6/5		1/1	4/3	10/8	7/4	5/4	17/4	40/8	27/5
Blechnum spicant	18	2/2	1/1			4/4			3/2	6/6	1/1	1/1	1/1	3/3	2/2
Recent fire, poorly drained (3)															
Festuca rubra	92 **			63/5		1/1	34/3	2/1	1/1						
Chamerion angustifolium	75 **			4/4	1/1										
Carex nigra	60 **			3/3											
Trientalis europaea	41 **			4/3	2/2					1/1	2/1				2/2
Dryopteris dilatata	30 **			2/2					2/2		1/1				
Luzula pilosa	21	2/2	4/4	3/3	1/1				2/2	1/1	3/3		3/3	3/3	1/1
Aulacomnium palustre	20			1/1											
Epilobium montanum	20			1/1											
Stellaria media	20			1/1											
Carex echinata	18			1/1					1/1						
Rubus chamaemorus	16		1/1	1/1											
Cerastium fontanum	13			1/1									1/1		
Fire and windthrow (3,4,5)															
No species															
Fire: deep peat, flat (4)															
Carex ovalis	23				3/3				1/1						
Polytrichum commune	20		6/6	1/1	14/4				1/1	3/3				3/3	2/2
Cirsium palustre	10				2/2				1/1				1/1		
Oreopteris limbosperma	10				1/1										
Rosa canina	10				1/1										
Stellaria holostea	10				1/1										
Carex binervis	7				2/1								1/1		
Fire: Mineral soil (4,5)															
Pinus sylvestris : >2 m	40	2/1	10/3		59/9	30/7	60/1	1/1	10/4	25/6			10/5	1/1	
Cladonia portentosa	35 **	1/1	11/8		25/5	12/5			2/2	1/1			1/1		
Fire, Wind-throw: podzol (5)															
No species															
Open canopy (3-9)															
Hypnum spp.	60 **	40/5	8/3	1/1	365/9	720/10	25/2	212/10	200/5	320/10	36/4		130/5	47/5	90/4
Scarified thin mineral soil (6)															
Pinus sylvestris : 1-2 m	84 **	1/1	13/5		81/10	23/8	201/5	2/2	4/2	29/7			4/2	4/3	
Erica cinera	53 **				3/3		10/5	35/5	9/5	3/3					1/1
Rhytidiadelphus triquestrus	33 **	1/1			1/1	1/1	6/4		1/1	2/2	5/4		1/1	15/5	6/3

Species	Indval		Habitats													
	%		1	2	3	4	5	6	7	8	9	10	11	12	13	14
icranum scoparium	30			1/1		1/1		3/3		2/2	2/2	1/1		2/2	1/1	1/1
alium saxatile	27			1/1	1/1			4/3		1/1	1/1	3/2		2/2	4/4	1/1
carified (6-9) and open heath																
o species																
o canopy, dry heath (7)																
o species																
7,8,9)																
o species																
carified slope, thin soil (8)																
lex europaeus	20									2/2						
grostis canina	16					1/1				2/2						
ymnocarpium dryopteris	7									2/1				1/1		
carified slope (8,9)																
igitalis purpurea	10									1/1	1/1					
carified peaty soil (9)																
o species																
ry habitats (3-14)																
o species																
ense plantaion style (10)																
o species																
ense Forest (10,11)																
eschampsia flexuosa	78	**	31/2	31/4	2/1	2/1	4/4	3/2	6/4	4/3	23/7	375/9	140/4	65/7	18/6	8/4
icranum majus	72	**			1/1				1/1			47/7	22/4	5/2	5/4	2/2
hytidiadelphus loreus	58	**	1/1	1/1		1/1		1/1				14/7	2/2			
ense semi-natural (11)																
mpetrum nigrum	59	**	1/1									1/1	3/3		1/1	
ycopodium annotinum	50	**											3/2			
ıniperus communis	44	**	40/1	5/1					3/2		1/1		20/3	5/2	5/1	2/2
elampyrum pratense	29								1/1			2/2	2/2			1/1
ucobryum glaucum	18								1/1				1/1			
edium - Dense canopy (10-14)																
rbus aucuparia	24	**	1/1		1/1						4/4	3/3		4/4	4/4	4/4
edium density semi-natural (12)																
pericum pulchrum	50	**								1/1				211/5		
ercurialis perennis	40	**												5/4		
eridium aquilinum	34	**	3/2			6/3				31/8		10/5		21/8	10/4	8/3
'ola riviniana	31	**				2/1				1/1				4/4		
ostis capillaris	22	**										1/1		8/3		5/1
tentilla sterilis	21	**								1/1				3/3	2/2	0/1
ıemone nemorosa	15					1/1					1/1			2/2		
ucrium scorodonia	14	**								3/2				2/2		
lcus lanatus	10					3/2								2/2	1/1	
dera helix	10													1/1		
nicera periclymenum	10													1/1		
thyrus montanus	7													1/1		
nella vulgaris	7					1/1								1/1		
,13,14)																
alis acetosella	21	**				5/3	3/1			1/1	2/2	2/2		8/4	4/4	3/3
dium density, flat (13)																
agnum squarrosum	9														1/1	
champsia caespitosa	9										1/1				20/1	
agnum recurvum	6					1/1					1/1				2/1	
dium density (13,14)																
species																
dium density: slope (14)																
species																
ıber of sites			6	14	5	10	10	5	10	10	10	10	5	10	10	10

tatistically significant ($P < 0.01$)

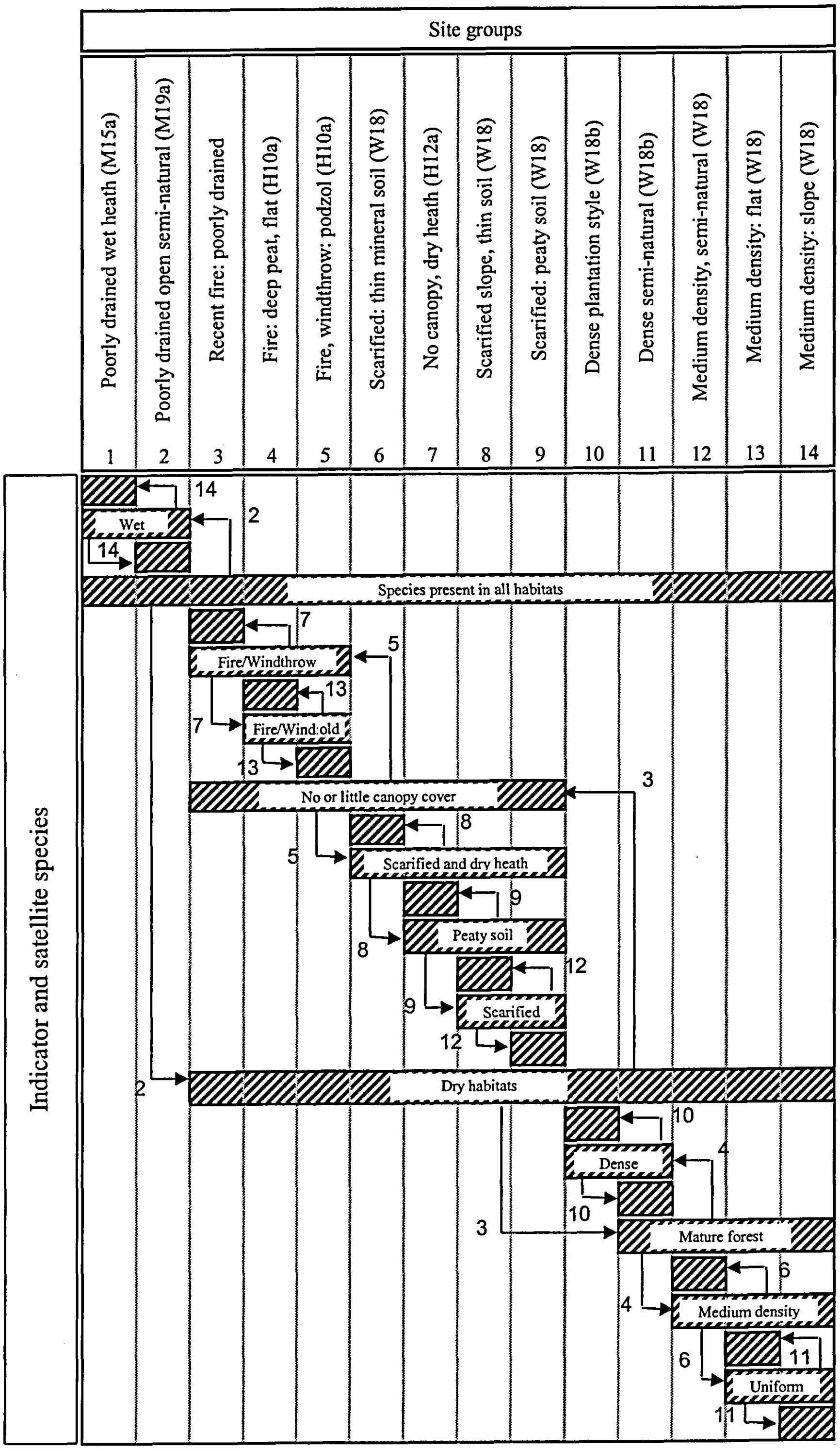


Figure 6.8. The structure of the two-way table presented in Table 6.5.

Discussion 6.5

6.5.1 Plant communitities and indicator species

Contrary to Peterken (1993) who stated 'that the relative poverty and uniformity of the shrub layers makes it meaningless to split the pinewoods into stand types to correspond to them,' this study has shown that major differences do exist even within a single site. The plant distribution of species within the pine forest is likely to be more extreme than the broadleaf woodland that is typical of the UK, as it contains a few dominant species and a large number of rare species, giving the appearance of a species-poor community. Mosses make up a large proportion of the total ground flora. Out of the eighty-two species recorded nineteen species were mosses. The mean percentage cover of moss was found to equal 52% of the ground cover. Any study involved in looking for indicator species in pine forest habitats must therefore take mosses into consideration. Figure 6.10 shows a rank-abundance diagram for the ground cover data. Only nine species have a mean cover greater than 1%. The values for diversity and equitability are very low indicating that that the distribution is closest to a geometric series.

This study has shown the existence of two very important habitat-types i.e. the dense semi-natural pine forest with *Lycopodium annotinum* and also the pine forest habitat of more base rich soils with *Mercurialis perennis* which could be considered as new divisions of the W18 NVC community. The former may be very rare in Scotland but is given some recognition in Scandinavian forests (Pitkin *et. al*, 1994). The species found to be significantly associated with ancient woodland were compared to the lists of ancient woodland indicator lists derived by Peterken (1993) for southeastern England, Honnay *et al.*, (1998) for western Belgium and Wulf (1997) for northwestern Germany. The list of species produced from the present study was found to be most similar to the list produced by Honnay *et al.*, (1998). The list produced by Honnay (1998) for the species found to be significantly associated with ancient woodland includes *Hypericum pulchrum*, *Melampyrum pratense*, *Mercurialis perennis*, *Potentilla sterilis*, *Pteridium aquilinum*, *Teucrium scorodonia* and *Viola riviniana*. The only species in common with the list produced by Wulf (1997) is *Mercurialis perennis* and that published by Peterken (1993) are *Mercurialis perennis* and *Viola riviniana*.

A number of species associated with the semi-natural pine forest communities appear to be have a strong or mild affinity (Tables 3.24 and 3,25) for ancient woodland in the South Yorkshire region. These species include *Anemone nemorosa, Mercurialis perennis, Potentilla sterilis, Lonicera periclymenum* and *Viola riviniana*. However, these are the species are probably associated with semi-natural pine forest on fairly base-rich soils. A tentative list of some of the species associated with more acid soils, probably podzols, is *Empetrum nigrum, Lycopodium annotinum, Juniperus communis* and *Melampyrum pratense*. Considering how different this latter list is to those in the literature and those listed in Chapter 3, it would appear than none of the mentioned lists of indicator species can be applied to this particular environment except perhaps *Melamyrum pratense*.

Scarification appears to have a small but significant effect on the ground cover. The dominant effects appear to be canopy cover and soil wetness. The soil depth and whether the soil is a brown earth, podzol or contains a great depth of peat will also be very important factors in determining the vegetation cover. This is in agreement with the conclusions reached by Humphrey *et al.* (1995).

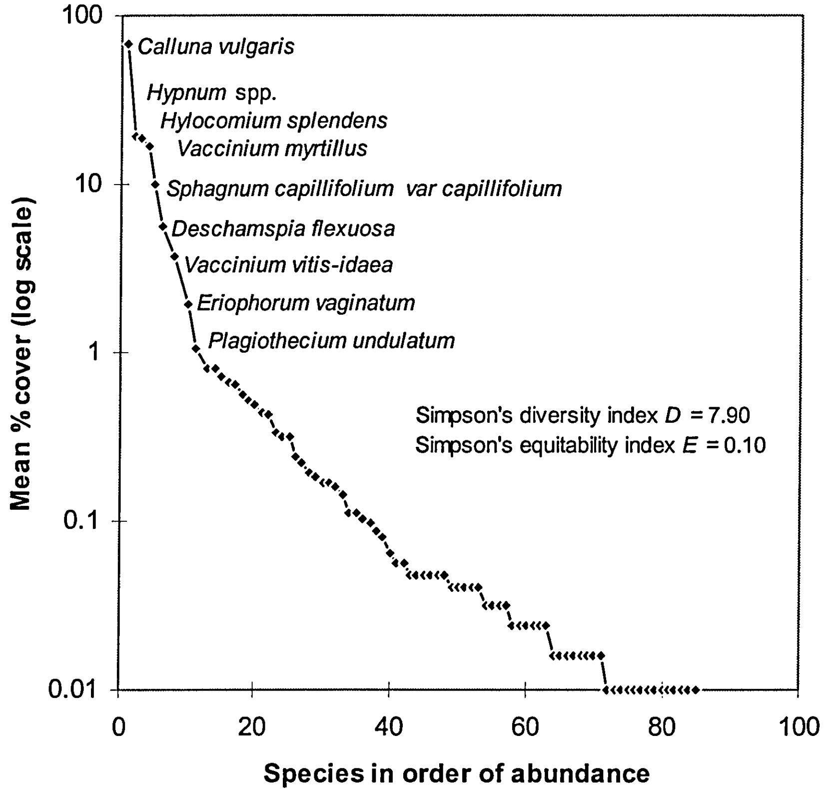


Figure 6.9. Rank-abundance diagram of the pine forest vegetation data. Species names are given for species with >1 % mean cover. Indices were calculated using log (% cover) values.

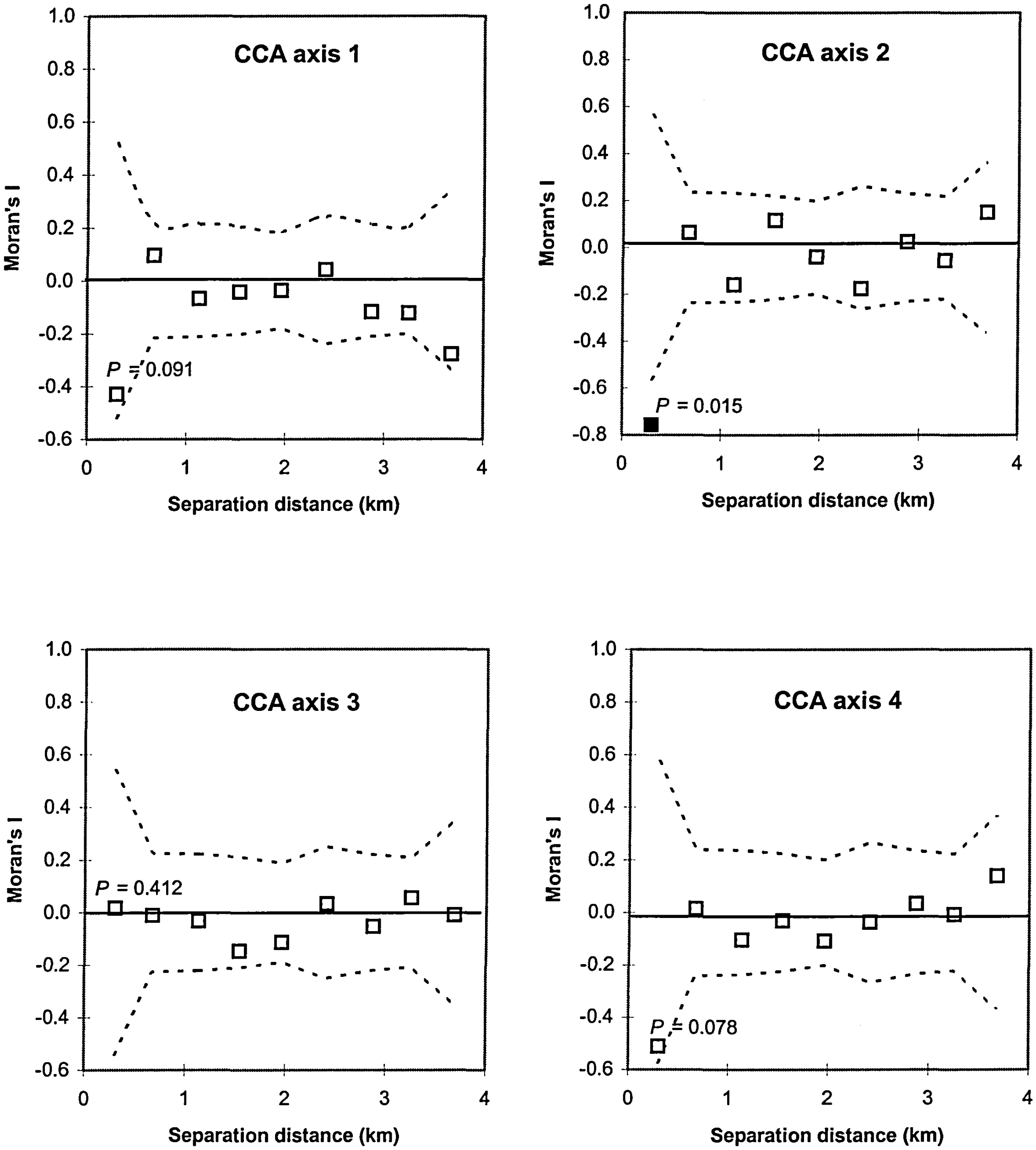


Figure 6.10. Spatial correlograms using Moran's I on the residuals produced from the GLMMs for the relatin the CCA axes to the environmental variables. Confidence intervals were calculated using a correction for sma sample size. The *P*-values for Moran's I at the first lag interval were derived using a Monte Carlo permutatio test. The smallest distance between any two plots was approximately 0.2 km.

6.5.2 Testing for any unaccounted for spatial auto-correlation

The mean values of the residuals, produced from the GLMMs for each of the CCA axes, were calculated for each plot. Correlograms using Moran's I, were then determined for all of the axes, which was conducted using Geostatistics for the Environmental Sciences (Gamma Design Software, 1999). This is equivalent to using the predicted values for each plot.

A corrected value for the confidence intervals, for small samples, was used to test for any unaccounted for spatial auto-correlation by producing spatial correlograms of the residuals produced by the GLMM of standardised values from the CCA analysis. The correction used was that described by Cliff & Ord (1981).

$$I_\alpha = z_\alpha \sqrt{\mathrm{Var}(I)} - k_\alpha (n-1)^{-1} \tag{6.1}$$

Where I is the confidence interval at probability α and k is the lag distance.

The variances (Var (I)) were computed under a randomisation assumption, which simply states that, under Ho, the observations are independent of their positions in space and, are thus exchangeable (Legendre & Legendre, 1998). The Excel 97 add-on – Rook's Case (Sawada, 1998) was used to calculate Var (I) and also to calculate the exact probabilities at the first lag class for each of the correlograms using a Monte Carlo permutation tests with 10 000 permutations.

The correlograms produced from the residuals of the GLMMs, calculated using Moran's I, are shown in Figure 6.10. The first lag interval class for the residuals from CCA axes 1,2 and 3 all show negative correlations, although this is only significant for axis 2. Axis 2 represented the variation of a soil wetness gradient. The poorly drained areas were small and the minimum distance between plots was 200 m, it is therefore not surprising that at the smallest lag there is a significant difference. However, it does suggest that the inclusion of the dummy variable for poorly drained may be an over-simplification, since differences between the wet sites were found to occur.

6.5.3 Forest management

This study has highlighted the importance of the older stand types in terms of their species composition. These habitat-types are likely to be at most risk from fire and overgrazing. Fire has the ability to remove the seed source as well as established plants. The burnt areas studied generally lacked the pine forest species and were more typical of heather moorland. Deer selectively eat each species such as *Vaccinium myrtillus*, which is one of the important components of the ground vegetation under denser canopy cover. Other species may also be severely reduced or lost due to overgrazing. The original extent of the pine forest would have been sufficiently large for fires to never be catastrophic enough for species extinctions. Furthermore fire may be a very important factor required for pine regeneration and maintaining a great diversity of different habitat-types. However, the areas of old mature pine forest now in existence are a very scare resource. They therefore need protecting from catastrophic fire.

A number of important potential indicator have been identified by this study. The same underlying principals that have were identified in Chapters 2 and 3 appear to hold true for the pine forest. This is the importance of factors that may limit the colonisation ability of different species. In Chapter 3 it was found that fewer species were required to be confident that a site is ancient on the Bunter Sandstone than on either the Coal Measures Series or Magnesian Limestone Series. The main limiting factor was suggested as being soil moisture content. The Bunter Sandstone vegetation could be described as isolated species rich communities in flushed areas or alongside streams amongst a very harsh environment formed by freely draining deep sand rich soil. In a similar way the pine forest could be described as having isolated species rich communities in base rich areas amongst a very harsh acidic substrate.

The presence of isolated communities in these situations poses great problems of how to conserve, restore and create new habitats if we wish to replicate semi-natural ones. Once species have been lost from a site as a result of draining, over-grazing *etc.* it may be very difficult to establish these species in sites where they have once flourished. If these sites are to be restored or new sites created, then a great deal needs to understood in terms of what factors limit particular species. The use of geographical models to suggest suitable locations for the planting of these species may greatly enhance this process. If these communities are left to develop with no intervention, it may take considerable time to produce communities that resemble those which formerly occurred.

CHAPTER 7

7. DISCUSSION

7.1 Summary of findings

The reasons why a particular species maybe a useful indicator are very complex. Although the methods that species use to spread vegetatively and produce and disperse seed, play an important factor, the complexity of response to a particular set of environmental parameters is great. It is of such magnitude that it is impossible to predict with accuracy which species should be considered as ancient woodland indicator species, based purely on laboratory work, or computer models. The complexity of response of *Hyacinthoides non-scripta* to climatic fluctuations illustrates this point.

The collection of data from a large number of sites does result in a dataset where species can be tested for their association with ancient woodland. However, the sites are often very isolated, have slightly different soil properties with different degrees of slope and different networks of water channels *etc.* The selection of indicator species from such data only serves to generalise what would appear to be a very complex question. The present study has been able to quantify the benefits of using the number of indicator species to predict whether a site is ancient or not. This study has been able to show considerable differences between the potential indicator species in different regions. It suggests that producing separate indicator lists for different regions or habitat-types is a very important and worthwhile exercise. In some circumstances it may also be possible to predict that a species may be a very good indicator. For example, this may be the case if it produces very little seed with a poor dispersal mechanism, and if it also grows very slowly.

Species that spread particularly slowly such as *Anemone nemorosa*, often appear to also have a highly competitive nature. These two properties result in a species that may only occur in sites of great antiquity. When present they may also be found in abundance. It may therefore be assumed that in any environmental gradient that is concerned with the age of a site, there will be a suite of species that will be almost restricted to the sites of greatest antiquity. These will be reliable indicators of the age of a site. However, there are often even exceptions to simple rules such as this. For example, *Anemone nemorosa* can also be found in meadows and also in heather moorland in Scotland. This species was found to have only a weak association with semi-natural pine forest in Chapter 6. It is possible that this species is particularly persistent and so may represent a relict species of a former ancient woodland. However, in many cases its occurence may be the result of chance or a result of sexual reproduction under certain conditions which as yet are poorly understood.

Problems associated with the use of different strategies for surveying woodlands have also been raised. In particular the bias towards detailed surveys of only ancient sites has been shown to cause a number of potential statistical problems. If more ancient woodland sites are surveyed or these sites surveyed more thoroughly than recent sites, then there is a greater risk of finding a species in only ancient sites. This may appear to be an indicator species when it may actually be present in an equal umber of ancient and recent sites if they were sampled in equal proportions. If a larger number of recent sites is surveyed or is surveyed more thoroughly, then this would give a more conservative list of indicator species. Peterken's (1974) list is based on data that includes a greater proportion of recent sites and so should be regarded as a fairly conservative list of indicator species. In order to produce a statistical test for the association of each species with ancient woodland separate tests need to be done. This results in multiple tests being carried out which weakens the power of the test. It is therefore suggested that including more recent sites than ancient sites in such a study is likely to produce a much more reliable list of indicator species than using equal numbers of both ancient and recent sites.

Perhaps the greatest advance of this work has been the refinement of Peterken's (1974) method for determining indicator species in Chapter 3. The advantages of the new method over Peterken's method are:-

- the ease with which different indicator lists can be compared with each other even where the ratio of ancient to recent sites may differ considerably
- the use of a statistical test to give enhanced confidence in the species that have been found to be associated with the ancient woodland

The advantages of this method compared to methods used by many of the ecologists on the Continent such as Honnay *et al.* (1998) who relied solely on Chi-square tests are:-

- the removal of species with only a very weak association with ancient woodland
- quantifying a species affinity for ancient woodland.

7.2 Surveying sites for research and nature conservation

There are a number of problems encountered when trying to standardise the method for surveying sites, whether for management plans, or for surveying and resurveying of Sites of Scientific Interest and Sites of Special Scientific Interest. Even Phase 1 Habitat surveys differ considerably from a few key species to complete species lists being given for sites. Some recorders just record whether a species is present, whilst others often use the DAFOR system for recording species abundance. Other major differences of recording include how sites are partitioned. Recorders may produce a number of different species lists with abundance values for a large number of areas within a single site. Others just give a single species list for the whole site. Sites can be divided up into conveniently sized areas perhaps separated by linear features such as paths, roads or streams. Other recorders will try to split a site up according to the vegetation type or they will do a mixture of these two strategies. Recording different habitat-types is very useful and can provide data on subtle differences between plant communities e.g. an area of recent woodland close to an existing ancient woodland provides a direct comparison of the two types of woodland without the difficulties of interpretation due to location. However, if a site is split up according to convenient features within the site then it is difficult to know what to do with such data in terms of analysis. It would appear firstly that there are far more samples in the data than there really are. This is because the separate species lists from a single site, would be expected to be highly correlated with each other. This will increase the risk of wrongly selecting species as indicators. There may be a number of species that might be restricted and widespread in a particular site for reasons other that of continuity and antiquity. If such a site was to be classified as ancient woodland then it would appear that a number of species may have a strong affinity to ancient woodland when their occurrence may be explained by a completely different factor. This study has not tried to address this problem directly. Even if statistical tests that include spatial dependency are used, there is currently no way of building this into any species indicator analysis, other than to pool these lists from compartments together to form a single species list for that site.

Other sources of error can originate in unequal survey effort of different types of environments or in different locations. In some cases very little time is spent on a site once the surveyor feels that the site has little biodiversity or conservation value. The only area where all sites had been thoroughly surveyed within the South Yorkshire region was for Doncaster. The sites surveyed in this area were all done by a single survey team, which were given no prior information about the sites they visited and had to assess the site in terms of their status as Sites of Scientific Interest.

The most species rich sites have often been surveyed thoroughly at least twice and species added to their list by local enthusiasts. There is therefore a high risk of bias in the data, if all the information available is included in any analysis. If the sites that are considered as having the highest conservation status have received the greatest survey effort, there is a greater risk that any species being found in one of these sites being considered to be an indicator species when it may not be so. The present study attempted to remove some of this bias by only using a single species list for a site. However, where sites did contain a number of different aged sites within a single site then the separate species lists for the different aged sites were used. This approach does suffer from a number of problems. There may be a significant amount of error in the data which is a result of seasonal differences. In Chapter 2 the variance explained by the temporal effects was found to be considerably less than that for the environmental and spatial effects. However, it is still preferable to have a data set that contains species lists that have been recorded at three different times of year i.e. spring, mid summer and late summer. If all the data had been used it is likely that more species would have been found to be indicators. However, there would be much less confidence in the results. Arguments could have been given using previous published data to try and decide on the likeliness of each species being an indicator species. However, such a list would tend to go against the criteria listed by Ferris and Humphrey (1999) i.e. they need to be easy to assess, even for non-specialists; they must be repeatable; they must be cost-effective; and they must be ecologically meaningful, providing data that is easy to interpret. The rarer species may be important in terms of the nature conservation importance of a site, but may contain very little further information in terms of the antiquity of a site. This is because if a site contains a number of very rare species which are perhaps largely dependent on a site's antiquity, then a large number of more easy to assess species would be expected to be found.

The Chi-square tests appeared to give very similar probability values as those calculated by INDVAL in Chapter 3. There appears to be little information contained in the DAFOR rating of species abundance. This is perhaps not surprising, since a large number of factors will influence species abundance and there are complications in estimating abundance values. This is particularly so where species are present in a number of different forms, such as trees, shrubs, grasses *etc.* A more useful dataset would probably be one that contains just the presence/absence of species but does contain combined species lists for sites surveyed twice. The differences between the flowering periods of different species are well known (Kirby, 1986). If all sites in a dataset have had least two surveys, one in early spring, the other in late summer, then this would help reduce any bias due to seasonal differences considerably. However, the temporal component of the data set used in Chapter 2, was found to be considerably smaller than the spatial and environmental components. This is perhaps not surprising since the dead stems of plant species, such as *Hyacinthoides non-scripta*, do persist for a great length of time, so can be recorded outside the main flowering period.

The importance of recording the length of time spent on site has also been emphasised by this study. The equation and graph presented in Chapter 4, relating survey effort and the proportion of species recorded, can be used to give an estimate as to the proportion of species found for a given length of time and a given area surveyed. This can then be used to give an estimate of the number of species missed for a given survey.

7.3 Implications for woodland conservation and management

Indicator species are primarily used to produce a shortlist of species which can be used to reduce the time required for assessing a number of sites, so that different species lists can be compared quickly and easily with each other. Typically those sites with more indicator species will be those considered for a conservation status. The rare species that are found to be restricted to ancient woodland may also be considered at this stage. The list of sites is often further refined to make sure that all relevant habitat-types and communities are included in this list of sites (Peterken, 1993). The refinement of indicator lists should also help build a clearer picture of relic environments and communities and where the species associated with these habitats may once have occurred.

Perhaps the greatest implication of the results presented in this study is the importance of determining different lists of indicator species for different geographical locations. The factor that was found to explain most of the variation in the data presented in Chapter 3 was the geology type.

The shorter list of indicator species on the Bunter Sandstone does not mean the woods are of less conservation value. In fact it could be taken to indicate the opposite. Since colonisation rates may be considerably slower the recovery time of such an ecosystem may be much greater. If communities are valued in terms of their antiquity and the time required for them to form, then it may be these species-poor sites that need the greatest protection. The main limiting factors on the Bunter Sandstone were suggested, in Chapter 2, as being soil moisture content and soil pH. Ancient woodlands on this geology type would be expected to vary considerably, reflecting the moisture and organic content of the soil. If the flushed areas of ground are isolated, then none of the ancient woodland indicator species may be able to colonise these areas. This can be compared to the situation with sites on the Magnesian Limestone and Permian Marl, which can form extensive areas with soil types favourable to many woodland species and so they do not present barriers for the recolonisation of woodland species. The pine forest area studied may also be described as small patches of species-rich communities, existing in favourable conditions, amongst a generally very much harsher environment. The species-rich areas exist where the soil type may be a brown earth or in boggy areas or along the banks of streams. The more harsh environment consists mainly of acidic soils, often prone to becoming very dry. These sites may have an open canopy that produces a highly competitive community, dominated by *Calluna vulgaris*, or under a dense canopy consisting of a dominant cover of *Vaccinium myrtillus* and *Deschampsia flexuosa*. The movement of key species may therefore be very limited, and result in a community which may take a very long time to recover from disturbance or inappropriate management.

It is therefore likely that a great deal more care and effort is needed to establish mature plant communities when recreating new woodlands in more patchy environments. This is especially so if they contain barriers, which may hinder or stop natural recolonisation.

7.4 Statistical rationale

A number of different datasets has been analysed and different types of statistical methods used. One of the most complex questions still remains. This is: 'How many data are needed to allow reliable analysis when samples are not independent of each other?' This study has shown the importance of large datasets for certain survey designs (especially in Chapter 5). This is particularly important when random plot effects are recognised to occur within the survey design or some sort of spatial or time series error structure exists within the data.

The inclusion of multiple errors often means that approximate degrees of freedom need to be used which can become very small depending upon what structures exist within the data. The advantage of using these types of statistical tests is that datasets with spatial data or data that contain no control, can be analysed by an appropriate test. This is the case even if the power of the test is greatly reduced.

There also remains the problem of how to combine these types of analysis with those available in ordination programs such as CANOCO. The latter use matrices to determine the solution of multiple regression equations. Although it would be ideal to be able to test environmental effects and simultaneously include random effects or other dependent error structures, such tests can still be used after the ordinations have been produced, and as a more conservative test than that provided by CANOCO. The further use of constrained ordination scores has also been shown to be a useful exercise. Ecological data are often 'noisy' with habitat-types varying considerably. Therefore ordinations produced from species abundance values can be difficult to interpret. However, if the sample scores are constrained by effects such as longitude, latitude and altitude or other effects of interest, then the sample scores are likely to be separated according to geographic location first, before being split according to antiquity *etc.* This means that the technique can be used to decide which communities are common to a particular geographic location or environmental characteristic. It would otherwise be lost if the sample scores were not constrained. Such an output from analysis may be particularly relevant for the work of local authorities, environmental cosultancies and research departments focussing primarily on a local regional area.

The use of INDVAL has also been shown to greatly aid the methods used in describing plant communities. This technique has a number of advantages over TWINSPAN and COINSPAN. The only drawback to INDVAL is perhaps its great flexibility. Care is needed when using cluster analysis to make sure that sensible clusters are produced. Although Dufrene and Legendre (1997) used a fairly conservative method by not assuming a hierarchical structure, the collection of ecological data does often follow a hierarchical structure. However, this is not always balanced. Since CANOCO can output sample scores that are based on chi-square distances, a large number of metric hierarchical clustering methods can be used. However, care still needs to be used in terms of judging when to stop forming new clusters. Confusion tables are a useful way of showing the amount of noise in the data set and identifying where the algorithm is having difficulties in forming separate groups. Attention should also be drawn to any areas of a dataset that maybe unbalanced, as this can lead to species being wrongly assigned significant indicator values. This may be for a particular group of samples which contain other important environmental characteristics that have not been recorded. This study has also shown that INDVAL can be applied to very simple hierarchical structures and used instead of chi-square tests and Fisher's exact tests. The main advantage of this is the property of INDVAL of producing a separate list for species that are deemed to be common to two groups, so producing a more reliable list of indicator species. Species are not assumed to belong to one of two groups, instead they may be equally common to both groups. Taken a step further by using threshold values for the percentage occurrence of a species in a particular habitat-type allows grouping according to affinity for a particular habitat-type. This appears to be particularly useful when the presence of particular species is required to assess sites for their antiquity in the absence of other evidence.

7.5 Future research

A number of important questions which still need answers have been highlighted by this study. These are:

1) How do species colonisation rates change under different environmental conditions for species and environmental combinations that have not been previously studied?
2) How will rates of colonisation differ with climatic change?
3) What are the mechanisms and limitations in seed dispersal and rates of spread for the slow colonising species?
4) How do the indicator species change according to other geology types or regions, such as the Natural Areas devised by English Nature?
5) How can survey methods be standardised so that both ancient and secondary woodland receive the same effort?
6) Should a single method be used for determining ancient woodland indicator species?

7.5.1 How do species colonisation rates change under different environmental conditions for species and environmental combinations that have not been previously studied?

Sudies of species colonisation rates have only been published for work carried out in a few locations. This includes a few sites in Belgium (Bossuyt *et al.*, 1999; Honnay *et al.*, 1999), Sweden (Brunet and Von Oheimb, 1998), North America (Matlack, 1994) and the UK (Vickers and Rotherham, 2000). No study as yet has tried to quantify how these colonisation rates change with different environmental conditions. There is also the need to identify conditions that might completely restrict colonisation. If this information is known then site maps can be produced based on known environmental conditions such as canopy cover, soil types and drainage ditches and stream courses. Statistical models using this information can then be displayed using geographical information systems to predict the rates of spread of selected species. This type of information would be of great value in the restoration of environments and the creation of new woodlands. It would help predict the time required to create a habitat-type perhaps with the manipulation of different initial plant introductions. It would also provide guidance for land managers in terms of where to plant species to give them the best chance of becoming established throughout the whole of a site. It would improve understanding of what plant communities are likely to form (given sufficient time), and when different management regimes such as coppicing may have the most beneficial effects in terms of stopping certain species becoming dominant and maintaining desirable communities and species.

7.5.2 How will these rates of colonisation differ with climatic change?

Weather has been shown to greatly influence the growth of *Hyacinthoides non-scripta*. It is likely that the growth of many woodland species will change under different climatic conditions. If the climatic change in the UK produces wetter milder winters, which is what has been predicted, then *Hyacinthoides non-scripta* will be at a disadvantage since this species appears to prefer cold dry winters and sunny dry springs. Since the colonisation ability of this species is directly related to the height of stem produced its colonisation rate is likely to be reduced in the future. Further work is required to assess how other species are likely to respond to climatic change and which species may be under threat. Many species rely on insects for pollination and/or for seed dispersal. An understanding of how these insect species will respond to climatic change is therefore very important.

7.5.3 What are the mechanisms and limitations in seed dispersal and rates of spread for the slow colonising species?

The seed dispersal of many species is still often poorly understood (Spencer, 1990). This is particularly true of species, which are thought to rely on ants (myrmechores). This includes species such as *Mercurialis perennis* and *Paris quadrifolia* and many *Carex* species. Many species are still listed in the literature as unknown (Honnay *et al.*, 1998). Without knowing how species disperse seeds, it is very hard to predict how they will respond to changes in environmental conditions. This will be predominantly relevant to the conservation of many of the rarer species.

7.5.4 How do the indicator species change according to other geology types or regions, such as the Natural Areas devised by English Nature?

The present study has emphasised the need to produce separate lists of species according to geology. The Natural Areas proposed by English Nature may be a very useful way of deciding upon where separate lists of indicator species are required. Although this study concentrated on areas to the east of Sheffield, it has also highlighted a major difference of sites to the west of Sheffield typical of wet upland conditions situated on the Millstone Grit Series.

7.5.5 How can survey methods be standardised so that both ancient and secondary woodland receives the same effort?

At present there are considerable differences between surveyors as to how they have recorded species and partitioned sites. This study has also highlighted the problem of restricting detailed surveys to just the sites, that at present, are thought to have the highest conservation status. In order to be able to justify the high conservation status of these sites there needs to be just as much data collected for more recent sites such as secondary woodland. The amount of time spent on a site is also a very important piece of information.

It would be very difficult and impractical to try and standardise the amount of time spent on a site, since the complexity of sites may vary considerably. However, the amount of time spent on a site, if recorded, can be partialled out in a multivariate analysis which should give a clearer picture as to which sites should be given the highest priority in terms of nature conservation.

7.5.6 Should a single method be used for determining ancient woodland indicator species?

In 1974, over twenty-five years ago, Peterken devised a method for determining indicator species. This could be easily used, without any statistical knowledge, in any region of this country or elsewhere in the world, to derive a list of species with a high affinity for ancient woodland. The present study has generated refinements to this method. These include correcting for unequal numbers of ancient and secondary sites, as well suggesting the types of statistical tests that can be used on this type of data. However, during this 25-year period, there seem to be no published attempts that either repeat or refine this original method. Another appraoch has been to decide upon a list from their own experience and knowledge (Rose, 1999; Peak National Park, unpublished); whilst other authors have devised their own methods deciding to solely rely upon chi-square tests.

However, it is very difficult to fault the basic method devised by Peterken. This is able to give information on how confident a surveyor should be when a species is found from his or her list. No other author has achieved this level of information, from the methods they have used for determining indicator species. The INDVAL procedure clearly shows the danger of relying upon chi-square type tests since a chi-square test makes no allowance for a species being common to two groups such as ancient and secondary woodland. The lists produced by authors in the rest of Europe such as Honnay *et al.* (1998) are therefore likely to include a mixture of species with a high affinity for ancient woodland and species with a general preference for woodland environments. These papers therefore provide far less useful information, when compared to Peterken's method.

If separate lists of indicator species are to be produced for different regions in the UK, it is hoped that a similar method to that devised by this study will be adopted. This will include the assumption that species common to both ancient and secondary woodland should not be regarded as indicators of ancient woodland, and to also recognise the importance of listing species with greater than 90% and those species with 75-90% of their occurrence in ancient woodland separately. Of course these are arbitrary values but for the purposes of producing a standard method they seem useful and pragmatic thresholds. The IndVal procedure with abundance data or presence/absence data or chi-square and Fisher's exact tests appear to both be adequate tests of significance after the species have been separated according to their percentage occurrence in ancient woodland.

7.6 Conclusion

The current study has identified a number of new techniques that can be applied to ecological data, and also provides scientific and statistical evidence to support several important ecological principles. The most important are listed below:-

- A list of ancient woodland indicator lists has been produced for the South Yorkshire Region. For the first time a list of ancient woodland indicators species has been produced using a procedure that combines a method for detecting reliable indicator species and also an appropriate statistical test.
- This research has demonstrated important differences between indicator lists in different regions (taking examples from the South Yorkshire study area and the Scottish Pine Forest), and so provides support for the production of separate lists throughout the UK.
- The rate of colonisation and succession has been determined for some sites in South Yorkshire
- Some of the complexities of why a species may be an indicator and how this might alter with climate change *etc.* is now better understood, in particular for *Hyacinthoides non-scripta.*
- The technique of using CCA with GLMs and Kriging to identify a species niche and overlap with other species / hybrids *etc.* has been used and in particular has given insight into the niche separation of *Holcus lanatus*, *Holcus mollis* and their hybrid.
- The importance of carrying out thorough surveys, repeat surveys and surveys at different times of year has been shown, and the implications of not doing so has been discussed.

In the words of Begon, Harper and Townsend (1990) –

"Ecology is not an easy science, and has particular subtlety and complexity. It must deal explicitly with three levels of the biological hierarchy – the organisms, the populations and the communities of populations, and it ignores at its peril the details of individuals, or the pervading influences of historical, evolutionary and geological events."

This thesis has attempted to draw together details of selected species, and the influence of various factors both at the species and community level. This study has also tried to achieve a good sampling methodology and statistical design for the collection of data, which are the two components of scientific progress according to Krebs (1994).

This thesis has also attempted to elucidate many complex ecological processes through the application of statistical modelling to ecological problems. Legendre and Legendre (1998) argue that when used appropriately, scientists can derive a deeper understanding of natural phenomena from their mathematical calculations. Begon, Harper, Townsend (1990) also argue that ecology will remain a meeting-ground for the naturalist, the experimentalist, the field biologist and the mathematical modeller, and that ecologists should try to some extent to combine all these facets.

The complexity of the interactions between communities and their environment will mean that mathematics/statistics will never alone solve ecological problems. The application of modern statistical methodologies will often only serve to emphasise the limitations of collecting ecological data. Modern statistical programs now allow a great number of variables to be entered. Also for the first time dependency can be assumed among the samples *i.e.* the samples may be spatially correlated resulting in less confidence in a statistical test than would otherwise have been expected. In practice it would be impossible to collect all the data that would be needed in order to be able to separate out the effects of all environmental variables and their interactions. It is therefore essential that the experience of field ecologists and environmental managers is relied upon in order to ask sensible questions before applying complex statistical models to any data set. This experience is also essential in order to interpret or refine any statistical analysis of ecological data.

It is hoped that the reader will be both come away with a fuller knowledge of woodland ecology and a greater understanding of statistical modelling to ecological data. The data collected and used in the different studies are all provided on a CD at the end of the thesis. All the SAS syntax programs written are listed in the Appendices.

References

Abbate, T. (1992) Which bird is the better indicator species of old-growth forest? *Bioscience*, **42**, (1) 8-9.

Anderson, M.J. and Gribble, N.A. (1998) Partitioning the variation among spatial, temporal and environmental components in a multivariate data set. *Australian Journal of Ecology*, **23**, 158-167.

Ash, J.E. and Barkman, J.P (1976) Chages and variability in the field layer of a coppiced woodland in Norfolk, England. *Journal of Ecology*, **64**, 697-712.

Begon, M., Harper, J.L. and Townsend, C.R. (1990) *Ecology. Individuals, Populations and Communities*. Blackwell Scientific Publications, London.

Belbin, L. and McDonald, C. (1993) Comparing three classification strategies for use in ecology. *Journal of Vegetation Science*, **4**, 341-348.

Bennett, K.D. (1994) Post-glacial dynamics of pine (*Pinus syvestris* L.) and pinewoods in Scotland. In: *Our Pinewood Heritage* (Ed. J.R. Aldhous). pp 23-39. Forestry Comission, Farnham, Surrey.

Bierzychudek, P. (1982) Life histories and demography of shade-tolerant temperate herbs: a review. *New Phytologist*, **90**, 757-776.

Blackman, G.E. and Rutter, A.J. (1954) *Endymion non-scriptus* (l.) Garcke. Biological flora of the British Isles. *Journal of Ecology*, **42**, 629-638.

Bocard, D. and Legendre, P. (1994) Environmental control and spatial structure in ecological communities: an example using oribatid mites (*Acari*, *Oribatei*). *Environmental and Ecological Statistics*, **1**, 37-61.

Bocard, D., Legendre, P. and Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045-1055.

Bossuyt, B. Hermy, M., Deckers, J. (1999) Migration of herbaceous plant species across ancient-recent forest ecotones in central Belgium. *Journal of Ecology*, **87**, 628-638.

Bradfield, G.E. and Scagel, A. (1984) Correlations among vegetation strata and environmental variables in subalpine spruce-fir forests, southeastern British Columbia. *Vegetatio*, **55**, 105-114.

Brunet, J. (1993) Environmental and historical factors limiting the distribution of rare forest grasses in south Sweden. *Forest Ecology and Management*, **61**, 263-275.

Brunet, J. and Von Oheimb, G. (1998) Migration of vascular plants to secondary woodlands in southern Sweden. *Journal of Ecology*, **86**, 429-438.

Bunge, J. and Fitzpatrick, M. (1993) Estimating the number of species: A Review. *Journal of the American Statistical Association*, **88**, (421) 365-373.

Carleton, T.J., Sitt, R.H. and Nieppola, J. (1996) Constrained indicator analysis (COINSPAN): An extension of TWINSPAN. *Journal of Vegetation Science*, **7**, 125-130.

Clench, H. (1979) How to make regional lists of butterflies: Some thoughts. *Journal of Lepidopterists' Society*, 33, **4**, 216-231.

Cliff, A.D. and Ord, J.K. (1981) *Spatial processes – Models and applications*. Pion, London.

Colwell, R.K. (1995) *User's Guide to the Estimator Program EstiMateS*. Unpublished.

Corbet, S.A. (1998) Fuit and seed production in relation to pollination and resources in *Hyacinthoides non-scripta*. *Oecologia*, **114**, (3) 349-360.

Cressie, N.A.C. (1991) *Statistics for Spatial Data*. John Wiley & Sons Inc., New York.

Dickson, J.H. (1992) Scottish woodlands: their ancient past and precarious present. *Botanical Journal of Scotland*, *46*, 155-165.

Diekmann, M., Eilertsen, O., Fremstad, E., Lawesson, J.E., Aude, E. (1999) Beech forest communities in the Nordic countries – a multivariate analysis. *Plant Ecology*, **140**, 203-220.

Diggle, P.J. (1992) *Time Series: A Biostatistical Introduction*. Oxford University Press, Oxford.

Dufrene, M. and Legendre, P. (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*, **67**, 345-366.

Dzwonko, Z. (1993) Relations between the floristic composition of isolated young woods and their proximity to ancient woodland. *Journal of Vegetation Science*, **4**, 693-698.

Dzwonko, Z. and Gawronski, S. (1994) The role of woodland fragments, soil types, and dominant species in secondary succession on the western Carpathian foothills. *Vegetatio*, **111**, 149-160.

Dzwonko, Z. and Loster, S. (1992) Species richness and seed dispersal to secondary woods in southern Poland. *Journal of Biogeography*, 19, 195-204.

Eriksson, O. and Ehrlen, J. (1992) Seed and microsite limitation of recruitment in plant populations. *Oecologia*, **91**, 360-364.

Evans, J. (1991) Forestry in Britain. In: *Forestry Comission Handbook 6: Forestry Practice* (Ed. B.G. Hibberd). HMSO Publications Centre, London.

Evans, M.N. and Barkham, J.P. (1992) Coppicing and natural disturbance in temperate woodlands – a review. In: *Ecology and Management of Coppice Woodlands* (Ed. G.P Buckley). Chapman & Hall, London.

Everitt, B.S. (1993) *Cluster Analysis. Third Edition*. Halsted Press, New York.

Ferris, R. and Humphrey, J.W. (1999) A review of potential biodiversity indicators for application in British forests. *Forestry*, **72** (4) 313-328.

Field, J.G., Clarke, K.R. and Warwick, R.M. (1982) A practical strategy for analysing multispecies distribution patterns. *Marine Ecology Progress Series*, **8**, 37-52.

Forestry Commission (1994) *Native pinewoods. Forestry Practice Guide No. 7.* Forestry Commission, Edinburgh.

Fuller, R.J. (1992) Effects of coppice management on woodland breeding birds. In: *Ecology and Management of Coppice Woodlands* (Ed. Buckley, G.P.), pp. 169-192. Chapman & Hall, London.

Friendly, M.L. (1991) *SAS System for Statistical Graphics. First Edition.* SAS Institute Inc. Cary, NC, USA.

Gamma Design Software (1999) *GS+ Geostatisics for the Environmental Sciences (version 3.11.14)*. Plainwell, Michigan, USA.

Goudie, I.B.J., and Goldie, C.M. (1981) Initial size estimation for the linear pure death process. *Biometrika*, **68**, 543-550.

Grabham, P.W. and Packham, J.R. (1983) A comparative study of the *Hyacinthoides non-scripta* (L.) Chouard in two different woodland situations in the West Midlands, England. *Biological Conservation*, **26**, 105-126.

Grashof-Bokdam, C.J. and Geertsema, W. (1998) The effect of isolation and history on colonization patterns of plant species in secondary woodlands. *Journal of Biogeography*, 25, 837-846.

Grime, J.P., Hodgson, J.G. and Hunt, R. (1988) *Comparative Plant Ecology*. Unwin Hyman, London.

Hammond, P.M. (1974) Changes in the British fauna. In: *The Changing Flora and Fauna of Britain* (Ed. D.L. Hawksworth) pp. 323-369. Academic Press, London.

Helliwell, D.R. (1975) The distribution of woodland plant species in some Shropshire hedgerows. *Biological Conservation*, **7**, 61-72.

Hermy M. and Stieperaere, H. (1981) An indirect gradient analysis of the ecological relationships between ancient and recent riverine woodlands to the south of Bruges (Flanders, Belgium). *Vegetatio*, **44**, 43-49.

Hermy, M., Honnay, O., Firbank, L., Grashof-Bokdam, C., Lawesson, J.E. (1999) An ecological comparison between ancient and other forest plant species of Europe, and the implications for forest conservation. *Biological Conservation*, **91**, 9-22.

Hill, M.O. (1976) *TWINSPAN: A Fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes*. Ecology and Sytematics, Cornell University, Ithaca, NY.

Honnay, O., Degroote, B. and Hermy, M. (1998) Ancient-forest plant species in western Belgium: A species list and possible ecological mechanisms. *Belgium Journal of Botany*, **130** (2) 139-154.

Honnay, O. Hermy, M., Coppin, P. (1999) Impact of habitat quality on forest plant species colonization. Forest Ecology and Management, 115, 157-170.

Humphrey, J.W., Glimmerveen, I. and Mason, W.L. (1995) The effect of soil cultivation techniques on vegetation communities and tree growth in an upland pine forest. I: Vegetation Responses. *Scottish Forestry*, 49, **4**, 198-205.

Jardine, N. and Sibson, R. (1971) *Mathematical taxonomy*. Wiley, London,

Jones, M. (1984) Woodland origins in a South Yorkshire parish. *Local Historian*, **16**, (2) 73-82.

Jones, M. (1993) *Sheffield's Woodland Heritage*. Second edition. Rotherham: Green Tree Publications.

Jones, M. and Talbot, E. (1995) Coppicing in urban woodlands: A progress report on a multi-purpose feasibility study in the City of Sheffield. Journal of Practical Ecology and Conservation. *Journal of Practical Ecology and Conservation*, **1**, (1) 46-51.

Kent, M. and Coker, P. (1996) *Vegetation Description and Analysis*. John Wiley & Sons Inc., New York.

Kirby, K.J. (1997) The distribution and growth of bramble (*Rubus fruticosus*) in British semi-natural woodland and the implications for nature conservation. *Journal of Practical Ecology and Conservation*, **2**, (1), 31-45.

Kirby, K.J., Bines, T. Burn, J., Mackintosh, P., Pitkin, P. and Smith, I. (1986) Seasonal and observer differences in vascular plant records from British woodlands. *Journal of Ecology*, **74**, 123-131.

Krebs, C.J. (1985) *Ecology. The Experimental Analysis of Distribution and Abundance. Third Edition.* Harper and Row, New York.

Koerner, W., Dupouey, J.L., Dambrine, E. and Benoit, M. (1997) Influence of past land use on the vegetation and soils of present day forests in the Vosges mountains, France. *Journal of Ecology*, **85**, 351-358.

Legendre, P. and Legendre, L. (1998) *Numerical Ecology. Second Edition.* Elsevier Science, Amsterdam.

Lewontin, R.C. and Prout, T. (1956) Estimation of the number of different classes in a population. *Biometrics*, **12**, 211-223.

Linhart, Y.B. and Whelan, R.J. (1980) Woodland regeneration in relation to grazing and fencing in Coed Gorswen, North Wales. *Journal of Applied Ecology*, **17**, 827-840.

Littell, R. C., Milliken, G. A., Stroup, W, W., Wolfinger, R. D. (1996) *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC, USA.

Malloch, A.J.C. (1997) MATCH version 2: A computer program to aid the assignment of vegetation data to the communties and sub-communities of the National Vegetation.

MapInfo Corporation (1996) *MapInfo Professional (Version 4.1) User's Guide*. MapInfo Corporation, Troy, New York.

Matlack, G.R. (1994) Plant species migration in a mixed-history forest landscape in eastern North America. *Ecology*, **75**, (5) 1491-1502.

Marquiss, M., Vickers, A.D. and Picozzi, N. (1997) *Baseline survey of pinewood birds in Cairngorms proposed special protection area, Glenmore Forest SSI and Ballochbuie Forest*. Report produced by ITE, contract number T08050u5, for SNH on contract number RASD/116/97 N2K.

McGeoch, M.A. and Chown, S.L. (1998) Scaling up the value of bioindicators. *Trends in Ecology and Evolutionary Time*, **13**, (2), 46-47.

Medforth, P. and Rotherham, I.D. (1997) The birds of Ecclesall Woods. Peak District *Journal of History and Archeology, Special Publication, No. 1 July 1997*, 21-33.

Meot, A., Legendre, P. and Bocard, D. (1998) Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework. *Environmental and Ecological Statistics*, **5**, 1-27.

Merryweather, J. and Fitter, A. (1998) The arbuscular mycorrhizal fungi of *Hyacinthoides non-scripta* – II. Seasonal and spatial patterns of fungal populations. *New Phytologist*, **138**, (1) 131-142.

Microsoft Corporation (1997) *Excel 97 Users Guide.* Microsoft corporation.

Miles, J. and Kinnaird, JW. (1979) The establishment and regeneration of birch, juniper and Scots pine in the Scottish Highlands. *Scottish Forestry*, **33,** 102-119.

Mitchell, F.J.G. & Kirby, K.J. (1990) The Impact of large herbivores on the conservation of semi-natural woods in the British uplands. *Forestry*, 63, (4), 333-353.

Moreno, C.E and Halffter, G. (2000) Assessing the completeness of bat biodiversity inventories using species accumulation curves. *Journal of Applied Ecology*, **37**, 149-158.

NCC (1986) *South Yorkshire Inventory of Ancient Woodland (Provisional).* Nature Conservancy Council.

Ney-Nifle, M. and Mangel, M. (1999) Species-area curves based on geographic range and occupancy. *Journal of Theoretical Biology*, **196**, 327-342.

Northwood Geoscience Ltd. (1996) *Vertical Mapper. Contour modeling and display software for MapInfo Desktop Software (Version 1.5).* Northwood Geoscience, Ontario, Canada.

Okland, R. H. (1999) On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, **10**, 131-136.

Oksanen, J. and Minchin, P.R. (1997) Instability of ordination results under changes in input data order: explanations and remedies. *Journal of Vegetation Science*, **8**, 447-454.

Okubu, A. and Levin, S.A. (1989) A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology*, **61**, 431-540.

Orton, P.D. (1986) Fungi of northern pine and birch woods. *Bulletin of the British Mycological Society*, **20**, (2), 130-145.

Packham, J.R. and Coin, E.V.J. (1990) Ecology of the woodland field layer. *Arboricultural Journal*, **24**, 357-371.

Packham, J.R. and Willis, A.J. (1982) The influence of shading and soil type on the growth of *Galeobdolon Luteum*. *Journal of Ecology*, **70**, 491-512.

Packham, J.R. and Willis, A.J. (1976) Aspects of the ecological amplitude of two woodland herbs, *Oxalis acetosella* L. and *Galeobdolon luteum* Huds. *Journal of Ecology*, **64**, 485-510.

Packham, J.R. and Willis, A.J. (1977) The effects of shading on *Oxalis acetosella*. *Journal of Ecology*, **65**, 619-642.

Palmer, M.W. (1990) The estimation of species richness by extrapolation. *Ecology*, **71**, (3), 1195-1198.

Palmer, M.W. (1993) Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology*, **74**, 2215-2230.

Palmer, M.W. (1999) *The Ordination Web Page*.http://www.okstate.edu/artsci/botany/ordinate/varpar.htm

Peterken, G.F. (1974) A method for assessing woodland flora for conservation using indicator species. *Biological Conservation*, **6**, 239-45.

Peterken, G.F. (1993) *Woodland conservation and Management*. Second edition. *Chapman & Hall, London.*

Peterken, G.F. and Game, M. (1981) Historical factors affecting the distribution of *Mercurialis perennis* in central Lincolnshire. *Journal of Ecology*, **69**, 781-796.

Peterken, G.F. and Game, M. (1984) Historical factors affecting the number and distribution of vascular plant species in the woodlands of central Lincolnshire. *Journal of Ecology*, **72**, 155-182.

Petersen, P.M. (1994) Flora, vegetation and soil in broadleaved ancient and planted and scrub on Rosnaes, Denmark. *Nordic Journal of Botany*, **14**, 693-709.

Picozzi, N., Catt, D.C. & Moss, R. (1992) Evaluation of Capercaillie habitat. *Journal of Applied Ecology*, **29**, 751-762.

Pigott, C.D. (1969) The status of *Tilia cordata* and *T. platyphyllos* on the Derbyshire limestone. *Journal of Ecology*, 57, 491-504.

Pigott, C.D. (1983) Regeneration of oak-birch woodland following exclusion of sheep. *Journal of Ecology*, 71, 629-646.

Pigott, C.D. (1990) The influence of evergreen coniferous nurse-crops on the field layer in two woodland communities. *Journal of Applied Ecology*, **27**, 448-459.

Pigott, C. D. and Huntley, J. P. (1981) Factors controlling the distribution of *Tilia cordata* Mill at the northern limits of its geographical range. III, Nature and causes of seed sterility. *New Phytology*, 87, 817-839.

Pitkin, P.H., Lusby, P.S. and Wright, J. (1994) Biodiversity and the ecology of pinewood plants. In: *Our Pinewood Heritage* (Ed. J. R. Aldhous).Forestry Commission, Surrey.

Pollard, E. and Yates, T.J. (1993) *Monitoring Butterflies for Ecology and Conservatio*n. Chapman & Hall, London.

Rackham, O. (1995) *Trees and Woodland in the British Landscape; The complete history of Britain's trees, woods and hedgerows*. Weidenfeld & Nicolson.

Rodwell, J.S. (1991) *British Plant Communities. 1. Woodlands and Scrub*. Cambridge University Press, Cambridge.

Rodwell, J.S. and Cooper, E.A. (1995) Scottish pinewoods in a European context. In: *Our Pinewood Heritage* (Ed. J. R. Aldhous). Forestry Commission, Surrey.

Rose, F. (1976) Lichenological indicators of age and environmental continuity in woodlands. In: *Lichenology: Progress and problemss (Systematics Special Volume 8)* (Eds. D.H. Brown, D.L. Hawksworth and R.H. Bailey) pp. 297-307. Academic Press, Lomndon.

Rose, F. (1992) Temperate forest management: its effect on bryophyte and lichen floras and habitats. In: *Bryophytes and Lichens in a Changing Environment* (Eds.J.W. Bales and A.M. Farmer) pp.211-233. Clarendon Press, Oxford.

Rose, F. (1999) Indicators of ancient woodland: The use of vascular plants in evaluating ancient woods for nature conservation. *British Wildlife,* **11**, 241-251.

Rotherham, I.D. and Doram, G.P. (1990) A preliminary study of the vegetation of Ecclesall Woods in relation to former management. *Sorby Record*, **27**, 60-70.

Salisbury, E.J. (1942) The Reproductive Capacity of Plants. Bell, London.

SAS Institute (1996a*) SAS/STAT Software: Changes and Enhancements through 6.11*. SAS Institute Inc., Cary, NC, USA.

SAS Institute (1996b) *SAS/STAT Technical Report: Spatial Prediction Using the SAS System*. SAS Institute Inc., Cary, NC, USA.

Sawada, M. (2000) *Rook's Case*. http://aix1.uottowa.ca/academic/ar…ections1/reports /moransi/moran.

Simberloff, D. (1998) Flagships, umbrellas, and keystones: is single-species management passe in the landscape era? *Biological Conservation*, **83**, 247-257.

Shirreffs, D.A. (1985) Biological Flora of the British Isles: *Anemone nemorosa* L. *Journal of Ecology*, **73**, 1005-1020.

Smilauer, P. (1992) *CanoDraw User's Guide (Version 3.0)*. Microcomputer Power. Ithaca, NY, USA.

Sneath, P.H.A. and Sokal, R.R. (1973) Numerical taxonomy – The principals and practice of numerical classification. W. H. Freeman, San Francisco.

Soberon J.M. and Llorente, B. (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, **7**, (3) 480-488.

Spencer, J.W. (1990) Indicators of antiquity; Some observations on the nature of plants associated with ancient woodland. *British Wildlife*, **2**, (2), 90-102.

Spencer, J.W. and Kirby, K.J. (1992) An inventory of ancient woodland for England and Wales. *Biological Conservation*, **62**, 77-93.

Stamp, N.E. and Lucas, J.R. (1983) Ecological correlates of explosive seed dispersal. *Oecologia*, **59**, 272-278.

Steven, H.M. and Carlisle, A. (1959) *The Native Pinewoods of Scotland*. Oliver and Boyd, Edinburgh.

Struik, G.J. (1965) Growth patterns of some native annual and perennial herbs in southern Wisconsin. *Ecology*, **46**, 401-420.

Stubbs, A.E. (1982) Hoverflies as primary woodland indicators with reference to Wharncliff Wood. *Sorby Record*, **20**, 62-67.

Summers, R.W., Moss, R. and Helliwell, E.C. (1994) Scotland's native pinewoods – the requirements of birds and mammals. In: *Our Pinewood Heritage* (Ed. J. R. Aldhous). Forestry Commission, Surrey.

Sykes, J.J., Horrill, A.D., Mountford, M.D. (1983) Use of visual cover assessments as quantitative estimators of some British woodland taxa. *Journal of Ecology*, **71**, 437-450.

Ter Braak, (1986) Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, (5), 1167-1179.

Ter Braak, C.J.F. and Prentice, I.C. (1988) A theory of gradient analysis. *Advances in Ecological Research,* **18**, 271-317.

Ter Braak, C.J.F. and Smilauer, P. (1998) CANOCO Reference Manual and User's Guide to Canoco for Windows. Centre for Biometry Wageningen, Wageningen, The Netherlands.

Terrell-Nield, C. (1990) Is it possible to age woodlands on the basis of their carabid diversity? The Entomologist, 109, (3) 136-145.

Usher, M. B. (ed.) (1986) *Wildlife Conservation Evaluation*. Chapman & Hall, London.

Verbeke, G. (1997) Linear mixed models for longitudinal data. In: *Linear Mixed Models in Practice. Lecture notes in statistics. No. 126.* (Eds. G. Verbeke and G. Molenberghs). Springer-Verlag, New York.

Vickers, A.D. and Palmer S.C.F. (2000) The influence of canopy cover and other factors upon the regeneration of Scots pine ant its associated ground flora within Glen Tanar National Nature Reserve. *Forestry*, **73**, (1) 37-49.

Vickers, A.D. and Rotherham, I.D. (2000) The response of *Bluebell (Hyacinthoides non-scripta*) to seasonal differences between years and woodland management. *Aspects of Applied Biology*, **58**, 351-356.

Vickers, A.D., Rotherham, I.D. and Rose, J.C. (2000) Vegetation succession and colonisation rates at the forest edge under different environmental conditions. *Aspects of Applied Biology*, **58**, 1-8.

Warhaugh, A.A.(1997) Terrestrial Molluscan fauna of Wilton Wood, north-east Yorkshire. *Yorkshire Naturalist*, **122**, 19-23.

Waring, P. (1989) Comparison of light trap catches in deciduous and coniferous woodland habitats. *Entomologists Record*, **101**, 1-10.

Watkins, C. (1990) *Britain's Ancient Woodland. Woodland Management and Conservation*. David and Charles, Newton Abbot.

Waugh, D. (1990) *Geography, An Integrated Approach*. Thomas Nelson and Sons Ltd.

Whitney, G.G. and Foster, D.R. (1988) Overstorey composition and age as determinantsof the understorey flora of woods of central New England. *Journal of Ecology*, **76**, 867-877.

Willson, M.F. (1993) Dispersal mode, seed shaddows, and colonization patterns. *Vegetatio*, **107/108**, 261-280.

Wulf, M. (1997) Plant species as indicators of ancient woodland in northwestern Germany. *Journal of Vegetation Science*, **8**, 635-642.

Young, M. and Armstrong, G. (1994) The effect of age, stand density and variability on insect communities in native pine woodlands. In: *Our Pinewood Heritage* (Ed. J.R. Aldhous).

Zar, J.H. (1996) *Biostatistical Analysis. Third Edition*. Prentice-Hall International (UK) Limited, London.

APPENDIX 1
SAS SYNTAX PROGRAM TO PERFORM KRIGING

SAS syntax program to calculate semi-variance, fit an exponential model to the data and perform kriging. The data is from Chapter 1 and simply contains the CCA sample scores of axes 1 and 2 and the log%cover values of *Hyacinthoides non-scripta*

```
proc g3d data=transect.semivar;
     scatter cca1*cca2=hyacnon /
       rotate=55
       zmin=0.00
       zmax=5.0
       grid
 xticknum=5 yticknum=5 zticknum=6;
run;


proc variogram data=transect.semivar
     outdistance = outd;
     compute nhc=15 novariogram;
     coordinates xc=cca2 yc=cca1;
     var hyacnon;
run;


proc print data=outd;
run;

data outd; set outd;
   mdpt = round((lb+ub)/2,.1);
run;

proc gchart data=outd;
  vbar mdpt / type=sum sumvar=count discrete;
run;


proc variogram data=transect.semivar
     outv = outv;
     compute lagd=0.3341 maxlag=10;
     coordinates xc=cca1 yc=cca2;
     var hyacnon;
run;

proc print data=outv;
run;

data outv2; set outv;
   vari = variog; type = 'regular'; output;
run;

symbol1 i=join l=1 v=star c=black;

axis1 minor=none label=(c=black 'lag distance') offset=(3,3);
axis2 minor=(number=1)
    label=(c=black 'variogram') offset=(3,3);

proc gplot data=outv2;
```

```
  plot vari*distance=type / vaxis=axis2 haxis=axis1;
run;

proc nlin data=outv2 method=gauss;
  parms b0=0.3
        b1=4.0
        b2=5.0;
  model variog = b0+b1*(1-exp(-lag/b2));
  output out=b p=yhat r=yresid;
run;


data outv3; set outv;
  c = 0.2704; co=4.0543; ao=4.4587;
  vari = c+co*(1-exp(-lag/ao));
  type = 'Exponential'; output;
  vari = variog; type = 'regular'; output;
run;

symbol1 i=join l=1 v=star c=black;
symbol2 i=join l=1 v=square c=blue;

proc gplot data=outv3;
  plot vari*distance=type / vaxis=axis2 haxis=axis1;
run;

proc krige2d data=transect.semivar outest=est;
   pred var=hyacnon radius=4.1758 maxpoints=35 minpoints=16;
   model
         nugget=0.2704
         scale=4.0543
         range=4.4587
         form=exponential;
coord xc=cca2 yc=cca1;
  grid x=-1.5 to 2 by 0.1 y=-1.5 to 2 by 0.1;
run;

proc g3d data=est;
  plot gyc*gxc=estimate /
     grid
     ctop=black
     cbottom=grey
     ctext=black
     rotate=55
     zmin=0.00
     zmax=5.0
     caxis=black
     xticknum=6
     yticknum=7
     zticknum=10
     ;
  label gyc='CCA axis 1    '
  label gxc='CCA axis 2    '
  estimate ='log Hyacinthoides';
run;


proc g3d data=est;
   plot gyc*gxc=stderr /
     grid
```

```
     ctop=black
     cbottom=grey
     ctext=black
     rotate=55
     zmin=0.00
     zmax=0.5
     caxis=black
     xticknum=8
     yticknum=9
     zticknum=6
     ;
  label gyc='CCA axis 1      '
  label gxc='CCA axis 2      '
  estimate = 'Standard Error   ';
run;
```

APPENDIX 2A
SAS SYNTAX PROGRAMS USED IN CHAPTER 2

1.) Testing for linear and quadratic relationships of *Hyacinthoides non-scripta* with distance; treating distance as a categorical variable. Diagnostic plots are included.

```
proc mixed data=transect.alldata ratio covtest;
  class transect wood distance;
  model hyacnon = wood distance distance*wood  /
    ddfm=satterth solution predicted;
  random transect(wood);
  repeated distance / sub=transect group=wood type=ar(1) rcorr=1,6,11,16,21;
  lsmeans wood / adjust=tukey;
  lsmeans distance / adjust=tukey;
  lsmeans distance*wood;
  contrast 'linear' distance -9 -7 -5 -3 -1 1 3 5 7 9;
  contrast 'quadratic' distance -6 -2 1 3 4 4 3 1 -2 -6;
  make 'solutionr' out = random noprint;
  make 'predicted' out = pred noprint;
  make  'rcorr1' out = rcorr1 noprint;
  make  'rcorr6' out = rcorr6 noprint;
  make  'rcorr11' out = rcorr11 noprint;
  make  'rcorr16' out = rcorr16 noprint;
  make  'rcorr21' out = rcorr21 noprint;
  make 'lsmeans' out = lsmeans;
run;

/*Test data are statistically different from normal distribution and draw simple plots*/
proc univariate plot data = pred normal;
  var _resid_;
run;

/*Draw normal Q-Q plots with 2*standard error lines and also a plot of the deviation from
noramal with error lines (Friendly, 1991)*/
%inc 'nqplot.sas';
%nqplot (data=pred, var=_RESID_, mu=MEAN, sigma=STD);

/*Draw histogram of residuals with fitted normal distribution*/
proc capability data=pred noprint graphics;
  var _resid_;
histogram _resid_/
  normal
  midpoints =-2 to 2 by 0.5
  haxis=axis1 vaxis=axis2
  noframe legend=legend1
  caxis=blue
  ctext=blue
  cbarline=black
  name='hist_normal';
axis1 label=(h=1.5)  value=(h=1.3);
axis2 label=(h=1.5)  value=(h=1.3);
legend1 label=(h=1.5) value=(h=1.3);
run;


  axis1
    color=blue
```

```
      width=2.0
      ;
   axis2
      color=blue
     width=2.0
      ;
   axis3
     color=blue
      width=2.0
     ;

  symbol1 c=DEFAULT
    i=NONE
    v=STAR;

 proc gplot data=pred;
   plot _resid_ * _pred_ /
         haxis=axis1
         vaxis=axis2;
 run;
 symbol1 c=DEFAULT
    i=RLCLI95
    v=STAR;

   proc gplot data=WORK.PRED ;
     plot hyacnon * _PRED_ /
         haxis=axis1
        vaxis=axis2
          ;
   run;

 data work.out;
 merge transect.alldata work.pred;

 proc sort data=work.out;
 by wood;
 run;

  goptions device=WIN  ctext=blue
          graphrc interpol=join;

   symbol1 c=DEFAULT
      i=JOIN
      v=DIAMOND
     ;
  symbol2 c=DEFAULT
      i=JOIN
      v=DOT
      ;
   symbol3 c=DEFAULT
      i=JOIN
      v==
      ;
   symbol4 c=DEFAULT
      i=JOIN
      v=DIAMOND
      ;
   symbol5 c=DEFAULT
      i=JOIN
      v=CIRCLE
```

```
      ;
 /*plot residuals for values along each transect by wood to check for any visible
 autocorrelation*/
   proc gplot data=WORK.OUT ;
      by WOOD;
      plot _RESID_ * DISTANCE  = TRANSECT /
         haxis=axis1
         vaxis=axis2
         ;
   run;

  pattern1 value=SOLID;

   axis1
      color=blue
      width=2.0
      ;
   axis2
      color=blue
     width=2.0
      ;
   axis3
     color=blue
      width=2.0
     ;

 /*plot predicted values for each site*/
 proc gchart data=WORK.LSMEANS;
      vbar WOOD /

         maxis=axis1
         raxis=axis2
         patternid=midpoint
         type=MEAN
         sumvar=_LSMEAN_
         ;
   run;

 symbol1 c=DEFAULT
      i=SPLINE
      l=1
      v=DIAMOND;


 /*plot Autoregressive error structure for each site*/
 title1 "GP";
  proc gplot data=WORK.RCORR1;
     plot COL1 * ROW /
         haxis=axis1
         vaxis=axis2;
 run;

 title1 "LMW";
 proc gplot data=WORK.RCORR6;
     plot COL1 * ROW /
         haxis=axis1
         vaxis=axis2;
 run;

  title1 "NW";
```

```
proc gplot data=WORK.RCORR11;
    plot COL1 * ROW /
        haxis=axis1
        vaxis=axis2;
run;

title1 "OSW";
proc gplot data=WORK.RCORR16;
    plot COL1 * ROW /
        haxis=axis1
        vaxis=axis2;
run;

title1 "SW";
proc gplot data=WORK.RCORR21;
    plot COL1 * ROW /
        haxis=axis1
        vaxis=axis2;
run;
```

2.) Testing for quadratic relation of Hyacinthoides non-scripta with distance in each wood; treating distance as a categorical variable. Nested model.

(The diagnostic tests carried out are not shown – same as in the previous model)

```
proc mixed data=transect.alldata ratio covtest;
  class transect wood distance;
  model hyacnon = wood distance(wood)  /
    ddfm=satterth solution predicted;
  random transect(wood) / s;
  repeated distance / sub=transect group=wood type=ar(1) rcorr=1,6,11,16,21;
  lsmeans wood / adjust=tukey;
  contrast 'linear gp' distance(wood)     -9 -7 -5 -3 -1 1 3 5 7 9;
  contrast 'quadratic gp' distance(wood)  -6 -2 1 3 4 4 3 1 -2 -6;
  contrast 'linear lmw' distance(wood)    0 0 0 0 0 0 0 0 0 0 -9 -7 -5 -3 -1 1 3 5 7 9;
  contrast 'quadratic lmw' distance(wood) 0 0 0 0 0 0 0 0 0 0 -6 -2 1 3 4 4 3 1 -2 -6;
  contrast 'linear nw' distance(wood)     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -9 -7 -5
-3 -1 1 3 5 7 9;
  contrast 'quadratic nw' distance(wood)  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -6 -2 1 3
4 4 3 1 -2 -6;
  contrast 'linear osw' distance(wood)    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
                                          0 0 0 0 0 0 0 0 0 0 -9 -7 -5 -3 -1 1 3 5 7 9;
  contrast 'quadratic osw' distance(wood) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
                                          0 0 0 0 0 0 0 0 0 0 -6 -2 1 3 4 4 3 1 -2 -6;
  contrast 'linear sw' distance(wood)     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
                                          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -9 -7 -5
-3 -1 1 3 5 7 9;
  contrast 'quadratic sw' distance(wood)  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
                                          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -6 -2 1 3
4 4 3 1 -2 -6;
  make 'solutionr' out = random noprint;
  make 'predicted' out = pred noprint;
  make  'rcorr1' out = rcorr1 noprint;
  make  'rcorr6' out = rcorr6 noprint;
  make  'rcorr11' out = rcorr11 noprint;
  make  'rcorr16' out = rcorr16 noprint;
  make  'rcorr21' out = rcorr21 noprint;
```

```
  make 'lsmeans' out = lsmeans;
run;
```

3.) Unequal slopes model for *Hyacinthoides non-scripta.*

(Transects were coded 1 to 25 so transect(wood) is not needed in the repeated statement.) Diagnostic tests are given in the model 1, Appendix 2A)

```
proc mixed data=transect.alldata ratio covtest;
  class transect wood dist;
  model hyacnon = wood distance distance*wood distance*distance distance*distance*wood /
     ddfm=satterth solution predicted htype=1,3;
  lsmeans wood / adjust=tukey;
  estimate 'b1gp-b2lmw' distance*wood  1 -1 0 0 0;
  estimate 'b1gp-b3nw' distance*wood   1 0 -1 0 0;
  estimate 'b1gp-b4osw' distance*wood  1 0 0 -1 0;
  estimate 'b1gp-b5sw' distance*wood   1 0 0 0 -1;
  estimate 'b2lmw-b3nw' distance*wood  0 1 -1 0 0;
  estimate 'b2lmw-b4osw' distance*wood 0 1 0 -1 0;
  estimate 'b2lmw-b5sw' distance*wood  0 1 0 0 -1;
  estimate 'b3nw-b4osw' distance*wood  0 0 1 -1 0;
  estimate 'b3nw-b5sw' distance*wood   0 0 1 0 -1;
  estimate 'b4osw-b5sw' distance*wood  0 0 0 1 -1;
  estimate 'b1gp-b2lmw' distance*distance*wood  1 -1 0 0 0;
  estimate 'b1gp-b3nw' distance*distance*wood   1 0 -1 0 0;
  estimate 'b1gp-b4osw' distance*distance*wood  1 0 0 -1 0;
  estimate 'b1gp-b5sw' distance*distance*wood   1 0 0 0 -1;
  estimate 'b2lmw-b3nw' distance*distance*wood  0 1 -1 0 0;
  estimate 'b2lmw-b4osw' distance*distance*wood 0 1 0 -1 0;
  estimate 'b2lmw-b5sw' distance*distance*wood  0 1 0 0 -1;
  estimate 'b3nw-b4osw' distance*distance*wood  0 0 1 -1 0;
  estimate 'b3nw-b5sw' distance*distance*wood   0 0 1 0 -1;
  estimate 'b4osw-b5sw' distance*distance*wood  0 0 0 1 -1;
  random transect(wood) / s;
  repeated dist / subject=transect group=wood type=ar(1) rcorr=1,6,11,16,21;
  make 'solutionr' out = random noprint;
  make 'predicted' out = pred noprint;
  make  'rcorr1' out = rcorr1 noprint;
  make  'rcorr6' out = rcorr6 noprint;
  make  'rcorr11' out = rcorr11 noprint;
  make  'rcorr16' out = rcorr16 noprint;
  make  'rcorr21' out = rcorr21 noprint;
  make 'lsmeans' out = lsmeans;
run;
```

4.) Fitting poisson regression models using the generalised linear model procedure and compensating for over-dispersion.

Diagnostic tests are given in the model 1, Appendix 2A.

```
proc genmod data=transect.species;
model alliurs = distance / dist=poisson link=log scale=pearson;
run;

proc genmod data=transect.species;
model alliurs = distance distance*distance / type1 dist=poisson link=log scale=pearson;
```

run;

5.) Non-linear model used to fit a negative exponential model to the estimated canopy cover and tree density

```
proc nlin data=transect.alldata;
 parms b0=-0.5 b1=2 b2=1.0;
 model arcsine = b0+b1*(1-exp(-b2*sqrttree));
 output out=b p=yhat r=yresid;
run;

proc sort data=b;
  by sqrttree;
run;

   symbol1 c=DEFAULT
     i=NONE
      v=STAR
      ;
   symbol2 c=DEFAULT
      i=SPLINE
      l=1
      v=NONE
      ;
  proc gplot data=WORK.B;
    plot (ARCSINE YHAT  ) * SQRTTREE / overlay
        haxis=axis1
        vaxis=axis2
         ;
  run;

  proc gplot data=WORK.B ;
     plot YRESID * SQRTTREE /
       haxis=axis1
       vaxis=axis2
       ;
  run;
```

6.) Testing CCA axes using standardised variables and a dependent error structure.

Weight equals the weights given to each sample by CANOCO, which is related to the total abundance of species in each sample, after down weighting of rare species.
Diagnostic tests are given in the model 1, Appendix 2A.

```
proc mixed data=transect.alldata ratio covtest info ic;
  class transect wood distance texture;
  model s_cca1 = s_distan s_arcsin s_lnslop s_sine s_cosine s_ph s_depth s_p_wate
    s_p_orga s_sand s_silt s_clay s_loam / noint ddfm=satterth solution predicted;
  weight weight;
  random wood transect(wood) / s;
  repeated distance / subject=transect group=wood type=ar(1) rcorr=1,6,11,16,21;
  estimate 'gp'  | wood 1 0 0 0 0;
  estimate 'lmw' | wood 0 1 0 0 0;
  estimate 'nw'  | wood 0 0 1 0 0;
  estimate 'osw' | wood 0 0 0 1 0;
```

```
   estimate 'sw'   | wood 0 0 0 0 1;
   estimate 'gp-lmw' | wood 1 -1 0 0 0;
   estimate 'gp-nw ' | wood 1 0 -1 0 0;
   estimate 'gp-osw' | wood 1 0 0 -1 0;
   estimate 'gp-sw'  | wood 1 0 0 0 -1;
   estimate 'lmv-nw' | wood 0 1 -1 0 0;
   estimate 'lmv-osw'| wood 0 1 0 -1 0;
   estimate 'lmv-sw' | wood 0 1 0 0 -1;
   estimate 'nw-osw' | wood 0 0 1 -1 0;
   estimate 'nw-sw'  | wood 0 0 1 0 -1;
   estimate 'osw-sw' | wood 0 0 0 1 -1;
   make 'solutionr' out = random noprint;
   make 'predicted' out = pred noprint;
   make  'rcorr1' out = rcorr1 noprint;
   make  'rcorr6' out = rcorr6 noprint;
   make  'rcorr11' out = rcorr11 noprint;
   make  'rcorr16' out = rcorr16 noprint;
   make  'rcorr21' out = rcorr21 noprint;
run;
```

7.) Fitting a quadratic regression surface model for *Allium ursinum* to the first two CCA axes

```
proc genmod data=transect.cubicsur;
  model alliurs = cca1 cca2 cca1x2 cca1x1 cca2x2
   / dist=poisson link=log scale=pearson wald type1 maxit=50;

run;
```

8.) Drawing a quadratic regression surface for *Allium usinum*

```
data transect.allium; set transect.gridfile;
  a = -2.2165 ; b1 = -3.4368 ; b3 = -1.279; b5 = 1.154 ;
cover = exp(a+(b1*cca1)+(b3*cca1*cca2)+(b5*cca2*cca2))
        ;
label cover = '% COVER';
run;


 proc g3d data=TRANSECT.allium ;
     plot cca2 * cca1 = COVER /
       rotate=35

       zmin=0.00
       zmax=100
       grid

       xticknum=8
       yticknum=8
       zticknum=6
           ;
  run;
```

9.) Flexible clustering of CCA axes 1-8.

Plots of the CCA axes are included showing the clusters to which the sample scores belong.

```
proc cluster data=transect.cca18 method=flexible beta=-0.5 outtree=work.tree rsquare
nonorm;
  var cca1--cca8;
  id sample;
run;

proc tree data=work.tree out=work.treeout n=15 graphics height=r  ntick=5
name=adv inc=0.2 tickpos=5;
id sample;
run;

proc sort;
by sample;
run;

data scatter;
merge transect.cca18 work.treeout;
run;

proc sort;
  by cluster;
run;

goptions reset=(axis, legend, pattern, symbol, title, footnote) norotate
    hpos=0 vpos=0 htext= ftext= ctext= target= gaccess= gsfmode= ;
    goptions device=WIN  ctext=blue
    graphrc interpol=join;

    symbol1 c=DEFAULT
    i=NONE
    v=DOT
    cv=RED
   ;
   symbol2 c=DEFAULT
   i=NONE
   v=CIRCLE
   cv=BLUE
   ;
  symbol3 c=DEFAULT
    i=NONE
    v=PLUS
   cv=BLUE
    ;
  symbol4 c=DEFAULT
   i=NONE
   v=STAR
   cv=MAGENTA
   ;
  symbol5 c=DEFAULT
   i=NONE
   v=square
  cv=black
   ;
  symbol6 c=DEFAULT
    i=NONE
    v=DOT
    cv=CYAN
   ;
```

```
  symbol7 c=DEFAULT
  i=NONE
  v=CIRCLE
  cv=RED
  ;
 symbol8 c=DEFAULT
   i=NONE
   v=PLUS
  cv=RED
   ;
 symbol9 c=DEFAULT
  i=NONE
  v=DOT
  cv=BLUE
  ;
 symbol10 c=DEFAULT
  i=NONE
  v=square
 cv=BLUE
  ;
 symbol11 c=DEFAULT
  i=NONE
  v=square
 cv=black
  ;
 symbol12 c=DEFAULT
   i=NONE
   v=DOT
   cv=cyan
  ;
  symbol13 c=DEFAULT
  i=NONE
  v=CIRCLE
  cv=green
  ;
 symbol14 c=DEFAULT
  i=NONE
  v=PLUS
 cv=cyan
  ;
symbol15 c=DEFAULT
 i=NONE
 v=STAR
 cv=BLACK
 ;

axis1
length=60 PCT
color=blue
width=2.0
style=1
;
axis2
 color=blue
 width=2.0
 ;
axis3
color=blue
width=2.0
;
```

```
   proc gplot data=WORK.SCATTER ;
      plot cca2 * cca1  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;


    proc gplot data=WORK.SCATTER ;
      plot cca4 * cca3  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;

    proc gplot data=WORK.SCATTER ;
      plot cca6 * cca5  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;

proc gplot data=WORK.SCATTER ;
      plot cca8 * cca7  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;
```

APPENDIX 2B
COMPARING GLMMs FROM CHAPTER 2

Full table giving the model fitting statistics produced by PROC MIXED for the random slopes models as well for the models with time series error structures.
s_dist refers to the standardised distance with mean zero and variance of one.
AIC and SBC values closest to zero are in **bold type**.

Random statement	Repeated statement	No. of Covariance Parameters	Criteria	CCA			
				axis 1	axis 2	axis 3	axis
wood transect(wood)	distance / sub=transect	17	AIC	**-127.5**	**-97.2**	-193.4	**-221**
	group=wood		SBC	-157.0	-126.7	-223.0	-250
	type=arma(1,1)		-2REML	221.0	160.3	352.9	408
wood transect(wood)	distance / sub=transect	5	AIC	-154.1	-105.1	-195.7	-231
	type=arma(1,1)		SBC	-162.7	-113.8	-204.4	-240
			-2REML	298.1	200.2	381.3	453
wood transect(wood)	distance /sub=transect	12	AIC	**-127.5**	-101.6	**-190.6**	-224
	group=wood		SBC	**-148.3**	**-122.5**	**-211.5**	**-245**
	type = ar(1)		-2REML	231.0	179.3	357.3	424
wood transect(wood)	distance / sub=transect	4	AIC	-153.4	-104.7	-196.3	-230
	type= ar(1)		SBC	-160.4	-111.6	-203.2	-237
			-2REML	298.9	201.3	384.6	453
wood	distance / sub=transect	11	AIC	-138.0	-126.0	-201.2	-233
	group=wood		SBC	-157.1	-145.1	-220.3	-252
	type=cs		-2REML	253.9	-229.9	380.5	445
wood transect(wood)	__________	5	AIC	-158.1	-112.1	-202.7	-234
s_dist(transect)			SBC	-166.8	-120.7	-211.4	-243
s_dist*s_dist(transect)			-2REML	306.1	214.1	395.4	459
wood transect(wood)	__________	4	AIC	-158.2	-111.1	-202.6	-235
s_dist(transect)			SBC	-165.1	-118.0	-209.6	-241
			-2REML	308.4	214.1	397.2	461
wood transect(wood)	__________	3	AIC	-161.8	-124.8	-209.7	-243
			SBC	-167.0	-130.0	-214.9	-249
			-2REML	317.7	243.7	413.4	481
wood	__________	2	AIC	-179.6	-140.8	-212.6	-245
			SBC	-183.0	-144.2	-215.6	-249
			-2REML	355.1	277.6	420.3	487
__________	__________	1	AIC	-181.7	-156.4	-279.4	-265
			SBC	-183.4	-158.2	-218.1	-267
			-2REML	361.3	310.9	556.7	529

APPENDIX 3A
SAS SYNTAX PROGRAMS USED IN CHAPTER 3

1.) Spatial exponential GLM

```
proc mixed data=ancind.geology ratio scoring=50 covtest info ic;
  class age geology;
  model s_cca1 = s_altitu s_east s_north s_en age geology age*geology
 / s predicted htype=1,3;
  lsmeans age geology age*geology / adjust=tukey;
  weight weight;
  repeated / subject=intercept local type=sp(exp) (long lat);
  make 'predicted' out = pred noprint;
run;

proc univariate plot data = pred normal;
  var _resid_;
run;


%inc 'nqplot.sas';
%nqplot (data=pred, var=_RESID_, mu=MEAN, sigma=STD);


proc capability data=pred noprint graphics;
  var _resid_;
histogram _resid_/
  normal
  midpoints =-2 to 2 by 0.5
  haxis=axis1 vaxis=axis2
  noframe legend=legend1
  caxis=blue
  ctext=blue
  cbarline=black
  name='hist_normal';
axis1 label=(h=1.5)  value=(h=1.3);
axis2 label=(h=1.5)  value=(h=1.3);
legend1 label=(h=1.5) value=(h=1.3);
run;


proc insight data=pred;
  dist _resid_;
  fit s_cca1 = _pred_;
run;
```

2.) Generalised linear model for unequal slopes with binomial error distribution and logistic link function

```
%inc 'glmm612.sas';
%glimmix(data=ancind.logist, procopt=method=reml ratio covtest,
    stmts=%str(
      class geology;
      model ancient = geology indicato indicato*geology / solution;
      lsmeans geology / pdiff adjust=tukey;
```

```
       estimate 'bunter-coal' indicato*geology 1 -1 0;
       estimate 'bunter-mgp'    indicato*geology 1 0 -1;
       estimate 'coal-mgp' indicato*geology 0 1 -1;
     ),
     error=binomial,
     link=logit,
     output=reschi=_reschi_
     )
run;

proc univariate plot data=work._outfile normal;
   var _reschi_;
run;
```

3.) Cluster analysis

```
proc cluster data=ancind.geology method=ward outtree=work.tree rsquare nonorm;
  var lc1--lc8;
  id sample;
run;


proc tree data=work.tree out=work.treeout n=7 hpages=2 graphics height=r ntick=5
name=indicator inc=0.2 tickpos=5;
id sample;
run;

proc sort;
by sample;
run;

data scatter;
merge ancind.geology work.treeout;
run;


proc sort;
  by cluster;
run;


goptions reset=(axis, legend, pattern, symbol, title, footnote) norotate
    hpos=0 vpos=0 htext= ftext= ctext= target= gaccess= gsfmode= ;
    goptions device=WIN  ctext=blue
    graphrc interpol=join;

    symbol1 c=DEFAULT
    i=NONE
    v=DOT
    cv=RED
   ;
   symbol2 c=DEFAULT
   i=NONE
   v=CIRCLE
   cv=BLUE
   ;
  symbol3 c=DEFAULT
    i=NONE
```

```
    v=PLUS
   cv=BLUE
    ;
  symbol4 c=DEFAULT
   i=NONE
   v=STAR
   cv=MAGENTA
   ;
  symbol5 c=DEFAULT
   i=NONE
   v=square
  cv=black
   ;
  symbol6 c=DEFAULT
    i=NONE
    v=DOT
    cv=CYAN
   ;
   symbol7 c=DEFAULT
   i=NONE
   v=CIRCLE
   cv=RED
   ;
  symbol8 c=DEFAULT
    i=NONE
    v=PLUS
   cv=RED
    ;
  symbol9 c=DEFAULT
   i=NONE
   v=DOT
   cv=BLUE
   ;
  symbol10 c=DEFAULT
   i=NONE
   v=square
  cv=BLUE
   ;
  symbol11 c=DEFAULT
   i=NONE
   v=square
  cv=black
   ;

axis1
length=60 PCT
color=blue
width=2.0
style=1
;
axis2
 color=blue
 width=2.0
 ;
axis3
color=blue
width=2.0
;

proc gplot data=WORK.SCATTER ;
```

```
     plot lc2 * lc1  = CLUSTER /
        haxis=axis1
        vaxis=axis2
        ;
  run;


   proc gplot data=WORK.SCATTER ;
     plot lc4 * lc3  = CLUSTER /
        haxis=axis1
        vaxis=axis2
        ;
  run;

   proc gplot data=WORK.SCATTER ;
     plot lc6 * lc5  = CLUSTER /
        haxis=axis1
        vaxis=axis2
        ;
  run;

proc gplot data=WORK.SCATTER ;
     plot lc8 * lc7  = CLUSTER /
        haxis=axis1
        vaxis=axis2
        ;
  run;
```

Appendix 3B List of sites used in Chapter 3 with grid references

Agden Wood	42483937
Almond Road Plantation	46154014
Alport Bridge	41413899
Alverly Spring Wood	45553989
Ant Wood	45623952
Ash Holt	45533960
Ash Holt	46804004
Back Wood	46183999
Balby Little Moor	45654021
Banktop Hey Wood	41473937
Barnburgh Cliff	44994036
Barnsdale Wood	45244142
Barnurgh Park Quarry Wood	44864045
Barwick Bank	44104367
Beeston plantation	45904000
Bentley Railway Embankment	45734058
Bilham Park and Summerhouse Plantation	44834063
Bilham Park Fishpond Plantation	44844068
Bitholmes Wood	42913961
Blaxton Common	46874015
Bleakley Plantation	42143629
Bowden Head	40723821
Bowden Housestead (1)	43953867
Bowden Housestead (2)	43973868
Bowden Housestead (3)	43993868
Bowden Housestead (4)	43993868
Bowden Housestead (5)	43993868
Bowden Housestead (6)	43993868
Bowsen Wood	42663932
Brackens's Plantation	46394021
Bramwith Hall	46254115
Bramwith Lock Woods	46154110
Brecks Common	46234078
Brick Kiln Plantation	45574108
Broad Wood	42454072
Bunfold Shaw	46094159
Burberry's Holt	45603927
Burgwallis Grange Bank	45284113
Cadeby Common	45124001
Calf Croft	44744056
Calf Hall Wood	42834032
Calfhey Wood	41643916
Cambridge Wood	42593615
Campsmount Park	45374140
Carr Lodge Plantation	45814002
Castle Hill	44804055
Challenger Wood	44434081
Chapel Hole	45453949
ChurchLane Railway Cutting	45494011
Cliff Wood	42174002
Cock Wood	46624036
Cockhill Plantation	45513965
Conisborough Mill Piece	45163988
Conisbrough Viaduct Area (a)	45223995
Coombs Dale	42203742
Cozen Croft Wood	46354063
Cressbrook / Cramside Wood	41673730
Crown Pool Plantation	46384007
Crowther Wood	46224030
Cuckold Carr	42494074
Daw Lane Plantation	45694072
Dearne Valley Railway Embankment	45503996
Denaby Wood	44863990
Duck Holt	45604096
Duckholt Plantation	45604097
Dudwood	42233619
Ewden Coppice and Sun Wood	42333967
Ewden Wood	42513967
Ewden, Bottom and Broomhead Woods	42463967
Finningly Big Woods and Gravel Pits	46523982
Finningly Gravel Pits	46853997
Finningly School Sand Pit	46763992
Firbeck Jucntion	45803934
First-Third Plantation	44804071
Fishpond Wood	42003620
Flat Wood	42454074
Four Lane Ends Plantation	46204016
Fox Covert	46024086
Fryston Park (grassland)	44624255
Fryston Park (woodland)	44624254
Gelster Lane Holt	46283987
Goistock Wood	40804372
Golden Clough	39804280
Great Gate Wood	46704045
Green Busks Wood	46134000
Green Hills - Pickburn	45144076
Grimbocar Wood	41803866
Hagg Wood	41504105
Hagg Wood	46114052
Hall Royd Wood	42974044
Hamphole Dike	45204100
Hampole Wood	45004090
Hanging Wood	45354069
Hangman Stone Wood	45014032
Hatchell Wood - West	46234004
Hatfield Lings	46554075
Haw Park	43654155
Heather Wood	46054046
Hexthorpe Flats (a)	45454015
Hexthorpe Flats (b)	45584018

Cliff Wood	43644068
Clough Wood	42384062
Hickleton Spring	44794061
Highlow Wood	42233799
Hill Carr Wood	42573635
Hill Top Wood	42934050
Hoblehoy Wood	46264138
Holes Wood	42373905
Hollin Bridge Farm	46844085
Hollin Wood	42264005
Hollinhurst Wood (a)	44034288
Hollinhurst Wood (b)	44024288
Hollins Clough	40523988
Holme House Woods	40394407
Holme Wood	46594061
Holmes Carr Great Wood	46023978
Holt Rocher	42333967
Hooton Thorn Covert	44374093
Horse Carr Wood	43904064
Houghton Common (4)	44254087
Howell Wood	44374095
Hoyland Bank Wood	42704105
Hugsett Wood	43044067
Hurst Wood	46483986
Jone's Cable	47004149
Kennels Plantation	45004070
King's Wood, Bawtry	46503948
Kirk Moor Plantation (a)	46184020
Ladies Spring Wood	43253815
Lady Wood (1)	44204087
Levels Lane Plantation	46874016
Levitt Hagg Quariires	45384009
Leys Hill Plantation	45284065
Limestone Plantation	45263990
Lindholme Hall Area	47084063
Littlewood's Plantation	45834000
Lodge Pantations	46144022
Long Plantation	45324055
Long Plantation	45404044
Long Plantation, Edenthorpe	46304063
Lound Hill Quarry	44984080
Low Hollins Wood	45014122
Machin's Plantation	46784010
Mappleyard Plantation	44804032
Margery Wood	42754096
Marr - Ducker Holt	45284052
Marr Flats Wood	46393989
Marr Grange Holt	45304040
Marr Hills & Holes	45124052
Marsden Clough	41004070
Meadow Wood	41223733
Hexthorpe Flats (c)	45454020
Hezlock Field Plantation	45204024
New Close Wood	46414043
New Hall Wood	42593985
New Monkton	43954138
Newsholme Dean	40154403
North Soak Drain	47364122
Oak Moor & Chadwick Dike Area	47024164
Ogden Reservoir	40634309
Old Spring's Wood	46394012
Orchard, woodhouse	44273847
Owston Wood	45704101
Ox Carr Wood	46254033
Park Crow trees	46214020
Park Wood	46173989
Parkin Clough	41953850
Parkland Plantation (a)	46343974
Parlington Hollins	44154353
Pears House Clough	42213898
Pickle Wood	46773981
Pot Ridings Wood	45264002
Pot Ridings Wood	45304005
Pridock Wood	42063863
Randall Croft Wood	45584101
Raven Hill	45354040
Redhouse Plantation	46084029
Reedy Holme Plantation	46093968
River Went Oxbow	45734155
Rocher Wood	42603936
Rossington Brick Pond	46243986
Rossington Plantation	46363970
Rother Valley Country Park	44483839
Rough Wood	41803866
Sandal Beat	46124036
Scabba Wood	45254015
Scarbottom Mill Dam	40764212
Scorcher Hills Wood	45214123
Shaftholme	45784081
Sheepwalk Wood	42463645
Shirebrook Valley	44243838
Shirebrook Valley	44263839
Shittern Clough	40573955
Shrogs Plantation	44814057
Silkstone Fall	42984055
Sixteen Acre Plantation	45624100
Sixteen Acre Plantation	46323968
Skell Woods	45224118
Skiers Spring Wood	43683995
Smelting Hill Wood	42673665
Spout House Wood	42813951
Spring Rein	45384042

Melton College	45014010
Melton Wood	45104035
Moor Leys Wood	43324025
St Catherine's Plantation (a)	45664003
St Catherine's Railway	45704000
Spring Wood	41324124
Spring Wood	42304004
Squirrel Wood	45424114
Stables Holt	44964044
Stainton Little Wood	45563943
Stane Hill Plantation	45254058
Stangstry Wood	41284220
Stones Wood	41864105
Strines Wood	42243907
Sunny Bank Wood East	43884068
Sunny Bank Wood West	43834069
Swift Wood	42304066
Swinnow Wood	46323928
The Ashes	44854093
The Wilderness	44784066
Thorne Railway Delves	46804145
Thorne Railway Kirton Lane Area	46834122
Thorpe Arch Disused Railway	44394465
Tinker's Pond	46543976
Tinkers house Wood	41933874
Tithe Barn Plantaion	45354065
Toecroft Little Spring	45264030
Tranmoor Wood	46254038
Treeton	44273878
Ughill Wood	42623902
Vale Head Marsh	44264143
Valehouse Wood	40353979
Wadsworth Wood	45553975
Warmsworth Plantation	45473991
Went Hill Grassland	44764184
West End Wood	46143988
West Haigh Wood (10)	44234082
West Haigh Wood (13)	44254082
West Haigh Wood (17)	44274082
West Haigh Wood - heath (9)	44234083
West Haigh Wood SE (7)	44264082
Whincover Wood	42744059
Whinny Lane Plantation	46383977
Whitaker's Plantation	47264127
White Mires Wood	46343960
Woodfield Plantation	45723998
Woodhouse	44213853
Woodhouse	44213854
Wothersome Woods	44004436
Wrancarr Drain	46104122
Yateholme Plantations	41154050
Yew Tree Rein	45573983

APPENDIX 4
SAS SYNTAX PROGRAMS USED IN CHAPTER 4

1.) Simple exponential decline in the rate of finding new species

```
proc nlin data=time.buckwood;
 parms b1=80 b2=0.1;
 model species = b1*(1-exp(-b2*time));
 output out=b p=yhat r=yresid;
run;


symbol1 c=DEFAULT
    i=NONE
      v=STAR
     ;
  symbol2 c=DEFAULT
     i=SPLINE
     v=NONE
     ;

 proc gplot data=WORK.b;
      plot (species yhat) * TIME / overlay
        haxis=axis1
        vaxis=axis2
        ;
  run;

 proc gplot data=work.b;
     plot yresid * t;
  run;
```

2.) Decline in the rate of finding new speciesusing new model, which assumes that species are found in the order of their abundance and thatspecies abundance declines exponentially.

```
proc nlin data=time.buckwood;
 parms b1=0.01 b2=90 b3=0.03;
 model species = log(exp(b1*b2)-(exp(b1*b2)-1)*exp(-b3*t))/b1;
 output out=b p=yhat r=yresid;
run;


   symbol1 c=DEFAULT
    i=NONE
      v=STAR
     ;
  symbol2 c=DEFAULT
     i=SPLINE
     v=NONE
     ;

 proc gplot data=WORK.b;
      plot (species yhat) * TIME / overlay
```

```
         haxis=axis1
         vaxis=axis2
         ;
   run;

  proc gplot data=work.b;
      plot yresid * t;
   run;
```

3.) The Clench Equation

```
parms b1=3 b2=0.03;
 model species = b1*t/(1+b2*t);
 output out=b p=yhat r=yresid;
run;


    symbol1 c=DEFAULT
     i=NONE
       v=STAR
      ;
   symbol2 c=DEFAULT
      i=SPLINE
      v=NONE
      ;

  proc gplot data=WORK.b;
       plot (species yhat) * TIME / overlay
         haxis=axis1
         vaxis=axis2
         ;
   run;

  proc gplot data=work.b;
      plot yresid * t;
   run;
```

4.) Unequal slopes model

```
proc sort data=time.models2;
 by model;
run;

proc mixed data=time.models2;
  class model;
  model expected=actual model actual*model / s;
  lsmeans model / pdiff;
  estimate 'new-single' actual*model     0 1 -1;
  estimate 'new-clench' actual*model     -1 1 0;
  estimate 'clench-single' actual*model  1 0 -1;
run;
```

5.) 3D graph

```
  proc g3d data=TIME.PREDICT ;
     plot  HRS * n = prop /
```

```
        cbottom=BLACK
        ctop=BLACK
        ctext=BLACK
        xytype=1
        grid

        rotate=305
        tilt = 65

        caxis=BLACK
        zticknum=6
        xticknum=5
        yticknum=6

          ;
run;
```

APPENDIX 5
SAS SYNTAX PROGRAMS USED IN CHAPTER 5

1.) Density of stems

```
%inc 'glmm612.sas';
%glimmix(data=bluebell.denplot4, procopt=method=reml ratio covtest,
   stmts=%str(
      class plot year time;
      model number = year time / ddfm=satterth;
      random plot plot*year plot*time;
      lsmeans year time / pdiff adjust=tukey;
   ),
   error=poisson,
   link=log,
   output=reschi=_reschi_
)
run;

data work.out;
merge work._outfile work._pred;

   symbol1 c=DEFAULT
     i=NONE
     v=STAR
     ;

proc gplot data=work.out ;
    plot  _reschi_ * _pred_ /
      haxis=axis1
       vaxis=axis2
      ;
  run;

data work.errbar;
  set work._lsm;
  xsys='2'; ysys='2'; color='red';
  length function $8;
  midpoint=year;
  t = 1;
  y = _lsmean_ + t * _se_;
  function = 'move'; output;
  y = _lsmean_ - t * _se_;
  function = 'draw';  output;
proc gchart data=WORK._LSM;
    vbar YEAR /
       maxis=axis1
       raxis=axis2
       type=MEAN
       sumvar=_LSMEAN_
       anno=errbar
       caxis=blue
       ctext=blue
      ;
   pattern v=empty c=black
;
```

```
   symbol1 c=DEFAULT
      i=SPLINE
      v=STAR
     ;

 proc gplot data=WORK._LSM ;
      plot _LSMEAN_ * TIME /
       haxis=axis1
       vaxis=axis2
       caxis=blue
       ctext=blue
     ;
   run;
```

2.) Height of stems

```
proc mixed data=bluebell.complet5 ratio covtest;
  class plot year time;
  model height = year time / ddfm=satterth p;
  random plot plot*year plot*time;
  lsmeans year time / pdiff adjust=tukey;
  make 'predicted' out = pred noprint;
  make 'lsmeans' out = lsmeans;
  make 'solutionr' out = random noprint;
run;


proc univariate plot normal data = pred;
var _resid_;
run;


%inc 'nqplot.sas';
%nqplot (data=pred, var=_RESID_, mu=MEAN, sigma=STD);


proc capability data=pred noprint graphics;
  var _resid_;
histogram _resid_/
  normal
  midpoints =-2 to 2 by 0.25
  haxis=axis1 vaxis=axis2
  noframe legend=legend1
  caxis=blue
  ctext=blue
  cbarline=black
  name='hist_normal';
axis1 label=(h=1.5)  value=(h=1.3);
axis2 label=(h=1.5)  value=(h=1.3);
legend1 label=(h=1.5) value=(h=1.3);
run;


  symbol1 c=DEFAULT
     i=NONE
     v=STAR
```

```
      ;

 proc gplot data=work.pred ;
     plot  _resid_ * _pred_ /
       haxis=axis1
        vaxis=axis2
       ;
   run;

 proc gchart data=WORK.lsmeans;
     vbar YEAR /
        maxis=axis1
        raxis=axis2
        patternid=midpoint
        type=MEAN
        sumvar=_LSMEAN_
       ;
 run;

   symbol1 c=DEFAULT
      i=SPLINE
      v=STAR
     ;

 proc gplot data=WORK.lsmeans ;
      plot _LSMEAN_ * TIME /
       haxis=axis1
       vaxis=axis2
     ;
   run;
```

3.) Total number of capsules per stem

```
%inc 'glmm612.sas';
%glimmix(data=bluebell.complet5, procopt=method=reml ratio covtest,
   stmts=%str(
  class plot year time;
  model total = year time / ddfm=satterth;
  random plot plot*year plot*time;
  lsmeans year time / pdiff adjust=tukey;
   ),
   error=poisson,
   link=log,
   output=reschi=_reschi_
)
run;

data work.out;
merge work._outfile work._pred;

proc univariate plot normal data=out;
var _reschi_;
run;

  symbol1 c=DEFAULT
     i=NONE
     v=STAR
```

```
      ;

 proc gplot data=work.out ;
     plot  _reschi_ * _pred_ /
       haxis=axis1
        vaxis=axis2
       ;
   run;

 proc gchart data=WORK._LSM;
     vbar YEAR /
        maxis=axis1
        raxis=axis2
        patternid=midpoint
        type=MEAN
        sumvar=_LSMEAN_
       ;
 run;

   symbol1 c=DEFAULT
      i=SPLINE
      v=STAR
     ;

 proc gplot data=WORK._LSM ;
      plot _LSMEAN_ * TIME /
       haxis=axis1
       vaxis=axis2
     ;
   run;
```

4.) Proportion of fertile capsules

```
proc mixed data=bluebell.complet5 ratio covtest;
  class plot year time;
  model by = year time / ddfm=satterth p ;
  random plot plot*year plot*time;
  weight s_total;
  lsmeans year time / pdiff adjust=tukey;
  make 'predicted' out = p noprint;
  make 'predmeans' out = pm noprint;
  make 'lsmeans' out = lsmeans;
  make 'solutionr' out = random noprint;
run;


proc univariate data=p plot normal;
var _resid_;
run;


%inc 'nqplot.sas';
%nqplot (data=p, var=_RESID_, mu=MEAN, sigma=STD);


proc capability data=p noprint graphics;
  var _resid_;
histogram _resid_/
  normal
```

```
   midpoints =-2 to 2 by 0.5
   haxis=axis1 vaxis=axis2
   noframe legend=legend1
   caxis=blue
   ctext=blue
   cbarline=black
   name='hist_normal';
 axis1 label=(h=1.5)  value=(h=1.3);
 axis2 label=(h=1.5)  value=(h=1.3);
 legend1 label=(h=1.5) value=(h=1.3);
 run;


   symbol1 c=DEFAULT
      i=NONE
      v=STAR
      ;

 proc gplot data=work.p ;
     plot  _resid_ * _pred_ /
       haxis=axis1
        vaxis=axis2
       ;
   run;

 proc gchart data=WORK.lsmeans;
     vbar YEAR /
        maxis=axis1
        raxis=axis2
        patternid=midpoint
        type=MEAN
        sumvar=_LSMEAN_
       ;
 run;

   symbol1 c=DEFAULT
      i=SPLINE
      v=STAR
     ;

 proc gplot data=WORK.lsmeans ;
      plot _LSMEAN_ * TIME /
       haxis=axis1
       vaxis=axis2
     ;
   run;
```

APPENDIX 6
SAS SYNTAX PROGRAMS USED IN CHAPTER 6

1.) Analysis of CCA axes using a random plot effect

```
proc mixed data=pine.alldata covtest;
      class plot;
      model s_cca1 = stdcan stdwet stdalt stdeast stdnorth
      stdslope stdfire stdwind stdscar / noint ddfm=satterth solution predicted;
      random plot;
      make 'solutionr' out = random noprint;
      make 'predicted' out = pred noprint;
run;


proc univariate plot data = pred normal;
  var _resid_;
run;


%inc 'nqplot.sas';
%nqplot (data=pred, var=_RESID_, mu=MEAN, sigma=STD);

proc capability data=pred noprint graphics;
  var _resid_;
histogram _resid_/
  normal
  midpoints =-2 to 2 by 0.5
  haxis=axis1 vaxis=axis2
  noframe legend=legend1
  caxis=blue
  ctext=blue
  cbarline=black
  name='hist_normal';
axis1 label=(h=1.5)  value=(h=1.3);
axis2 label=(h=1.5)  value=(h=1.3);
legend1 label=(h=1.5) value=(h=1.3);
run;

proc insight data=pred;
   dist _resid_;
   fit s_cca1 = _pred_;
run;


proc mixed data=pine.alldata covtest info ic;
      class plot;
      model s_cca1 = stdcan stdwet stdalt stdeast stdnorth
      stdslope stdfire stdwind stdscar / noint ddfm=satterth solution predicted;
      repeated / subject=plot type=csh;
      make 'solutionr' out = random;
      make 'predicted' out = pred noprint;
run;


proc univariate plot data = pred normal;
```

```
  var _resid_;
run;

%inc 'nqplot.sas';
%nqplot (data=pred, var=_RESID_, mu=MEAN, sigma=STD);

proc capability data=pred noprint graphics;
  var _resid_;
histogram _resid_/
  normal
  midpoints =-2 to 2 by 0.5
  haxis=axis1 vaxis=axis2
  noframe legend=legend1
  caxis=blue
  ctext=blue
  cbarline=black
  name='hist_normal';
axis1 label=(h=1.5)  value=(h=1.3);
axis2 label=(h=1.5)  value=(h=1.3);
legend1 label=(h=1.5) value=(h=1.3);
run;


proc insight data=pred;
   dist _resid_;
   fit s_cca1 = _pred_;
run;
```

2.) Backward stepwise logisitic regression quadratic surface model

```
  proc logistic data=pine.alldata order=data;
    model pine1 = cca1 cca2 cca1sq  cca1x2 cca2sq / selection=backward;
  run;
```

3.) Program used to draw 3D surface response curves

```
data pine.pine1; set pine.logistic;
  a = 1.2646 ; b1 = -1.2918 ; b2 = -6.1131 ; b3 = -2.0224 ;
prob = exp(a+(b1*cca1)+(b2*cca1*cca1)+(b3*cca1*cca2))
       /(1+exp(a+(b1*cca1)+(b2*cca1*cca1)+(b3*cca1*cca2)));
label prob = 'Probability';
run;

 proc g3d data=pine.pine1 ;
     plot cca2 * cca1 = prob /
       ctop=black
       cbottom=grey
       ctext=black
       rotate=35
       zmin=0.00
       zmax=1
       grid
       caxis=BLACK
       xticknum=6
       yticknum=6
       zticknum=6
```

```
         ;
  run;
```

4.) Program used to perform cluster analysis

```
proc cluster data=pine.alldata method=ward outtree=work.tree rsquare nonorm;
  var ccalc1--ccalc8;
  id sample;
run;

proc tree data=work.tree out=work.treeout n=15 graphics height=r  ntick=5
name=adv inc=0.2 tickpos=5;
id sample;
run;

proc sort;
by sample;
run;

data scatter;
merge pine.alldata work.treeout;
run;

proc sort;
  by cluster;
run;

goptions reset=(axis, legend, pattern, symbol, title, footnote) norotate
    hpos=0 vpos=0 htext= ftext= ctext= target= gaccess= gsfmode= ;
    goptions device=WIN  ctext=blue
    graphrc interpol=join;

    symbol1 c=DEFAULT
    i=NONE
    v=DOT
    cv=RED
   ;
   symbol2 c=DEFAULT
   i=NONE
   v=CIRCLE
   cv=BLUE
   ;
  symbol3 c=DEFAULT
    i=NONE
    v=PLUS
   cv=BLUE
    ;
  symbol4 c=DEFAULT
   i=NONE
   v=STAR
   cv=MAGENTA
   ;
  symbol5 c=DEFAULT
   i=NONE
   v=square
  cv=black
   ;
  symbol6 c=DEFAULT
    i=NONE
```

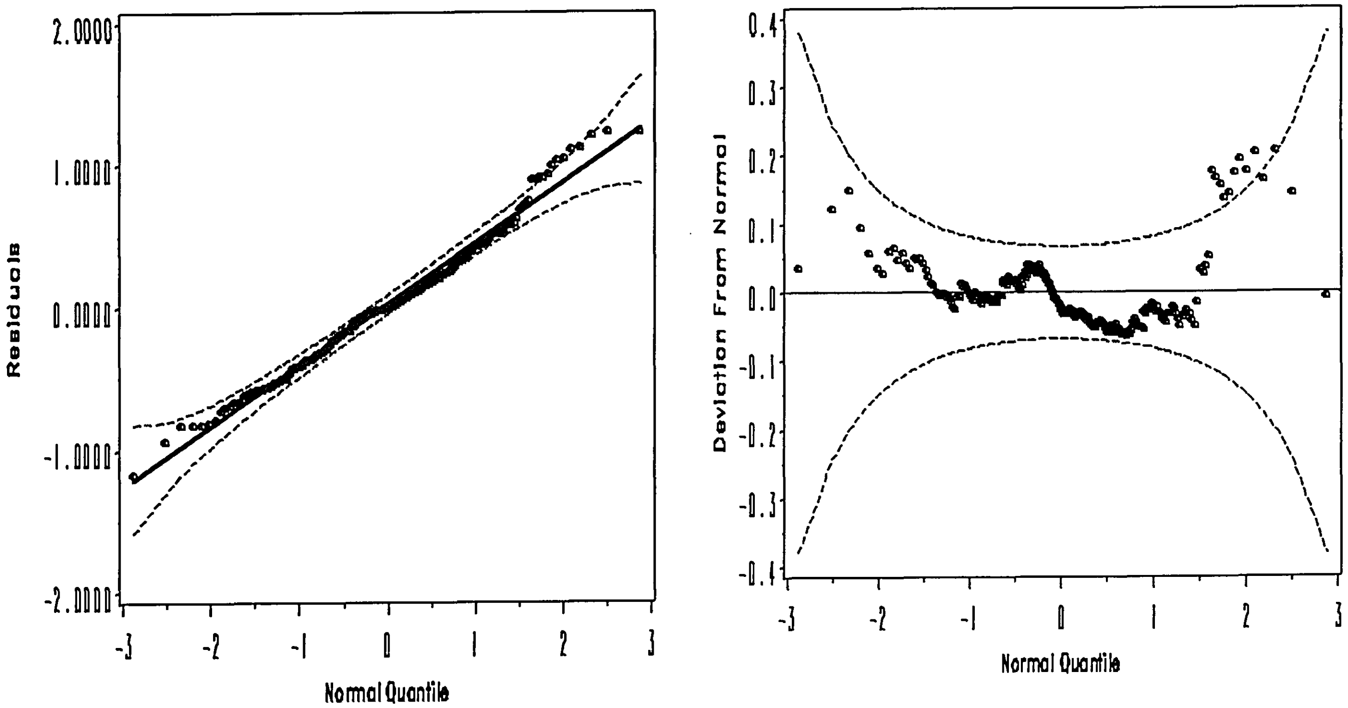
```
    v=DOT
    cv=CYAN
   ;
   symbol7 c=DEFAULT
   i=NONE
   v=CIRCLE
   cv=RED
   ;
  symbol8 c=DEFAULT
    i=NONE
    v=PLUS
   cv=RED
    ;
  symbol9 c=DEFAULT
   i=NONE
   v=DOT
   cv=BLUE
   ;
  symbol10 c=DEFAULT
   i=NONE
   v=square
  cv=BLUE
   ;
  symbol11 c=DEFAULT
   i=NONE
   v=square
  cv=black
   ;
  symbol12 c=DEFAULT
    i=NONE
    v=DOT
    cv=cyan
   ;
   symbol13 c=DEFAULT
   i=NONE
   v=CIRCLE
   cv=green
   ;
  symbol14 c=DEFAULT
    i=NONE
    v=PLUS
   cv=cyan
    ;
 symbol15 c=DEFAULT
  i=NONE
  v=STAR
  cv=BLACK
  ;


axis1
length=60 PCT
color=blue
width=2.0
style=1
;
axis2
 color=blue
 width=2.0
 ;
```

```
   axis3
   color=blue
   width=2.0
   ;

   proc gplot data=WORK.SCATTER ;
      plot ccalc2 * ccalc1  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;


    proc gplot data=WORK.SCATTER ;
      plot ccalc4 * ccalc3  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;

    proc gplot data=WORK.SCATTER ;
      plot ccalc6 * ccalc5  = CLUSTER /
         haxis=axis1
         vaxis=axis2
         ;
   run;

proc gplot data=WORK.SCATTER ;
     plot ccalc8 * ccalc7  = CLUSTER /
        haxis=axis1
        vaxis=axis2
        ;
  run;
```

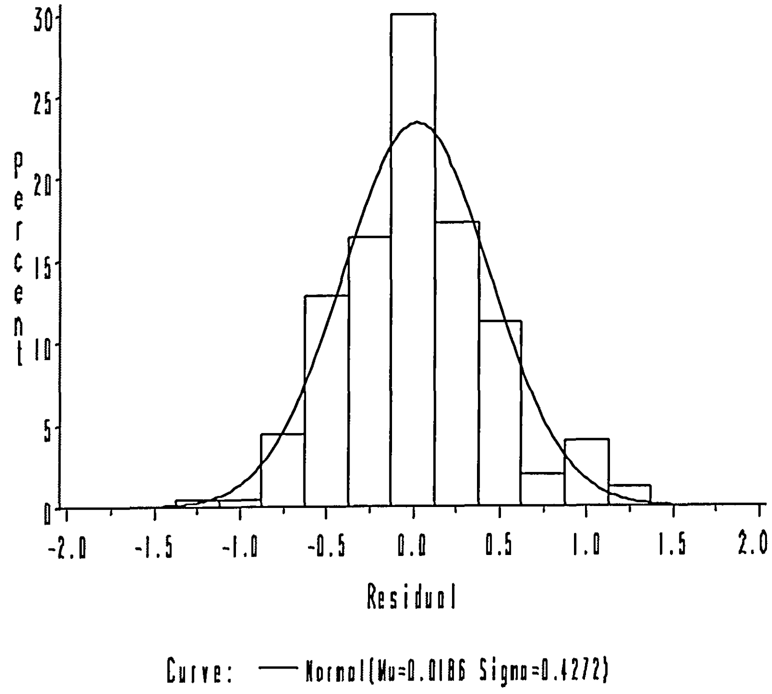
APPENDIX 7
EXAMPLES OF DIAGNOSTIC PLOTS

Output from model 1, Appendix 3A Spatial exponential GLM

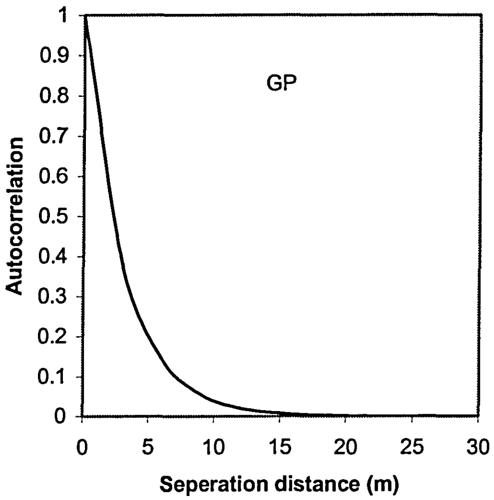
Normal Q-Q Plot with Standard Errors (NQPLOT macro)

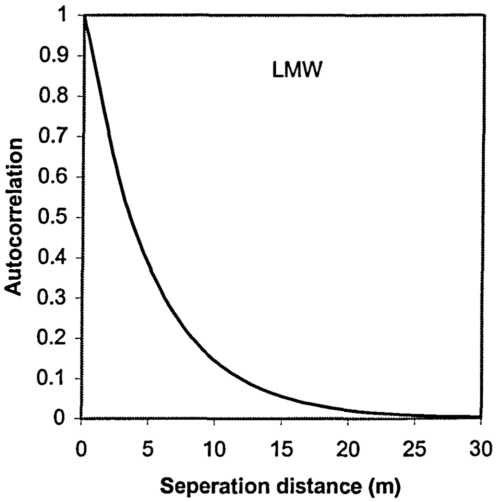


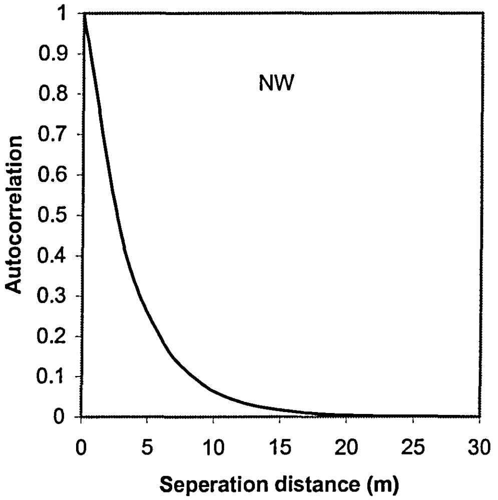
Histogram of residuals with fitted normal distribution (CAPABILITY procedure).

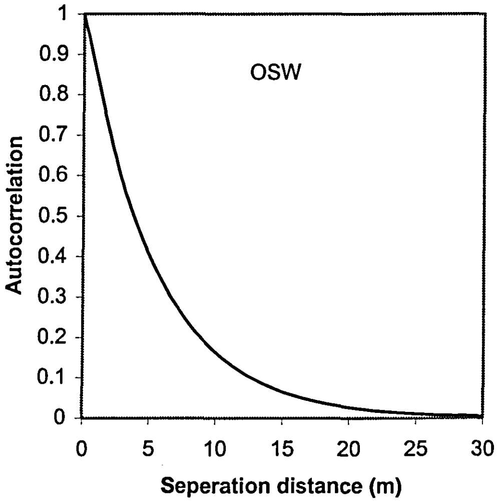


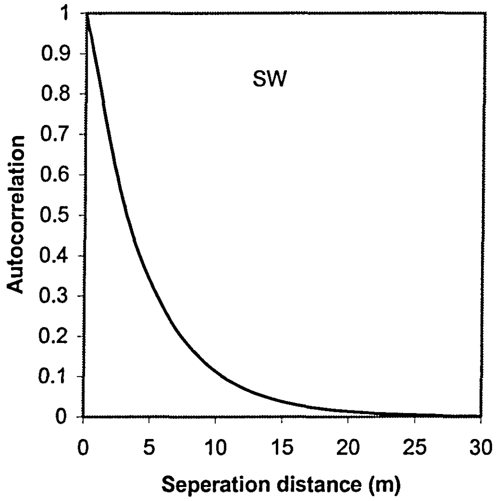
Auto-regressive error structures





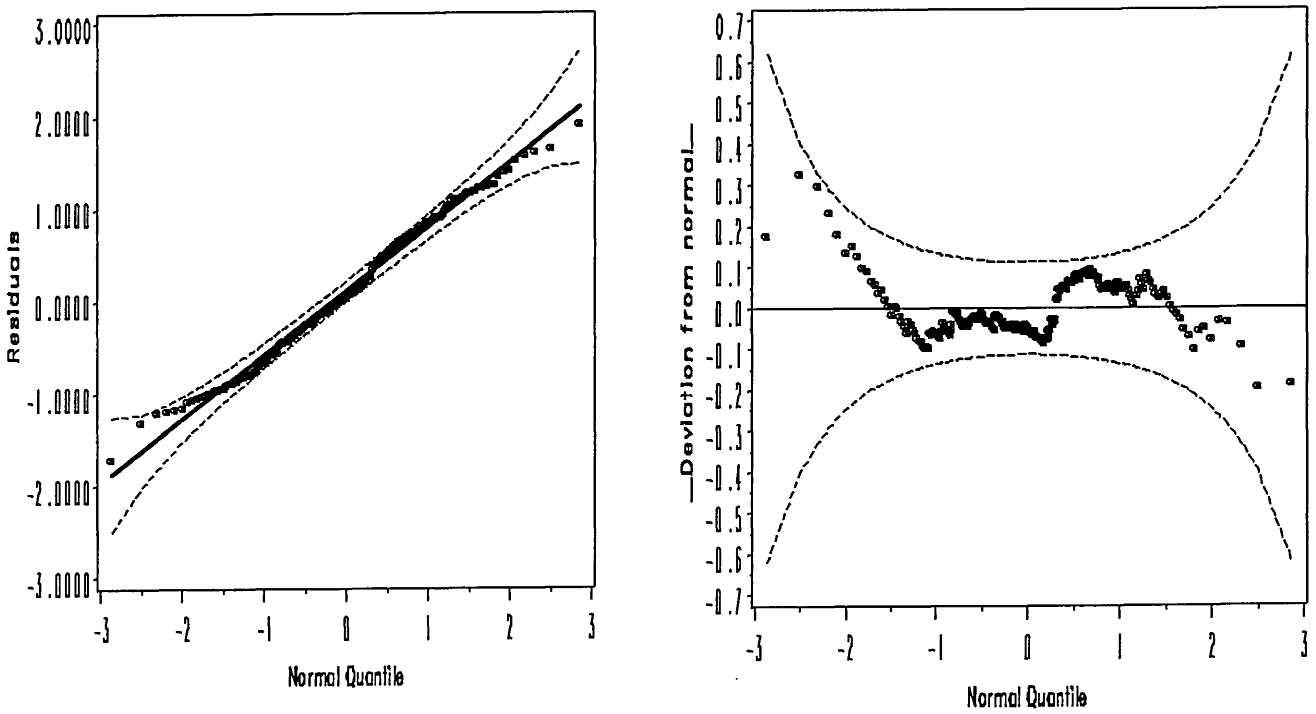




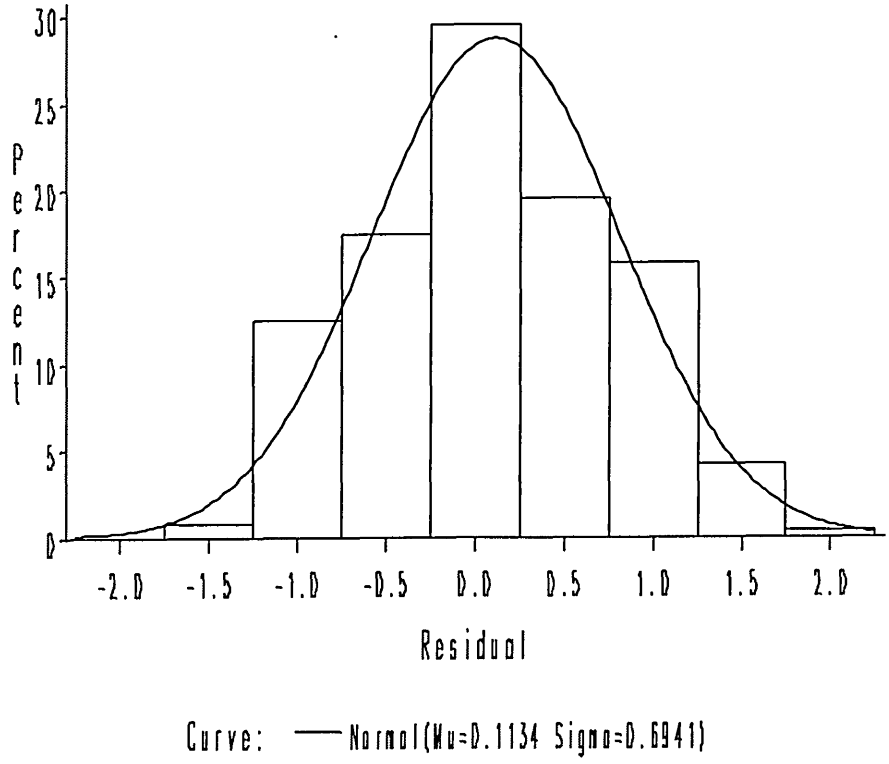


Output from model 1, Appendix 3A Spatial exponential GLM

Normal Q-Q Plot with Standard Errors

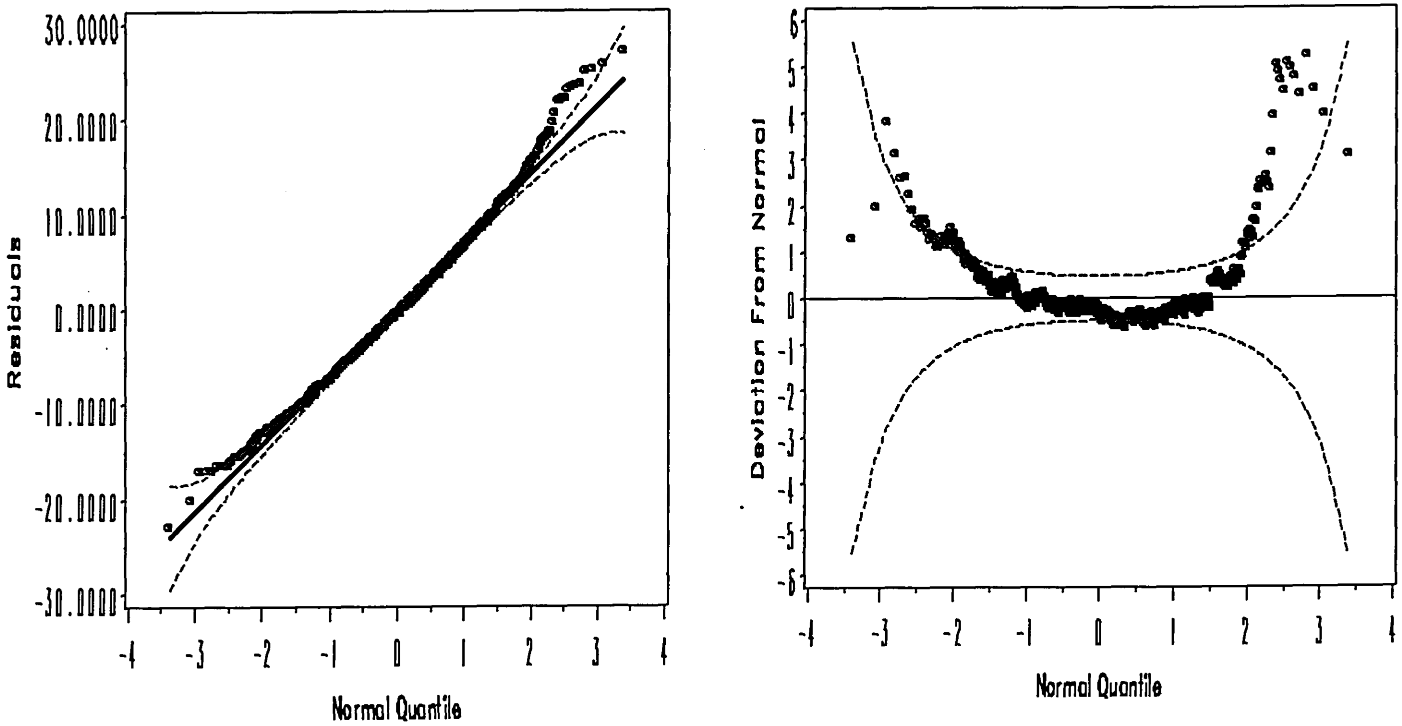


Histogram of residuals with fitted normal distribution (CAPABILITY procedure).

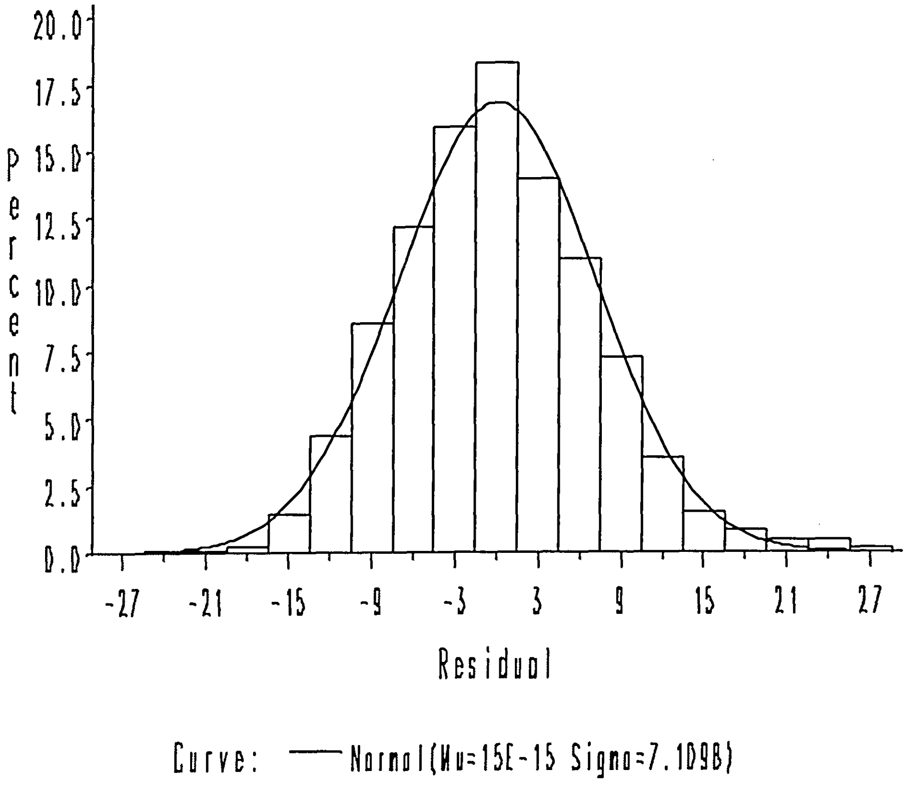


Output from model 2, Appendix 5; Height of Bluebell stems

Normal Q-Q Plot with Standard Errors

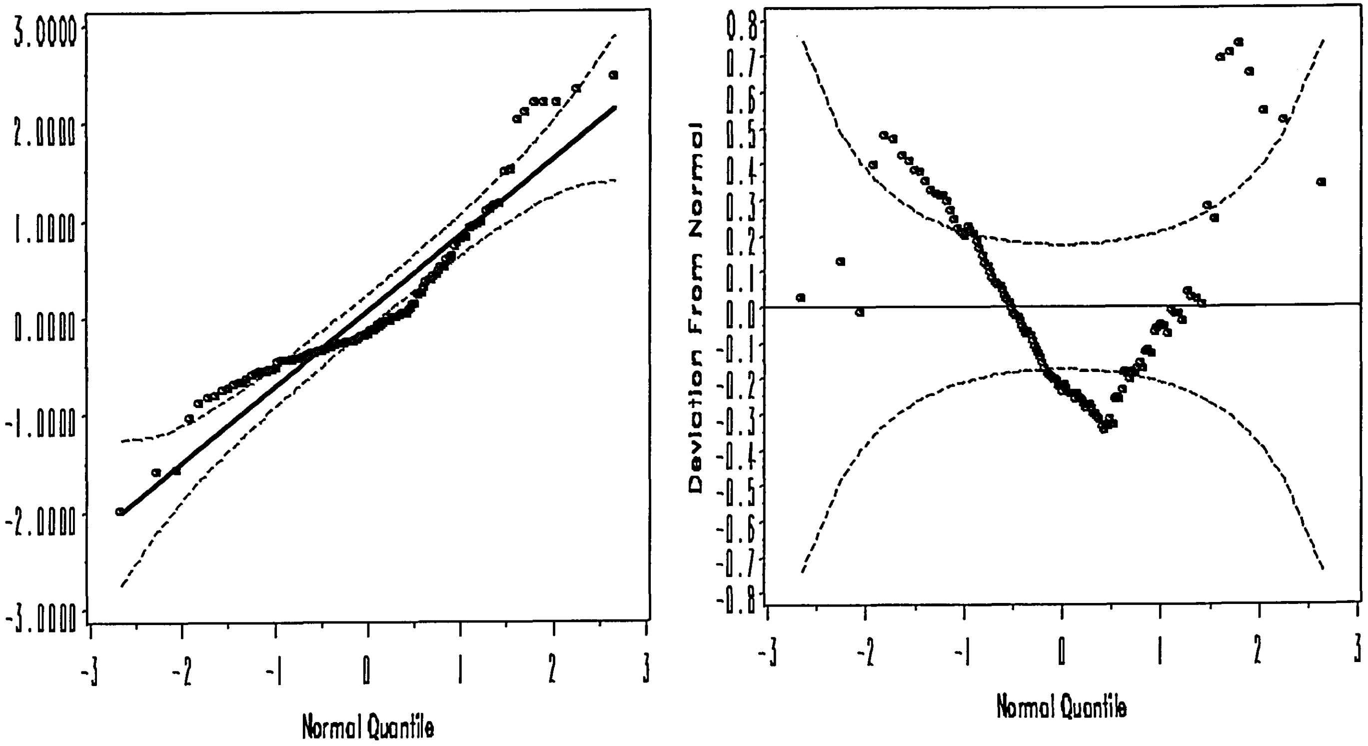


Histogram of residuals with fitted normal distribution (CAPABILITY procedure).



Output from model 1, Appendix 6; CCA1 against standardised variables for pine vegetation data

Normal Q-Q Plot with Standard Errors



Histogram of residuals with fitted normal distribution (CAPABILITY procedure).