



Diptera diversity in Lancashire and Cheshire - a quantitative survey and analysis

by

Phil Brighton

32, Wadeson Way, Croft, Warrington WA3 7JS

helophilus@hotmail.co.uk

Version 1.2

March 2020



CONTENTS

Abstract	3
1. Introduction.....	3
2. Recording method	4
3. Geographical scope	5
4. Taxonomic and temporal scope	6
5. Alpha and gamma diversity	7
6. Inter-site diversity	10
7. Comparisons of sites with overall dataset	10
8. PANTHEON results.....	11
9. Discussion	13
REFERENCES	15
Appendix: Derivation of the Fisher curve from the log series	16
TABLES.....	18
FIGURES.....	26

Diptera diversity in Lancashire and Cheshire - a quantitative survey and analysis

Abstract

Diptera surveys have been carried out over a wide range of sites in Cheshire (VC58) and Lancashire (VCs 59 and 60). Over the period 2012-2019 just over 20,000 records representing the occurrence of a species in a specific 100m square on a specific day were collected using a sweep-net and aspirator, together with field observations and individual capture of some species. The taxonomic scope was 3,300 species from families of larger and more readily identifiable species, of which 1,292 were recorded. Six sites with mixed open, wooded and wetland sites and areas between 19 and 972ha were selected for intensive survey consisting of at least 6 visits spread over the period April-September.

Rank-abundance data have been analysed to produce estimates of biodiversity measures both for the full dataset and for the six sites. They approximate to the log-series distribution. Non-parametric methods have been used to estimate Hill numbers of order 0, 1 and 2. It is concluded that for an adequate sample from a site around 1,200 records need to be gathered leading to around 300-400 species. Pair-wise inter-site comparisons using the Sørensen and Morisita-Horn indices show that the strongest diversity difference is between the single coastal site and the other 5 inland sites. The numbers of the 96 species constituting 50% of the overall data over the six sites is assessed against the hypothesis of a uniform distribution. The diversity differences correlate qualitatively with the habit assemblages determined from the site species lists using the PANTHEON database.

1. INTRODUCTION

In recent decades, interest in biodiversity has inspired the establishment of local and national schemes for the collection, storage and dissemination of natural history records. This has been greatly aided by the development of personal computers, enabling larger amounts of data to be easily handled and analysed than was possible with the traditional tools of pens, notebooks and record cards. The Internet has further enabled data from individual recorders to be pooled and checked centrally and has also made available discussion fora and other resources to assist with identification. Recording of as wide a range of the biodiversity as possible is also being encouraged by national conservation organisations to support assessment and monitoring of wildlife reserves.

At its simplest level biodiversity can be interpreted as the number of species present in an area or location. This means that for the terrestrial fauna biodiversity will be dominated by the insects at the macroscopic level. Amongst these, there are over 2000 British species of lepidoptera and over 4000 species of coleoptera, which are probably the most well-recorded in the UK. However the biggest single British order is the diptera, the true or two-winged flies, for which the British list now stands at over 7000 species (Chandler, 2020). Another substantial group is the hemiptera or true bugs with over 1600 species.

This report presents a summary and analysis of the data from surveys of diptera over a large range of sites in Cheshire and Lancashire (vice-counties 58, 59 and 60): these range from a small wildlife-friendly garden, through various local wildlife sites, to major expanses such as the Birkdale Dunes and the Delamere Forest.

By the end of 2019, over 20,000 diptera records had been accumulated. These have been analysed using quantitative biodiversity measurement methods described by Magurran (2004) and Gotelli and Chao (2012). Six of the larger sites were selected for individual analysis and comparison to assess the adequacy of the extent of sampling to reveal ecological differences. The diptera species lists from these six sites have also been analysed using the PANTHEON database of ecological associations and assemblages, as a further check on the consistency of the survey results.

2. RECORDING METHOD

The main collecting technique was sweeping with a net with a rigid pentagonal frame of width 40cm and a handle of length 40 cm. Specimens were retrieved from the net using an aspirator (or “pooter”) consisting of a rubber bung with two plastic tubes inserted into interchangeable collecting tubes of 50ml volume and 30mm diameter. Large individual specimens were also caught from flowers or flat surfaces such as fences directly into individual pots of 30mm diameter and 50mm length approximately.



On a full day’s survey, 12 tubes would be used, each numbered and recorded for a single six-figure grid reference. More than one tube might be used in particularly productive locations. The aspirator was used to select the larger diptera from the catch down to a minimum size of 2-3mm, while avoiding obvious duplicates. The number of specimens per tube was limited to avoid damage, and would vary according to their size. Where fly specimens were well settled on flowers, leaves or fence posts, screw-top pots (30mm diameter and 50mm length) were used to capture them individually – not always successfully. The pots were also needed to secure large craneflies and calypterates captured in the net. A few species could also be recorded on sight in the field without capture.

Captured specimens were killed by storage in a freezer for a few hours at least or for several days. Following defrosting the specimens were pinned and staged for identification under a low-power binocular zoom microscope within a week or two of the day of collection.

All records have been compiled into a standard Excel spreadsheet format, with the following fields

- Recording Scheme (see below)
- Species
- Date
- Site – typically location name as given on the Ordnance Survey maps, as described below
- Location details – indication of more specific locations or particular habitats within the site
- Six-figure grid reference
- Vice-county number
- Observer –collector or photographer of specimen (some records from other observers are included in the data)
- Determiner
- Stage
- Sex for adults
- Number of specimens caught
- Comments

Up to the end of 2016, spreadsheets were submitted both to recording scheme organisers and to the Cheshire local records centre (www.record-lrc.co.uk/). Since then the spreadsheets have been uploaded to IRECORD (www.brc.ac.uk/irecord/) as the sole record depository. All occurrences of a given species in an individual six-figure grid square on a particular date have been logged, so multiple records for a site can be generated in the course of a survey walk. Data for abundance in a single record generally refer to the number of specimens actually caught and identified, though for larger

species such as hoverflies a quantitative or qualitative count of those present, on flower heads for example, can sometimes be made.

3. GEOGRAPHICAL SCOPE

Figure 1 is a map of all the locations visited up to the end of 2018. Some additional sites were added in 2019 though without making a significant difference to the overall range. The map shows recording locations as 2km tetrads with the colour indicating the number of records on a heat scale from blue as low through green, yellow and orange to red as the highest.

The following six sites were selected for repeated sampling for their interest as containing distinctive habitats under management by conservation organisations. Their locations are indicated by the letters on Figure 1.

Astley Moss SSSI (A)

This Lancashire Wildlife Trust (LWT) reserve represents a significant part of the Manchester Mosses SAC. It consists mainly of cut-over peat bog, until recently dominated by birch and *Molinia*, but now being managed by removal of birch scrub and raising of water levels by provision of plastic and peat bunds. There are also significant areas of remaining woodland, mainly birch, willow and oak, and an area of fen with pools, known as the “carrot field” from its previous agricultural use.

The area is 37 Ha, mainly within the monad SJ6997. Together with the adjacent Bedford Moss, it was designated as an SSSI for its botanical interest as acid mire, one of the largest remaining fragments of the original 25km² extent of Chat Moss, now mainly drained for agriculture. The site forms part of the Manchester Mosses Special Area of Conservation (SAC).

Birkdale Green Beach and Hills (B)

The Ainsdale and Birkdale Hills Local Nature Reserve (LNR) is owned and managed by Sefton Metropolitan Borough Council. It is the northernmost portion of the Sefton Coast dune system adjoining the Ainsdale National Nature Reserve (NNR). The dune system is accreting leading to a seaward belt of low dunes and slacks some of which are only a decade or so old. There is a long belt mainly of alder separating this part from higher more established dunes with increasing scrub and woodland beyond the coastal road which traverses the whole site. The overall area of the LNR is 296 hectares mainly within the hectad SD31. The whole Sefton Coast is designated as an SAC.

Cholmondeley Estate (C)

The privately owned Cholmondeley estate in south-west Cheshire comprises farmland and parkland around the eponymous castle at SJ536513. The parkland covers approximately 240ha and includes a SSSI, Chapel Mere, a eutrophic lake formed in a depression in the glacial drift deposits. The land around this mere is damp pasture, much of it grazed by cattle with small patches of damp woodland. The estate includes the 86ha of Bickley Hall Farm, leased by Cheshire Wildlife Trust who have their headquarters there and practice mixed wildlife-friendly farming. Other points sampled in the surveys were damp wood land and carr around Deer Park Mere and south of Moss Wood near Bickley church.

Delamere Forest (SJ57) (D)

Delamere Forest covers a large area of undulating glacial deposits with sandy mounds and hollows filled with basin mires. At 972ha it is the largest area of woodland in Cheshire. Its medieval designation as a royal hunting forest was revoked in 1812. During the 19th century parts were drained and planted with oak, sweet chestnut and conifers. The bulk of the Forest is now under the stewardship of the Forestry Commission, but Cheshire Wildlife Trust owns an area of mossland and damp woodland together with the lake at Hatchmere (SJ57L) at the north-eastern edge, designated as a SSSI. Work to reclaim the Delamere wetlands from adverse effects of forestry began in the 1990s with the flooding of the large expanse of Blakemere Moss. During the last decade the “Lost Mosses” project has been removing trees and raising water levels in a number of the smaller basins.

Holcroft Moss (H)

Holcroft Moss (SJ6893) is a lowland raised bog, thought to be the only known example in Cheshire that has not been cut for peat. Commercial peat extraction from the bog's immediate surroundings has however lowered the water table, and this subsequently damaged the hydrology of Holcroft Moss. The Cheshire Wildlife Trust began work in restoring the bog's ability to hold water by inserting sheets of plastic piling around the perimeter of the site in about 2000, when birch was also cleared from the central mossland area. Further works to clear areas dominated by bracken and birch and convert them to mossland with water-retaining peat bunds have been carried out in recent years. The 19ha of the reserve retain a wooded fringe, mostly birch, but with damp woodland dominated by willow on clay deposited during excavations for the M62 which runs along the northern edge.

The Smithills Estate (S)

The Smithills Estate (SD61) became the largest land-holding of the Woodland Trust when it was bought from Bolton Council in 2015. The area of 686 hectares includes large expanses of grazed grassland, deep wooded cloughs traversed by fast-flowing stony streams, heathland and a large area of moorland on Winter Hill, scene of extensive fires in 2018. The latter falls within the area of the West Pennines SSSI. The estate now forms a nucleus of the Northern Forest, with extensive planting of a range of broad-leaved native trees.

Detailed species list and habitat analyses for the above sites are contained in separate individual reports (Brighton 2017, 2018, 2020a, b, c).

4. TAXONOMIC AND TEMPORAL SCOPE

The initial focus of the surveys was on taxa in the main diptera recording schemes, particularly the craneflies and hoverflies, for which extensive identification resources are available. However, it was soon possible to extend this scope by acquisition of available RES handbooks, the 2nd edition of Stubbs and Drake's soldierflies book and Collin's 1961 monograph on the Empididae. The Dipterists Forum workshops in recent years have provided new identification keys for several of the Calyptrate families, so that these can now be comprehensively recorded with relative ease. Information on other families has been gathered piecemeal from contacts in the Dipterists Forum, and internet sources such as the Biological Heritage Library.

Table 1 shows a list of the families covered. For convenience of data handling, they have been divided into five groups. Four of these follow standard taxonomic divisions, though the cranefly recording scheme covered the Trichoceridae (winter gnats) and Ptychopteridae in addition to the superfamily Tipuloidea of craneflies proper. The fifth group of other diptera is very varied, and Table 1 separates out subgroups covered by recording schemes.

Table 1 also shows the number of species in each group or subgroup according to the British checklist (Chandler, 2020). There is a grand total of 3300 species within the scope of the surveys, nearly half the current British total of 7193.

Inclusion of families in Table 1 does not necessarily imply that species have been recorded from that family, only that they should be identifiable if encountered. The major families excluded from the surveys are the following:

- Mycetophilidae (Fungus gnats): 493 species
- Sciaridae (Black fungus gnats): 267 species
- Cecidomyiidae (Gall midges): 693 species
- Psychodidae (Owlet-midges or mothflies): 100 species
- Chironomidae (non-biting midges): 625 species
- Phoridae (Scuttle-flies): 356 species
- Agromyzidae (Leaf-miner flies): 400 species

- Chloropidae: 177 species
- Ephydriidae: 151 species

These add up to 3394 species. A few distinctive common species have been recorded from some of these families.

A total of 20249 records have been collected covering 1292 species or 39% of the British species within the taxonomic scope (see Table 2). Figure 2 shows the numbers of records for each group for each year. There was a progressive expansion in numbers of records up to 2015, partly as a result of increased availability of effort and partly because familiarity was gained with a wider range of families. There is no particular reason why the numbers of records peaked in 2016 and has declined steadily thereafter. Certainly no inference should be drawn about overall diptera abundance from year to year. Variations in relative numbers of the different groups from 2016 may be the result of a different range of sites being sampled each year. However, more attention on calyptrates and a greater range of other diptera may have detracted from the recording of the other three groups.

The data have been analysed by month of the year in Figure 3. The cumulative chart shows that during the peak recording months the numbers of records is reasonably constant, being limited mainly by recorder effort. In the spring months of March and April, and early autumn in September and October, the numbers may reflect fewer days with favourable collecting weather but also be limited by the numbers of insects in flight. The individual lines for the different groups show markedly different variations through the seasons with craneflies peaking strongly in May, while hoverflies increase rather slowly over the spring and summer to reach a peak in August. The Empidoidea increase rapidly from April and June and then remain level over the summer months. The calyptrates are already building in numbers in March to reach strong peaks both in May and August, with a significant gap between: there is a similar variation for the heterogeneous group of other diptera.

The data from the six selected sites have been aggregated in a similar fashion to show the taxonomic and temporal coverage in each case. Figure 4 shows the proportions of each group recorded at the six main sites. They fall into two groups: Astley Moss, Delamere and Holcroft Moss which were visited between 2012 and 2017; while the bulk of the records from Birkdale, Cholmondeley and Smithills are from 2017 to 2019. Thus the relative proportions reflect the year-to-year variations seen in Figure 2.

The monthly charts in Figure 5 show that in no case was an even distribution of records achieved across the main collecting months of May-September. However, efforts were made to ensure that return visits to a particular site were several weeks away from the anniversary of visits in previous years, so the overall temporal coverage is not as erratic as the Figure appears to show. The specific dates are given in the individual site reports.

5. ALPHA AND GAMMA DIVERSITY

It is striking feature of the complete dataset that while the most frequent species *Scathophaga stercoraria* was recorded 458 times, there were 284 species, 22% of the list, with a single record apiece. The latter constituted just 1.4% of the number of records while 50% was provided by the most frequent 96 species, accounting for only 7.4% of the full species list. This very uneven distribution is visualised in Figure 6 as a standard Whittaker plot in which the number of records n per species is plotted against species position r in an overall ranking in descending order of n . This is the Whittaker plot as discussed by Magurran (2004, Chapter 2), in which the vertical axis is logarithmic because of the large variation in the values of n .

This very uneven distribution of species abundances is very widespread in insect surveys. Fisher *et al* (1943) developed a statistical theory to predict this variation from knowledge of the total number of individuals N and the number of species S in a samples obtained from an environment with a large number of species. The results are normally expressed in terms of the “log series” which gives the numbers of singletons, doubletons, etc in the sample. May (1975) gives details of the derivation of a

continuous distribution curve, which I will call the Fisher curve (see Appendix). Figure 6 shows that this curve follows the data closely (on a logarithmic scale at least) even for the most frequent species. The predicted numbers of singletons and doubletons for the full dataset are 303 and 149 compared with actual numbers of 284 and 170.

Fisher's analysis was just the first of many and varied methods for what has become known as biodiversity measurement. The parameter (α) alpha which is calculated from N and S in this theory has given its name to embrace the overall species richness and the evenness of the abundance distribution within a single habitat in the concept of alpha diversity (Magurran, 2004, Ch. 6). Beta (β) diversity covers variation between habitats. The combination of alpha and beta diversity within a landscape has been termed gamma (γ) diversity. Thus the distribution shown in Figure 6 is a visualisation of the latter concept for diptera across Lancashire and Cheshire.

Our six main sites are somewhat wider than individual sites, but for the purposes of this report they are treated as single units of alpha diversity. Inter-comparisons between these sites and between each site and the overall dataset are presented in the next section under the heading of beta diversity.

Magurran (2004) discusses the problems which arise in making inferences from biodiversity samples and surveys which capture only an unknown proportion of the overall diversity present, as exemplified by the large numbers of singletons and doubletons. In a consensus which emerged since publication of that book, the empirical distribution has been used as the basis for biodiversity measures and for interpolating and extrapolating these measures to compare different samples on a consistent basis. In these analyses f_1 the number of singletons and f_2 the number of doubletons play a key role as inputs, rather than being predicted as in the approach of Fisher *et al* (1943). These non-parametric statistical techniques have been developed and applied in an extensive literature by Anne Chao and co-workers, available from the web-page <http://chao.stat.nthu.edu.tw/wordpress/publication-list/>.

The data samples from the six sites show very similar patterns in the distribution of numbers of records for different species. Figure 7 shows the results for Astley Moss and although the lower quantity of data results in a more irregular stepped curve, the overall trend is again clearly captured by the theoretical Fisher curve. The results for the other five main sites are very similar, so not illustrated here. The key parameters are listed in Table 3.

Table 3 also includes two quantities derived from the Chao suite of methods. Chao (1984) provided a statistical analysis deriving a lower bound statistical estimator for the eventual total number of species that could be found by continuing sampling indefinitely from given population. This is a simple formula based just on f_1 , f_2 and N . The result is listed under the heading "Chao1" in Table 3. This asymptotic number of species ranks the species-richness of the sites in the order $S>D >A>H>C>B$ with the maximum 634 being 40% greater than the minimum 450.

Of course its application in the current situation must be hedged round with many qualifications. It can be imagined as the result if the numbers of records gathered had been increased *ad infinitum* for the same pattern of geographical, taxonomic and temporal coverage described above. While some of these could be replicated in future surveys, a complete coverage must remain unattainable in view of the year-to-year variability of insect populations, the development of the vegetation in response to agricultural or conservation management, or lack thereof, and other factors such as climate or environmental pollution. These issues are covered further in the light of the overall results of the study in the discussion section below. It is the hypothesis of this report that a meaningful comparison can at least be made for the sites as they existed in the conditions prevailing over the period that data was collected in each case.

Since different numbers of records were gathered at the six sites, the actual number of species recorded does not provide a valid comparison of the sites. Fisher's α can be interpreted as an intrinsic measure of alpha diversity in units of numbers of species. It ranks the six sites in the order $S>D>C>B>A>H$, with the highest being 44% greater than the lowest.

The non-parametric methodology gives two methods of making a fair comparison. The first is to standardise the samples to a fixed number of records by means of species accumulation curves derived from the data by interpolation, also known as rarefaction, and extrapolation (Chao *et al* 2014). The empirical species accumulation curve is an alternative to the Whittaker curve in which the number of species recorded is plotted against the cumulative number of records in the order they were gathered (Magurran, 2004, Chapter 1). In rarefaction or interpolation, a smooth curve is derived as an average over all possible permutations in the order of the individual records. This can be done in Excel by using the formulae for the zero-order Hill number in Table 1 of Chao *et al* (2014)*. The results for the six sites are plotted in Figure 8, in which the actual data points (N, S) are marked with a spot. It can be seen that these curves display the same ranking as found from the values of Fisher's α , except towards the right, where the slopes curves B and C start decreasing more rapidly and cross below their neighbours. Extrapolating this trend produces the different ranking seen in the Chao1 figures in Table 3.

Magurran (2004, Chapter 4) reviewed the large number of other indices which had been developed to express different aspects of the species abundance distribution. More recently, the work by Chao *et al* (2014) has focussed attention on the Hill numbers qD which rationalise a whole set of these indices within a single framework defined by a single parameter, the order q . The order is a power index applied to the individual proportional frequencies of all the species before averaging them in such a way as to equate to a number of species. As already mentioned 0D equates to the actual number of observed species in the sample, but can be interpolated or extrapolated to different sample sizes. For comparing the six main sites, 1200 has been selected as a standard number of records.

The Hill numbers for orders 0, 1 and 2 have been evaluated for this size sample using the on-line INEXT facility provided by Anne Chao's group at <https://chao.shinyapps.io/iNEXTOnline/> (Hsieh, Ma & Chao, 2016). Figure 9 shows the results including error bars generated by 200 bootstraps. The Hill number of order 1 is equivalent to the Shannon index and that of order 2 to the Simpson index. As q increases the Hill Number becomes increasing weighted towards the most frequent species. If all species were recorded in equal numbers the Hill numbers would all reduce to the same value. For $q>0$, the numbers decrease progressively, each representing an equivalent number of species for an even distribution.

In Figure 9, the least and most diverse sites remain Holcroft and Smithills at each order with their 95% confidence intervals non-overlapping throughout. The Smithills results exceed the Holcroft ones by 25%, 38% and 45% at the three orders, though the order 2 results also have a greater variance. The ratio of Hill numbers 0 and 2 as a relative measure of evenness ranges from 2.4 at Smithills to 3.0 at Astley as the most uneven distribution.

Chao & Jost (2012) recommended standardising diversity measures at different sites by coverage rather than by absolute numbers in the samples. The coverage is the proportion of the total population of organisms represented by the actually recorded species. In our case "population" is to be understood in terms of numbers of records than individual flies, of course. Like Chao1, a statistical estimate of each sample is provided by a formula involving just N, f_1 and f_2 . The values for each site sample and for the overall dataset are given in Table 3. As with Chao1, these figures can refer only to the potential coverage which could be achieved by the particular collecting technique at the locations and times sampled if there had been no limit on the numbers of specimens caught and identified. The proportion represented by undetected species varies approximately as $1/N$ giving a high figure for the overall dataset.

* In some cases the Excel COMBIN function did not produce a result because of the size of the arguments. In these cases Stirling's formula for the factorial function (Abramowitz & Stegun, 1972, §6.1.37) was used to produce a tractable result.

Like the Hill numbers the coverage can be interpolated or extrapolated to different sample sizes (Chao *et al*, 2014, Table 1). When comparing different samples, this is used to find the sample sizes needed to give equal coverage. Figure 10 shows the curves calculated for each site. The curves all flatten out much more rapidly than the species accumulation curves in Figure 9, so that coverage of 90% or even 95% should be achievable at most sites without substantially more effort than in the present surveys. Given the range of achieved coverage to be seen in Table 3, 85% has been selected as a suitable value for evaluating the set of Hill numbers, with the results shown in Figure 11.

Chao and Jost demonstrated that standardisation of samples on coverage method removes a degree of bias inherent in comparing communities of different richness at the sample size. In the present case, comparison with Figure 9 shows that the ranking of sites is preserved but the range is accentuated. The Hill numbers for Smithills exceed those at Holcroft by 51-52% for each of the three orders.

6. INTER-SITE DIVERSITY

While the alpha diversity analysis reveals significant differences between the six sites in terms of the number of species and the evenness of the abundance distribution, it says nothing about the specific species involved. The least diverse site could contain just a subset of the species at the most diverse, or there might be no species in common. As with alpha diversity, Magurran (2004, Chapter 6) reviews a wide range of methods of measuring the similarities and differences between two samples from different sites or habitats. Two of these, the Sørensen and the Morisita-Horn (M-H) indices have been selected to make pairwise comparisons between the six sites.

The Sørensen index involves just the two lists of species, and is formulated to have the value 0 if there are no species in common and unity if the species lists are identical. The M-H index also has this characteristic, but involves differences in the squares of the relative frequencies. Chao *et al* (2008) showed that these two indices could be regarded as examples of a general index involving powers of the relative frequencies, analogous to the Hill numbers. The Sørensen index corresponds to species richness with order 0, and the M-H index has order 2 making it the counterpart of the Simpson index. It appears that no method has yet been developed for standardising sample sizes before applying these indices or for estimating the statistical variance in the index values (Chao & Chiu 2016). While the order 0 method is rather sensitive to sampling size, the use of the order 2 index reduces this as it is weighted towards the commoner species.

There are 15 different pairs of 6 sites, but the calculations are relatively simple to set up in a spreadsheet. Figure 12 includes all the results in a diagrammatic form. There is a good deal of consistency between the two indices, but the range of the M-H index is greater so that it gives a more definite indication of differences and resemblances between each pair of communities. This fits in with the recommendation by Magurran of the M-H index as among the most satisfactory.

The results appear to make ecological sense in terms of the overall character of the sites. The coastal Birkdale site has the least in common with any of the others. The neighbouring mossland sites have the closest affinity. Delamere has fairly strong connections with the two mossland sites, perhaps reflecting the acid mire element of the habitat, but also a strong link to Smithills. This may reflect the important woodland element at these two sites, but Smithills also tends to give higher results because it has the highest alpha diversity and therefore chance of species in common with other sites.

7. COMPARISONS OF SITES WITH OVERALL DATASET

As well as comparing sites with one another, one may ask how far the individual site abundance distributions reflect the overall regional data. Table 4 shows various indicators of the divergence of the sites from the overall distribution. The first parameter, $1-MH$, may be interpreted as a distance measure from the average, in opposed to the affinity between pairs of sites shown in Fig. 11. Just as the coastal Birkdale site has the least affinity with other sites, so it turns out to be most distinct from

the average. It is at first sight rather surprising that the acid mire sites of Astley and Holcroft Mosses should be much closer to the average, but this may be partly due to their being relatively small sites embedded in an agricultural expanse, particularly for Holcroft Moss at only 19ha. Cholmondeley itself might be regarded as predominantly agricultural and has a similarly low divergence from the average. The most divergent sites are the large forested expanse of Delamere, and the Smithills Estate with its extensive moorland and wooded areas, though also extensive areas given over to grazing.

To highlight the contribution made by different species, a comparison is made between the observed number of records (O) at a site for each species with the number expected (E) if the chance of recording it were uniform and equal to that in the overall dataset. Under this assumption, O would follow a Poisson distribution with standard deviation \sqrt{E} . A species is considered to be significantly over-represented at a site if O exceeds E by 2.35 standard deviations or more, ie $(O-E)/\sqrt{E} > 2.35$. This is equivalent to a 99% confidence level, $p < 0.01$, for a normal distribution, which is a reasonable approximation when $E > 5$. This procedure flags up species which contribute significantly to making the site diptera fauna distinct from the overall average. Table 4 lists for each site the number of such characteristic species, and also the number of those species which are singletons both for the site and in the overall dataset. These numbers roughly correlate with the overall divergence measure. Details of the individual species found in this way to be characteristic of each site are given in the individual site reports.

Table 4 also lists how many of the 96 most frequent species in the overall dataset were recorded at each main survey site. It will be recalled that these provide 50% of the total number of records. The figures range from a low of 68 at Birkdale, the most distinct site, to 89 at Holcroft as the site nearest the average. The distribution of records amongst the sites provides insight into possible habitat preferences of these species, many of which were regarded as nationally “ubiquitous” or “abundant” in Skidmore’s (2008) classification. Table 5 shows the 96 species from the top 50% in order of descending overall number of records, together with the numbers of records for the six main sites. The sums of the squares of the residuals $(O-E)^2/E$ for each species across the six sites is also listed in Table 6. This may be compared with the χ^2 statistic in a test of whether the numbers show significant deviation from a uniform distribution across the sites and overall dataset. For six degrees of freedom, the 99% significance level falls at $\chi^2 = 16.81$. 56 of the 96 species do show a significant deviation from a uniform distribution. In many cases, this can be linked to known habitat preferences (eg *Hercostomus aerosus* with 177 records overall, *Euphyllidorea meigenii* with 81 records, *Pherbellia cinerella* with 69 records), but significant divergences may also result from the overall selection of sites and timing of surveys – as with the winter gnat *Trichocera regelationis* (64 records) recorded only on visits outside the main recording season.

Finally, it is worth noting that the combined data for the six main sites amount to 7062 records covering 936 species. The rarefaction or species accumulation curve for the full dataset ($N = 20249$, $S = 1292$) indicates that 977 species would have been expected if sampling from the complete set of locations in Figure 1 had been limited to a total of 7062 records, indicating that the six sites together constitute almost a completely representative set of locations within the wider landscape. Scaling back further to 1200 records as the standardised figure for comparing species numbers between sites (0D) gives 470 species, compared with individual sites totals in Figure ranging from 312 to 390.

8. PANTHEON RESULTS

Webb and Lott (2006) described the development by Natural England (and its predecessor English Nature) of the Invertebrate Species-habitat Information System (ISIS), a database linking invertebrate species to a range of ecological attributes such as feeding guild, hosts and prey, habitat, and conservation status. Species are grouped into assemblages characteristic of habitats, with particular application to defining site quality for nature conservation. Alexander *et al* (2004) and Drake *et al* (2007) provide further details of its development and application. The database has been further developed and made accessible to public online use via the PANTHEON website

(<https://www.brc.ac.uk/pantheon/>), which also contains details of the current scope of the database and a comprehensive glossary of terms. From the input of a species list for a site, or indeed for a taxonomic groups, PANTHEON produces a wide range of statistics on the numbers of species associated with the various ecological attributes.

The individual site reports (Brighton 2017-2020) include a variety of PANTHEON outputs, which are summarised in Table 6. This shows the numbers of species* associated with habitats grouped under broad biotopes. These are the first two of a three-level hierarchy which also included specific assemblage types (SATs) such as fungal fruiting bodies and heartwood decay in the arboreal habitat. In the earlier references, the PANTHEON habitats were called broad assemblage types (BATs) and this term is retained here to avoid confusion with more general uses of the term “habitat”. Also included within the PANTHEON results are the numbers of species with a conservation designation within each BAT.

The results show perhaps more directly than the species comparisons above how differences in the diptera fauna arise between sites. It should be emphasised that the assignment of species to the BATs is based entirely on PANTHEON, and there has been no attempt to compare with the actual habitats in which the species were recorded. Not surprisingly Birkdale is the only site with a significant proportion of species within the coastal BAT, but even so they account for only 11% of the species recorded at the site. The more wooded sites Delamere and Smithills have the smallest contribution from open habitats, while again Birkdale has the largest. This is consistent with the general character of the sites, but the dominance of “tall sward & scrub” species in all cases certainly does not reflect the actual proportion of the habitat at the sites. However, the collecting method and focus on larger diptera species are probably a significant factor in this finding. It is surprising that Smithills does register the upland BAT more strongly, but this may reflect the relatively small number of species assigned to this BAT in the PANTHEON database.

With the tree-associated BAT, the overall proportions are again in line with the amount of woodland sampled at each site during the surveys. The numbers of species in the shaded woodland floor is dominant for all sites including, rather surprisingly, Birkdale. The woodland habitat there is limited almost entirely to the alder scrub belt along the inner edge of the slacks, which can certainly be described as wet woodland. The wet woodland BAT in PANTHEON is populated with rather few species†. There are clearly numerous diptera species assigned to shaded woodland floor which are also included in other BATs.

The wetland scores also run counter to broad expectations, with the lowest contribution appearing for the acid mire site at Holcroft Moss. There is little wetland vegetation at this site apart from the dominant sward of *Molinia* with cotton-grass and various *Sphagnum* species. Also, it has already been noted above that Holcroft shows the least divergence from the overall species distribution. In contrast the nearby Astley Moss has the second-highest relative contribution from the wetland BAT, attributable to a fen area developed by “re-wilding” a carrot field.

The last row of Table 6 shows the numbers of species with a conservation designation detected by PANTHEON, again summed over the BATs so that some may be counted more than once. These figures mainly go to show that overall diversity at each sites bears little relation to the small proportions of nationally scarce or rare species which have been detected by the current survey methods.

* This includes heteroptera which were collected alongside the diptera.

† See <https://www.brc.ac.uk/pantheon/explore/traits>

9. DISCUSSION

Sampling

Trying to measure diptera diversity across the landscape presents conceptual as well as practical problems. Insect populations vary as they progress through the successive stages of metamorphosis possibly giving very large variations in numbers of individuals, or at a more fundamental ecological level, in biomass. Year-to-year changes in weather affects breeding success, while predator-prey and parasite-host relations are well-known to liable to very large and erratic cycles from generation to generation. Swarming behaviour will result in spatial heterogeneity, and day-to-day weather variations, and even time of day, affect the detectability from diptera by methods such as sweep-netting. Wilder landscapes present a random and heterogeneous pattern of vegetation on a range of length-scales. Hence even defining a suitable scale of abundance is problematic, let alone actually measuring it. The concept implicit in this report is a count of the number of 100m², that is hectare, squares in which the adults of a given species are present in a site of up to few km² at any time throughout a period of 2 to 3 years: this count is carried out repeatedly to produce a sum over the temporal as well as the spatial variation. Thus a species may achieve the same conceptual level of abundance by being present in a few areas over the site for a large proportion of the time or by present over a large area for only a short season.

Detailed and comprehensive guidance on conducting invertebrate surveys and analysing the results for the purposes of nature conservation has been provided by Drake et al (2007). This is mainly oriented at the evaluation of the invertebrate fauna of specific rare habitats, usually designated or potential SSSIs, but also envisages surveys at a landscape or regional level. Whatever the purpose of the study, resources will be limited and a sampling strategy needs to be defined to find sufficient specialist species to characterise the site. In para 2.3, Drake *et al* go on state:

“Therefore the main driver in selecting points to sample becomes a deliberate bias towards ‘promising’ patches of habitat – the features of greatest potential. In these cases, representativeness and randomness in point selection are unhelpful.”

This in contrast to “standardised scientific protocols [which] often result in an emphasis on common and widespread species”. The present surveys might perhaps be regarded as striking a balance between these opposing approaches. In progressing around a site with a sweep-net I have certainly targeted “promising” patches of vegetation, while also ensuring that a wide range of locations and variety of micro-habitats are covered on any one day. Sweep-netting is well-known to have a randomising tendency, with widely differing catches being recorded by dipterists working side-by-side. Even differing styles and sweep actions are reckoned to have an effect. My method has further inherent biases such as avoiding certain kinds of microhabitat such as bramble patches, and selecting the more unfamiliar and larger flies out of the net before the others. On the other hand, limiting the coverage of each 100m² square to a single fixed-volume aspirator container prevents over-sampling of a single location. Drake et al (2007, para 36) provide a detailed protocol for sweep-netting, including a limit of 10 minutes at each location, but split between different parts of the area. My method approximates to this protocol, but with many detailed differences.

Analysis

Given all the considerations in the last sub-section, it will be appreciated that it is not a foregone conclusion that the present survey method can produce representative and reproducible estimates of diptera biodiversity across the regional landscape covered. Even if a robotic device had been deployed to capture a sample at the locations and times of the surveys, a high degree of variability would seem likely. The intervention of the recorder in selecting areas for attention and in selecting specimens from the net introduces further uncertainty. However, the results as expressed in the various biodiversity measures presented above do show a degree of self-consistency indicating that they provide a valid comparison between the six main sites and between those sites and the wider region sampled less intensively and less systematically.

The first element of the analysis is the comparison of the individual site's species-abundance curves with the log series and the Fisher curve. The individual site reports reveal a fairly close degree of fit to the theoretical distribution, and this is also the case for the overall dataset. Thus the data seem consistent as regards this basic statistical pattern.

The analysis based on the methods of Chao *et al* (2014) is complementary as it does not fit a distribution but provides non-parametric statistical estimates of the coverage in the sampling. Their whole theory is erected on the simple concept of random sampling from a large, effectively infinite, collection of objects of different types, where the relative numbers of each type are unknown and to be estimated. At the level of the individual type or species this is equivalent sampling from a Poisson distribution, which was also the basis of Fisher *et al*'s (1943) theory. Chao *et al* distinguish between abundance and incidence statistics. In the first case, sampling proceeds indefinitely building up counts of individuals in each class. The size of the sample is defined by the total number of individuals collected, N . With incidence sampling, repeated samples of a specified size are drawn from the population and only the presence or absence of each type is recorded, not the number of individuals in that particular sample. In this approach, the count for each type or species is limited to the number of samples, and for the more frequent species this can be considerably less than the number of individuals collected.

The present method of recording is by incidence in 100m² squares, without counting numbers of individuals. However, it is appropriate to use the abundance model because although several repeat visits were made to each of the main sites, the same squares were not necessarily re-visited and in any case the differences in time of year and between year would make them effectively different contributions to a single large sample from the population as defined by the extent of the site and period of the visits. The incidence model would only apply if the populations within each hectare square on each visit had been the same. The typical 12 squares visited on each of 6 days in the main site surveys amounted to 72 samples. Use of Chao *et al*'s incidence model with our site results in fact makes very little numerical difference to the estimates of species numbers and coverage.

It will be appreciated that the application of Chao *et al*'s interpolation and extrapolation formulae to standardise the size of samples for comparisons between the sites and to estimate a total species richness represents the effects of shrinking or indefinitely extending the surveys within the original geographic, taxonomic and temporal scope. Being based on the number of records rather than a direct measure of effort as recommended by Drake *et al* (2007), this removes some elements of the surveyor's skill and efficiency from affecting the results.

Results

These surveys give an indication of the amount of effort needed to provide a meaningful measures of diptera diversity at typical nature conservation sites with a mixture of habitats. For sites ranging from 19 to over 600 ha in area, surveys spread over 2 or more years and sampling the months from April to September have produced between 850 and 1500 records in 100m squares, with between 300-400 species in each case. This is for a large range of diptera families with c. 3,300 species recorded for the British Isles. The statistical analysis of the numbers of records per species indicates that the coverage achieved in terms of abundance is between 80 and 90%, while the projected numbers of species that could be achieved by exhaustive sampling with the same methodology range from 450 to 630.

Standardising the numbers of species to 1200 records or 85% has enabled the α -diversity of the sites to be compared. The lowest values are associated with small acid mire sites with a limited amount of woodland set in an extensive agricultural landscape, while the most diverse site is one of the largest with more varied wetland habitats, woodland, upland heath and moorland.

The comparisons between pairs of sites gave a consistent picture with the two nearby acid mire sites having the most similar biodiversity profile. Although being relatively low on the α -diversity scale, the coastal Birkdale site stood out as clearly distinct from all the other sites on the β -diversity indices. The

geographical separation of sites (Figure 1) does not seem to be a major factor: with Cholmondeley being reasonably close to Smithills on this measure but with a separation of 63km.

Examination of the full dataset indicates that the six main sites contain a large proportion of the regional biodiversity in terms of number of species: adjusted for equal amounts of recording, the six sites' total of 936 species compares with 977 from the full range of sites. For the 96 species providing over 50% of the records, only 44 have been recorded at all of the 6 sites. Conversely, 56 of these species had a distribution across these sites departing significantly from uniformity, indicating a degree of localisation mostly associated with habitat.

The proportions of species associated with the four broad biotopes according to PANTHEON were qualitatively consistent with the general character of the six main sites, except for the wetlands. This may reflect the small area and varying numbers of such features such as ponds and streams. Efforts were made during each of the surveys to visit known examples of water and boggy patches, while not detracting from a more comprehensive coverage.

The PANTHEON results showed no correlation between site diptera biodiversity and the number of species with a national conservation designation. This could be a purely statistical effect for species which are not abundant even in their particular habitat and fly for only a very limited period. Of course, when such species are detected by general surveys, more targeted searches or special techniques can be used to obtain data more relevant to assessment of conservation value.

While the survey methodology involved conscious choices of sites and habitats within them, the overall internal consistency of the results suggests that a representative sample has been obtained. It is difficult to see how this could be improved. Random selection of 100m squares within a site would be liable to miss small but significant features. One might attempt to weight samples according to the proportion of different habitats at each site but relevant data is usually not readily available. However, it would be informative to gather and analyse data from more uniform sites such as stretches of upland moor or larger acid mire sites.

The use of number of records as a measure of effort reduces but does not eliminate the possible influence of the observer's skill and efficiency in collecting specimens. Comparative data may be available from the records of dipterists who have carried out general sweep-netting with systematic comprehensive recording in their home areas for many years. This need not cover the same range of taxa as in this report, as all our analyses could have been carried out for individual families or recording scheme groups.

REFERENCES

- Abramowitz, M. & Stegun, I. A. (1972) *Handbook of Mathematical Functions*, 8th printing, Dover Publications.
- Alexander, K. N. A. *et al.* (2004) Invertebrate assemblages on English SSSIs, *English Nature Research Reports*, No. 618
- Brighton, P. (2017) *Heteroptera and Diptera Surveys in the Delamere Forest, 2013-2017*, Report to Cheshire Wildlife Trust
- Brighton, P. (2018) *Heteroptera and Diptera Surveys on the Manchester Mosses with PANTHEON analysis*, Report to Lancashire and Cheshire Wildlife Trusts.
- Brighton, P. (2020a) *Heteroptera and Diptera Surveys on the Ainsdale and Birkdale Hills LNR*, Report to Natural England and Sefton Council.
- Brighton, P. (2020b) *Heteroptera and Diptera Surveys on the Cholmondeley Estate*, Report to the Cholmondeley Estate and Cheshire Wildlife Trust.

- Brighton, P. (2020c) *Heteroptera and Diptera Surveys on the Smithills Estate*, Report to the Woodland Trust.
- Chandler, P. J. (2020) *An Update of the 1998 Checklist of Diptera of the British Isles*, <https://www.dipterists.org.uk/checklist>
- Chao, A. (1984) Non-parametric estimation of the number of classes in a population, *Scand. J. Stat.* **11**, 265-270.
- Chao, A. & Chiu, C.-H. (2016) Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures, *Methods in Ecology and Evolution* **7**, 919-928.
- Chao, A. & Jost, L. (2013) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size, *Ecology*, **93**, 2533-2547.
- Chao, A. *et al.* (2008) A two-stage probabilistic approach to multiple-community similarity indices, *Biometrics* **64**, 1178-1186.
- Chao, A. *et al.* (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies, *Ecological Monographs*, **84**(1)45–67.
- Collin, J. E. (1961) *British Flies*, Volume VI: Empididae, Cambridge University Press.
- Drake, C. M. *et al.* (2007) Surveying terrestrial and freshwater invertebrates for conservation evaluation, *Natural England Research Report*, NERR005.
- Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population, *J. Anim. Ecology* **12**, 42-58.
- Gotelli, N. J. & Chao, A. (2013) Measuring and estimating species richness, species diversity, and biotic similarity from sampling data, in: Levin S.A. (ed.) *Encyclopedia of Biodiversity*, 2nd edition, Volume 5, pp. 195-211, Academic Press.
- Hsieh, T.C., Ma, K. H. & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, **7**, 1451-1456.
- Magurran, A. E. (2004) *Measuring Biological Diversity*, Blackwell.
- May, R. M. (1975) *Patterns of Species Abundance and Diversity*. Typescript, 89pp, Biology Department, Princeton University [available from British Library Document Supply Centre, Boston Spa].
- Skidmore, P. (2008) A review of the Diptera of the Western Isles of Scotland, *Dipterists Digest* **15**, 99-194.
- Stubbs, A. E. & Drake, M. (2014) *British Soldierflies and their Allies*, 2nd Edition, British Entomological and Natural History Society.
- Webb, J. R. & Lott, D. A. (2006) The development of ISIS: a habitat-based invertebrate assemblage classification system for assessing conservation interest in England, *J. Insect Cons.* (2006) **10**,179–188.

APPENDIX: DERIVATION OF THE FISHER CURVE FROM THE LOG SERIES

Part 3 of Fisher *et al* (1943), by R. A. Fisher himself, contains a detailed derivation of the theory. The fundamental assumptions are quite few. Random samples are taken from some specified environment in which there is a very large number of species and individual organisms. However numbers of each species present vary very widely from the commonest to the rarest. For each species the numbers obtained if the sampling were repeated on different occasions would follow the Poisson distribution. By assuming a certain distribution for the Poisson parameter itself over the whole range of an infinite number of species, Fisher obtains the result that in a sample of N individuals with S species detected, the numbers of species with 1, 2,... n individuals in the sample follows a simple sequence $\alpha x, \alpha x^2/2, \dots, \alpha x^n/n$: this is known as the logarithmic series (Magurran, 2004) because these

are the terms of the infinite series expansion of $-\alpha \ln(1-x)$. The parameters α and x are derived from N and S in the simultaneous equations

$$S = \alpha \ln(1 + N/\alpha), \quad x = N/(N + \alpha).$$

These equations cannot be solved in a closed form. Fisher *et al* (1943) provide tables from which they may be solved by interpolation, a procedure which can be easily implemented in an Excel spreadsheet. x is always close to 1 in practical situations so α itself approximates to the predicted number of singletons.

While the Fisher sequence can be directly read off as the numbers of species with small numbers of individuals, a difficulty arises in interpreting it for the common species. With the example of the dataset analysed in this report the log series predicts the number of species with $n = 200$ records as 0.045. Clearly this must be interpreted as the chance of there being a species with exactly this number of records. So we might expect that in a range of 20 or so values of n around 200, we would typically get one occurrence of a species at one particular value. In fact there are three species in this range on 195, 204, and 207 records, but that is not particularly unlikely with the statistics of low numbers.

The resolution of this problem is as follows. The theoretical Whittaker plot is built up starting from the right hand end where the species rank $r = S$ and $n = 1$, and adding a step up of 1 up as r is reduced by the next term in the log series. While this could be done explicitly in the spreadsheet, it is simpler and more revealing to replace the summation by the equivalent integral over n . The result is (May 1975):

$$r = \alpha E_1(-n \ln x)$$

where E_1 is the exponential integral function as defined in §5.1.1 of Abramowitz and Stegun (1972) - §§5.1.53 and 5.1.54 of that work provide approximations to this function which can be coded in Excel formulae to generate the theoretical “Fisher curves” plotted in red in the Figures. For the higher rank values, the curves approximate to a straight line passing through the steps of the log series.

May (1975) discussed several other abundance distributions representing various ways in which resources or ecological niches might be allocated amongst the species in the community. The log-series could be derived from a hypothesis of extreme “niche pre-emption”. At the other extreme, a community comprising a limited number of taxonomically similar species in competitive contact with each other in a relatively homogeneous would give rise to the so-called “broken-stick” distribution. Intermediate cases involving a large number of interacting random factors were associated with log-normal distributions.

The range of Diptera in the present surveys represent a large range of life-styles and clearly do not represent a closed ecological system. One might therefore expect the various random effects in the underlying distribution of species and their detectability to result in a log-normal distribution, which might however be “veiled” if the sampling is insufficient (Magurran, 2004, chapter 2). In fact, May (1975) also stated that the log-series could arise from limited sampling. In the present context, the log-series and the Fisher curve have been found to provide a good representation of the data at both regional or site level to guide sampling strategy.

TABLES

Table 1: coverage of diptera taxa

Group	Families	No. of British species
CRANEFLIES	Tipulidae, Cylindrotomidae, Pediciidae, Limoniidae, Trichoceridae, Ptychopteridae	352
HOVERFLIES	Syrphidae	283
EMPIDS AND DOLLIES	Atelestidae, Hybotidae, Empididae, Brachystomatidae, Dolichopodidae	700
CALYPTRATES	Hippoboscidae, Nycteribiidae, Scathophagidae, Anthomyiidae, Fanniidae, Muscidae, Calliphoridae, Rhiniidae, Rhinophoridae, Sarcophagidae, Tachinidae, Oestridae	1056
OTHER DIPTERA		
Mosquitoes	Culicidae (35)	909
Soldierflies & allies	Xylophagidae, Athericidae, Rhagionidae, Tabanidae, Xylomyidae, Stratiomyidae, Acroceridae, Bombyliidae, Therevidae, Scenopinidae, Asilidae (160)	
Platypezidae	Opetiidae, Platypezidae (35)	
Conopidae & others	Lonchopteridae, Conopidae, Pallopteridae, Ulidiidae, Platystomatidae (66)	
Tephritidae	Tephritidae (77)	
Stilt & stalk flies	Pseudopomyzidae, Micropezidae, Tanypezidae, Strongylophthalmyiidae, Megamerinidae, Psilidae (41)	
Sciomyzidae	Sciomyzidae (72)	
Sepsidae	Sepsidae (29)	
No Recording Scheme	Bibionidae, Anisopodidae, Dixidae, Lauxaniidae, Dryomyzidae, Opomyzidae, Asteiidae, Heleomyzidae, Trixoscelididae, Sphaeroceridae, Drosophilidae, Campichoetidae, Diastatidae (394)	
TOTALS		3300

Table 2: overall composition of the full dataset

Group	Number of records	Number of species	% of British list
Craneflies	3002	176	50.0%
Hoverflies	2886	120	42.4%
Empidoidea	3692	255	36.4%
Calyptrates	5386	368	34.8%
Selected other diptera	5283	373	41.0%
TOTALS	20249	1292	39.2%

Table 3: biodiversity statistics for the overall dataset and for the six main sites

Dataset	<i>N</i>	<i>S</i>	Fisher's α	f_1	f_2	Chao1	Coverage
Overall	20249	1292	307.4	284	170	1766	98.6%
Astley (A)	1026	313	153.5	155	51	548.3	84.9%
Birkdale (B)	985	325	169.3	138	79	450.5	85.8%
Cholmondeley (C)	842	312	179.4	153	69	481.4	81.8%
Delamere (D)	1477	399	179.6	162	74	578.4	89.0%
Holcroft (H)	1477	341	139.0	143	51	541.3	90.3%
Smithills (S)	1255	398	201.0	182	70	634.4	85.5%

Table 4: measures of divergence of sites from the overall abundance distribution. *MH* is the Morisita-Horn index formed from the individual site and overall dataset frequencies. The 96 commonest species providing 50% of the records in the overall dataset are listed in Table 5.

Site	1- <i>MH</i>	Number of species with residual >2.35	Number of singletons unique to site	Number of species from overall top 50%
Astley (A)	0.27	67	12	87
Birkdale (B)	0.42	136	25	68
Cholmondeley (C)	0.24	82	15	81
Delamere (D)	0.31	114	22	86
Holcroft (H)	0.16	64	9	89
Smithills (S)	0.34	132	32	82

Table 5. The numbers of records in the overall dataset (ALL) and at the main six sites for the 96 species providing 50% of all the records. The χ^2 test statistic has been evaluated for the deviations of the six site values from the frequency in the overall dataset. For $\nu = 6$, the χ^2 distribution has the value 16.81 at the 99% significance level.

Species	ALL	A	B	C	D	H	S	χ^2
<i>Scathophaga stercoraria</i>	458	14	13	30	29	43	32	17.6
<i>Lonchoptera lutea</i>	397	26	10	18	29	24	17	9.6
<i>Melanostoma mellinum</i>	235	22	6	11	6	24	9	23.4
<i>Melanostoma scalare</i>	223	14	4	4	13	13	12	9.5
<i>Eristalis pertinax</i>	213	9	1	2	21	21	5	23.0
<i>Episyrphus balteatus</i>	208	8	9	4	12	23	5	12.8
<i>Opomyza florum</i>	204	15		10	25	38	1	65.8
<i>Campsicnemus curvipes</i>	183	13	3		24	14	19	26.7
<i>Coenosia tigrina</i>	182	9	3	9	4	27	2	32.5
<i>Hercostomus aerosus</i>	177	23	1		41	15	19	103.4
<i>Opomyza germinationis</i>	174	4	1	5	16	6	16	16.8
<i>Sepsis fulgens</i>	167	15	2	15	19	12	10	22.8
<i>Sepsis cynipsea</i>	164	8	7	11	21	11	10	9.6
<i>Campsicnemus scambus</i>	160	36	5	1	31	28	4	160.2
<i>Geomyza tripunctata</i>	159	18	4	4	4	16	7	22.6
<i>Limonia nubeculosa</i>	157	2		5	21	23	13	33.2
<i>Fannia serena</i>	154	4	1	15	7	6	10	23.1
<i>Helophilus pendulus</i>	152	9	2	4	8	23	3	23.0
<i>Lonchoptera bifurcata</i>	152	8	6	10	6	13	5	7.2
<i>Platycheirus albimanus</i>	148	5		1	12	12	5	14.5
<i>Dicranomyia modesta</i>	141	10	4	2	10	40	9	90.8
<i>Hybos culiciformis</i>	140	5	9	2	30	7	7	43.5

Species	ALL	A	B	C	D	H	S	χ^2
<i>Scathophaga furcata</i>	137	13	2	1	6	15	23	41.3
<i>Dilophus febrilis</i>	135	5	5	10	3	8	13	12.0
<i>Sympycnus desoutteri</i>	134	10	21	4	9	6	4	37.9
<i>Syritta pipiens</i>	134	4		1	6	8	2	18.0
<i>Tricyphona immaculata</i>	132	11	4	7	9	10	6	4.8
<i>Eriocnopa trivialis</i>	132	8		3	4	9	11	12.1
<i>Dicranomyia chorea</i>	118	4	1		3	16		26.8
<i>Bicellaria vana</i>	117	4	3	6	7		12	14.1
<i>Sepsis orthocnemis</i>	113	3			14	9	11	17.9
<i>Botanophila fugax</i>	113	3	6	11	4	5	8	13.4
<i>Delia platura</i>	112	1	21	7	6	3	3	55.5
<i>Cheilotrichia cinerascens</i>	112	6	1	5	4	10	3	8.5
<i>Tipula oleracea</i>	109	9	12	3	3	11	3	17.5
<i>Tipula paludosa</i>	107	6	4	7	2	13	2	12.8
<i>Platycheirus clypeatus</i>	105	5	2	5	9	14	2	10.6
<i>Eristalis tenax</i>	103			2	6	6	1	16.6
<i>Lispocephala erythrocerata</i>	101	24	9		8	11		85.4
<i>Ocydromia glabricula</i>	98	3		13	20	10	5	49.5
<i>Opomyza petrei</i>	97	2	7	12	2		11	33.4
<i>Pegoplata infirma</i>	91	1	2	13	1	7	12	38.6
<i>Calliphora vicina</i>	91		1			5		23.7
<i>Siphona geniculata</i>	90	3	14	9	2	3	6	34.2
<i>Euphyllidorea meigenii</i>	81	2			6	35		156.7
<i>Dolichopus plumipes</i>	81	2	2	5	4	1	14	23.6
<i>Fannia mollissima</i>	81	2		1	2	9	8	12.7
<i>Syrphus ribesii</i>	80	2		1	6	2	2	10.9
<i>Sepsis punctum</i>	79	6		5	6	11	1	13.6
<i>Sylvicola cinctus</i>	78	1	2	2	6	10	7	7.8
<i>Symplecta stictica</i>	78	1	14		2	2		42.5
<i>Phaonia tuguriorum</i>	78	1		2	3	5	2	9.5
<i>Hylemya vagans</i>	75			9	6	9	7	22.1
<i>Chrysotus gramineus</i>	75	4		4	5	3	3	5.7
<i>Rhagio lineola</i>	72	5	6	2	10	3	4	7.9
<i>Molophilus griseus</i>	71	9	2	6	9	3	5	15.7
<i>Leptocera fontinalis</i>	71	5	1	6	5	1	2	10.1
<i>Helina depuncta</i>	70	4	3	4	7	3	4	2.1
<i>Helina reversio</i>	70	1	12	6	2	1	1	34.6
<i>Campsicnemus loripes</i>	69	23		2	13	8	12	140.7
<i>Pherbellia cinerella</i>	69	1	31			2		243.5
<i>Hybos femoratus</i>	68	5	3	1	10	2	19	60.7
<i>Dolichopus simplex</i>	68	15	1		1	8		52.4
<i>Haematopota pluvialis</i>	68	9		1	3	8	2	17.2

Species	ALL	A	B	C	D	H	S	χ^2
<i>Austrolimnophila ochracea</i>	67	1		2	15	5	7	28.1
<i>Hercostomus cupreus</i>	65	5			24	11	3	93.5
<i>Empis nigripes</i>	65	3	1	4	1	2	6	7.6
<i>Eristalis arbustorum</i>	65	4		2	2	3		9.8
<i>Trichocera regelationis</i>	64	14		1	1	19		90.7
<i>Neoascia tenur</i>	63	16	3	2	4		1	58.4
<i>Lotophila atra</i>	63	1	4	6	2	5	2	8.6
<i>Hylemya variata</i>	61		3	9	1	7	8	28.4
<i>Hebecnema vespertina</i>	61	5	2	5	1	12		23.2
<i>Dolichopus trivialis</i>	61	1		2		3	3	9.6
<i>Meiosimyza rorida</i>	60		1	7	2	2	7	17.9
<i>Opetia nigra</i>	59	5	1	2	6	11	2	14.5
<i>Beris chalybata</i>	59	1		5	4	4	8	12.0
<i>Morellia aenescens</i>	59	3					8	19.1
<i>Chrysopilus cristatus</i>	58	3	12		1	1	2	37.9
<i>Copromyza nigrina</i>	58	4	1	1	2	2	7	8.0
<i>Tephritis neesii</i>	58	5		2	3	5	2	5.5
<i>Sphaerophoria scripta</i>	58	1	9	2		1		25.2
<i>Dolichopus popularis</i>	57	6		1	1	4	14	40.3
<i>Rhagio scolopaceus</i>	56	2	1	1	1	6	9	14.1
<i>Molophilus obscurus</i>	56	1	11	1		1		37.0
<i>Empis livida</i>	56	9				3	1	24.6
<i>Tetanocera elata</i>	55	10	2	2	2	4	7	23.7
<i>Anthomyia liturata</i>	55		4	11	1		2	43.5
<i>Hilara maura</i>	55		3	2	3	5	1	5.1
<i>Lasiomma seminitidum</i>	54		1	2	3	3	1	5.9
<i>Minettia fasciata</i>	53		23	2	1		1	172.1
<i>Chloromyia formosa</i>	53	4	1	2		6		10.0
<i>Paradelia intersecta</i>	53	1	4	1	5	1	1	6.6
<i>Prionocera turcica</i>	52	10	1		16	2		67.0
<i>Nanna fasciata</i>	52	2		5			4	14.2
<i>Platycheirus angustatus</i>	51	4	1	5		5	3	9.7

Table 6. Summary of results from PANTHEON (Version 3.7.6) for the six main sites, for numbers of species by broad assemblage type (BAT) and proportion of the species at the sites in the four overarching biotopes. The proportions add up to more than 100% for each site because species may belong to more than one BAT and wet woodland occurs in more than one biotope. The number of species with a conservation designation is also given as a sum of the results for the individual BATs, which for the same reason may exceed the number actually recorded.

Site:	H	A	B	C	D	S
coastal						
brackish pools & ditches	1	1	10		1	
rocky shore			3			
saltmarsh	2	1	19		1	
sandy beach			5			
Total coastal (%)	0.9%	0.6%	11.4%	0.0%	0.5%	0.0%
open habitats						
short sward & bare ground	10	11	30	6	10	5
tall sward & scrub	104	104	101	85	86	84
upland	2	2	2	2	3	6
Total open habitats (%)	34.0%	37.4%	40.9%	29.8%	24.8%	23.9%
tree-associated						
arboreal	20	18	7	8	20	16
decaying wood	15	7	3	9	22	9
shaded woodland floor	120	98	69	96	161	160
wet woodland	19	25	13	20	54	47
Total tree-associated (%)	51.0%	47.3%	28.3%	42.6%	64.4%	58.3%
wetland						
lake			1			
marshland	38	51	48	45	43	34
peatland	50	72	52	55	81	54
running water	31	36	28	25	60	64
wet woodland	19	25	15	20	55	47
Total wetland (%)	40.5%	58.8%	44.0%	46.5%	59.9%	50.0%
Number of species conservation designations	18	20	27	14	37	6

FIGURES

Figure 1. Relief map showing vice-county boundaries of VCs 58, 59 and 60. Coloured squares are tetrads (2km square) with numbers of records shown on a “heat scale” (blue low to red high). Letters indicate the locations of the six sites for detailed analysis (see text).

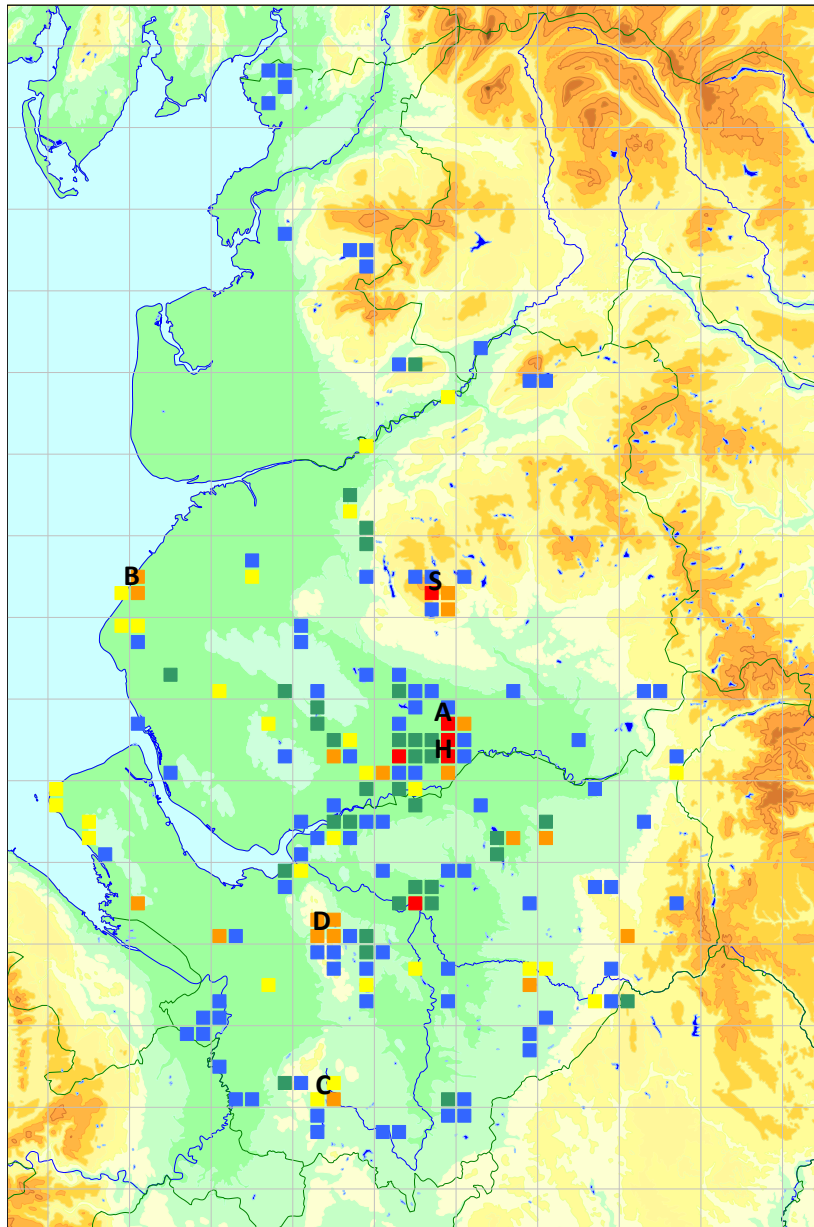


Figure 2. Stacked-area chart showing the numbers of records gathered each year for the five diptera groups

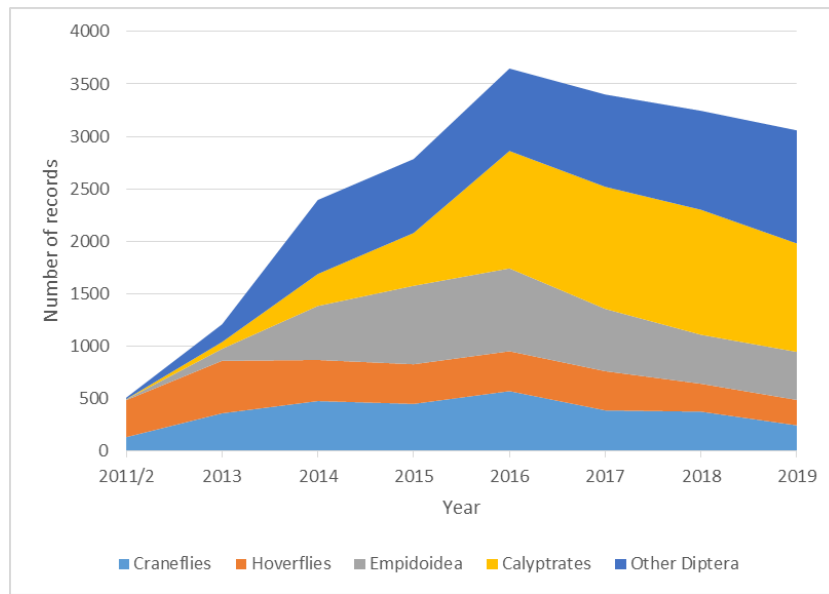


Figure 3. Numbers of records per group by calendar month (a) stacked and (b) individual variation

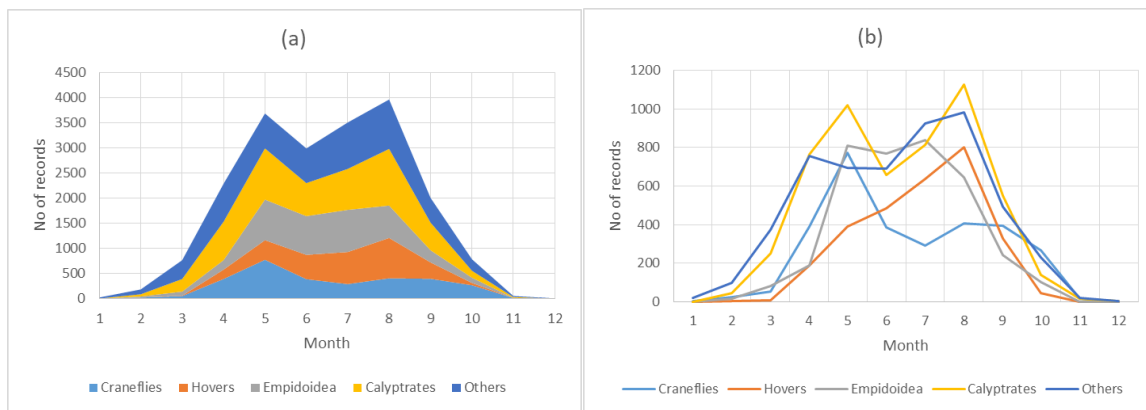


Figure 4. Proportions of each diptera group by records at each of the 6 main sites

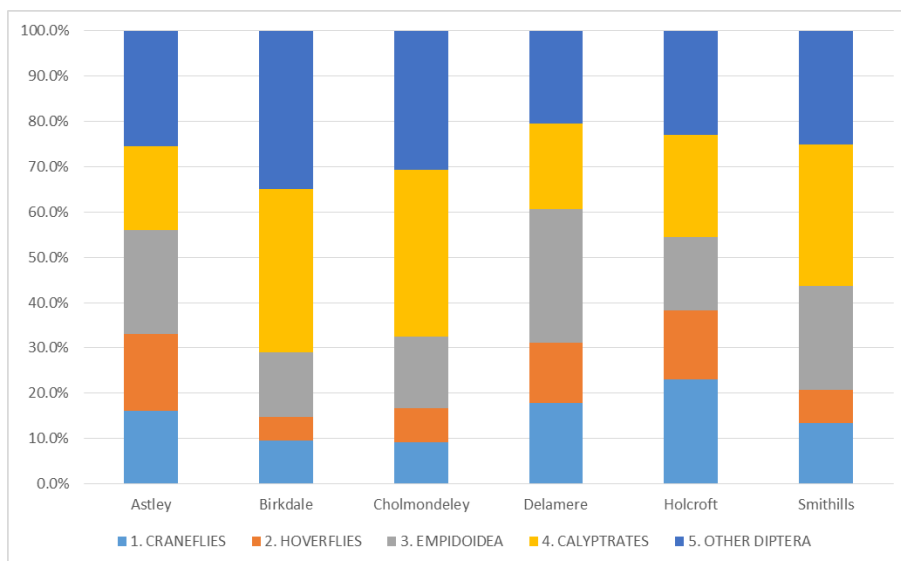


Figure 5. Numbers of records by month at each of the 6 main sites

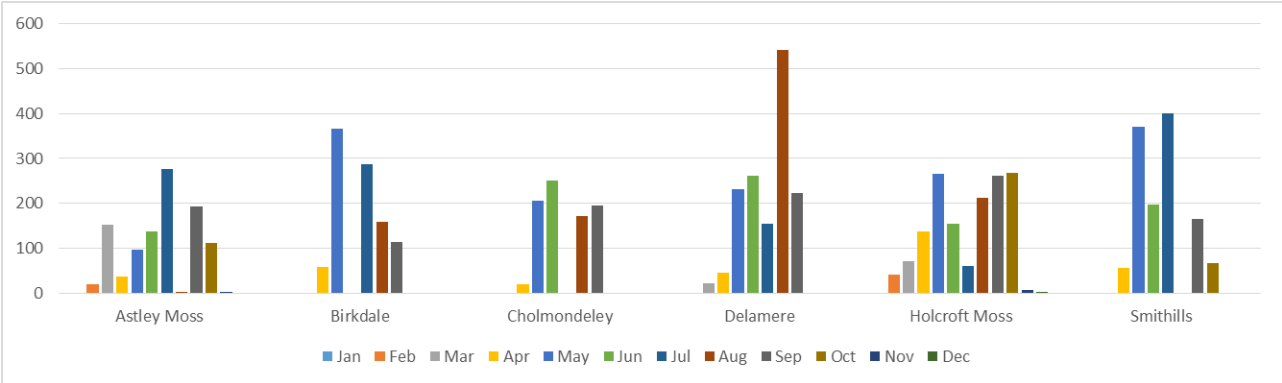


Figure 6. Whittaker plot for the full dataset of $N = 20249$ records covering $S = 1292$ species in vice-counties 58, 59 and 60. The blue curve is the smoothed theoretical distribution of Fisher *et al* (1943), given by $r = \alpha E_1(-n \ln x)$ (see Appendix)

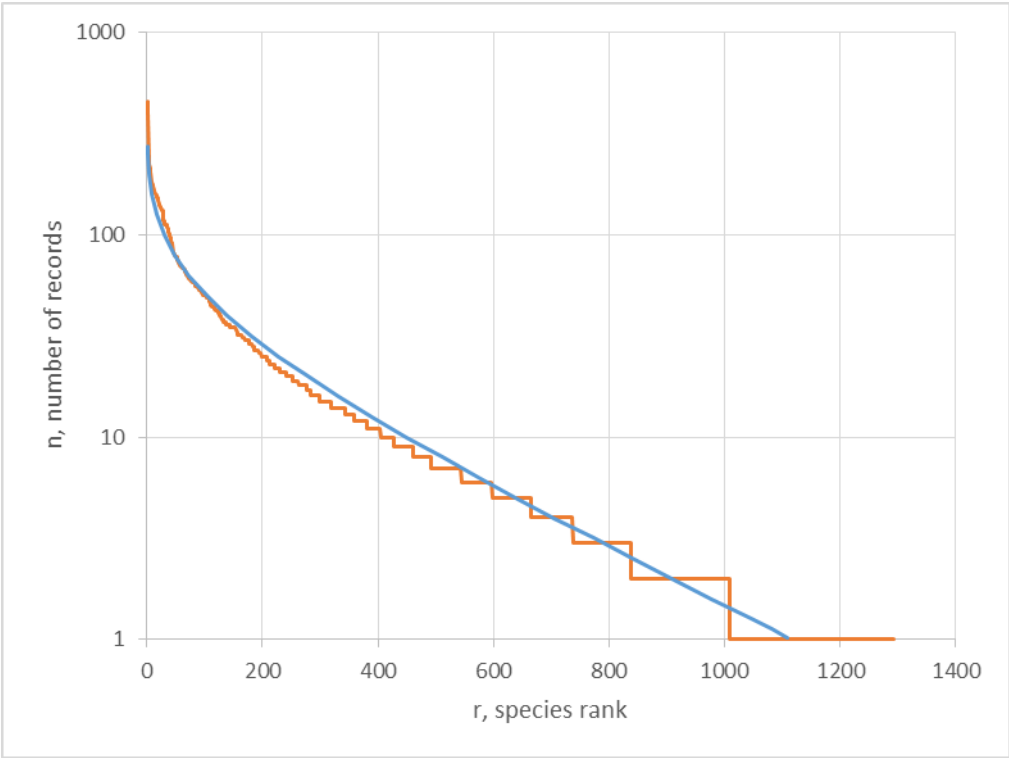


Figure 7. Whittaker plot for Astley Moss. The orange curve is the smoothed theoretical distribution of Fisher *et al* (1943), given by $r = \alpha E_1(-n \ln x)$ (see Appendix)

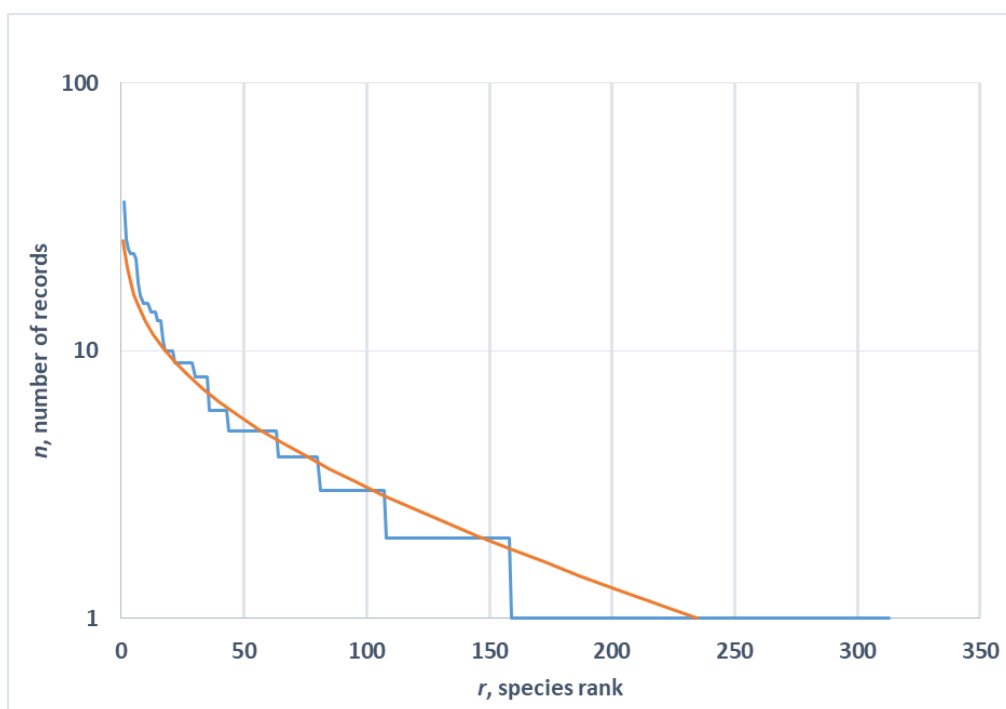


Figure 8. Species accumulation curves derived from the empirical numbers of records of species for the six main sites

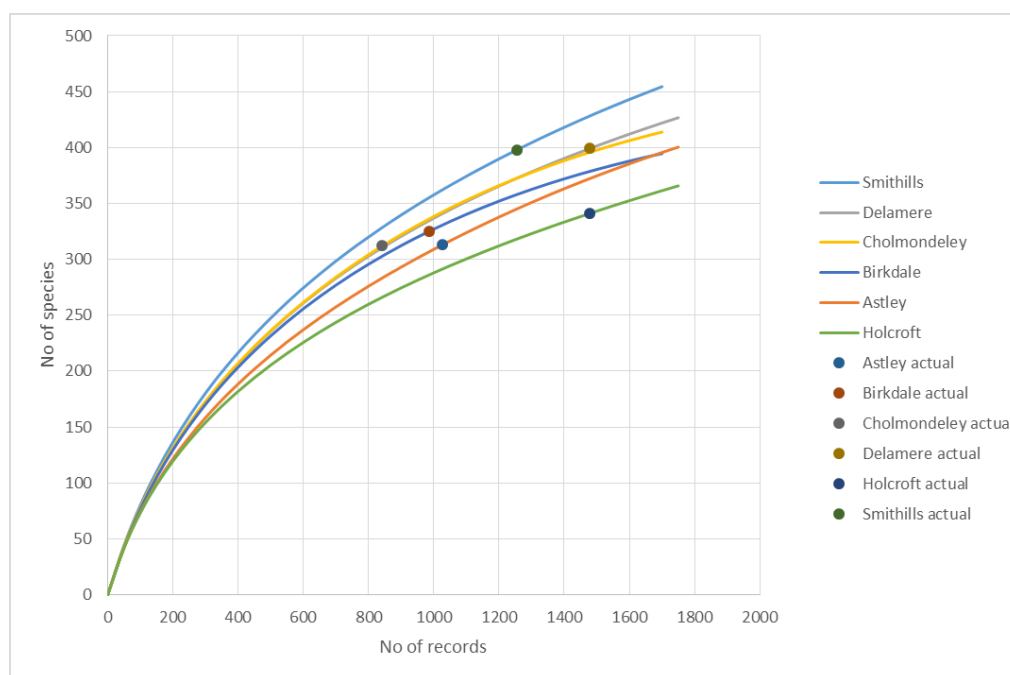


Figure 9. Hill numbers (effective numbers of species) for the six sites at sample size of 1200 records with 95% error bars evaluated by INEXT using 200 bootstraps.



Figure 10. Coverage as a function of sample size derived from the empirical numbers of records of species for the six main sites.

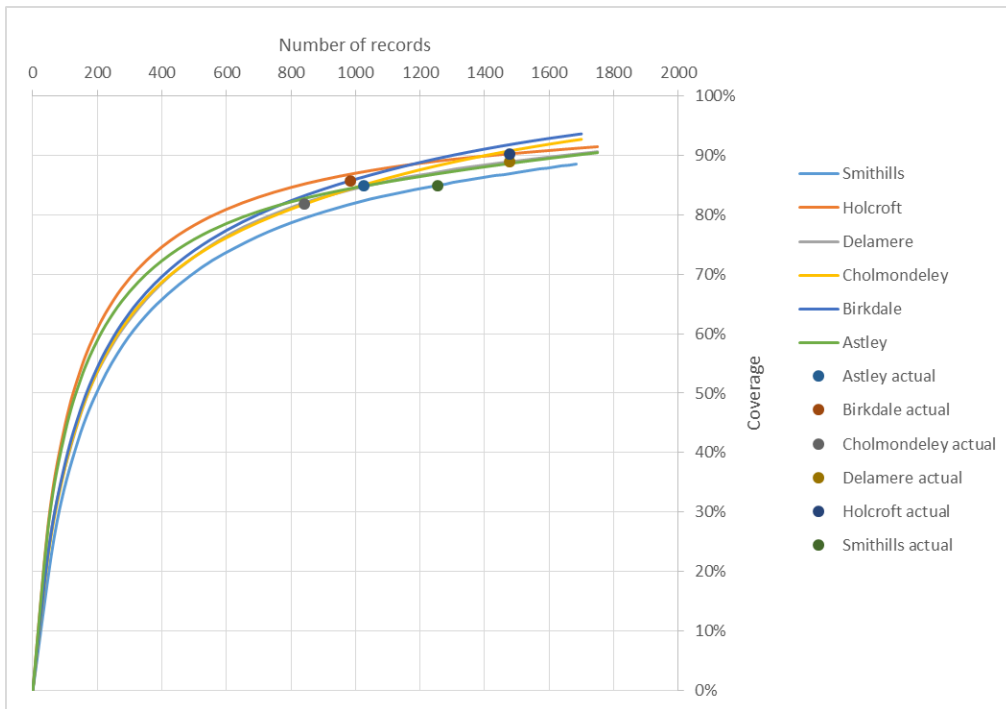


Figure 11. Hill numbers (effective numbers of species) for the six sites at sample coverage of 85% with 95% error bars evaluated by INEXT using 200 bootstraps.

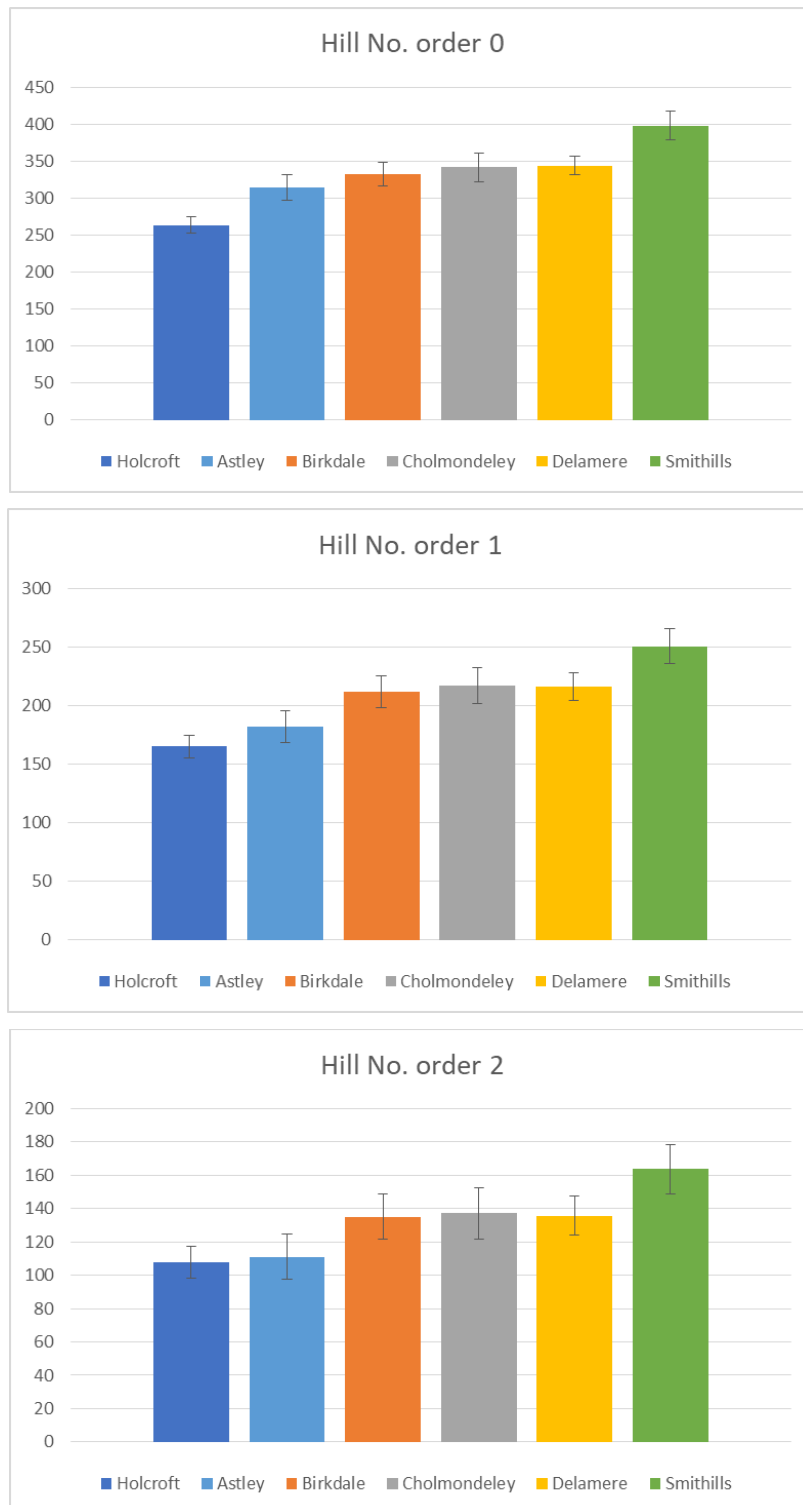


Figure 12a. Pairwise site comparisons with the Sørensen index

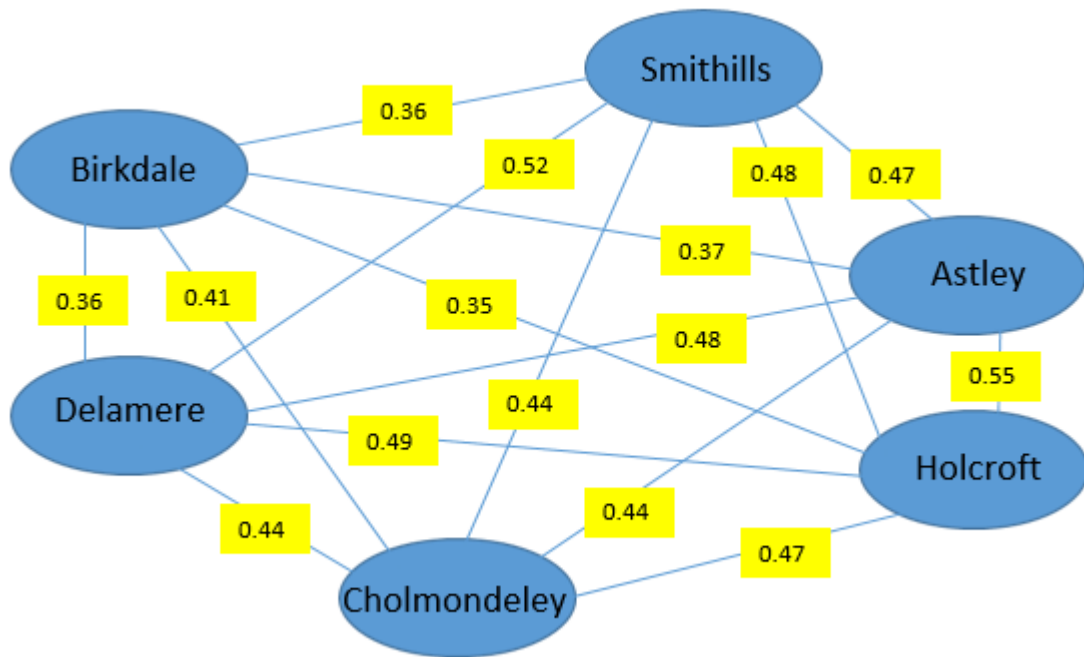


Figure 12b. Pairwise site comparisons with the Morisita-Horn index

