USING TWINSPAN TO EXTRACT ECOLOGICAL TRENDS FROM OPPORTUNISTIC DATA GATHERED BY AMATEURS: A CASE STUDY OF SHROPSHIRE MICROLEPIDOPTERA

A. G. BLUNT, I. C. TRUEMAN and R. J. BURKMAR

This study explores the use of TWINSPAN (Hill, 1979; Hill & Šmilauer, 2005) to analyse microlepidoptera data gathered through opportunistic recording by amateurs in 1991-2015 in the vice-county of Shropshire (VC40). Biases in the database are explored and found to be consistent with those identified for opportunistic data by the literature. Following an initial test the database was divided into two files, a Day file containing data obtained by daytime recording methods (searching, sweep-netting, beating vegetation) and a Night file containing data obtained by light-trapping. A minimum number of species per sample was established experimentally. Each file was then analysed separately using TWINSPAN.

Most groups in the Day file output may be interpreted along ecological lines, and suggest the influences chiefly of seasonality, soil type and associated vegetation, altitude and anthropogenic factors including habitat management. Indicator species in the Day file output are mostly monophagous or near-monophagous moths, allowing the interpretation of groups through the ecologies of larval host plants. Ecological trends suggested by the Night file output are not strong, and light-trapping appears to offer little potential for extracting ecological information. The main influence found to affect the Night file output is sample size: samples with fewer species are separated by the program from those with more species. This finding may have implications for the use of light-trapping data in understanding and mapping species distributions. Some methodological issues arising from the use of TWINSPAN in this study are also discussed.

INTRODUCTION

The recent rapid growth of biological recording carried out by amateurs has produced very large data sets for many groups of organisms. The Shropshire microlepidoptera database is one such. Originally maintained regionally for the Watsonian vice-county of Shropshire (VC40) that includes Telford & Wrekin, since 2016 it has been located within Butterfly Conservation's National Moth Recording Scheme. Concomitant with this growth of recording has been the development of citizen science programmes seeking to use volunteer data in monitoring national or international trends, particularly in biogeography, climate change ecology and conservation biology (Powney & Isaac, 2015). For example, Väisänen *et al*. (1991) used the computer program TWINSPAN (Hill, 1979; Hill & Šmilauer, 2005) to investigate biogeographical variations in species lists including one family of Lepidoptera (Sesiidae) in Fennoscandia and Denmark. The present study explores the data for one group of species, the microlepidoptera, in a region of west-central England, Shropshire, whose topography has an altitudinal range of below 50m to 540m. Its principal focus is to assess the extent to which TWINSPAN may extract regional ecological trends among the microlepidoptera: a focus that appears to be new.

The Shropshire microlepidoptera database

'Microlepidoptera' refers to a group of families of mostly very small moths. It is not a strict taxonomic group but a conventional treatment of part of the Lepidoptera, and its composition is not interpreted alike by all authors. Here microlepidoptera are conceived as comprising the 50 families allocated conventionally to the group by British authors, i.e. families numbered 1-2, 4-49 and 62-63 in Agassiz *et al*. (2013) and described by Sterling *et al*. (2012). This paper therefore excludes from microlepidoptera the families Hepialidae, Cossidae, Sesiidae, Limacodidae and Zygaenidae, that may be included in this group by some continental authors.

The database used in this study covers the years 1991-2015 inclusive, and comprises records published by Blunt (2014) along with other records to the end of 2015. It contains 24,783 records for 745 species at 2,680 locations in 931 tetrads.

The data are typical of those gathered by amateur volunteers in originating from opportunistic recording in which locations, times and methods are determined by individual recorders. Some degree of structure may be involved, e.g. for this database an attempt has been made to visit and record almost all tetrads in the vice-county (Figure 1). However, the spatial and temporal patterns of such recording, and of the data produced, are very different from those of stratified random sampling protocols (Tulloch *et al*., 2012), and any apparent 'design' in them cannot easily be captured as metadata (Pocock *et al*., 2015). Only three Shropshire locations have been sampled systematically in a fully standardized manner designed to provide data for scientific analysis: at Ludlow, Pennerley and FSC Preston Montford (Figure 2), where light-trapping has been conducted as part of the Rothamsted Insect Survey (Gould & Woiwod, 2009). Elsewhere the database is affected by biases typical of opportunistic sampling: uneven recording intensity over time, uneven spatial coverage, uneven sampling effort per visit, and uneven detectability across space and time (Isaac *et al*., 2014; Isaac & Pocock, 2015). Sites of high conservation value have attracted most attention from recorders, with 28% of the data coming from just four such sites: Wyre Forest, Whixall Moss, Bettisfield Moss, Prees Heath (Figure 2). Recorders' gardens have also been sampled extensively and contribute a further 20% of data from 13 regularly-operated static light-traps. This level of effort contrasts with that in almost half the tetrads (459 = 49.3%) where sampling has been carried out on only one occasion each, often for less than an hour. Uneven spatial coverage is demonstrated by Figure 1, which shows the number of monads sampled between 1991 and 2015 in each of 805 tetrads in which all four monads have public access: 479 tetrads (59.5%) were sampled in one monad, 221 (27.4%) in two, 77 (9.6%) in three and 28 (3.5%) in four. Uneven detectability of taxa and difficulties of identification have led to disproportionately few records of Psychidae, Coleophoridae, Elachistidae and Gelechidae in the database (Blunt, *loc.cit.*).

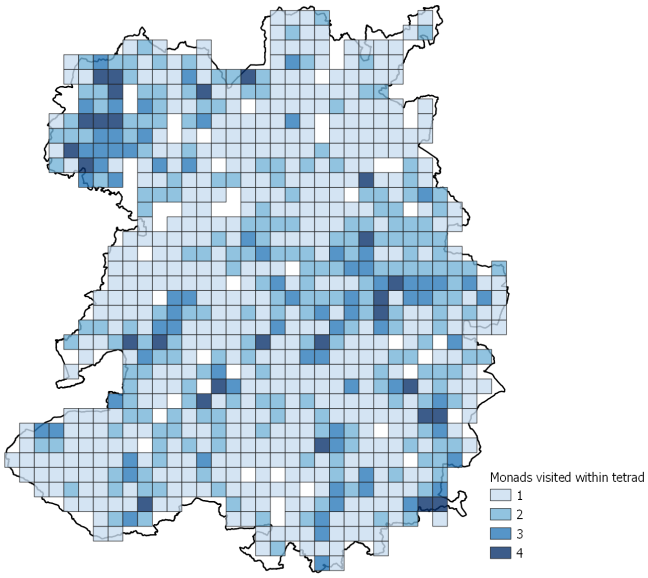


FIGURE 1. Map of VC40 Shropshire showing the number of monads sampled in 1991-2015 in tetrads where all four monads have public access. Uncoloured tetrads have one or more monads without public access, or were not sampled in 1991-2015.

Recording intensity is hard to quantify accurately, as site visits and light-trapping sessions that fail to record microlepidoptera have not been captured by the database. Of the records for which a month is specified (94.6% of all records), 2.7% are from the period November to March, 43.3% from April to mid-July, and 48.6% from mid-July to October, figures which suggest that overall recording activity across the year is reasonably

consistent with the natural seasonality of the fauna. Recording intensity between years, however, is very uneven: 7.8% of data is from the decade 1991-2000, 33.1% from 2001-2010 and 59.1% from 2011-2015; the year 2013 alone accounts for 25.1% of database records.

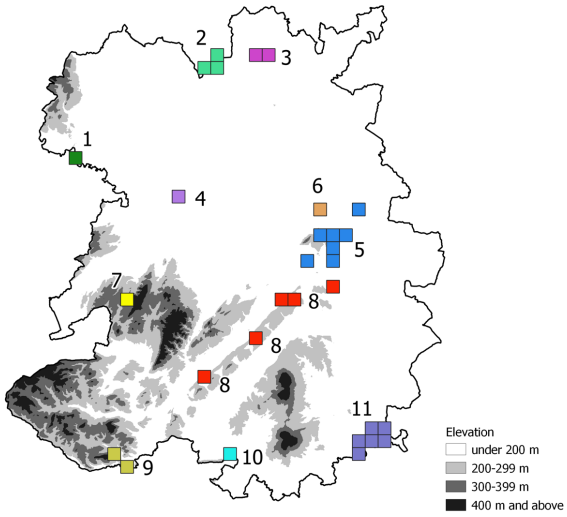


FIGURE 2. Map of VC40 Shropshire showing locations mentioned in the text.
1 Llanymynech 2 Shropshire Mosses (Whixall, Bettisfield, Wem) 3 Prees Heath 4 FSC Preston Montford 5 Telford post-industrial sites 6 Apley Castle 7 Pennerley 8 Wenlock Edge 9 Bucknell Wood 10 Ludlow 11 Wyre Forest

TWINSPAN

TWINSPAN (Two-Way INdicator SPecies ANalysis) is a computer program designed to compare samples on the basis of their component taxa and plots simultaneously (Hill, 1979; Hill & Šmilauer, 2005). A classification technique originally developed for analysing quadrat samples of vegetation, its use has been extended to compare larger vegetation units. It has been used less in entomological studies, mainly on data from structured programmes (e.g. Eyre *et al*., 1989; Hutcheson, 1990; Carter *et al*., 1996).

TWINSPAN is a divisive, polythetic method that aims to maximize the information in a data set to generate a satisfactory hierarchical classification of samples. In the current investigation samples are sites, i.e. tetrad samples. Using reiteration of RA or DCA ordination (the version of TWINSPAN employed in this study uses DCA ordination) the program first constructs a classification of a complete set of samples from the raw data, by identifying the strongest numerical trend in the sample set and dividing the samples into two groups at the centroid position on the trend. It also identifies, in approximate order of effectiveness, the best species to differentiate between the two groups: these are 'indicator species'. Other species that are at least twice as likely to occur on one or other side of the division are identified as 'preferential species'.

For each division the program calculates an indicator threshold, which is used to place samples on the positive or negative side of the division; this is done by comparing a sample's indicator score (to which each positive indicator species contributes +1 and each negative indicator species contributes -1) to the indicator threshold. Samples with an indicator score equal to or greater than the indicator threshold are placed in the positive group, otherwise in the negative group. In a division some samples may be defined as 'borderline' or 'misclassified' where either the refined ordination or indicator ordination is indecisive. The strength of a division is expressed as an eigenvalue (λ), with a maximum possible value of 1, which indicates a completely different species composition of the samples in the two groups of the division. Good groups have eigenvalues greater than about 0.20-0.25, values that correspond approximately to a 50% difference in species composition (M.O. Hill,

pers. comm. per D. Curtis). Each resulting positive and negative group is then re-analysed and re-divided hierarchically in turn, down to the required level of the classification

TWINSPAN simultaneously uses the classification of samples to obtain a classification of species. The two classifications are then used together to obtain an ordered two-way table, in which species are placed in groups based on the samples in which they occur, and showing the constancy (i.e. percentage frequency) with which they occur in those groups. When interpreted, some groups may express community and synecological relationships (Hill & Šmilauer, 2005; Curtis, 2010).

A TWINSPAN output of the classification of samples may be presented as a dendrogram showing the negative group on the left and positive group on the right, along with the number of samples and list of indicator species for each side of the division. The output of the species classification is an ordered table with species classes in rows and TWINSPAN classes (end-groups) in columns.

A TWINSPAN output may be interpreted on the basis of the species composition of classes in the ordered table, and of the location of species and samples within each division of the sample classification as shown in the dendrogram. Where no indicator species are identified for one side of a division, preferential species may be helpful. A division may invite an ecological interpretation, but other factors may be involved.

TWINSPAN has been selected as the analytical tool for this study as its methodology and outputs are more accessible to those amateurs whose data form the basis of this investigation than are the highly complex statistical tools, including ordination methods, such as those contained within the R statistical environment; it is acknowledged, however, that one or other of these methods may be preferred by some scientists for the analysis of such data.

METHODS

To prepare the Shropshire microlepidoptera database for TWINSPAN analysis 531 records identified only to species aggregates and 35 records located only to hectads were removed. The remaining records were entered into an EXCEL file with each row representing a species and each column a tetrad. The presence or absence of a species in a tetrad was shown in binary form: this constitutes a single sample in the analysis. For uploading to TWINSPAN the data were held in Cornell Condensed format created using CANOCO 4.5 program WCanoImp (Ter Braak & Šmilauer, 2002).

In classification and ordination methods it is advisable to remove species that are rare in samples as they may distort the analysis (Ezcurra, 1987; Legendre & Legendre, 1998). In our study a minimum number of species per tetrad had to be established experimentally, as that used by Trueman (2015) for analysis of the Shropshire flora (10 species per tetrad after removing many essentially ubiquitous species) was considered too severe for microlepidoptera, whose species are typically less numerous than plant species in samples, and very few may be classed as "essentially ubiquitous" in Shropshire.

An initial test made on the file with a minimum number of five species per tetrad did not sufficiently control small sample size bias and clearly separated the data by recording method used (i.e. daytime and night-time methods). The data were therefore separated into two files, a Day file representing daytime sampling (by searching, sweep-netting, beating vegetation) and a Night file representing night-time sampling (by light-trapping). Each file was reduced again by applying a minimum number of 10 species per tetrad, but the resulting numbers of species (176 in the Day file and 131 in the Night file) were considered still rather too large for an effective analysis. A third reduction, to 12 species per tetrad, was therefore applied; this produced a Day file with 129 species and 157 tetrads and a Night file with 113 species and 68 tetrads: these files form the basis of the analysis presented in this paper. Scientific nomenclature follows Stace (2010) for vascular plants and Agassiz *et al*. (2013) for Lepidoptera. Comments on the distribution and ecology of species in Shropshire follow Sinker *et al*. (1985) and Lockton & Whild (2015) for vascular plants, and Blunt (2014) for microlepidoptera.

RESULTS

The following sections separately describe and analyse the TWINSPAN outputs for the Day and Night files, up to the third level of classification. At this level most samples in the Night file output and some in the Day file output were considered too small for meaningful further analysis, so both outputs were terminated there.

In each section the TWINSPAN output is shown as a dendrogram (Figures 3 and 5) and as an ordered two-way table in summarized form (Tables 1 and 3). The ordered tables show the number of samples in each end-group, and species constancy is expressed as a percentage of samples in which each species occurs in each end-group. The dendrograms give the conventional TWINSPAN numbering for each group (in bold type), the number of samples in each group (in normal type), and the indicator species (in italics), eigenvalue (λ) and indicator threshold (both in red), for each division. End-groups are shown using the same alphabetical descriptors in block capitals in the dendrograms and ordered tables, allowing the relationship between the two to be seen. In each section relief maps of VC40 Shropshire (Figures 4 and 6) show the distribution of samples for selected divisions. Negative samples are shown in blue and positive samples in orange. The maps in Figures 4 and 6 exclude borderline and misclassified samples (as defined by the program) to give a better definition of groups in the divisions.

Appendix A gives the larval foods (mostly vascular plants) of all indicator species in each output, and shows the foods utilized by those species in Shropshire (where known), along with any other main foods recorded for those species elsewhere in Britain (after Langmaid *et al*., 2018). Appendices B and C give the expanded versions of the two-way tables that are shown in abbreviated form in Tables 1 and 3 respectively.

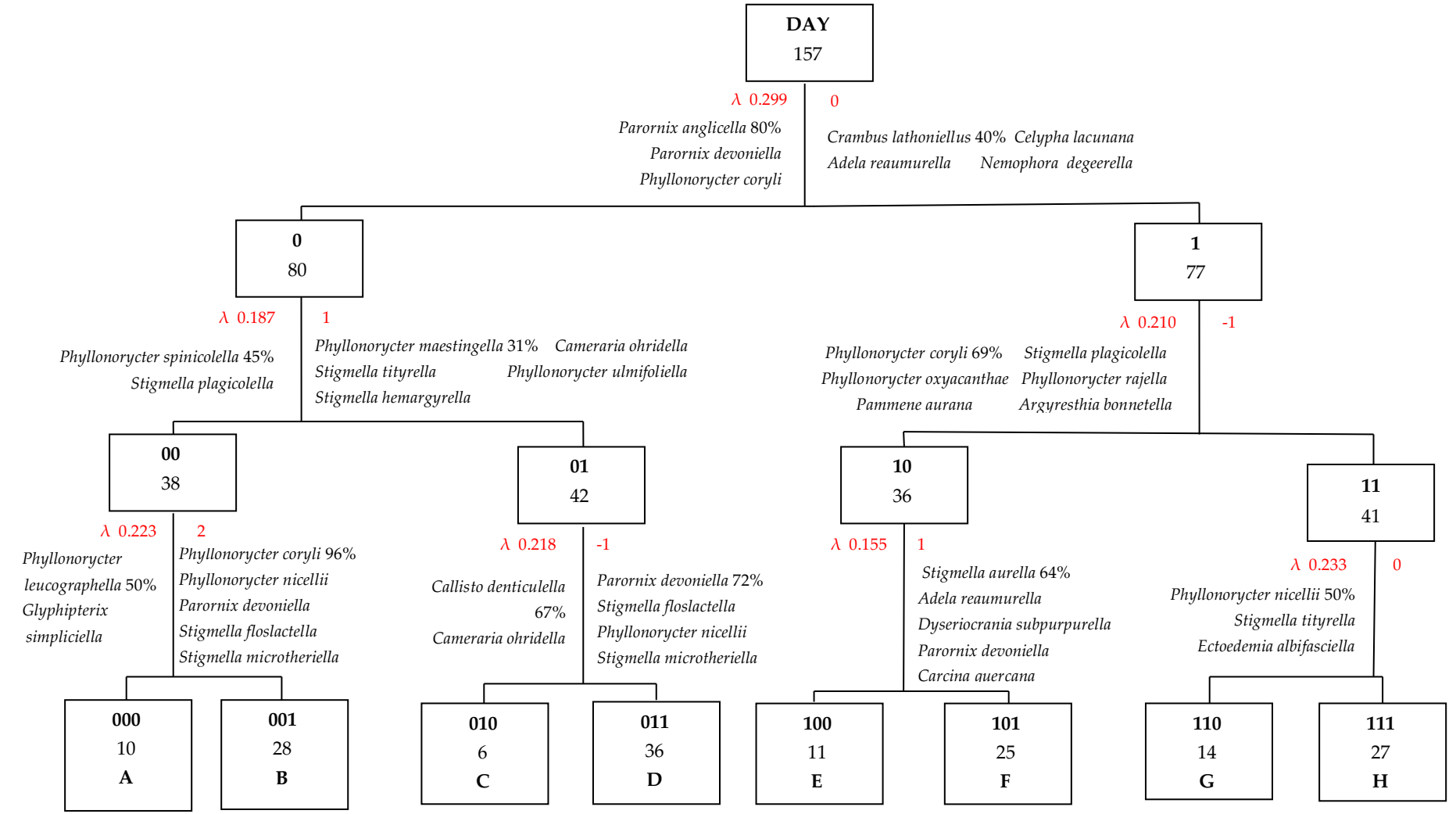
Outputs are interpreted on the basis of the distribution of species in the ordered tables and the distribution of species and samples in the TWINSPAN groups in the dendrograms.

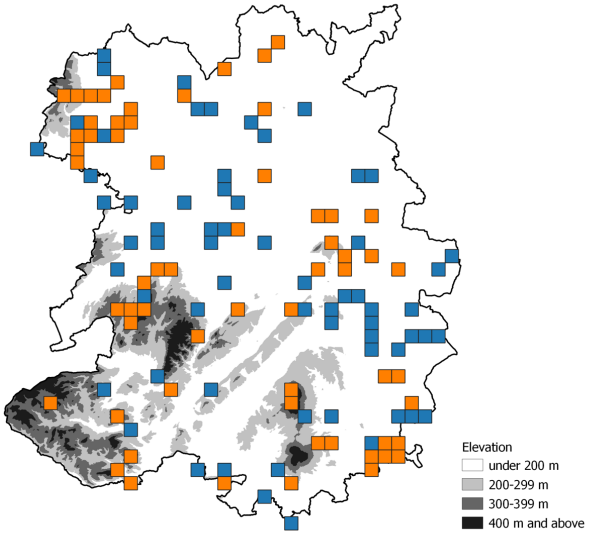
ANALYSIS OF THE DAY FILE

Figure 3 shows the dendrogram of the Day file output. The eigenvalue for the primary division (0.299) is quite strong, as are eigenvalues for the divisions of Day Groups (DG) 1, 00, 01 and 11, suggesting that each division accounts for around or above 50% of variation in the data analysed within it. The lower eigenvalues for divisions DG 0 and 10 show that a smaller amount of variation in the data is explained by these divisions. Figure 4 maps the distribution of samples for selected divisions in the Day file output.

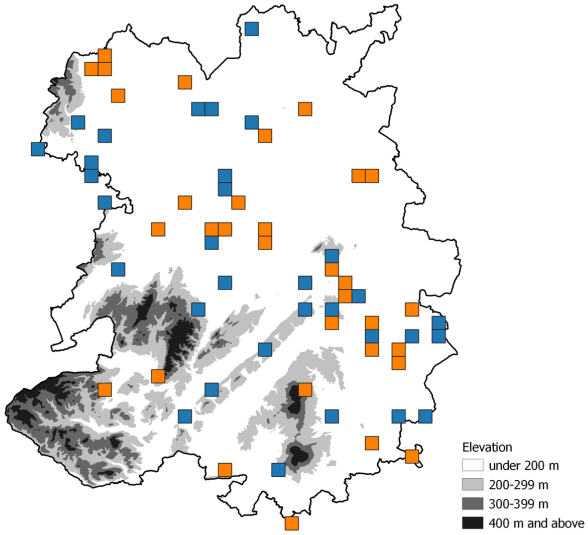
Figure 3 implies that seasonality is a main factor influencing the primary division: DG 0 contains species that have been recorded mostly from mid-July to October, while DG 1 species have been recorded mostly from April to early July. This seasonality is partly a natural one, as adults of some moths (e.g. *Adela reaumurella*, *Nemophora degeerella*) fly only in spring and their early stages are seldom recorded; but some other indicator species on both sides of the division may potentially be recorded in both time-frames as either adults or larvae. Also appearing to influence this division is a habitat management factor, resulting from many samples having been made in managed roadside habitats. The mechanical cutting of verges in Shropshire mostly eliminates from samples after early July moths whose larvae feed on herbaceous plants, whereas hedgerow cutting has a lesser (though not negligible) effect on populations of moths that feed on roadside shrubs (Blunt, 2014). The indicator species, plus all six other preferential species allocated by the program to DG 0 (*Leucoptera malifoliella, Phyllonorycter schreberella, Stigmella anomalella, S. floslactella, S, lemniscella, S, plagicolella*), feed on trees or shrubs; while two indicator species (*Celypha lacunana*, *Crambus lathoniellus*) and 16 of 23 other preferential species allocated to DG 1 feed on herbaceous plants. Figure 4a maps the samples for this division and suggests that altitude is also a factor: all but three DG 0 samples lie at or below 200m, whereas 15 DG 1 samples lie above that altitude. In summary, the primary division of the Day file separates DG 0 species, which feed on hedgerow trees and shrubs and are recorded mostly after early July at lower altitudes, from DG 1 species, which feed mainly on herbaceous plants and are recorded mostly before mid-July at both higher and lower altitudes.

FIGURE 3. Dendrogram of output of the TWINSPAN analysis of the Day file. Groups are shown in bold type, number of samples in ordinary type. Red figures show the eigenvalue (λ) on the left and threshold score on the right for the division of the group in the box above. Indicator species as identified by the program are shown for each division, with negative indicator species on the left and positive on the right. Percentages are the percentage of samples in which occur the most constant indicator species identified on each side of each division. Letters in bold capitals show end-groups

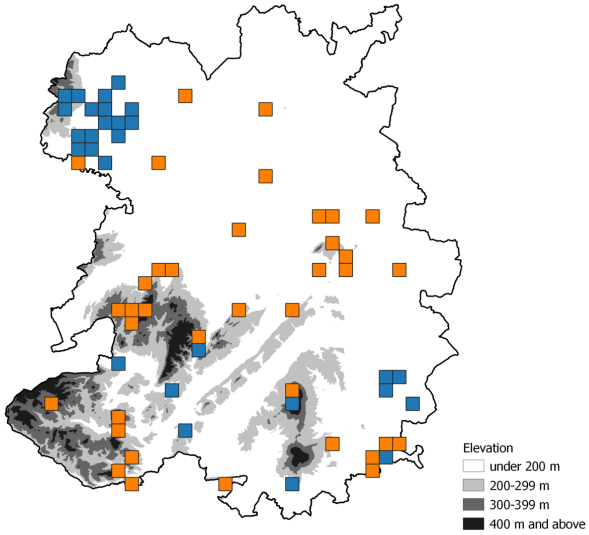




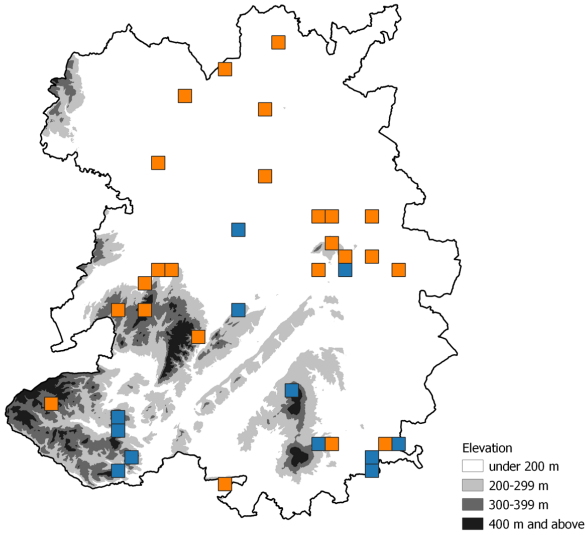
4a



4b



4c



4d

FIGURE 4. Maps showing the distribution of samples in selected divisions of the TWINSPAN output of the Day file. Negative samples are shown in blue, positive samples in orange. Borderline and misclassified samples are omitted.
Fig. 4a: primary division. Fig. 4b: division of DG 0. Fig. 4c: division of DG 1 Fig. 4d: division of DG 11.

In the division of DG 0 the two negative indicator species have blackthorn *Prunus spinosa* and wild plum *P. domestica* as their host plants in Shropshire, while three of the positive indicators feed on beech *Fagus sylvatica*, one on horse chestnut *Aesculus hippocastanum* and one on birches *Betula* spp. (Appendix A). The file output gives two further preferential species for DG 00: *Parornix torquillella*, whose larval food-plants are also blackthorn and wild plum, and *Phyllonorycter acerifoliella*, which feeds on field maple *Acer campestre*. Figure 4b shows that samples for DG 00 are lowland samples, up to around 200m altitude; those for DG 01 are mostly lowland but include two above 250m in the southern hills. Habitats represented by DG 00 samples are hedgerows and

woodland margins on basic soils, including the ancient semi-natural limestone woodlands of Wenlock Edge (Figures 2, 4b) and potential remnants of ancient woodlands in the south-east of the vice-county (Sinker *et al*., 1985; Blunt, *loc. cit*.); those represented by DG 01 samples are villages, parks, restored or regenerating post-industrial sites, and plantation woodlands, including some on acidic soils. Field maple and blackthorn are characteristic species of the W8 *Fraxinus excelsior-Acer campestre-Mercurialis perennis* woodland community in the National Vegetation Classification (Rodwell, 1991), and wild plum is usually confined in Shropshire to W8 vegetation, or to W21 *Crataegus monogyna-Hedera helix* scrub that is frequently a successional stage towards W8 woodland (Lockton & Whild, 2015). This division, therefore, appears to distinguish DG 00 samples that represent habitats on basic soils, some potentially ancient, from DG 01 samples that represent more disturbed and developed sites where beech (probably not native to Shropshire: Sinker *et al*., *loc. cit*; Lockton & Whild, *loc. cit*.) and horse chestnut have been planted, and birch grows as both a planted tree and a pioneer of natural regeneration.

Further divisions on the negative side of the dendrogram amplify this interpretation. In the division of DG 00 the positive indicator species constitute all five regularly-occurring feeders on hazel *Corylus avellana* in Shropshire (where W8 woodland is hazel's main habitat), along with a further seven preferential species, five of which (*Gracillaria syringella, Phyllonorycter acerifoliella, P. tristrigella, Stigmella anomalella, Parornix finitimella*) feed on other trees and shrubs typical of W8 woodland: blackthorn, wild plum, ash *Fraxinus excelsior*, elm *Ulmus* spp. (mainly *U. glabra*), and roses *Rosa canina, R. arvensis* (Langmaid *et al*., 2018; Lockton & Whild, *loc. cit*.). In this division separating DG 000 from DG 001 the food-plants of the negative indicator species are more associated with anthropogenic habitats: chiefly cock's-foot *Dactylis glomerata* for *Glyphipterix simpliciella*, and firethorn *Pyracantha* spp., hawthorn *Crataegus monogyna*, and occasionally other Rosaceae (including cultivated apple *Malus pumila* and Swedish whitebeam *Sorbus intermedia* agg.) for *Phyllonorycter leucographella* (Appendix A). In the division of DG 01 into DG 010 and DG 011, the two negative indicator species are associated with horse chestnut in one case and apple, including cultivated apple, in the other, whereas the four positive indicator species feed on hazel.

On the positive (right) side of the dendrogram DG 1 is divided into DG 10 and 11. This division has six negative indicator species and no positive ones, but the program identifies five preferential species for DG 11: *Crambus lathoniellus, C. pascuella, Cydia ulicetana, Micropterix aureatella, Scoparia ambigualis*. These species form a reasonably coherent assemblage associated in Shropshire particularly with areas of unimproved acid grassland, often where one or both gorses *Ulex europaeus* and *U. gallii* grow; their distributions include the southern hills, rides in old oakwoods such as the Wyre Forest, post-industrial sites in Telford, and the Shropshire Mosses, with an extension by some but not all species to abandoned limestone quarries at Llanymynech and Wenlock Edge (Blunt, *loc. cit*.). The Shropshire Mosses do not feature in samples for this division, perhaps because recent management has mostly eliminated gorses there; this ecology is otherwise quite well reflected by the tetrads mapped for DG 11 samples (Figure 4c). It contrasts with that of the six negative indicator species for this division, any one of which would place a sample in DG 10; these species are associated with host plants typical of basic and more nutrient-rich soils: hawthorn, hazel, blackthorn, wild plum, hogweed *Heracleum sphondylium* and alder *Alnus glutinosa* (Appendix A; Lockton & Whild, *loc. cit*.). In Shropshire these DG 10 species have a predominantly lowland distribution, though some do extend occasionally to higher altitudes.

The division of DG 11 has no positive indicator species, but DG 111 has ten preferential species (*Agriphila geniculea, Alucita hexadactyla, Argyresthia brockeella, Cauchas rufimitrella, Chrysoteuchia culmella, Coleophora serratella, Coptotriche marginea, Crambus perlella, Pyrausta aurata, Scoparia ambigualis*) that are associated with more open and little-managed sites with herbaceous plants, birches, alder and honeysuckle *Lonicera periclymenum* as larval host plants (Blunt, *loc cit*., Langmaid *et al*., *loc. cit.*). Eight samples for this group represent restored or regenerating post-industrial sites such as quarries and mining spoil-heaps, and four more have a wetland habitat element. The three negative indicator species, any one of which would place a sample in DG 110, feed separately on hazel, oaks *Quercus* spp., and beech (Appendix A). Fifteen other microlepidoptera species are identified as preferentials for this group, of which four feed on oaks, three on hazel, two on beech, two on ash, and one each on birch and hawthorn. The habitats represented by DG 110 samples are typically more mature woodland and

woodland margins, including areas planted with beech. Samples for both groups are located mainly on acidic soils but include some on neutral or basic soils, and both groups occur across a range of altitudes (Figure 4d).

The division of DG 10 has quite a low eigenvalue, implying that the distinction between the resulting groups is fairly weak. There are no negative indicator species, but DG 100 has two preferentials (*Ectoedemia subbimaculella*, *Pammene regiana*). The division has five positive indicator species, any one of which would allocate a sample to DG 101, which has 49 preferential species. Tetrads on both sides of this division have data for two or more monads, typically from multiple recording visits (Figure 1). DG 100 samples are all in the north-west of Shropshire, its most intensively recorded area. The division may therefore mainly reflect recording intensity.

Table 1 shows the two-way ordered table (in abbreviated form) obtained from the Day file output. The full table is in Appendix B.

Table 1. Ordered two-way table (abbreviated) obtained from the Day file output. Letters A-H in bold indicate the TWINSPAN classes (end-groups) as shown in Figure 3. Numbers are percentages of the number of samples in which each species occurs in each end-group. The full table is in Appendix B.

Species	**A** *n=10*	**B** *n=28*	**C** *n=6*	**D** *n=36*	**E** *n=11*	**F** *n=25*	**G** *n=14*	**H** *n=27*	**Species class**
Phyllonorycter tristrigella		29	33	39	27	20		15	*a*
5 more species									*a*
Caloptilia rufipennella		11		25		8			*a*
Stigmella plagicolella	60	68	33	17	45	36	7		*b*
16 more species									*b*
Leucoptera malifoliella	40	36	50	17	9	20		4	*b*
Tischeria ekebladella	10	14	33	36	9	40	36		*c*
5 more species									*c*
Acrolepia autumnitella	10	4	50	8	9	20		4	*c*
Stigmella oxyacanthella	50	7		17	18	44	7		*d*
8 more species									*d*
Anthophila fabriciana	70	89	83	75	73	84	79	67	*d*
Udea lutealis		32	17	11	9	56	21	33	*e*
7 more species									*e*
Agonopterix arenella		4		11		28	14	7	*e*
Syndemis musculana	10	4		3	9	44	7	15	*f*
13 more species									*f*
Agonopterix heracliana		14	17			36		7	*f*
Scoparia ambigualis				8		16	14	30	*g*
26 more species									*g*
Adela reaumurella		4		11	9	60	50	63	*g*
Phyllonorycter harrisella		4			9	24	29	7	*h*
4 more species									*h*
Crambus lathoniellus		4		3		36	57	52	*h*

This table clearly separates moths whose larvae feed on trees and shrubs, which constitute all but two species in Classes *a* to *d*, from those which feed on herbaceous plants, which constitute 66% of species in Classes *e* to *h*. It is apparent that the latter four classes are much less associated with end-groups A to D than the former classes are with end-groups E to H. This probably reflects the many samples in the file that represent roadside localities, where regular management differentially affects the fauna of woody plants and herbaceous plants, as discussed above. Species in Class *h*, which are typically found in Shropshire in drier, little-managed grassland with gorse and oak scrub on acid soils (Blunt, *loc. cit.*), are associated almost completely with end-groups F, G and H.

At the other end of the table, species in Class *b* feed as larvae mostly on roadside shrubs characteristic of W8 or W21 vegetation communities of basic soils: hazel, hawthorn, blackthorn, wild plum, field maple, roses and apple, with one species (*Stigmella aurella*) associated with bramble *Rubus fruticosus* agg.. Classes *a*, *c* and *d* contain

moths that feed on forest trees, though none of these classes is entirely coherent in this respect, and none shows a strong affinity with any group of samples. There is some difference in the trees represented by the more monophagous or near-monophagous moths in these classes: those feeding on elms, sweet chestnut *Castanea sativa* or sycamore *Acer pseudoplatanus* appear in Class *a*, those on beech or horse chestnut in Class *c*, and those on alder or sallow *Salix* spp. in Class *d*.

Except for Class *h* discussed above, classes that are associated more strongly with end-groups E to H are hard to interpret as coherent assemblages of species in terms of their larval food-plants or known ecologies in Shropshire (Langmaid *et al*., 2018; Blunt, *loc.cit*.). Nearly all species that are more associated with end-group E than F are in Class *g* (*Pammene aurana, Gypsonoma dealbana, Mompha epilobiella, Ectoedemia subbimaculella, E. intimella, Depressaria radiella*) and may suggest that end-group E represents rather moister or more nutrient-rich habitats than F; but the difference between these end-groups and species classes is not strong.

In summary, Table 1 clusters together some assemblages of species on the basis of their food-plant preferences, and thus reveals some similarities in terms of the ecology of these assemblages. There is a clear trend across end-groups from tree- and shrub-feeding cohorts to those associated with grasses, forbs and woody plants in more open habitats. Moths more associated with host plants of damper habitats tend to occur towards the middle of this trend. Except for Class *h* species, assemblages are not strongly aligned with any particular end-groups. The influence of habitat management is perhaps an important factor in the table.

ANALYSIS OF THE NIGHT FILE

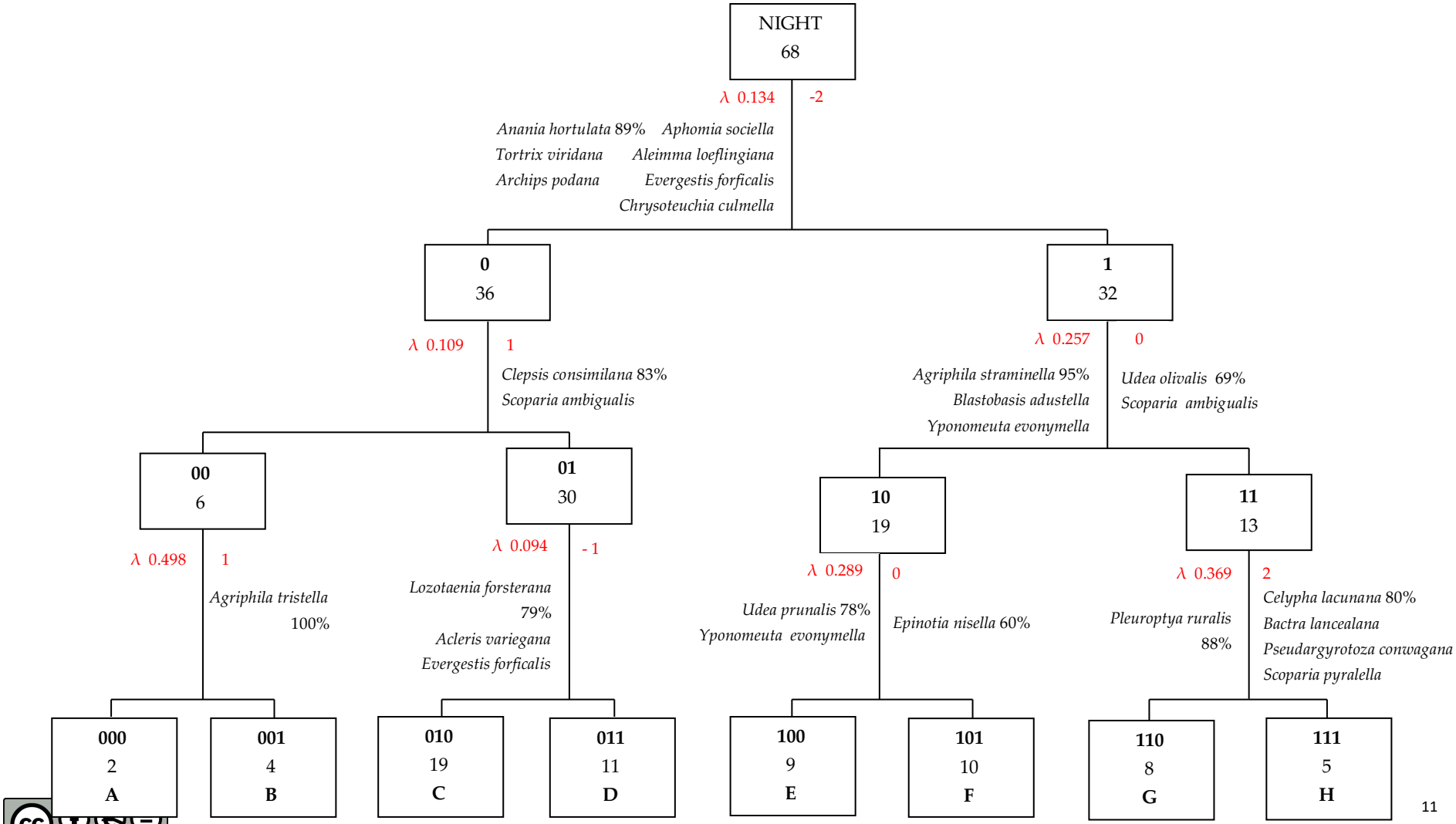
Figure 5 shows the dendrogram of the Night file output. There are low eigenvalues for the first division of the analysis and for the divisions of Night Groups (NG) 0 and 01, showing that only a relatively small amount of variation in the data is accounted for by each of these divisions. Eigenvalues for the division of NG 1 and further divisions on that side of the dendrogram are stronger, suggesting a somewhat better differentiation between samples in these divisions. Stronger eigenvalues for later divisions on both sides of the dendrogram perhaps reflect the influence of small sample sizes more than any other factor. The distribution of samples in selected divisions of this output is mapped in Figure 6.

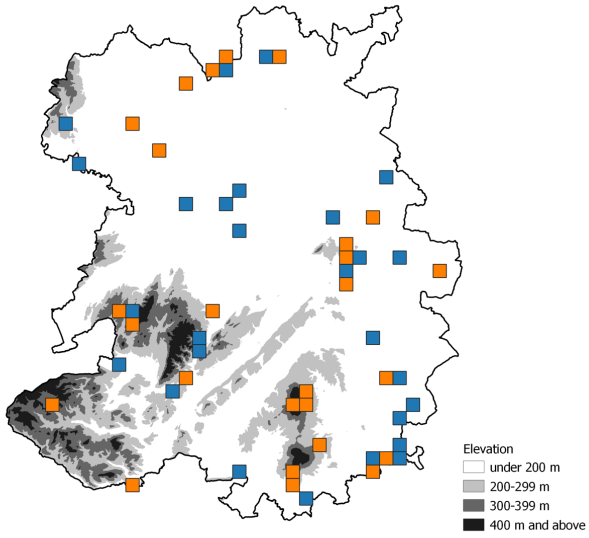
The major factor influencing the primary division (Figure 5) can be interpreted as the number of species per sample: the program separates samples with more species (all in NG 0), obtained from many trapping sessions, from those with fewer (all in NG 1), from one or a few trapping sessions. Figures for these groups, after removal of borderline and misclassified samples, are: NG 0, mean species per sample 66.26, range 25-103, n=27; NG 1, mean species per sample 20.74, range 12-40, n=27. Median values are close to mean values for both groups. If borderline and misclassified samples are included, the mean for NG 0 samples falls to 55.94 (n=36); that for NG 1 remains at 20.72 (n=32). NG 0 sites include all 13 gardens in Shropshire where extensive light-trapping has taken place over several years, plus well-worked rural sites; these latter include the Shropshire Mosses, Prees Heath and Wyre Forest, and the placement of tetrad samples from these locations clearly shows the influence of sample size: tetrads with more species are in NG 0; those with fewer are in NG 1 (Table 2).

Table 2. Comparison of samples from the Shropshire Mosses, Prees Heath and the Wyre Forest that are placed in different groups in the primary division of the TWINSPAN output of the Night file.

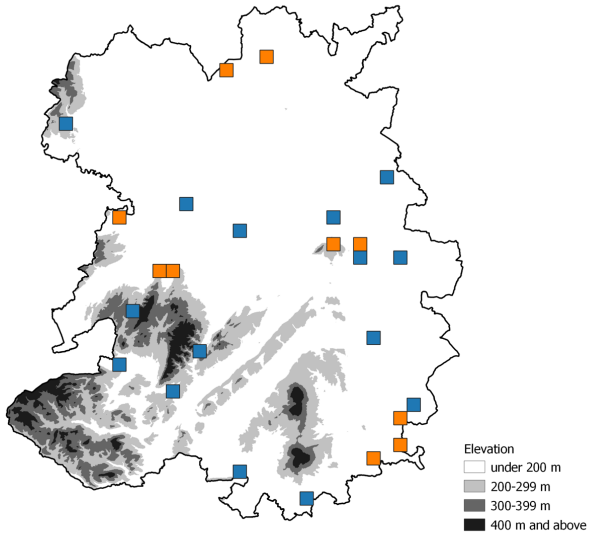
Night Group 0			Night Group 1		
Tetrad	Location	No. species	Tetrad	Location	No. species
SJ43X	Shropshire Mosses	79	SJ43S	Shropshire Mosses	27
			SJ43Y	Shropshire Mosses	40
SJ53N	Prees Heath	67	SJ53T	Prees Heath	24
SO77D	Wyre Forest	80	SO77C	Wyre Forest	20
SO77N	Wyre Forest	103	SO77I	Wyre Forest	32
SO77P	Wyre Forest	78			

FIGURE 5. Dendrogram of output of the TWINSPAN analysis of the Night file. Groups are shown in bold type, number of samples in ordinary type. Red figures show the eigenvalue (λ) on the left and threshold score on the right for the division of the group in the box above. Indicator species as identified by the program are shown for each division, with negative indicator species on the left and positive on the right. Percentages are the percentage of samples in which occur the most constant indicator species identified on each side of each division. Letters in bold capitals show end-groups.

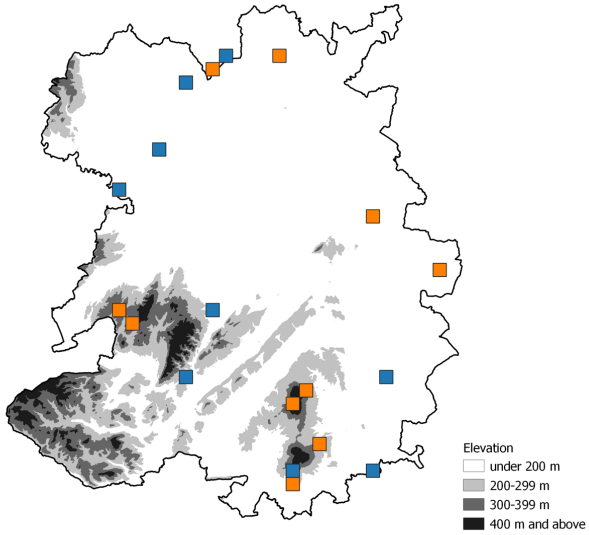




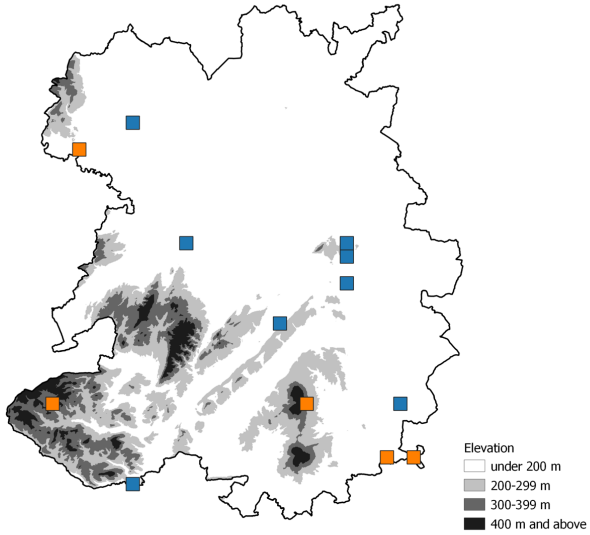
6a



6b



6c



6d

FIGURE 6. Maps showing the distribution of samples in selected divisions of the TWINSPAN output of the Night file. Negative samples are shown in blue, positive samples in orange. Borderline and misclassified samples are omitted. Fig. 6a: primary division. Fig. 6b: division of NG 01. Fig. 6c: division of NG 10. Fig, 6d: division of NG 11.

The primary division has no positive indicator species but has seven negative indicator species, any two (or more) of which would place a sample in NG 0. Species associated with NG 0 do not form a coherent assemblage in respect of their larval food requirements (Appendix A), but their Shropshire distributions are mostly at lower altitudes (Figure 6a). The output suggests that these species tend to occur less within smaller samples, and enter samples later as the species complement of those increases.

In the division of NG 01 into NG 010 and 011 the 13 intensively-trapped gardens, in widespread locations and sampling a range of habitats, constitute NG 010, together with sites at FSC Preston Montford and Apley Castle which have some characteristics of rural gardens. NG 011 samples are non-garden habitats: woodland, heathland, farmland, and two old post-industrial sites. Sample size difference is less marked in this division: NG 010 samples have a mean of 70.07 species and NG 011 samples a mean of 52.0; moreover, the Wyre Forest, Shropshire Mosses and Prees Heath samples allocated to NG 011 are some of the more species-rich samples from these localities (tetrads SJ43X, SJ53N, SO77D, SO77P: see Table 2). NG 010 samples include some at or above 200m (Figure 6b). The eigenvalue for the division (0.094) is very low and any differentiation between the two groups in terms of habitat is not clear-cut. The division implies at best a small difference between extensive moth faunas of gardens and non-garden habitats, that difference being represented particularly by the negative indicator species *Lozotaenia forsterana*, *Acleris variegana* and *Evergestis forficalis*; altitude may perhaps also influence this division to a small degree.

The division of NG 1 into NG 10 and 11 suggests a trend towards differentiating samples in damp and dry habitats at various altitudes in NG 10 from those representing drier lowland habitats in NG 11; the trend is, however, rather weak, and the subsequent divisions of NG 10 and 11 offer somewhat better definitions. Figure 6c maps the division of NG 10 into NG 100 and 101 and suggests an influence of altitude: all samples in NG 100 lie below 220m, whereas NG 101 represents all samples above this altitude, together with some lowland samples. Both groups represent a mix of damp and dry, acidic and basic habitats. The indicator species for this division show altitudinal differences in their Shropshire distributions: the negative indicator species *Udea prunalis* and *Yponomeuta evonymella* occur principally in lowland Shropshire, while the positive indicator species *Epinotia nisella* is found at both low and high altitudes (Blunt, *loc. cit.*).

The division of NG 11 into NG 110 and 111 appears to distinguish rather more disturbed habitats (NG 110) from rather less disturbed ones (NG 111). The placement in NG 110 of Bucknell Wood, a mixed woodland with old native oak stands and conifer plantations, seems anomalous, but may be at least partly explained by the presence of ruderal vegetation and plantation woodland at the trapping site. A small altitudinal influence may possibly also be present, as the two samples at higher altitudes in this division feature in NG 111 (Figure 6d).

Table 3 is the ordered two-way table of the output from the Night file, in abbreviated form: the full table is in Appendix C. This table suggests that most species in the Night file analysis have poor differential potential, as they show no great affinity for any particular group of samples. End-groups C and D contain all species in the table. Based on the interpretation of the division of NG 01 above, Table 3 may imply that a few species (e.g. *Eudonia angustea, Lozotaenia forsterana, Nomophila noctuella*), occur proportionately more in garden habitats (end-group C) than non-garden ones (end-group D).

Species in Class *a* are clearly associated with the negative side of the TWINSPAN output that was interpreted above as representing more extensive light-trapping programmes. Class *a* species do not appear to form a coherent group either in terms of habitat, as implied by their larval food-plant preferences (Langmaid *et al.*, 2018) or of their known distributions and ecology in Shropshire (Blunt, *loc. cit.*); and some, notably *Tortrix viridana*, can form large local populations. Class *a* may perhaps be interpreted as consisting of species that are less attracted to light, so that they are generally absent from less intensive light-trapping programmes, and arrive in traps only when these are run with greater frequency over time.

The one species assemblage in this table that is reasonably coherent in ecological terms is Class *g*. This comprises moths that are associated in Shropshire mainly with heathy woods, unimproved grassland and old post-industrial sites, but are seldom found in gardens (Blunt, *loc. cit.*). The table implies that these species are readily recorded by lower intensity light-trapping in older habitats, including those represented by end-group D samples, but are weakly attracted to light-trapping in other, more disturbed habitats including gardens (end-group C), even when trapping is at higher intensity. This class may, in fact, say more about the relatively lower mobility of species of older or less disturbed habitats than about the specific nature of those habitats.

In summary, this interpretation of the Night file output suggests that the number of species in samples is much the most important influence. As samples accrue more species they tend towards convergence in their species complements, and divisions generally show low levels of differentiation. Smaller groups are better

differentiated in later divisions, with some effects, not particularly strong, of altitude and habitat being suggested. In general, this analysis of light-trapping data offers rather little by way of ecological interpretation; it does, however, provide insights into the nature of those data, and of the potential mobility of microlepidoptera species as reflected in light-trapping programmes.

Table 3. Ordered two-way table (abbreviated) obtained from the Night file output. Letters A-H in bold indicate the TWINSPAN classes (end-groups) as shown in Figure 5. Numbers are percentages of the number of samples in which each species occurs in each end-group. The full table is in Appendix C.

Species	A *n=2*	B *n=4*	C *n=19*	D *n=11*	E *n=9*	F *n=10*	G *n=8*	H *n=5*	Species class
Aleimma loeflingiana			74	82					*a*
6 more species									*a*
Tortrix viridana	50	50	68	73			13		*a*
Acleris variegana		75	84	18		20	13		*b*
15 more species									*b*
Ypsolopha scabrella			63	36	11	10			*b*
Acleris rhombana			53	9		20		20	*c*
35 more species									*c*
Yponomeuta evonymella		25	89	64	89	20			*c*
Aethes cnicana			37	36		10		20	*d*
3 more species									*d*
Spilonota ocellana			68	73	33	10			*d*
Apotomis betuletana			37	55	22	30	13		*e*
3 more species									*e*
Rhopobota naevana			26	45	22	10	13	20	*e*
Acentria ephemerella		25	26	45	33	10	13	20	*f*
14 more species									*f*
Udea olivalis	100	50	100	55	22		63	80	*f*
Agriphila inquinatella			32	27	22	30		20	*g*
5 more species									*g*
Ypsolopha parenthesella			21	55	22	40		40	*g*
Agriphila straminella	50	100	89	91	89	100	25	40	*h*
3 more species									*h*
Pleuroptya ruralis	50	100	95	91	100	50	88	20	*h*

DISCUSSION

The Shropshire microlepidoptera database used in this study, consisting of opportunistic records gathered by amateurs, exhibits the range of biases identified for such data by the literature. It may be considered reasonably typical of amateur databases for a wide range of taxa, save in one respect: moth records, including those used in this study, are normally gathered by two very different techniques, i.e. daytime methods such as searching, sweep-netting and beating vegetation, which record both adults and earlier stages (predominantly larvae) and capture much *in situ* breeding data; and night-time methods using static light-traps, which record mobile adults attracted to traps from variable but uncertain distances. After an exploratory test of the database it was decided to analyse separately the data gathered by these two techniques. The division of the database into a Day file and a Night file, accounting for around 40% and 60% of total records respectively, and the application of a minimum number of 12 species per sample, produced files of quite different character, in which only 16% of species and 21% of tetrads were in common, justifying this division of the database.

The level of sample size reduction, by removing rarer species to avoid the distorting effect of very small samples on outputs, was decided experimentally. A problem in doing so was that higher altitude tetrads in Shropshire, with typically fewer species compared to lowland tetrads, were disproportionately eliminated from

files by applying larger minimum sample sizes. Blunt (2014) hypothesized that climate related to altitude is an important influence on the Shropshire distributions of many microlepidoptera species; it was therefore felt desirable to retain enough samples at higher altitudes to allow this factor to be tested in the TWINSPAN outputs. Applying a minimum number of 12 species per sample therefore represents a cautious approach; in different studies and contexts a greater level of reduction might offer a clearer definition of groups in TWINSPAN outputs.

A main purpose of this study was to investigate the extent to which TWINSPAN analysis may throw light on the ecology of Shropshire microlepidoptera. The Day file output has produced some groups that are reasonably coherent in ecological terms. An influence of altitude is suggested by the primary division of the Day file and by the divisions of DG 0 and DG 1, though altitude is not the only factor influencing these divisions: the primary division appears also to be much influenced by seasonality and habitat management, and it differentiates between species that feed on trees and shrubs and those that feed on herbaceous plants. The species assemblages of DG 00 and 001 are largely associated with trees and shrubs of NVC W8 woodland or W21 scrub on basic soils, with some samples potentially representing relict and semi-natural woodland; these groups are separated from DG 01 and 000, whose component species are more associated with anthropogenic habitats. The positive side of the Day file output appears to distinguish chiefly between assemblages of species associated with unimproved or little-managed grassland, including on post-industrial sites, and those of woodland and scrub, including plantation woodland. The distribution of species and samples in the ordered table further reflects this interpretation of the primary division of the Day file output; it also throws additional light on the species complements of some groups that may be interpreted along ecological lines similar to those described for groups in the Day file dendrogram. The species classes in the ordered table may also point towards ecologically meaningful sets of species which could be helpful in synecological studies; and as a TWINSPAN output can also be expressed as a dichotomous key, the classification could be used to classify new samples from Shropshire for comparison with those in the database.

It is instructive that all but one indicator species identified by the program on the negative side of the Day file output, and many on the positive side, are monophagous or near-monophagous moths, i.e. each feeds on only one or very few closely-related vascular plant species. These moths are represented in the file almost entirely by larval records. This implies that such records, which indicate *in situ* breeding, offer a greater interpretive potential than records of adult moths; and it may suggest that larval records for other phytophagous taxa would have a similar potential in TWINSPAN analyses. The ecologies of larval food-plants and associated vegetation, derived from the literature, were used in this study as an important interpretive resource; for non-phytophagous invertebrate groups, interpretations based on the vegetation of sampling sites would be effective only if good botanical data were available for those sites.

The Night file analysis suggests that an altitudinal influence may be present in the division of NG 10, and to a lesser extent in that of NG 11. The output of this file also produces groups that imply some differences between the fauna of more and less disturbed habitats, and between garden and non-garden habitats. None of these differences is great, however, and all are subject to the main trend in the Night file analysis, that of sample size. It strongly appears that TWINSPAN has separated Night file samples with more species, resulting from greater trapping intensity, from those with fewer species, resulting from lower trapping intensity, and that the species compositions of samples tend towards greater similarity with more trapping, at least for the more widespread species that comprise the samples. This factor appears to override others, including habitat; and apparent influences of altitude and habitat in the Night file may in fact be products more of the amount of trapping carried out at different locations. In summary, the Night file output suggests that light-trapping has limited ability to offer ecological information about more widespread species. There are, however, implications for understanding moth distributions; for if this study shows that light-trapping captures much transience of species at trapping sites, then distribution maps and accounts based on light-trapping data risk giving over-optimistic pictures of breeding densities. The matter lies beyond the scope of this study, but invites further detailed investigation.

The extensive databases now in existence, compiled with great effort by enthusiastic amateurs, involve biases that restrict effective analyses. Complex statistical methods for dealing with those biases, such as explored

by Isaac *et al*. (2014), require specialist training in their use and interpretation. While the authors do not claim that TWINSPAN is necessarily the best analytical technique for such databases, it does provide objectively-derived sample and species groupings which may be used as entities in more rigorous multivariate statistical analyses, to explore, for example, how they may be affected by, or correlate with, sets of environmental factors. TWINSPAN also offers amateurs a greater measure of accessibility for understanding analyses of their data at more than headline levels; and in an age of citizen science, that is no negligible consideration.

ACKNOWLEDGEMENTS

The authors thank Professor David Curtis and Professor Rob Marrs for much valuable criticism, guidance and helpful suggestions and corrections for improving this paper. Prof. Curtis gave extensive help over TWINSPAN. Thanks are also due to Mike Shurmer for help with the microlepidoptera database, and to the many recorders whose assiduous fieldwork has accumulated the data used in this study.

REFERENCES

Agassiz, D.J.L., Beavan, S. D. & Heckford, R. J. (2013). *A checklist of the Lepidoptera of the British Isles*. Field Studies Council, Telford.

Blunt, G. (2014). *The smaller moths of Shropshire: their status, distribution and ecology*. Field Studies Council, Telford.

Carter, J.L., Fend, S.V. & Kennelly, S.S. (1996). The relationships among three habitat scales and stream benthic invertebrate community structure. *Freshwater Biology* **35**, 109-124.

Curtis D.J. (2010). Communities of opiliones. www.david.curtis.care4free.net/opilcomms3.htm (accessed 10 September 2020).

Eyre, M.D., Luff, M.L., Rushton, S.P. & Topping, C.J. (1989). Ground beetles and weevils (Carabidae and Curculionoidea) as indicators of grassland management practices. *Journal of Applied Entomology* **107**, 508-517.

Ezcurra, E. (1987). A comparison of reciprocal averaging and non-centred principal component analysis. *Vegetatio* **71**, 41-47.

Gould, P.J.L. & Woiwod, I.P. (2009). The Rothamsted Insect Survey light-trap network. *Biodiversity News* **46**, 12-13.

Hill, M.O. (1979). *TWINSPAN – a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes*. Cornell University, Ithaca, New York.

Hill, M.O. & Šmilauer, P. (2005). *TWINSPAN for Windows version 2.3*. Centre for Hydrology and Ecology, Swindon.

Hutcheson, J. (1990). Characterization of terrestrial insect communities using quantified, Malaise-trapped Coleoptera. *Ecological Entomology* **15**(2), 143-151.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P. & Roy, D.B. (2014). Extracting robust trends in species' distributions from unstructured opportunistic data: a comparison of methods. www.biorxiv.org/content/biorxiv/early/2014/07/10/006999.full.pdf (accessed 12 December 2018)

Isaac, N.J.B. & Pocock, M.J.O. (2015). Bias and information in biological records. *Biological Journal of the Linnaean Society* **115**(2), 522-531.

Langmaid, J.R., Palmer, S.M. & Young, M.R. (eds) (2018). *A field guide to the smaller moths of Great Britain and Ireland, 3rd edition*. The British Entomological and Natural History Society.

Legendre, P. & Legendre, L. (1998). *Numerical ecology*. Elsevier, Amsterdam.

Lockton, A. & Whild, S. (2015). *The flora and vegetation of Shropshire*. Shropshire Botanical Society, Shrewsbury.

Pocock, M.J.O., Roy, H.E., Preston, C.D. & Roy, D.B. (2015). The Biological Records Centre: a pioneer of citizen science. *Biological Journal of the Linnaean Society* **115**(2), 475-493.

Powney, G.B. & Isaac, N.J.B. (2015). Beyond maps: a review of the applications of biological records. *Biological Journal of the Linnaean Society* **115**(2), 532-542.

Rodwell, J.S. (ed.) (1991). *British plant communities, vol. 1: woodlands and scrub*. Cambridge University Press, Cambridge.

Sinker, C.A., Packham, J.R., Trueman, I.C., Oswald, P.H., Perring, F.H. & Prestwood, W.V. (1985). *Ecological flora of the Shropshire region*. Shropshire Trust for Nature Conservation, Shrewsbury.

Stace, C.A. (2010). *New flora of the British Isles, third edition.* Cambridge University Press, Cambridge.

Sterling, P., Parsons, M. & Lewington, R. (2012). *Field guide to the micro moths of Great Britain and Ireland*. British Wildlife Publishing, Dorset.

Ter Braak, C.J.. & Šmilauer, P. (2002). CANOCO 4.5, Software for canonical community ordination (version 4.5). Centre for Biometry, Wageningen, The Netherlands.

Trueman, I. (2015). Changes in the vascular plant flora of Shropshire. In Lockton, A. & Whild, S. (2015). *The flora and vegetation of Shropshire*, pp. 11-38. Shropshire Botanical Society, Shrewsbury.

Tulloch, A.I.T., Mustin, K., Possingham, H.P., Szabo, J. K. & Wilson, K.A. (2012). To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. *Diversity and Distributions* **19**, 465-480.

Väisänen, R., Heliövaara, K. & Immonen, A. (1991). Biogeography of northern European insects: province records in multivariate analyses (Saltatoria; Lepidoptera: Sesiidae; Coleoptera: Buprestidae, Cerambycidae). *Annales Zoologici Fennici* **28**(2), 57-81.

APPENDIX A: Larval foods of indicator species in the TWINSPAN outputs.

Species	Foods recorded in Shropshire ` (after Blunt, 2014) (**bold text** = main foods)	Other foods recorded in the UK (after Langmaid, Palmer & Young, 2018)
DAY FILE ANALYSIS		
Adela reaumurella	leaf litter	oak & birch leaf litter
Argyresthia bonnetella	*Crataegus monogyna*	
Callisto denticulella	***Malus sylvestris*, *M. pumila***	
Cameraria ohridella	***Aesculus hippocastanum***	occ. *Aesculus carnea, Acer pseudoplatanus, A. platanoides*
Carcina quercana	*Rubus fruticosus* agg., *Acer pseudoplatanus*	many trees & shrubs especially *Quercus* and *Fagus*
Celypha lacunana	***Urtica dioica, Epilobium angustifolium***, *E, hirsutum, Rumex obtusifolium, R. acetosa, Artemisia vulgaris, Rhinanthus minor, Teucrium scorodonia, Corylus avellana*	most species of herbaceous plants, occ. trees & shrubs
Crambus lathoniellus		Poaceae, esp. *Deschampsia cespitosa*
Dyseriocrania subpurpurella	***Quercus robur***, *Q. petraea*	occ. *Castanea*
Ectoedemia albifasciella	***Quercus robur***, *Q. petraea*	
Glyphipterix simpliciella		seeds of *Dactylis, Schedonurus arundinaceus, S. pratensis*
Nemophora degeerella		dead leaves
Pammene aurana		*Heracleum* seeds
Parornix anglicella	***Crataegus monogyna***, *C. laevigata*	*Sorbus torminalis*, occ. *S. aucuparia, Fragaria vesca*
Parornix devoniella	***Corylus avellana***	
Phyllonorycter coryli	***Corylus avellana***	
Phyllonorycter leucographella	***Crataegus monogyna, Pyracantha* sp.**, *Malus pumila, M. sylvestris, Prunus avium, Sorbus aucuparia, S. intermedia* agg.	*Pyrus, Prunus spinosa, Chaenomeles japonica, Cotoneaster frigidus*
Phyllonorycter maestingella	***Fagus sylvatica***	
Phyllonorycter nicellii	***Corylus avellana***	
Phyllonorycter oxyacanthae	***Crataegus monogyna***, *C. laevigata*	*Pyrus, Sorbus torminalis*
Phyllonorycter rajella	***Alnus glutinosa***, *A. cordata*	*Alnus incana*
Phyllonorycter spinicolella	***Prunus spinosa*, *P. domestica***	
Phyllonorycter ulmifoliella	***Betula pendula***, *B. utilis*	
Stigmella aurella	***Rubus fruticosus* agg.**, *R. idaeus, Geum urbanum*	*Fragaria, Agrimonia*
Stigmella floslactella	***Corylus avellana***	*Carpinus*
Stigmella hemargyrella	***Fagus sylvatica***	
Stigmella microtheriella	***Corylus avellana***, *Carpinus betulus*	*Ostrya carpinifolia*
Stigmella plagicolella	***Prunus spinosa, P. domestica***	
Stigmella tityrella	***Fagus sylvatica***	
NIGHT FILE ANALYSIS		
Acleris variegana	*Rosa arvensis, Rubus idaeus*	*Crataegus, Prunus, Malus, Pyrus, Poterium,* various other plants
Agriphila straminella		Poaceae, inc. *Festuca ovina*
Agriphila tristella		Poaceae, inc. *Deschampsia cespitosa, Poa*
Aleimma loeflingiana	*Quercus* sp.	*Acer, Carpinus*

Continued overleaf.

APPENDIX A continued:

Species	Foods recorded in Shropshire ` (after Blunt, 2014) (**bold text**= main foods)	Other foods recorded in the UK (after Langmaid, Palmer & Young, 2018)
NIGHT FILE ANALYSIS continued		
Anania hortulata		*Urtica dioica*, occ. *Marrubium, Ballota, Stachys, Mentha, Convolvulus arvensis*
Aphomia sociella	*Bombus hypnorum* nest	nests of other Hymenoptera species
Archips podana	*Acer palmatum*	polyphagous on trees & shrubs, dead insects
Bactra lancealana		*Juncus, Schoenoplectus lacustris, Cyperus longus, Trichophorum cespitosum, Eriophorum angustifolium*
Blastobasis adustella		diverse fresh & dry vegetable matter inc.galls, spun fruits
Celypha lacunana	***Urtica dioica, Epilobium angustifolium***, *E, hirsutum, Rumex obtusifolium, R. acetosa, Artemisia vulgaris, Rhinanthus minor, Teucrium scorodonia, Corylus avellana*	most species of herbaceous plants, occ. trees & shrubs
Chrysoteuchia culmella		various Poaceae
Clepsis consimilana		polyphagous on trees & bushes, esp. *Ligustrum*
Epinotia nisella	*Salix* sp. (sallow)	*Populus*
Evergestis forficalis		Brassicaceae inc. *Brassica, Raphanus, Armoracia, Alliaria, Sisymbrium, Crambe maritima, Sinapis arvensis*
Lozotaenia forsterana	*Hedera helix*	many trees & shrubs, herbaceous plants esp. *Ligustrum, Vaccinium*
Pleuroptya ruralis	***Urtica dioica***	*Ulmus procera, U. glabra, Atriplex, Chenopodium, Filipendula ulmaria, Humulus*
Pseudargyrotoza conwagana		*Ligustrum, Fraxinus*, occ. *Syringa*
Scoparia ambigualis		mosses inc.*Polytrichum commune;* also *Valeriana officinalis*
Scoparia pyralella		decaying plant material
Tortrix viridana	*Quercus* sp., *Salix* sp. (willow)	occ. other deciduous trees
Udea olivalis		many herbaceous plants inc. *Urtica dioica, Symphytum officinale, Stachys, Lamiastrum, Mercurialis perennis, Glechoma, Silene dioica, Humulus*
Udea prunalis	*Urtica dioica*	many other plants inc., *Ballota, Centaurea nigra, Lamium, Sambucus nigra, Ulmus, Prunus spinosa*
Yponomeuta evonymella	***Prunus padus***	

APPENDIX B: Ordered two-way table (in full) obtained from the Day file output. Letters A-H in bold indicate the TWINSPAN classes (end-groups) as shown in Figure 3. Numbers are percentages of the number of samples in which each species occurs in each end-group.

Species	A *n=10*	B *n=28*	C *n=6*	D *n=36*	E *n=11*	F *n=25*	G *n=14*	H *n=27*	Species class
Phyllonorycter tristrigella		29	33	39	27	20		15	*a*
Phyllonorycter schreberella	10	18		33	18	4		4	*a*
Phyllonorycter messaniella	10	4	33	31	9	28	14		*a*
Phyllonorycter leucographella	50	11	50	28	9	32		7	*a*
Phyllonorycter geniculella	10	14		28	18	20		4	*a*
Phyllonorycter corylifoliella	10	14	17	19		24			*a*
Caloptilia rufipennella		11		25		8			*a*
Stigmella plagicolella	60	68	33	17	45	36	7		*b*
Stigmella perpygmaeella	20	21	33	11		12	7		*b*
Stigmella microtheriella	10	54		39	18	36	21	4	*b*
Stigmella lemniscella	20	39	17	19	27	20			*b*
Stigmella hybnerella	30	18	17	14	9	28	7		*b*
Stigmella floslactella	10	68		58		36	36	4	*b*
Stigmella aurella	50	86	50	86		64	57	67	*b*
Stigmella anomalella	10	39	83	22	9	32		4	*b*
Phyllonorycter spinicolella	40	46		3	18	28			*b*
Phyllonorycter oxyacanthae	50	50	17	47	27	64	21		*b*
Phyllonorycter nicellii		64		56	36	48	50		*b*
Phyllonorycter coryli	30	96	67	83	45	80	29	4	*b*
Phyllonorycter acerifoliella		32		8	9	36			*b*
Parornix torquillella	20	25	17	6	27	16			*b*
Parornix devoniella	30	89		72		44	21	15	*b*
Parornix anglicella	90	89	50	75	45	56	29	15	*b*
Lyonetia clerkella	70	57	83	69	45	72	43	30	*b*
Leucoptera malifoliella	40	36	50	17	9	20		4	*b*
Tischeria ekebladella	10	14	33	36	9	40	36		*c*
Stigmella tityrella	10	4	17	47	18	32	64	4	*c*
Stigmella hemargyrella		4	33	36	9	24	43	4	*c*
Phyllonorycter maestingella	10	4	67	61	9	36	50	7	*c*
Phyllonorycter heegeriella		7	17	33		28	21	4	*c*
Cameraria ohridella	20	14	100	47	64	32	14	26	*c*
Acrolepia autumnitella	10	4	50	8	9	20		4	*c*
Stigmella oxyacanthella	50	7		17	18	44	7		*d*
Phyllonorycter ulmifoliella		4	17	42	9	44	29	15	*d*
Phyllonorycter rajella	40	7	17	22	45	40		15	*d*
Phyllonorycter froehlichiella	20	4	33	8	9	32	7		*d*
Gracillaria syringella	10	29	33	42	27	60	29	11	*d*
Ectoedemia atricollis	10	18		17	18	36	7		*d*
Coptotriche marginea	10	29	17	39	9	40	21	44	*d*
Caloptilia stigmatella	10	14		31	27	28	14	15	*d*
Callisto denticulella	20	4	67	11	18	32			*d*
Anthophila fabriciana	70	89	83	75	73	84	79	67	*d*
Udea lutealis		32	17	11	9	56	21	33	*e*
Phyllonorycter stettinensis	40	4		8	27	24		19	*e*
Ectoedemia septembrella	10		33	8	9	32	7		*e*
Ectoedemia occultella	10		33	19	9	48	7	4	*e*
Ectoedemia albifasciella		4		25	9	48	36		*e*
Depressaria daucella	10	4	17	3	27	24	7	4	*e*

Continued overleaf.

APPENDIX B continued:

Species	A *n=10*	B *n=28*	C *n=6*	D *n=36*	E *n=11*	F *n=25*	G *n=14*	H *n=27*	Species class
Celypha lacunana	40	21	33	25		84	86	56	*e*
Anania hortulata	10	4	17	6	27	28	7	11	*e*
Agonopterix arenella		4		11		28	14	7	*e*
Syndemis musculana	10	4		3	9	44	7	15	*f*
Psychodes verhuella	10			6	27	24		7	*f*
Pammene aurana				6	45	32		4	*f*
Nematopogon swammerdamella		4		3		36	14	4	*f*
Mompha epilobiella			17	3	45	24	7	7	*f*
Gypsonoma dealbana	10			3	36	32			*f*
Emmelina monodactyla				3	9	40	7	4	*f*
Elophila nymphaeata			17	3	27	44	7	7	*f*
Ectoedemia subbimaculella			17	3	55	24	7		*f*
Ectoedemia intimella				8	55	28	7		*f*
Depressaria radiella		7		3	45	24		11	*f*
Coleophora gryphipennella	20	4		8	9	40	7		*f*
Carcina quercana			17		9	60	7	7	*f*
Argyresthia bonnetella		7		3	27	44		4	*f*
Agonopterix heracliana		14	17			36		7	*f*
Scoparia ambigualis				8		16	14	30	*g*
Pyrausta aurata		4	33	3	9	24		22	*g*
Psyche casta				8		12	29	15	*g*
Pseudargyrotoza conwagana	20	4		8	27	44	21	22	*g*
Pleuroptya ruralis	10	4		19	36	48	29	41	*g*
Notocelia uddmanniana	10	4		6	9	24	21	11	*g*
Nemophora degeerella	10	11		11	36	60	64	48	*g*
Mompha raschkiella		11		14	9	44	29	30	*g*
Micropterix calthella	10	11		8	36	32	43	30	*g*
Micropterix aureatella				6		12	21	22	*g*
Lathronympha strigana		4		8	9	32	21	19	*g*
Glyphipterix simpliciella	50			8	27	64	50	30	*g*
Esperia sulphurella			17	6	9	24	7	15	*g*
Epinotia tenerana		4		6	18	28	21	19	*g*
Endrosis sarcitrella			33	3	9	24	7	11	*g*
Dyseriocrania subpurpurella	10	4		11		48	43	33	*g*
Coleophora serratella		4		8	18	44	7	26	*g*
Chrysoteuchia culmella		7	17	14	36	40	21	48	*g*
Cauchas rufimitrella	20			3	9	16	14	33	*g*
Bactra lancealana				8	9	24	21	11	*g*
Argyresthia goedartella	10	4	17	8	45	52	29	37	*g*
Ancylis badiana	10	4		8		32	14	22	*g*
Alucita hexadactyla			33	3		28	7	26	*g*
Agriphila tristella	10	4	17	6	36	44	29	48	*g*
Agriphila straminella		11		17	27	56	43	56	*g*
Agriphila geniculea				8		12	7	26	*g*
Agapeta hamana			17	3		36	21	15	*g*
Adela reaumurella		4		11	9	60	50	63	*g*
Phyllonorycter harrisella		4			9	24	29	7	*h*
Micropterix aruncella		7				36	64	22	*h*
Glyphipterix fuscoviridella				3		32	21	15	*h*
Cydia ulicetana						20	43	33	*h*
Crambus pascuella		4				12	14	41	*h*
Crambus lathoniellus		4		3		36	57	52	*h*

APPENDIX C: Ordered two-way table (in full) obtained from the Night file output. Letters A-H in bold indicate the TWINSPAN classes (end-groups) as shown in Figure 5. Numbers are percentages of the number of samples in which each species occurs in each end-group.

Species	A *n=2*	B *n=4*	C *n=19*	D *n=11*	E *n=9*	F *n=10*	G *n=8*	H *n=5*	Species class
Aleimma loeflingiana			74	82					*a*
Amblyptilia acanthadactyla			58	18					*a*
Anania coronata		100	74	36					*a*
Aphelia paleana	50	25	63	45			13		*a*
Archips podana		50	74	82					*a*
Hypsopygia glaucinalis		25	58	36					*a*
Notocelia roborana			47	45					*a*
Tortrix viridana	50	50	68	73			13		*a*
Acleris variegana		75	84	18		20	13		*b*
Alucita hexadactyla		50	79	36				40	*b*
Ancylis badiana			47	36			13	20	*b*
Celypha striana			79	55	33				*b*
Clepsis consimilana			95	64	22	10	13		*b*
Crambus pascuella		50	63	73	22		13		*b*
Cydia pomonella			58	18				20	*b*
Ditula angustiorana			74	64	33				*b*
Epinotia bilunana			58	45		10			*b*
Epiphyes postvittana	50	75	63	55		10	13		*b*
Eudonia angustea			63	9		10			*b*
Lozotaenia forsterana			79	9		10			*b*
Nematopogon swammerdamella			47	36				20	*b*
Nomophila noctuella		100	53	9	11			20	*b*
Phycita binaevella			58	45			13		*b*
Ypsolopha dentella			47	36	11				*b*
Ypsolopha scabrella			63	36	11	10			*b*
Acleris rhombana			53	9		20		20	*c*
Acrobasis advenella		25	89	9	56	20	13		*c*
Agonopterix heracliana			58	9	11	20		20	*c*
Agriphila selasella			37	27	44				*c*
Anania hortulata	50	75	95	91	33	10	25	20	*c*
Aphomia sociella	50	75	95	64	22	10	25	20	*c*
Batia unitella			37	64	33				*c*
Carpatolechia proximella			21	55	11		13	40	*c*
Catoptria falsella		25	68	18	33				*c*
Chrysoteuchia culmella	100	100	95	100	33	10	50	60	*c*
Clepsis spectrana			47	36	33	20			*c*
Crambus lathoniellus			84	73	22	10	13	60	*c*
Crambus perlella			53	55	44		13		*c*
Cydia splendana		25	68	45	33		25	40	*c*
Diurnea fagella			79	27		10	13	60	*c*
Elophila nymphaeata	50		68	73	22	20	38	40	*c*
Emmelina monodactyla		25	63	18		20		20	*c*
Endrosis sarcitrella		75	63	36	11	20	25	20	*c*
Eucosma cana	50	25	53	82	22	20	13	20	*c*
Euzophera pinguis		50	68	27	33		25	20	*c*
Evergestis forficalis	100	75	95	36	11	10	38	40	*c*
Hedya nubiferana		25	74	82	22	10	25		*c*
Hedya pruniana			68	64	11		25	60	*c*

Continued overleaf.

APPENDIX C continued:

Species	A *n=2*	B *n=4*	C *n=19*	D *n=11*	E *n=9*	F *n=10*	G *n=8*	H *n=5*	Species class
Hofmannophila pseudospretella		75	79	27	22	10			*c*
Hypsopygia costalis		50	74	45	33	10	38		*c*
Notocelia trimaculana	50		42	45	11		25	40	*c*
Notocelia uddmanniana	100	25	84	100	33	10	75	20	*c*
Pandemis corylana		25	84	82	44	10	13	20	*c*
Pandemis heparana		25	68	73	44	20	25		*c*
Phycita roborella			63	64	44		13		*c*
Pyrausta aurata		50	58	9			38		*c*
Pyrausta purpuralis		50	42	36			13	20	*c*
Scoparia pyralella	50		37	45		10		60	*c*
Syndemis musculana			42	64	11	20		40	*c*
Tinea trinotella			58	27	22			40	*c*
Udea prunalis		25	89	73	78		25	20	*c*
Yponomeuta evonymella		25	89	64	89	20			*c*
Aethes cnicana			37	36		10		20	*d*
Notocelia cynosbatella			58	55			13	40	*d*
Orthotaenia undulana		25	26	64	11			40	*d*
Pseudargyrotoza conwagana			68	64		10		60	*d*
Spilonota ocellana			68	73	33	10			*d*
Apotomis betuletana			37	55	22	30	13		*e*
Catoptria pinella	50		37	55	33	20			*e*
Eudonia lacustrata	50		95	73	67	30	25	20	*e*
Pandemis cerasana	50		100	91	33	40	63	20	*e*
Rhopobota naevana			26	45	22	10	13	20	*e*
Acentria ephemerella		25	26	45	33	10	13	20	*f*
Acleris forsskaleana		25	63	27	33	10	38		*f*
Agapeta hamana	100	50	95	100	56	30	75	60	*f*
Agapeta zoegana	50		42	27	11		38	20	*f*
Agonopterix arenella	50		74	27	11	40	13	60	*f*
Agriphila geniculea		25	84	36	22	60	38		*f*
Apotomis turbidana		25	16	55	11	30	25		*f*
Blastobasis lacticolella			37	55	11	30	38	20	*f*
Carcina quercana		75	79	82	56	20	63	40	*f*
Celypha lacunana	50	25	84	91	56	40	13	80	*f*
Eudonia truncicolella		25	53	55		50	25	60	*f*
Gypsonoma dealbana		25	32	45	44		13		*f*
Plutella xylostella		25	79	73	44	40	13	60	*f*
Scoparia ambigualis			79	100	22	20	75	100	*f*
Udea lutealis		50	84	55	44	40	50		*f*
Udea olivalis	100	50	100	55	22		63	80	*f*
Agriphila inquinatella			32	27	22	30		20	*g*
Argyresthia brockeella			21	45	22	20	13	20	*g*
Bactra lancealana			26	64	44	50		60	*g*
Epinotia nisella			16	55		60		20	*g*
Epinotia ramella			11	45	11	50	13	60	*g*
Eudonia mercurella	50		79	82	67	50	63	60	*g*
Ypsolopha parenthesella			21	55	22	40		40	*g*
Agriphila straminella	50	100	89	91	89	100	25	40	*h*
Agriphila tristella		100	100	73	56	90	75	40	*h*
Argyresthia goedartella			58	55	44	90	38	20	*h*
Blastobasis adustella			74	64	67	80		40	*h*
Pleuroptya ruralis	50	100	95	91	100	50	88	20	*h*