

# THE SYRPH THE NET DATABASE OF EUROPEAN SYRPHIDAE (DIPTERA), PAST, PRESENT AND FUTURE

Martin C.D. Speight  
Dept. of Zoology, Trinity College, Dublin 2, Ireland  
speightm@gmail.com

## Introduction

Development of the Syrph the Net (StN) database began in 1990. From the beginning, the primary objective was to produce a predictive tool, to aid in evaluation of the biodiversity maintenance potential of sites and in biodiversity management. Biodiversity maintenance issues, as addressed by use of the StN database, relate to species-level biodiversity within ecosystems, i.e. the species is the unit of biodiversity employed in StN procedures and the general objective is to establish what proportion of the predicted syrphid biodiversity is present within a particular area. A secondary objective was that the database would be published, so that it would be available to others who might wish to use it. The primary objective determined to a considerable extent the sort of information that was coded into the database and the secondary objective largely determined the sort of software used for both spreadsheets and text files. There was also a third objective, that the database should have expert system capability, so that its use wasn't confined to specialists in the study of syrphids. This resulted in clear delimitation of the database procedures that require expert knowledge from those that do not.

A central premise of the StN database is that syrphid species are sufficiently closely associated with particular habitats for each habitat to have its own characteristic assemblage of syrphids, making it possible to predict the potential syrphid fauna of a site from the habitats occurring there. In this context the term "habitat" is used in the sense of the CORINE European Habitats Classification System (Devillers et al, 1991) and the EU's Habitats Directive. For many entomologists the "habitat" of an insect is the sort of landscape component referred to in StN as a microhabitat e.g. muddy pond-bottom substrate, tree foliage, rotten wood, cow dung etc. Indeed, in various insect groups the species seem to respond to the environment largely at the microhabitat level and it is difficult to identify habitats, in the CORINE sense, with which they are associated. The ground beetles (Coleoptera: Carabidae) are a case in point – they may have well-defined microhabitats, but only in a very general way are they associated with landscape variability on any larger scale, like the CORINE habitat scale. Syrphids respond at both that scale and at microhabitat scale. The landscape components called habitats in CORINE are referred to as Macrohabitats in StN, to distinguish them from the Microhabitats that are also very important in characterising syrphid species. Essentially, the StN microhabitats are where the larvae of the species occur and are small parts of a Macrohabitat. Syrphid larval biology is unusually diverse and different syrphid species have larvae in most parts of nearly all freshwater and terrestrial macrohabitats. They are only absent from the deeper water of large water bodies like lakes and rivers and from cave systems. In consequence, a site species list of syrphids can provide information about most parts of nearly all the macrohabitats found there. The links between syrphids and Macrohabitats and syrphids and Microhabitats make prediction of site faunas possible. Comparison between the observed and predicted syrphid fauna can then be used to identify which parts of a site are "underperforming" in maintaining its potential biodiversity and which parts are apparently in good condition. This predictive capability is an almost unique feature of the StN database, in respect of terrestrial habitats.

The sheer novelty of the procedures, that had to be developed for use with the StN database, resulted in a rather long period of experimentation before the first published version appeared, in 1997. Since then, the database content and species coverage has been progressively expanded, with

updated versions appearing every 2 or so years. The most recent StN version (2010) codes information for 700 of Europe's syrphid species (more than 80% of the known European syrphid fauna).

### The anatomy of the StN database

The StN database is a set of spreadsheets and text files using Microsoft Office software. So the spreadsheets are Excel files, and the text files use Word. The core of the database is four spreadsheets, into which are coded data on various species attributes, including macrohabitat, microhabitat, traits, range and status. Coding entails digitisation. Digitisation of some attributes, like whether or no a given species is known from a particular part of Europe, is straightforward. Coding "1" for presence and "blank" (i.e. no coding) for absence is immediately comprehensible. But digitisation of many of the other attributes, for instance the degree to which a species is associated with a particular macrohabitat or microhabitat, cannot be achieved so simply. To digitise information of this type a form of fuzzy coding has been adopted (Castella and Speight, 1996).

#### Coding the spreadsheets

The most complex usage of fuzzy coding in StN is found in the macrohabitats spreadsheets. Four degrees of macrohabitat-association are recognised, coded 3, 2, 1 or blank (see Fig.1).

	MACROHABITAT		Supplementary habitat
	Humid, unimproved, lowland grassland		Edge of permanent pool
SPECIES	eutrophic	oligotrophic	
Cheilosia albitarsis	3	3	
Melanogaster aerea		1	1
Melanogaster hirtella	1	1	1
Platycheirus clypeatus	2	3	
Sphaerophoria taeniata	2		

**Figure 1:** extract from StN Macrohabitats spreadsheets, to show fuzzy coding

A coding of "blank" for a particular macrohabitat indicates that there is no known association between a species and that macrohabitat. A coding of "2" indicates that the species is predicted to occur in association with that macrohabitat and a coding of "3" indicates not only that the species is predicted to occur, but also that macrohabitat is the preferred, perhaps the primary, macrohabitat of the species. In the macrohabitats spreadsheets, a coding of "1" indicates a very particular situation, namely that the species is only predicted to occur in association with that macrohabitat when it occurs in combination with another macrohabitat. There are quite a lot of syrphids requiring a combination of macrohabitats. Experience has shown that the habitat combinations required involve one macrohabitat of large extent (e.g. grassland) and another of much smaller extent that is essentially nested within it, like a stream or pond. The approach adopted to these combinations in the StN macrohabitats spreadsheet is to regard a macrohabitat nested within another as a "supplementary" habitat, the species requiring the supplementary habitat coded "1" not only for the macrohabitat of large extent, but also "1" for the supplementary habitat. The extract from the macrohabitats spreadsheet shown in Figure 1 indicates that the species *Melanogaster hirtella* is not predicted to occur in humid, unimproved grassland except when a "supplementary" habitat such as a standing water body (pool) is also present.

The precise meaning of the coding of an attribute in the StN spreadsheets depends on the attribute involved, necessitating explanation of the coding in each case. There is also a need to define each of the attributes. These requirements gave rise to one of the main text files associated with the spreadsheets, the “Content and Glossary” volume (Speight and Castella, 2010a).

#### The Selection Tool spreadsheet

A feature central to the procedure of assessment of the biodiversity maintenance potential of a site, using the StN database, is generation of a list of syrphid species predicted to occur on that site. This is achieved by putting together the lists of species predicted to occur in association with the macrohabitats observed on the site. It is by no means unusual to find 5 or 6 macrohabitats represented, each with one or more supplementary habitat. To make the generation of predicted lists for sites easier and more rapid the “Selection Tool” spreadsheet has been added to the database. The Selection Tool provides a series of drop-down habitat lists, allowing the user to click on each habitat represented on a site. The Selection tool then combines the species lists for the selected habitats, taking the data from the macrohabitats spreadsheets. This is especially useful for dealing with the species that need to be added to a site list due to the presence of supplementary habitats. Without the Selection Tool, generating the lists of species predicted to occur because of the presence of supplementary habitats is a time-consuming process. Inevitably, use of the Selection Tool is not an entirely intuitive process and requires some explanation. This has given rise to a text file instruction manual (Monteil et al, 2008) on use of the Selection Tool spreadsheet, produced now in English, French and Italian, and currently being translated also into Spanish.

#### The Use of the Database text file

Use of the database in assessment of the biodiversity maintenance potential of a site requires both predicted and observed species lists for the site. To obtain comparability between lists of observed species from different locations, whether in different habitats on the same site or on different sites, requires a degree of standardisation of collection effort. That, in turn, gives rise to a need for description of trapping methods and regimes that can provide species lists of a quality appropriate for use with the database. That bundle of information is effectively a fieldwork protocol. Together with description of procedures to follow using the spreadsheets themselves, the fieldwork protocol is incorporated into another StN text file, known as the Use of the Database volume (Speight et al, 2000). Latterly, the information provided on use of the database has been expanded to include a series of case studies of use of the database in site evaluation and management. Each of these case studies has been published as a separate StN volume (Gharet, 2010; Sarthou and Sarthou, 2010; Speight and Castella, 2005, 2010b, 2011).

#### The Species Accounts

Early in the development of the StN database it was recognised that users would find it helpful to have access to an overview of the information on each of the species covered by the spreadsheets, incorporating elements that the spreadsheets do not treat, like what identification literature is available for the species. This resulted in production of the text file known as the StN Species Accounts volume (Speight, 2011 ). The Species Accounts also function to provide reference to published sources of data incorporated into the spreadsheets, though it has to be said that much of what is coded into the spreadsheets has not been published elsewhere, having been provided by European syrphidologists, from their own, unpublished observations. In consequence, there are many “pers.comms.” cited in the Species Accounts. An unexpected outcome of production of the StN Species Accounts was that this volume became sought after by people interested in Syrphidae, but who had no requirement to use the StN spreadsheets. Effectively, the Species Accounts volume developed a life of its own, independent of the rest of the database. Initially, a number of identification keys were included in the Species Accounts volume. But as these became more numerous and the Species Accounts volume expanded due to progressive addition of species

accounts, the identification keys were separated into a volume of their own. The StN Keys volume (Speight and Sarthou, 2011) was the result. It now contains a key to all known European syrphid genera, as well as many keys to the European species of individual genera, and is produced in both English and French.

At the outset, the need for a diverse set of text files to support and explain use of the spreadsheets was not envisaged. It comes as quite a surprise to realise how many of these text files there now are and how important are the roles they play. To use an automotive analogy, if the spreadsheets are the engine which powers the database the text files are the equivalent of the car body that, in conjunction with the spreadsheets, make the database functional.

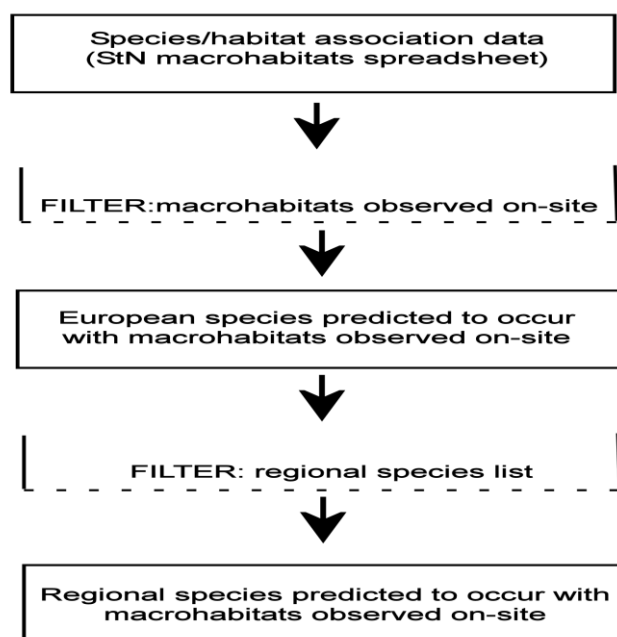
## **The Predicted Species**

### Prediction and the species pool

There are two factors which largely determine which species can be expected in any given area. The first is the “species pool” for the region in which the target area is located. As employed for purposes of generating predicted lists of species using StN, a region is usually the administrative unit (county, canton, province, département, Land etc) of the national territory of the geopolitical entity (State, Nation, Principality etc.) in which the target site is located. There are practical reasons for this application of the term “region”. Firstly, such administrative units are usually delimited on maps, cited in publications of distribution data, referred to on data labels and included as part of each record in biological records centres, greatly facilitating compilation of a regional species list. Secondly, such administrative units have their own biodiversity maintenance objectives and agendas and often require information on the fauna of their territory and on the regional significance of the fauna of sites within their territory. Once a regional list has been compiled, this acts as the “species pool” from which predicted lists are generated for target sites within that region. The principle involved in this use of a regional list is that it represents the maximum possible syrphid biodiversity that could occur on any site within that region, i.e. the syrphid biodiversity to be maintained in that region is finite, and defined by the regional list. Clearly, it is advisable to ensure that a regional list employed in this way is reliable – a certain amount of professional judgement is required to determine whether or no a regional list is adequate.

### Prediction and habitat

The second factor determining which species can be expected on any given site is the macrohabitats observed within that site. The way in which these habitats are used to generate the list of species predicted for an area has already been touched upon in describing the StN macrohabitats spreadsheet. Essentially, the observed macrohabitats are used to select, from among the European syrphids, a list of the species predicted to occur in association with those habitats. The regional species list is then used to further reduce that list, to the species associated with those macrohabitats and also known to occur within that region. It might be anticipated that a region that is extremely heterogeneous, ecologically, could give rise to a regional list liable to exhibit over-prediction i.e. species inappropriately predicted to occur on a site in that region. However, use of observed macrohabitats in conjunction with a regional list reduces such over-prediction to a minimum, because the habitats represented on a site normally provide an integrated, holistic reflection of local topography, climate, geology and biogeography. The species list for a region which encompasses both alpine and Mediterranean conditions, and all stages in between, will reflect that environmental heterogeneity. But the list of species predicted to occur on a site in that region, generated using the StN database, will not include Mediterranean species if Mediterranean habitats are not represented on that site. Similarly, the predicted list will not contain subalpine species if subalpine habitats are not represented on the site.



**Figure 2:** diagrammatic representation of the process of generating a list of predicted species for a site.

### Prediction and biogeography

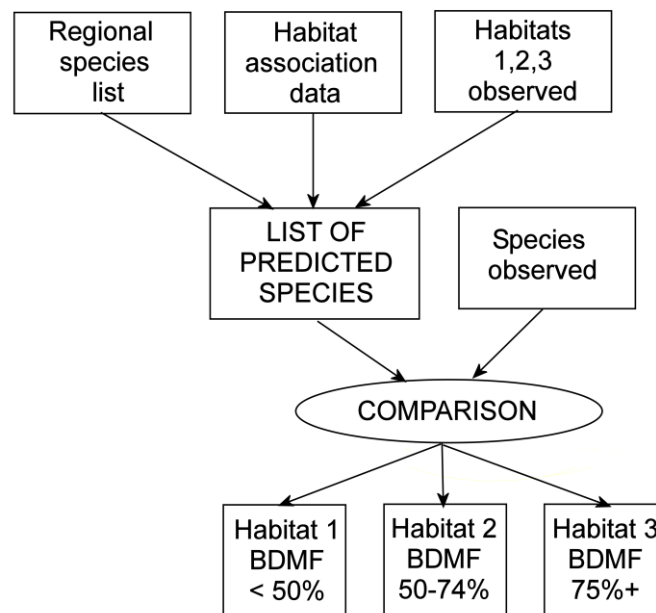
A third factor which can impact on the list of species predicted to occur on a site is the biogeographical zonation of Europe. As a generality, syrphids are highly mobile organisms and can be expected to occur wherever in the continent they find appropriate habitat. However, our understanding of how to define and recognise different habitats remains imperfect, with implications for the coding of species-habitat associations. For instance, the StN macrohabitat category “montane, unimproved grassland” is applicable in both the Pyrenees and the Alps. The two closely-related syrphids *Merodon aeneus* and *M. unicolor* are both predicted to occur in montane, unimproved grassland. *M. aeneus* occurs in the Alps, but not in the Pyrenees. *M. unicolor* occurs in the Pyrenees, but not in the Alps. Both of these *Merodon* species occur in France. So, using the French species list as the species pool for a French site on which montane, unimproved grassland is present, it would in theory be possible to predict the occurrence of both *M. aeneus* and *M. unicolor* there, although in practise these species appear to be geographically isolated from one another and cannot be expected to occur on the same site. So, within countries with a very large surface area, like France or Spain, the regional list performs an additional function in the process of generating a list of species predicted to occur on a site: by basing the prediction on the species known from the region in which a site is located, the regional list ensures that a species which might occur in association with a macrohabitat present on the target site, but whose range doesn’t include that region, will not be predicted for the site. Taking France once again as an example, a site in the French Alps, where montane, unimproved grassland is present, will not generate a list of predicted species including *Merodon unicolor* if the regional list used as the species pool is the species list for the alpine Département in which the site is located, because *M. unicolor* is not listed for any of those Départements. Similarly, the list of species predicted to occur on a montane, unimproved grassland site in the French Pyrenees will not include *M. aeneus*, if the regional list used for generating the predicted list is for the Pyrenean Département in which the site is located.

It would be over-optimistic to imply that anomalous predictions are impossible. The fauna of macrohabitats like many of those found in farmed and urban landscapes, whose characteristics are driven by intensive management by man, does not reflect local topography, climate, geology and biogeography so closely as does the fauna of natural/semi-natural habitats. Anomalous predictions

could in theory result, in cases where a species is present only in intensively-managed habitats in a particular region, but is associated in other parts of its European range with natural/semi-natural habitats. Taking again the example of a montane/subalpine grassland species, *Cheilosia caeruleascens* has been introduced with its foodplant to suburban gardens in Atlantic parts of Europe far outside its natural range and natural habitat. Its presence on the regional list for some part of Atlantic Europe could result in prediction of *Cheilosia caeruleascens* for a montane site in that region, even though neither *C.caeruleascens* nor its foodplant (*Sempervivum*) occurred there. Such instances appear to be few and far between.

### Predicted species, observed species and site evaluation

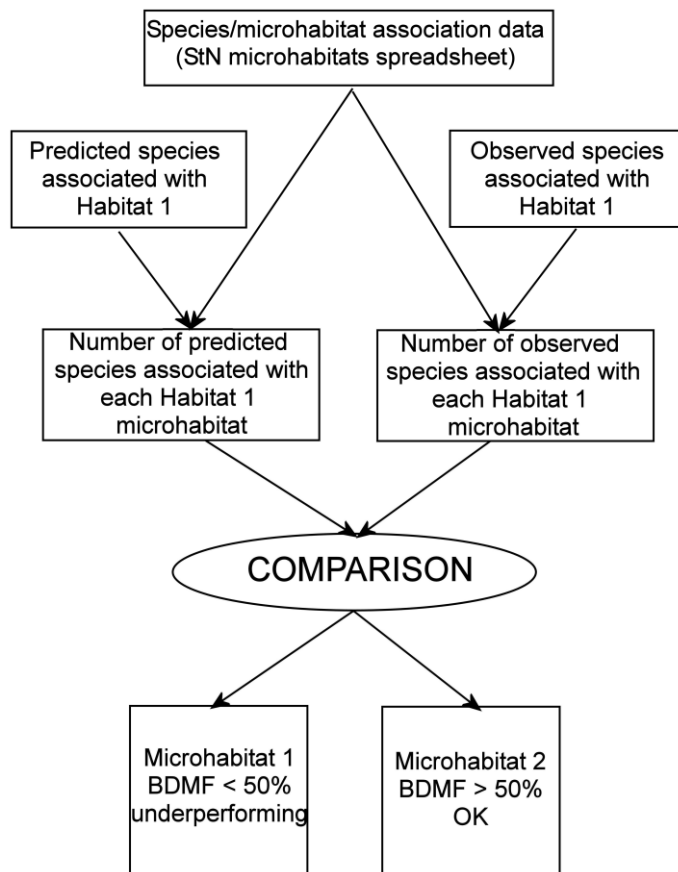
If all of the regional species predicted to occur with a particular macrohabitat are found on a site, that habitat's biodiversity maintenance function is taken to be performing optimally there. If none of its predicted species were found, a habitat would be assessed as totally dysfunctional. In reality, those extremes almost never occur, and representation of predicted species tends to vary between 30% and 80%. There is no absolute measure of the biodiversity maintenance potential of a site. Neither are there agreed international or national evaluation systems. In using the StN database to evaluate the biodiversity maintenance potential of a site a pragmatic approach has been adopted. The higher the percentage of predicted species observed on a site, the higher the biodiversity maintenance potential of that site is taken to be. Experience of using the database has shown that there is a tendency for sites of recognised international significance for nature conservation, such as areas selected for protection under the provisions of the Habitats Directive, to exhibit representation of predicted species at levels of 75% and higher. The same is true for sites maintained for nature conservation under national legislation. But overall site evaluation is less informative than evaluation of the biodiversity maintenance potential exhibited by individual macrohabitats on a site.



**Figure 3:** diagrammatic representation of the basic procedure followed in calculating the condition of the biodiversity maintenance function of a site where three macrohabitats are represented, with each macrohabitat treated separately. This shows a situation where less than 50% of the predicted species associated with one of the habitats (habitat 1) are observed, taken to indicate this habitat is less than 50% functional, in maintaining its potential biodiversity. By contrast, more than 75% of the predicted species are observed for another habitat (habitat 3), taken to indicate this habitat is more than 75% functional.

BDMF = biodiversity maintenance function

Representation of predicted species can differ noticeably between macrohabitats on the same site, and is almost never the same for all macrohabitats on a site. As a “rule of thumb”, the biodiversity maintenance potential of a macrohabitat whose associated syrphids are represented by less than 50% of the predicted species would be subject to further investigation, to see if there are grounds for recommending changes in management of that habitat, in order to increase its biodiversity maintenance potential. Further investigation means comparing the expected representation of species associated with each microhabitat in the under-performing macrohabitat with the observed species associated with the same microhabitats, using the StN microhabitats spreadsheet. This comparison provides information on which parts of the macrohabitat are performing well and which are not. Those parts of the macrohabitat highlighted as not performing their biodiversity maintenance function well can then be targeted for management aimed at improving their performance. The sequence of examining representation of predicted species firstly at macrohabitat level and then at microhabitat level is basic to use of the StN database in investigating the biodiversity maintenance potential of sites. The results do not show *why* the syrphid fauna of a macrohabitat or microhabitat is under-represented in comparison with the syrphid fauna of other macrohabitats or microhabitats, even if the reasons for such underperformance can be strongly inferred from the attributes of the species involved. A site syrphid list frequently totals more than a 100 species, each of which represents an independent data set. Being able to use all of the observed species, or divide them into functional groups to consider separately or in various combinations provides a greater diversity of insights into the biodiversity maintenance characteristics of a site than can be achieved by considering only threatened species as measures of site quality. And, of course, if there is a need to consider threatened syrphids using the database, that can be done anyway, just as any other group of selected species can be investigated.



**Figure 4:** diagrammatic representation of the basic procedure followed in identifying underperforming microhabitats in a macrohabitat (habitat 1) itself recognised as underperforming in maintaining its potential biodiversity. A microhabitat for which less than 50% of the predicted species are observed is taken to be performing its biodiversity maintenance function with less than 50% efficiency.

BDMF = biodiversity maintenance function

#### The spectre of trap selectivity

It might be postulated that the reason for apparent under-representation of the syrphids associated with some macro-habitat, or microhabitat, could be due to selectivity of trapping devices used in compiling a list of the observed species. In the case of the type of field survey programme recommended for use with the StN database this would mean consistent failure of Malaise traps to collect syrphid species associated with particular macrohabitats or microhabitats. It would not be realistic to claim that Malaise traps collect all syrphid species with equal facility. Indeed it has been shown that some species are collected by Malaise trap more easily than others. But that is a quite different matter from demonstrating that the “trap-shy” syrphids are associated more with one macrohabitat or microhabitat, than with another. Unpublished results of using the database to assess the same site using a list of observed species collected by hand-net and a list collected by Malaise trap demonstrated the same results from both sets of data, although the two species lists were not identical. Using the database it is not differences in the names of the species collected that is of significance, but differences in their attributes. In this case the bundle of attributes exhibited by the species in each list were the same. So the fact that some species collected by hand net were not collected by Malaise trap, and vice versa, did not impact on the results.

#### The pre-eminence of species presence/absence data in StN site evaluation procedures

An issue related to trap selectivity is species abundance data. When species are used as integers in statistical analysis, without reference to coded attribute information, the only information available for use is the number of species involved, the combinations in which they occur and the number of specimens of each species. In such statistical analyses relative abundance of species is difficult to ignore, given the few inputs available to fuel the analysis, and trap selectivity can be an issue of some import. When coded attribute data are available, presence-absence data become of far greater value, because each species name is not just a label but carries with it all of that digitised attribute information. It is a general feature of statistical treatment of survey data that the rarest species collected tend to be left aside from analysis, because they are represented by too few individuals for the products of analysis to be meaningful if they are included. At the opposite extreme, abundant species provide a solid basis for analysis, and are so included. Use of the StN database does not depend on quantitative data. Species collected only as single individuals have the same information content as species collected in large numbers and both can be employed in analysis and evaluation together. Indeed, interrogation of the database in analysing/evaluating the fauna of a site depends only on presence/absence data. This can be quite important in circumstances where survey procedures yield only small numbers of individuals of each species, as demonstrated in Speight and Castella (2005). Co-incidentally, the capacity of the database to function using presence/absence data reduces any possible impact of trap selectivity on interpretation of results, since it doesn't matter if the trap collects one species less efficiently than another, so long as at least one specimen of each is collected. More importantly, the species likely to have originated from a site under investigation can be separated from those likely to have originated elsewhere, using the coded attribute data. For instance, species like *Episyrphus balteatus* or *Sphaerophoria scripta*, which could be collected in large numbers on a site where they probably represent migrant populations originating in neighbouring crops, can be excluded from consideration when the biodiversity maintenance potential of that site is under scrutiny, because they are coded in the Traits file as migratory. It is not that relative abundance data cannot be used in conjunction with the StN database, but that they are not necessary in order to use the database.



Also, the database can show when super-abundant species are more appropriate to exclude from analysis than species represented in the same survey results by only a single specimen.

#### Extrapolation from results achieved with the StN database

The Stn database enables comparison between predicted and observed syrphid faunas and provides information about the attributes of species both predicted to occur and observed and species predicted but not observed, linked to the habitats with which they are associated. In this way it delivers insights into the potential of a site to maintain syrphid biodiversity and on macrohabitats and microhabitats potentially requiring management modifications in order to maximise syrphid biodiversity. But are these insights of more general relevance – with the StN database, can one use syrphids as bioindicators of biodiversity maintenance for other invertebrates? Using criteria for selection of invertebrates for use in assessment of the state of the biodiversity maintenance function of habitats, syrphids rate quite highly (Speight, 2008). Even so, they represent only approximately 1% of the invertebrate fauna of terrestrial/freshwater landscapes, a very small proportion of the total invertebrate fauna, and a reluctance to extrapolate to other invertebrates the results of such a small sample of the invertebrate fauna is easily understandable. Unfortunately, European-level databases providing equivalent information for other taxonomic groups of terrestrial invertebrates are almost non-existent. So simple comparison between the results obtained for syrphids, using StN, with results obtained for other taxonomic groups of invertebrates, using similar databases, are not available. Indeed, there are few studies in which the StN database has been employed and other taxonomic groups of invertebrates have been investigated at the same time, by whatever means. One example is provided by Goeldlin et al (2003), which shows quite clearly the limitations on deductions made about the Coleoptera studied, imposed by the lack of databased information, in comparison with what could be achieved for syrphids using the StN database. In present circumstances, syrphids are becoming *de facto* bioindicators of the state of the biodiversity maintenance function of habitats, not only because they seem to be a taxonomic group peculiarly suited to that role, but also because of the existence of the StN database, which makes them more accessible to interpretation and analysis than other taxonomic groups.

### **The Unpredicted Species**

The capacity to predict which species will occur there is central to use of the StN database in assessing the biodiversity maintenance potential of a site. But the observed syrphid fauna of a site normally contains two elements, species that are predicted to occur and species that are not predicted to occur. The unpredicted species can originate either on-site or elsewhere. While they may not be directly involved in the site assessment procedure they can nonetheless provide valuable information and cannot be regarded as irrelevant.

#### Overlooked habitats

Adopting the practice of leaving the observed but unpredicted species on one side until after a site assessment/interpretation process has been completed is not advisable. All too frequently, examination of the known macrohabitats of the unpredicted species collected on a site indicates that one or more macrohabitats present on-site may have been overlooked during habitat survey, necessitating critical re-appraisal of the site macrohabitat list. Macrohabitats which characteristically occupy only small areas, such as springs and flushes, are particularly susceptible to being overlooked, as are seasonal macrohabitats, like temporary pools and streams. Macrohabitats occurring together in mosaic form are also susceptible to misinterpretation, the presence of one component of the mosaic being recognised while the other goes un-recorded. Macrohabitats, whose presence is discovered from checking on the unpredicted species observed on a site, then require to be included in the site assessment/interpretation process.

### Poorly-known habitats

A second source of unpredicted species originating on-site is species associated with one or other of the macrohabitats present, but not coded for those macrohabitats in the database. Sites on which macrohabitats not covered by the database are present might be expected to support some species not coded into the macrohabitats spreadsheets, and such sites can be valuable sources of information leading to incorporation of additional macrohabitat categories into the database. The basic principle that has been adopted, in relation to adding a macrohabitat to the database, is to wait until a basic list is available for that habitat, and then incorporate it. This approach anticipates the addition of further species to those coded for that macrohabitat, as and when more information becomes available. So the unpredicted species also have to be scrutinised for potential additions to habitat-association lists already coded into the Macrohabitats spreadsheets.

### Off-site habitats

Unpredicted species originating off-site can come from habitats in the vicinity of the site or even further afield. The group of syrphids categorised as migrants, because of their propensity for long distance flight, can be found a long way from macrohabitats in which they can develop, on sites that couldn't possibly support them. Other syrphids, even if more susceptible to being restricted in their movements by ecological barriers, nonetheless undertake local movements into habitats in which they cannot sustain populations. A well-recognised phenomenon of this sort is syrphids from surrounding habitats visiting fields for flower-feeding purposes. A Malaise trap installed in a flowering meadow, close to a forest margin, can thus collect not only syrphids associated with the meadow, but also syrphids associated with the adjacent forest. Using the database to generate a list of predicted species for the meadow will then indicate the presence of unpredicted forest syrphids in the observed species list for the meadow. In today's Europe, areas of natural/semi-natural habitat are frequently reduced to islets within a sea of cultivated land. In those circumstances, Malaise-trap survey of the syrphid fauna of an area of natural/semi-natural habitat carries with it the almost inevitable consequence that unpredicted species originating in the surrounding farmland will make up part of the observed fauna, as collected in the traps. Indeed, in terms of numbers of individuals collected, the syrphids derived from the surrounding farmland may far outnumber the species associated with the habitats represented on the site surveyed. In recording habitats represented on a site, it is in consequence necessary to record also the habitats predominating in the vicinity of that site, in order to better understand the potential origins of unpredicted species in the site's observed fauna.

### Landscape permeability issues

The apparent reluctance of many syrphids to cross relatively minor barriers provides another source of unpredicted species. A Malaise trap installed against a hedge, on a river bank, or on a motorway margin can collect non-migrant syrphids associated with habitats quite distant from the location of the trap, the syrphids evidently having followed along the edge of the barrier, presumably in an attempt to circumnavigate it. In mountainous terrain another source of unpredicted species is the tendency for syrphids to fly uphill. Because of this tendency, Malaise traps installed in subalpine habitats on the top of a massif can collect syrphids from various habitats found only at lower altitudes, as well as species from subalpine habitats. This phenomenon is illustrated in Speight and Castella (2010b). The tendency for not only syrphids, but also other Diptera, to accumulate on high points in the landscape, has given rise to the term "hill-topping".

## **Exploring the potential of the database**

After a quick glance at the user's manual, the owner of a new car wants simply to turn on the ignition and drive the car away, without first having to find out how its engine works. Database users demonstrate similar tendencies – following the procedures described in StN text files the

spreadsheets get used for investigating site biodiversity maintenance potential, but rarely for other purposes. Further applications of spreadsheet use cannot easily be developed without a reasonably comprehensive knowledge of spreadsheet content and that requires both time and motivation. Even so, some extensions of database use can be achieved without the need for much additional knowledge. For instance, the predictive capacity of the spreadsheets is not restricted to indicating which species might be expected on a site right now. Using the same principles, they can be used to predict both past and future syrphid faunas. An example of their use in investigating the syrphid fauna of the past is provided by Speight (2004), who used them to gain an overview of how the Irish syrphid fauna developed over 10,000 years or more in the post-glacial. Using the database to predict how the fauna of a farm might be expected to change under the influence of future management changes is shown by Speight (2001).

Studies of the range and distribution of species are not novel. Indeed they have given rise to the term biogeography. What is novel, in using the StN spreadsheets in biogeographical studies, is that the range and distribution of species attributes can be explored. In Speight and Good (2003), the microhabitats of saproxylic syrphids in different zones of Europe were compared using the database spreadsheets, showing that fallen, dead wood microhabitats support a higher proportion of the saproxylic species in northern parts of the continent than in the Mediterranean zone, where the microhabitats found on overmature, living trees are of greater significance. Considering syrphids associated with forest macrohabitats in Europe's Atlantic zone, and once again using the spreadsheets, Speight (1996) showed that wherever a deciduous forest was replaced by a conifer plantation a reduction in syrphid biodiversity could be anticipated. Similarly, in Speight (2004) it was shown that saproxylics comprise the most poorly represented component of the syrphid fauna of forest macrohabitats in Ireland, linking this to the virtual eradication of indigenous forest cover achieved in previous centuries. Gittings et al (2006) employ the database in demonstrating that the greater part of the syrphid diversity in conifer plantations in Ireland is dependent on open areas within the plantations, rather than the tree crop. In Biesmeijer et al (2006) syrphid traits drawn from the database were used to explore similarities and differences in the syrphid fauna of Britain and the Netherlands, in relation to changes in distribution and status of pollinators. Compilation of the necessary biological information about syrphids, in the absence of a tool like the StN spreadsheets, would be so time consuming as to make such biogeographical comparisons virtually impossible. For syrphids, the possibility of conducting such comparisons is now more limited by the restricted availability of reliable regional species lists. Even so, the capacity of the database to provide biological insights into distributional and range data could be more widely used than it has been to date.

So far, no study has been carried out that used the StN database spreadsheets in their entirety. It is probably impractical, if not impossible, to do so. Most uses of the spreadsheets made so far require only a small fraction of the coded information, even data for only a single species attribute. And users of the database often don't use the spreadsheets at all, confining their attention to the Species Accounts text file!

## **Where to now?**

### Training workshops

One of the objectives of setting up the StN database, identified at the outset of this account, was that it should not only be usable by specialists in Syrphidae. The sort of scenario envisaged was that, if a syrphid specialist provided a nature reserve manager with species lists, the reserve manager, even if knowing little about syrphids, could then use the database spreadsheets to extract from the species lists information useful to management of the reserve. Experience has shown that, although the database can be used in this way, it is more usual for a syrphid specialist to prefer also to carry

out the analysis using the spreadsheets, and for a site manager to want to learn sufficient about syrphids to also carry out the site survey work. More problematic has been the discovery that neither syrphid specialist nor site manager can be expected to have experience in habitat recognition, habitat survey or use of Excel spreadsheet software! The net result is that, in order to bring the StN database into use in a country, it is proving necessary to hold instruction workshops that cover design and implementation of field survey programmes, including habitat survey; syrphid identification; sample sorting; Excel use and procedures to follow in interrogation of the database spreadsheets. As interest in the database and its use grows and spreads, so does the demand for training workshops .

In various ways, use of the StN database files has become easier as successive versions of the database have been produced. But this trend is not universal. The changes in presentation of Excel files that have accompanied release of each new version of the software have added little to the performance of the software, in carrying out the rather simple calculations and manipulations required to use StN. On the contrary, these changes in presentation have made it necessary to “re-learn” how to use the software following release of each new Excel version and have made it impractical to produce a pictorial instruction manual on use of Excel spreadsheets with the database, using frames imported from Excel, since this would have to be revised for each new version of the software. These changes also complicate training workshops, since participants of necessity use their own computers and can find themselves using different versions of Excel from each other and from the version used for demonstration purposes. There is no indication that this problem will diminish in the foreseeable future, and so will have to be accommodated in production and use of the database.

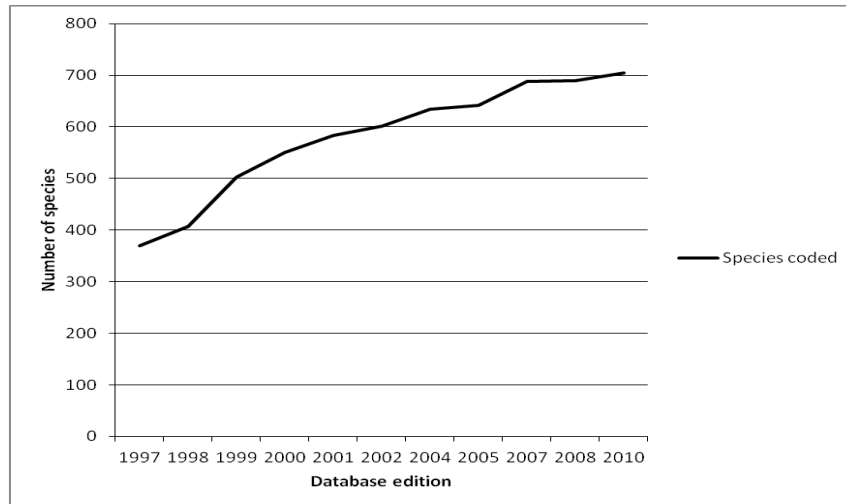
#### StN as a teaching tool

Excel spreadsheet software is now widely used in third-level teaching institutions. That fact, coupled with the reality that use of the StN spreadsheets does not require specialist knowledge of syrphids, has resulted in a developing use of the StN database as a teaching tool in Environmental Science courses, particularly at the M.Sc level. Based on previously-gathered data sets (derived from Malaise trapping in reasonably identifiable habitats on a site accessible to the University) students are able to compare observed syrphid lists with lists of predicted species they have themselves generated. In its more comprehensive form the course unit incorporates most parts of the sequence of events that lead to site evaluation/management recommendations, from a site visit demonstrating habitat survey and Malaise trapping through to report production and presentation. Results so far show this to be an effective way of introducing the concept of invertebrates as bioindicators in biodiversity management. It is also proving popular with students.

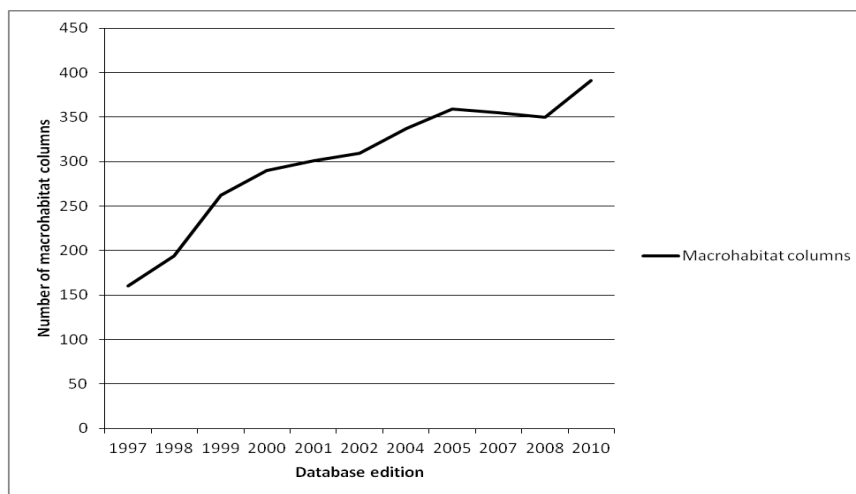
#### Database maintenance and expansion

The database continues to expand, albeit more slowly than it did initially. In the 1997 version, less than half of the European species were coded into the spreadsheets and various of the supporting text files were yet to be produced. Virtually all of the known European species for which there is sufficient information are now coded and the set of text files is near completion. However, the number of syrphid species known from Europe still changes each year, with a few relegated into synonymy and a larger number discovered. The annual additions to the European syrphid fauna are a combination of newly described species and species described elsewhere previously, newly found in Europe. Each year, there is a net increase of 10 or more species in the European list. This is reflected in the current rate of increase in species coverage in the spreadsheets, shown in Fig. 5. The coverage of macrohabitats coded into the StN spreadsheets has not yet reached a point where only a few habitats remain to add. There remain many, especially in eastern and southern Europe, not yet coded into the StN spreadsheets. But reliable information allowing the coding of additional macrohabitats is hard to come by. The progress in habitat coverage is shown in Fig. 6. Review of

habitat categories, carried out in the period 2005-2008, led to a small decrease in the number of categories, caused by amalgamation of some with others. The steeper incline of the graph shown recently is due to a focus on addition of Mediterranean zone habitats, a process which still continues.



**Figure 5:** number of syrphid species coded into each version of the spreadsheets of the StN database

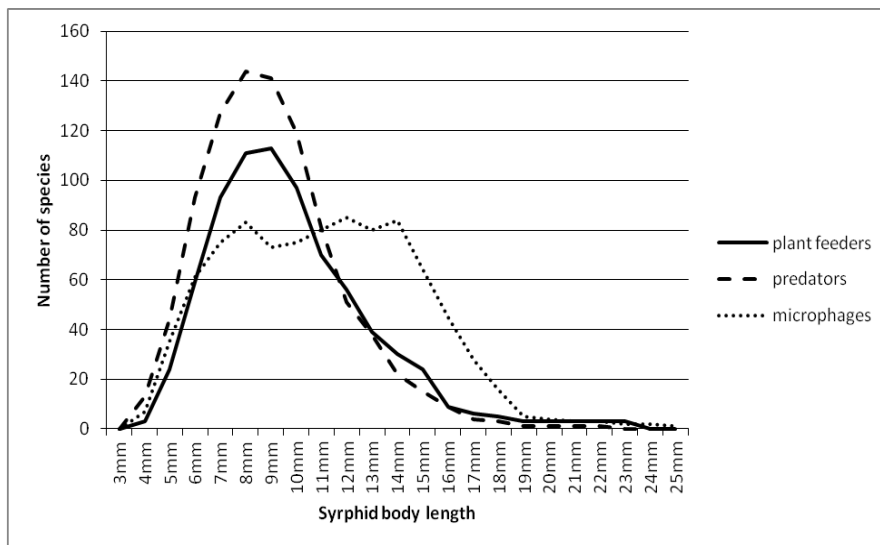


**Figure 6:** number of macrohabitat columns coded into each version of the spreadsheets of the StN database

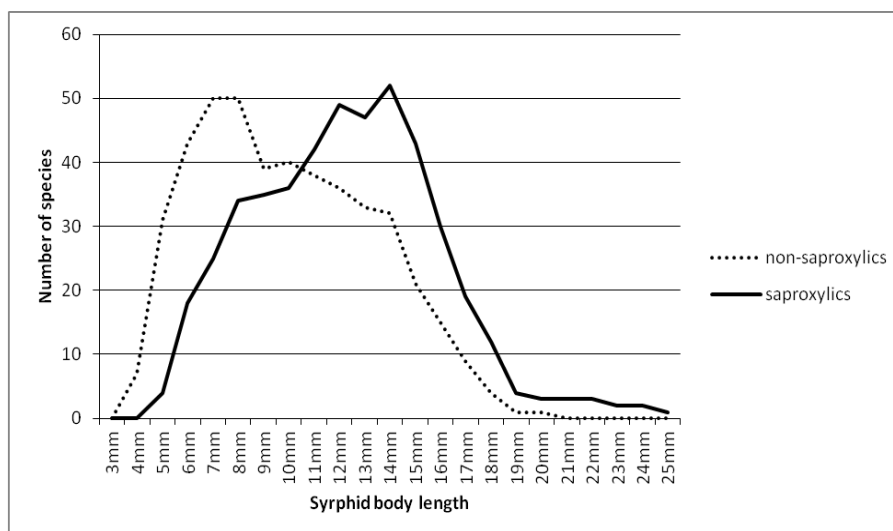
Coding of an additional macrohabitat is characteristically a consequence of consultation with a European syrphidologist with knowledge of that habitat – the necessary information does not appear in publications. Whether the lack of published sources of species/habitat association data is because syrphid specialists don't think to publish such material or journals are not interested in publishing it, or both, is not clear. Many syrphidologists are now conscious of which habitats they are collecting in. Initially, asking someone where a particular species had been found was as likely to elicit the response "on a flower" as a habitat description! Further expansion of macrohabitat coverage by the database would clearly be desirable – use of the database in Mediterranean and eastern parts of Europe remains restricted due to limited habitat coverage.

The microhabitats array incorporated into the first version of the StN spreadsheets has hardly changed since, making the microhabitats spreadsheet one of the most static elements of the entire database. The number of traits covered has increased slowly. But availability of traits data is highly

variable, with information on a particular trait existing for some species but not for others. One trait not coded into existing versions of the database, but for which data are available, is body length of the adult insect. At the time of writing (Spring, 2012) this trait has now been coded but awaits publication in the next version of the database. Using the coded body length data, with the European syrphids divided into three groups according to larval feeding habits, produces the result shown in Fig.7. A surprising similarity in the range of body lengths is apparently exhibited by syrphids with plant-feeding and predatory larvae, contrasting sharply with the range of body lengths exhibited by species with microphagous larvae. Fig.8 demonstrates that saproxylic microphages are responsible for much of the difference, including, as they do, so many species with a body length greater than found among the plant-feeders or predators.



**Figure 7:** Body length of adult European syrphids, based on 820 species, with the species divided into three categories according to larval feeding habits. Total number of species with plant-feeding larvae = 291; total number of species with predatory larvae = 328; total number of species with microphagous larvae = 298.



**Figure 8:** Body length of adult European syrphids with microphagous larvae, divided into saproxylic and non-saproxylic species. Total number of saproxylic microphages = 129; total number of non-saproxylic saprophages = 169

The future development of the Range and Status spreadsheet is subject to the same influences as the other spreadsheets. Essentially, the greater part of the information available for coding is now in the spreadsheet, so its rate of expansion is largely dependent on the rate at which new data are generated. All recently-published lists of syrphids for European States and similar entities are now coded. Within the EU, only Austria, Bulgaria, Cyprus and Greece still lack published species lists. The provision of national lists within the Range and Status spreadsheet is now primarily a matter of trying to keep abreast of updates to existing lists, rather than adding lists for parts of the continent not covered previously. It is to be hoped that European syrphidologists will now turn their attention to listing the species requiring designation as threatened in their part of Europe. Because such undertakings require a more comprehensive knowledge of the species than is needed for compiling a national species list, generating lists of threatened species is likely to prove a slower process. A certain number of regional lists (i.e. lists for parts of countries) are also incorporated into the Range and Status spreadsheet, but significant expansion of this component of the spreadsheet is unlikely. Regional lists are now becoming available on nationally-maintained websites, so there is less need for StN to supply them. An example is the SYRFID website, from which the list of syrphids known from any French Département can be downloaded.

From issue of the first published version of the database, in 1997, dissemination of the database files to users has been carried out electronically. Initially, this was a cumbersome procedure, due to the small size of files that could be sent via e-mail. The considerable advances that have been made in data transfer since then have changed radically both the quantity and the quality of what can be sent, particularly following introduction of broadband technology. Further, downloading facilities can now be set up that allow dissemination of even larger bodies of information in individual files. During the same period, computers have become standard household items and internet access has likewise become almost universal, throughout Europe. There are now recipients of the database in 24 European countries and also in a few countries outside Europe.

Overall, it could be said that the StN database is entering a phase of consolidation, when adding to the material coded for attributes already incorporated into its spreadsheets predominates over increasing the number of attributes coded. A major exception is represented by the macrohabitats – information is still becoming available for macrohabitats as yet not covered by the database, so further increases in macrohabitat coverage can be anticipated. The text files mirror the situation of the spreadsheets – updating of content, rather than production of text files on new topics, is becoming the order of the day. But this is hardly a suggestion that little remains to be done! And however innovative and ahead of its time the database may be, however impressive its data content, it is only as its utility is manifested, in publications and presentations, that it becomes truly the tool it is made to be.

### **Acknowledgements**

I am grateful to Ruth Blackith, Michael de Courcy Williams and Phil Withers for their helpful comments on an earlier version of this text. Emmanuel Castella, Jean-Pierre Sarthou and Veronique Sarthou have contributed much to the concept and content of the database over the years. Their inspiration and encouragement have ensured it continues to develop. And without the active participation of Europe's syrphidologists, who have been so generous with their information – often unpublished, the database could never have "taken flight". Finally a thank you to the users of the database, who make the effort worthwhile.

### **References**

Biesmeijer, J. C., Roberts, S. P. M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A. P., Potts, S. G., Kleukers, R., Thomas, C. D., Settele, J. & Kunin, W. E. (2006) Parallel Declines

- in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands. *Science*.313: 351-354.
- Castella, E. & Speight, M.C.D. 1996. Knowledge representation using fuzzy-coded variables: an example with Syrphidae (Diptera) and the assessment of riverine wetlands. *Ecological Modelling*, 85: 13- 25.
- Convention on Biological Diversity (1992). United Nations.
- Devillers, P., Devillers-Terschuren, J. & Ledant, J.-P. (eds.) (1991) *CORINE biotopes manual : Habitats of the European Community, Data specifications part 2*. 300pp. Commission of the European Communities, Luxembourg.
- Gharet, S. (2010) Mise en application du diagnostic base sur l'étude des dipteres syrphidae sur les tourbieres du lac des rouges truites et de l'entrecotes (39). *Syrph the Net, the database of European Syrphidae (Diptera)*. Vol. 64, 73 pp + Ann., Syrph the Net publications, Dublin.
- Gittings, T., O'Halloran, J., Kelly, T. and Giller, P.S. (2006) The contribution of open spaces to the maintenance of hoverfly (Diptera, Syrphidae) biodiversity in Irish plantation forests. *Forest Ecology & Management*, 237: 290-300.
- Goeldlin, P., Delarze, R., Castella, E. & Speight, M.C.D. (2003)Projet-pilote de gestion écologique des forêts de Montricher (Jura vaudois, Suisse): les insectes indicateurs. *Mém.Soc.vaud.Sc.nat*, 20: 159-267.
- Monteil, C, Speight, M.C.D., Sommaggio, D., and Sarthou, J.P. (2008). Species Selection Tool for the Syrph the Net Database, 2008. *Syrph the Net, the database of European Syrphidae*, Vol. 58, 32 pp, Syrph the Net publications, Dublin.
- Sarthou, V. & Sarthou, J.P. (2010) Evaluation écologique d'écosystèmes forestiers de Réserves Naturelles de Haute-Savoie à l'aide des Diptères Syrphidés. *Syrph the Net, the database of European Syrphidae*, Vol. 62, 131 pp, Syrph the Net publications, Dublin.
- Speight, M.C.D. (1996) Syrphidae (Diptera) of Central France. *Volucella*, 2: 20-34.
- Speight, M.C.D. (2001) Farms as biogeographical units: 2, the potential role of different parts of the case-study farm in maintaining its present fauna of Sciomyzidae and Syrphidae (Diptera). *Bull. Ir.Biogeog.Soc.*, 25: 248-278.
- Speight, M.C.D. (2004) Towards an understanding of the development and constitution of the Irish post-glacial syrphid fauna (Diptera: Syrphidae). *Volucella*, 7: 125-155.
- Speight, M.C.D. (2008) Database of Irish Syrphidae (Diptera). *Irish Wildlife Manuals*, No.36, 338pp. NPWS, Dept of Environment, Heritage and Local Government, Dublin.
- Speight, M.C.D. (2011) Species accounts of European Syrphidae (Diptera), Glasgow 2011. *Syrph the Net, the database of European Syrphidae*, Vol. 65, 285 pp., Syrph the Net publications, Dublin.
- Speight, M.C.D. & Castella, E. (2005) Assessment of subalpine grassland and heath sites in Haute-Savoie using Syrphidae (Diptera). *Syrph the Net, the database of European Syrphidae (Diptera)*. Vol. 46, 37 pp., Syrph the Net publications, Dublin.
- Speight, M.C.D. & Castella, E. (2010a) StN Database: content and glossary of terms. 2010. *Syrph the Net, the database of European Syrphidae*, Vol. 61, 83 pp , Syrph the Net publications, Dublin.
- Speight, M.C.D. & Castella, E. (2010b) Diagnostic de quelques habitats des Hauts de Chartreuse (Isère, France) à l'aide des Syrphidés (Diptères). *Syrph the Net, the database of European Syrphidae*, Vol. 63, 36 pp, Syrph the Net publications, Dublin.
- Speight, M.C.D. & Castella, E. (2011) Analyse des Syrphidae (Dipteres) de deux réserves alluviales de Haute-Savoie, en relation avec l'évaluation et la gestion des sites. *Syrph the Net, the database of European Syrphidae*, vol. 67, 35 pp., Syrph the Net publications, Dublin.
- Speight, M.C.D. & Good, J.A. (2003) Development of eco-friendly forestry practices in Europe and the maintenance of saproxylic biodiversity. In: Mason, F., Nardi, G. & Tisato, M. (eds.) Proc. Internat.Symposium "Legno Morto: una chiave per la biodiversità", Mantova, May 2003, *Sherwood*, 95, Suppl.2, 73-77.
- Speight, M.C.D. & Sarthou, J.-P. (2011) StN keys for the identification of adult European Syrphidae (Diptera), Glasgow 2011/Clés StN pour la détermination des adultes des Syrphidae Européens



(Diptères), Glasgow 2011. *Syrph the Net, the database of European Syrphidae*, Vol. 66, 120 pp, Syrph the Net publications, Dublin.

Speight, M.C.D., Castella, E. & Obrdlik, P. (2000) Use of the Syrph the Net database 2000. *Syrph the Net, the database of European Syrphidae*, Vol.25, 99 pp., Syrph the Net publications, Dublin.