The phytosociological database SOPHY as the basis of plant socio-ecology and phytoclimatology in France

Emmanuel Garbolino, Patrice De Ruffray, Henry Brisse & Gilles Grandjouan

Abstract: This article draws attention to the botanical and ecological database SOPHY (GIVD ID EU-FR-003), which is hosted at the University Paul Cézanne at Marseille, France. Initiated in the 1970s, this database was first dedicated to the study of the relationships between plants and climate (phytoclimatology) at the scale of France. In the early 1980s it was central to the development of socio-ecology, which studies the statistical relationships between plant species. At present the SOPHY database contains more than 200,000 plots located in France and in some areas close to the French border. The managers of the database have developed methods and algorithms dedicated to the control of data quality and to the characterisation of socio-ecology and phytoclimatology. The principles and results obtained in these two domains using the SOPHY database are presented.

Keywords: fidelity; mesotype; paradigm shift; phytotype.

Received: 29 October 2010 – Accepted: 19 April 2012 – Co-ordinating Editor: Jörg Ewald.

Introduction

The objective of the database SOPHY is not only to collect and supply phytosociological data, but to supply the field botanists with methods for the ecological characterisation of plant species and plots at the community level. The database combines all available plot data from France and some data from the neighbouring countries Belgium, Luxemburg, Southern Germany, Switzerland, Italy, Andorra and Spain. Figure 1 shows the general localization of the plots recorded in the SOPHY database.

Digital plot data from the SOPHY database have been used by collaborators for research on the dynamic of vegetation and the consequences of global climate change on plant distribution. Among these works, the articles by Gegout et al. (2005) about the EcoPlant database that integrates a part of SOPHY database’s plots, Albert et al. (2008) on vegetation dynamics, Lenoir et al. (2009, 2010) and Bertrand et al. (2011) on climate change impacts are notable.

In this paper we present the structure of the database SOPHY and two main examples of methodologies and results in the domain of socio-ecology and phytoclimatology using our database.

Database structure

The data is stored in four groups of tables, which are usually realised in text format (Fig. 2):

Bibliography

Bibliographic sources are managed in dBase tables. Each source record has at least one vegetation plot assigned to it. SOPHY is based on more than 4,500 references, which have been assigned ids in the order of their integration into the bibliographic table. Among them, 3,600 references are in digital form. The bibliography is displayed with localisation of each dataset (as of 2009). Each reference has an ID attached to it, which can be found at the SOPHY website (http://SOPHY.univ-cezanne.fr).

Taxonomy tables

The original taxonomic reference list was based on the form for floristic surveys of vascular plants of France (Brisse & Grandjouan, 1971), which was developed from the flora by Fournier (1963). It was replaced by the Flora Europaea list (Brisse & Rasmont unpubl.), later on by the digital code of the flora of France (Brisse & Kerguelen, 1994) and finally by the database of plant nomenclature of the flora of France (Bock 2004). A reference list of bryophytes based on the flora by de Augier (1966) has been added. Each reference list contains an indexed list of taxa. These species ids are used in coding vegetation plots.

Plot observations

Vegetation plots are arranged in cross-tables of the same format as found in the source data. Header data are added in the same format at the bottom of the table. Each table consists of three parts: A title, the table itself and a marker for the end of the table consisting of one line consisting of the number 9999. The title is made up of four parts: ID of the source, ID of the phytosociological table, number of plots and name of the displayed vegetation type as given by the original author (if available). A plot can thus be identified by an 8-digit composite number of the format RéfTbRITb, consisting of reference ID (Réf), table ID (Tb) and plot ID (Rl). To date, more than 200,000 plots have been digitised (Fig. 1).
Data control

There are control procedures for each data type. For references and plant names proofreading is usually sufficient to clean tables from errors. Control of phytosociological data is more complex, but has been increasingly automated. First of all, on each data addition, the structural identity of phytosociological and location tables is checked.

With locations, the correspondence of coordinates and place names is checked by comparing distances between the centroid of the community and plot coordinates. Records with distances > 10km are reported by an error message. As only community names are indexed, inspection of maps allows to detect the most severe errors.

As for phytosociological data, beyond proofreading an automated procedure has been developed. It consists of measuring the distance between plot ordination (see below) and ordination of each component plant species. Inspection of distances allows returning to data sources and check whether a coding error has occurred or whether occurrence of the species in an obviously untypical habitat is plausible.

The essential method that distinguishes SOPHY from similar databases is that it generates two main scientific contents:
- A characterisation of the socio-ecological behaviour of each plant species represented in the data. Species then serve to define the habitat conditions in the plot. From these two basic computations a large array of results is generated. The classification of these behaviours provides a hierarchy where it is possible to identify the main environments defined by the plants.
- A characterization of the climatic behaviour for each plant species in order to formalize climatic bio-indicators. After the application of a calibration between plants and climate data (climatic data were provided by MeteoFrance), the plants can be used as indicators of the climate variables. It is also possible to classify these bio-indicators and to map their clusters in order to identify the main climatic factors that contribute to plant distribution in France.

Methodology in socio-ecology

The theoretical foundation and a discussion of the methodology in socio-ecology is exposed in the document "Changement de paradigme en écologie végétale" (A change of paradigm in vegetation ecology) at the SOPHY website as well as in the paper published by Grandjouan (1998). The ecological information contained in the abundance of plants is taken into account (Brisse & Grandjouan 1980), by subdividing species into pseudo-taxa, when their abundance exceeds a certain threshold. Thus, a species may be replaced by two (or three) pseudo-species defined by threshold levels of abundance (Brisse & Grandjouan 1977, Hill 1979).

Socio-ecological characterisation of plant species

The proposed method generalises the concept of fidelity, which Braun-Blanquet (1932) recognised as a fundamental feature. Originally, computation of fidelity of plant species had been applied to plot groups, with a faithful species being restricted to one single or a small group of vegetation types. Before computational algorithms became available in large databases, fidelity values were hardly ever reported. These algorithms were first applied to the relationship between plant species and climate (Brisse & Grandjouan 1977), thus expressing the apparent dependence of a plant species on an ecological condition, in this case a category of a climatic variable.

However, as traditional phytosociology data reported on species composition following a standardised method, but did not deliver corresponding environmental data such as the lime content of soils, it was not possible to calculate the fidelity of a plant towards calcareous soils. Long before the advent of large databases of joint vegetation plots and soils measurements like EcoPlant (Gégout et al. 2005) ecological indication based on species composition had to be achieved by calculating the fidelity of a plant species to-
Towards a calciphilous species (Brisse et al. 1995a, 1995b).

Generally speaking, this method considers all plant species as indicators of the environment. Without being able to gain data on the actual environment, it at least transforms a purely floristic characterisation into an implicit, yet quantitative ecological characterisation. This had been the devotion of Pavillard (1935) who had the vision of weighing all plant species. The same plant species is treated simultaneously with regard to its behaviour and as an indicator of the environment. A space of fidelities is defined with as many dimensions as there are indicator plants, i.e. a cross matrix linking 8,003 types of socio-ecological behaviour to 8,003 index variables (Fig. 3). Ewald (2002) has pointed out that a corresponding method termed "Beals smoothing" has been developed by Beals (1984). This fidelity table constitutes the backbone of the database allowing the ecological interpretation of phytosociological data.

Results on the ecology of plant species

The table of mutual fidelities of plant species (considered as index variables) has two applications. (1) The comparison of plant species behaviour, as measured by global differences in their behaviour in fidelity space, leads to a catalogue of ecologically similar species. (2) The relative importance of an indicator plant in overall compositional space or its diagnostic power is measured as the distance between its individual behaviour and the ensemble of behaviours captured by the database. The two catalogues can be viewed at the SOPHY website.

Among the 4,600 taxa that have thus been characterised many behave similarly. It is therefore desirable to summarise behaviour types by forming groups called "phytotypes".

Socio-ecological characterisation of plots

The site is the basic unit of observation in phytosociology. Each plot recorded at a site can be regarded as a sample of an environment of which the constituent plant species give testimony. The environment of the plot is situated at the centre of gravity of the behaviour of all constituent plant species. Thus, the environment of a plot containing $n$ plant species is composed of as many values of indices of variables (8,003 taxa and pseudo-taxa), and the plot position is defined by the average index of these $n$ plant species (Fig. 3).

The transformation of plots into environmental conditions yields a new table of 200,000 environments characterised by 8,003 average fidelities, of which 1,000 on average are larger than zero. The fact that each plot is characterised by the same number of numerical values allows mutual comparisons of the environments, even if they have largely different species richness, different plot sizes and have been recorded by different authors. It is even possible to compare plots that do not have a single species in common (Brisse et al. 1995a, 1995b). In fact the comparison of plots does no longer depend on the list of observed species but on the fidelities with respect to variable indices. It is no longer floristical, but has become ecological. Results concerning plot environments.
Probable flora

Average fidelities correspond to the probability of finding the environmental conditions at a site that is suitable to the plant species. A map of these fidelities for a given variable index displays the probability of finding a plant species at a site. Probability maps display concentrations of plant distributions in a very general fashion, thus visualising ecological gradients as well as many more sites that could be favourable to the plant species (Fig. 4). Likewise, species may be exposed to extinction risks at certain marginal sites, because average fidelities are too low.

Classification of 200,000 plots

As plant species are similar, so are numerous plots which reflect similar environments. This calls for a classification of plots (WPGM, Sokal & Sneath 1963). Meanwhile, several hundreds of thousands of plots cannot be classified straightaway. The number of objects has to be reduced. Whichever the method of choice, it requires the definition of some sort of kernels consisting of plots with maximum ecological similarity, which can be performed on a maximum of 15,000 objects within one day of computation. A trial classification was realised with 11,365 kernels, the classification of which yielded 890 types of environment (mesotypes). Table 1 shows the main environments identified based on the socio-ecological classification of all plots in the SOPHY database.

Discussion

General

One can now assemble a database that represents the knowledge that phytosociologists have accumulated since the discipline exists. It follows, that if the method matches their criteria and the database contains their knowledge, a good portion of the phytosociologists’ work can be replaced by a numerical treatment of the socio-ecological type. In fact, the socio-ecological classification meets the requirement of phytosociologists to found their discipline on their own ideas and achieve a hierarchy based on their own data. Database tools deliver a geographical representation of vegetation types, lists of their diagnostic species and a complete list of the constituent species. Furthermore, these tools allow to compare groups of the same level (twin groups) in order to clarify the reasons for their separation and to propose interpretations that are far more general because they account for the observations made by close to 2,000 botanists, more precise because they are numerical, more stable because of the completeness of the database, and more complete, because they treat different domains (geographical, ecological, floristical, phytosociological). It equally proposes criteria to define the most important groups in the hierarchy (absence of diagnostic plant species shared by two twin groups) as well as other criteria for stopping further subdivision of the hierarchy (insufficient distance of discriminant values of plant species in two twin groups). It also demonstrates that in the socio-ecology of plants there are no discrete limits, but only gradients.
Methods in phytoclimatology

In its origins, the SOPHY database was developed to study the relationships between plants and climate (temperature, precipitation) at the scale of France, in a period where French ecologists were more interested in plant-soil relationships. This aim was studied by means of a probabilistic calibration between 12,000 vegetation plots situated close to 574 climatic stations. The calibration measures the climatic optimum (position) and the indicator power (concentration) of 1,874 plant species for six climatic variables (monthly averages and extreme values of minimal and maximal temperatures, amount of precipitation and number of rainy days).

The probabilistic calibration takes into account three main ecological assumptions:

- unimodal response of plant species frequency along a climatic gradient (Fig. 5);
- gradual effects, even where plant species occur intermittently along the gradient (Fig. 6);
- ranking of indicator plant performance according to their concentration, i.e. if two plants are distributed in the same part of the range of a climatic variable, the most indicative plant is the one showing the highest frequencies at one or more levels of the range, even though the two plants may have the same optimum (Fig. 7).

Therefore, the calibration defines a position parameter of the plant in the range of the climatic variable, named ‘optimum’, and a dispersion parameter named ‘indicator power’ for each calibrated plant. These two parameters characterize the climatic behaviour of a plant by indicating its climatic position and the strength of the bond between the plant and the climatic variable. ‘Optimum’ and ‘indicator power’ are both expressed in percent in order to compare the climatic behaviour of plant taxa.

This calibration produces a list of 1,874 bio-indicators of climatic variables in France (Garbolino et al. 2007). The validation of this calibration is based on the difference between the measured climate by the meteorological stations and the climate estimated by the plants surrounding these stations. The result of this validation underlines that plants are accurate (accuracy = 88.5%) and stable (stability = 96.5%) bio-indicators of the climatic parameters in France.

The results show that the geographic distribution of some bio-indicators coincides with the distribution of some well-described climates in France. For example, Pistacia lentiscus indicates a Mediterranean climate characterised by warm and dry summers and autumns and mild and relatively rainy winters. This species has been found to be extremely indicative for high temperatures throughout the year, underlining the thermal aspect of the

---

Fig. 4: Example of a comparison between the observed plots of Viola biflora and its probable spatial distribution.

Fig. 5: Effect of a factor on the frequencies of a plant.

Fig. 6: Intermittence of plant’s frequencies in the range of an environmental variable.

Fig. 7: Comparison of two power indicators between an abundant plant and non-abundant plant.
Mediterranean climate (Fig. 8; see Plate A).

Impatiens parviflora, on the other hand, indicates a temperate subcontinental climate characterised by very low optima of the temperatures in winter and average and high optima in summer (Fig. 9). However, this recent neophyte of East-Siberian origin may not yet have filled its potential range.

Because this calibration between plants and climate is based on large data sets and a probabilistic model, it gives accurate information of the climatic behaviour of plants in France and an analytical underpinning of indicator values for temperature and continentality based on expert judgement by Ellenberg (1974) and Llandolt (1977). In doing this, it must be borne in mind that SOPHY covers only a fraction of climatic niche space.

Conclusion

The use of large vegetation databases allows to study the ecology of plant species and vegetation types, as well as to characterize their environments. But, even if the data are essential, the methodology to characterize the relationships among plant species and with respect to environmental parameters is decisive in understanding the ecology of plants. Respect for the nature of the data and the development of specific algorithms have been combined in the design of the SOPHY database. The obtained results underline the efficiency of applying numerical methods based on ecological assumptions.

Acknowledgements

The authors are very grateful to Prof. Jörg Ewald for its pertinent suggestions to improve the quality of this paper.

Fig. 8: Climatic behaviour of *Pistacia lentiscus* in France. The abscissa represents the months, starting from September to August. The ordinate represents the value of the climatic optimum of a plant according to the national average value of the variable. The colours represent the power indicator.

Fig. 9: Climatic behaviour of *Impatiens parviflora* in France.
Table 1: Hierarchy of the main environments for plants in France.

<table>
<thead>
<tr>
<th>Number of plots</th>
<th>Main ecological factor</th>
<th>Vegetation types</th>
</tr>
</thead>
<tbody>
<tr>
<td>52,731</td>
<td>Shadow</td>
<td>Forests of temperate Europe with sciaphilous plants</td>
</tr>
<tr>
<td>42,486</td>
<td>Full light</td>
<td>Meadows, grasslands of temperate Europe with heliophilous plants, including Mediter-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ranean forests</td>
</tr>
<tr>
<td>95,217</td>
<td>Temperate climate</td>
<td>All the vegetation of temperate Europe</td>
</tr>
<tr>
<td>36,376</td>
<td>Warm</td>
<td>Sub-Mediterranean calcareous environments</td>
</tr>
<tr>
<td>10,655</td>
<td>Cold</td>
<td>Mountains (see plate C)</td>
</tr>
<tr>
<td>10,313</td>
<td>Crops</td>
<td>Cultures</td>
</tr>
<tr>
<td>9,725</td>
<td>Salt</td>
<td>Salty environment, coastal and inland ones</td>
</tr>
<tr>
<td>7,222</td>
<td>Water</td>
<td>Aquatic environments (see plate B)</td>
</tr>
</tbody>
</table>

Plate: Vegetation types featured by the vegetation-plot database GIVD EU-FR-003.

A: Mediterranean scrub on acidophilic soil characterised by woody vegetation with many bushes. This vegetation type contains many protected species like *Serapis cordigera* (Photos: B. Bock et al.).

B: Mediterranean marshes: this type of environment is fairly represented in the French Mediterranean area. These environments are often rich in protected species like *Narcissus tazetta* (Photos: B. Bock et al.).

C: High mountains grasslands: this kind of vegetation formation is widely spread in the French territory. Some places contain endemic species like *Berardia subacaulis*, which is also a protected plant (Photos: B. Bock et al.).
References


Emmanuel Garbolino* (emmanuel.garbolino@mines-paristech.fr)

Crisis and Risk research Centre, MINES ParisTech

rue Claude Daunesse, BP 207
06904 Sophia-Antipolis, FRANCE

Patrice De Rufrray (patrice.derrufray@ibmp-upl.u-strasbg.fr)

IBMP

12, rue du Général-Zimmer

67084 Strasbourg cedex, FRANCE

Henry Brisse (brisse.henry@orange.fr)

Faculté des Sciences, MEP av. Escadrille Normandie-Niemen - Boite 441

13397 Marseille cedex 20, FRANCE

Gilles Grandjouan (†)

UMR 5554 Isém

place Eugène-Bataillon

34095 Montpellier cedex 5, FRANCE

*Corresponding author