STATISTICAL EVALUATION OF BIOCOENOTIC RECORDS: SOFIS[1] AND THE AUCHENORRHYNCHA FILE

Miroslava Dulova and Ilja Okali, Slovak Entomological Society, Bratislava, CSSR*

Abstract. A data file structure is described and the advantage of computer processing is discussed.

The experimental Auchenorrhyncha file consists of 3493 faunistical records, each defined by an entity and an attribute. An entity represents a concrete object. The entity "species" is defined by a Latin name, name of taxon's author and year of taxon's description. An attribute is a property of the entity, representing it qualitatively and quantitatively. The entity "species" may have attribute "locality" in which it was found. Alternatively, the entity "locality" may have attribute "species". Other attributes, describing the occurence of species in given localities and time were rarely recorded in the past. As a result of advances in automated record processing, it is proposed that a fuller set of attributes should be recorded and stored.

The Auchenorrhyncha file is an example. This file contains information on 378 species, out of an estimated 482 species in Slovakia. These species were found in 330 localities where the biotop is classified into four classes. Literature data abstracted from 61 sources are incorporated. The insect development stage is itself divided into 5 stages and the number of species is recorded in 4 possible ways. 83 collectors were involved during the period 1869-1980 and the material is stored.

The structure of the Auchenorrhyncha file under the SOFIS data base management system is as follows:

Species File

Species code (6)
Species name (60)

Localities File

Locality code (6):
Orographic unit code (3)
Locality code within that
orographic unit (3)
Plotting co-ordinates (4, 4)
Mapping square code (5)
Geographical co-ordinates (4, 4)
Altitude above sea level (8):
Minimum (4)
Maximum (4)
River basin code (8)

Biotopes File

Biotope code (6)
Biotope name (50)

Literature File

Literature code (10)
Bibliographical citation (60)

Faunistical Records File

Species code
Locality code
Biotope code
Collector code
Literature code
Deposit code
Living plant code
Date of finding (10)
Stage and species number code (12)
Altitude above sea level (4)
Collection method code (3)

Substrate code (3)

Living Plants File

Living plant code (6)
Living plant name (50)

Collectors File

Collector code (3)
Collector name (20)

Deposits File

Deposit code (7)
Deposit name (50)

---

(1) SOFIS is a database software product of the Computer Research Centre of Bratislava for mass data processing on the unified system and IBM using the operating system OS.
(*) Mailing address: Bazovského 14, 84101 Bratislava, Czechoslovakia.

Note 1: Numbers in brackets denote characters per item.

Note 2: Altitude in the faunistical records file is the

value at that point and falls within the range recorded in the localities file.
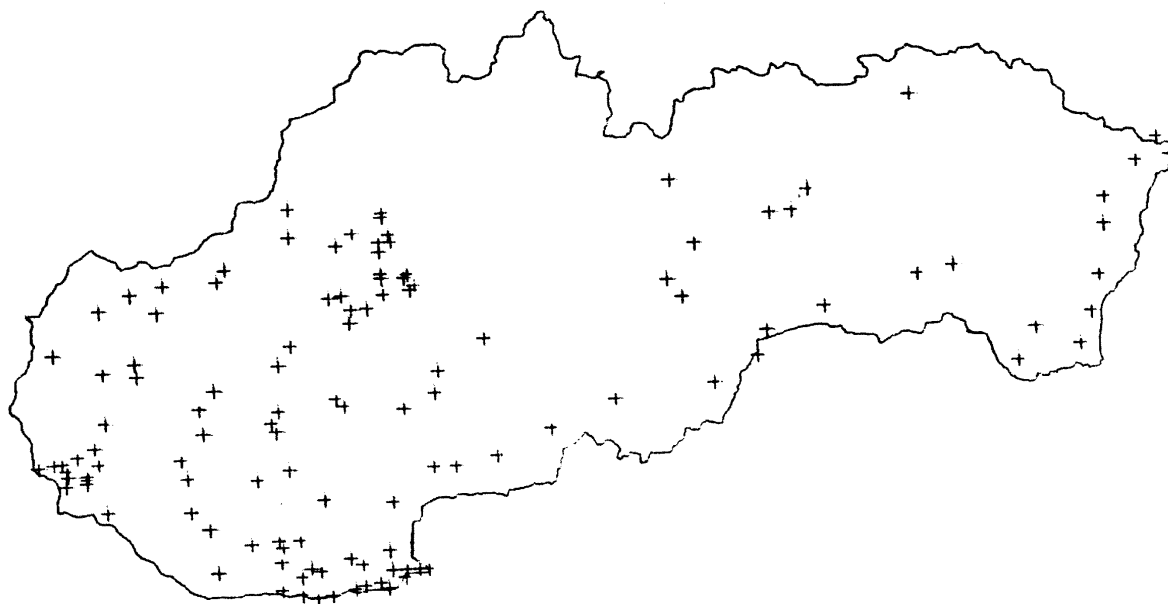
Note 3: A composite item is represented by: ":"

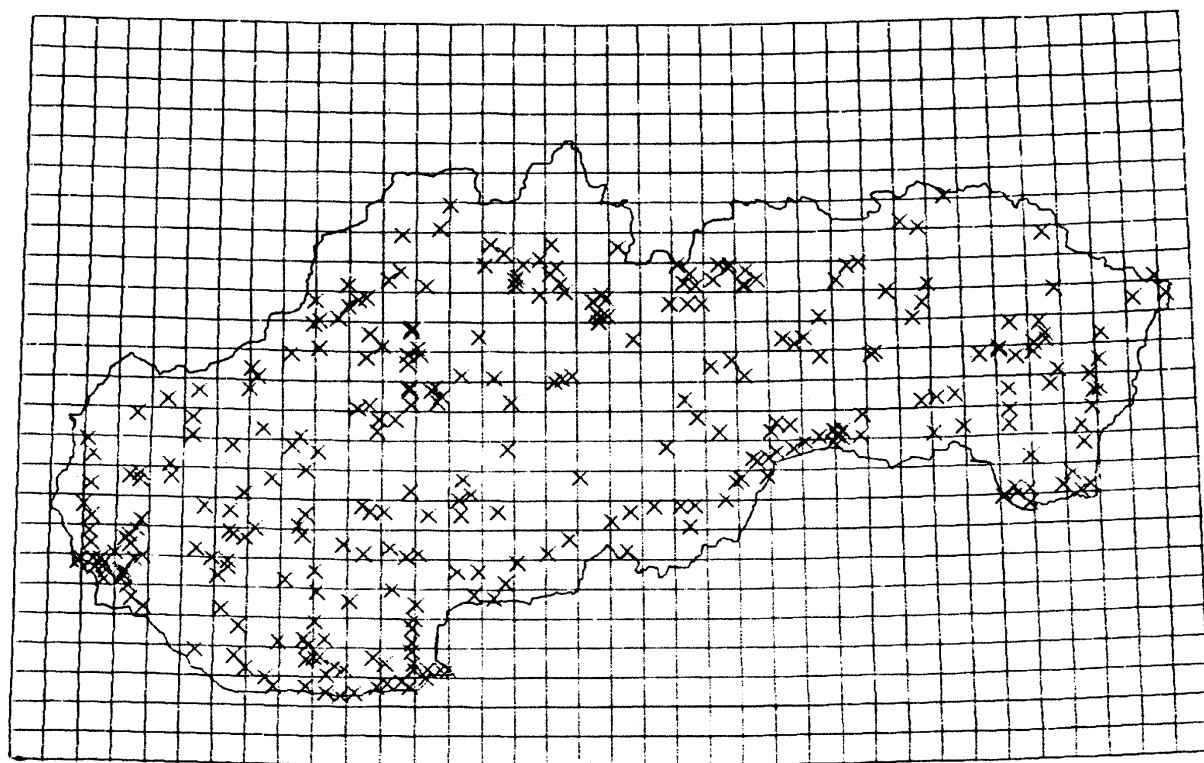Fig. 1. Collection sites in Slovakia for the period 1955-1975.

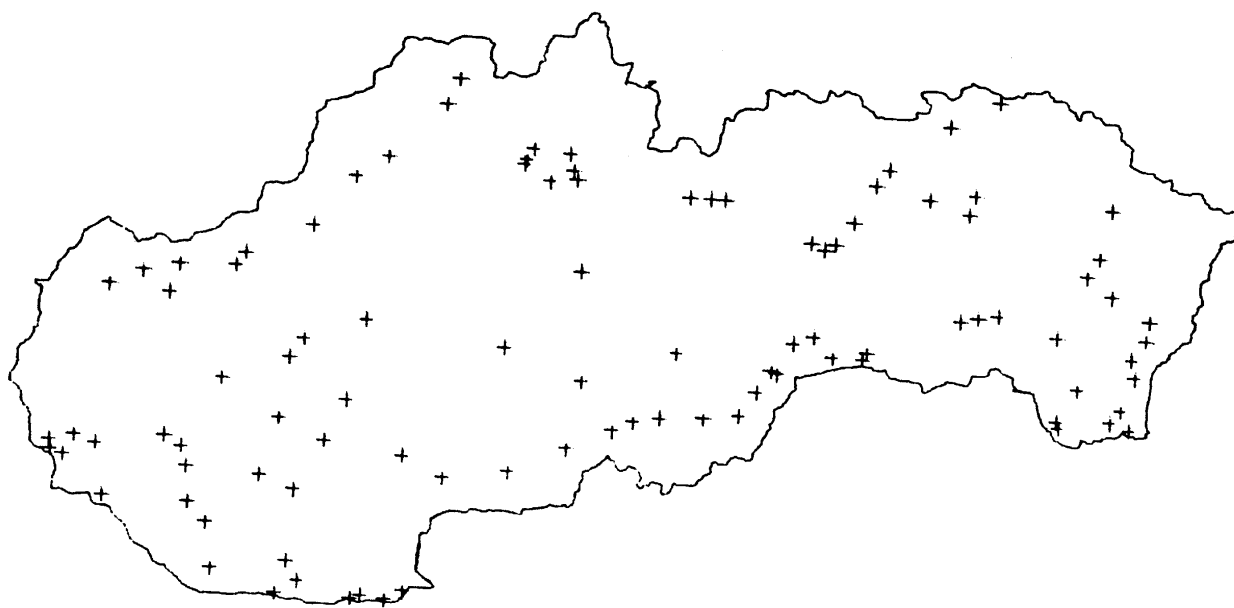Figure 2. Collection sites in Slovakia with mapping grid superimposed.

Fig. 3. Collection sites, collector 62.

The automatic processing results are represented by outputs from the faunistical records file. Some examples of types of output are presented. For each output the attribute specified in the search condition is stated, followed by other attributes contained in the faunistical record.

Faunistical records file:

- attribute "Species":
species code - locality code
species code - locality code - orographic unit code
species code - locality code - biotope code
species code - locality code - date of finding
species code - biotope code
species code - date

These combinations are considered as basic for evaluation of the time and space distribution, but other combinations are also possible.
- attribute "Locality"
locality code - species code
locality code - species code - mapping grid unit code
- date
- attribute "Biotope"
biotope name - species name
biotope code - species code
- attribute "Date"
date - species code
month - species code

The attributes "Literature", "Collector" and "Deposit" have become valid search criteria as a result of the automation of the data processing. These outputs can give information on presentation of individual species in specific journals, on a particular collector's activity, or on storage of materials in a particular museum collection.

From other files (Species, Biotopes, Localities, Literature, Deposits) so-called standard outputs of two basic types have been developed:
- systematic list (search criterion, code of entity)
- alphabetic list (search criterion, name of entity)

Graphical outputs are of particular importance for the evaluation of records. Twenty types of maps were created for the attributes "locality", "species", "biotope", "date", "deposit", "literature" and "collector".

Some results from automated processing of the Auchenorrhyncha file are presented below:

Attribute "species":

From 3493 faunistical records the most numerous species recorded were *Philaenus spumarius* (170), *Aphrodes bisinctus* (132), *Aphrophora alni* (104) and *Macrosteles laevis* (95) which together made up 14.34 per cent of all collections. These are not necessarily the most abundant species, however, just the most collected.

Attribute "locality":

The most common localities recorded were Bratislava (258), Bojnic (147), Sturovo (121) and Nitra (121), which together made up 18.75 per cent of the total collection. This does not mean that the observed species were most common here. It simply reflects the collectors choice of locality, which may have been influenced by many factors.

Attribute "biotope":

21.47 per cent of all records included this attribute in same form and in 89.68 per cent of these, it was specifically hierarchical class no 3: the meadow biotope.

Attribute "date":

Date was given for 35.76 per cent of all the faunistical records and 64.61 per cent of these were in the period 1955-1975, with the dates split into 5 periods of history, as follows: Pre-1918, 1918-1945, 1945-1955, 1955-1975, 1975-1985. In terms of individual years, 40.65 per cent of all records where the date was noted were collected in 3 years, 1956 (21.45 per cent), 1955 (10.00 per cent) and 1962 (9.20 per cent). In terms of seasons, the month of finding was specified in 27.36 per cent of all faunistical records, with 78.45 per cent of these being in June, July, or August. Again, this reflects the ability of the collectors to go out collecting, rather than the species distribution.

Attribute "literature":

Literature was given in 71.74 per cent of all records, of which 32.11 per cent came from one work (Dlabola).

Attribute "deposit":

Nine depositories are represented by 27.75 per cent of faunistical records, of which 37.11 per cent are from the Slovak National Museum (the Central Museum Institution of the Slovak Socialist Republic, with approximately 75 per cent of the collections in Slovakia.

The methodology of systematic automated evaluation of insects, based on experience with the Auchenorrhyncha file, includes the preparation of basic materials, the structure of a formalised attribute description, a software solution of a natural science file, and evaluation from different viewpoints. The proposal that a full set of attributes should be recorded, concerning natural science material, carries with it suggested attributes:

Species (name, code)

Locality (name, code, mapping grid unit, geographical co-ordinates, plotting co-ordinates, maximum and minimum elevation above sea level for the region, altitude of the finding, river basin code)

Biotope (name, code)

Date (complete year-month-day)

Collector (name, code)

Stage and number of species.

Also included are the literature data (code, citation), name of deposit, living plant, method of collection, and substrate, if known.

The experimental Auchenorrhyncha file is one of the first, concerning faunistical records processed by databank technology linked to a graphical output. If the method of systematic automated data evaluation of insects, presented here, were to be used in all institutions processing this type of material, it would be possible to process in a short period of time all the data necessary for a synthetic study.