

Application of PCA in Taxonomy Research – Thrips (Insecta, Thysanoptera) as a Model Group

Halina Kucharczyk¹, Marek Kucharczyk²,
Kinga Stanisławek¹ and Peter Fedor³

¹*Department of Zoology, Maria Curie-Skłodowska University, Lublin,*

²*Department of Nature Conservation, Maria Curie-Skłodowska University, Lublin,*

³*Department of Ecosozology, Faculty of Natural Sciences, Comenius University, Bratislava,*

^{1,2}*Poland*

³*Slovakia*

1. Introduction

Phenetic taxonomy is based on analysis of many unweighted characters. The number of variables that can be analyzed for a plant or animal species is so high that it is necessary to use a mathematical tool for grouping them into units corresponding to taxa.

Principal Component Analysis enables researchers to reduce the number of possible groupings. It is significant due to the occurrence of some redundancy in variables. In this case, redundancy means that some of the variables are correlated with one another, because they are measuring the same construct.

Principal Component Analysis replaces many original characters with only a few most significant principal components (PCs) which represent combinations of closely correlated original characters.

Principal Component Analysis was first described by Pearson (1901); in the 1930's Hotteling (1933) prepared a fully functional method that generates a set of orthogonal axes, placed in decreasing order and determining the main directions of variability of samples.

In analysis of principal components, eigenvalues represent the relative participation of each principal component in presenting the general variability of sampled material. The numerical value of a given eigenvalue is a direct indicator of the weight of a particular component in the general characteristics of the variability of a set of data. In practice, the distribution of the elements of the analyzed set in the space of the first three or four components allows one to present almost the complete diversity of the set.

Eigenvectors are sets of numbers which show the weights of individual characters for each principal component. Like the correlation coefficient, eigenvalues are scaled from -1 to +1. The higher the value, the more closely a given trait is connected with a component. On the

basis of eigenvectors it is then possible to interpret principle components, e.g. determine which character (or characters) are the most representative.

Before principal components are determined, the data can be processed in many ways. It is necessary if variables (features) are expressed in different units or the range of their variability is different. The two methods used are centering and standardization. In the former, the variable axes are moved so that the beginning of one axis is in the centre of inertia of the axis. Standardization of data involves changing the values of characters in such a way that their mean is 0 and standard deviation is 1.

If analysis covers calculating the covariance matrix, the variables are centred; if, on the other hand, the correlation coefficient matrix needs to be calculated, the variables are centred and standardized at the same time.

The Principal Component Analysis (PCA) is widely used in taxonomic research of plants and animals (Apuan et al., 2010; Chiapella, 2000; Kucharczyk & Kucharczyk 2009; Lilburn & Garrity, 2004; Sahuquillo et al., 1997; Wolff et al., 1997; Vela et al., 2007). Below we present application of PCA in taxonomy of the order Thysanoptera (Insecta). The numerical calculations were made with the Multi-Variate Statistical Package (MVSP) (Kovach, 2005) and Statistica PL, version 6 (StatSoft Inc. 1984-2001).

1.1 Characteristic of the order Thysanoptera

Up to now almost 6000 species of the order Thysanoptera have been described worldwide but many other are added to this list every year. The asymmetric, with only one left mandibule, mouth cone is a synapomorphic character state which differs thrips from the other insect orders (Mound et al., 1980; Mound & Morris, 2004, 2007). In currently accepted systematic position the order Thysanoptera is divided into two suborders: Terebrantia and Tubulifera. The former consists of eight families which comprise very small and tiny insects, most of them reaching 1-3 mm in length. They are mostly herbivorous insects feeding both on dicotyledonous and monocotyledonous plants, only a small part of them are facultative or obligate predators (fam. Aeolothripidae). The family Thripidae with nearly 2,500 species is the largest within Terebrantia but the relationships within this group is not clear (Mound & Morris, 2007). Only one tubuliferan family – Phlaeothripidae with almost 3,500 known species includes larger thrips, the biggest, mostly tropical taxa, reach up to 15 mm. Excepting herbivores most of them live on dead wood or in leaf litter and feed on fungi. Some phytophagous species are regarded as pests, feeding and breeding on different parts of plants they cause deformations of leaves, flowers and fruits, and in the final result stop their development. A limited number of terebrantians may transmit fungi, bacteria and viruses, which may infect the host plants reducing the quality of yields and their market value (Lewis, 1997; Tommasini & Maini, 1995).

Originally taxonomy referred to the description and naming of the organisms (alpha taxonomy). Currently it is the science based on different fields of knowledge and that uses various tools to classify the organisms and determine the relationships amongst them. There are two systems of the Thysanoptera order classification: phenetic based mainly on the morphological characteristics of adults' specimens and phylogenetic one based on the evolutionary relationships (Mound, 2010). The former one is more practical and widely used

in construction of the identification keys. The latter based on the molecular and genetic data is not satisfactory enough because of insufficient data. Therefore in practice, morphology and other biological aspects, e.g. observations of developmental stages and relations with host plants provide more data and may be useful in comprehension of the relationships amongst taxa (Crespi et al., 2004; Mound & Morris, 2007).

Some disagreements exist concerning classification system of Thysanoptera on different taxonomic levels. Because of the great differences in body structures between Terebrantia and Tubulifera Bhatti (1988) proposed to raise them to the order rank in new superorder Thysanopteroidea composed of 40 families. In the next works this author divided Terebrantia into 28 families (Bhatti, 2005, 2006). However, this classification is not accepted by most of thysanopterologists now. On the other hand the current state of Terebrantia with eight families has not been taken under consideration by zur Strassen (2003) in his latest key. In contrast to the currently accepted division zur Strassen classified the species of the genera *Melanthrips* and *Ankothrips* into the family Aeolothripidae. Many revisions at the genus level took place in the past, e.g. changes within the genera: *Thrips* Linnaeus 1758 and *Taeniothrips* Amyot & Serville 1843, *Mycterothrips* Trybom 1910 and *Taeniothrips*, *Anaphothrips* Uzel 1895 and *Rubiothrips* Schliephake 1975 etc. (Strassen zur 2003).

The correct identification of specimens to the species level is a basis for further taxonomic study. Often there are many problems with recognition of adults, as well as immature stages because of diversity of variation within and between species, particularly in the species rich genera, e.g. *Thrips* (Terebrantia) and *Haplothrips* Amyot & Serville 1843 (Tubulifera). These genera include mainly the Holarctic species feeding and breeding on dicotyledonous plants, though occasionally they are graminicolous. Very numerous species representing both of them may suggest that these genera are relatively young in evolutionary history. Morphologically, many species are very similar and are treated by some researchers as the same species in two forms, e.g. *Thrips atratus* Haliday 1836 and *Thrips montanus* Priesner 1921, *Thrips sambuci* Heeger 1854 and *Thrips fuscipennis* Haliday 1836, *Thrips fuscipennis* and *Thrips menyanthidis* Bagnall 1923, *Haplothrips leucanthemi* (Schrank 1781) and *Haplothrips niger* (Osborn 1883), *Haplothrips tritici* (Kurdjumov 1912) and *Haplothrips cerealis* Priesner 1939 (Mound & Minaei, 2007, 2010; Strassen zur, 2003).

There are even more problems with recognizing larval stages. The larvae are less mobile therefore have a stronger relationship with their host plants. Morphological dissimilarity amongst larvae of different species are often larger than amongst their adults (Kucharczyk, 2010). So, the detailed study both on adults and immature stages with application of the statistical tools, within PCA method, allows to solve several problems in the taxonomy of thrips and at least partially explain their phylogenetic relationships.

2. Application of Principal Component Analysis method in taxa recognizing

2.1 Identification of the second larval instar of the *Haplothrips* genus species

The Principal Component Analysis method (PCA) may be useful in selecting from among the great number of morphometric characters, especially those that have some taxonomical value. Such a necessity is occurring within genera which species are very uniform in morphological structure and there are weak qualitative characters differentiating them. The only potential differences are related to the measurements of some body parts. The genus

Haplothrips (Thysanoptera, Phlaeothripidae) is a good example for the discussed case. To this genus belong species that are very similar in body structure and therefore they are difficult to identify.

The *Haplothrips* genus is one of the most numerous in species of the Phlaeothripidae family (about 230 species). It is distributed worldwide but mainly in northern hemisphere, almost 70% known hitherto species of this genus have been noted in Holarctic. Most of them are phytophagous, feeding and breeding in flowers of dicotyledons, mostly of *Asteraceae* family, only a few taxa are connected with monocotyledons (Mound & Minaei, 2007; Pitkin, 1976; Zawirska & Wałkowski, 2000). Because of the fact that the research on larval morphology is very scarce (especially on species belonging to Phlaeothripidae) the morphological analysis of the second instar larvae of *Haplothrips* species have been undertaken.

The measurements were made for 165 specimens of larvae belonging to 11 species (each species was represented by 15 individuals) collected from various habitats and regions of Poland. Among examined species were: *Haplothrips aculeatus* (Fabricius 1803) (inhabitant of various *Poaceae* plants, the ubiquitous species); *H. tritici* (Kurdjumov 1912) (also *Poaceae*-related species, but mainly with *Triticum vulgare*; occurring in south-eastern part of Europe and Asia Minor); *H. setiger* Priesner 1921 (the polyphagous species living on flowers of *Asteraceae* for example *Senecio fuchsii*, *Crepis* spp., *Matricaria* spp., *Achillea* spp., *Anthemis arvensis*; having Palaearctic range of occurrence); *H. subtilissimus* (Haliday 1852) (present on leaves of bushes and trees such as *Quercus* spp. and *Fagus* spp.; considered as predatory species on mites, lepidopteran eggs and larval stages of *Coccidae* and *Aleyrodidae*; noted from Europe). The rest of species are regarded as mono- or oligophags: *H. arenarius* Priesner 1920 (especially connected with *Helichrysum arenarium*; with Palaearctic range); *H. dianthinus* Priesner 1924 (on *Dianthus carthusianorum*; noted from central and south Europe); *H. jasionis* Priesner 1950 (on *Jasione montana*; known from Europe only); *H. leucanthemi* (Schrank 1781) (on *Leucanthemum* spp.; reaching Holarctic, Oriental and Australian regions); *H. setigeriformis* Fabian 1938 (on *Potentilla argentea*, known from Europe and Near East); *H. stacies* (Haliday 1836) (on *Armeria elongata*; occurring in Europe) and *H. angusticornis* Priesner 1921 (connected with *Asteraceae* but mainly with *Achillea millefolium*; noted from Palaearctic).

Each individual of larva have been measured in respect of 72 potentially important features, most of them concerned lengths of selected both dorsal (d) and ventral (v) setae on all parts of body (h – head, pro – pronotum, mes – mesonotum, meta – metanotum), distances between setae, measurements of apical abdominal segments (8-11) and antennal segments (ant, III-VII). Finally, the data matrix consisted of 11880 measurements was constructed (comprised of 72 character states of 165 individuals belonging to 11 species) (Fig. 1).

The specimens were ordinated along first two PCA axes (transformed data). The results of PCA showed that the cumulative variance of the two principal components reached 62.4%: Axis 1 - 38.5%, Axis 2 - 23.9%. Figure 2 shows two groups of specimens belonging to different species which are clearly isolated (*H. aculeatus* and *H. subtilissimus*). The most discriminative features with the highest eigenvalues are the lengths of abdominal setae (9-d2, 9-v2, 11-d1, 11-v2). On the mentioned PCA graph the most numerous group of specimens, belonging to the nine other and listed above species, is creating compact cloud. To choose more selective character states discriminating the rest of examined species we can remove the measurements of the two first separated species (*H. subtilissimus* and *H. aculeatus*) from the primary data matrix.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
1	acul1	ant-d1	ant-d2	ant-d3	ant-d4	ant-d5	ant-d6	ant-d7	ant-d8	ant-d9	ant-d10	ant-d11	ant-d12	ant-d13	ant-d14	ant-d15	ant-d16	ant-d17	ant-d18	ant-d19	ant-d20	ant-d21	ant-d22	ant-d23	ant-d24	ant-d25	ant-d26	ant-d27	ant-d28
2	acul1	230.0	25.0	20.0	32.5	0.8	30.0	21.3	1.4	50.0	22.5	2.2	47.5	25.0	1.9	32.5	22.5	1.4	25.0	17.5	1.4	22.5	8.8	2.6	10.0	12.5	5.0	10.0	13.8
3	acul2	220.0	22.5	20.0	30.0	0.7	30.0	21.3	1.4	47.5	22.5	2.1	47.5	23.8	2.0	32.5	22.5	1.4	22.5	15.0	1.5	22.5	8.8	2.6	10.0	13.8	5.0	11.3	17.5
4	acul3	230.0	22.5	20.0	32.5	0.8	30.0	20.0	1.5	50.0	22.5	2.2	47.5	25.0	1.9	35.0	22.5	1.6	25.0	16.3	1.5	22.5	7.5	3.0	10.0	12.5	5.0	11.3	15.0
5	acul4	205.0	22.5	17.5	27.5	0.6	27.5	18.8	1.5	45.0	21.3	2.1	42.5	22.5	1.9	30.0	21.3	1.4	22.5	15.0	1.6	20.0	7.5	2.7	7.5	12.5	5.0	10.0	12.5
6	acul5	225.0	22.5	22.5	28.8	0.8	30.0	21.3	1.4	50.0	21.3	2.4	45.0	22.5	2.0	32.5	20.0	1.6	25.0	15.0	1.7	22.5	7.5	3.0	10.0	12.5	7.5	10.0	15.0
7	acul6	205.0	22.5	20.0	30.0	0.7	27.5	20.0	1.4	45.0	20.0	2.3	40.0	22.5	1.8	27.5	20.0	1.4	20.0	16.3	1.2	20.0	8.8	2.3	10.0	12.5	6.3	10.0	12.5
8	acul7	225.0	22.5	25.0	32.5	0.8	30.0	21.3	1.4	50.0	22.5	2.2	45.0	25.0	1.8	32.5	22.5	1.4	25.0	16.3	1.5	22.5	8.8	2.6	10.0	12.5	5.0	10.0	15.0
9	acul8	235.0	22.5	20.0	30.0	1.0	30.0	22.5	1.3	50.0	22.5	2.2	47.5	23.8	2.0	35.0	22.5	1.6	25.0	16.3	1.5	25.0	8.8	2.9	8.8	12.5	5.0	10.0	12.5
10	acul9	225.0	22.5	25.0	30.0	0.8	30.0	20.0	1.5	52.5	22.5	2.3	45.0	25.0	1.8	35.0	22.5	1.6	25.0	15.0	1.7	22.5	7.5	3.0	7.5	12.5	5.0	8.8	13.8
11	acul10	200.0	22.5	22.5	27.5	0.8	27.5	20.0	1.4	42.5	20.0	2.1	42.5	21.3	2.0	30.0	18.8	1.6	25.0	16.3	1.5	17.5	7.5	2.3	8.8	12.5	3.8	10.0	13.8
12	acul11	220.0	22.5	25.0	30.0	0.8	30.0	20.0	1.5	42.5	22.5	1.9	47.5	25.0	1.9	37.5	22.5	1.7	25.0	15.0	1.7	22.5	7.5	3.0	10.0	12.5	5.0	11.3	12.5
13	acul12	215.0	22.5	25.0	30.0	0.8	30.0	20.0	1.5	47.5	22.5	2.1	42.5	25.0	1.7	32.5	22.5	1.4	25.0	15.0	1.7	20.0	7.5	2.7	7.5	12.5	5.0	11.3	13.8
14	acul13	215.0	22.5	20.0	30.0	0.7	30.0	21.3	1.4	47.5	22.5	2.1	42.5	25.0	1.7	30.0	22.5	1.3	22.5	15.0	1.5	20.0	8.8	2.3	7.5	12.5	5.0	10.0	15.0
15	acul14	215.0	22.5	20.0	30.0	0.7	27.5	20.0	1.4	47.5	21.3	2.2	42.5	22.5	1.9	32.5	20.0	1.6	22.5	15.0	1.6	20.0	8.8	2.3	10.0	12.5	5.0	11.3	12.5
16	acul15	210.0	25.0	17.5	30.0	0.6	27.5	20.0	1.4	45.0	20.0	2.3	42.5	22.5	1.9	32.5	20.0	1.6	22.5	15.0	1.5	20.0	8.8	2.3	7.5	12.5	5.0	11.3	12.5
17	aren1	205.0	10.0	20.0	27.5	0.7	27.5	21.3	1.3	35.0	21.3	1.6	37.5	22.5	1.7	35.0	20.0	1.8	27.5	15.0	1.8	20.0	7.5	2.7	5.0	8.8	3.8	6.3	8.3
18	aren2	205.0	15.0	20.0	32.5	0.6	30.0	22.5	1.3	37.5	22.5	1.7	37.5	22.5	1.7	35.0	20.0	1.8	27.5	15.0	1.8	22.5	7.5	3.0	6.3	10.0	2.5	7.5	7.5
19	aren3	210.0	15.0	20.0	32.5	0.8	32.5	22.5	1.4	37.5	22.5	1.7	40.0	22.5	1.8	35.0	20.0	1.8	27.5	15.0	1.8	25.0	7.5	3.3	6.3	10.0	5.0	8.8	10.0
20	aren4	205.0	10.0	20.0	27.5	0.7	27.5	20.0	1.4	40.0	21.3	1.9	37.5	22.5	1.7	32.5	20.0	1.6	27.5	15.0	1.8	22.5	7.5	3.0	5.0	10.0	3.8	7.5	7.5
21	aren5	200.0	10.0	20.0	27.5	0.7	30.0	20.0	1.5	35.0	22.5	1.6	37.5	22.5	1.7	32.5	22.5	1.4	27.5	15.0	1.9	22.5	8.8	2.6	6.3	8.8	3.8	7.5	8.8
22	aren6	211.0	12.4	21.8	32.3	0.7	30.0	22.1	1.4	36.3	23.9	1.5	38.3	24.5	1.6	33.8	21.5	1.6	28.0	16.7	1.7	23.9	9.8	2.4	5.0	8.7	4.0	7.0	8.3
23	aren7	227.7	15.0	23.9	32.2	0.7	33.0	23.5	1.4	42.7	24.0	1.8	43.2	24.6	1.8	37.5	22.4	1.7	29.8	16.7	1.8	24.9	9.5	2.6	5.9	10.7	5.3	9.0	9.2
24	aren8	213.8	14.0	20.4	29.3	0.7	30.2	21.8	1.4	36.9	22.7	1.6	40.0	24.5	1.6	36.2	21.1	1.7	29.9	15.8	1.9	23.3	8.5	2.7	5.8	8.1	3.3	6.5	7.3
25	aren9	199.5	9.8	21.1	28.8	0.7	29.5	20.0	1.4	34.8	21.5	1.6	35.3	22.2	1.8	32.8	18.6	1.8	27.3	14.7	1.9	22.7	9.0	2.5	5.1	7.8	3.7	6.6	9.9
26	aren10	207.9	15.0	20.1	30.9	0.7	31.0	21.9	1.4	32.9	24.0	1.4	40.8	25.6	1.6	35.6	22.5	1.6	28.9	15.7	1.9	23.9	9.9	2.7	5.7	9.9	4.5	7.4	9.7
27	aren11	197.8	9.8	19.9	29.7	0.7	29.9	20.6	1.3	35.0	21.9	1.6	36.2	22.4	1.6	31.1	19.9	1.6	28.3	15.8	1.8	23.2	9.2	2.5	5.3	7.4	4.8	6.0	7.5
28	aren12	205.8	13.5	19.3	28.7	0.7	30.0	21.6	1.4	34.8	24.0	1.5	35.6	24.3	1.5	33.2	20.6	1.6	29.5	16.0	1.8	23.0	8.8	2.6	6.0	8.9	4.8	8.2	8.9
29	aren13	204.6	10.0	18.8	28.9	0.6	29.2	21.0	1.4	36.8	21.1	1.7	37.0	22.9	1.6	31.9	20.3	1.6	28.0	15.5	1.8	25.3	8.8	2.9	5.0	8.5	4.6	7.0	6.9
30	aren14	201.5	9.8	19.0	28.4	0.7	29.2	20.0	1.5	36.1	20.2	1.6	38.7	21.8	1.8	32.4	19.2	1.7	27.0	14.0	1.9	21.9	8.9	2.5	6.3	9.2	4.0	7.4	8.2
31	aren15	214.2	13.3	21.3	30.9	0.7	29.8	22.3	1.3	37.0	22.9	1.6	38.8	24.7	1.6	34.8	20.8	1.7	29.5	16.2	1.9	25.0	8.3	3.1	4.6	8.3	3.8	8.2	9.0
32	dianth1	265.0	17.5	22.5	35.0	0.6	35.0	23.8	1.5	57.5	25.0	2.3	60.0	25.0	2.0	45.0	22.5	2.0	32.5	17.5	1.9	22.5	10.0	2.3	7.5	12.5	2.5	10.0	10.0
33	dianth2	265.0	25.0	22.5	35.0	0.6	37.5	22.5	1.7	57.5	25.0	2.3	60.0	27.5	1.8	45.0	25.0	1.8	32.5	17.5	1.9	22.5	10.0	2.3	7.5	12.5	2.5	10.0	10.0
34	dianth3	272.5	17.5	25.0	35.0	0.7	37.5	22.5	1.7	60.0	25.0	2.4	62.5	26.0	2.1	42.5	22.5	1.9	35.0	17.5	2.0	22.5	8.8	2.6	5.0	12.5	2.5	7.5	10.0
35	dianth4	270.0	17.5	25.0	37.5	0.7	37.5	22.5	1.7	60.0	25.0	2.4	62.5	26.0	2.1	45.0	22.5	2.0	35.0	17.5	2.0	22.5	8.8	2.6	7.5	12.5	2.5	10.0	10.0
36	dianth5	275.0	20.0	25.0	35.0	0.7	35.0	23.8	1.5	57.5	25.0	2.3	60.0	25.0	2.2	45.0	23.8	1.9	32.5	17.5	1.9	22.5	10.0	2.3	7.5	12.5	3.8	9.8	10.0
37	dianth6	260.0	22.5	25.0	33.8	0.7	35.0	23.8	1.5	62.5	26.3	2.4	62.5	27.5	1.9	42.5	23.8	1.8	32.5	17.5	1.9	20.0	10.0	2.0	7.5	12.5	3.8	7.5	12.5
38	dianth7	270.0	20.0	22.5	35.0	0.6	35.0	25.0	1.4	60.0	25.0	2.4	62.5	26.3	2.0	42.5	23.8	1.8	35.0	17.5	2.0	22.5	10.0	2.3	7.5	12.5	2.5	7.5	10.0
39	dianth8	270.0	20.0	22.5	40.0	0.6	37.5	23.8	1.6	62.5	25.0	2.5	62.5	26.0	2.1	43.8	22.5	1.9	33.8	17.5	1.9	22.5	10.0	2.3	7.5	12.5	5.0	10.0	10.0
40	dianth9	265.0	22.5	22.5	37.5	0.6	37.5	22.5	1.7	60.0	25.0	2.4	60.0	25.0	2.0	42.5	22.5	1.9	32.5	17.5	1.9	22.5	10.0	2.3	7.5	12.5	3.8	10.0	7.5

Fig. 1. A piece of data matrix comprising (in the whole) 72 metric characters of the second larval instar of eleven *Haplothrips* species

The result of the next PCA is separation of the next group of individuals belonging to *H. dianthinus* species (Axis 1 – 40.5%, Axis 2 – 18.1%; Fig. 3). The most discriminative features are short setae 9-v1 and 11-d1. There are some additional selective features (e.g. 8-v2, h-s2) but they are less significant. After elimination of *H. dianthinus* data from the matrix next two species are emerging: *H. arenarius* and *H. angusticornis* (Fig. 4) and three main characters are discriminated (setae: 8-v2, 9-d2, 9-v2). Now we can establish the value-range of selective features for examined species both with the help of data matrix and PCA graph.

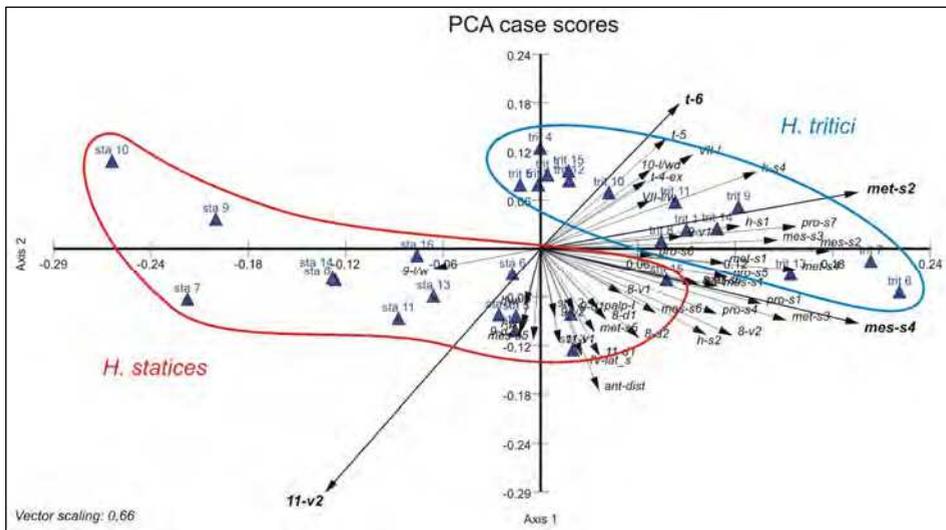


Fig. 7. PCA – scatter diagram of *H. tritici* and *H. statures* specimens as OTUs along PC1 and PC2 based on 72 quantitative features

Next steps of the analysis are similar to the previous ones (Figs. 5-7). During the analysis, it may turn out, that some specimens have ambiguous position (like “jas3, jas6, jas8” Fig. 5) and this may be a hint to check the data of the doubtful ones. At the end of the elimination, two species which are the most resembling (*H. tritici* and *H. statures*) remain. The selected distinguishing features can be used to construct an identification key. Additional benefit of presented method is elimination of characters which have no taxonomical value (in this case, out of 72 tested features there are 22 unimportant states). That would speed up further measurements of more specimens and improve the precision of determined features range.

2.2 Identification of the second larval instar of the *Thrips* genus species

The study on the morphology of the second larval instar of Central European *Thrips* species is another example of using PCA method in taxonomic research (Kucharczyk, 2010). In contrast to the previously discussed larvae of the genus *Haplothrips*, larvae of 34 researched species of the genus *Thrips* may be recognized mainly on the basis of qualitative characters. The data matrix covered 26 multistate discontinuous characters, amongst them the most important related to the sclerotisation and sculpture of integument, the structure of spiracles and antennae (Tab. 1). This analysis was conducted in two steps. At the beginning all researched specimens were treated as operational taxonomic units (OTUs) which have been characterized by 26 variables. On the graph OTUs form the clouds which are corresponding to the studied species. During the second step not specimens but species were treated as OTUs. In these cases the PCA was applied as a method for ordination and reducing the number of variables, the characters which had the highest loadings to PC1 and PC2 were extracted and they were regarded as the

most important and useful in distinguishing the studied taxa (Fig. 8). Finally, these selected features have been used in constructing the identification key to second larval instar of studied *Thrips* species.

The characteristics tested in PCA method have been also used as variables in Cluster Analysis (CA). The results in the two-dimension ordination of PCA were consistent with the results in the hierarchical clustering analysis (Fig. 9). The results of numerical analysis sheds some new light on the relationships amongst studied species. Moreover, the received dendrogram showed the similarity within studied *Thrips* species and allowed to propose the ancestral (plesiomorphic) and advanced (apomorphic) characteristics of immature stage which have not been studied hitherto.

Feature	Abbrev.	Feature	Abbrev.
head sclerotization	HPLATE	sculpture of tergite VIII	T8SCP
pronotum sclerotization	PRPLATE	tergite IX - sculpture anterior to and between campaniform sensilla (Cs)	T9SCP
mesonotum sclerotization	MSPLATE	tergite IX - sculpture between setae level and Cs	T9DSCP
metanotum sclerotization	MTPLATE	tergite IX - sculpture, teeth between D1 setae	T9D1TH
tergite IX sclerotization	T9SCL	sculpture of tergite X	T10SCP
tergite X sclerotization	T10SCL	microtrichial comb at base of hind legs	COXAT
sternite XI sclerotization	S11SCL	spiracles on abdominal segment II	AT2SP
posteromarginal comb on tergite IX	T9COMB	number of facets in spiracle on abdominal segment II	NOOFF
length of teeth on posteromarginal comb	LCOMB	width of spiracle on abdominal segment II	WOFSP
sculpture of mesonotum	MSSCP	microtrichia on antennal segment III	ANS3MT
sculpture of metanotum	MTSCP	length of trichome on antennal segment VI	ANS6TR
sculpture of tergites III-VII	T3-7SC	shape of abdominal tergal setae D1, D2	TSETA
sculpture of sternites III-VII	S3-7SC	furca on metasternum	FURCA

Table 1. Features of second larval instar of *Thrips* species used in numerical analyses (after Kucharczyk, 2010)

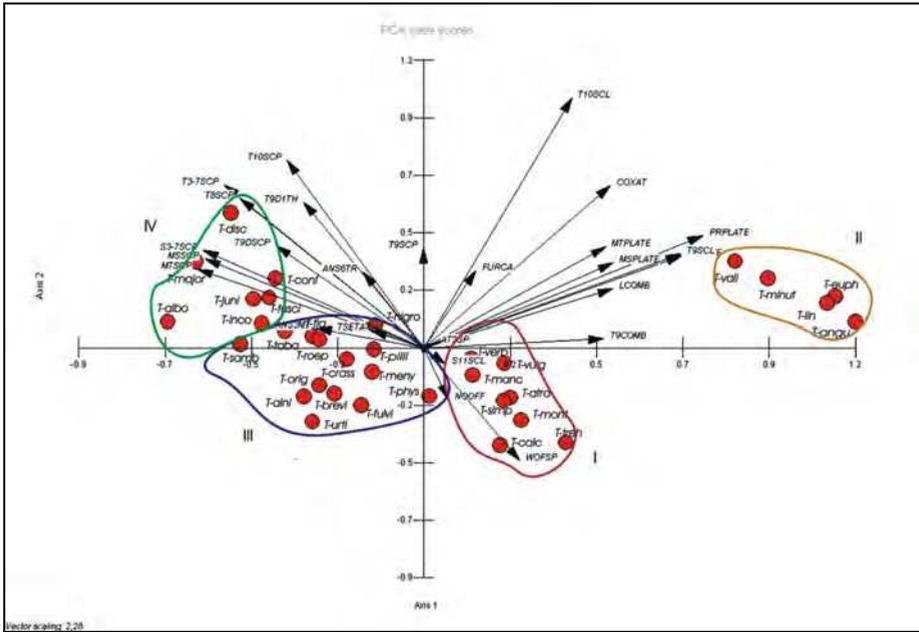


Fig. 8. PCA - scatter diagram of *Thrips* genus species as OTUs along PC1 and PC2 based on 26 qualitative features (names of species abbreviated, abbreviations of characters as in Table 1)

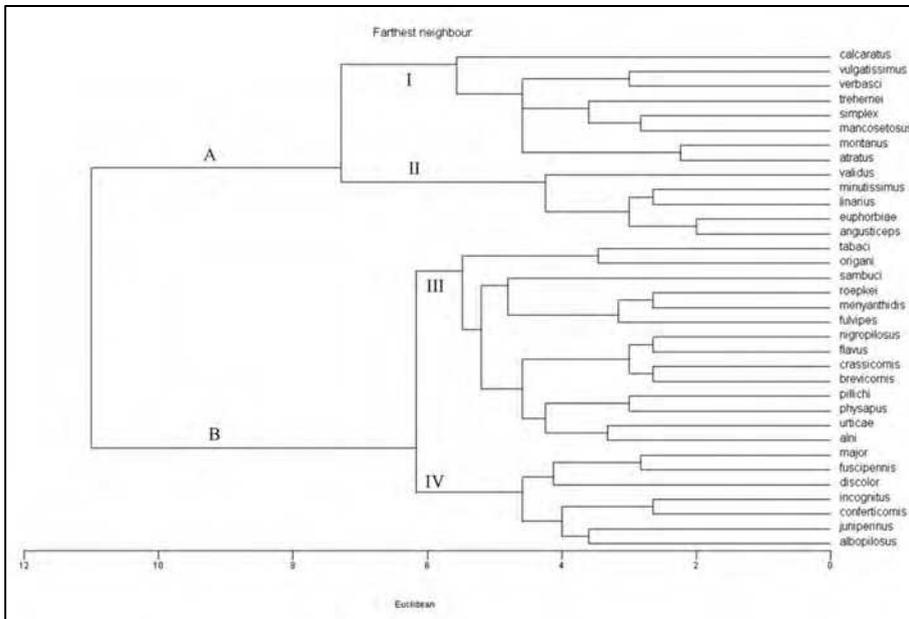


Fig. 9. Dendrogram (CA) of the similarity amongst analyzed *Thrips* species based on the morphological characters of the second larval instar (after Kucharczyk, 2010)

2.3 Identification of *T. atratus* Haliday 1836 and *T. montanus* Priesner 1920 species

For the first time Kucharczyk & Kucharczyk (2009) have used PCA method in the study on two similar *Thrips* species: *T. atratus* and *T. montanus*. Since the description, both of them were re-classified to *Thrips* Linnaeus 1758, *Taeniothrips* Amayot et Serville 1843 or *Similothrips* Schliephake 1972 genera (Priesner, 1964; Schliephake, Klimt, 1979; Schliephake, 2001). Zur Strassen (2003) replaced the latter species by *T. atratus* as its mountainous form.

During the study on the *Thrips* genus larvae the morphological differences between second larval instar of *T. atratus* and *T. montanus* were observed (Kucharczyk, 2010). These species also showed different food preferences. PCA method was applied for distinguishing the most important morphological, measurable features (8 for females and 12 for males) which may be useful in recognizing these species.

On the graphs prepared for females and males separately the specimens of both species were ordered along the first two principal components, the lengths of vectors were correlated with the features significant in recognizing the studied taxa (Figs. 10, 11, Tab. 2). Two principal components sequentially accounted for cumulatively 76.0% for females (62.7% and 13.3% for PC1 and PC2 respectively) and 63.6% for males (41.4% and 22.0% for PC1 and PC2 respectively). The quantitative characters together with the qualitative ones both of adults and larvae and the host preferences: polyphagy for *T. atratus* and monophagy (on *Rhinanthus* spp.) for *T. montanus* allowed to consider them as valid taxa.

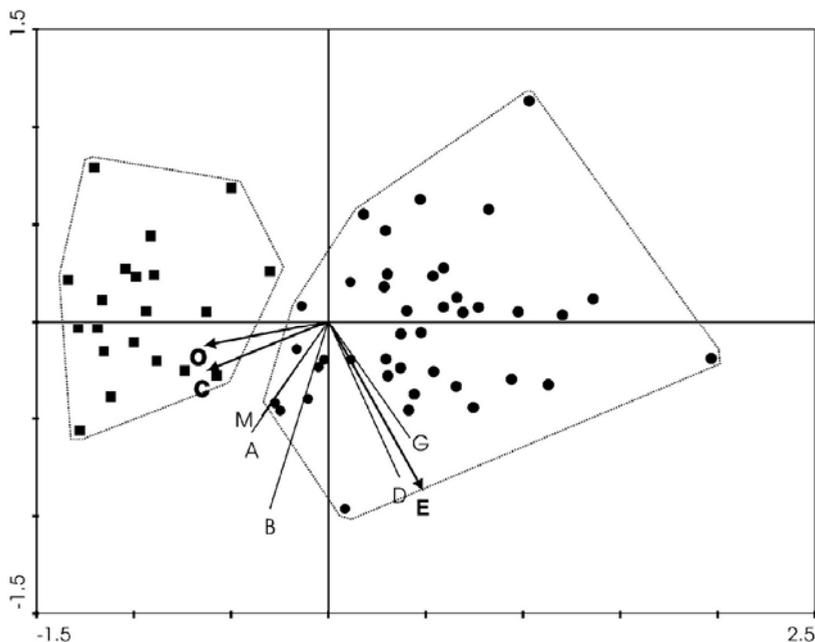


Fig. 10. PCA scatter diagram of female specimens of *Thrips atratus* (squares) and *Thrips montanus* (circles) as OTUs along PC1 and PC2 based on 8 quantitative characters (symbols as in Table 2) (after Kucharczyk & Kucharczyk, 2009)

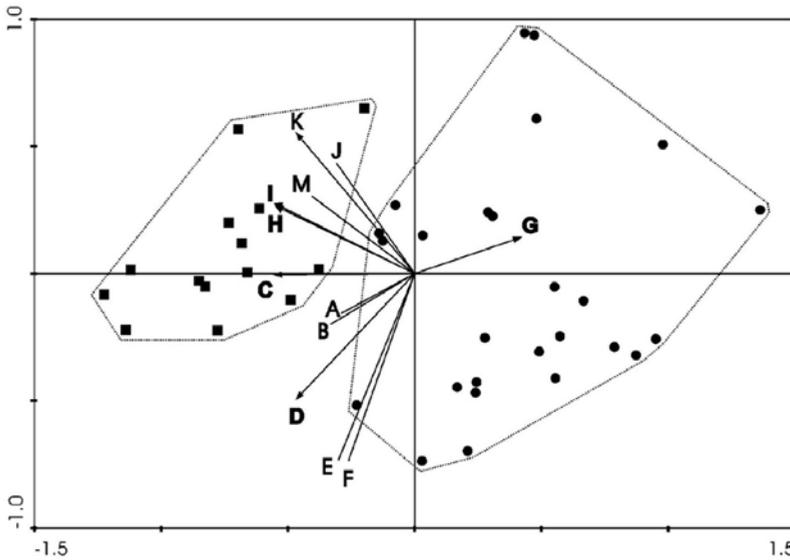


Fig. 11. PCA scatter diagram of male specimens of *Thrips atratus* (squares) and *Thrips montanus* (circles) as OTUs along PC1 and PC2 based on 12 quantitative characters (symbols as in Table 2) (after Kucharczyk & Kucharczyk, 2009)

Features	Female	Male
length of pronotal posteroangular seta interna	A	A
length of pronotal posteroangular seta externa	B	B
length of pronotal posteromarginal median seta	C	C
number of discal setae on sternum V	D	D
number of discal setae on sternum VII	E	E
number of discal setae on sternum VIII	-	F
max. number of distal setae on first vine of forewing	G	G
width of pore plate on sternum V	-	H
width of pore plate on sternum VII	-	I
length of antennal segment VII	-	J
length of antennal segment VIII	-	K
length of antennal segment III	M	M
max. length of median microtrichia on tergum VIII	O	-

Table 2. Features of *Thrips atratus* and *Thrips montanus* adults used in quantitative analyses (after Kucharczyk & Kucharczyk 2009)

2.4 Identification of *Thrips fuscipennis* Haliday 1836 and *T. sambuci* Heeger 1854 species

Similar problem exist in distinguishing *Thrips fuscipennis* Haliday 1836 and *T. sambuci* Heeger 1854. The former is a polyphagous species most often feeding in flowers while the latter is a monophagous insect feeding, breeding and developing on the lower side of *Sambucus* spp. leaves. In spite of both of the species being recognized as valid, their adults may be

distinguished mainly by differences in color of antennal segments (Schliephake & Klimt, 1979; zur Strassen, 2003). Due to the fact that the color characters are very variable in specimens originated from different populations and stations, it is not possible to accurately identify species by using only their color features (Strassen zur, 1997; Mound & Minaei, 2010).

The aim of this task was to find new characters and test their usefulness in distinguishing these taxa. Fedor et al. (2008) proposed to use an artificial neural network method (ANN) for identifying species. This method was successfully applied according to 18 species of four genera: *Aeolothrips*, *Chirothrips*, *Dendrothrips* and *Limothrips*. Finally the authors selected 19 morphometric features which have been used in ANN analysis. In the current study seven of them were selected to distinguish *T. fuscipennis* and *T. sambuci*. Additionally five new quantitative features present in both sexes and three typical for males, and one qualitative feature were proposed for using in comparative study (Tab. 3).

Features	Female	Male
head width	H-w	H-w
head length (dorsal)	H-l-d	H-l-d
head length (ventral)	H-l-v	H-l-v
eye length	E-l	E-l
ovipositor length	Ov-l	-
antenal s. V length	A-V-l	A-V-l
antenal s. VI length	A-VI-l	A-VI-l
distance between anterior and posterior ocelli	D-oc	D-oc
distance between CS - metanotum	D-Cs-mt	D-Cs-mt
distance between D1 - metanotum	D-D1-mt	D-D1-mt
length of posteroangular seta interna	L-p-s-int	L-p-s-int
length of posteroangular seta externa	L-p-s-ext	L-p-s-ext
number of campaniform sensillae - mesonotum	N-cs-ms	N-cs-ms
distance between setae D1 and fore edge of metanotum	D-D1-e-mt	D-D1-e-mt
width of area porosae on sternum V	-	W-ap-sV
width of area porosae on sternum VI	-	W-ap-sVI
width of area porosae on sternum VII	-	W-ap-sVII

Table 3. Features of *Thrips fuscipennis* and *Thrips sambuci* adults used in quantitative analyses

Similarly as in *T. atratus* and *T. montanus* characters of females and males were analyzed separately. On the graphs being the result of PCA method the specimens of both species are well separated and are located on opposite sides of Axis 2 (Figs 12, 13). The number of campaniform sensilla on the mesonotum – two (sporadically one) in *T. fuscipennis* and lack of them in *T. sambuci* is the characteristic which in the highest degree differentiate these species. Additionally the measured setae are shorter in specimens of the former species. Males of the latter one are characterized by narrower area porosae on abdominal sternites V and VI and very often lack of them on sternite VII. Moreover, this analysis shows low significance such characteristics in recognizing these species as: eye length, distance between ocelli and length of antennal segments. The characteristics mentioned above were used in ANN analysis, the results obtained with PCA method tend to reflect on their usefulness in further similar studies, particularly on *Thrips* species identification.

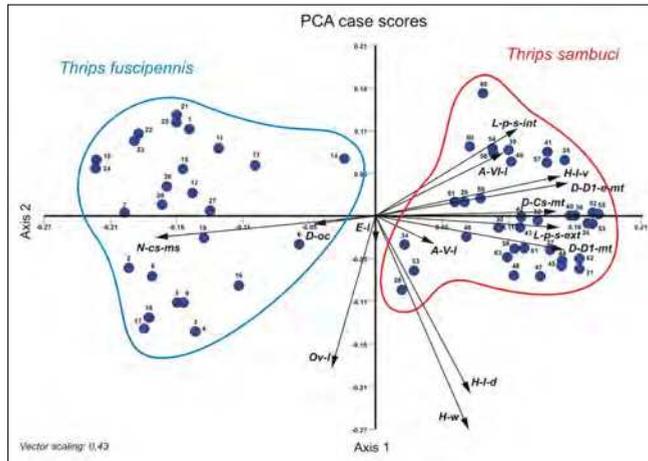


Fig. 12. PCA scatter diagram of female specimens of *Thrips fuscipennis* and *Thrips sambuci* as OTUs along PC1 and PC2 based on 14 quantitative characters (abbreviations as in Table 3)

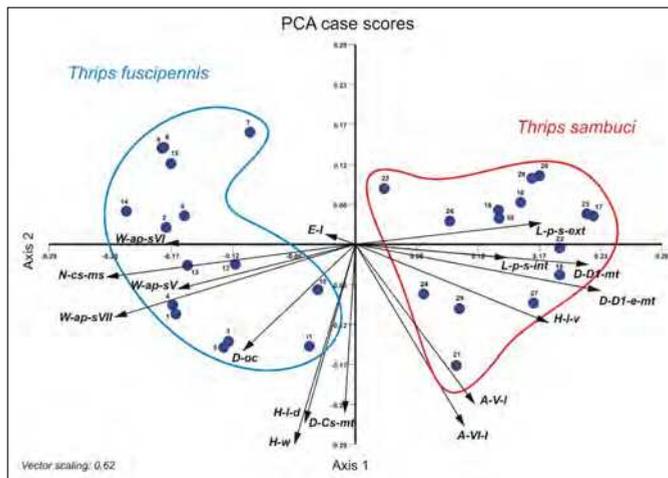


Fig. 13. PCA scatter diagram of male specimens of *Thrips fuscipennis* and *Thrips sambuci* as OTUs along PC1 and PC2 based on 16 quantitative characters (abbreviations as in Table 3)

3. Conclusion

The above mentioned examples show that PCA method is a valuable tool in identifying species. Its' application allows to select morphometric and qualitative characteristics which discriminate taxa and may be useful in construction of identification keys. This method also allows to verify the significance of some characteristics in taxonomic study and select these most relevant from large set of data.

The results obtained using the numerical taxonomy methods could objectively reflect the taxonomic position of studied taxa.

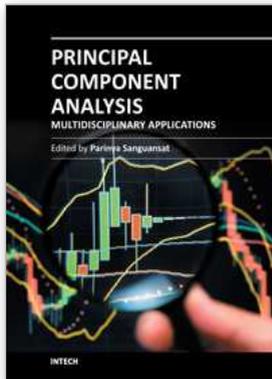
4. Acknowledgements

Financial support for this project was partially provided by the Slovak Science Foundation VEGA 1/0137/11.

5. References

- Apuan D.A.; Torres M.A.J. & Demayo C. G. (2010). Describing variations and taxonomic status of earthworms collected from selected areas in Misamis Oriental, Philippines using principal component and parsimony analysis. *Egyptian Academic Journal of Biological Sciences, B Zoology*, Vol. 2, No. 1, pp. 27-36, ISSN 2090-0759
- Bhatti J.S. (1988). The orders Terebrantia and Tubulifera of the superorder Thysanoptera (Insecta). A Critical Appraisal, *Zoology (Journal of Pure and Applied Zoology)*, Vol. 1: pp. 167-240, ISSN 0970-1516
- Bhatti J.S. (2005). Fifteen new families in the Order Terebrantia (Insecta). *Thysanoptera*, Vol. 1, pp. 27-49.
- Bhatti J.S. (2006). The classification of Terebrantia (Insecta) into families. *Oriental Insects*, Vol. 40, pp. 339-375, ISSN 0030-5316
- Chiapella J. (2000). The *Deschampsia cespitosa* complex in central and northern Europe: morphological analysis. *Botanical Journal of the Linnean Society*, Vol.134. No. 4, (August 2006), pp. 495-512, ISSN 1095-8339
- Crespi, B.J.; Morris, D.C. & Mound, L.A. (2004). *Evolution of Ecological and Behavioural Diversity: Australian Acacia Thrips as Model Organisms*. Australian Biological Resources Study & Australian National Insect Collection. ISBN: 9780975020616, Canberra, Australia
- Fedor P.; Malenovský J.; Vaňhara J.; Sierka W. & Havel J. (2008). Thrips (Thysanoptera) identification using artificial neural networks. *Bulletin of Entomological Research*, Vol. 98, No.5, (April 2008), pp. 437-447, ISSN 0007-4853
- Hotelling, H. (1933). Analysis of a complex of statistical variable into principal components. *Journal of Educational Psychology*, Vol. 24, No. 6, (September 1933), pp. 417-441, ISSN 0022-0663
- Kovach Computing Services (2005). *Multi-Variate Statistical Package Plus, Version 3.1*. Kovach Computing Services, Pentraeth, Wales, U.K.: 137 pp.
- Kucharczyk H. (2004). Larvae of the Genus Thrips – Morphological Features in Taxonomy. *Acta Phytopathologica et Entomologica Hungarica*, Vol. 39 No. 1-3, (May 2004), pp. 211-219, ISSN 0238-1249
- Kucharczyk H. (2010). *Comparative morphology of the second larval instar of the Thrips genus species (Thysanoptera: Thripidae) occurring in Poland*. Mantis Publishing Company, ISBN 978-83-929997-7-5, Olsztyn, Poland
- Kucharczyk H. & Kucharczyk M. (2009). *Thrips atratus* HALIDAY, 1836 and *Thrips montanus* PRIESNER, 1920 (Thysanoptera: Thripidae) – one or two species? Comparative morphological studies. *Acta Zoologica Academiae Scientiarum Hungaricae*, Vol. 55, No. 4, (November 2009), pp. 349-364, ISSN 1217-8837
- Lilburn T.G. & Garrity G.M. (2004). Exploring prokaryotic taxonomy. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 54, No. 1, (January 2004), pp. 7-13, ISSN 1466-5026
- Lewis T. (Ed.). (1997). *Thrips as Crop Pests*. CAB International, ISBN 0 85199 178 5, Wallingford, UK
- Mound L.,A.; Heming C., S. & Palmer J., M., (1980). Phylogenetic relationships between the families of recent Thysanoptera (Insecta). *Zoological Journal of the Linnean Society*, Vol. 69, No. 2, (June 1980), pp. 111-141, ISSN 1096-3642

- Mound L.A. & Morris D.C. (2004). Thysanoptera Phylogeny - the Morphological Background. *Acta Phytopathologica et Entomologica Hungarica*, Vol. 39, No. 1-3, (May 2004), pp. 101-113, ISSN 0238-1249
- Mound L.A. & Morris D.C. (2007). The insect order Thysanoptera: Classification versus Systematics. *Zootaxa*, Vol. 1668, (December 2007), pp. 395-411, ISSN 1175-5326
- Mound L.A. (2010). Classifying thrips into families. http://anic.ento.csiro.au/thrips/identifying_thrips/classification.html [accessed October 2011].
- Mound, L. A., Minaei, K. (2007). Australian insects of the Haplothrips lineage (Thysanoptera - Phlaeothripinae). *Journal of Natural History*, Vol. 41, pp. 2919-2978, ISSN 0022-2933
- Mound, L.A. & Minaei, K. (2010). Taxonomic problems in character state interpretation: variation in the wheat thrips *Haplothrips tritici* (Kurdjumov) (Thysanoptera, Phlaeothripidae) in Iran. *Deutsche Entomologische Zeitschrift*, Vol. 57, No. 2, (November 2010), pp. 233-241, ISSN 1435-1951
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Vol. 2, pp. 559-572, ISSN 1478-6443
<http://pbil.univ-lyon1.fr/R/pearson1901.pdf> [accessed October 2011].
- Pitkin, B. R. (1976). A revision of the Indian species of *Haplothrips* and related genera (Thysanoptera: Phlaeothripidae). *Bulletin of the British Museum (Natural History), Entomology series*, Vol. 34, pp. 223-280, ISSN 0524-6431
- Sahuquillo E.; Fraga M.I. & Martinez Cortizas A. (1997). Comparative study of classical and numerical taxonomic methods for infraspecific taxa of *Triticum aestivum* L. traditionally cultivated in Galicia (NW Spain), pp. 919-926. Valdes B. & Pastor J. (Eds.), *Proceedings of the VIII optima meeting*. Sevilla 25 Sep.-1 Oct. 1995.
- Schliephake G. (2001). Verzeichnis der Thysanoptera (Fransenflügler) - Physopoda (Blasenfüße). [In] *Entomofauna Germanica* Klausnitzer B. (Ed.). Vol. 5, pp. 91-106, ISSN 0232-5535
- Schliephake G. & Klimt K. (1979). *Thysanoptera, Fransenflügler*. Die Tierwelt Deutschlands, Vol. 66. Veb Gustav Fischer Verlag Jena, Germany
- Strassen zur R. (1997). How to classify the species of the genus *Thrips* (Thysanoptera)? *Folia Entomologica Hungarica*, Vol. 58, pp. 227-235, ISSN 0373-9465
- Strassen zur R. (2003). *Die terebranten Thysanopteren Europas und des Mittelmeer-Gebietes*. [In:] Die Tierwelt Deutschlands Vol. 74 (Goecke & Evers eds). ISBN 3-931374-58-0, Keltern, Germany
- Tommasini M. G. & Maini S. (1995). Frankliniella occidentalis and other thrips harmful to vegetable and ornamental crops in Europe, In: *Biological control of thrips pests* Loomans A., J.; M., van Lenteren J.,C.; Tomasini M., G.; Maini S. & Riudavets J. (Eds.), pp. 1-42, Wageningen Agricultural University Papers, 95.1, ISBN 90-6754-395-0, Wageningen, The Netherlands
- Wolff R. L.; Comps B.; Marpeau A. M. & Deluc L.G. (1997). Taxonomy of *Pinus* species based on the seed oil fatty acid compositions. *Trees – Structure and Function*, Vol. 12, No. 2, (December 1977), pp. 113-118, ISSN 0931-1890
- Véla E.; Tirard A.; Rinucci M.; Suesh C.M. & Provost E. (2007). Floral Chemical Signatures in the Genus *Ophrys* L. (Orchidaceae): A Preliminary Test of a New Tool for Taxonomy and Evolution. *Plant Molecular Biology Reporter*, Vol. 25, No. 3-4, (December 2007), pp. 83-97, ISSN 0735-9640
- Zawirska I. & Wałkowski W. (2000). Fauna and importance of Thrips (Thysanoptera) for rye and winter wheat in Poland. *Journal of Plant Protection Research*, Vol. 40, No. 1, pp. 35-55, ISSN 1427-4345



Principal Component Analysis - Multidisciplinary Applications

Edited by Dr. Parinya Sanguansat

ISBN 978-953-51-0129-1

Hard cover, 212 pages

Publisher InTech

Published online 29, February, 2012

Published in print edition February, 2012

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as taxonomy, biology, pharmacy, finance, agriculture, ecology, health and architecture.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Halina Kucharczyk, Marek Kucharczyk, Kinga Stanislawek and Peter Fedor (2012). Application of PCA in Taxonomy Research – Thrips (Insecta, Thysanoptera) as a Model Group, Principal Component Analysis - Multidisciplinary Applications, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0129-1, InTech, Available from: <http://www.intechopen.com/books/principal-component-analysis-multidisciplinary-applications/application-of-pca-in-taxonomy-research-thrips-insecta-thysanoptera-as-a-model-group>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.