

# UNIVERZITA PARDUBICE

Fakulta chemicko-technologická  
Katedra analytické chemie

*Licenční studium*

*„Statistické zpracování dat při managementu jakosti“*

*Semestrální práce:*

## **METODY S LATENTNÍMI PROMĚNNÝMI A KLASIFIKAČNÍ METODY**

*Vypracoval:*

**Ing. Jiří Novák**

Výzkumný ústav lesního hospodářství a myslivosti Jíloviště-Strnady  
Výzkumná stanice Opočno

**Květen 2003**

## **OBSAH**

<b>ZADÁNÍ .....</b>	<b>3</b>
<b>DATA .....</b>	<b>3</b>
<b>ŘEŠENÍ POMOCÍ PROGRAMU OPGM</b>	
<b>EDA .....</b>	<b>5</b>
<b>TVORBA HVĚZDICOVÝCH GRAFŮ .....</b>	<b>6</b>
<b>METODA HLAVNÍCH KOMPONENT .....</b>	<b>8</b>
<b>FAKTOROVÁ ANALÝZA .....</b>	<b>8</b>
<b>ANALÝZA SHLUKŮ .....</b>	<b>9</b>
<b>ZÁVĚR .....</b>	<b>10</b>

## ZADÁNÍ

Na příkladu z vlastního pracoviště provedte analýzu vícerozměrných dat statistického softwaru OPGM<sup>®</sup>.

## DATA

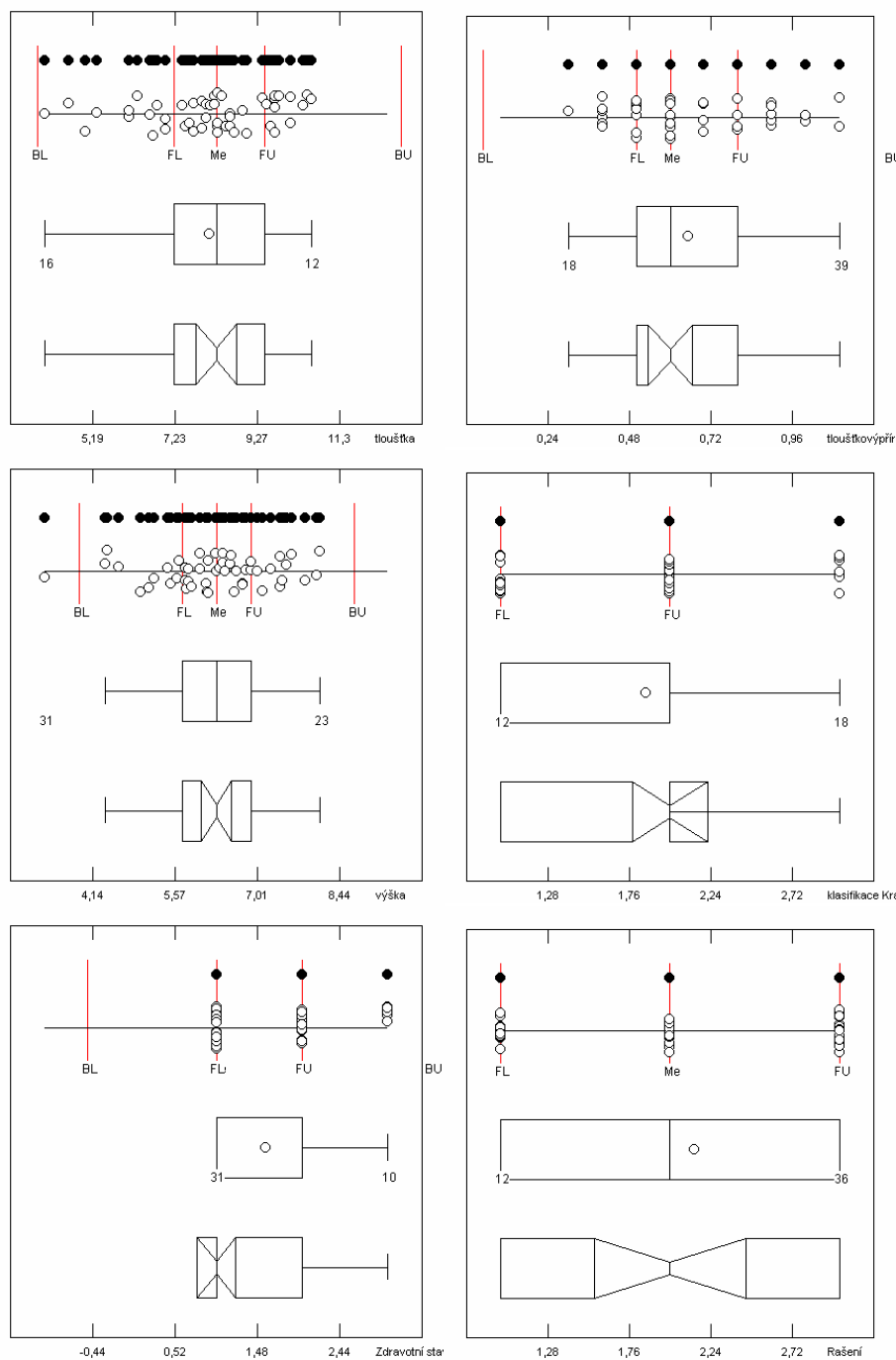
Soubor dat pro 50 jedinců smrku ztepilého (*Picea abies* L.) z pokusné plochy (bez úmyslných těžebních zásahů) Polom v Orlických horách obsahuje šest charakteristik sledovaných pro každý strom zvlášť. Jedná se o: **d** (cm) – výčetní (měřená ve výšce 1.3 m nad zemí) tloušťka stromu, **id** (cm) – roční tloušťkový přírůst (tamtéž), **h** (m) - výška stromu, **Kraftova** klasifikace stromů v porostu (1 – předrůstavý, 2 – úrovňový, 3 – částečně úrovňový, 4 – podúrovňový, 5 – potlačený (odumírající)), **zdravotní stav** (1 – olistění 90 – 100 %, 2 – olistění 70 – 80 %, 3 – olistění pod 60 %), **rašení** letorostů (1 – pozdní, 2 – střední, 3 – časně).

index	tloušťka d (cm)	tloušťkový přírůst id (cm)	výška h (m)	Kraftova klasifikace	zdravotní stav	Rašení letorostů
1	8.9	0.8	4.60	2	1	2
2	9.6	0.6	6.60	1	2	2
3	9.7	0.4	7.24	1	2	3
4	7.5	0.5	5.12	2	1	1
5	10.4	1.0	7.40	1	2	3
6	8.3	0.7	5.80	2	2	1
7	4.6	0.5	6.13	3	1	3
8	8.6	0.5	5.87	1	3	3
9	9.8	0.6	8.05	1	2	1
10	8.6	0.6	6.45	1	3	2
11	8.0	0.9	6.00	2	1	2
12	10.6	0.9	7.60	1	2	1
13	8.0	0.6	6.30	2	2	1
14	7.7	0.5	6.35	2	2	3
15	7.9	0.8	6.90	2	2	2
16	4.0	0.4	4.35	3	2	3
17	7.6	0.4	5.75	2	2	2
18	5.3	0.3	4.40	3	1	1
19	8.2	0.6	5.60	2	1	1
20	7.0	0.5	5.75	2	1	1
21	10.1	0.8	7.85	1	3	2
22	8.7	0.7	6.65	2	1	2
23	9.7	0.7	8.10	1	2	1
24	9.5	0.7	6.54	2	1	1
25	6.8	0.6	6.12	2	1	2
26	9.4	0.9	7.00	1	2	2
27	8.4	0.6	6.15	2	1	3
28	7.5	0.4	5.45	2	2	3
29	6.3	0.7	5.20	3	1	1
30	9.7	0.8	6.00	2	1	3
31	5.0	0.4	3.30	3	1	1
32	6.1	0.4	4.98	3	1	2
33	8.3	0.7	6.75	2	1	3
34	10.5	1.0	7.45	2	1	3
35	7.4	0.6	5.45	2	2	2
36	9.7	0.9	7.40	1	1	3
37	7.9	0.6	6.75	2	3	3
38	8.3	0.6	6.90	1	2	3
39	9.0	1.1	7.10	2	1	2
40	10.1	0.9	6.55	2	1	2
41	8.3	0.8	6.28	2	1	3
42	7.7	0.6	6.40	1	1	1
43	6.6	0.5	5.65	3	1	3
44	8.6	1.1	6.40	2	1	2
45	8.2	0.6	7.52	2	1	2
46	6.1	0.5	5.50	2	1	3
47	8.1	0.9	6.80	1	1	3
48	8.5	0.6	6.15	2	2	3
49	6.7	0.4	5.77	2	3	2
50	7.0	0.5	5.80	2	3	3

## ŘEŠENÍ POMOCÍ PROGRAMU OPGM

### EDA

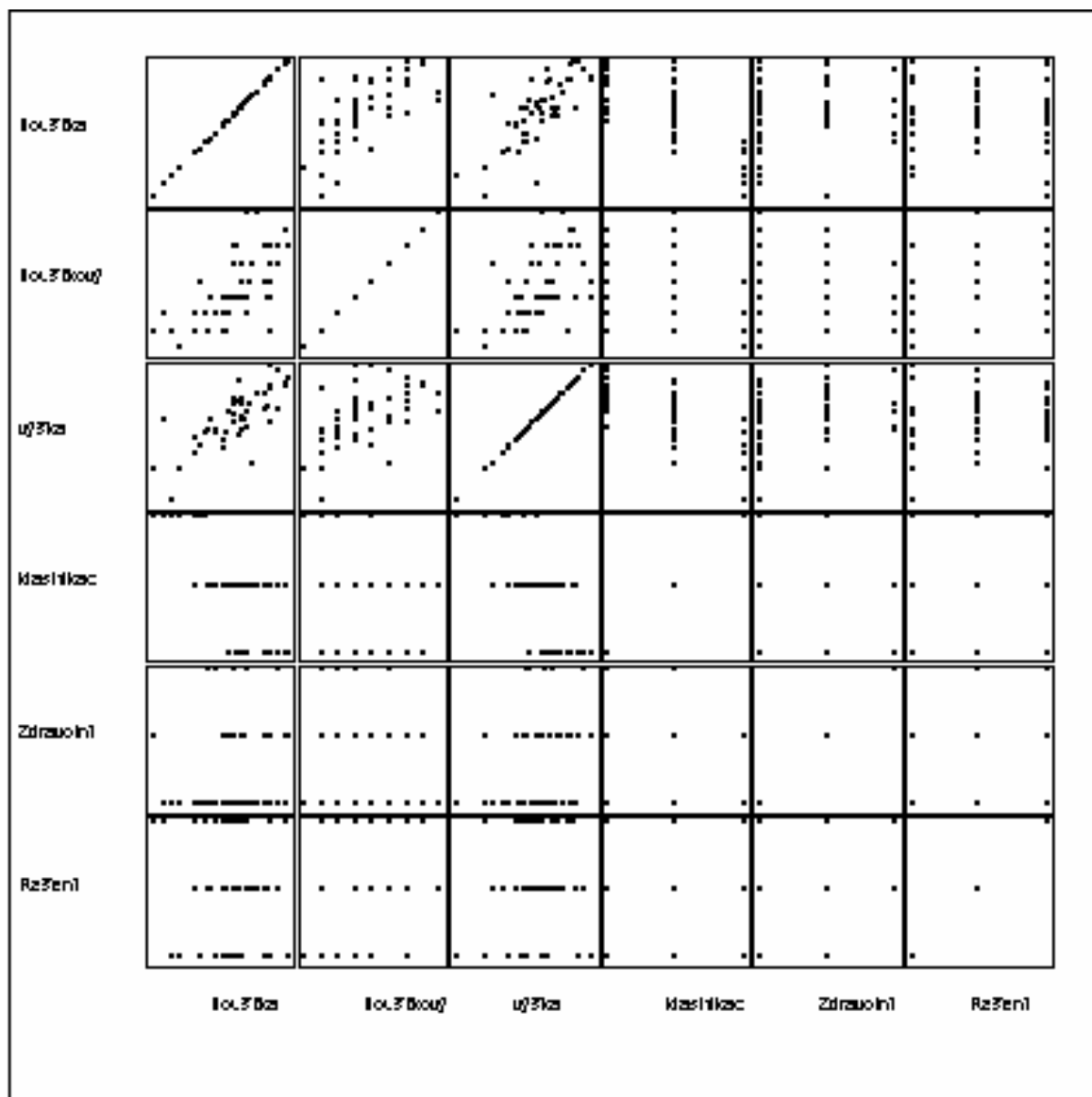
Základní šetření bylo u tohoto programu provedeno pomocí krabicových grafů (viz obr. 1 – grafy jsou sestrojeny pro každou proměnnou zvlášť). Podobně jako u předchozích hodnocení lze říci, že nejlépe budou o zvoleném souboru informovat proměnné d (tloušťka), id (tloušťkový přírůst) a h (výška).



Obr. 1: Krabicové grafy pro jednotlivé proměnné

Možná korelovanost proměnných byla vyšetřena pomocí výpočtu jednotlivých korelačních koeficientů (obr. 2)

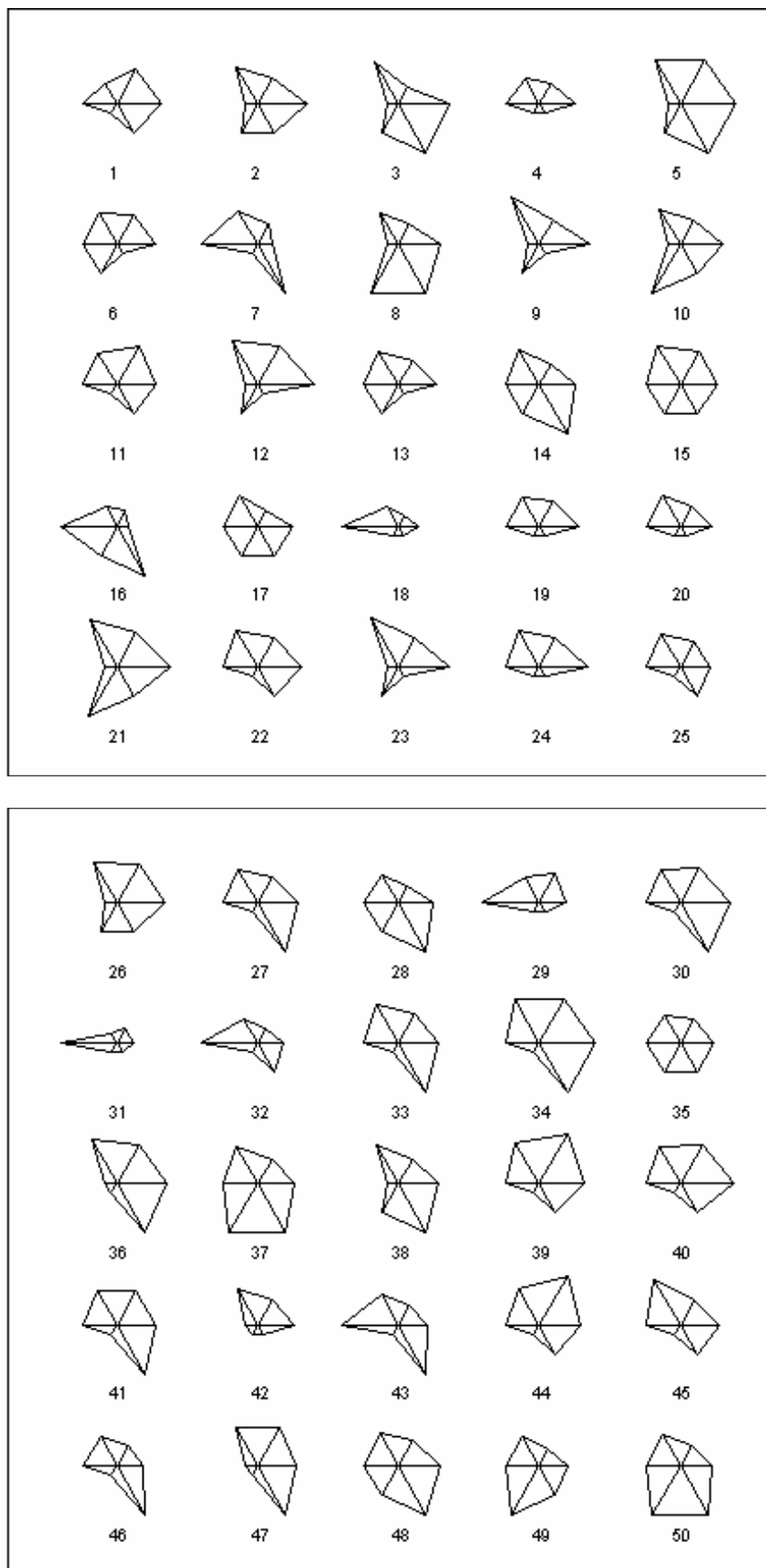
Grafické znázornění ukazuje především na korelovanost proměnných d (tloušťka) a h (výška) a také d (tloušťka) a id (přírůst).



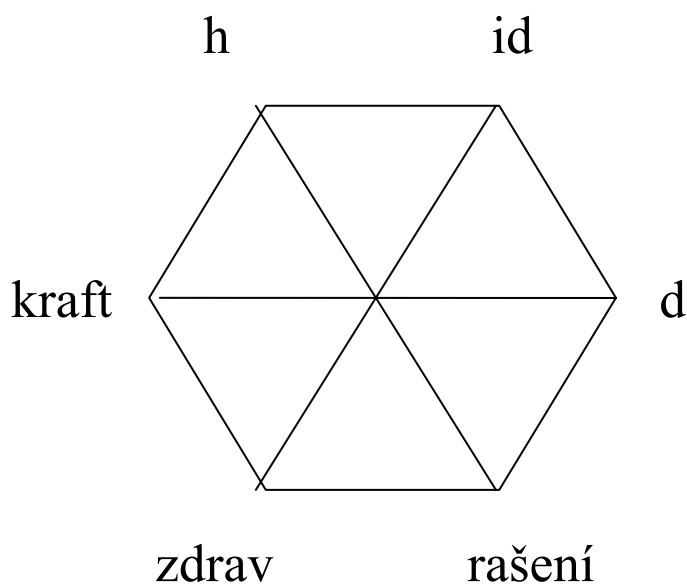
Obr. 2: Matrix plot

#### TVORBA HVĚZDICOVÝCH GRAFŮ

Program OPGM umožňuje před vlastním použitím metody hlavních komponent vytvořit tzv. hvězdicový graf (obr. 3a, b). Pohledem do grafu můžeme identifikovat podobné objekty. Např. 4, 19 a 20, nebo 39 a 44 apod.



Obr. 3a: Hvězdicový graf



Obr.3b: Klíč – legenda – ke hvězdicovému grafu

#### METODA HLAVNÍCH KOMPONENT

Pomocí této metody lze redukovat počet původních proměnných. V programu OPGM byla provedena analýza s následujícím výstupem. Program provedl standardizaci dat a analýza byla spuštěna pro 4 latentní proměnné.

j	p1	p2	p3	p4
1	9,247E-01	-1,379E-01	-6,029E-02	-6,687E-03
2	6,844E-01	-5,589E-01	2,165E-01	3,915E-01
3	8,962E-01	1,144E-02	3,074E-02	-1,832E-01
4	-8,616E-01	-2,523E-01	1,855E-01	2,153E-01
5	2,775E-01	8,720E-01	-2,122E-01	3,371E-01
6	1,077E-01	3,752E-01	9,172E-01	-5,226E-02
<b>vysvětlená variabilita %</b>				
pro 1 proměnnou	49.292	pro 3 proměnné	87.102	
pro 2 proměnné	70.898	pro 4 proměnné	92.927	

Redukcí počtu proměnných jsme dosáhli pro čtyři latentní proměnné maxima.

#### FAKTOROVÁ ANALÝZA

Analýza proběhla se standardizací dat pro 4 latentní proměnné s použitím rotace:

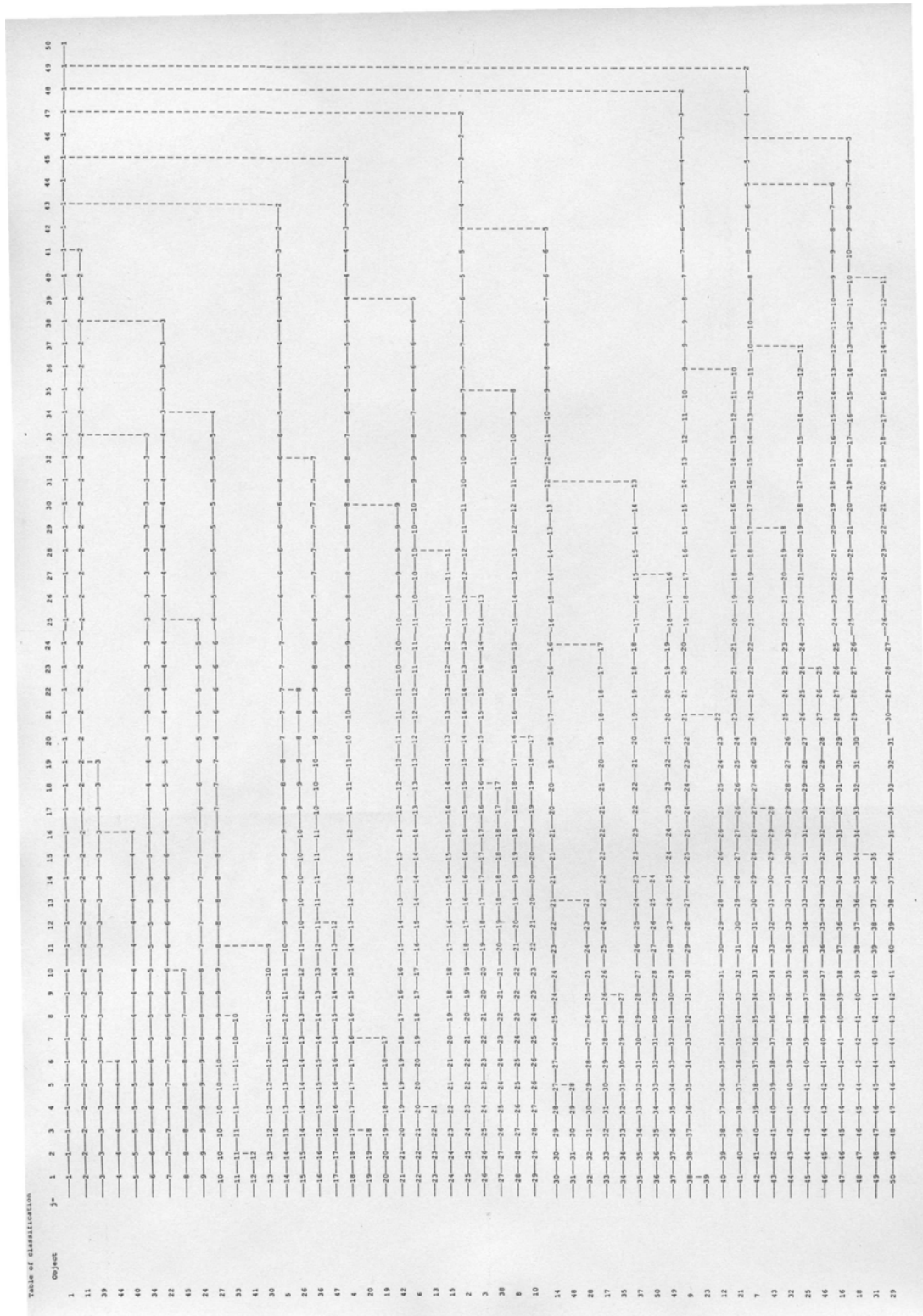
j	f1	f2	f3	f4
1	8,413E-01	2,823E-02	-2,460E-02	-3,609E-01
2	4,064E-01	-1,832E-01	2,602E-02	-8,784E-01
3	7,893E-01	8,987E-02	8,403E-02	-2,640E-01
4	-8,469E-01	-2,556E-01	2,742E-05	6,907E-02
5	2,000E-01	9,643E-01	7,489E-02	1,383E-01
6	2,581E-02	6,699E-02	9,948E-01	-1,926E-02

Ve zvoleném souboru nelze separovat faktorově čisté proměnné. Lze říci, že zvolený soubor můžeme dostatečně popsat pomocí čtyřech (88.26 % variability) faktorů.



## ANALÝZA SHLUKŮ

Grafický výstup hierarchického klastrování programu OPGM je znázorněn na obr. 4. Byla zvolena technika tvorby dendrogramu metodou shlukování „Skupinový průměr“, typu vzdálenosti „Eukleidovská“, směrodatná odchylka.



Obr. 4: Dendrogram

Na základě analýzy shluků lze konstatovat (podobně jako u předchozích analýz), že objekty (stromy) 7, 16, 18, 25, 29, 31, 32, 43 a 46 se svými vlastnostmi výrazně odlišují od ostatních objektů experimentálního souboru.

### **ZÁVĚR**

Pomocí statistického softwaru OPGM byla provedena analýza vícerozměrných dat pro soubor 50 jedinců smrku ztepilého (*Picea abies* L.) z pokusné plochy (bez úmyslných těžebních zásahů) Polom v Orlických horách. Klastrováním byl celý soubor rozdělen do celkem osmi shluků, přičemž devět jedinců (stromy 7, 16, 18, 25, 29, 31, 32, 43 a 46) se od ostatních svými vlastnostmi odlišuje velmi výrazně.