



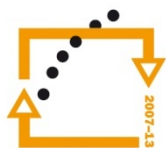
evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

„Straka v říši entropie“

– možnosti informační entropie v geoinformaticce

Vít PÁSZTO

vit.paszto@gmail.com

StatGis Team

Reg. č.: CZ.1.07/2.3.00/20.0170



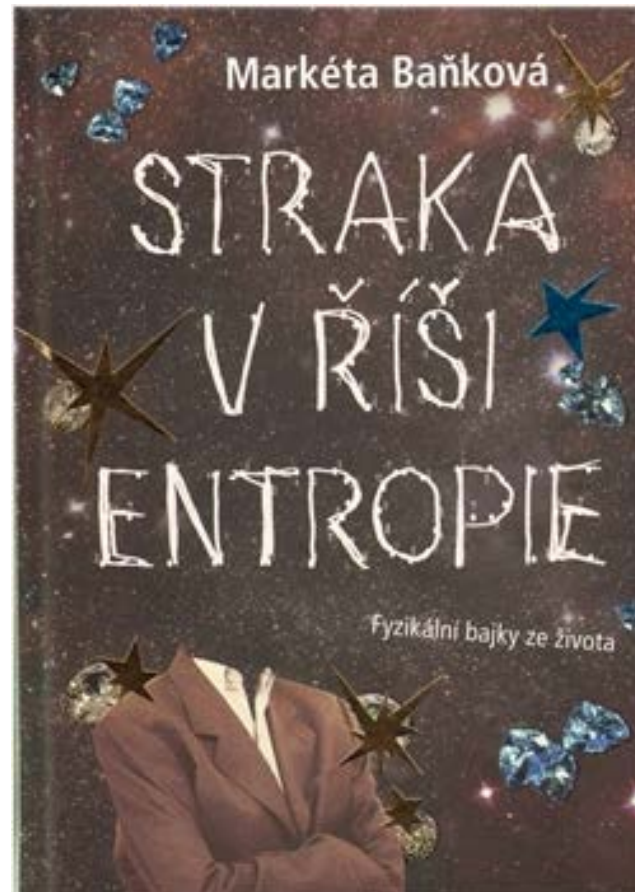
Outline

- „Straka v říší entropie“
- Koncept informační entropie
- Hodnocení uspořádanosti a podobnosti
- Další aplikace
- Návrh

„Straka v říši entropie“



„Straka v říši entropie“



Koncept informační entropie

- Vychází z teorie informace (C.E. Shannon, 1948)
- Entropie \approx mírou informace jakéhokoliv sdělení
- Míra entropie kvantifikuje množství informace
- Informace omezuje nebo odstraňuje nejistotu či nevědomost o daném jevu z dané množiny jevů (Kučerová, 2006)
- požadavky na míru neurčitosti situace respektuje veličina zvaná entropie (Komenda, 1991)

Koncept informační entropie

„Entropie je střední hodnota míry informace k odstranění neurčitosti, která je dána konečným počtem vzájemně vylučujících se jevů.“

C.E.Shannon (1948)

Koncept informační entropie

- Pro systém s konečným počtem možných stavů $S \in \{s_1, s_2, \dots, s_n\}$, $n \leq \infty$ a pravděpodobnosti jejich výskytu $P(s_i)$ je informační entropie definována:

$$H(S) = -\sum_{i=1}^n P(s_i) \log_2 P(s_i)$$

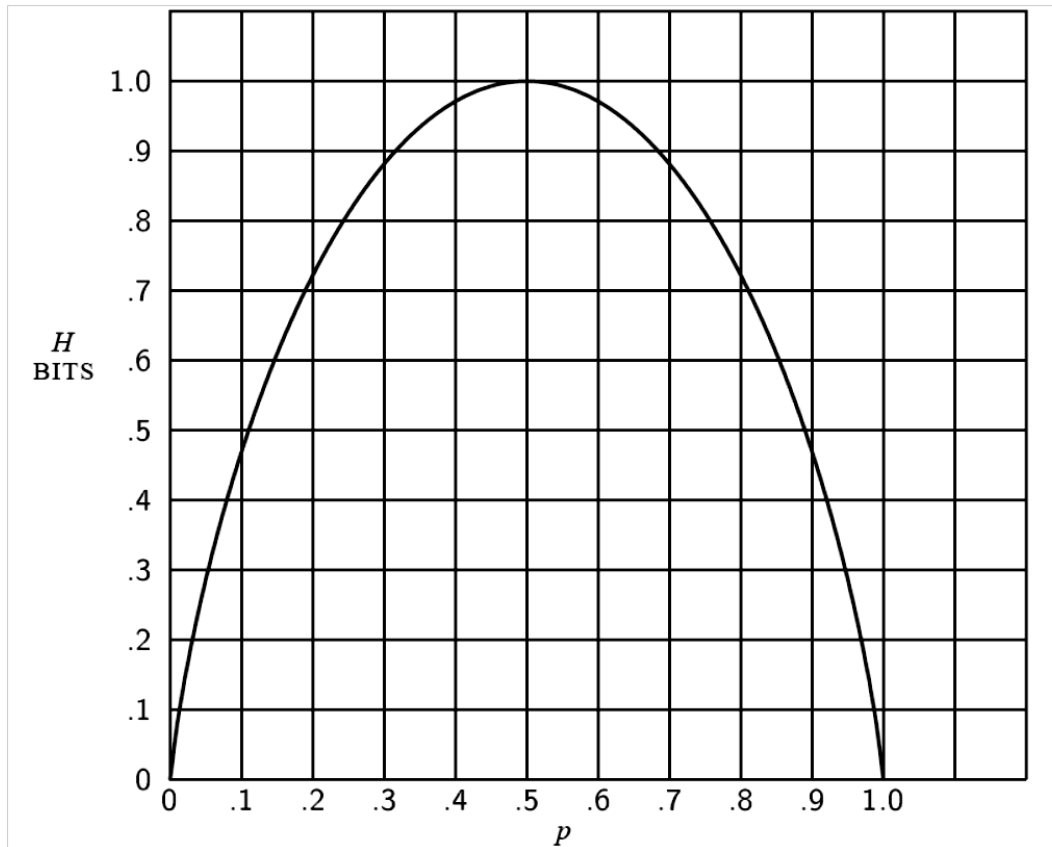
$$H(S_{\max}) = \log_2 n; P(s_i) = \frac{1}{n}$$

pro $\forall i$

$$H(S_{\min}) = 0; \exists P(s_k) = 1 \wedge P(s_i) = 0$$

pro $\forall i \neq k$

Koncept informační entropie

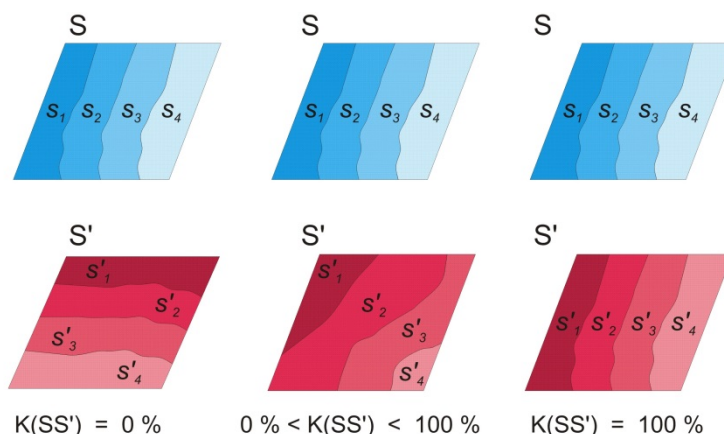


Koncept informační entropie

- V geoinformatice/geografii:
 - Hodnocení variability souboru
 - Hodnocení rozptýlení (např. kategorií, intervalů, klasifikací, ...)
 - Hodnocení (bio)diverzity (tzv. Shannon Index)
 - Hodnocení uspořádanosti jevů
 - Hodnocení podobnosti dat/map
 - Hodnocení nejistoty
 - Hodnocení restrikce dat (dělení dat do intervalů)

Hodnocení uspořádanosti a podobnosti klimatických dat pomocí informační entropie

- Metody – informační entropie (sdružená), koeficient vzájemné shody
- Data – průměrná měsíční teplota vzduchu v lednu a červenci 1991 a 2001



$$H_r = \frac{H}{H_{\max}} \quad H_{(A+B)} = H_A + H_B = -\sum_{i=1} p_{a_i} \ln p_{a_i} - \sum_{j=1} p_{b_j} \ln p_{b_j} \quad H_{(AB)} = -\sum_{j=1} \sum_{i=1} p_{a_i b_j} \ln p_{a_i b_j} \quad H_{(A+B)} \geq H_{(AB)} \quad K_{(AB)} = \frac{T_{(AB)}}{H_{(AB)}}$$

Hodnocení uspořádanosti a podobnosti klimatických dat pomocí informační entropie

- Reklasifikace (standardizace) dat

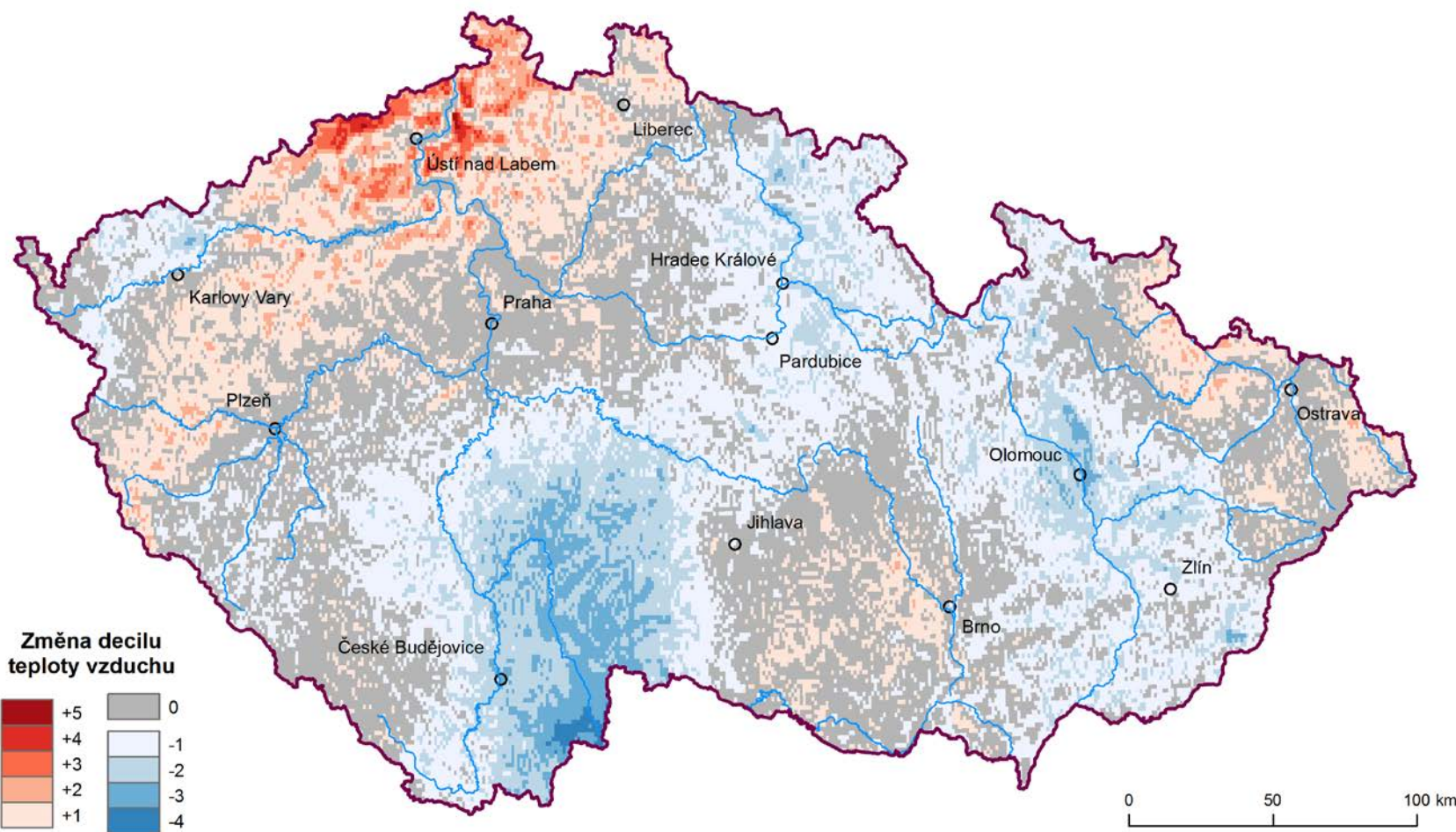
Měsíc a rok	Rozsah hodnot (°C)
Červenec 1991	13,889 až 22,855
Červenec 2001	10,349 až 19,378
Leden 1991	-3,961 až 3,049
Leden 2001	-6,186 až 1,844

- Kombinace gridů pro hodnocení uspořádanosti
- Hodnocení uspořádanosti pro každý sloupec a řádek
- Relativní hodnocení podobnosti dat/map (vizuálně)
- Výpočet koeficientu shody

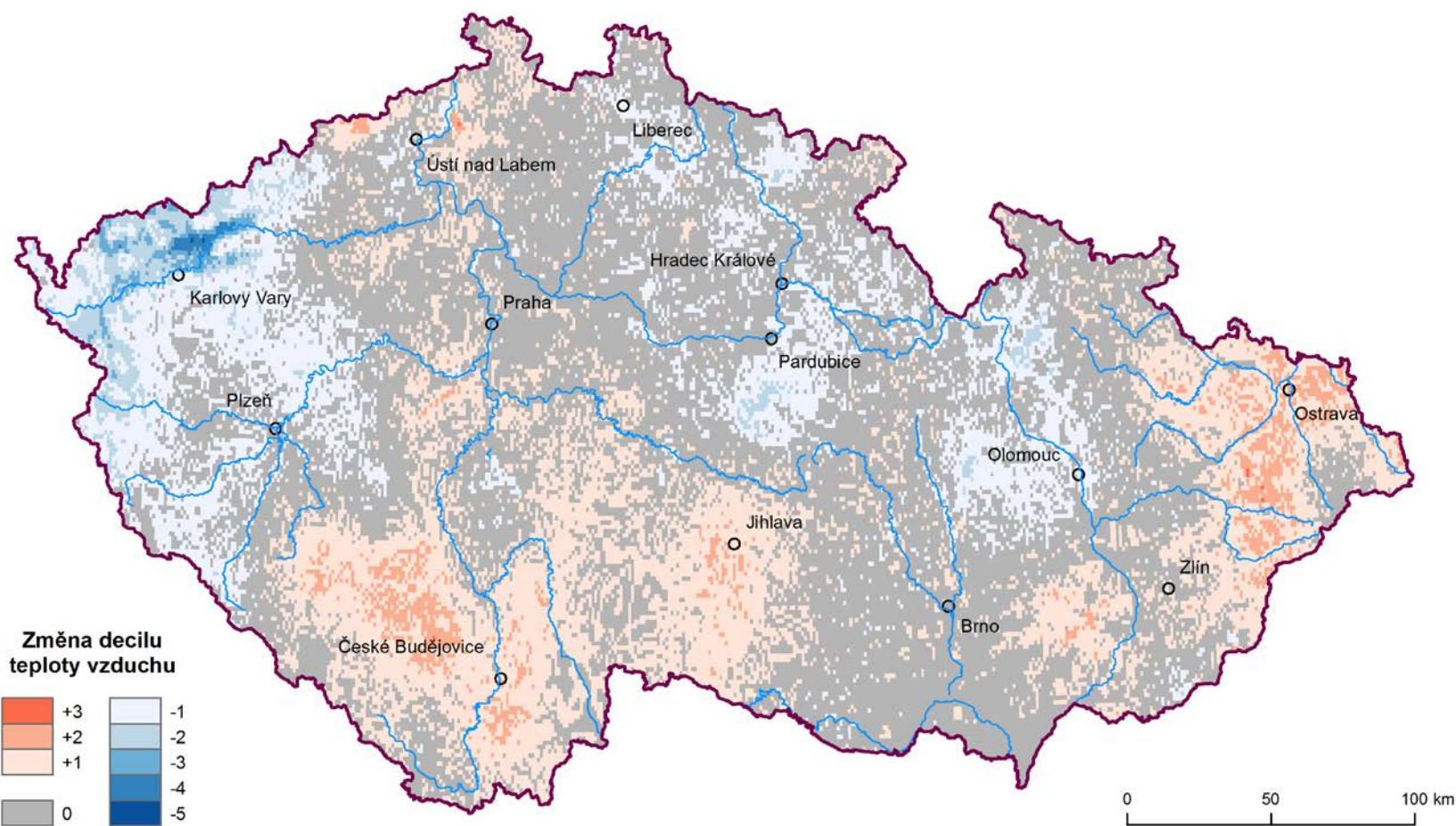
Hodnocení uspořádanosti a podobnosti klimatických dat pomocí informační entropie

		Leden 2001											
Leden 1991	Decil	1	2	3	4	5	6	7	8	9	10	H_r	
	1	649	141	4	1								0,321
	2	262	1534	413	45	37	29	4					0,532
	3	4	1043	2561	684	77	71	52	4				0,551
	4		176	2486	2534	1078	194	115	25				0,680
	5		4	995	2801	2168	1202	285	165	2			0,719
	6			345	1511	2416	2288	1206	492	51			0,840
	7			7	1006	1807	2720	2405	1323	418	1		0,801
	8				89	294	1281	2450	2323	1360	54		0,787
	9						79	404	1357	2209	694		0,791
	10							3	21	271	979		0,443
	H_r	0,569	0,641	0,693	0,748	0,782	0,688	0,659	0,678	0,649	0,578	-----	

Hodnocení uspořádanosti a podobnosti klimatických dat pomocí informační entropie



Hodnocení uspořádanosti a podobnosti klimatických dat pomocí informační entropie



Hodnocení uspořádanosti a podobnosti klimatických dat pomocí informační entropie

$$H_r = \frac{H}{H_{\max}}$$

$$H_{(A+B)} = H_A + H_B = -\sum_{i=1}^n p_{a_i} \ln p_{a_i} - \sum_{j=1}^m p_{b_j} \ln p_{b_j}$$

$$H_{(AB)} = -\sum_{j=1}^m \sum_{i=1}^n p_{a,b_j} \ln p_{a,b_j}$$

$$H_{(A+B)} \geq H_{(AB)}$$

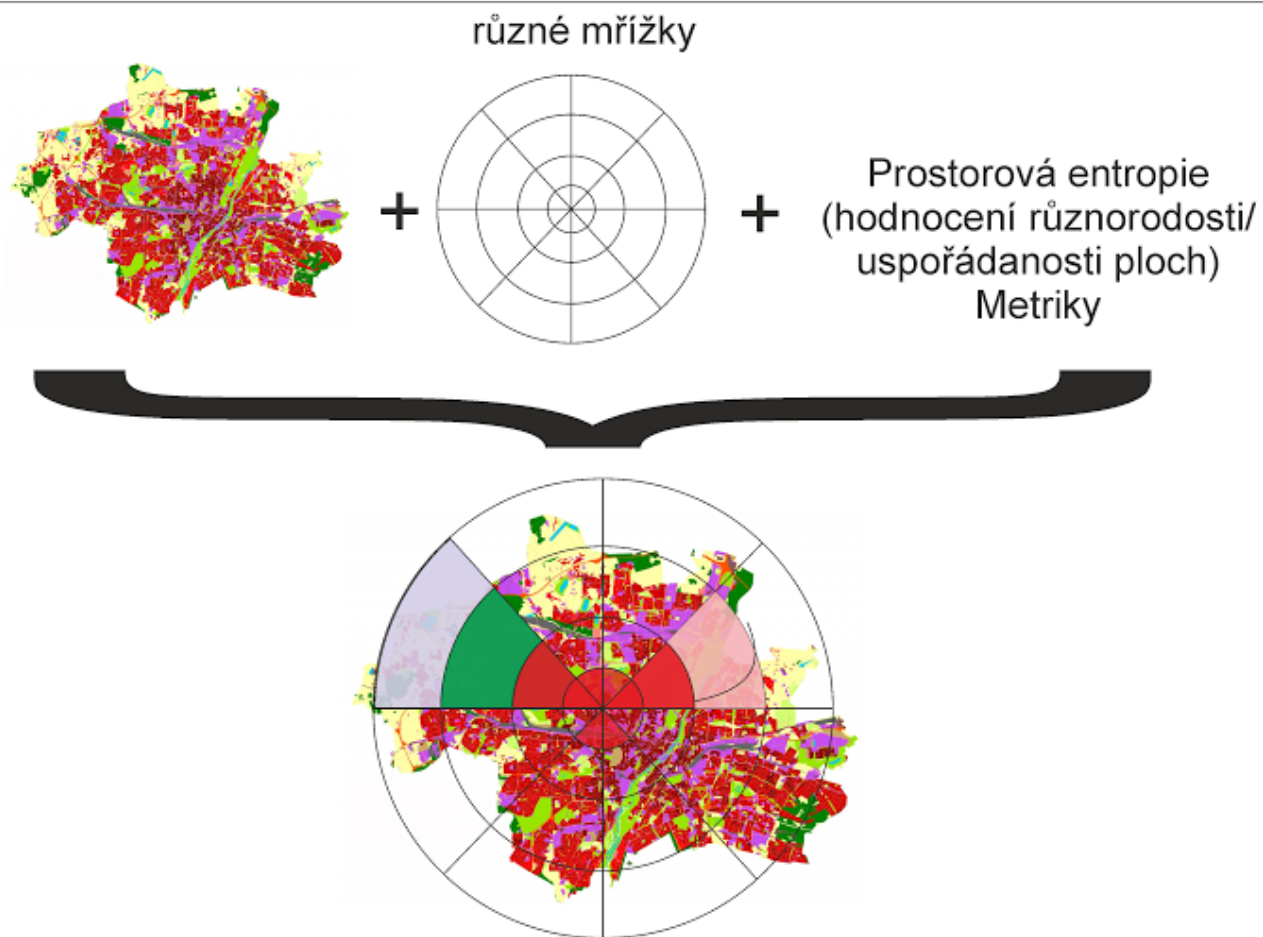
$$K_{(AB)} = \frac{T_{(AB)}}{H_{(AB)}}$$

Datová sada	Součet entropií $H_{(A+B)}$	Sdružená entropie $H_{(AB)}$	Koeficient vzájemné shody $K_{(AB)}$ (%)	Zastoupení shodných pixelů (%)
Leden 1991 a 2001	4,284	3,550	20,686	36,585
Červenec 1991 a 2001	4,243	3,256	30,329	52,759

Další aplikace...

- **Pászto, V. (2009):** Geoinformatické zpracování prostorové entropie klimatických jevů. Magisterská práce, UP Olomouc
- **Dvorský, J., Pászto, V., Skanderová, L. (2013):** Geodata Scale Restriction using Genetic Algorithm, IBICA 2013, Ostrava, Springer AICS, (Thomson ISTP), pp. 215-223, DOI 10.1007/978-3-319-01781-5_20
- **Brus, J. (2013):** Vizualizace nejistoty v environmentálních studiích. Disertační práce, UP Olomouc
- **Batty, M. (1974):** Spatial Entropy, Geographical Analysis, vol 6
- **Komenda, S. (1991):** Základy statistiky ve zdravotnictví. Olomouc, Univerzita Palackého v Olomouci, 53 s.

Návrh



„Straka v říši entropie“

– možnosti informační entropie v geoinformatice



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Vít PÁSZTO

vit.paszto@gmail.com

StutGis Team

