

# Stanovení optimálního počtu vzorků v geoekologii

Jana KITZBERGEROVÁ – Ivan LANDA

Fakulta životního prostředí

Katedra environmentálního inženýrství a ochrany prostředí

3.12. 2008

## Význam

- stanovení vhodného počtu měření, zkoušek, testů, pozorování při zajištění maximální informační hodnoty při minimální finanční náročnosti

## Cíle

- přehled přístupů stanovování počtu vzorků nezbytných pro řešení typových úloh v geoekologii
- základní principy při stanovení rozsahu vzorkování (počtu vzorků a případně jejich velikost) v návaznosti na přijetí inženýrského řešení a na náklady na získání potřebných informací v průběhu výzkumných prací

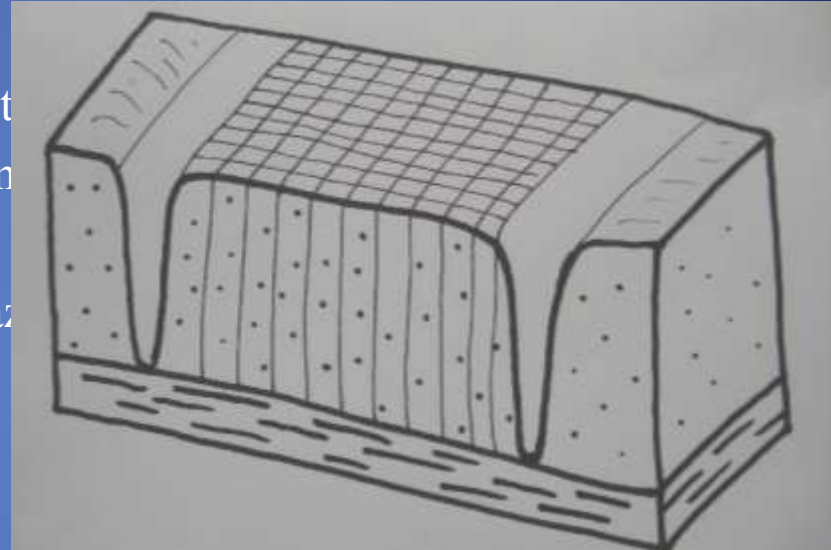
# METODIKA

1. rešerše hlavní statistické literatury
  - vztahy použitelné jen pro statistickou nezávislost
  - existence jen legislativních a normativních předpisů, a priori stanovujících počty vzorků – v geologii /ha, v geoekologii /jednotkovou plochu ohniska znečištění
2. metoda hodnocení užité hodnoty informace - sledování přírůstku užitečné informace v modelových situacích
  - 5 modelových souborů o různé heterogenitě
  - sestavení matice čísel a výpočet výběrového průměru hodnot
  - určení užité hodnoty informace
3. posouzení metody vykreslení map izolinií pomocí programu Surfer za podmínek, kdy:
  - mřížka 10 x 10, za parametr Z dosazeny modelované hodnoty jednotlivých souborů
  - možnosti metod tvorby mřížky

## 4. Modelování hydrodynamických a hydrochemických polí - programu Modflow

podmínky – prostředí je:

- blokově homogenní
- hladina je napjatá
- propustnost vertikálně a horizontálně izotropní
- okraj. podmínky na řekách  $H = \text{konst}$  (sinusoidální)
- řeky jsou hydraulicky úplné
- rozhraní mezi jednotlivými bloky je výrazné
- infiltrace (srážky) je nulová tj.  $w = 0$
- pórovitost  $n = 0,01$
- režim proudění je ustálený



Modelované hodnoty souborů dosazeny za horizontální hydraulickou vodivost.

# VÝSLEDKY

## 1. STATISTICKÉ POSTUPY

### určování počtu vzorků podle vzorců různých autorů

Autor	Vzorec	Parametry
Damaška (1986)	$n = (W * Z_{\alpha} / \delta)^2$	$n$ = počet vzorků, $W$ = variační koeficient, $Z_{\alpha}$ = zvolená spolehlivost, $\delta$ = zvolená relativní chyba
Lepš (1996)	$n > (\sigma_x / q)^2$	$\sigma_x$ = směrodatná odchylka základního souboru, $q$ = střední chyba průměru
Brabenec et al. (2004)	$n = (u_{\alpha} * \sigma / \Delta_{\max})^2$	$u_{\alpha}$ = kritická hodnota normovaného normálního rozdělení pro hladinu významnosti $\alpha$ , $\sigma$ = směrodatná odchylka základního souboru, $\Delta_{\max}$ = maximální předem stanovená přípustná chyba
Militký (1996)	$n = (t_{1-\alpha/2}(a) / d)^2 * s^2$	$t_{1-\alpha/2}(a)$ = kvantil Studentova rozdělení s $a = N_1 - 1$ stupni volnosti, $s^2$ = odhad rozptylu získaný z $N_1$ předběžných experimentů, $d$ = střední chyba průměru
Likeš et Laga (1978)	$k = 2 \Delta / \sigma$ pro $k$ jsou tabulované hodnoty $n$	$\Delta$ = předem stanovená přípustná chyba, $\sigma$ = směrodatná odchylka

- stanovení počtu vzorků - podle jednotlivých autorů

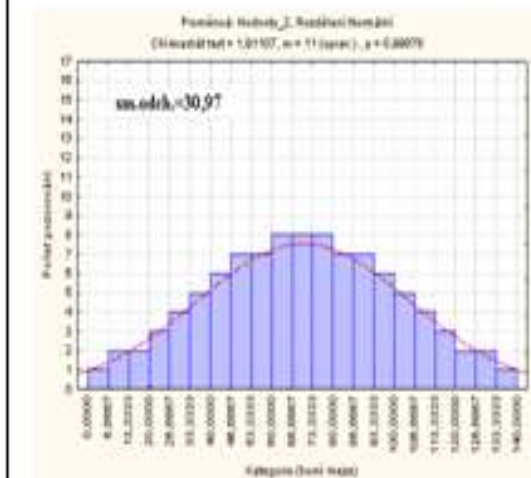
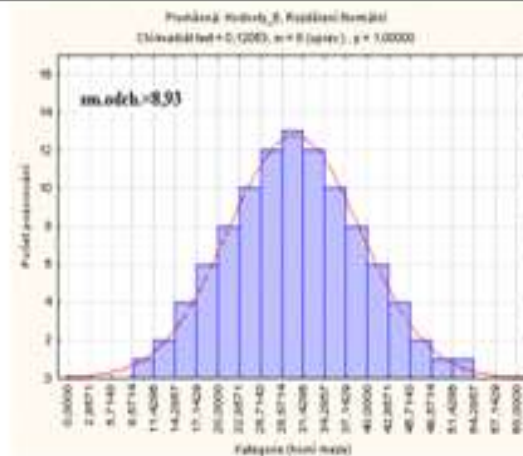
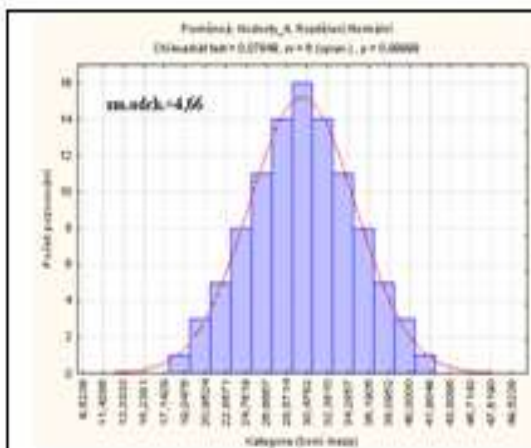
Prostředí	Sm.odch.	Damaška $Z_{\alpha} = 0,9$	Damaška $Z_{\alpha} = 0,7$	Lepš	Brabenec $\alpha=0,05$	Brabenec $\alpha=0,1$	Likeš et Laga
4	4,66	7,8	4,7	5,4	14,7	8,9	15
6	8,93	28,2	17,1	8,9	24	14,6	25
1	18,51	44,4	26,8	21,4	57,9	35,2	59
2	30,97	62,7	37,9	26,6	72,1	43,8	72
5	66,56	65	39,3	44,3	119,9	72,8	121

- funkční závislosti pro výpočet potřebného počtu vzorků

Autor	Předpis funkce	Proměnná x
Experiment DP	$y = -36,1 + 22,4 * \ln(x)$	směrodatná odchylka
Damaška ( $Z_{\alpha}=0,7$ )	$y = 1,74 e^{7,15x}$	variační koeficient
Lepš	$y = -20,4 + 14,6 * \ln(x)$	směrodatná odchylka
Brabenec	$y = -33,6 + 24,04 * \ln(x)$	směrodatná odchylka
Broocs et Carusers	$y = 1,43 e^{0,0165x}$	rozsah
Likeš et Laga	$y = -54,9 + 39,69 * \ln(x)$	směrodatná odchylka

## 2. METODA HODNOCENÍ UŽITEČNÉ INFORMACE

- Vstupní podmínky: 5 modelovaných souborů s různou





- rozdělení testovaných hodnot v souřadnicích X, Y zájmové oblasti

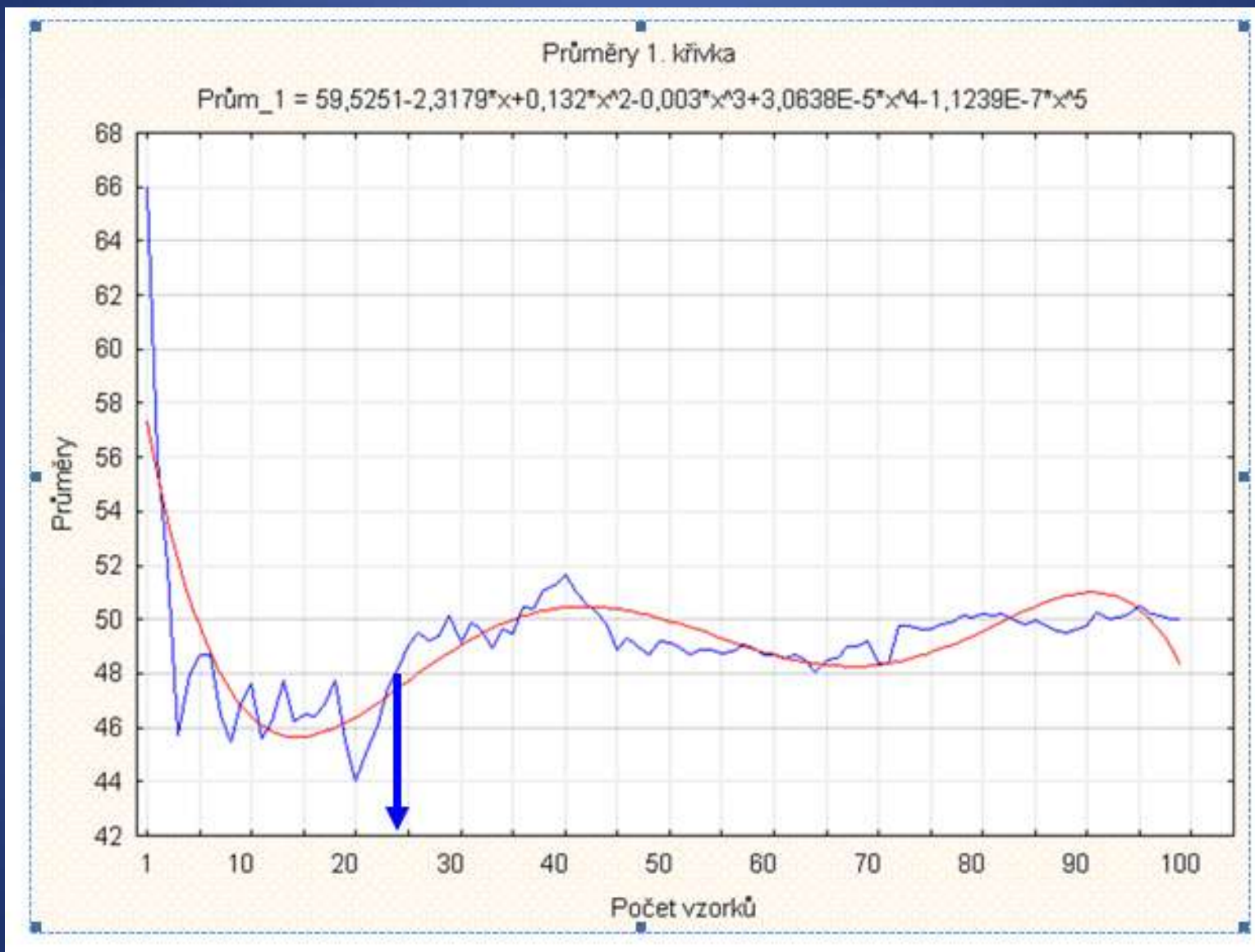
24	30	30	40	39	35	31	38	33	30
23	36	29	35	23	38	30	30	31	27
26	26	29	35	33	31	35	34	29	24
24	29	31	31	34	32	31	33	21	37
32	23	27	29	27	30	34	23	33	29
40	25	23	30	29	28	32	28	28	20
30	30	42	27	30	33	32	24	31	27
24	28	30	24	28	30	21	34	27	30
29	36	28	35	35	18	37	33	31	30
26	27	32	37	33	27	25	30	30	32

- přibližování výběrového průměru „průměru základního souboru“

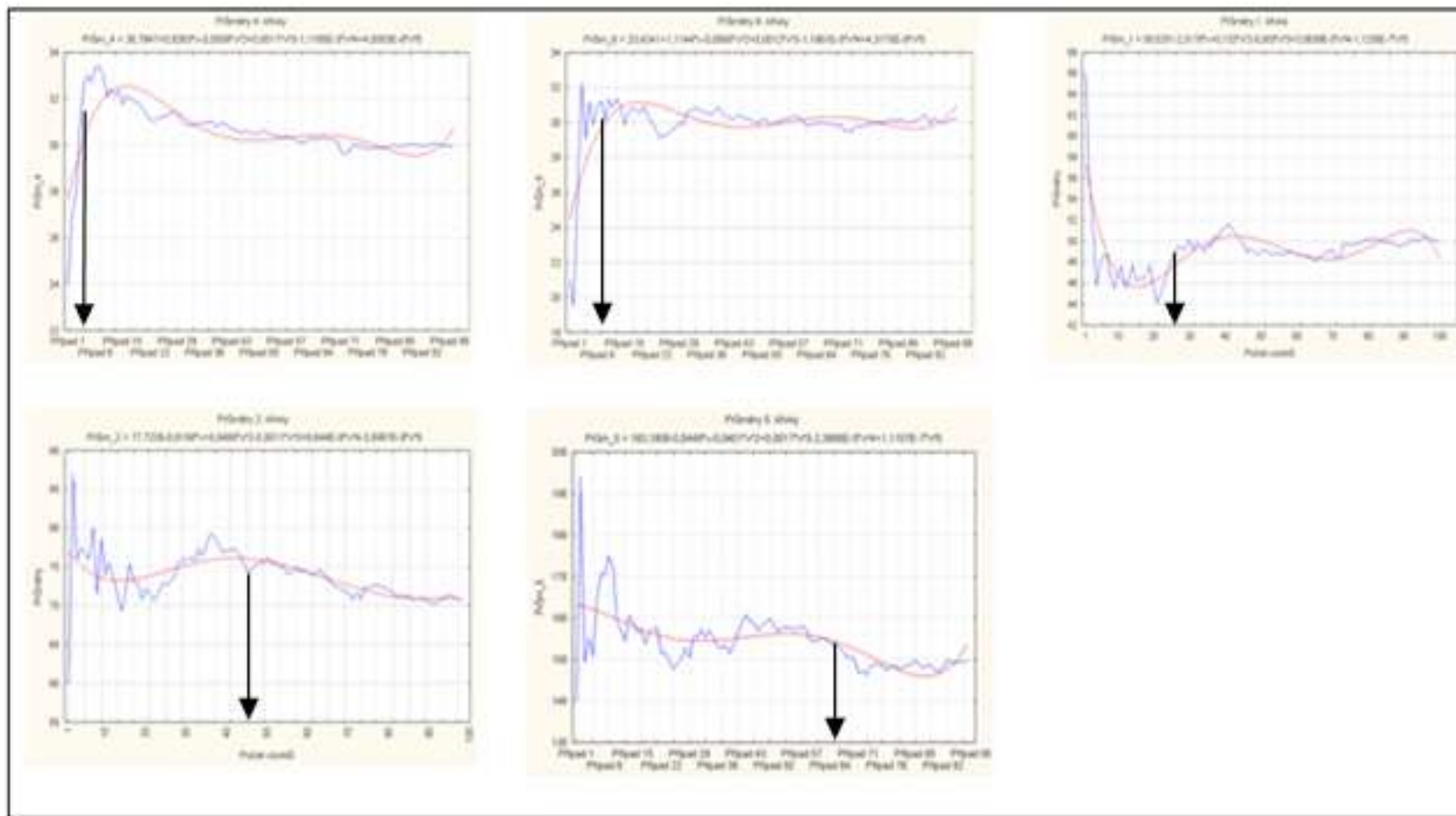
24,00	27,00	28,00	31,00	32,60	33,00	32,71	33,38	33,33	33,00
32,09	32,42	32,15	32,36	31,73	32,13	32,00	31,89	31,84	31,60
31,33	31,09	31,00	31,17	31,24	31,23	31,37	31,46	31,38	31,13
30,90	30,84	30,85	30,85	30,94	30,97	30,97	31,03	30,77	30,93
30,95	30,76	30,67	30,64	30,56	30,54	30,62	30,46	30,51	30,48
30,67	30,56	30,42	30,41	30,38	30,34	30,37	30,33	30,29	30,12
30,11	30,11	30,30	30,25	30,25	30,29	30,31	30,22	30,23	30,19
29,68	29,64	29,70	29,99	29,96	29,96	29,84	29,90	29,86	29,86
29,85	29,93	29,90	29,96	30,02	29,88	29,97	30,00	30,01	30,01
29,97	29,93	29,94	30,03	30,06	30,03	29,98	29,98	29,98	30,00



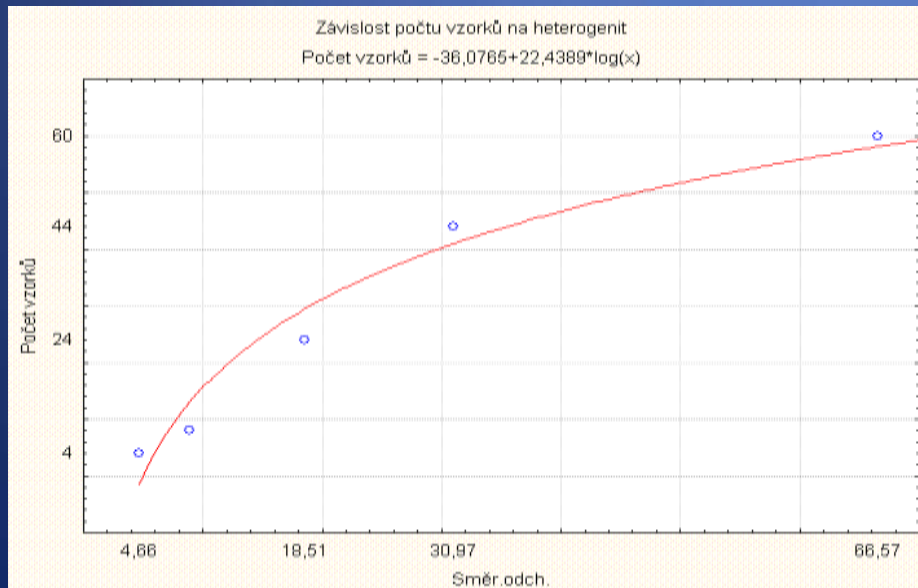
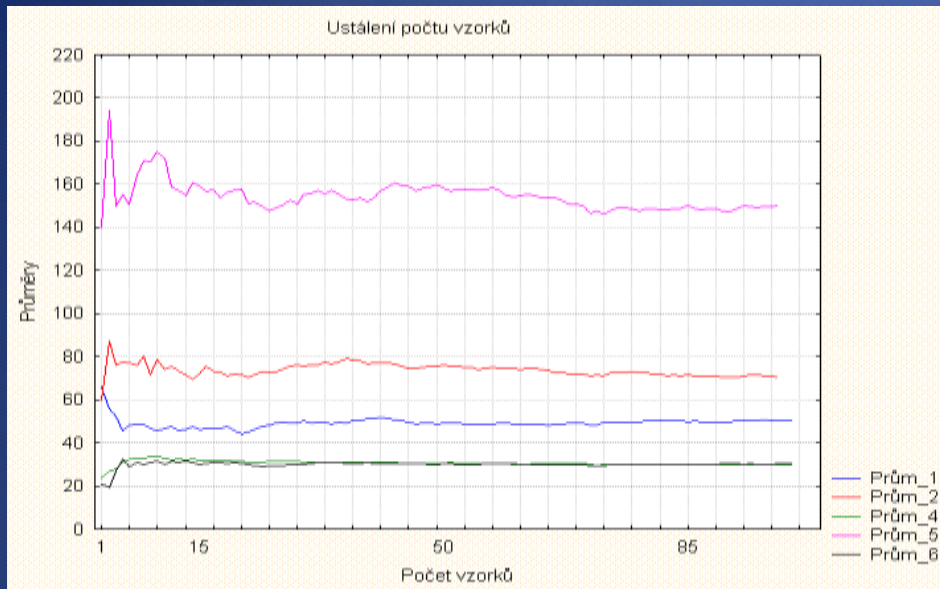
- změna užitečné informace v závislosti na zvyšování počtu vzorků



- hodnocení užitečné informace – potřebný počet vzorků se zvyšuje s rostoucí heterogenitou jednotlivých prostředí

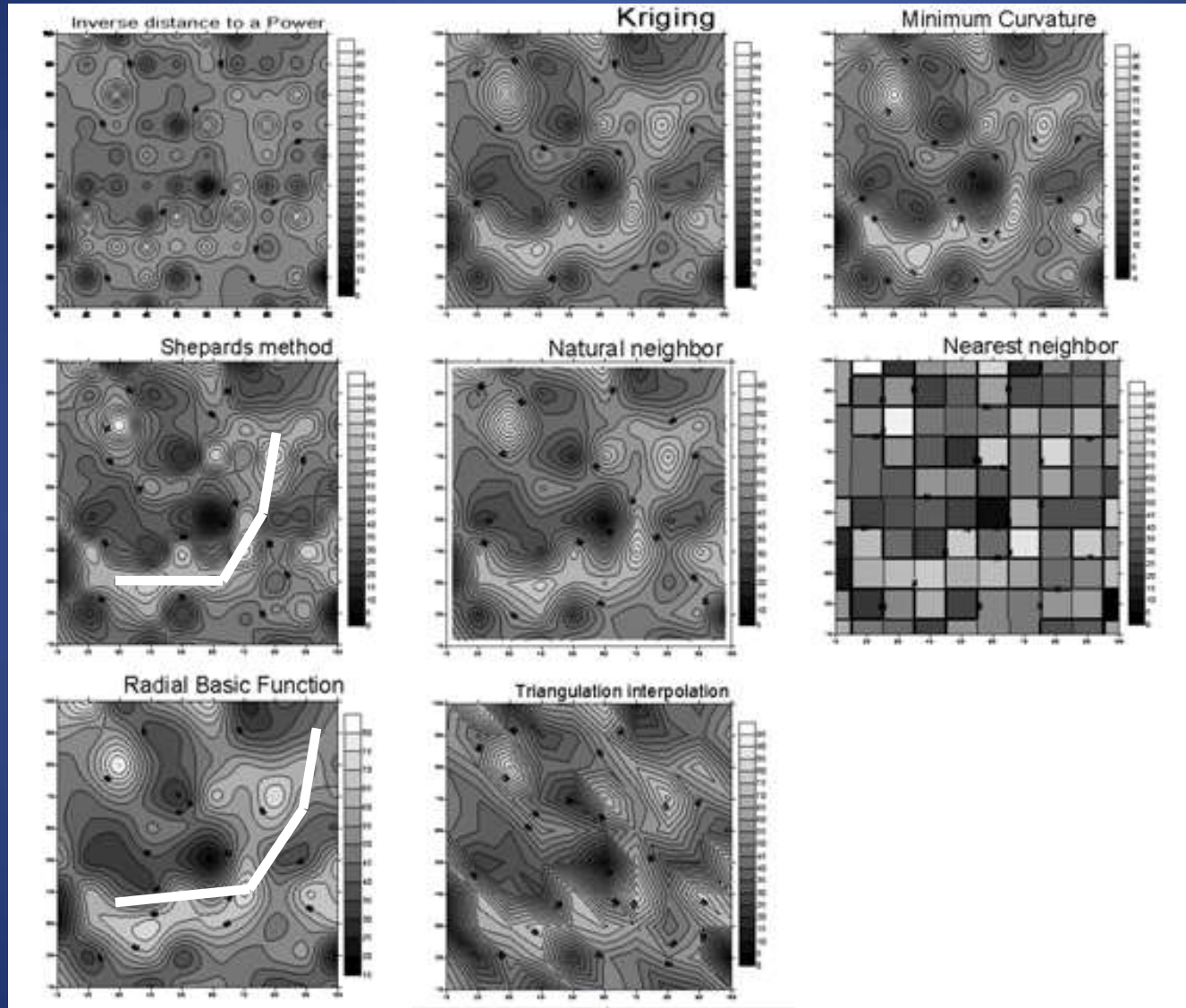


- stanovení počtu vzorků v závislosti na směrodatné odchylce prostředí



prostředí	směrodatná odchylka	počet vzorků
křivka 4	4,66	4
křivka 6	8,93	8
křivka 1	18,51	24
křivka 2	30,97	44
křivka 5	66,57	60

### 3. METODY KONSTRUKCE MAP - program Surfer

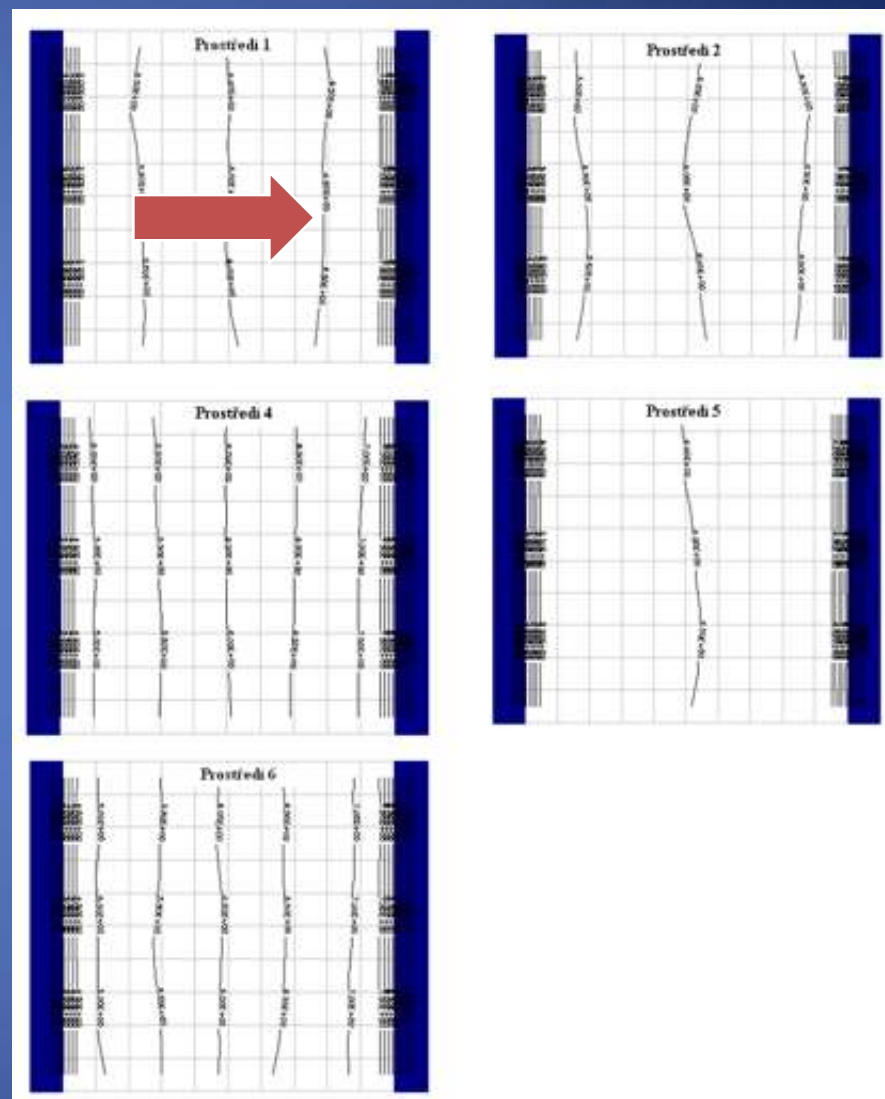


*Porovnání map izolinií parametru Z vytvořených pomocí různých gridovacích metod*

# 4. MODELOVÁNÍ PŘÍTOKU - program Modflow

Varianta prostředí	IN (m <sup>3</sup> /s) celkové množství vody, které přitéká do řeky (Q m <sup>3</sup> /s)	Aritmetický průměr koeficientů průtočnosti (T m <sup>2</sup> /s)	Směrodatná odchylka koeficientů průtočnosti (T m <sup>2</sup> /s)
1	8,085E-02	50E-04	18,51
2	8,520E-02	70E-04	30,97
4	7,462E-02	30E-04	4,66
5	9,233E-02	150E-04	66,57
6	7,364E-02	30E-04	8,93

*Přítok vody do řeky pro jednotlivá prostředí*

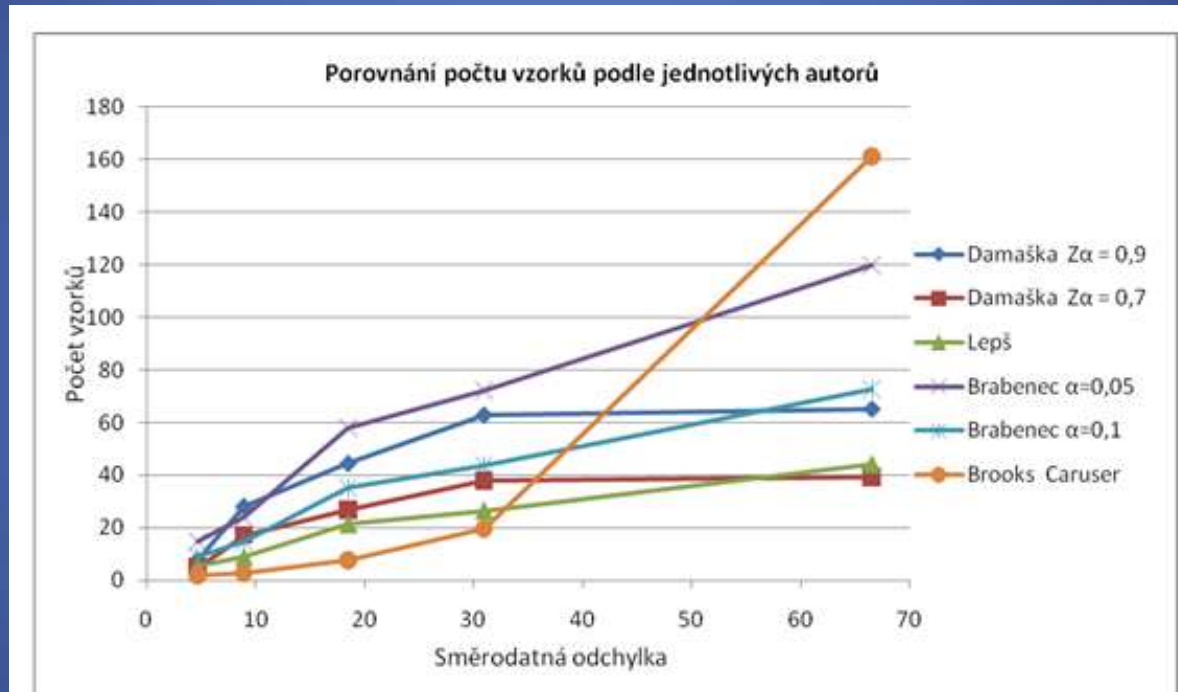


*2D vizualizace hydroizohyps*



# ZÁVĚR

- Publikované vztahy odvozené na základě statistických předpokladů vykazují značný rozptyl vypočtených hodnot počtu vzorků



- Rozpracovaná metoda sledování nárůstu užitečné informace s použitím matematického modelování s využitím získaných informací je nezaměnitelná hlavně v případech, kdy se zvyšuje prostorová závislost měřených dat tj. data jsou závislá
- Výsledky modelování umožnily definovat vztah pro výpočet počtu vzorků

$$n = -36,1 + 22,4 * \ln (\text{sm.odch.})$$

- Běžně používané interpolační postupy s využitím programových prostředků (např. programu SURFER) mohou poskytnout zavádějící výsledek a proto je nutno brát při konstrukci map i pomocné informace, např. o prostorové heterogenitě, okrajových podmínkách atp.



Děkuji za pozornost