



Kartografické modelování

X – Prediktivní modelování

jaro 2014

Petr Kubíček

kubicek@geogr.muni.cz

**Laboratory on Geoinformatics and Cartography (LGC)
Institute of Geography
Masaryk University
Czech Republic**



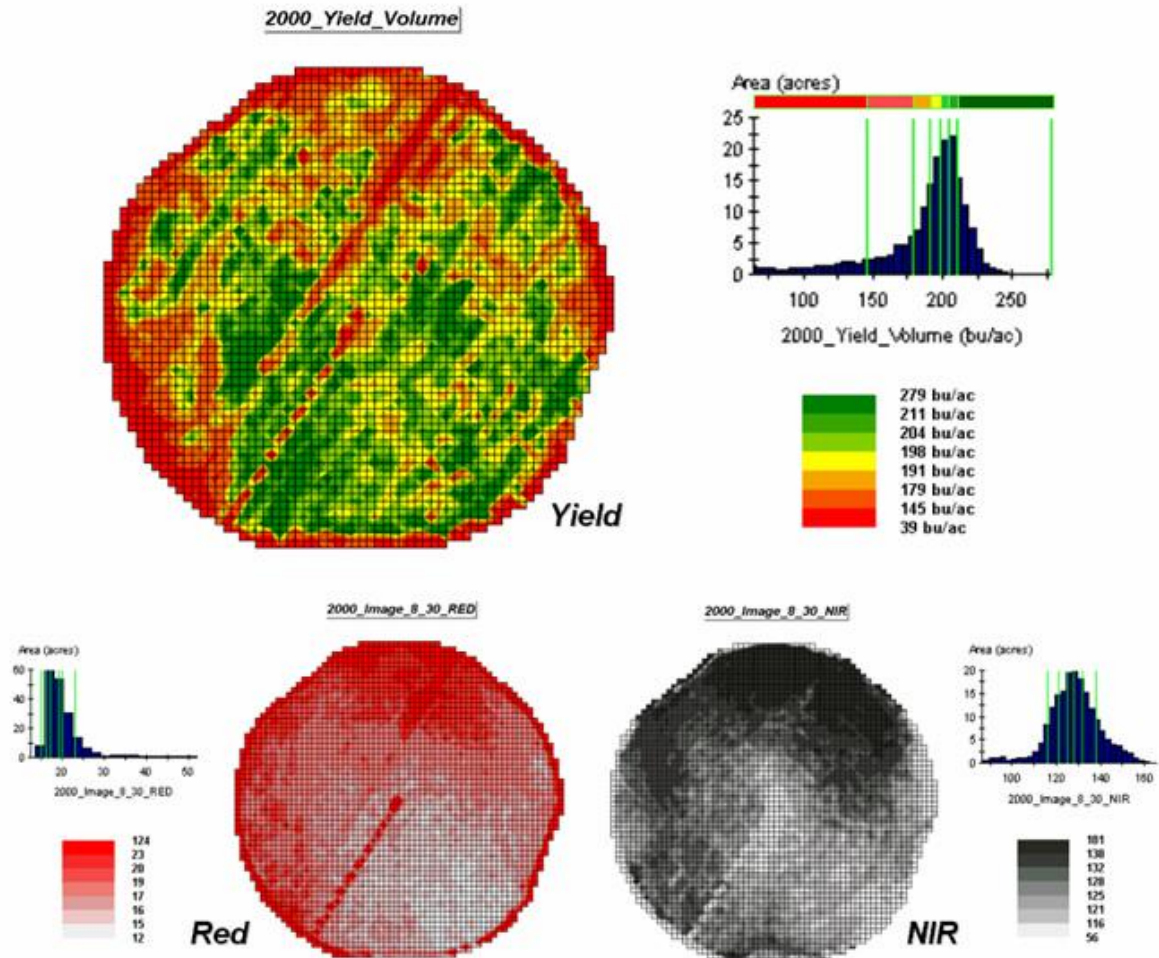
Podstata prediktivního modeování

- Doposud jsme se zabývali problémem, jak počítač „vidí“ geografická data prostřednictvím popisných (deskriptivních) technik a vytváří z nich oblasti s určitými vlastnostmi.
- Další logický krok je použití „**prediktivních – předpovědních**“ technik k vytvoření **extrapolačních map předvídajících budoucí podmínky**.
- Využití v řadě oblastí:
 - Predikce kriminality,
 - Archeologická naleziště
 - Výnosy plodin

Případová studie

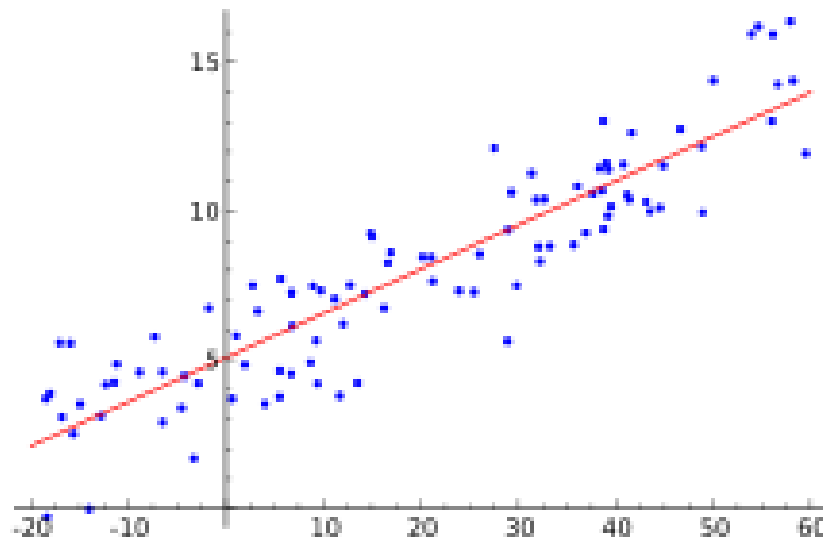
- **Využití prediktivního modelování pro precizní zemědělství (J. Berry).**
- **Výnosy kukuřice** – nízké (39 – červená) . Vysoké (279 – zelená) – **závislá proměnná** identifikující ten fenoméne (jev), který chceme predikovat.
- **Nezávislé proměnné** jsou použity pro to, aby bylo možné odhalit prostorové vztahy a vytvořit predikční rovnici.
- Využití data DPZ – **odrazivost povrchu rostlin** v červené a části spektra (RED) a v části blížíící se infračervené (NIR).

Kartografické modelování

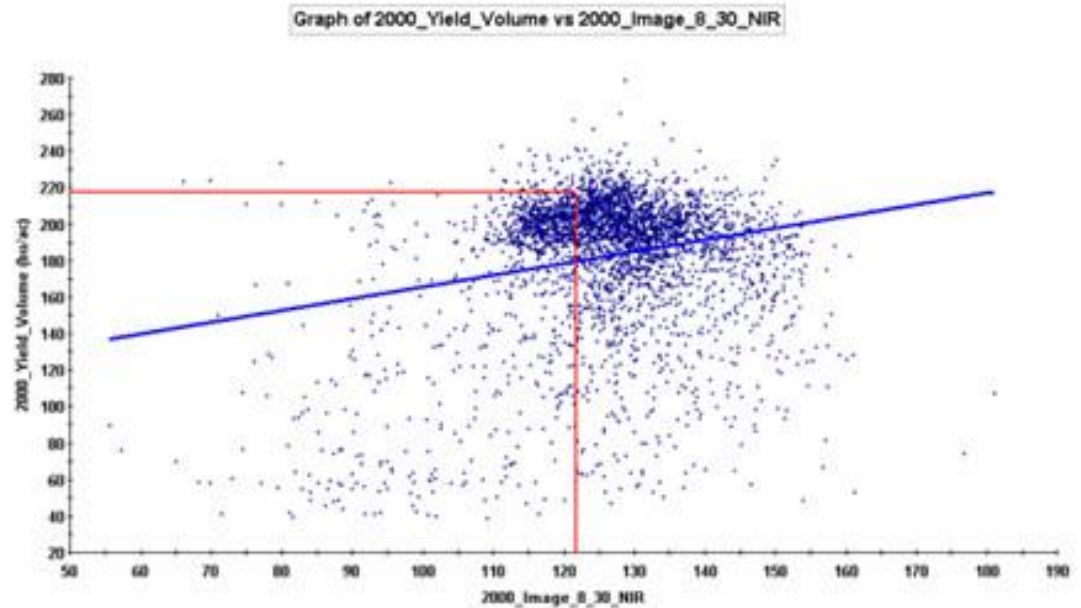
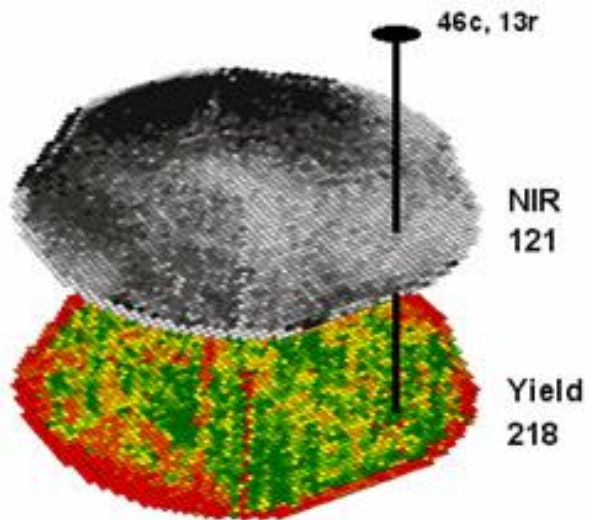
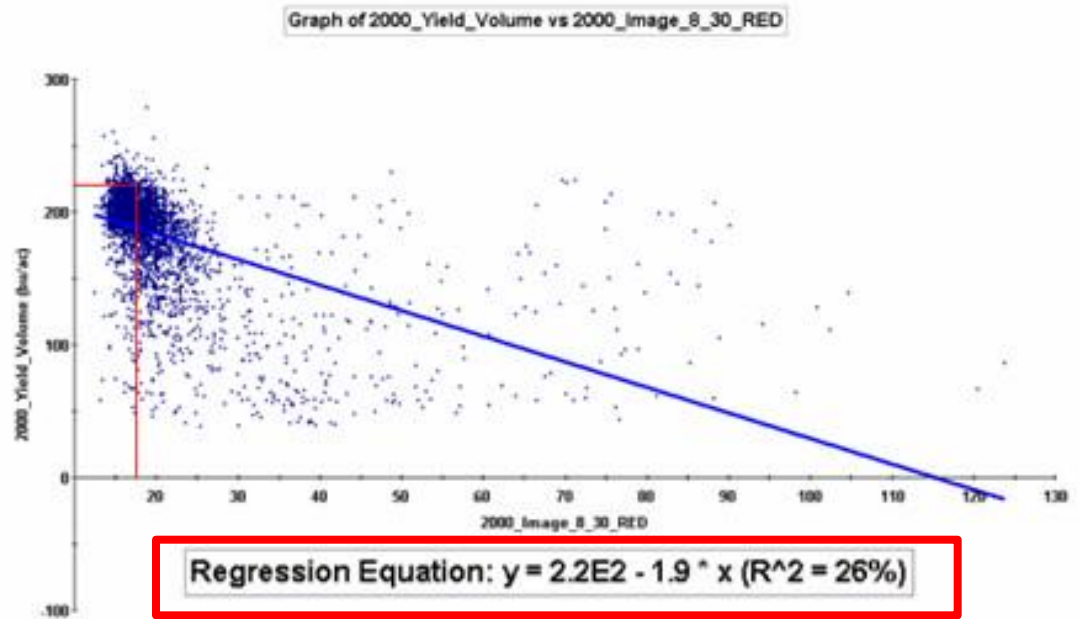
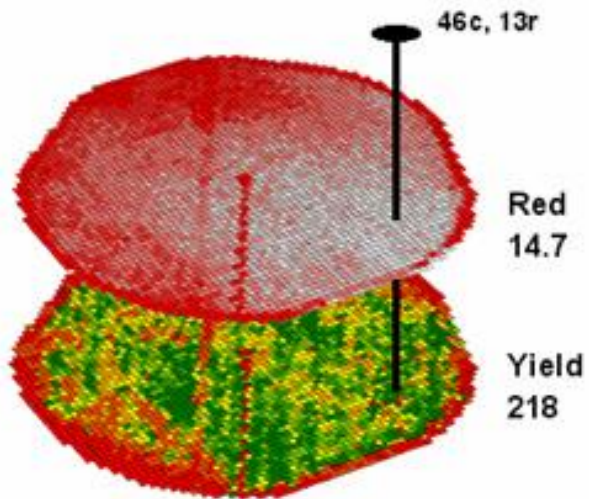


Případová studie II

- Korelační diagram (Scatter plot) pro všechny existující dvojice hodnot.
- Predikční rovnice vytvořená pomocí regresní analýzy – křivka nejlépe charakterizující datové rozložení.

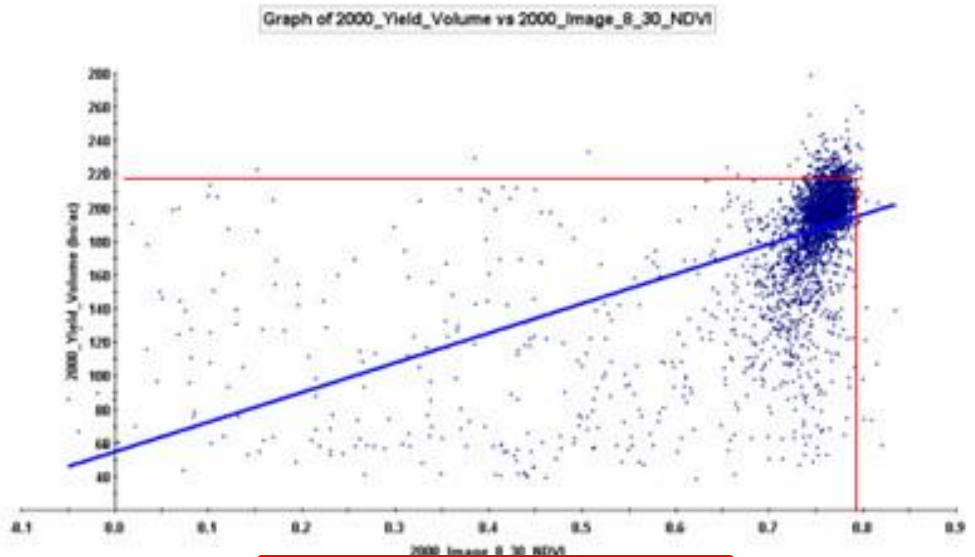
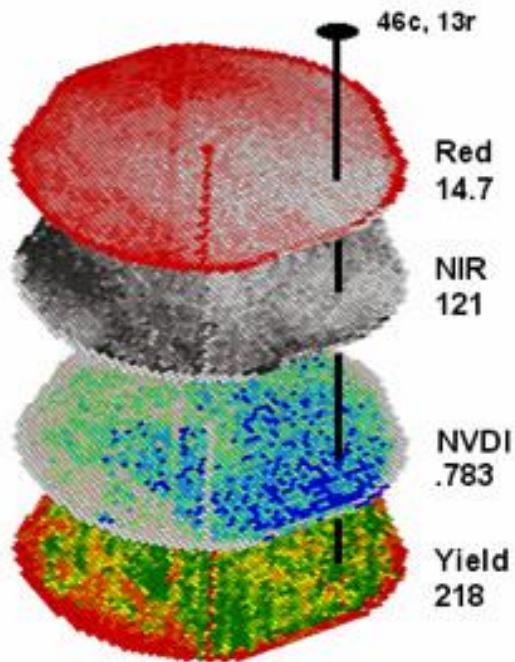


- Využití predikční rovnice pro další lokality.
Kartografické modelování



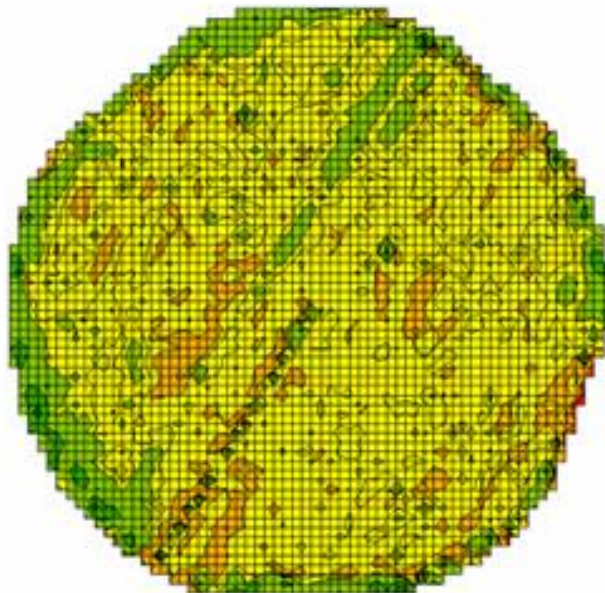
Případová studie IV

- **Problém?**
- **Predikční křivka nevystihuje rozdělení dat - nízké hodnoty R^2 (jaké hodnoty jsou vyhovující?)**
- **Možnosti využití kombinovaného indexu - NDVI**
- **Normalized Density Vegetation Index (NDVI)**
- **$NDVI = ((NIR - Red) / (NIR + Red))$**
- **Srovnání predikované a skutečné hodnoty (kalibrace modelu) - mapa odchylek.**
- **Průměrná chyba 2,26 q/ha.**

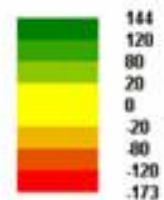


Regression Equation: $y = 55 + 1.8E2 \cdot x$ ($R^2 = 30\%$)

Error_Map

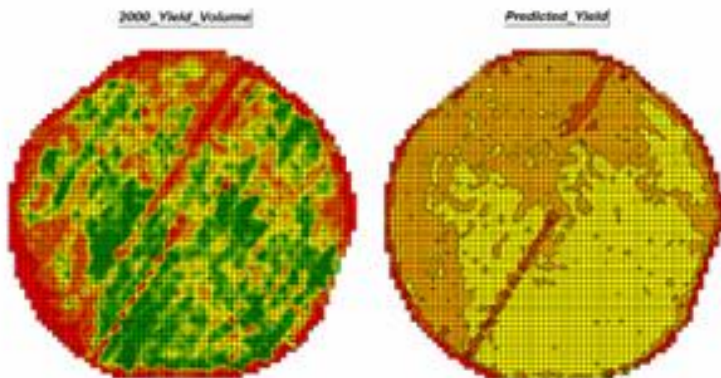


Statistics	
Min:	-173
Max:	144
Range:	318
Mean:	2.62
Median:	-4.23
Std. Dev.:	32
Variance:	1,025
Gridded Area:	189 acres



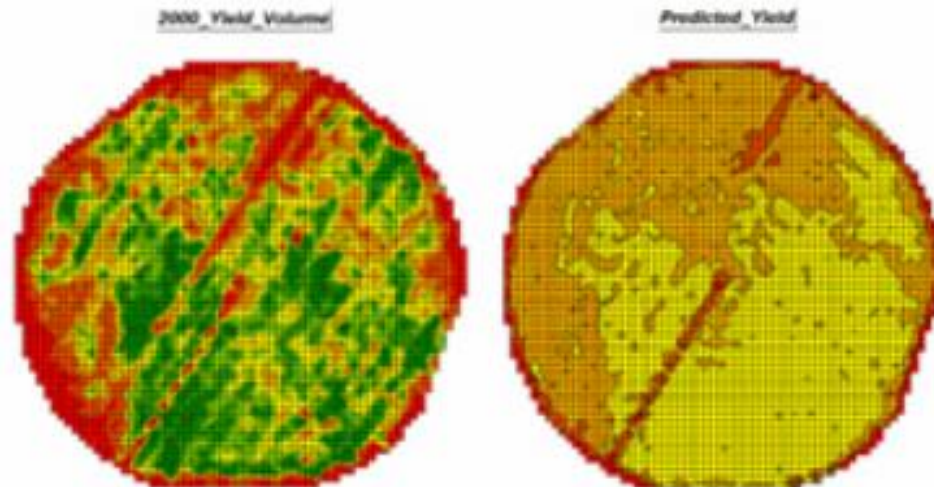
67%

Regression Equation: $y = 55 + 1.8E2 \cdot x$ ($R^2 = 30\%$)



Případová studie IV

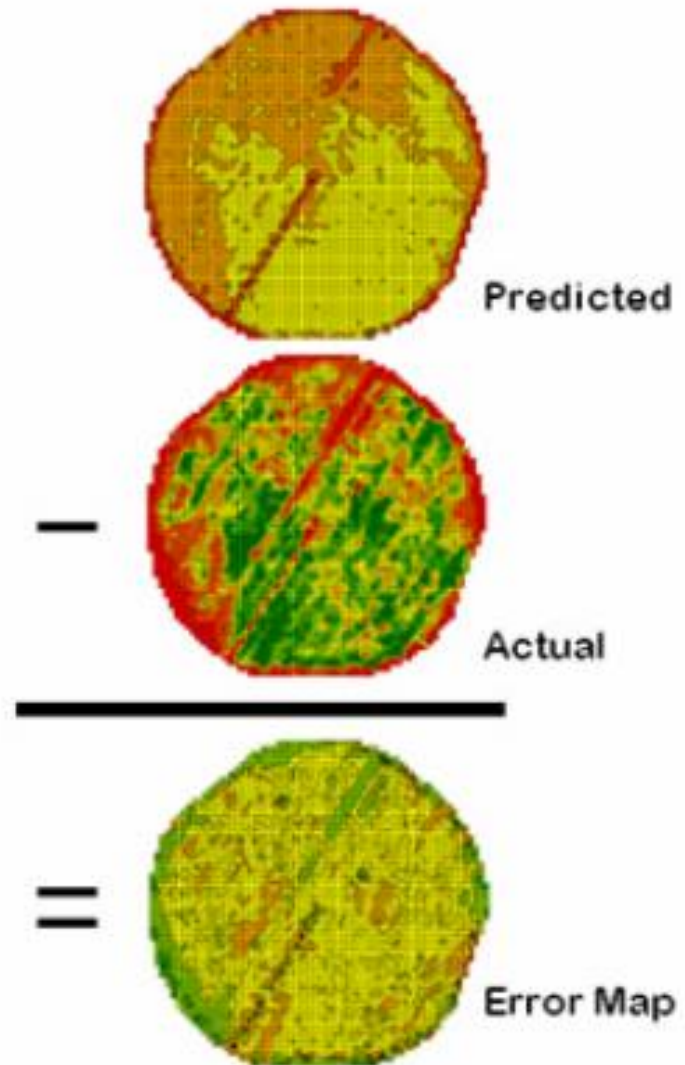
- **Výsledky shlazují skutečné výnosy.**



- **Nejedná se o skutečnou kalibraci modelu, ale spíše o první zjednodušený náhled, jaké by mohly výnosy být.**
- **Jak můžeme výsledný model dále zlepšit?**

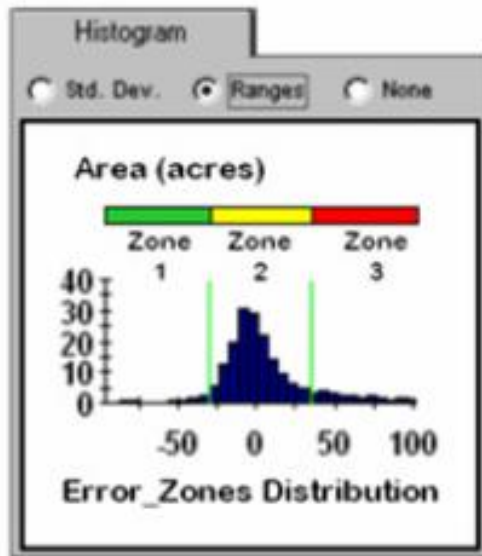
Případová studie V

- **Bližší pohled na mapu odchylek**
- **Průměrná chyba 2,62 q/ha.**
- **67% odhadu \pm 20 q/ha.**
- **ALE – některé lokality až +144 a -173 q/ha.**

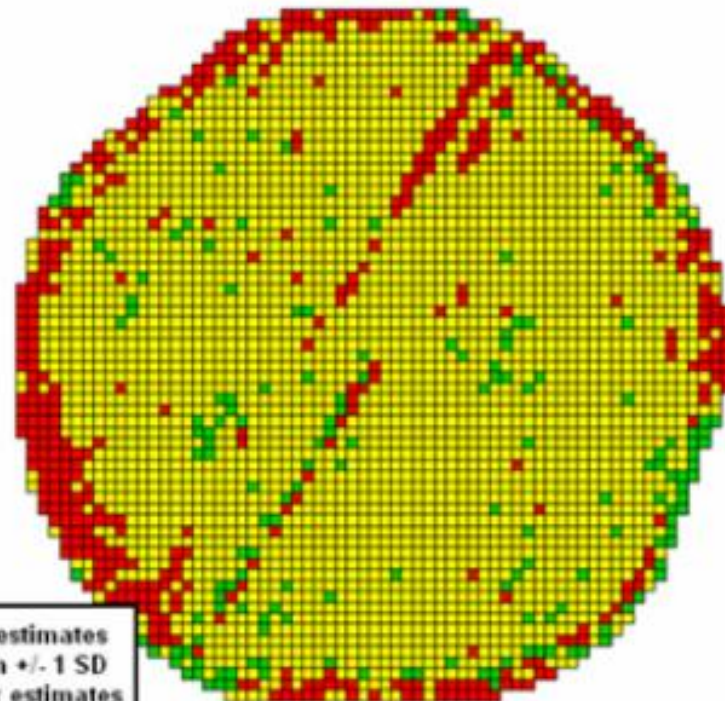


Případová studie VI

- Řešení?
- Stratifikace datové sady – rozdělení do skupin ze stejnými charakteristikami.



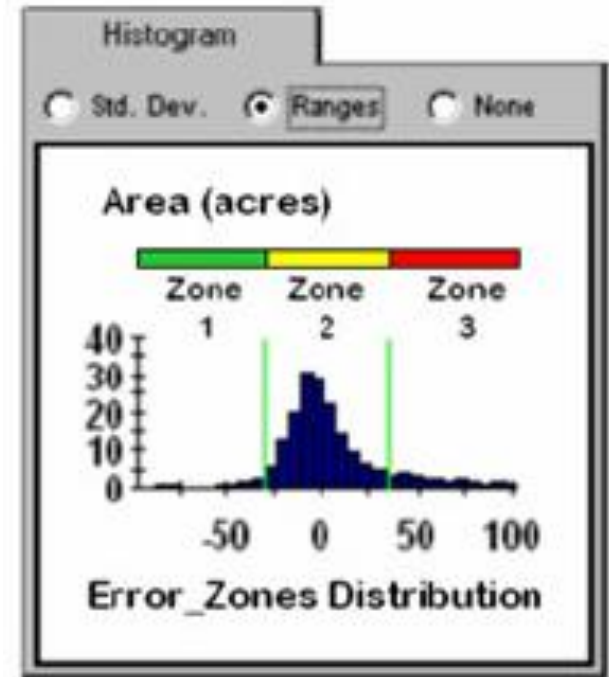
Error_zones



Případová studie VII

- Predikční rovnice bude lépe vystihovat jednotlivé vrstvy, než jedna rovnice pro celá data.
- Více technik pro stratifikaci.
- Využití histogramu – plus/minus směrodatná odchylka dělí histogram na 3 zóny.
- Predikce funguje pro zónu 2.
- Pro zóny 1 a 3 jsou výsledky pod a nadhodnocené.
- Specifická predikční rovnice pro každou zónu by měla dát lepší predikci.

Kartografické modelování

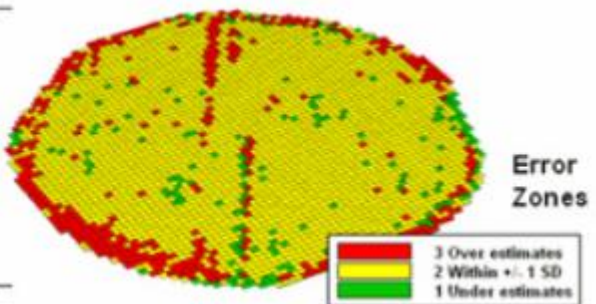




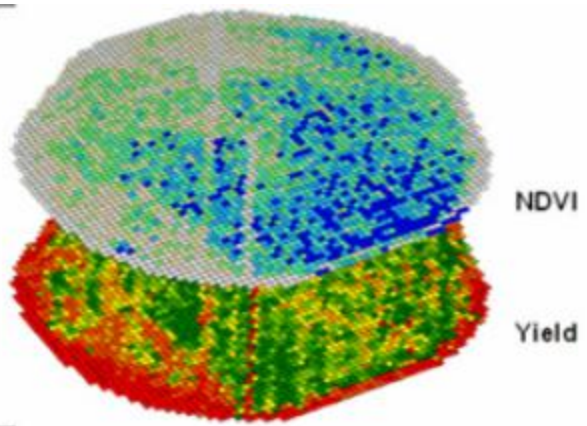
Výpočet predikce pomocí separovaných predikčních rovnic

- **Vstupy – NDVI a výnosová mapa.**
- **Algoritmus nejdříve zkontroluje mapu odchylek a určí, do které ze 3 zón daná oblast patří.**
- **Následně jsou použity regresní rovnice pro predikci po zónách.**
- **Složená predikční mapa vytvořena pomocí 3 rovnic a NDVI dat.**

Template Map

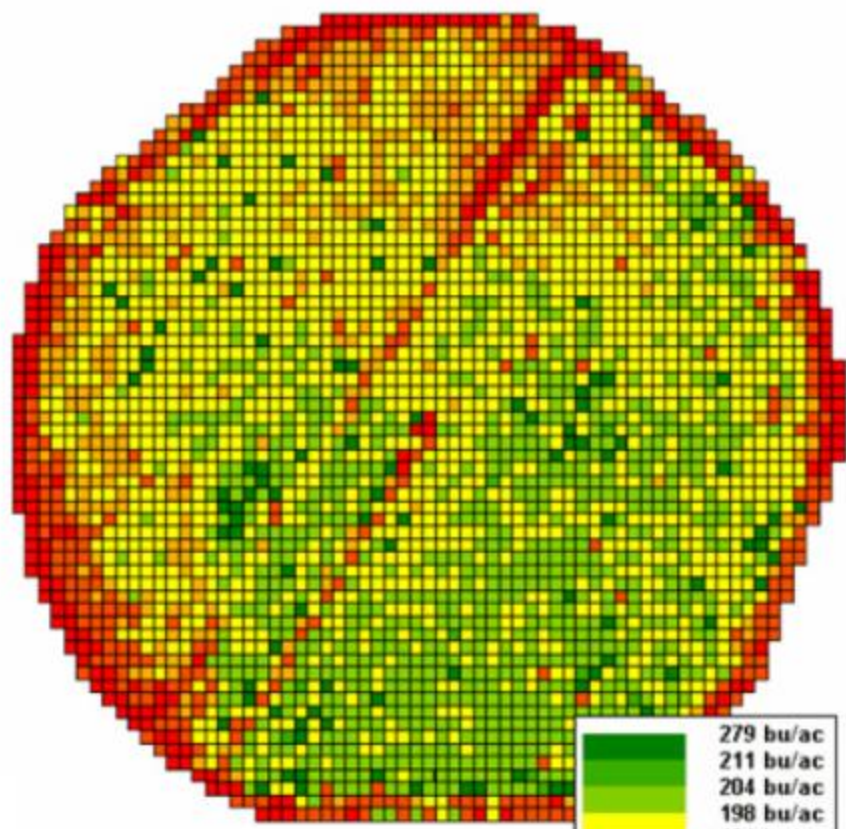


Data Maps

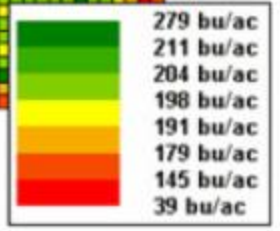


Prediction Equations

- Zone 1 ■ $Y = 145.40 + 110.79X$ ($R^2 = .68$)
- Zone 2 ■ $Y = 32.93 + 215.06X$ ($R^2 = .60$)
- Zone 3 ■ $Y = -4.85 + 169.38X$ ($R^2 = .42$)



Prediction_composite



Kartografické modelování

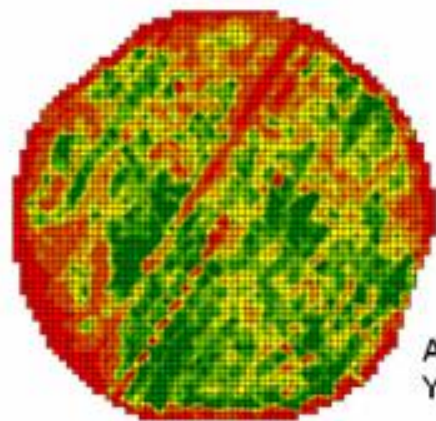
Případová studie VIII

- **Vizuální srovnání :**
 - Skutečné výnosy
 - Predikovaná mapa pro celou oblast
 - Predikovaná stratifikovaná mapa
- **Mapa odchylek pro stratifikovanou predikci – 80% odhadu je ± 20 q/ha.**
- **Průměrná chyba je pouze 4q/ha.**
- **Dobrá predikce úrody na základě DPZ více jak měsíc před sklizní 😊**

Min [>=]	Max [<]	Count	acres	% Gridded Area	Color
120	144	0	0	0	Dark Green
80	120	30	1.72	0.91	Light Green
20	80	477	27.4	15	Yellow-Green
0	20	1236	70.9	38	Yellow
-20	0	1380	79.2	42	Orange
-80	-20	165	9.47	5	Red-Orange
-120	-80	1	0.0574	0.03	Red
-173	-120	0	0	0	Dark Red

80%

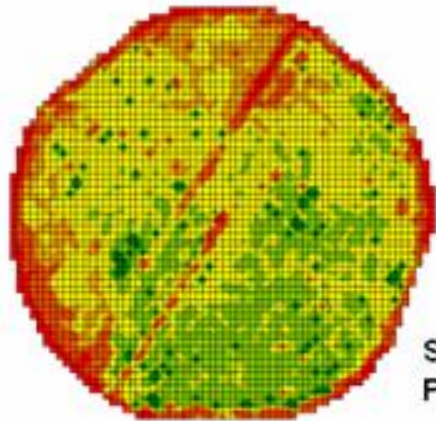
Statistics	
Min:	-81.2
Max:	113
Range:	194
Mean:	3.97
Median:	1.12
Std. Dev.:	19.2
Variance:	369
Gridded Area:	189 acres



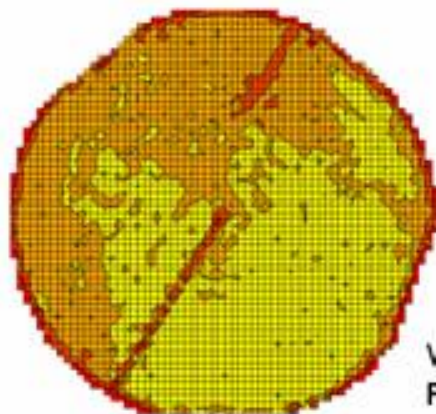
Actual Yield



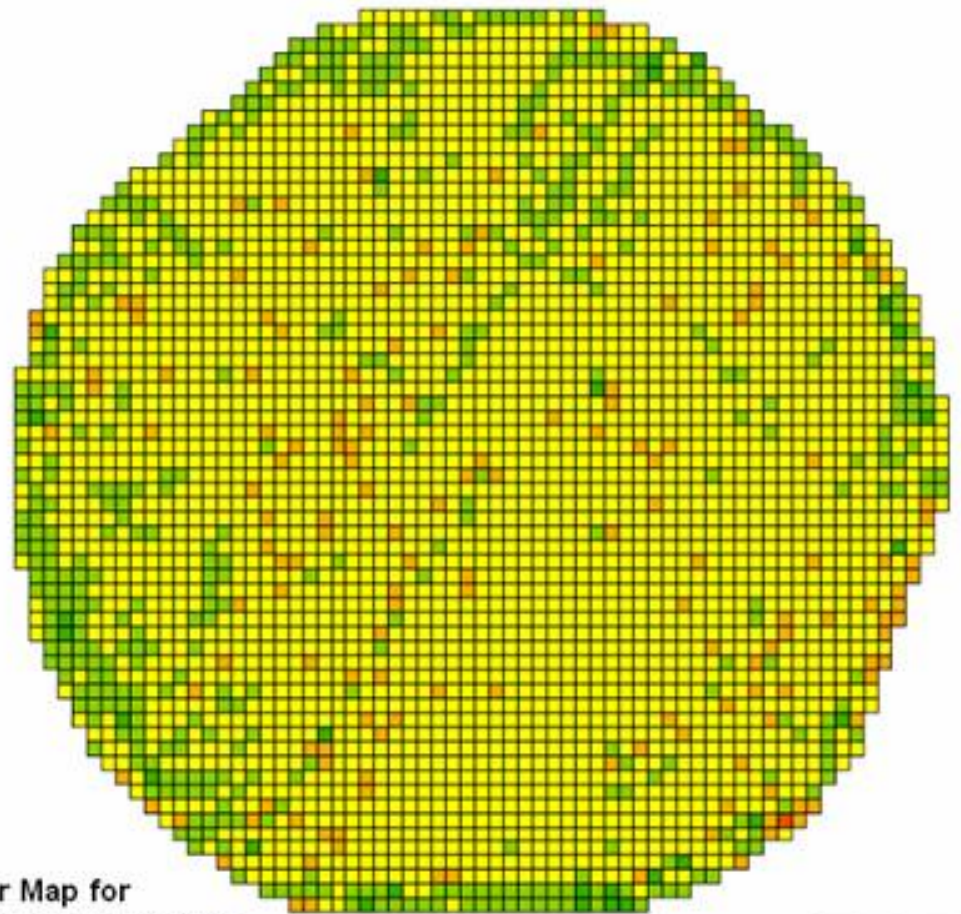
279 bu/ac
 243 bu/ac
 209 bu/ac
 175 bu/ac
 141 bu/ac
 107 bu/ac
 73 bu/ac
 39 bu/ac



Stratified Prediction



Whole Field Prediction



Error Map for Stratified Prediction

Min [>=]	Max [<]	Count	acres	% Gridded Area	Color
120	144	0	0	0	Dark Green
80	120	30	1.72	0.91	Light Green
20	80	477	27.4	15	Yellow-Green
0	20	1236	70.9	38	Yellow
-20	0	1380	79.2	42	Light Orange
-80	-20	155	9.47	5	Orange
-120	-80	1	0.0574	0.03	Red-Orange
-173	-120	0	0	0	Red

} 80%

Statistics

Min: -81.2
 Max: 113
 Range: 194
 Mean: 3.97
 Median: 1.12
 Std. Dev.: 19.2
 Variance: 369
 Gridded Area: 189 acres

Na co dávat pozor?

- **Odlišné způsoby stratifikace dat:**
 - *Prostorové zóny* – blízkost či vzdálenost od okraje pozemku.
 - *Závislé mapové zóny* – oblasti různé intenzity výnosů.
 - *Datové zóny* – půdní druhy, zrnitost, nutriční hodnoty.
 - *Korelační mapové zóny* – mikrorelief – hřbety a deprese.

Na co dávat pozor?

- Citlivé užití mapy odchylek – zejména pro další **extrapolaci v čase a prostoru**.
- Nutno využít dalších oblastí pro **kalibraci a validaci**.
- V oblasti precizního zemědělství například možnost kombinace detailního měření v relativně dlouhé periodě (DPZ) s častým měřením s omezeným prostorovým výskytem (vzorky, senzory).



Prediktivní modelování v ArcGIS

- **Obvykle se jednotlivé procedury modelování spouští samostatně a opakovaně.**

Možnost využít ModelBuilderu pro:

- 1) Zaznamenání všech **postupných kroků** v modelování;
- 2) Snadná **opakovatelnost** modelování a **sdílení** s dalšími uživateli;
- 3) Lepší **vizuální reprezentace**, která vede k lepšímu pochopení celého průběhu modelování..



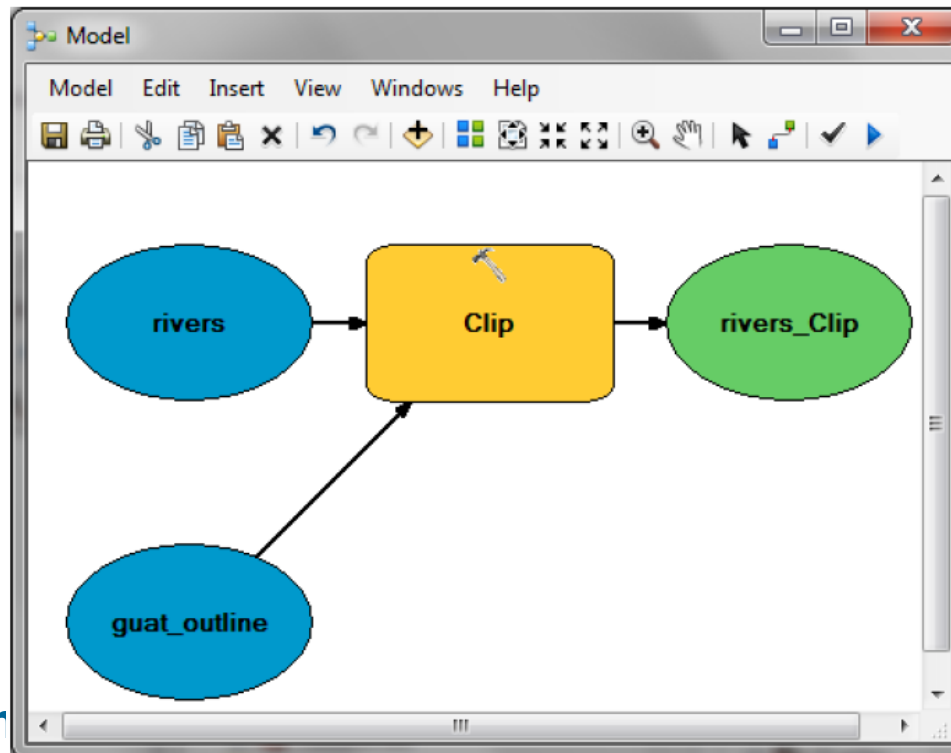
Prediktivní modelování archeologického naleziště

- Prediktivní modelování v archeologii – „nástroj pro vyjádření pravděpodobnosti výskytu archeologického naleziště kdekoliv v krajině“.
- Snaha určit pravidla a preference pro výběr lokality danou kulturou.
- Zahrnuje deskriptivní analýzu přírodních faktorů pro známé lokality a snahu najít společné opakující se kombinace.
- Příklad: vybraná kultura (Mayové) preferovala historicky známá místa v **blízkosti** oceánu a mokřadů s výskytem porostů endemita *Salvia apiana*.
- Která místa ve zkoumané oblasti odpovídají podmínkám??

Kartografické modelování

1. Omezení zkoumané oblasti

- Omezení oblasti na severní Guatemalu a oříznutí vybraných vodních toků pomocí funkce Clip.
- Vstupní a výstupní soubory + funkce.



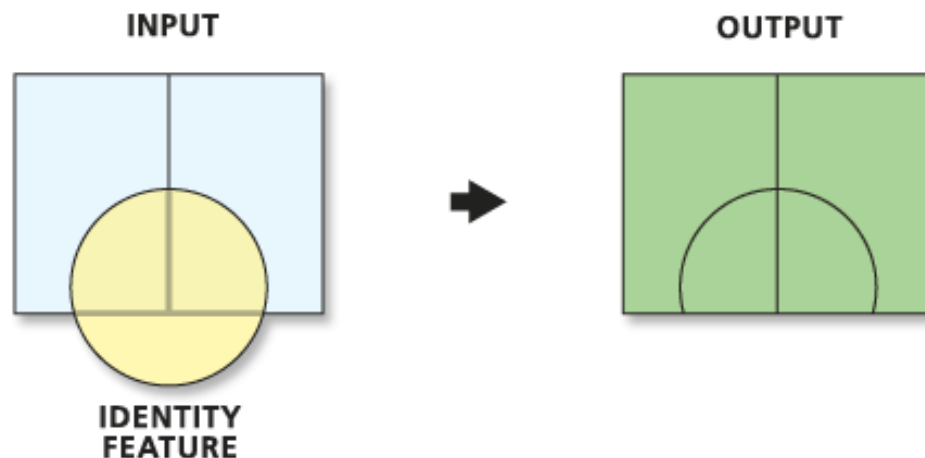


2. Změření vzdálenosti lokalit od řeky

- Říční síť nyní omezena na sledované území.
- Určení vzdálenosti potenciálních archeologických nalezišť od říční sítě – **Near**.
- Vyhledávací vzdálenost nastaveno na 5 km (=blízko).
- Všechny lokality blíže než 5 km mají určenou přesnou (vzdušnou) vzdálenost (**NEAR_DIST**).
- Ostatní lokality mají přiřazenu hodnotu -1.

3. Kombinace přírodních podmínek

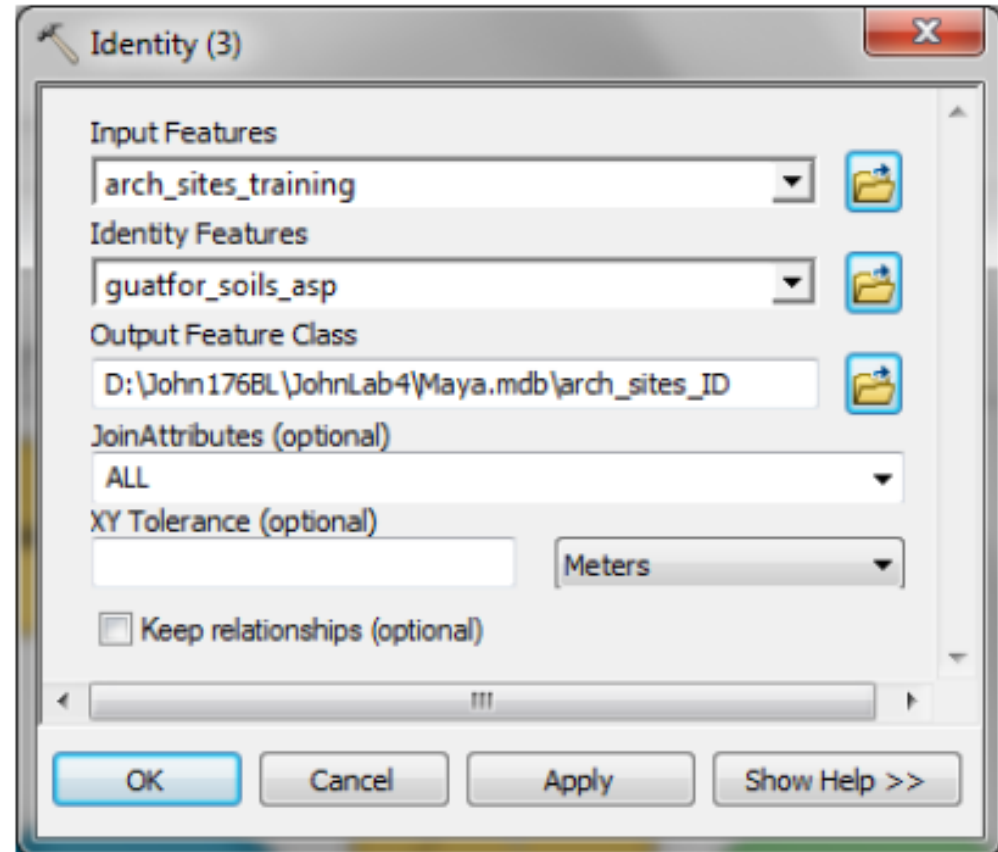
- Zjištění jaké přírodní podmínky obklopují naše archeologické lokality.
- **Vegetace – půdy – orientace svahu.**
- Nutná postupná analýza přírodních podmínek v několika krocích a postupné rozšíření atributové tabulky o přírodní ukazatele.
- Použití nástroje ***Identity***.
- Vegetace + půdy = PP1
- PP1 + orientace = PP2

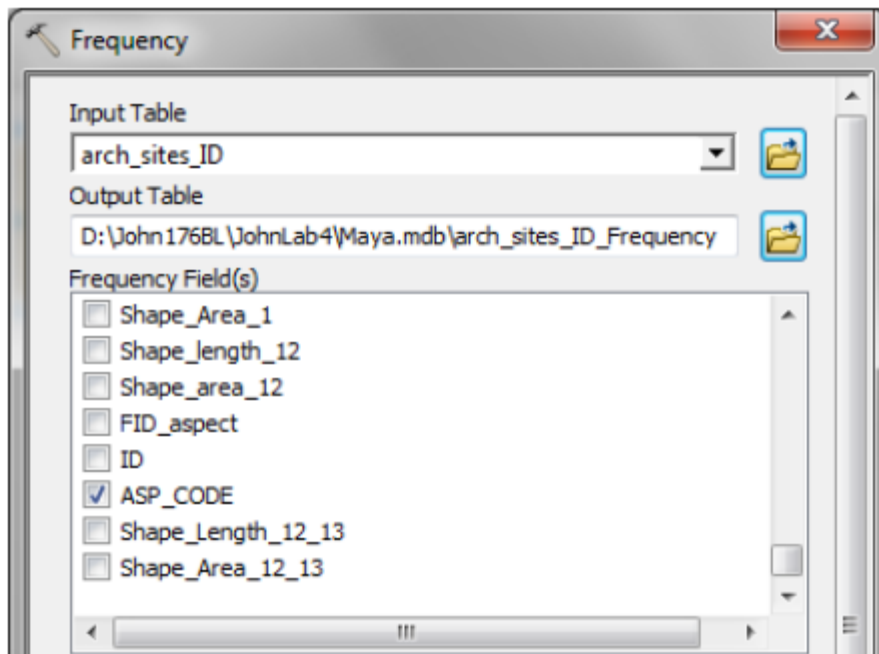




4. Přírodní podmínky pro archeologické lokality

- Spojení dat o archeologických lokalitách a PP2 pomocí nástroje **IDENTITY**.
- Následný výběr potřebných atributů z tabulky – nástroj Identity zachovává všechny atributy a vytváří další.
- Využití nástroje **Frequency**.

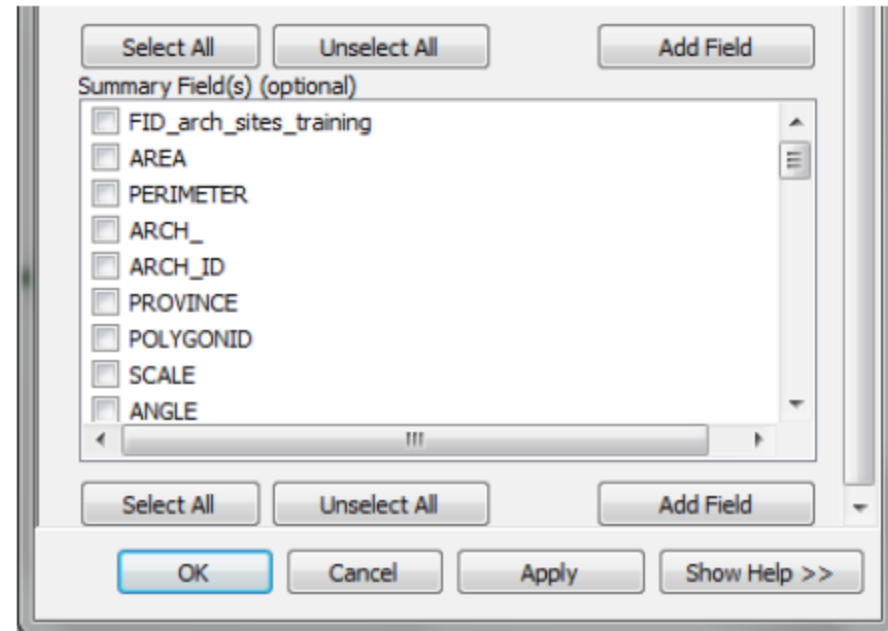




Výběr atributů

- **Nutno zachovat:**
 - NEAR_DIST - blízkost
 - DESC_vegetace
 - R_FERT - půda
 - ASP_CODE - orientace

Kartografické modelování





Finální model



Kartografické modelování

6. RUN a prozkoumání výsledků

- Určení hlavních shluků přírodních podmínek.
- Stanovení pracovních predikční hypotézy pro vybraná místa.
- Ověření hypotézy.

arch_sites_ID_Frequency

OBJECTID *	FREQUENCY	NEAR_DIST	DESC_	R_FERT	ASP_CODE
1	1	-1	Inland swamp forest	4	10
2	1	-1	Lowland rain forest	1	2
3	2	-1	Lowland rain forest	1	5
4	2	-1	Lowland rain forest	1	9
5	1	-1	Non forest	1	10
6	1	68.570929		2	10
7	1	177.68938		2	9
8	1	274.989335	Lowland rain forest	1	4
9	1	327.802407	Non forest	1	8
10	1	427.268735	Inland swamp forest	2	4
11	1	546.290435	Non forest	1	7
12	1	593.566121	Lowland rain forest	4	6