



Geoinformatika

V – Transformace dat

jaro 2023

Petr Kubíček

kubicek@geogr.muni.cz

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

IGC

SVĚTOVÝ DEN METEOROLOGIE 23. BŘEZNA



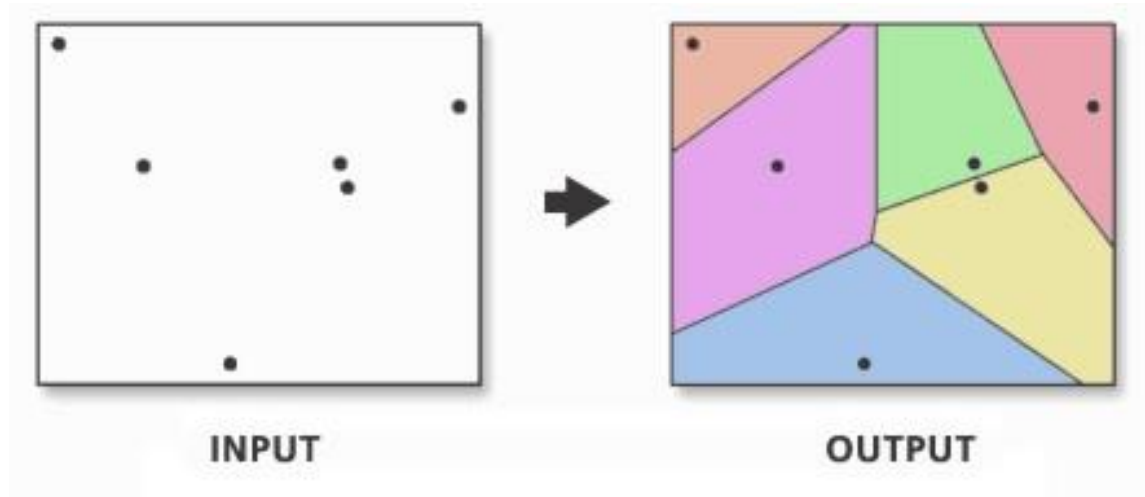
WORLD
METEOROLOGICAL
ORGANIZATION



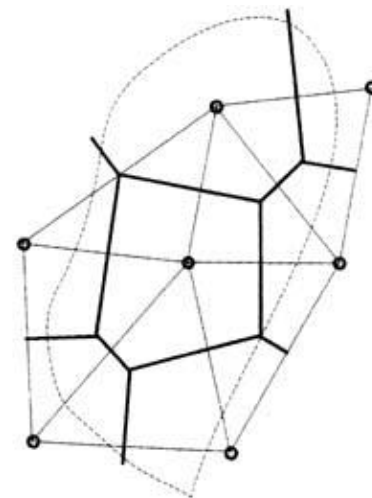
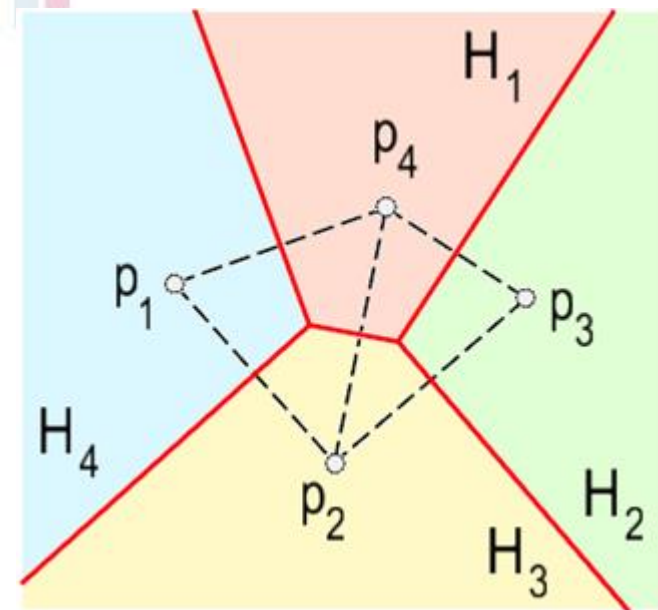
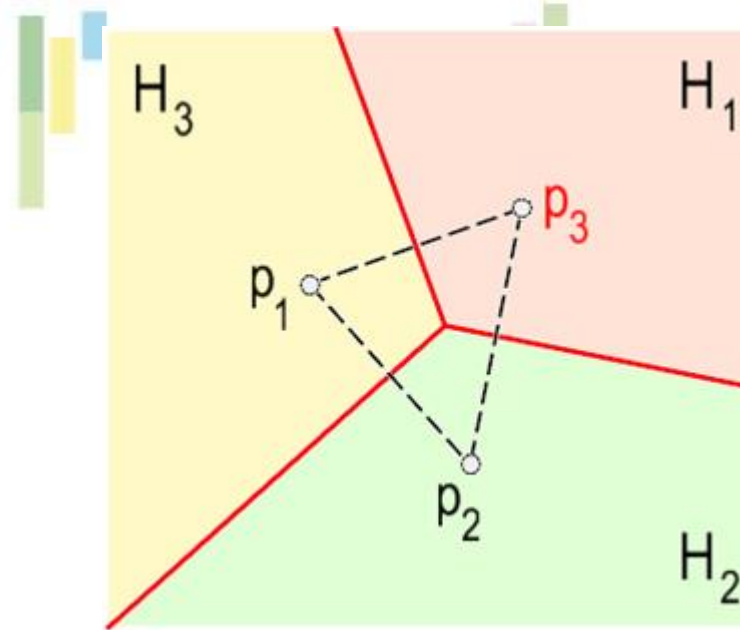
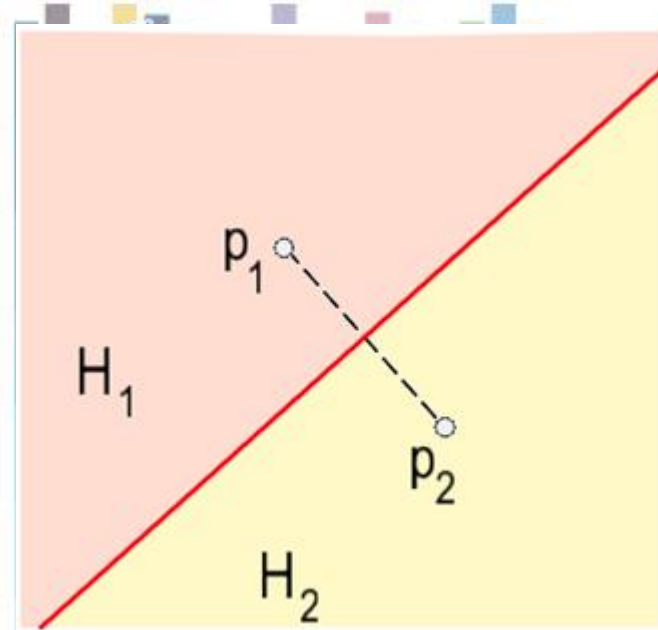
Dělení prostoru a Thiesenovy polygony



Analýzy sousedství (Proximity analysis) –



- Každá polygon obsahuje jeden bod vstupního souboru. Každé místo polygonu je blíže k tomuto bodu, než k jakémukoliv dalšímu bodu vstupního souboru.
- Thiesenovy polygony, Voronoi cell (Voroného tesalace) – konstrukce??



- rain gauges
- catchment boundary
- lines joining nearby stations
- Thiessen Polygon network

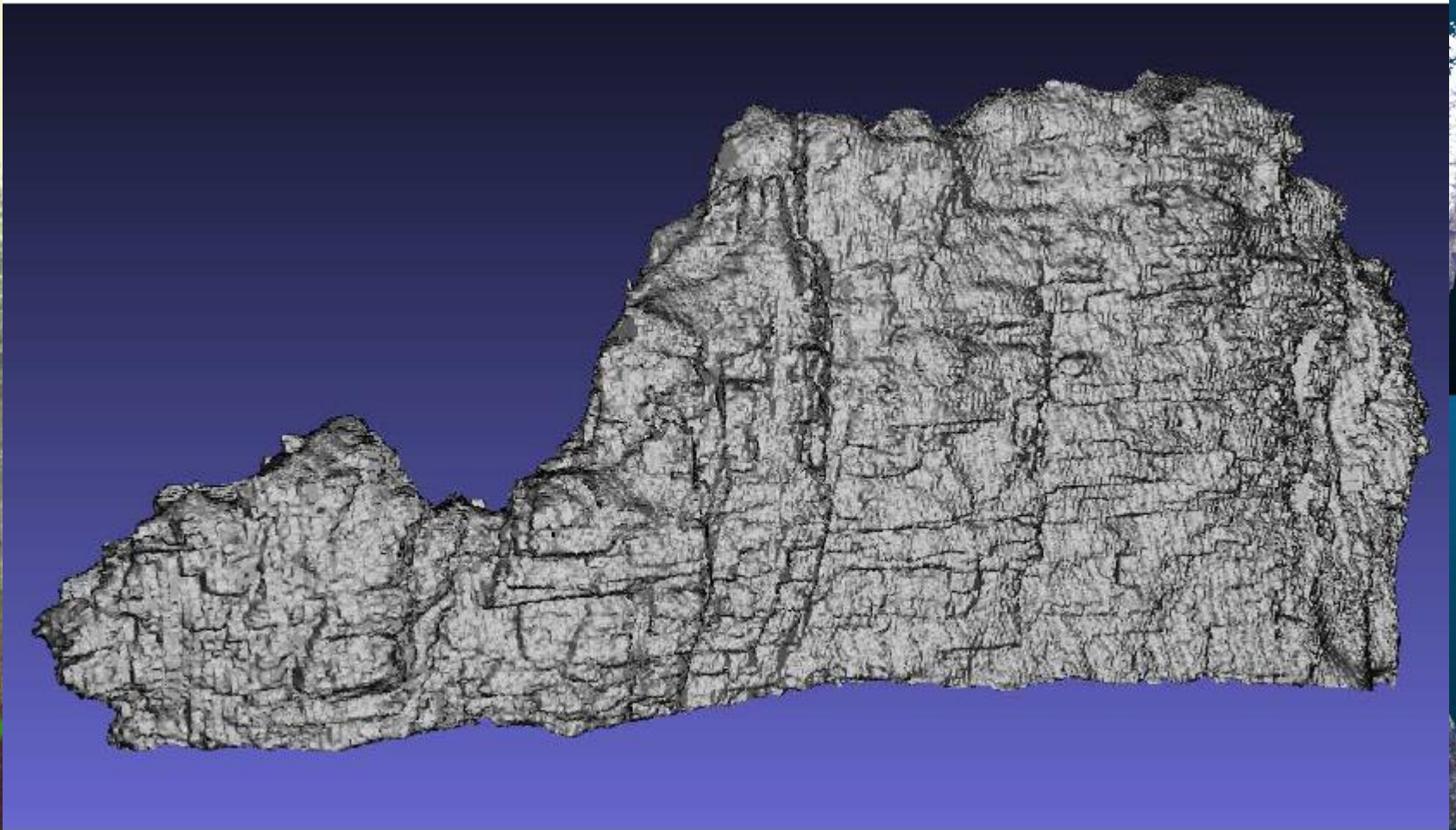
Thiessen, A.H., 1911.
 Precipitation for large areas, Monthly Weather Review, 39, 1082–1084.

Laserové skenování – LIDAR

- **Light Detection and Ranging (LIDAR)**
- Princip LIDAR je postaven na aktivním senzoru, který vyšle laserový paprsek a zaznamená jeho zpětný odraz.
- Odrazů může být několik, první je považován za digitální model povrchu (DMP, DSM – digital surface model). Poslední odraz je pak považován za digitální model reliéfu (DMR, DTM – digital terrain model).
- Teprve potom se senzor otočí a zaznamenává další bod!
- Existují letecké a pozemní scannery!



Horizontální skenování



Sphere8

Sekundární zdroje dat

- **Sekundární zdroje dat jsou již jednou zpracované primární zdroje**
- **=> jsou v nich obsaženy chyby získané již během prvního zpracování dat, tudíž nemohou být přesnější než zpracovávané primární zdroje.**
- **Možnosti jejich vstupu do GIS**
 - manuální vstup přes klávesnici (pracné, zdlouhavé)
 - digitalizace
 - skenování a vektorizace
 - import dat.

Vektorizace

- **Automatická** - vše dělá počítač. Je to velice rychlé (co se týče nároků na uživatele), ale je nutné provádět čištění vektorových dat.
- **Polautomatická** - interaktivní metoda, s tím že počítač sám vektorizuje, ale uživatel jej koriguje na sporných místech (ArcScan).
- **Ruční** (on screen digitizing) - interaktivní, kdy uživatel provádí sám vektorizaci na základě rastrového podkladu. Některé systémy umožňují automatizovat alespoň přichycení na rastr.



Manuální zadávání atributů

- Nejběžnější způsob zadávání atributových dat je manuálně, **pomocí klávesnice**, na což stačí pouze jednoduchý hardware.
- Možné problémy s **integritou** dat – lze kontrolovat.
- Atributy se následně navazují na prostorovou část pomocí **unikátního identifikátoru**, který prostorové prvky již obsahují (vytváří se obvykle již při jejich tvorbě).
- Kontrola **správnosti** zadaných údajů.
 - **Single Key Data Entry** - jeden operátor zadává atributová data a druhý operátor již zadaná data kontroluje (porovnává originál s vytištěnými výpisy, ...).
 - **Double Key Data Entry** - atributová data jsou zadávána dvěma na sobě nezávislými operátory (každý zadává stejná data) a poté se obě varianty v počítači porovnají. Při nalezení rozdílných hodnot se zadaný atribut překontroluje a opraví. Metoda se používá spíše na větší projekty, u kterých velice záleží na správnosti zadaných údajů.



Skenování + rozpoznávání textu

- Další možností je scannování textu obsahující žádané atributy a poté jeho automatizované rozpoznávání pomocí nějakého **OCR** (Optical Character Recognition - nástroje na rozpoznávání písma) software.
- Tato metoda, ačkoli relativně velice rychlá, je stále **úspěšná jen z části** a je možné ji aplikovat většinou pouze na již tištěný text (i z psacího stroje). Po automatickém převodu je navíc nutné vše pečlivě zkontrolovat (podobně jako u manuálního zadání pomocí metody Single Key Data Entry).
- Problémy s diakritikou.
- Další nevýhodou je obvyklá **nutnost ručního navazování atributů** na prostorovou část, podobně jako u ručního zadávání dat.

Chyby v datech

- Při vkládání dat do systému **není možné zabezpečit správnost 100% zadání dat.**
- **Identifikace chyb** je velice obtížná. Obvykle se data kontrolují **vizuálně**. Dalším způsobem kontroly chyb prostorových dat je proces vytváření topologie neboli **topologické čištění dat.**
- GIS mají většinou schopnosti **procházet místa s potenciální chybou** a umožní uživateli interaktivně odstranit případné chyby.



Možné chyby při zadávání

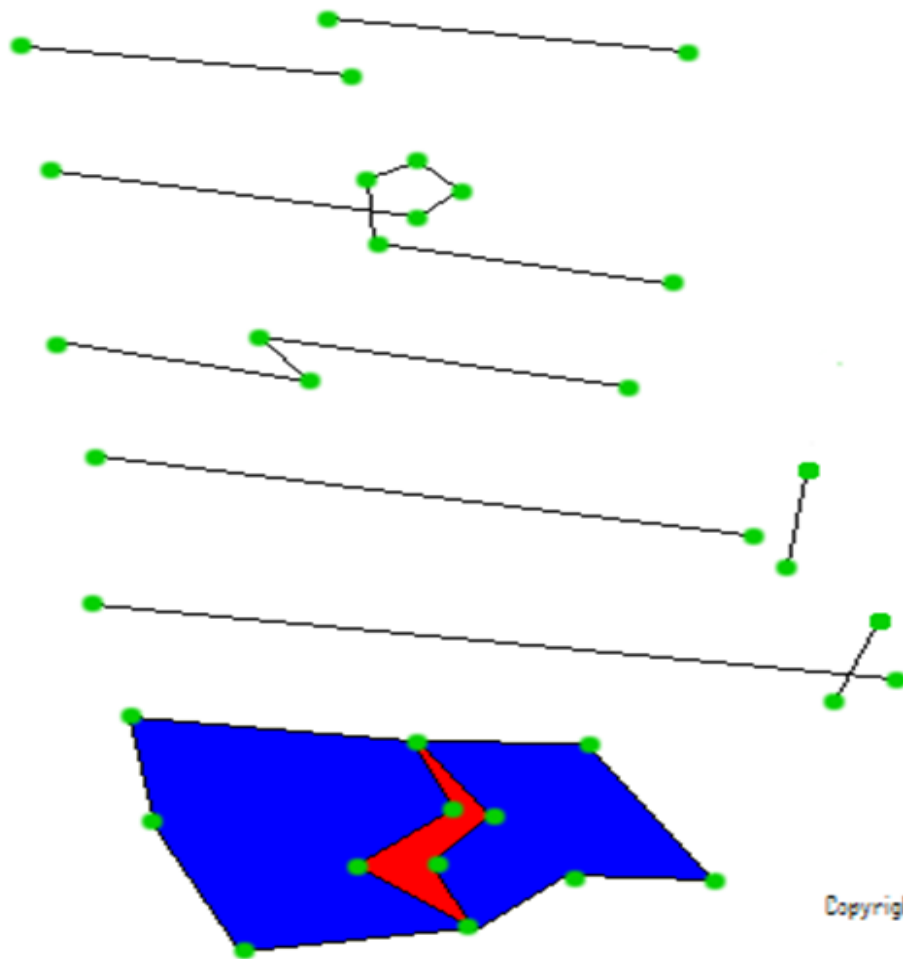
- **Nekompletnost dat** - scházejí body, linie, polygony.
- **Chybné umístění** prostorových dat - chyby vycházející ze špatné kvality vstupních dat nebo z nedostatečné přesnosti při digitalizování.
- **Zkreslení prostorových dat** - chyby z nepřesností vstupních dat (deformace podkladových dat, zkreslení již existující analogové kresby).
- **Špatná vazba** mezi prostorovými a atributovými daty.
- **Atributy jsou chybné** nebo nejsou kompletní – velice častá chyba zvláště pokud jsou atributy pořizovány z různých zdrojů v různých časech.



Chyby při vytváření topologie

- **Třísky a mezery (Sliver and gaps)** - jev nastává, když jsou dvě hranice digitalizovány z různých zdrojů, ačkoli v terénu představují jednu a tu samou. V takovém případě jsou linie představující tutéž hranici neidentické (nepřerývají se)
- **Volné konce (dead ends)** - nedotahy a přetahy.
- **Duplikátní linie** (hlavně v CAD, ale i u některých GIS, které z toho vytváří regulární polygon) reprezentující stejný objekt.
- Pokud se používá pro reprezentaci polygonů metoda hranic a centroidů, tak i **přiřazení více centroidů jednomu polygonu**.

Obvyklé chyby v GIS geometrii



A
B
C
D
E
F



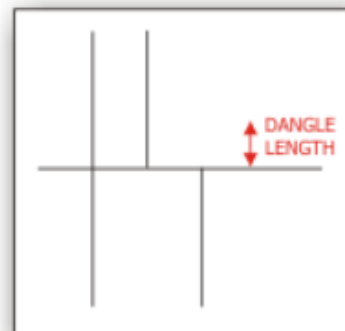
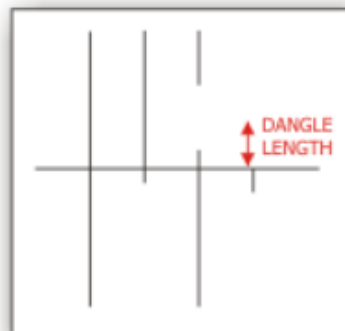
Copyright 2012 DMS-Mapping

Trim line (ArcGIS Pro)

Geoinformatika

INPUT FEATURES

OUTPUT
KEEP SHORT FEATURES



Chyby právního charakteru

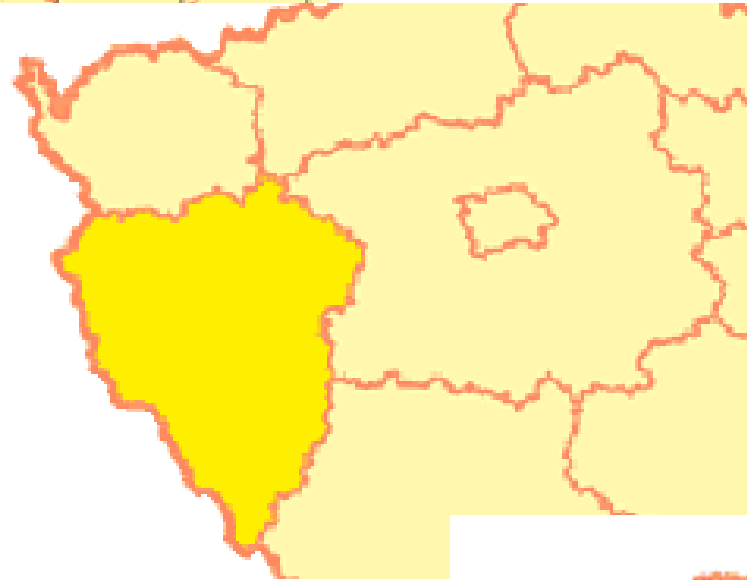
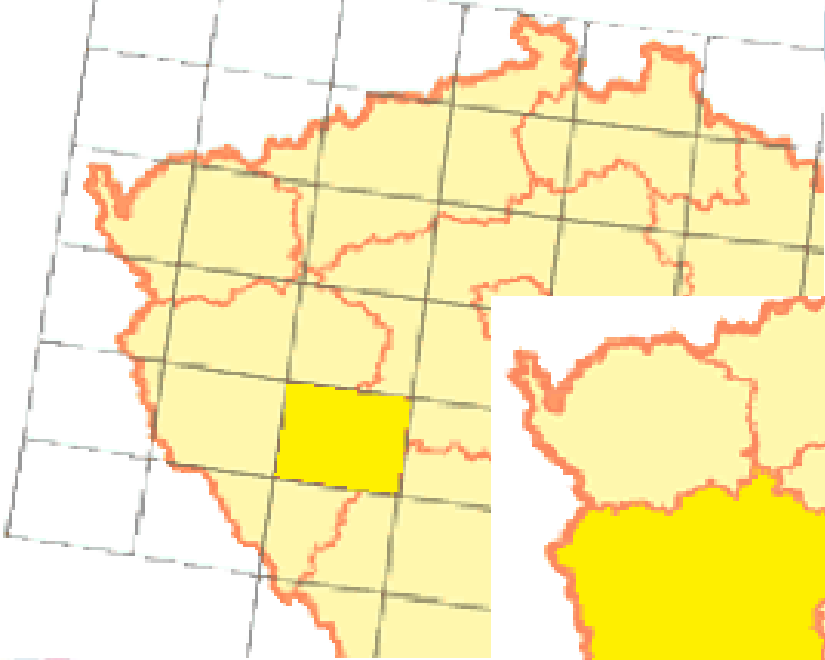
- Při pořizování dat je nutné brát v potaz i právní souvislosti problematiky, kdo má na data obchodní práva, zda je možné data využívat pro akademické, soukromé, či obchodní účely.
- Zdroje obvykle přesně popisují možnosti využití a omezují zejména komerční či veřejné použití dat (i jako podkladu).
- **Ochrana dat (vodotisk, záměrné chyby).**





Uchovávání a zpracování dat

- **Pravidelné (např. mapové listy).**
 - Na disku je každý mapový list v jednom souboru (resp. ve více souborech se stejným jménem, lišících se pouze příponou) či adresáři.
- **Nepravidelné (mapové listy, zájmové území - katastrální území, území národního parku, okresu, kraje ...).**
 - Na disku je každé zájmové území v jednom souboru (resp. ve více souborech se stejným jménem, lišících se pouze příponou) či adresáři.
- **Bezešvé (Seamless)**
 - Celé zájmové území je uloženo v jednom souboru, adresáři či databázi).



Geoinformatika



Komplexní GIS schéma

Transformace dat

- modelu
- polohy
- formátu

Sběr dat

- editace
- import

Uložení dat

- | | |
|-------------|------------------------------|
| Návrh | - struktura
- datové typy |
| Manipulace | - dotazování
- indexování |
| Dokumentace | - metadata |

Analýza dat

- průzkum
- modelování

Prezentace dat

Vizuální

- kartografická
 - statické mapy
 - dynamické mapy
 - uživatelské rozhraní
- nekartografická
 - grafická
 - textová

Nevizuální

- export
- řídicí povely

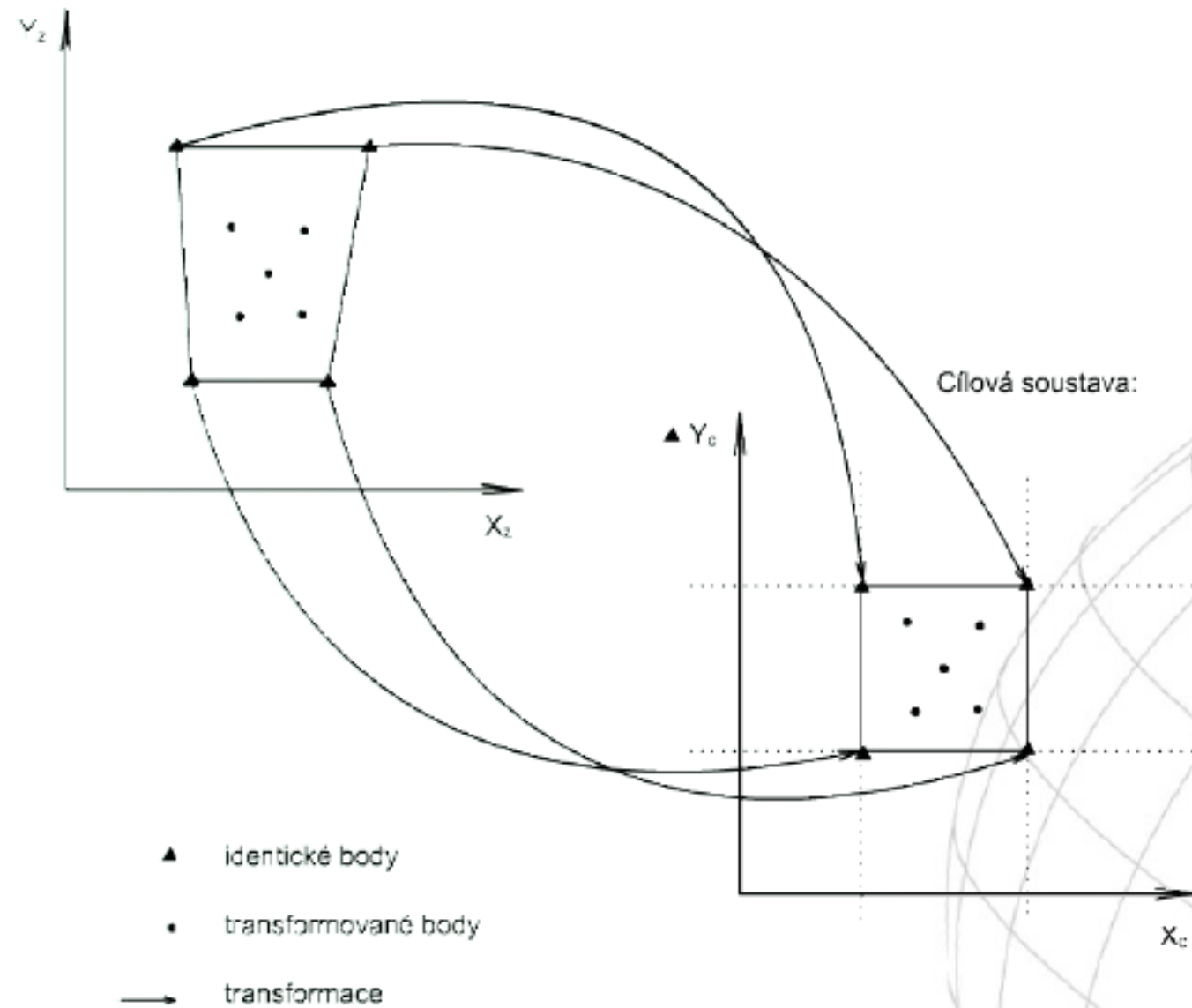


Transformace

- **Polohová – geometrická transformace**
 - Lineární
 - Afinní
 - Projektivní
- **Datového modelu**
 - RAVE – rastr to vector
 - VERA – vector to rastr
- **Formátu**

Geometrické transformace

Zdrojová soustava:

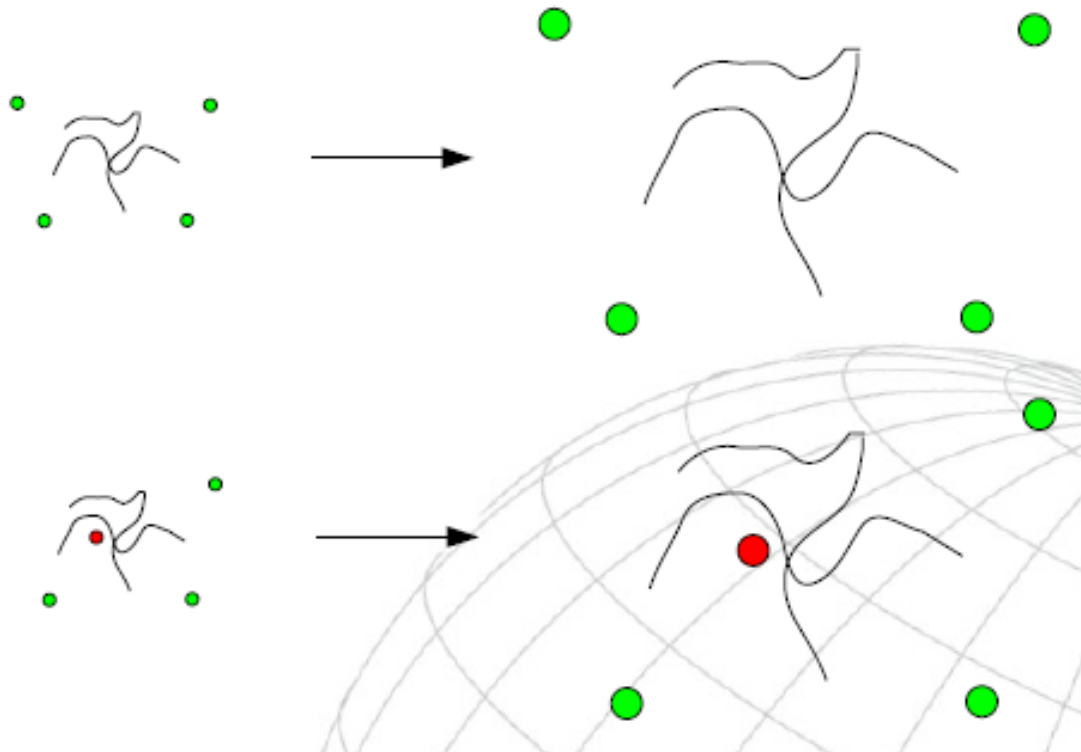


Transformace mezi rovinnými pravoúhlými souřadnicemi jsou založeny na poznání přesné polohy vybraných identických bodů.

Geometrické transformace - volba identických bodů

- U výběru dvojic identických bodů je také vhodné mít na paměti, že je nutné je vybírat co nejbližže okrajům transformovaného území, aby nebyly způsobeny nežádoucí deformace na okrajích.

» dobře





Geometrické transformace - identické body a transformační koeficienty

- **Transformační koeficienty** jsou hodnoty, vypočtené z dvojic identických bodů, kterými se vyjadřuje **přechod od zdrojové souřadnicové soustavy do cílové.**
- U transformace se ale obvykle **používá více identických bodů, než je nutné** pro výpočet transformačních koeficientů.
- Hodnoty transformačních koeficientů se pak vypočtou **metodou nejmenších čtverců**, kde se minimalizuje suma rozdílů v poloze mezi souřadnicemi transformovaných bodů (více – Matematická kartografie).
- **Transformace** je například **posun a změna měřítka.**



Geometrické transformace – typy transformací

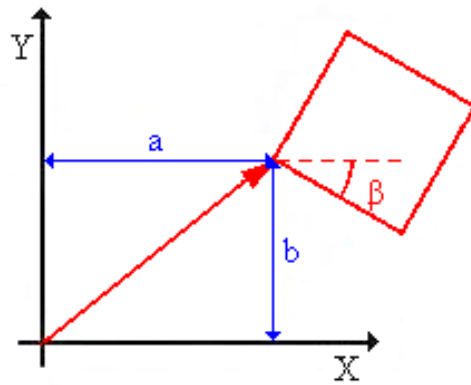
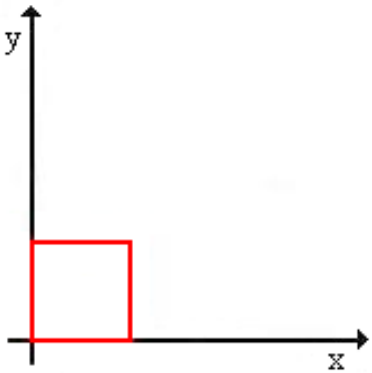
Transformace souřadnicového systému mezi rovinnými pravoúhlými souřadnicemi:

- Lineární konformní transformace (LKT)
- Afinní transformace (polynomická prvního řádu a polynomické transformace vyšších řádů)
- Projektivní transformace



Lineární konformní transformace

- $X(X,Y)$ - nové souřadnice
- $x(x,y)$ - staré souřadnice
- B - úhel otočení
- m - změna měřítka
- $p(a,b)$ - posun
- **Transformační koeficienty** (m, B, a, b) lze vypočítat již ze dvou dvojic identických bodů $(X1,Y1), (X2,Y2)$ a původní $(x1,y1), (x2, y2)$.



Zápis rovnicí

$$X = m \cdot \cos(B) \cdot x - m \cdot \sin(B) \cdot y + a$$

$$Y = m \cdot \sin(B) \cdot x + m \cdot \cos(B) \cdot y + b$$

Helmertova transformace – speciální případ LKT; $m = 1$

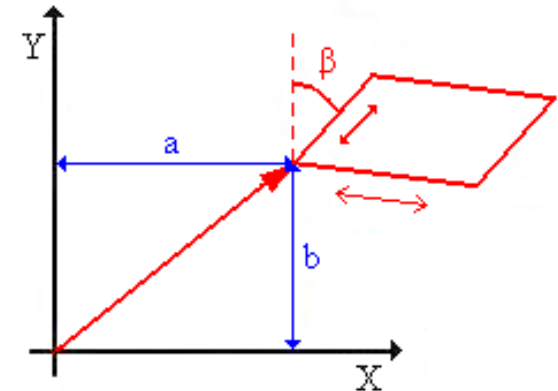
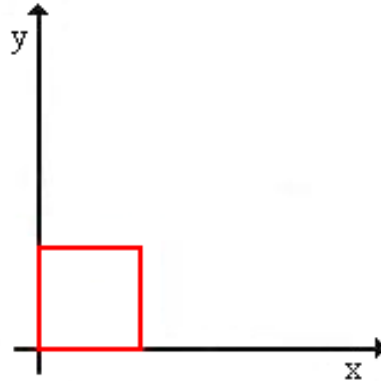


Lineární konformní transformace

- Posun
- Rotace
- Uniformní změna měřítka (v obou osách stejná)
- **Zachovává tvar objektu! (konformní)**
- Je potřeba dvou dvojic identických bodů

Afinní transformace

- **Jednotlivé souřadnice nejsou na sobě závislé – změna měřítka v různých směrech.**
- **$X(X,Y)$** - nové souřadnice
- **$x(x,y)$** - staré souřadnice
- **A** - regulární matice
- $p(c,f)$ - posun
- Transformační koeficienty (a, b, c, d, e, f) lze spočítat ze tří dvojic identických bodů.



Zápis rovnicí

$$X = a \cdot x + b \cdot y + c$$

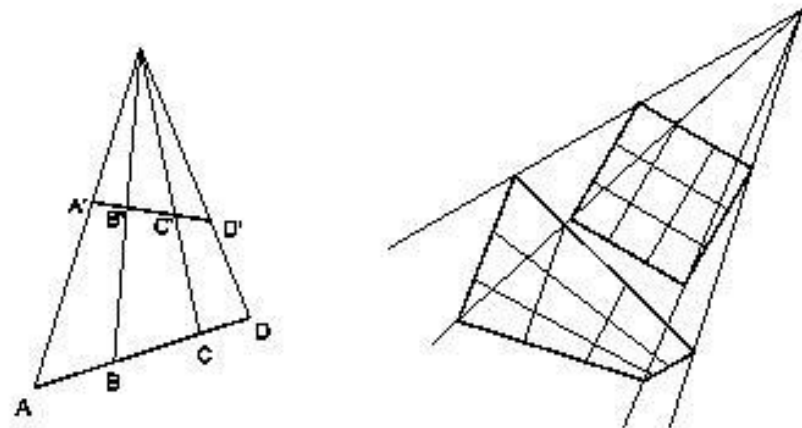
$$Y = d \cdot x + e \cdot y + f$$

Afinní transformace

- Posun
- Rotace
- Neuniformní změna měřítka (v každé ose jinak – zkosení)
- „Z obdélníka kosodélník“
- Je potřeba **tří dvojic identických bodů**

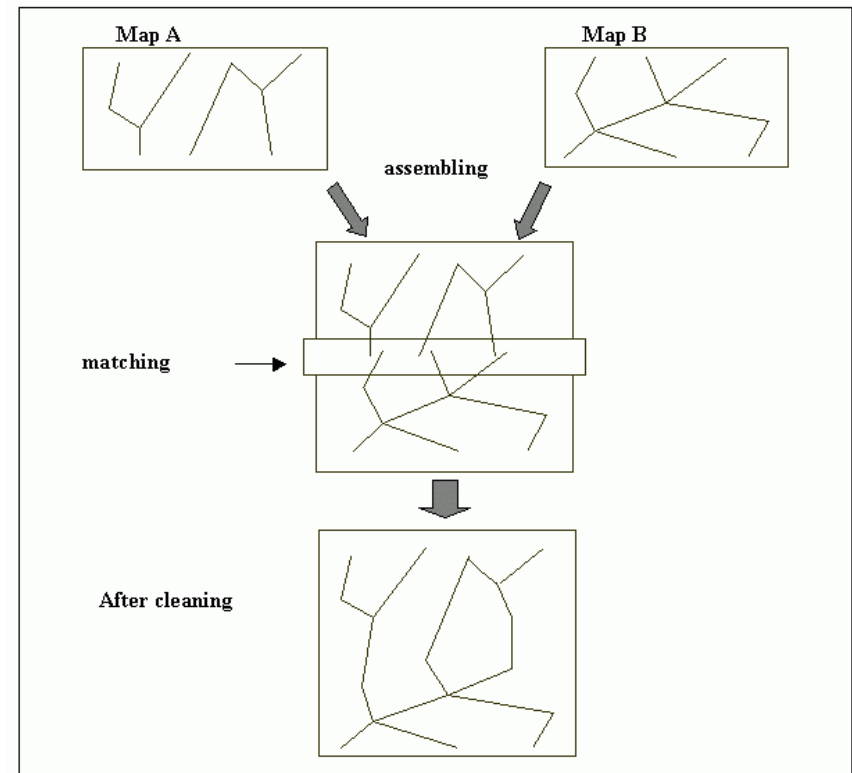
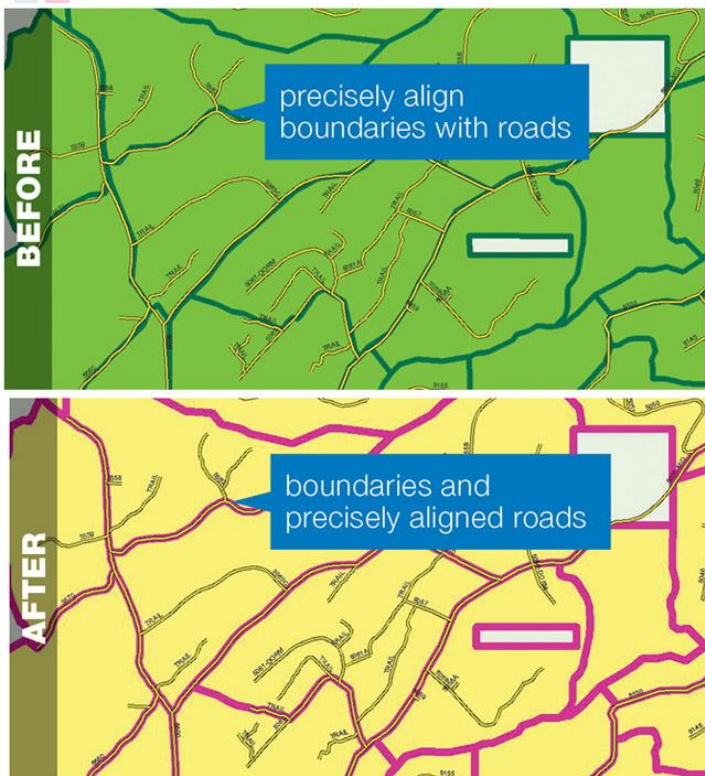
Projektivní transformace

- Transformace jednoho rovinného prostoru do druhého – vhodné pro data s menšími deformacemi.
- Posun
- Rotace
- „Z obdélníka lichoběžník“
- Je potřeba **čtyř dvojic identických bodů**

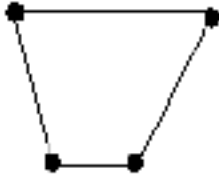
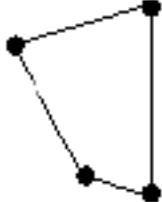





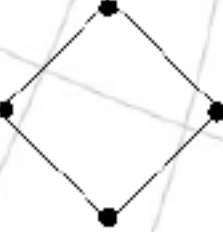


Dalsi typy geometrických transformací - Rubber sheeting, edge matching

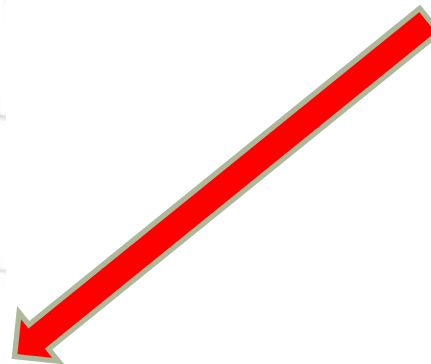
- **Rubber sheeting** - pro zdeformované mapy - lineární transformace po částech.
- **Edge Matching** - sjednocení okrajů mapy. V důsledku dělení plochy na mapové listy, odpovídá rubber sheetingu, ale platí pouze pro okraje mapových listů.



Projevy geometrických transformací

Transformace	Před	Po
Projekce		
Afinita		
Změna měřítka		
Rotace		

- **Helmertova?**
- **LKT, $m=1$**





Transformace datového modelu

- Jelikož pro některé analýzy jsou vhodnější vektorové reprezentace dat a pro jiné zase rastrové, GIS systémy pracující s oběma typy nabízejí nejrůznější nástroje umožňující a usnadňující převod mezi oběma reprezentacemi.
- Převod z rastrové do vektorové podoby se nazývá **vektORIZACE (RAVE)**, opačný proces z vektorové do rastrové podoby je **rasterizace (VERA)**.



Generalizace

Proč vůbec je generalizace v GIS potřebná:

- **Ekonomické požadavky** - svět nelze nikdy modelovat úplně přesně, vždy je to kompromis přesnost/cena.
- **Požadavky redukce objemu dat**
 - čím více je dat, tím je větší možnost udělat chybu a čím je přesnější (intenzivnější) měření, tím je větší šance ovlivnění dílčích měření individuální chybou.
 - generalizace slouží k odfiltrování těchto chyb a konsolidaci.
- **Víceúčelovost požadavků pro údaje** - z jedné digitální reprezentace dat je nutné vytvářet mapy s různými informacemi i v různých měřítkách, často velice rozdílných.
- **Požadavky zobrazování a komunikace (percepce-vnímání) dat**
 - vychází z kartografických doporučení některých limitů, při jejichž překročení se mapy stávají nečitelnými (př. max 10 gr. znaků na cm²).

Přehled metod

Vybrané generalizační metody užívané v kartografii a GIS

- **Selekce** (výběr prvků)
- **Eliminace** (odstranění prvků)
- **Zjednodušení** (zjednodušení prvků)
- **Agregace** (kombinování malých prvků do větších)
- **Prostorová redukce** (collapse)
- **Typifikace** (redukce hustoty prvků)
- **Zvýraznění** (přehnání, exaggeration)
- **Reklasifikace** a spojení (spojení prvků se stejnými vlastnostmi)
- **Řešení konfliktů** (posunutí méně důležitých prvků)
- **Redukce vrcholů** (Coordinate Thinning)

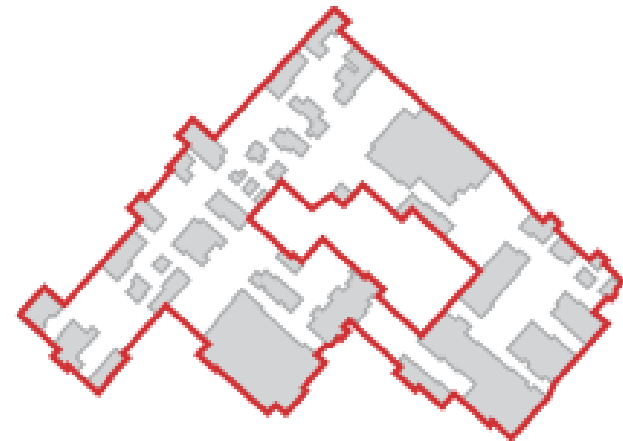
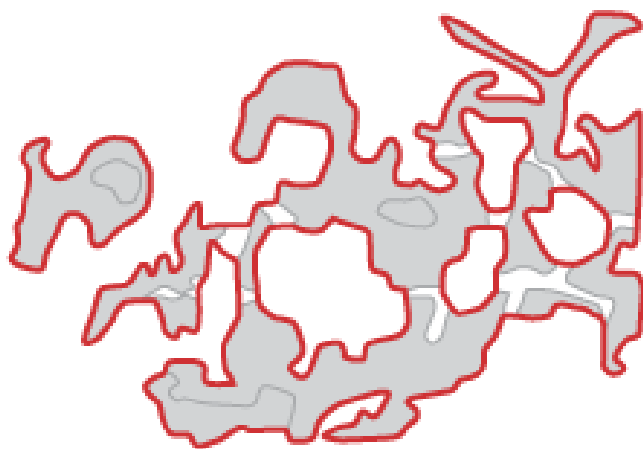


Nástroje generalizace v ArcGIS – příklady a užití

- **Aggregate** Points, Polygons
- **Collapse** Dual Lines To Centerline
- **Merge** Divided Roads
- **Simplify** Building, Line, Polygon
- **Smooth** Line, Polygon

Aggregate Points, Polygons

- **Kombinování menších prvků do větších – nahrazení shluku bodů či objektů (polygonů) jedním velkým objektem.**



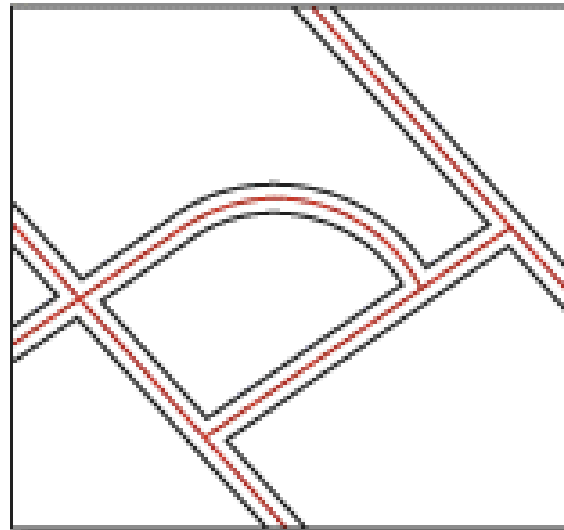
■ Input Feature
■ Aggregated Feature

A) Nonorthogonal features

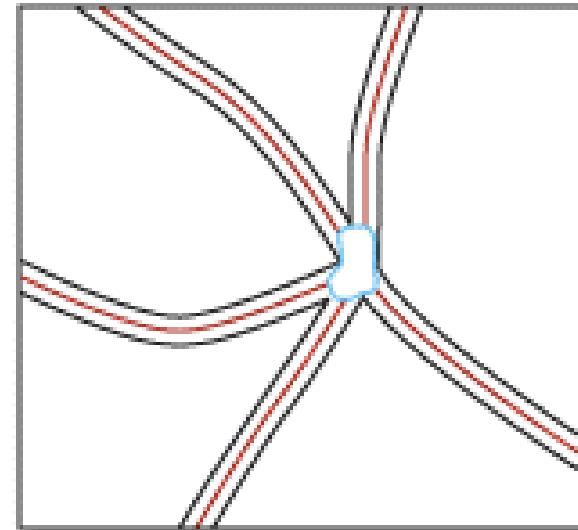
B) Orthogonal features

Collapse Dual Lines To Centerline

- **Prostorová redukce – obrysové linie nahrazeny centrální linií.**



Simple Case



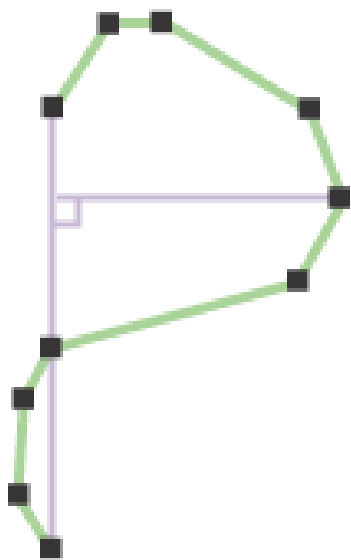
Complicated Intersection

- Casings
- Centerlines (LTYPE 1)
- Unresolved (LTYPE 2)

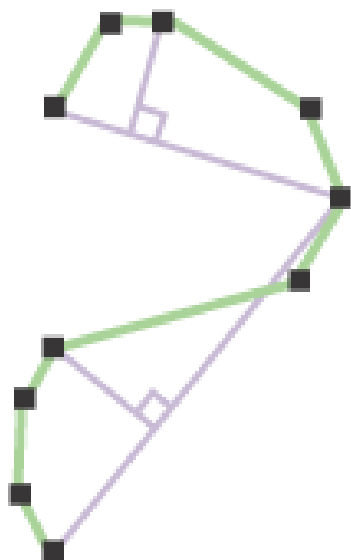


Simplify Building, Line, Polygon

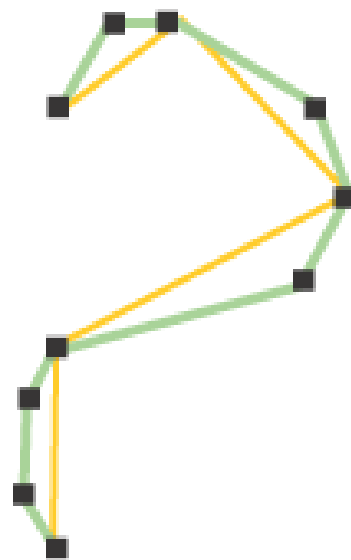
- **Douglas –Peucker algoritmus – zachovává základní tvar křivky. Čím větší tolerance, tím jednodušší křivka.**



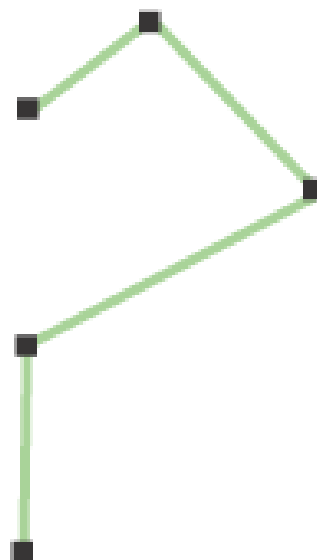
1ST TREND LINE



2ND TREND LINE

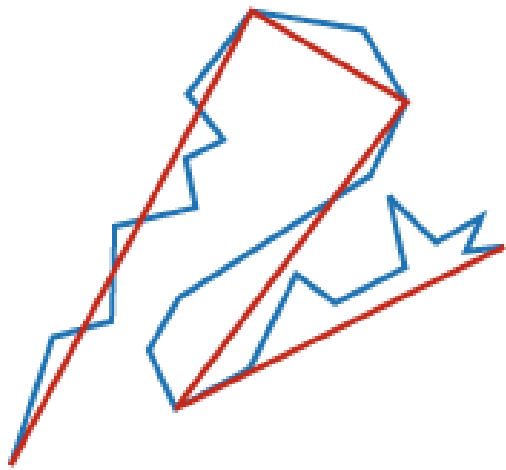


3RD TREND LINE

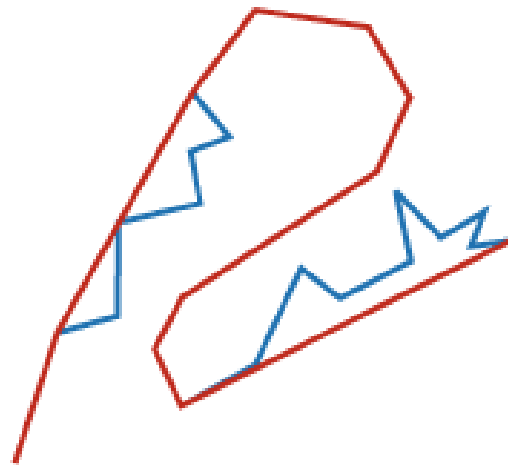


RESULTING ARC

 SIMPLIFICATION TOLERANCE



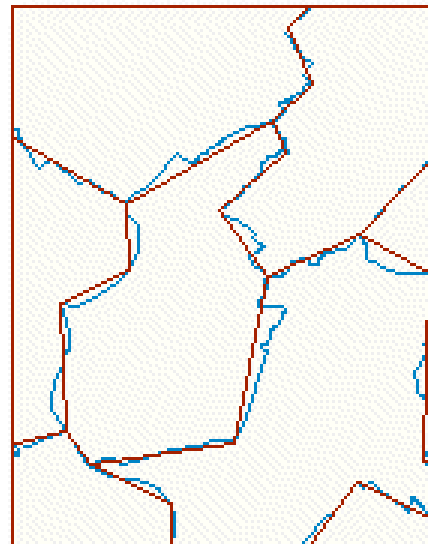
POINT REMOVE



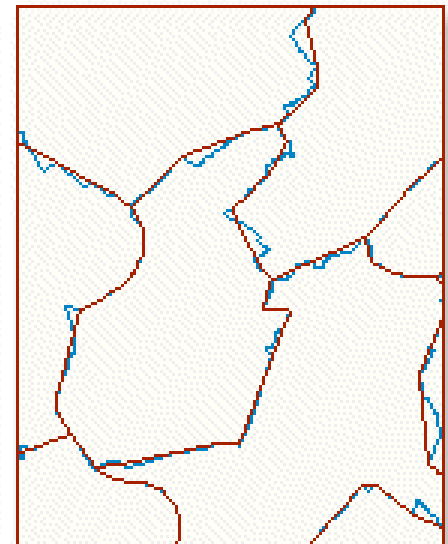
BEND SIMPLIFY

— ORIGINAL
— SIMPLIFIED

Point x
pásmo
(tvar!)



Point
Remove



Bend
Simplify

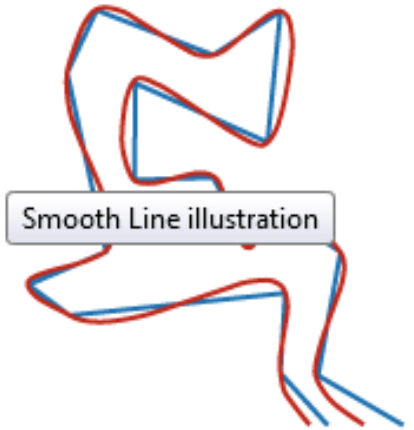


Smooth Line, Polygon

- **Shlazení (polynomální aproximace)**
- **Bézierovy křivky**

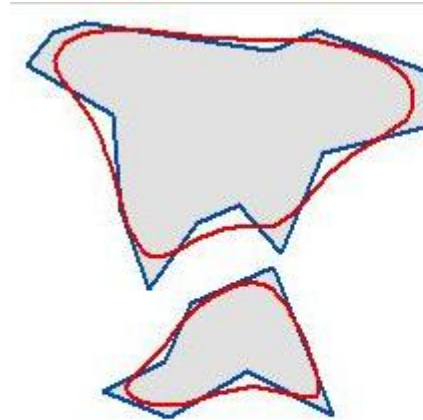
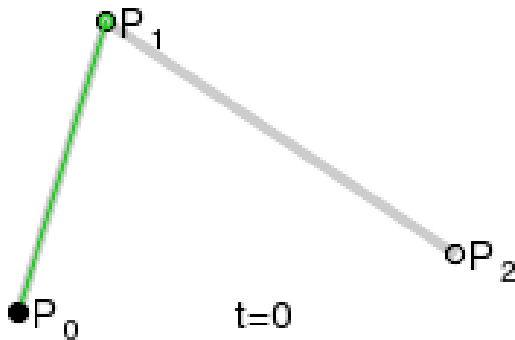


PAEK

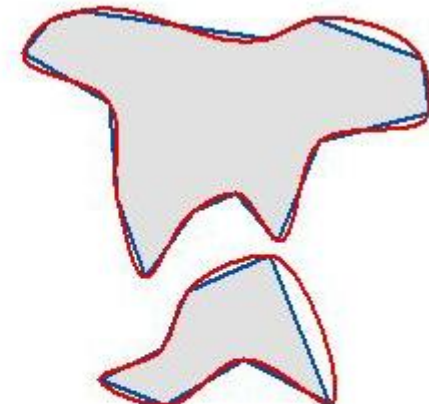


Bezier Interpolation

— ORIGINAL
— SMOOTHED



PAEK

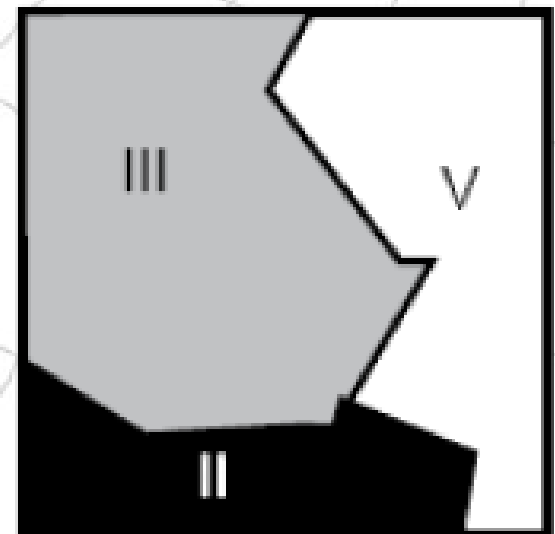
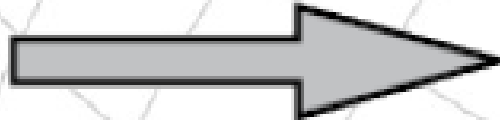
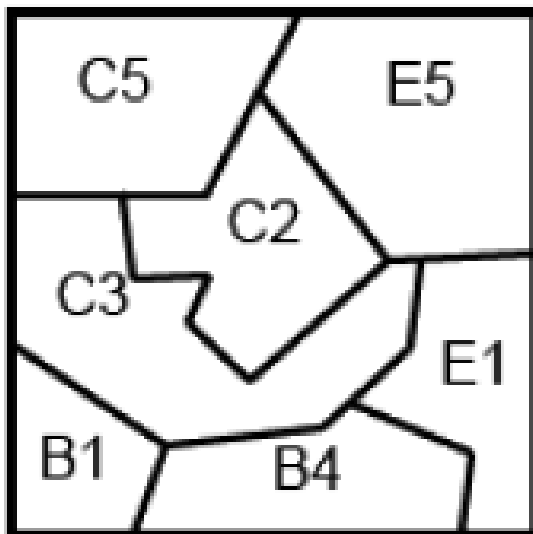


Bezier Interpolation

— ORIGINAL
— SMOOTHED

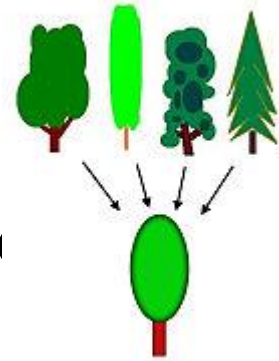
Vliv generalizace na kvalitu údajů

- Snižuje se **polohová** (prostorová) přesnost.
- Při snížení polohové přesnosti se může snížit i **atributová** přesnost!
- např. reklasifikace a spojení.



Reklasifikace a spojení

- Reklasifikace a spojení – spojení prvků se stejnými vlastnostmi do jednoho, například vrstvy listnatých a jehličnatých lesů spojit do vrstvy lesů – při změně měřítka.



Classification and Symbolization—Grouping features sharing similar geographic attributes into a new, higher-level feature class and representing it with a new symbol.

