

# Kontrola a zaručenie kvality HRVD údajov – teória a prax

Ing. Kamil FAKO, PhD.

Topografický ústav, Ružová 8  
974 01, Banská Bystrica, Slovensko  
[fako@topu.army.sk](mailto:fako@topu.army.sk)

**Abstrakt.** Článok pojednáva o metódach kontroly údajov vysokého detailu, ktoré sa v súčasnosti získavajú v Topografickom ústave. Vývoj a skúšanie rôznych metód viedlo k poznaniu najlepšej metódy, ktorou je metóda skusných plôch. Výber potrebného počtu objektov na základe štatistických metód vytvára stabilný základ pre vykonanie kvalitnej kontroly týchto údajov.

**Kľúčové slová:** HRVD dáta, štatistické metódy kontroly kvality, CPD,

**Abstract.** The article consider about methods of HRVD (high resolution vector data) checking, which are collecting by Topographic institute Banská Bystrica in present. Development and testing of various methods conduct in knowledge of best method – method of testing areas. Choice of needed objects quantity, based on statistical methods, forms stable foundation for execution good quality data checking.

**Keywords:** High resolution vector data, statistical methods, quality checking,

## 1 Čo sú HRVD údaje?

HRVD – High Resolution Vector Data – alebo vektorové údaje vysokého rozlíšenia sú údaje, ktoré svojim rozlíšením husto pokrývajú celý záujmový priestor. Nedá sa presne vyjadriť, čo vysoká hustota znamená, pretože každá oblasť má špecifiká, ktoré pojem „potrebná hustota“ determinujú.

Pre podmienky Slovenskej republiky možno povedať, že za údaje s vysokou hustotou môžeme považovať tie, ktoré sa svojou detailnosťou približujú obsahu máp 1: 25 000 a väčším. V praxi takáto detailnosť znamená aj veľkú presnosť zberu údajov ale na druhej strane aj veľkú objemnosť údajových sád.

Málokedy sa stretáme s údajmi vysokého rozlíšenia voľne dostupnými alebo komerčne ponúkanými. Tieto údaje sú vytvárané pre špecifickú skupinu podľa prísne definovaných pravidiel, technologických postupov a metód. Energia a náklady, ktoré sa vynakladajú na získavanie HRVD sú obrovské a je jednoduché ich zabezpečiť zo súkromných zdrojov pre väčšie územia.

Väčšina HRVD údajov patrí pod správu orgánov štátnej správy – ministerstvo obrany, ministerstvo vnútra, ktoré ich aj najviac využívajú a zabezpečujú ochranu údajov pred zneužitím neoprávnenými osobami.

## 2 Metódy zberu

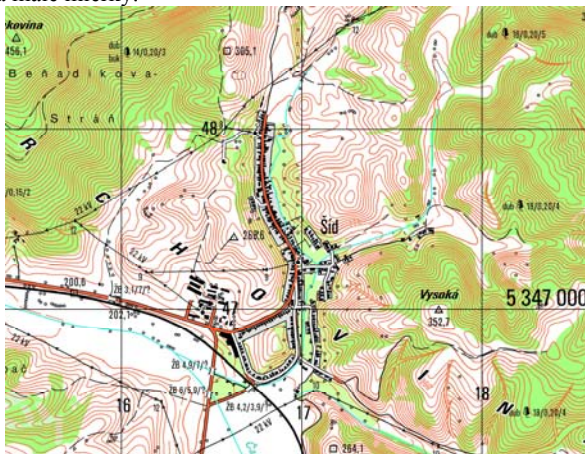
Metódy zberu sú odvodené od faktu, ako podrobné údaje potrebujeme. Zber pomocou metód DPZ je obmedzený na letecké meračské snímky (rozlišovacia schopnosť musí byť pod 40 cm na teréne). Aj keď už existujú satelity s veľkou rozlišovacou schopnosťou, údaje sa zväčša vyhodnocujú v 3D priestore a na to sú LMS vhodnejšie. Geodetické metódy zberu sú len ťažko použiteľné s ohľadom na množstvo potrebných údajov a náročnosť geodetického merania kdekoľvek v priestore.

Už pri návrhu štruktúry HRVD sa uvažuje aj o atribútovej stránke údajov. Bolo by veľmi nezodpovedné zbierať geometriu objektov a ich vlastnosti úplne zanedbať. Tento fakt znamená, že sa pri zbere údajov nezaobíjdeme bez overovania objektu jeho pozorovaním v teréne, teda miestnym preverovaním, šetrením.

## 3 Ako sa využívajú

Využitie údajov je veľmi rôznorodé. Dá sa povedať, že čím je vlastností objektov viac, tým širšie je aj využitie. Navyše, ak objekty predstavujú generálny pohľad na priestor bez zvädzania k špecifickému chápaniu objektov, je možné „pripájať“ stále ďalšie a ďalšie vlastnosti a tak poskytovať dobrý základ pre iné systémy. Hlavné ťažisko využitia je v tvorbe analýz. Či už priestorových, nákladových, sieťových a ďalších. Objektovo orientovaný model je preto nevyhnutnosťou.

Nezanedbateľnou oblasťou je tvorba máp rôznych mierok. Vždy platí zásada, že z veľa údajov sa málo dá spraviť, ale opačne to nefunguje. Vhodnými redaktorskými postupmi a za pomoci kvalitných softvérov sa dajú generovať mapy pre veľké, stredné a malé mierky.



Obr.1 Vizualizácia údajov CPD pomocou kľúča mapy UM25

Vizualizácia je často postranný efekt spracovania údajov, no aj surové údaje sa dajú zobrazovať a nechať na používateľa, nech si vytvorí vlastný obraz. Keďže takéto údaje za zbierajú zásadne v 3D priestore, nie je problém vytvárať virtuálne modely krajiny. Tieto sa v poslednom období začínajú často používať pri rôznych prezentáciách. Napomohol tomu aj fakt, že mnohé softvérové firmy už ponúkajú freewarové riešenia vizualizácie údajov v 3D.



Obr. 2 Vizualizácia údajov CPD - 3D pohľad

#### 4 Ako sa dajú skontrolovať?

Obrovské množstvo objektov, ktoré sa rôznymi metódami do modelu HRVD dostávajú, prináša so sebou aj riziko, že údaje nebudú správne. Je zrejme, že každý užívateľ chce mať údaje čo najpresnejšie, ale 100% bezchybnosť sa nedá nikdy zaručiť ani dosiahnuť. Dôležité je však poznať riziko chybovosti údajov, jeho istú časť akceptovať ako nevyhnutný fakt a zvyšok odhaliť a odstrániť.

Operátor, ktorý prichádza do kontaktu s údajmi ako editor môže a určite aj vnáša do procesu zberu subjektívny pohľad, isté množstvo omylov a občas aj úmyselných chýb. Aj keď technologický proces je kvalitne popísaný, musíme priznať, že sa vždy dajú nájsť „únikové uličky“, ktoré umožňujú operátorom obchádzať nastavené atribútové a topologické pravidlá.

Kontrola údajov musí mať vhodne zvolené kritériá, aby sa pri použití relatívne malých nákladov dalo s dostatočnou pravdepodobnosťou zaručiť buď nájdenie chýb,

alebo zaručiť, že údaje sú s prijateľným percentom chýb. Poznanie rizík chybovosti je prvým predpokladom na správne stanovenie metód kontroly.

HRVD sú geografické údaje, dá sa preto povedať, že ich priestorové rozvrstvenie je nerovnomerné. V oblastiach s vyššou hustotou obyvateľstva je objektov neporovnateľne viac, ako v oblastiach riedko osídlených. (Spôsobuje to fakt, že tvorcom objektov na povrchu Zeme je hlavne človek).

Ako teda vhodne zvoliť kritériá výberu objektov tak, aby vhodne pokryli celý záujmový priestor a zároveň zabezpečiť dostatočnú kontrolu rôznych objektov?

#### **4.1 1. metóda – rovnaká pravdepodobnosť**

Vychádza z predpokladu, že všetky objekty sú si rovné. a je v zásade jedno, či vyberieme na kontrolu budovu alebo les. Potom stačí objekty vo vhodnom počte rozvrstviť rovnomerne po kontrolovanej ploche, vykonať kontrolu a na základe jej výsledkov sa rozhodnúť, či údaje zodpovedajú požadovanej presnosti.

Výhodou je fakt, že k objektom pristupujeme bez ovplyvňovania výberu, čo je najvhodnejšie pre štatistickú kontrolu. Nevýhodou zase to, že pre žiadny systém nie je budova rovnako dôležitá ako les a kontrola bude musieť preverovať objekty po celej ploche. To prináša zvýšené náklady na presuny, zdĺhavý proces identifikácie a často kontrolu „prevládajúcich nepodstatných“ objektov.

#### **4.2 2. metóda – zamerané na sídla**

Vyberme do kontroly iba objekty, ktoré sú pre systém dôležité a vyskytujú sa v aglomeráciách. Budú to hlavne budovy, cesty, elektrické vedenia, mosty, ... (jednoducho objekty kritické). Na nich vykonajme podrobnú kontrolu a na základe jej výsledkov prijmime rozhodnutie o prijatí/neprijatí rizika chybovosti.

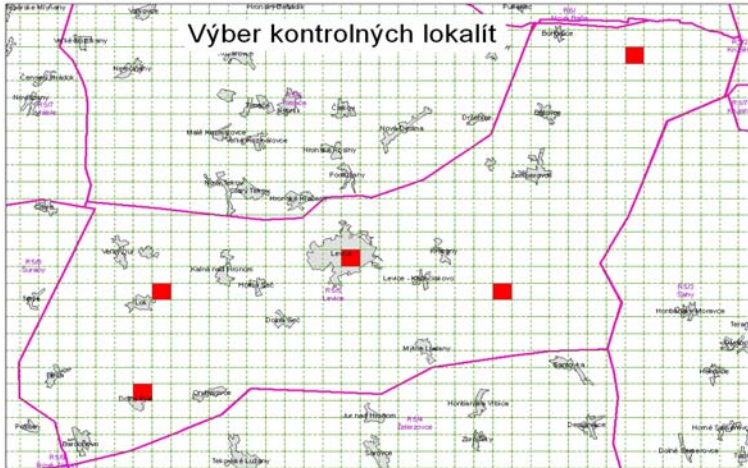
Výhody sú v znížení nákladov na kontrolu, keďže sa kontrolujú objekty husto pri sebe, v mestách a relatívne rýchlo. taktiež sa odhalilo viac chýb ako sa predpokladalo a to hlavne z toho dôvodu, že subjektívne kontrolovali objekty, kde je vyššia pravdepodobnosť, že sa chyby vyskytnú. Nevýhodou je vnášanie subjektívnosti do štatistického výberu, nerovnomerné rozvrstvenie objektov po kontrolovanej ploche.

#### **4.3 3. metóda – skusné plochy**

Dá sa povedať, že predstavuje istú kombináciu predchádzajúcich dvoch. Záujmová plocha sa pokryje mriežkou s rozmermi buniek zhodnými s rozmermi skusnej plochy. Zistí sa počet objektov v jednotlivých triedach v každej bunke mriežky a celkový počet objektov v bunke. Rozhodnutie o tom, či objekt patrí alebo nepatrí bunke je dané polohou jeho centroidu. Podľa početnosti objektov sa rozdelia bunky do 5 tried (každý po 20% početnosti). Náhodným výberom sa vyberie taký počet buniek, aby plocha bola 1% plochy regiónu a to v pomere určenom na základe stanovených pravidiel. (napr. 3-2-2-1-1) podľa počtu početnosti. Znamená to, že na

každé tri bunky z prvej triedy početnosti pripadnú dve bunky 2 a 3 triedy početnosti a 1 bunka z 4 a 5 triedy.

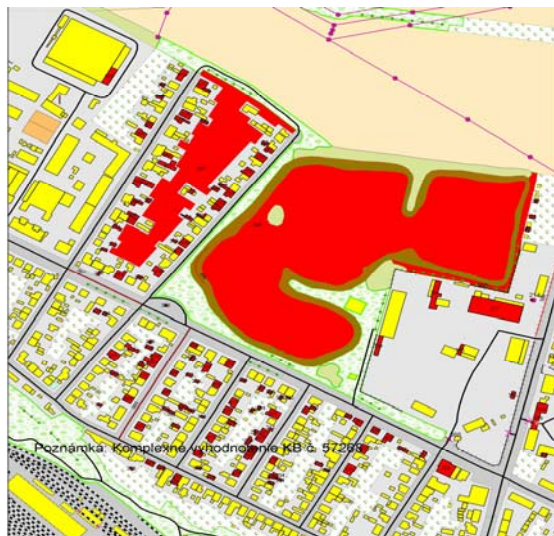
Preverované objekty v rámci skusnej plochy sa vyberú náhodným výberom cez všetky prvky triedy a to dovtedy, kým ich počet nie je minimálne 5 % celkového počtu, maximálne 134 objektov v triede. Tieto objekty predstavujú štatistický výber.



Obr. 3 Príklad výberu skusných plôch

Koľko objektov však do výberu vybrať? Vzorec, ktorý bol použitý vychádza z poznania prípustnej chybovosti a pravdepodobnosti, že sa zamietneme dodávka ak bude počet chybných menší ako nejaká hodnota a pravdepodobnosťou, ktorú sme ochotní tolerovať, že tento počet chýb bude vyšší. Na základe prvých skúseností vyšli kritické čísla pre objekty cca 160 objektov (alebo 5%) a z nich môže byť maximálne 6 chybných (alebo 3%). Tieto hodnoty platia pre 2 % prípustný podiel chybných údajov a 7 % neprípustný podiel. Riziko dodávateľa a odberateľa sme stanovili na 5%.

Čo sa za chybný objekt považuje a čo nie, je rozhodnutie porovnávatelia, ktorý získané údaje z kontroly porovnáva s údajmi v databáze z povolenými odchýlkami stanovenými pre jednotlivé atribúty jednotlivých objektov. Je totiž striktné oddelené poznávanie objektov kontrolórom na mieste (v teréne) a porovnávatelom alebo vyhodnocovateľom v kancelárii. Kontrolór sa nedostáva k údajom v databáze, aby nebol ovplyvnený subjektívnym rozhodnutím operátora alebo miestneho šetriča. Na druhej strane je jeho povinnosťou opísať objekt tak, aby vyhodnocovateľ prijal správne rozhodnutie.



Obr. 4 Detail vybraných objektov na SP

## 5 TOPÚ a jeho metóda na kontrolu

Topografický ústav pri zabezpečovaní kontroly kvality údajov CPD (ktoré predstavujú údaje typu HRVD) používal rôzne metódy. V začiatku sa pristupovalo k objektom podľa výberu vedúceho STK, ktorý volil trasu preverovania tak, aby pokryl územie čo najlepšie ale sústreďoval sa na aglomerácie (zhruba ako variant č.2). Takto bol výsledok veľmi zameraný na miesta s vysokou hustotou, teda dobre odhaľoval chyby dôležitých objektov a odhalil aj neúplnosti údajov, avšak nebol dostatočne reprezentatívny ani čo sa týka druhov objektov, ani čo sa týka skúmaného priestoru. Napriek tomu, že počet chýb nikdy nepresiahol kritickú hranicu a údaje vždy boli prijaté ako dobré, vyskytovali sa v nich chyby, ktoré pôsobili rušivo. Tie, ktoré sa podarilo nájsť však boli opravené. Aspoň malou časťou tak kontrola pomohla zlepšiť kvalitu údajov.

Od roku 2007 sa skúša metóda č.3, teda preverovanie náhodne vybraných objektoch na skusných plochách. Preverovanie je však zložitejšie, pretože sa preverujú všetky vlastnosti vybraného objektu a objekty sa často nachádzajú na neľahko prístupnom miesta (vzadu za domami, v družstvách). Napriek nárastu časového faktora sa zlepšil faktor náhodnosti a tým pádom aj štatistickej reprezentatívnosti. Ku každému objektu vybranému na preverovanie sa vedie samostatný list, kde sa popíšu všetky relevantné údaje zistené na kontrole aj s uvedením bližšieho popisu objektu v prípade nejednoznačnosti, dátumu vykonania kontroly a podpisu pracovníka, ktorý objekt fyzicky kontroloval.

Pri takomto spôsobe kontroly sa však potláča kontrola úplnosti a je potrebné ju nahradiť iným mechanizmom. V začiatku sa vykonávala kontrola úplnosti na jednej vybranej skusnej ploche. Ukázalo sa však, že toto riešenie nie je celkom vhodné a v súčasnosti neprebieha kontrola úplnosti na jednej bunke, ale vykonávatelia kontroly pohľadovo kontrolujú úplnosť údajov v priebehu celého procesu kontroly na celej ploche, po ktorej sa pohybujú. Tento fakt, síce zmenil prístup ku kontrole úplnosti, avšak nezvýšil náklady na celkový výkon kontroly. Po zaučení kontrolórov a zvýšení praktických skúseností bude aj kontrola úplnosti dostatočne podrobná.

## **6 Záver**

Kontrola údajov HRVD si vo všeobecnosti vyžaduje špecifický prístup a návrh konkrétnej metódy bude vždy podmienené spôsobom zberu a typom využitia údajov. Každý užívateľ si však musí uvedomiť a zmieriť sa s faktom, že tak obrovské množstvo údajov so sebou nesie aj isté percento chýb. Poznanie tohto percenta potom umožní lepšie rozhodnutie užívateľa o tom, na čo sa získané údaje dajú a na čo už nedajú použiť.