

**HODNOTENIE NÁCHYLNOSTI ÚZEMIA NA ZOSUV S VYUŽITÍM  
ŠTATISTICKÉHO PREDIKČNÉHO MODELOVANIA NÁHODNÝCH LESOV  
A LOGISTICKEJ REGRESIE VO VYBRANOM FLYŠOVOM ÚZEMÍ**

**ASSESSMENT OF LANDSLIDE SUSCEPTIBILITY USING STATISTICAL  
PREDICTIVE MODELLING OF RANDOM FORESTS AND LOGISTIC  
REGRESSION IN THE SELECTED FLYSCH ZONE**

Mária BARANČOKOVÁ<sup>1</sup>, Peter BARANČOK jr.<sup>2</sup>, Peter BARANČOK<sup>1</sup>,  
Matej ŠOŠOVIČKA<sup>2</sup>

<sup>1</sup>Ústav krajinej ekológie SAV, Štefánikova 3, P. O. Box 254, 814 99 Bratislava  
e-mail: maria.barancokova@savba.sk, peter.barancok@savba.sk

<sup>2</sup>Fakulta matematiky, fyziky a informatiky, UK, Mlynská dolina, 842 48 Bratislava  
e-mail: peter.barancok@gmail.com, matejsosovicka@gmail.com

**Abstract:** Landslides are the most common geodynamic phenomenon in Slovakia and the most affected area is the north-western part. In the paper we evaluate the susceptibility of the region to landslides using statistical predictive modelling. In evaluating, we used the method of logistic regression and the method of random forests, while various types of variables affecting the models were taken into consideration – continuous, categorical ordinal and categorical nominal variables. We selected 15 variables that were used as predictors of a dependent variable (landslide susceptibility). Categories of categorical nominal variables were reclassified from the original detailed division into more general categories based on similarities of their characteristics. To measure the association between the input variables, Spearman's rank correlation coefficient and Cramer's V were used. Logistic regression model can be said to be conservative in the majority of the area, as it very rarely predicts the probability of landslide lower than 0.25 or higher than 0.75. The accuracy of the classification based on this threshold value was 79 %. The random forest model had  $AUC_{val} = 0.967$  and the accuracy of the classification was 94 %. Random forests achieved better results on all performance metrics considered. The predicted probabilities from the logistic regression model and the random forests model were divided into five categories of landslide susceptibility. In the logistic regression model, the slopes with very low susceptibility cover 17.2 % of the area, in the random forests model it is 34.8 % of the area.

**Key words:** predictive modelling, correlation analysis, logistic regression, random forest, landslide susceptibility, flysch zone

## Úvod

Zosuvy sú deštruktívne a každoročne sa opakujúce javy, ktoré spôsobujú značné škody na majetku. V súčasnosti predstavujú zosuvy významnú environmentálnu hrozbu a u mnohých odborníkov vzbudzujú pozornosť, najmä pre analýzu faktorov ich vzniku. Metódy hodnotenia zosuvov sú založené na porovnaní a následnom štatistickom spracovaní a zhodnotení vzťahov medzi relevantnými faktormi vplyvujúcimi na stabilitu svahov s reálnym výskytom svahových deformácií. Ako dôležité faktory, ktoré odrážajú vhodné podmienky pre vznik zosuvov je na základe doterajších skúsenosti s danou problematikou, geologická stavba (litológia, seizmicko-tektonické a štruktúrne pomery), geomorfologické pomery, charakter georeliéfu, klimatické faktory, hydrogeologické pomery ako aj aktuálne využitie krajiny. Priestorové rozloženie litologických typov hornín v značnej miere určuje distribúciu svahových deformácií v území, z čoho vyplýva výrazná kladná korelácia medzi počtom svahových deformácií a horninovými typmi vysoko náchylnými na zosúvanie. Najohrozenejším územím z hľadiska výskytu svahových deformácií na Slovensku patrí flyšové pásmo, ktoré tvorí rozsiahle územie na severozápade a severe Slovenska.

Ako prví sa myšlienkou štatisticky modelovať zosuvnosť svahov v 70-tych rokoch zaoberal Neuland (1976), keď sa prostredníctvom dát o geologickom podloží svahov snažil popísať stabilitu svahov juhozápadného Nemecka, pomocou diskriminačnej analýzy dvoch premenných. V 80-tych rokoch Carrara (1983) využil viacrozmerné štatistické metódy na predikciu zosuvnosti svahov a výsledky získané z tejto analýzy boli po prvýkrát priestorovo vizualizované v počítačových verziách geografického informačného systému (GIS). Zhrnutie metodiky aplikácie GIS v analýze zosuvného hazardu je v prácach Carrara et al. (1995) a van Westen et al. (1993). Hodnotenie územia na základe inventarizácie zosuvnosti je dôležitým krokom na určenie citlivosti, nebezpečnosti a posúdenia rizika zosuvnosti svahov (napr. Guzzetti et al., 2005; Fell et al., 2008; Dou et al., 2015; Wang et al., 2020; Lee et al. 2016; Chen et al., 2018).

Vyhodnotenie priestorových vzťahov medzi zosuvmi a environmentálnymi faktormi, ktoré ich vyvolávajú, môže byť založené na digitálnych databázach a zahŕňa metódy ako je štatistika, analýza priestorových vzťahov a interaktívne mapovanie (Meten et al., 2015; Prefac et al., 2016; Sakkas et al., 2016). Štatistický histogram, priestorové prekrytie a metódy dynamického mapovania sú navzájom prepojené na interaktívne vyhodnotenie priestorového vzťahu medzi zosuvmi a environmentálnymi faktormi. V súvislosti s environmentálnymi faktormi používanými pri posudzovaní nebezpečenstva zosuvu pôdy existuje tendencia využívať tie vrstvy údajov, ktoré sa dajú ľahko získať z digitálneho modelu reliéfu a satelitných snímok, pričom menší dôraz sa kládol na vrstvy údajov, ktoré vyžadujú podrobnú analýzu v teréne (environmentálne faktory so zameraním na digitálny model terénu, geológia a pôda, geomorfológia, využitie územia a rizikové prvky), (van Westen et al., 2008; Ciampalini et al., 2016; Tseng et al., 2015). Hydrologické podmienky, ako sú silné zrážky, infiltrácia a exfiltrácia podzemnej vody, majú veľký vplyv na výskyt zosuvov (Krzeminska et al., 2013; Wiedenmann et al., 2016; Lee et al., 2016). Na nestabilných svahoch z jemnozrnných sedimentoch sa vytvára gravitačná deformácia (povrchové šmykové roztrhnutie a napínacie trhliny), ktorých

kontinuálne otvorenie môže byť zosilnené zmrašťovaním počas suchých období (Bièvre et al. 2012).

Litológia hornín a sklon svahu sú dôležitými faktormi, ktoré prispievajú k zvýšenej aktivite zosuvu. Väčšina autorov pri hodnotení a mapovaní citlivosti svahov na zosuvy využíva uhol sklonu svahu a litológiu ako nezávislé premenné (Ayalew, Yamagishi, 2005; Safaei et al., 2012). Wu, Qiao (2009) rozdeľujú faktory, ktoré ovplyvňujú výskyt zosuvu do dvoch typov, t.j. na vonkajšie a vnútorné faktory. Vonkajšie faktory zahŕňajú zrážky, zemetrasenia a činitele spôsobené človekom. Dôležité je však venovať pozornosť vnútorným faktorom, medzi ktoré patrí litológia hornín, uhol sklonu svahu, poloha svahu, profil svahu a výškový gradient.

Hodnotením zosuvnosti v Čechách a na Slovenku sa zaoberali mnohí autori, napr. Paudiš et al. (2005), Metelka, Kycl (2007), Bednarik et al. (2010), Havlín et al. (2011), Barančoková, Kenderessy (2014). Využitím predikčných modelov pri hodnotení náchylnosti na zosúvanie na Slovensku okrem Muňka (Muňko, 2015), ktorý aplikoval metódu Support Vector Machines a porovnával ju s bivariačnou analýzou sa zaoberali aj Pham et al. (2021) a Ali et al. (2021). Autori analyzovali 16 faktorov podmieňujúcich zosuvy a pri hodnotení citlivosti územia na zosuvy využili fuzzy modelovanie, bivariačný frekvenčný model, logistickú regresiu, náhodné lesy a iné metódy.

Jedným zo spôsobov prevencie zosuvných rizík môže byť návrh vhodných štatistických modelov na vytvorenie mapy potenciálnej zosuvnosti svahov. Cieľom predloženej práce je vytvorenie prognózy s využitím štatistického predikčného modelovania, na základe ktorej by bolo možné s vysokou pravdepodobnosťou predpovedať vznik zosuvov do budúcnosti. Za modelové územie bolo zvolené územie Kysúc, ktoré spadá do flyšového pásma Západných Karpát a je budované sedimentárnymi horninami tvorenými v prevažnej miere pieskovecami a ílovcami. Vlastnosti týchto hornín majú v území výrazný vplyv na reliéf a vodný režim. Rýchlo podliehajú erózii a svojou nízkou priepustnosťou môžu počas príválových dažďov spôsobovať časté zosuvy.

### **Charakteristika územia**

Sledované územie bolo vytvorené ako prirodzený a ucelený celok povodia rieky Kysuce (mapa 1). Zahŕňa okres Čadca (23 obcí), okres Kysucké Nové Mesto (14 obcí) a časť okresu Žilina (obec Lutiše). Na základe členenia Slovenska do geomorfologických jednotiek (Mazúr a Lukniš 1986) patrí sledované územie do troch oblastí: Stredné Beskydy (celky: Kysucké Beskydy, Oravské Beskydy, Kysucká vrchovina, Podbeskydská brázda, Podbeskydská vrchovina. Oravská Magura) zaberajú 28,21km<sup>2</sup>, Západné Beskydy (celky: Moravsko-sliezske Beskydy, Turzovská vrchovina, Jablunkovské medzihorie) 29,46 km<sup>2</sup> a Slovensko-moravské Karpaty (celok: Javorníky) zaberajú 22,98 km<sup>2</sup>. Výškové rozmedzie je od 325 m n. m. v Kysuckej bráne pri Radoli až po vrchol Veľkej Rače 1236 m n. m.

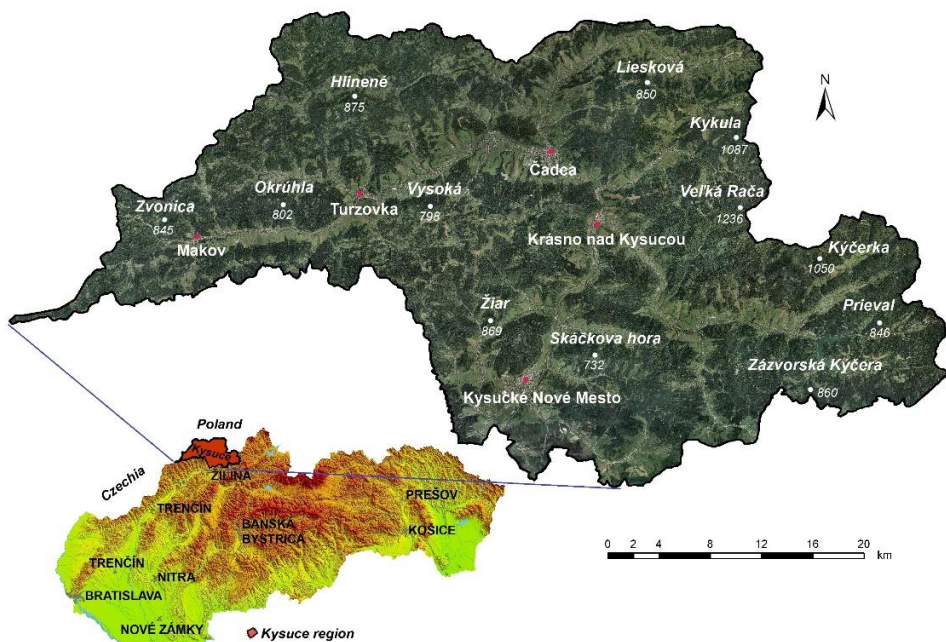
Na geologickej stavbe územia sa podieľajú tri základne tektonické geologické celky, v smere od severozápadu na juhovýchod sú to sliezsky príkrov, magurský príkrov

a bradlové pásmo. Najrozšírenejšou geologickou jednotkou v území je zlínske súvrstvie (355,65 km<sup>2</sup>), deluviálne sedimenty (231,62 km<sup>2</sup>), bystrická jednotka (229,26 km<sup>2</sup>) a solánske súvrstvie (182,63 km<sup>2</sup>). Významnou úlohu vo zvýšenej zosuvnosti na zlínskom súvrství zohráva rozdielna geomorfologická hodnota súvrství ílovcového a pieskovecového vývoja. Vyšší podiel ílovcov ako pieskovecov v zlínskom súvrství spôsobuje menšiu odolnosť voči exogénnym procesom.

Kysuce patria zrážkovo do vlhkej klimatickej oblasti. Priemerný ročný úhm zrážok v severných oblastiach doliny Kysuce je 700–1110 mm. Najvyšší priemerný mesačný úhm zrážok je v mesiacoch jún a júl. Z teplotného hľadiska patria Kysuce do studenej a mierne teplej klimatickej oblasti s priemernou ročnou teplotou za roky 2001 – 2020 v Čadci 7,9 °C.

V tomto území je evidovaných 1869 zosuvov (Šimeková, Martinčeková et al., 2006). Zosuvy sú registrované v databázach Štátneho geologického ústavu Dionýza Štúra a ich poloha bola verifikovaná v teréne. Podľa aktivity boli zosuvy rozdelené do troch kategórií: aktívne, potenciálne a stabilizované. V sledovanom území zaberajú zosuvy takmer 20 % (174,21 km<sup>2</sup>). Najviac je potenciálnych zosuvov, 1284 (122,36 km<sup>2</sup>), potom stabilizovaných, 544 (50,20 km<sup>2</sup>) a 41 aktívnych (1,64 km<sup>2</sup>). Potenciálne zosuvy zaberajú 70 % plochy zo všetkých zosuvov.

Mapa 1: Sledované územie



## Použité metódy

Neexistuje žiadny globálny návod na výber faktorov podmieňujúcich zosuvnosť, ale mali by sa vyberať na základe charakteristík sledovaného územia, dostupnosti údajov a prehľadu literatúry. Jedným z najdôležitejších krokov v modelovaní priestorovej predikcie je výber vhodných a účinných podmieňujúcich faktorov spomedzi všetkých dostupných faktorov (Chen et al., 2018). Pri výbere vstupných premenných sme sa zamerali na tie premenné, ktoré môžu byť spúšťačom zosuvov. Ide o geologické, hydrogeologické a pôdne vlastnosti, morfometrické parametre reliéfu, hydrofyzikálne vlastnosti pôdy a hydraulické vlastnosti hornín. Zvolili sme si dve metódy predikčného modelovania: logistická regresia a náhodné lesy (random forest – metóda pre klasifikáciu a regresiu). Pri štatistickom modelovaní a následnom hodnotení sme postupovali nasledovne:

1. Výber vstupných premenných a ich reklasifikácia,
2. Korelačná analýza premenných a ich výber do štatistických modelov,
3. Štatistické modelovanie metódami logistickej regresie a náhodných lesov,
4. Interpretácia jednotlivých metód,
5. Vyhodnotenie náchylnosti územia na zosúvanie.

Pri výbere vstupných premenných (krajinných parametrov) sme vychádzali z údajovej databázy prvkov krajinnokoekologického komplexu (ESPRIT, 2013), ktoré predstavujú vektorovú reprezentáciu syntetických jednotiek vyjadrujúce relevantné vlastnosti abiotickéj zložky krajinnnej sféry spolu s prvkami krajinnnej pokrývky.

Do modelovania vstupovalo 15 premenných so samostatnými vektorovými vrstvami. Kombináciou jednotlivých vrstiev vznikli tzv. kvázi-homogénne terénne jednotky (UCU – unique condition unit). Nadmorská výška vstupovala do hodnotenia v rastrovej vrstve s bunkami mriežky 15 m x 15 m. Na základe kombinácie nadmorskej výšky s ostatnými premennými bola celá sledovaná oblasť rozdelená na pixely (bunky mriežky) s veľkosťou 15 m x 15 m a dátové vlastnosti každého pixelu boli priradené podľa väčšinového zastúpenia jednotlivých tried premenných v danom pixeli. Dôvod prechodu na pixely bol ten, že vybrané metódy modelovania pracujú lepšie s pixelmi ako s jednotlivými vrstvami. Ich výhody a nevýhody v oboch prípadoch sú porovnateľné (Reichenbach et al., 2018).

Vstupné kategoriálne premenné boli reklasifikované do všeobecnejších kategórií, z dôvodu jednoduchšie aplikovateľného modelovania a rozdelené na nominálne premenné a ordinálne premenné. Jedinou spojitou premennou bola nadmorská výška. Zosuvnosť ako závislá premenná vystupovala ako binárna jednotka (0 – nezosuvný svah a 1 – zosuvný svah).

Vstupné premenné:

1. Kategoriálne nominálne premenné:
  - Morfológia reliéfu (SRM)
  - Hydrogeologické jednotky (HGU)
  - Pôdne jednotky (ST)
  - Krajinná pokrývka (LC)

- Geologické jednotky (GU)
  - Zrornosť (GSS)
  - Horizontálna krivosť reliéfu (PC)
  - Orientácia (Or)
2. Kategoriaálne ordinálne premenné:
    - Sklon svahu (Sln)
    - Koeficient prietočnosti (TC)
    - Koeficient filtrácie (FC)
    - Hladina podzemnej vody (GWL)
    - Hĺbka pôdy (SDe)
    - Skeletnosť pôdy (SSk)
  3. Spojitá premenná:
    - Nadmorská výška (Alt)
  4. Závislá premenná:
    - Zosuvnosť (LS)

Pri modelovaní náchylnosti na zosuvy pomocou logistickej regresie sa kategoriaálne nominálne premenné transformovali na  $c - 1$  binárne indikátorové premenné (kde  $c$  je počet tried pre danú premennú). Kategoriaálne ordinálne premenné boli transformované na spojité premenné tak, že hodnota tried bola braná ako stred intervalu, ktorý reprezentovali, ako bolo navrhnuté v Agresti (2018). Toto stále umožňovalo modelovanie nelineárneho vzťahu medzi náchylosťou na zosuv a vstupnými premennými sčítaním mocnín premenných (napr. sklonu). Súbor údajov obsahoval 4 223 653 pixelov, z ktorých 550 896 bolo označených ako zosuvné oblasti.

#### *Korelačná analýza vstupných premenných*

Spoločnou črtou reálnych údajov je, že vstupné premenné (nazývané aj prediktory) nie sú od seba nezávislé. To môže viesť k nestabilným odhadom parametrov, ktoré bránia identifikácii premenných relevantných pre modelovanie náchylnosti na zosuv pôdy a interpretáciu ich účinkov (Dorman et al., 2013). Ďalším krokom bolo vykonanie korelačnej analýzy pre vstupné premenné, aby sa určil vzťah medzi nimi. Spôsob merania asociácie sa líši v závislosti od typu premenných (spojité, kategoriaálne ordinálne alebo kategoriaálne nominálne).

Na testovanie závislosti medzi dvoma kategoriaálnymi premennými bola použitá štatistika *Cramer's V*, ktorá pre dve premenné  $X$  a  $Y$  s kategóriami  $A_i$  a  $B_j$ , kde  $i = 1, \dots, r$  a  $j = 1, \dots, k$ , je definovaná ako:

$$V_{X,Y} = \sqrt{\frac{X^2/n}{\min(r-1, k-1)}} \quad X^2 = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{\frac{n_i n_j}{n}}$$

kde  $n$  je počet pixelov,  $n_i$  a  $n_j$  sú počty pixelov v kategóriách  $A_i$  a  $B_j$ . Cramer's  $V$  dosahuje hodnoty z intervalu  $[0, 1]$ , pričom 0 znamená žiadnu asociáciu medzi premennými a 1 komplexnú asociáciu.

Na meranie asociácie medzi spojitými alebo kategoriálnymi ordinálnymi premennými bol použitý Spearmanov korelačný koeficient, ktorý je definovaný ako:

$$r_{xy} = \frac{\sum_{i=1}^n (x'_i - \bar{x}') (y'_i - \bar{y}')}{\sqrt{\sum_{i=1}^n (x'_i - \bar{x}')^2} \sqrt{\sum_{i=1}^n (y'_i - \bar{y}')^2}}$$

kde  $n$  je počet pixelov,  $x_i$  a  $y_i$  sú poradia  $i$ -tého pixelu vzhľadom na premenné  $X$  a  $Y$ ,  $\bar{x}$  a  $\bar{y}$  sú priemery premenných  $X$  a  $Y$ . Hodnoty Spearmanovho koeficientu poradovej korelácie ležia v intervale  $[-1, 1]$ , pričom hodnoty  $-1$  a  $1$  označujú dokonalú lineárnu koreláciu a  $0$  znamená žiadnu lineárnu koreláciu.

#### *Modelovanie náchylnosti územia na zosuv*

Hlavným cieľom štatistického modelovania je predpovedanie neznámej premennej z dostupných údajov alebo popísanie čo najzrozumiteľnejším spôsobom vzťah medzi vysvetlenými a vysvetľujúcimi premennými (James et al., 2013). Výber techniky modelovania závisí od problému, ktorému čelíme. Lineárne modely sa ľahko interpretujú, ale ich prediktívna sila nie je taká podstatná ako nelineárne modely. Na druhej strane, nelineárne modely majú vyššiu predikčnú schopnosť, je to však za cenu nižšej interpretovateľnosti. Pri našom hodnotení sme použili dve najpoužívanejšie metódy štatistického modelovania – logistickú regresiu a náhodné lesy.

#### *Logistická regresia*

Logistická regresia je jednou z najpopulárnejších a najpoužívanejších viacrozmerných štatistických metód na modelovanie náchylnosti na zosuvnosť (Reichenbach et al., 2018; Budimir et al., 2015). Ide o lineárny model a pokúša sa modelovať pravdepodobnosť  $p$  udalosti (v tomto prípade zosuvnosť), ktorá nastane na základe hodnôt vstupných premenných. Vzorec pre pravdepodobnosť je definovaný ako:

$$p(X) = \frac{1}{1 + \exp(-z)} \quad z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

kde  $n$  je počet vstupných premenných,  $X = (X_1, X_2, \dots, X_n)$  je vektor vstupných premenných a  $\beta_i$  ( $i = 0, 1, \dots, n$ ) sú koeficienty logistickej regresie, ktoré je potrebné odhadnúť z údajov.

Podrobnejšie o metóde je v prácach Agrestiho (Agresti, 2007; Agresti, 2010).

### Náhodné lesy (random forest)

Metódu náhodné lesy je možné zaradiť do skupiny tzv. skupinových metód (ensemble methods). Princíp skupinových modelov je jednoduchý a vychádza z toho, že skupine modelov (napr. rozhodovacích stromov) je zadán rovnaký problém, na ktorom sa učia. Následne sa výstupy naučených modelov kombinujú do jedného skupinového modelu. Dôležité je zistiť, či kombináciou výsledkov jednotlivých modelov je možné dosiahnuť zlepšenie výsledkov klasifikácie alebo predikcie. Rozhodovací strom sa snaží na základe vektora vstupných dát (pozorovaní, premenných) predpovedať hodnotu výstupnej premennej (v našom prípade zosuvnosti). Skladá sa z dvoch typov uzlov: neterminálnych a terminálnych. V neterminálnych uzloch sa nachádza kritérium, ktoré určí, do ktorého uzla sa vstupný vektor ďalej posunie. V terminálnych (koncových) uzloch sa určujú triedy, na aké sa vstupný vektor klasifikuje.

Na trénovanie náhodných lesov sa používajú viaceré techniky. Najjednoduchšia prebieha pomocou *bootstrappingu*. Pre každý rozhodovací strom sa náhodne (s opakovaním) vyberie podmnožina dát, ktorá bude použitá na trénovanie daného stromu. V prípade kategoriálnej ordinálnej závislej premennej je možné použiť relatívne nový prístup (Hornung, 2019) vylepšenia regresných náhodných lesov pre kategoriálnu ordinálnu závislú premennú. Závislá premenná bola chápaná ako spojitá premenná, rovnako ako pri regresných náhodných lesoch, avšak hlavným vylepšením tohto algoritmu oproti regresným náhodným lesoch je optimálne, a nie rovnomerné, rozdelenie zoradených kategórií závislej premennej do intervalu [0, 1].

Kvalitu (validáciu) modelov je možné merať viacerými spôsobmi. Môže to byť napr. pomocou štatistických metrik alebo testovanie robustnosti modelu, t.j. porovnávať, či sú predikcie modelu na neznámych dátach podobne úspešné ako na známych dátach. Vhodné je rozdelenie dát na viacero skupín, pričom na jednej skupine dát je model trénovaný a na ďalšej testovaný. V našej práci sme využili jednoduchší spôsob. Do trénovacej vzorky bola vybraná časť dát (pixelov), ktoré boli označené ako zosuvné a náhodne boli doplnené rovnakým počtom nezosuvných dát.

Vizualizácia výkonu binárneho klasifikačného modelu bola vytvorená na základe *confusion matrix*. Bunky *confusion matrix* predstavujú počet skutočne pozitívnych (TP), falošne pozitívnych (FP), skutočne negatívnych (TN) a falošne negatívnych (FN) pixelov. Existuje mnoho opatrení, ktoré možno vypočítať z týchto hodnôt v závislosti od cieľa štúdie. Najčastejšie používaným meradlom je presnosť, čo je pomer medzi počtom správne klasifikovaných pixelov a celkovým počtom pixelov:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$

Jeho spoľahlivosť je však značne znížená pre vysoko nevyvážené súbory údajov, ako je to v prípade modelovania náchylnosti na zosuvnosť. Problém nevyvážených tried rieši *Matthewsov korelačný koeficient (MCC)* (Hastie et al., 2009), ktorý je definovaný nasledovne:



$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Hodnoty MCC ležia v intervale [-1, 1], pričom -1 a 1 označujú dokonalú nesprávnu klasifikáciu a dokonalú klasifikáciu, 0 znamená, že model je len taký dobrý ako náhodné priradenie tried na základe ich pomeru. Okrem presnosti a MCC je potrebné uviesť aj skutočnú pozitívnu mieru (TPR) a skutočnú negatívnu mieru (TNR), nazývanú tiež citlivosť a špecifickosť, definovanú ako:

$$TNR = \frac{TN}{TN + FP} \quad TPR = \frac{TP}{TP + FN}$$

Matthewsov korelačný koeficient sa podľa Boughorbela (Boughorbel et al., 2017) považuje za vhodnú štatistiku na popisovanie kvality klasifikácie modelu s rozdielnymi počtami pozitívnych a negatívnych prípadov (v našom prípade zosuvných pixelov). Tento koeficient, ako aj samotná Confusion matrix, sú však závislé od zlomovej hodnoty z.

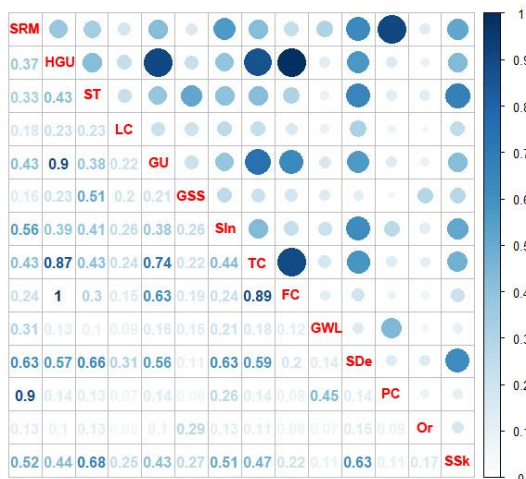
Všeobecnejšia metrika popisujúca počty prvkov na diagonále Confusion matrix bez nutnosti určenia z sa nazýva AUC (Area Under Curve) skóre. AUC je počítané ako obsah pod ROC krivkou, ktorá je definovaná ako krivka závislosti medzi TPR a TNR.

## Výsledky

### Korelačná analýza vstupných premenných

Do analýzy nevstupovali premenné (9 979 pixelov, 0,24 % plochy), ktoré nemali kompletne informácie. Tieto premenné predstavovali najmä plochu vodnej nádrže Nová Bystrica. Z ostatných premenných (okrem nadmorskej výšky), ktoré boli pôvodom kategoriálne (nominálne alebo ordinálne), bolo možné zostrojiť maticu korelácií Cramer's V (obr. 1).

Obr. 1: Matica korelácií Cramer's V



Ako vysoko korelované (*Cramer's V* > 0,6) vystupujú tri skupiny premenných, z ktorých do modelu logistickej regresie boli použité iba po jednej z nich:

- geologické jednotky (GU), koeficient prietochnosti (TC), hydrogeologické jednotky (HU), koeficient filtrácie (FC),
- sklon svahu (SIn), hĺbka pôdy (SDe), pôdny typ (ST), skeletnosť (SSk),
- morfológia reliéfu (SRM), horizontálna krivosť reliéfu (PC).

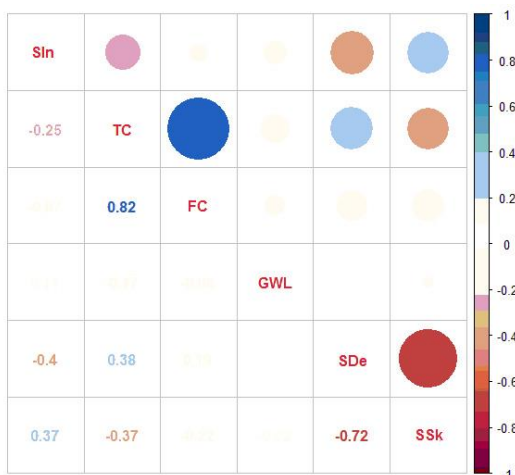
V prvej skupine vystupujú premenné s prevahou hydrogeologických vlastností. V tejto skupine majú najvyššie asociačné väzby FC – HGU (1,00), HGU – GU (0,9), FC – TC (0,89), TC – HGU (0,87), TC – GU (0,74) a FC – GU (0,63). Horninový komplex v pieskovcovom vývoji je hlavným kolektorom podzemných vôd, vyznačuje sa puklinovo-medzizrnovou priepustnosťou a odvodňovanie tohto komplexu je značne variabilné. Tieto komplexy sú tvorené horninami s nízkou priepustnosťou. Na rozdiel od tohto komplexu sa horniny v pieskovcovo-hlinitom vývoji vyznačujú rytmickým striedaním pieskovcov a ílu (resp. prevahou pieskovcov v niektorých častiach súvrstvia), t.j. horniny s vodonosnými vlastnosťami s izolátormi, ktoré obmedzujú cirkuláciu podzemnej vody v komplexe. Ílovitý substrát je slabo priepustný a najviac zrážkovej vody steká po povrchu s následným intenzívnejším rozvojom erózie (najmä na odlesnených územiach) s následnou tvorbou zosuvov. Striedanie geologických vrstiev s rôznymi hydrogeologickými parametrami vo flyšových oblastiach predstavuje spúšťač faktor zosuvov pôdy pri výdatných zrážkach.

V druhej skupine sú pôdne charakteristiky a sklon. Asociačné väzby medzi premennými sú SSk – ST (0,68), SDe - ST (0,66), SDe - SIn (0,63), SSk - SDe (0,63), SSk - SIn (0,51) a ST - SIn (0,41). Zosuvné procesy často vznikajú v dôsledku nasýtenia pôdnych pórov vodou pri hlbších pôdach a pôsobením gravitačných síl na nestabilných svahoch (napr. pri dlhodobých dažďoch alebo v dôsledku topenia snehu).

V tretej skupine sú reliéfné charakteristiky. Tieto dve premenné vykazujú veľmi vysokú korelačnú väzbu SRM - PC (0,90). Ide o prepojenie transportného svahu a lineárnej krivosti reliéfu.

Premenné, ako sklon svahu, koeficient transmisivity, koeficient filtrácie, hladina podzemnej vody, hĺbka pôdy a skeletnosť pôdy možno považovať za kategoriálne ordinálne premenné a spolu so spojitou premennou nadmorskou výškou bola vytvorená matica *Spearmanových korelácií* (obr. 2). Získané hodnoty potvrdzujú koreláciu medzi koeficientom transmisivity – koeficientom filtrácie (0,82) a hĺbkou pôdy – skeletnosťou pôdy (-0,72) na základe hodnôt *Cramer's V*. Okrem toho, *Spearmanové korelácie* udávajú povahu asociácie podľa toho, či je hodnota kladná alebo záporná: pixely s vyšším koeficientom priepustnosti majú zvyčajne aj vyšší koeficient filtrácie, zatiaľ, čo pixely s väčšou hĺbkou pôdy majú zvyčajne nižšiu skeletnosť pôdy.

Obr. 2: Matica Spearmanových korelácií



### Modelovanie náchylnosti na zosuv pôdy

#### Logistická regresia

Prítomnosť korelovaných premenných mala vplyv na nejednoznačnú interpretabilitu koeficientov  $\beta$  v logistickej regresii, preto bolo potrebné urobiť výber korelovaných premenných podľa logistickej regresie. Dáta boli rozdelené náhodne do trénovacej a testovacej vzorky tak, že v trénovacej vzorke bol rovnaký počet zosuvných a nezosuvných dát (pixelov) s tým, že zosuvných bolo 70 % všetkých zosuvných údajov (spolu 771 254 pozorovaní v trénovacej vzorke). Do testovacej vzorky išli všetky ostatné dáta (3 452 399). Do modelu logistickej regresie (pomocou funkcie *glm* v programe *R*) vstupovala jedna premenná zo skupiny korelovaných a ostatné premenné. Na dátach z trénovacej vzorky bol natrénovaný model a vypočítané  $AIC$ ,  $AUC_{tren}$ . Z dát z testovacej vzorky bola vypočítaná ich predikcia a  $AUC_{test}$ . Výsledky mediánov koeficientov zodpovedajúce výsledkom logistických regresii z 11-tich náhodne vygenerovaných trénovacích (a testovacích) vzoriek sú uvedené v tab. 1. Variancia  $AUC_{test}$  z 11-tich regresii sa pohybovala na úrovni  $4 \cdot 10^{-4}$ . Identifikátor ( $p$ ) pri spojitých premenných predstavoval premennú chápanú ako polynóm stupňa  $(n-1)$  u danej premennej s pôvodne  $n$  kategóriami.

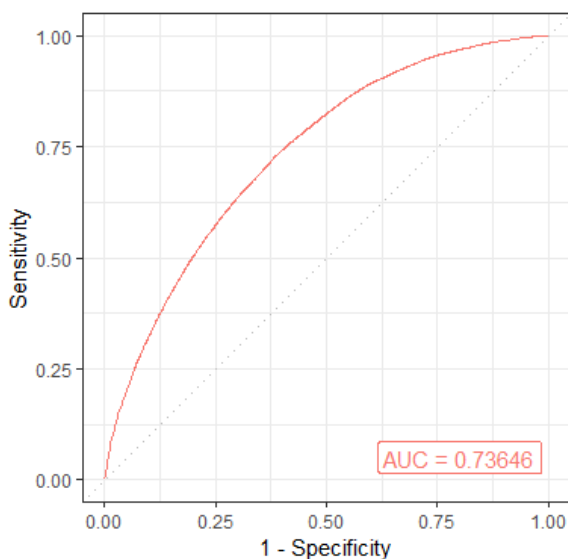
Tab. 1: Mediány koeficientov zodpovedajúcich za výsledky logistických regresii

Premenná	# regresorov	$AIC$	$AUC_{tren}$	$AUC_{test}$
HGU	58	926 048	0,73387	0,73359
GU	53	917 575	0,74104	0,74082
TC	48	944 417	0,71796	0,71764
TC (p)	52	933 722	0,72724	0,72697
FC	48	943 186	0,71693	0,71653
SRM	63	916 670	0,74082	0,74041

Premenná	# regresorov	AIC	$AUC_{tren}$	$AUC_{test}$
PC	56	918 870	0,74108	0,74084
ST	64	914 065	0,74366	0,74326
SDe	56	918 426	0,73994	0,73989
SDe (p)	57	918 279	0,73999	0,73974
SSK	56	918 221	0,73988	0,73987
SSK (p)	58	917 758	0,74001	0,73983

Po zohľadnení výkonnosti modelov ( $AIC$ ,  $AUC$ ), počtov stupňov voľnosti a interpretability premenných podľa tab. 1 boli vybraní reprezentanti korelovaných skupín, a to: geologické jednotky, horizontálna krivosť reliéfu a pôdne jednotky. ROC krivka modelu logistickej regresie z testovacích dát zodpovedala  $AUC_{test} = 0,73646$  (obr. 3).

Obr. 3: ROC krivka modelu logistickej regresie z testovacích dát

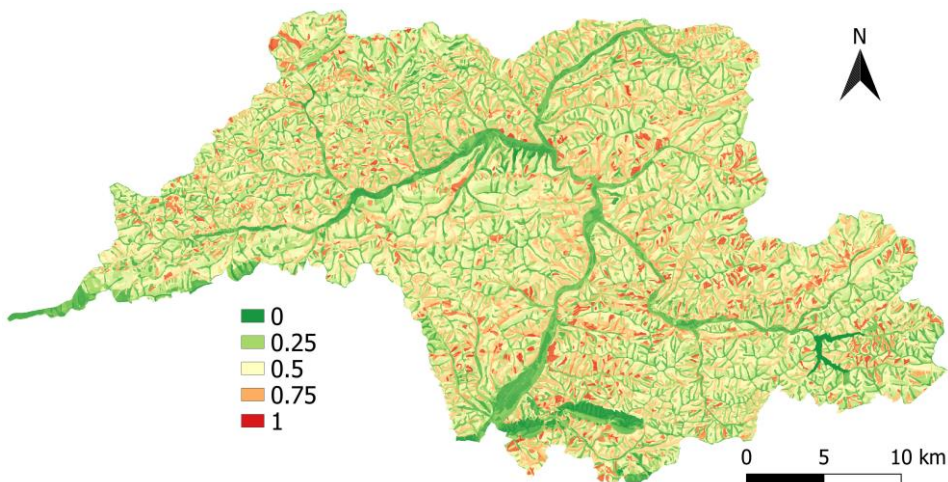


Optimálny zlomový bod z vzhľadom na *Matthewsov korelačný koeficient* na testovacích dátach bol 0,59 a hodnota  $MCC = 0,1582$ . Presnosť tejto klasifikácie bola 79 %. Ak bol pixel nezosuvný, logistická regresia ho pri tomto deliacom bode na 80 % klasifikovala ako nezosuvný, avšak ak bol pixel zosuvný, logistická regresia ho správne klasifikovala len v polovici prípadov.

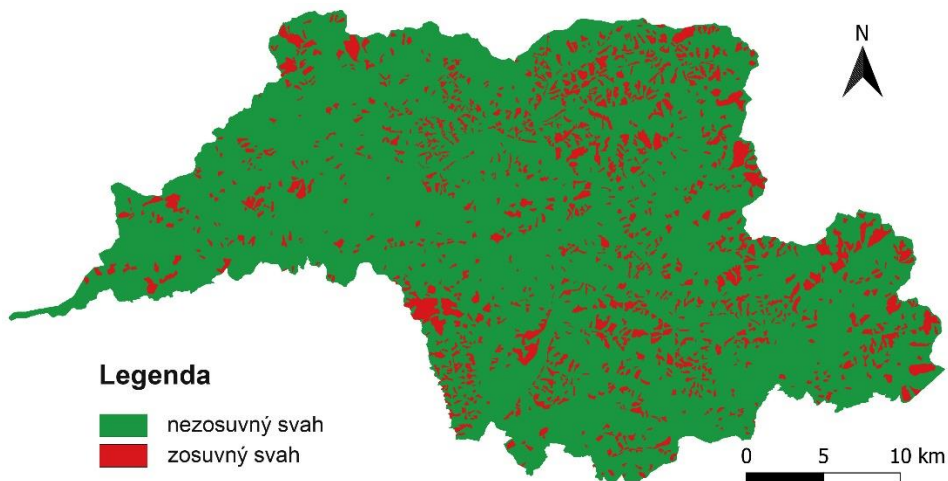
Vizualizácia predikovaných hodnôt získaná pomocou modelu logistickej regresie (mapa 2) bola porovnávaná so skutočnými zosuvmi (mapa 3). Model logistickej regresie možno na väčšine územia označiť ako opatrný, pretože len zriedka predikuje jednotlivým pixelom pravdepodobnosti na zosuv nižšie ako 0,25 alebo vyššie ako 0,75. Model logistickej regresie je možné rozšíriť o viac regresorov, a tým zlepšiť jeho výkonnosť.

V našom prípade sa väčšina pridaných premenných ukázala ako významná, výsledné hodnoty  $AUC_{test}$  aj pre viac ako 100 regresorov nikdy neprekročili hranicu  $AUC_{test} = 0,76$ .

Mapa 2: Mapa náchylnosti svahov na zosuv podľa modelu logistickej regresie



Mapa 3: Mapa zosuvnosti sledovaného územia

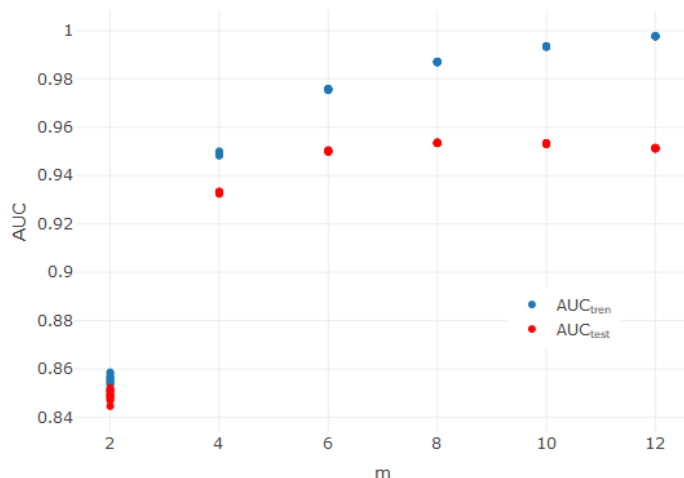


### Model náhodných lesov

Pri modelovaní náhodnými lesmi sme postupovali ako pri modelovaní logistickej regresie, s vylúčením výberu korelovaných premenných. Dáta boli rozdelené tak, že v trénovacej vzorke bol rovnaký počet zosuvných a nezosuvných dát s tým, že zosuvných bolo 70 % (zo všetkých zosuvných údajov). Do testovacej vzorky vstupovali všetky ostatné dáta.

Do modelu náhodných lesov vstupovali dva hlavné parametre:  $N$  - počet stromov v lese a  $m$  - počet premenných v strome. Keďže rastúce  $N$  má podľa Jamesa (2013) pozitívny efekt na výsledný model, tak sme zvolili maximálny počet, čo však neplatí pre parameter  $m$ . Na 11-tich náhodne vygenerovaných tréningových (a testovacích) vzorkách boli natréňované modely náhodných lesov (v programe R pomocou funkcie randomForest) pre rôzne hodnoty  $m$  a následne vypočítané výsledné  $AUC_{tren}$  a  $AUC_{test}$  (obr. 4). Disperzie  $AUC$  klesajú s rastúcim  $m$  z dôvodu vyšších korelácií stromov v lese ( $N = 50$ ).

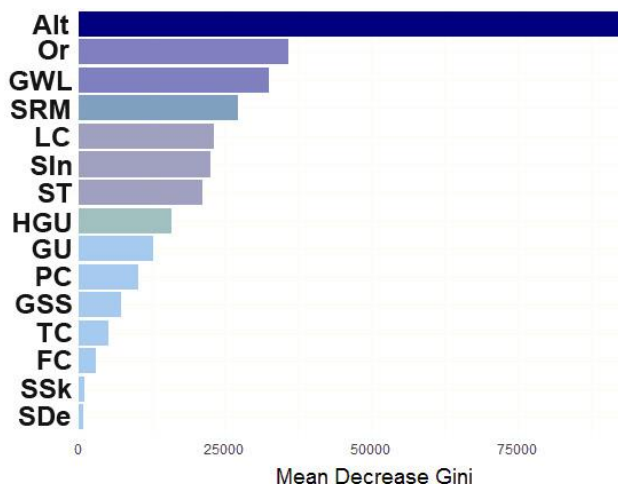
Obr. 4: Hodnoty  $AUC_{tren}$  a  $AUC_{test}$  modelu náhodných lesov so všetkými premennými



Najvyššie hodnoty  $AUC_{test}$  boli zaznamenané pri  $m = 8$ , a to priemerne 0,95351. Keďže hodnoty  $AUC_{tren}$  sa od hodnôt  $AUC_{test}$  veľmi nelíšia, nemožno teda hovoriť o pretrénovaní modelu, preto sme pracovali s parametrom  $m = 8$ .

Náhodné lesy neumožňujú testovať signifikanciu premenných v modeli tak, ako pri logistickej regresii. Umožňujú však sledovať dôležitosť premenných v modeli, a to napríklad pomocou metriky *Mean Decrease Gini*. Graf dôležitosti premenných pre model náhodných lesov s parametrami  $m = 8$ ,  $N = 60$ , so všetkými premennými a na náhodne vygenerovanej vyváženej tréningovej vzorke možno vidieť na obr. 5.

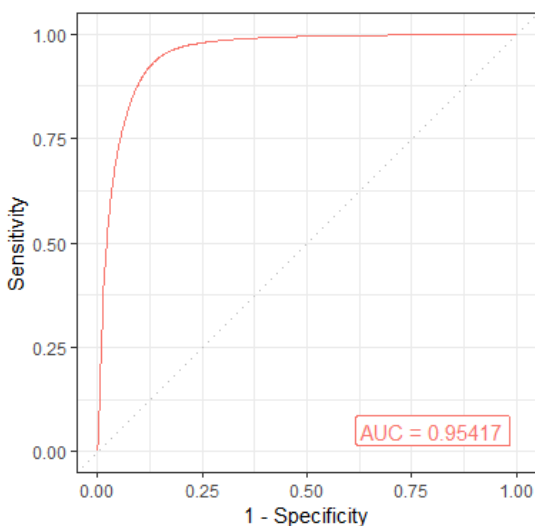
Obr. 5: Graf dôležitosti premenných podľa Mean Decrease Gini pre model náhodných lesov



Výrazne najdôležitejšou premennou v modeli náhodných lesov bola nadmorská výška, čo však mohlo byť spôsobené aj faktom, že je ako jediná v modeli pôvodne spojitá premenná (Strobl et al., 2007). Ďalej sa ako dôležitými ukazujú premenné orientácia, hladina podzemnej vody a morfológia reliéfu. Sklon je až na šiestom mieste. Premenné, ako hĺbka pôdy a skeletnosť pôdy sa ukázali ako najmenej dôležité.

Tento model dosahoval hodnoty  $AUC_{tren} = 0,98938$  na trénovacej vzorke a  $AUC_{test} = 0,95417$  na testovacej vzorke, čo je výrazne viac ako hodnoty pri logistickej regresii. ROC krivka modelu z testovacích dát je zobrazená na obr. 6.

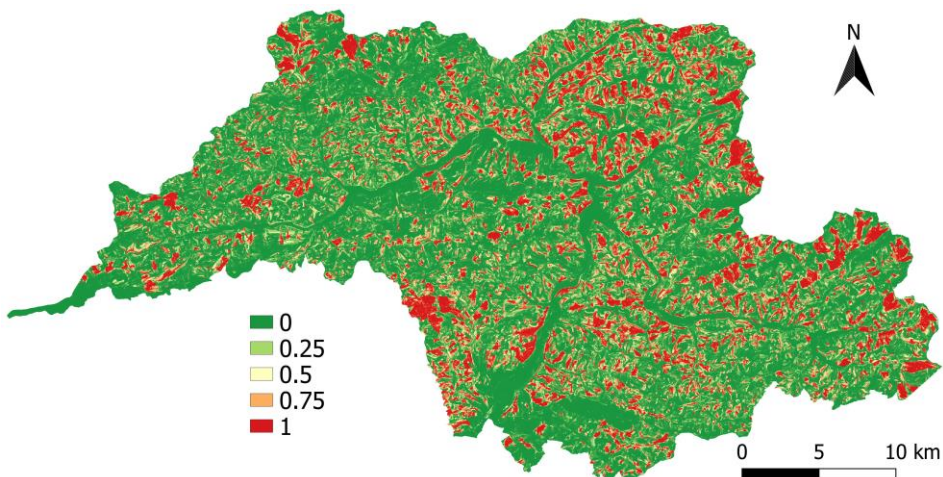
Obr. 6: ROC krivka modelu náhodných lesov z testovacích dát



Optimálny zlomový bod z vzhľadom na *Matthewsov korelačný koeficient* na testovacích dátach je  $z = 0,91$  a dosahuje v ňom hodnotu  $MCC = 0,52826$ . Ak je pixel nezosuvný, náhodné lesy ho pri tomto deliacom bode na 95 % klasifikujú ako nezosuvný a ak je pixel zosuvný, náhodné lesy ho na 71 % klasifikujú ako zosuvný. To možno označiť za výrazné zlepšenie oproti logistickej regresii, kde bola pri zosuvných pixeloch pravdepodobnosť správneho zaradenia len 50 %.

Vizualizáciu predikovaných hodnôt získanú pomocou modelu náhodných lesov (mapa 4) je možné porovnať so skutočnými hodnotami (mapa 3). Model náhodných lesov si aj vizuálne počínal značne lepšie ako model logistickej regresie a na väčšine pixelov správne predikoval výrazne nízke alebo výrazne vysoké pravdepodobnosti. Model náhodných lesov sme sa snažili vylepšiť viacerými spôsobmi, čiže postupne odoberať premenné s najnižšou informačnou hodnotou alebo korelované premenné a následne meniť parametre  $m$  a  $N$ . Výrazné zlepšenie modelu však nenastalo a hodnoty  $AUC_{test}$  neprekročili ani hranicu 0,96.

Mapa 4: Mapa náchylnosti svahov na zosuv podľa modelu náhodných lesov



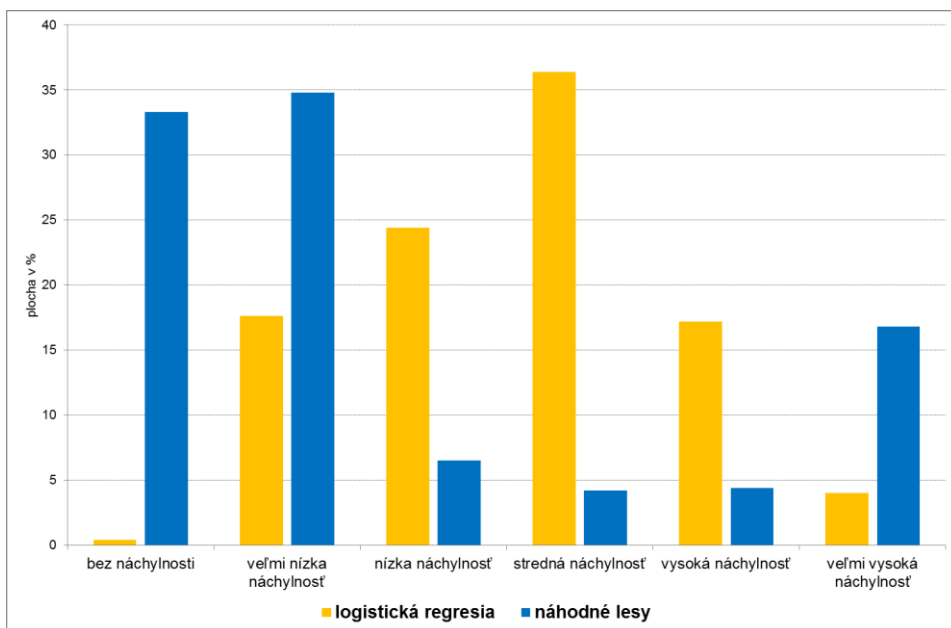
Pre zjednodušenie interpretácie výsledkov boli predpovedané pravdepodobnosti z modelu logistickej regresie a modelu náhodných lesov rozdelené do piatich kategórií náchylnosti na zosuvnosť (veľmi nízka, nízka, stredná, vysoká a veľmi vysoká). Tieto kategórie boli vytvorené na základe predikcií 0 – 1, t.j. predikcie 0 boli v kategórii bez náchylnosti; veľmi nízka náchylnosť 0,01 – 0,20; nízka náchylnosť 0,21 – 0,40; stredná náchylnosť 0,41 – 0,60; vysoká náchylnosť 0,61 – 0,80 a veľmi vysoká náchylnosť 0,81 – 1.

Podľa modelu logistickej regresie bolo bez náchylnosti na zosuvy klasifikovaných len 0,4 % územia Kysúc, podľa modelu náhodných lesov to bolo až 33,3 %. V modeli logistickej regresie pokrývajú svahy s veľmi nízkou náchylnosťou 17,2 % plochy, v modeli náhodných lesov je to 34,8 % plochy. Svahy s veľmi vysokou náchylnosťou pokrývajú



4,0 % plochy v modeli logistickej regresie a 16,8 % plochy v modeli náhodných lesov (graf 1). Pri porovnaní s aktuálne registrovanými zosuvmi, ktoré pokrývajú 18,26 % územia (1869 zosuvov), sa model náhodných lesov javí ako realistickejší, čo je možné vidieť aj na základe výsledkov (mapa 3 a 4).

Graf 1: Porovnanie modelov pri hodnotení náchylnosti územia na zosuv



## Záver

Naša práca predstavuje štúdium hodnotenia zosuvnosti v krajine na základe dvoch štatistických predikčných metód (metódou logistickej regresie a metódou náhodných lesov). Výsledky poukazujú na to, že ako vysoko korelované (na základe matice korelácií *Cramer's V*) sa ukázali tri skupiny premenných. Miera závislosti premenných bola určená na základe *Spearmanových korelácií*, kde sa potvrdila silná korelácia medzi premennými ako je koeficient prietočnosti - koeficient filtrácie, hĺbka pôdy – skeletnosť pôdy. Medzi dôležité premenné patrí nadmorská výška, orientácia a hĺbka podzemnej vody.

Geologické jednotky, horizontálna krivosť reliéfu a pôdny typ boli vybrané ako reprezentanti pri modeli logickej regresii. Presnosť klasifikácie pri tomto modeli bol len 79 %. Preto ho možno označiť ako opatrný, pretože len zriedka predikuje jednotlivým pixelom pravdepodobnosti na zosuv nižšie ako 0,25 alebo vyššie ako 0,75. *ROC* krivka zodpovedajúca  $AUC_{test}$  bola 0,73646 a *Matthewsov korelačný koeficient* na testovacích dátach bol 0,1582 ( $z = 0,59$ ). V modeli náhodných lesov sa ako dôležité premenné ukázali orientácia, hladina podzemnej vody a morfológia reliéfu. Najvyššie hodnoty  $AUC_{test}$  boli zaznamenané pri 8 premenných, priemerne 0,95351. Pri tomto modeli bol vyšší aj optimálny zlomový bod z vzhľadom na *Matthewsov korelačný koeficient* na

testovacích dátach, ktorý bol 0,91 s hodnotou  $MCC = 0,52826$ . V práci Pham et al. (2021) bola celková presnosť v danom modeli 90,5 % a AUC 91,2 %.

Celkovo výrazne lepšie výsledky v modelovaní priniesli modely náhodných lesov nielen pre binárnu, ale aj pre kategoriálnu ordinálnu závislú premennú. Presnosť tejto klasifikácie bola 94 % ( $AUC_{\text{test}} 0,96$ ). Pham et al. (2021) uvádzajú celkovú presnosť pri tomto modeli 98,3 % (AUC 97 %) a Ali et al. (2021) uvádzajú AUC 0,954. Dominancia náhodných lesov mohla byť spôsobená aj veľkosťou dátovej vzorky (približne 700 000 pozorovaní v tréningovej vzorke). Z výsledkov vyplýva, že závislú premennú je dobré používať ako kategoriálnu ordinálnu, ak sú všetky kategórie v dátach zastúpené dostatočným počtom pozorovaní. Ak tomu tak nie je a existuje prirodzený a dobre interpretovateľný zlomový bod, potom so závislou premennou možno pracovať ako s binárnou. Ukázalo sa, že je oveľa lepšie používať spojité a nie kategoriálne premenné.

Na modelovanie zosuvnosti svahov sa dnes úspešne používajú rôzne iné modelovacie techniky, najmä však neuronové siete, ktoré majú dobré predpoklady riešiť zložité modely. Na pixely je možné nepozerať sa ako na samostatné pozorovania, ale vhodným spôsobom využiť aj vlastnosti okolitých pixelov.

S postupnou urbanizáciou a zvyšujúcou sa potrebou vyššieho životného komfortu sú projektanti nútení podrobnejšie sa zaoberať hodnotením zložitých inžinierskogeologických pomerov pri posudzovaní pozemných, podzemných, líniových, vodných, ale aj iných druhov stavieb. Preto správne umiestnenie stavby s dôkladným poznaním súčasného stavu geologického prostredia, ale najmä predpokladaním geologických procesov v budúcnosti, s dôrazom na geobariéry, môže byť prostriedkom na ušetrenie vysokých finančných nákladov na možné sanácie v budúcnosti a v neposlednom rade zároveň zvýši bezpečnosť obyvateľstva.

## PodĎakovanie

*Tento príspevok vznikol s podporou grantového projektu VEGA 2/0132/18 „Historické a súčasné zmeny krajinej diverzity a biodiverzity vplyvom pôsobenia prírodných a antropogénnych faktorov“.*

## Literatúra

AYALEW, L., YAMAGISHI, H., 2005: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, 65, 1 – 2, 15 – 31.

AGRESTI, A., 2007: *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Hoboken, New Jersey.

AGRESTI, A., 2010: *Analysis of Ordinal Categorical Data*, John Wiley & Sons, Hoboken, New Jersey.

AGRESTI, A., 2018: *An introduction to categorical data analysis*; John Wiley & Sons.

- ALI, S. A., PARVINM, F., VOJTEKOVÁ, J., COSTACHE, R., LINH, N. T. T., PHAM, Q. B., VOJTEK, M., GIGOVIĆ, L., AHMAD, A., GHORBANI, M. A., 2021: GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. *Geoscience Frontiers*, 12, 2, 857 – 876.
- BARANČOKOVÁ, M., KENDERESSY, P., 2014: Assessment of landslide risk using GIS and statistical methods in Kysuce region. *Ekológia (Bratislava)*, 33, 1, p. 26 – 35.
- BIÈVRE, G., JONGMANS, D., WINIARSKI, T., ZUMBO, V., 2012: Application of geophysical measurements for assessing the role of fissures in water infiltration within a clay landslide (Trieves area, French Alps). *Hydrological Processes*, 26, 2128 – 2142.
- BEDNARIK M., MAGULOVÁ B., MATYS M., MARSCHALCO M., 2010: Landslide susceptibility assessment of the Kraľovany–Liptovský Mikuláš railway case study. In: *Physics and Chemistry of the Earth*, 35, 3 – 5, 162 – 171.
- BOUGHORBEL, S., JARRAY, F., EL-ANBARI, M., 2017: Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric, *PLoS One*, 12, 6.
- BUDIMIR, M., ATKINSON, P., LEWIS, H. A., 2015: systematic review of landslide probability mapping using logistic regression. *Landslides*, 12, 419 – 436.
- CARRARA, A., 1983: Multivariate models for landslide hazard evaluation. *Mathematical Geology*, 15, 3, 403 – 427.
- CARRARA, A., CARDINALI, M., GUZZETTI, F., REICHENBACH, P., 1995: GIS technology in mapping landslide hazard. *Geographical information systems in assessing natural hazards*. Kluwer Academic Publisher, Dordrecht, The Netherlands, 135 – 175.
- CIAMPALINI, A., RASPINI, F., LAGOMARSINO, D., CATANI, F., CASAGLI, N., 2016: Landslide susceptibility map refinement using PSInSAR data. *Remote Sensing of Environment*, 184, 302 – 315.
- DOU, J., TIEN BUI, D. P., YUNUS, A., JIA, K., SONG, X., REVHAUG, I., XIA, H., ZHU, Z., 2015: Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. *PLoS ONE*, 10, 7.
- DORMANN, C. F., ELITH, J., BACHER, S., BUCHMANN, C., CARL, G., CARRÉ, G. MARQUÉZ, J. R. G., GRUBER, B., LAFOURCADE, B., LEITAO, P. J., 2013: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27 – 46.
- ESPRIT, 2013: Obnova a budovanie technickej infraštruktúry výskumu a vývoja Ústavu krajinej ekológie SAV. Priestorový databázový systém. ÚKE SAV, Bratislava.
- FELL, R., COROMINAS, J., BONNARD, CH., CASCINI, L., LEROI, E., SAVAGE, W. Z., 2008: Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*, 102, 99 – 111.
- GUZZETTI, F., STARK, C. P., SALVATI, P., 2005: Evaluation of food and landslide risk to the population of Italy. *Environ. Manag.* 36, 1, 15 – 36.

- HAVLÍN, A., BEDNARIK, M., MAGULOVÁ, B., VLČKO, J., 2011: Použití logistické regrese pro hodnocení náchylnosti k sesouvání ve střední části Chřibu (Česká republika) *Acta Geologica Slovaca* 3, 2, 153 – 161.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., 2009: The elements of statistical learning. Cited on, 33.
- HORNUNG, R., 2019: Ordinal Forests. *Journal of Classification*, 37, 4 – 17.
- CHEN, W., SHAHABI, H., ZHANG, S., KHOSRAVI, K., SHIRZADI, A., CHAPI, K., PHAM, B. T., ZHANG, T., ZHANG, L., CHAI, H., MA, J., CHEN, Y., WANG, X., LI, R., AHMAD, B. B., 2018: Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Appl. Sci.*, 8, 2540.
- JAMES, G., WITTEN, D., HASTIE, T., TIBSHIRANI, R., 2013: An Introduction to Statistical Learning with Applications in R, Springer, New York.
- KRZEMINSKA, D. M., BOGAARD, T. A., MALET, J. P., VAN BEEK, L. P. H., 2013: A model of hydrological and mechanical feedbacks of preferential fissure flow in a slow-moving landslide. *Hydrology and Earth System Sciences*, 17, 947 – 959.
- LEE, M. J., PARK, I., WON, J. S., LEE, S., 2016: Landslide hazard mapping considering rainfall probability in Inje, Korea. *Geomatics, Natural Hazards and Risk*, 7, 1, 424 – 446.
- MAZÚR, E., LUKNIŠ, M., 1986: Geomorfologické členenie SSR a ČSSR. Časť SSR. 1:50 000. Slovenská kartografia, Bratislava.
- METEN, M., PRAKASHBHANDARY, N., YATABE, R., 2015: Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. *Geoenvironmental Disasters*, 2, 9.
- METELKA, V., KYCL, P., 2007: Mapování náchylností svahů k sesouvání v prostředí GIS, Miramar, Kostarika. In: I. Baroň, J. Klimeš, D. Kašperová (eds.) *Sborník Svahové deformace a pseudokras 2007 Vsetín: Česká geologická služba, Ústav struktury a mechaniky hornin, AVČR*, p. 1 – 15.
- NEULAND, H., 1976: A prediction model of landslips, *Catena*, 3, 2, 215 – 230.
- MUŇKO, M., 2015: Usability of Support Vector Machines method in prediction modelling in GIS. Diplomová práca, Slovenská technická univerzita, Bratislava.
- PAUDITŠ, P., VLČKO, J., JURKO, J., 2005: Využívanie štatistických metód pri hodnotení náchylností územia na zosúvanie. *Mineralia Slovaca*, 37, 4, 529 – 538.
- PHAM, Q. B., ACHOUR, Y., ALI, S. A., PARVIN, F., Vojtek, M., Vojteková, J., AL-ANSARI, N., ACHU, A. L., COSTACHE, R., KHEDHER, K. M., ANH, D. T., 2021: A comparison among fuzzy multi-criteria decision making, bivariate, multivariate and machine learning models in landslide susceptibility mapping. *Geomatics, Natural, Hazards and Risk*, 12:1, 1741 – 1777.
- PREFAC, Z., DUMITRU, S., CHENDEŞ, V., SÎRODOEV, I., CRACU, G., 2016: Assessment of landslide susceptibility using the certainty factor model: Răşcuţa

catchment (Curvature Subcarpathians) case study. *Carpathian Journal of Earth and Environmental Sciences*, 11, 2, 617 – 626.

REICHENBACH, P., ROSSI, M., MALAMUD, B. D., MIHIR, M., GUZZETTI, F., 2018: A review of statistically-based landslide susceptibility models, *Earth-Science Reviews*, 180, 60 – 91.

SAFAEI, M., OMAR, H., HUAT, B. K., YOUSOF, Z. B. M., 2012: Relationship between Lithology Factor and landslide occurrence based on Information Value (IV) and Frequency Ratio (FR) approaches - Case study in North of Iran. *Electronic journal of geotechnical engineering*, 17, 79 – 90.

SAKKAS, G., MISAILIDIS, I., SAKELLARIOU, N., KOUSKOUNA, V., KAVIRIS, G., 2016: Modeling landslide susceptibility in Greece: a weighted linear combination approach using analytic hierarchical process, validated with spatial and statistical analysis. *Natural Hazards*, 84, 1873 – 1904.

STROBL, C., BOULESTEIX, A-L, ZEILEIS, A., HORTHORN, T., 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics*, 8, 25.

ŠIMEKOVÁ, J., MARTINČEKOVÁ, T. et. al., 2006: Atlas máp stability svahov SR v mierke 1:50000. Záverečná správa. Geofond, Bratislava: MŽP SR, 155 p.

TSENG, C. M., LIN, C. W., HSIEH, W. D., 2015: Landslide susceptibility analysis by means of event-based multi-temporal landslide inventories. *Natural Hazards Earth System Sciences Discussion*, 3, 1137 – 1173.

VAN WESTEN, C. J., VAN DUREN, I., KRUSE, H. M. G., TERLIEN, M. T. J., 1993: GISSIZ: training package for geographic information systems in slope instability zonation. ITC Publication, 15. International Institute for Aerospace and Earth Resources Survey, ITC, Enschede.

VAN WESTEN, C. J., CASTELLANOS ABELLA, E. A., SEKHAR, L. K., 2008: Spatial data for landslide susceptibility, hazards and vulnerability assessment : an overview. *Engineering geology*, 102, 112 – 131.

WANG, G., CHEN, X., CHEN, W., 2020: Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions. *ISPRS Int. J. Geo-Inf.* 9, 144.

WIEDENMANN, J., ROHN, J., MOSER, M., 2016: The relationship between the landslide frequency and hydrogeological aspects: a case study from a hilly region in Northern Bavaria (Germany). *Environmental Earth Sciences*, 75, 609.

WU, C., QIAO, J., 2009: Relationship between landslides and lithology in the Three Gorges Reservoir area based on GIS and information value model. *Frontiers of Forestry in China*, 4, 165 – 170.