Random forest classification of morphology in the northern Gerecse (Hungary) to predict landslide-prone slopes

Gáspár Albert[1] and Dávid Gerzsenyi[2]

1 Eötvös Loránd University, Cartography and Geoinformatics, Budapest, Hungary (albert(at)ludens.elte.hu)

2 Eötvös Loránd University, Doctorate School of Earth Sciences, Budapest, Hungary (gerzsd(at)caesar.elte.hu)

Landslide susceptibility in the Gerecse Hills


Damaged house built above landslide scarp (Sárisáp)

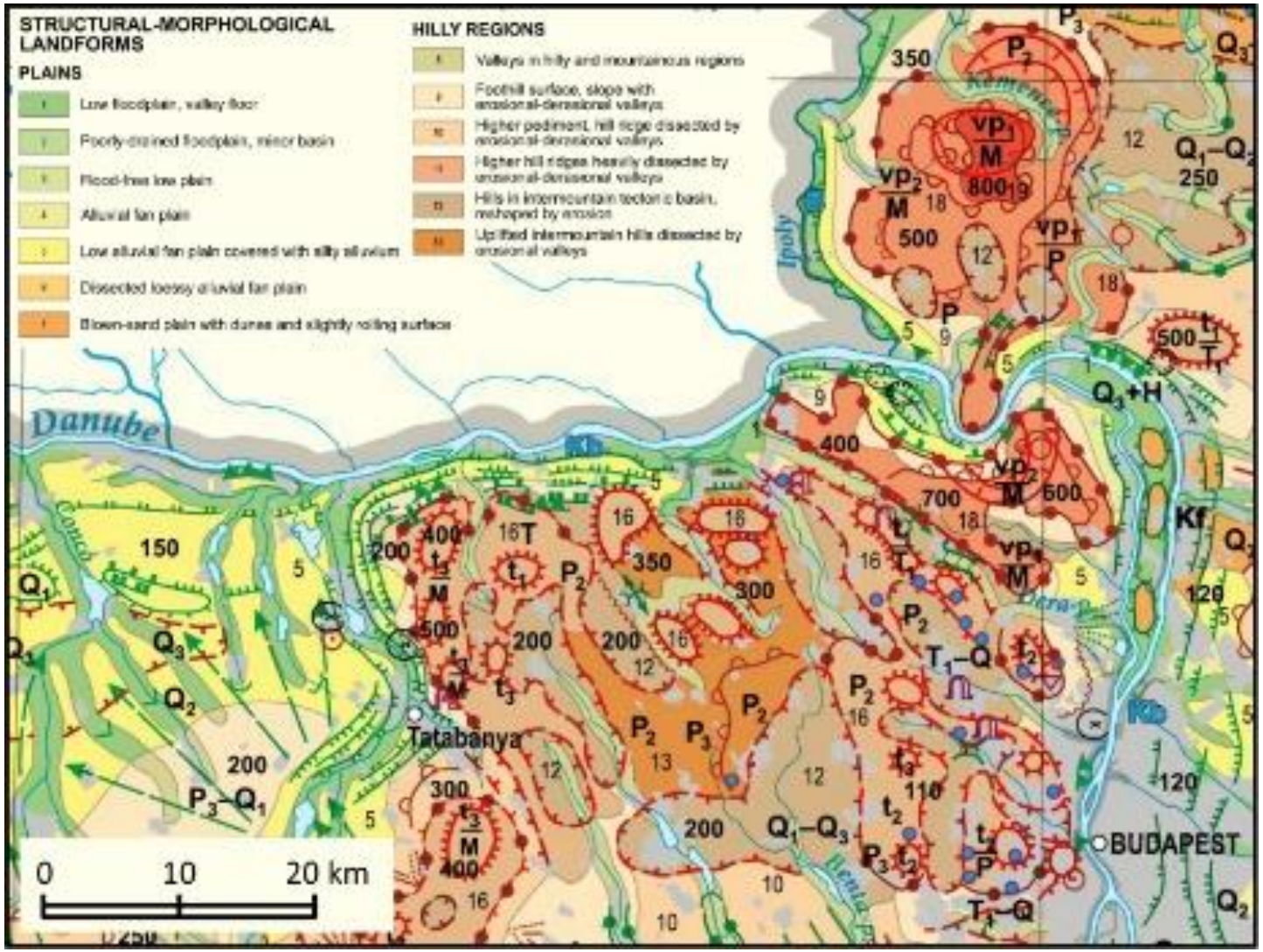

Back-propagating gully in the loess (Neszmély)

On the loess covered northern regions of the Gerecse Hills agricultural (vineyards) areas, built-in slopes and villages are the places where land movements are detected.


Loess-covered slopes prone to landslides (Mogyorósbánya)

General morphology of the area

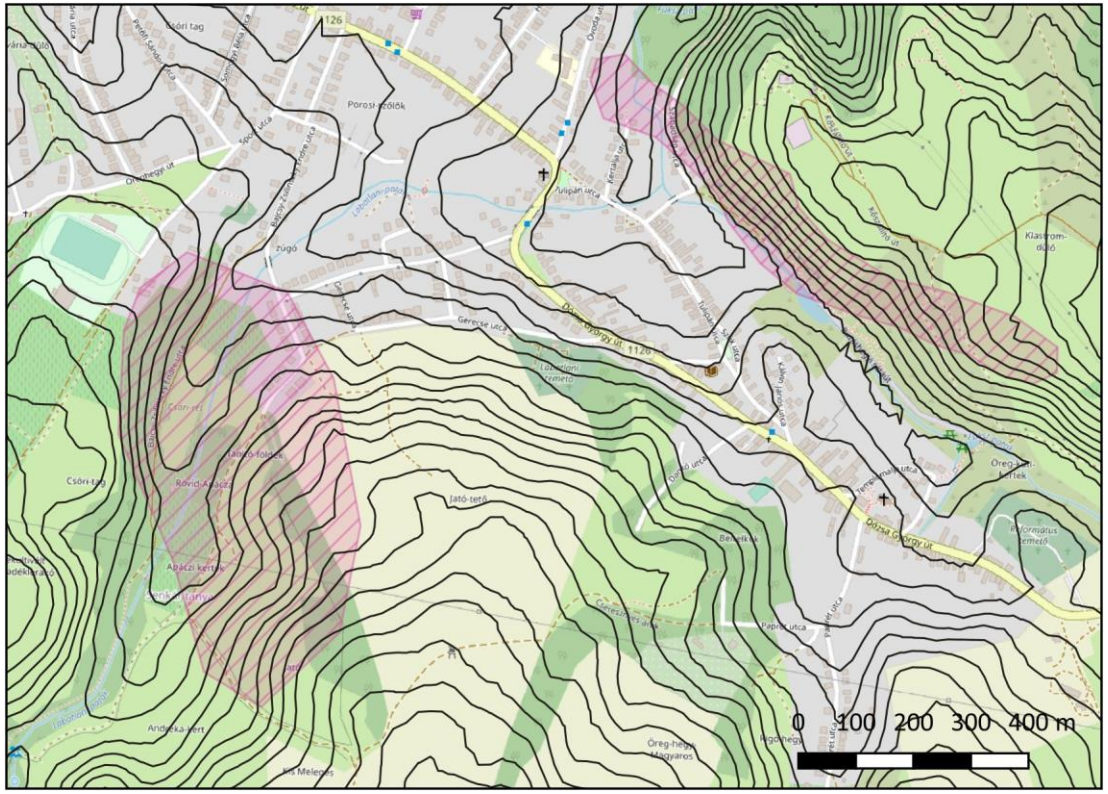
- The morphology of the Gerecse Hills bears the imprints of fluvial terraces of the *Danube River*, Neogene tectonism and Quaternary erosion.
- The solid bedrocks are composed of Mesozoic and Paleogene limestones, marls, and sandstones, and are covered by 115 m thick layers of unconsolidated Quaternary fluvial, lacustrine, and aeolian sediments.
- Hillslopes, stream valleys, and loessy riverside bluffs are prone to landslides, which caused serious damages in inhabited and agricultural areas in the past.



Geomorphological map from the National Atlas of Hungary (2018), Chapter IV. – link here

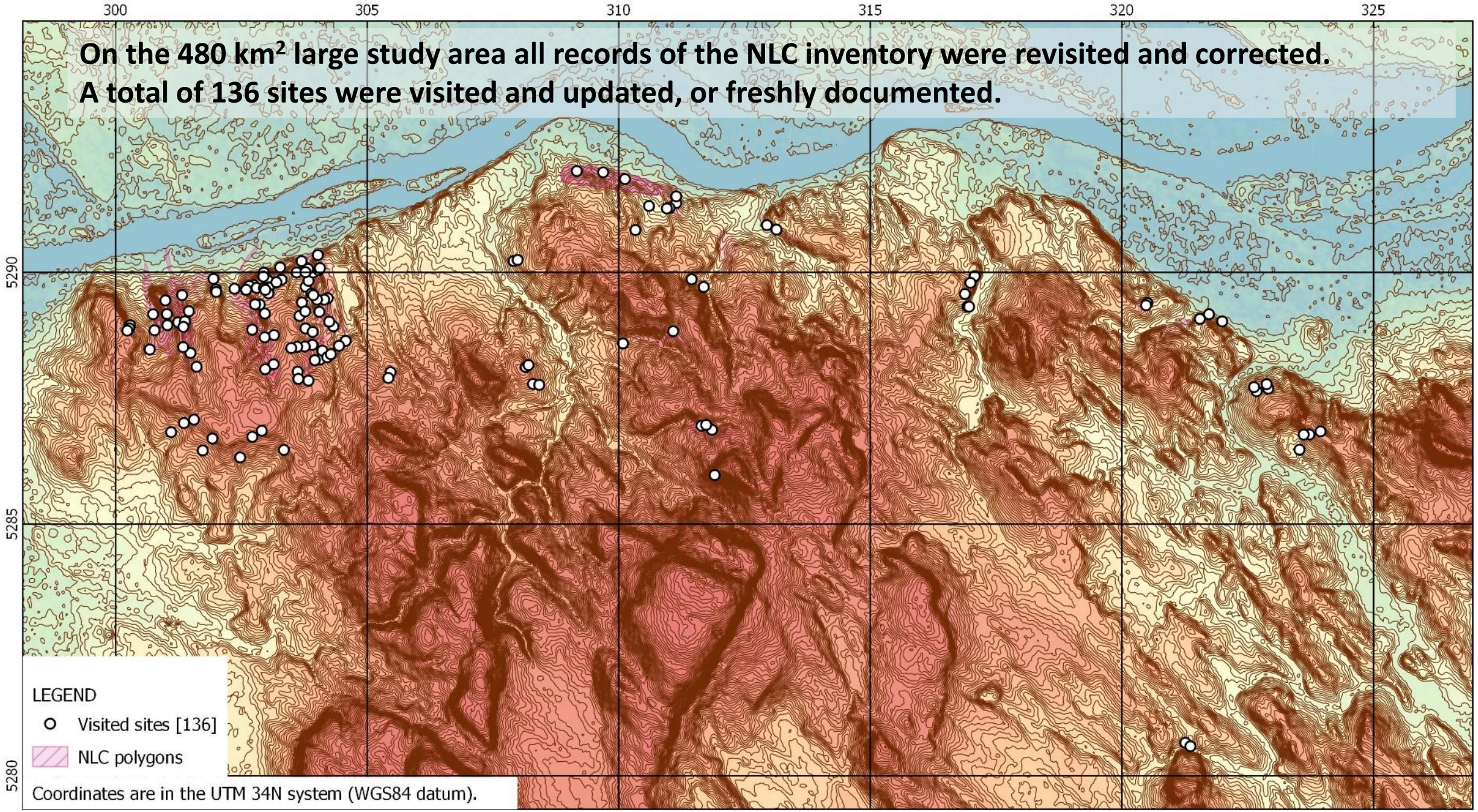
The NLC (National Landslide Cadastre)

- Attempts to map the landslides were made and the observations were documented in the National Landslide Cadastre (NLC) inventory since the 1970's.
- These documentations are sporadic, concentrating only on certain locations.
- They often refer inaccurately to the state and extent of the landslides.



The map shows two roughly outlined polygons of the NLC in the vicinity of Lábatlan village covering very different morphological types.

The aim of the present study was to complete and correct the landslide inventory by using quantitative modelling.


300
305
310
315
320
325
5290
5285
5280
On the 480 km² large study area all records of the NLC inventory were revisited and corrected.
A total of 136 sites were visited and updated, or freshly documented.
LEGEND
Visited sites [136]
NLC polygons
Coordinates are in the UTM 34N system (WGS84 datum).

Morphological categories in the model

- Using objective criteria, the renewed records of the NLC and additional sample locations (136 sites) were sorted into three morphological categories: **scarps, debris, transitional area**.
- Additionally, 45 pieces of stabile accumulation areas, hilltops, and stabile slopes were designated to the records.
- The categorized map of these observations served as training data for the random forest classification (RFC).

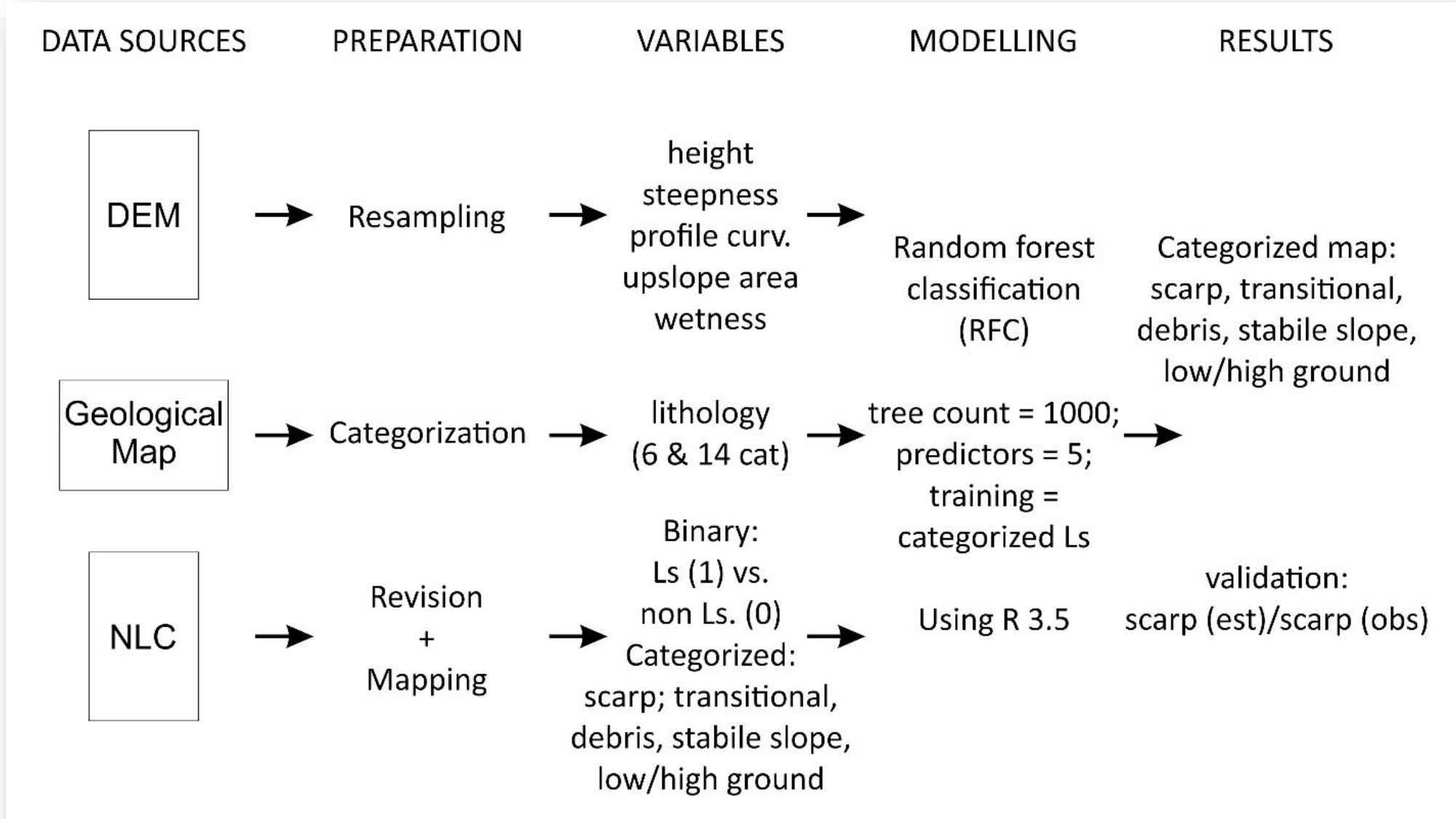
A total of 181 sites of 6 morphological categories were identified and checked on the field.

RFC methodology

- Random forest is a powerful tool for multivariate classification that uses several decision trees.
- The predictions were done for each pixel of medium-resolution (~10 m) rasters.
- The predictor variables of the decision trees were morphometric and geological indices.
- The terrain indices were derived from the MERIT DEM [1] with SAGA routines and the categorized geological data is from a medium-scale geological map [2].
- The predictor variables were packed in a multi-band raster and the random forest classification (RFC) method was executed using R 3.5 with RStudio.

[1] Yamazaki et al. A high accuracy map of global terrain elevations. Geophysical Research Letters, vol. 44, pp. 5844-5853, 2017

[2] Gyalog L., & Síkhegyi F., eds. *Geological map of Hungary* (scale: 1:100 000). Budapest, Hungary, Geological Institute of Hungary, 2005.



An overview of the modelling methodology using RFC.

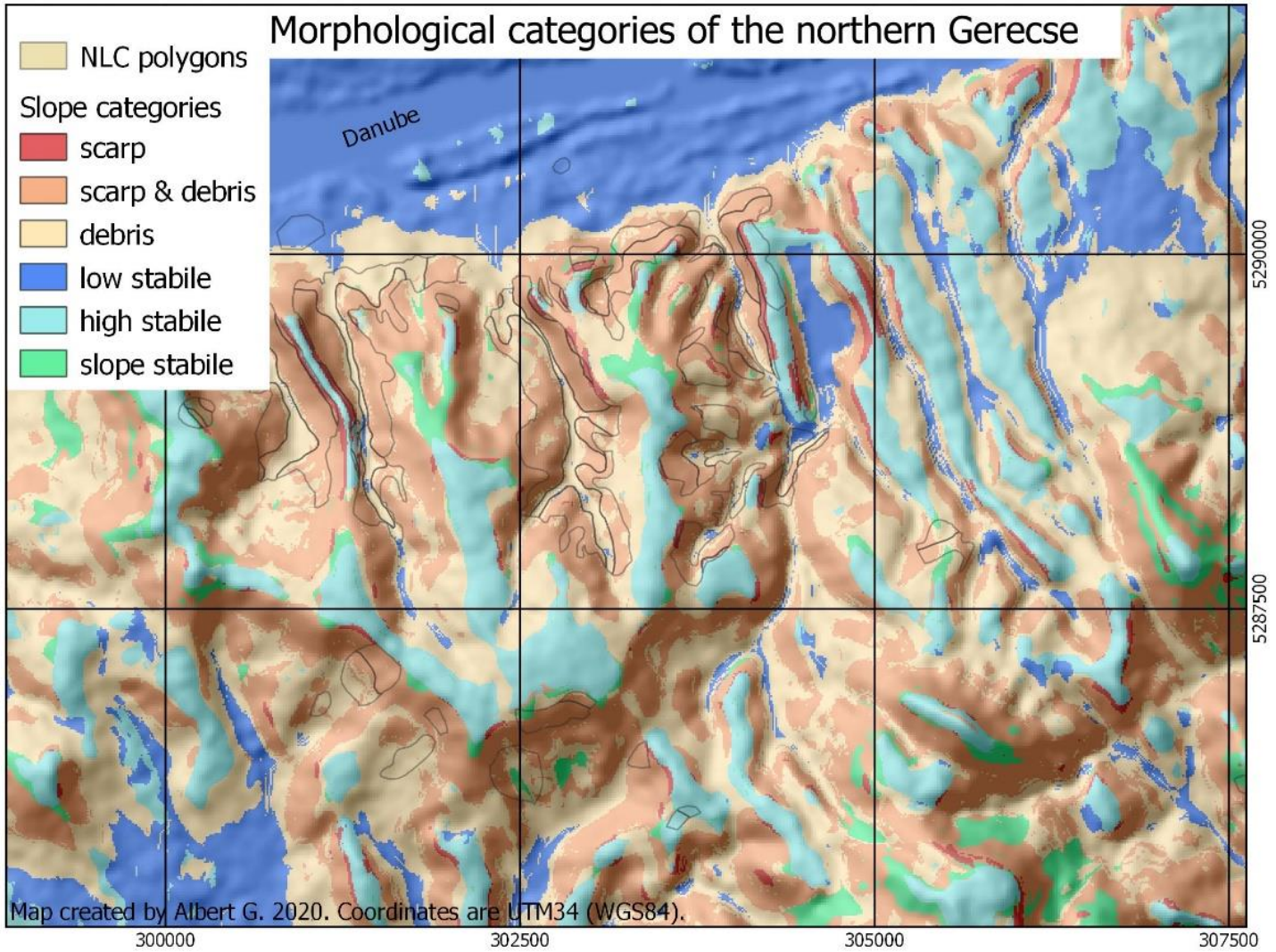


Figure shows the excerpt of the final categorization at the North-western part of the study area, where the landslides were the most frequent.

After testing several combinations of the predictor variables and two different categorization of the geological data, the best prediction has cca. 80% accuracy*.

The used morphometric variables:

TWI, mean upslope area, slope angle, slope curvature, normalized height.

The used geological categories:

1	clay
2	silt
3	travertine
4	sand
5	sand and silt
6	sand-silt-clay
7	old carbonates
8	gravel
9	gravel-sand
10	gravel-sand-silt-clay
11	consolidated mixed debris
12	layered sandstone
13	layered mixed debris
14	loose mixed debris

**cca 80% of the predicted pixels in the training polygons were in the right type.*

Results & conclusions

- The results showed that the probable location of landslide-prone slopes is not restricted to the areas recorded in the National Landslide Cadastre inventory.
- Only ~6% of the highly unstable slopes (scarps) fall within the NLC polygons in the study area according to the model.

_____

1. **The landslide inventory (NLC) records require refinement and corrections before using them for modelling.**
2. **The refinement primarily should include the distinction between scarp, debris, and transitional areas of the landslide locations.**

The results of our modelling serves as the base of an ongoing regional geohazard mapping project.

Acknowledgements

The project was partly supported by the Thematic Excellence Program, Industry and Digitization Subprogram, NRDI Office, project no. ED_18-1-2019-0030 (from the part of G. Albert) and,

The ÚNKP-19-3 New National Excellence Program of the Ministry for Innovation and Technology (from the part of D. Gerzsenyi).

The beautiful landscape of the Dorog Basin that lies to the East of the studied area is also prone to landslides! We look forward to see them!