Predicting the outcrop of pre-Quaternary formations in the Dorog Basin (Hungary) using random forest classification

Réka Pogácsás[1] and Gáspár Albert[2]

1 Eötvös Loránd University, Cartography and Geoinformatics, Budapest, Hungary (reka(at)pogacsas.hu)

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

Study area

- Study area:
  - The Dorog Basin had a huge importance 30-60 years ago, due to Eocene coal lens, but the area is mostly covered by Quaternary sediments.
  - Pre-Quarternery rocks such as Triassic dolomites and Eocene limestones, marls and sandstones has outcrops in between the erosion shaped valleys.
  - Tectonic forces also had a great impact on the area.
  - Currently the renewal of geological maps of this region is in progress.

The central study area

The Gete Hill (457 m a.s.l.) is mainly composed of pre-Quaternary (Triassic and Eocene) outcrops.

Used materials and training area

- The area was previously (30-50 years ago) mapped, with the scale of 1:10000. This was used as a reference.

- The digitalised map was classified into Quaternary and pre-Quaternary formations.

- The input variables of the modell were derived from MERIT DEM [1] of the area. This has a resolution of 3 arc seconds (~90m at the equator here about ~78 m)

- The training area refers to a diverse part of the whole scope, of about 20%.

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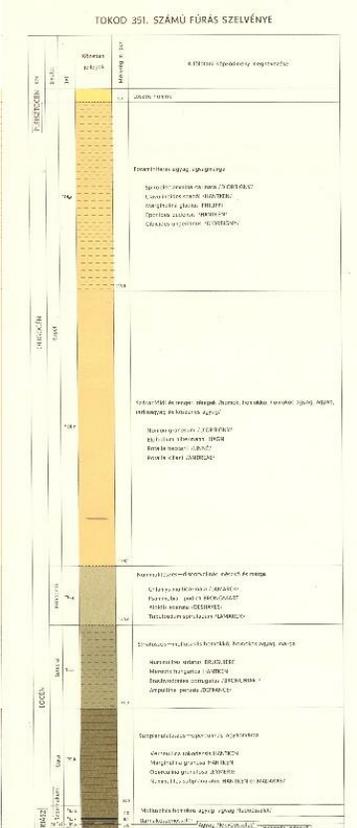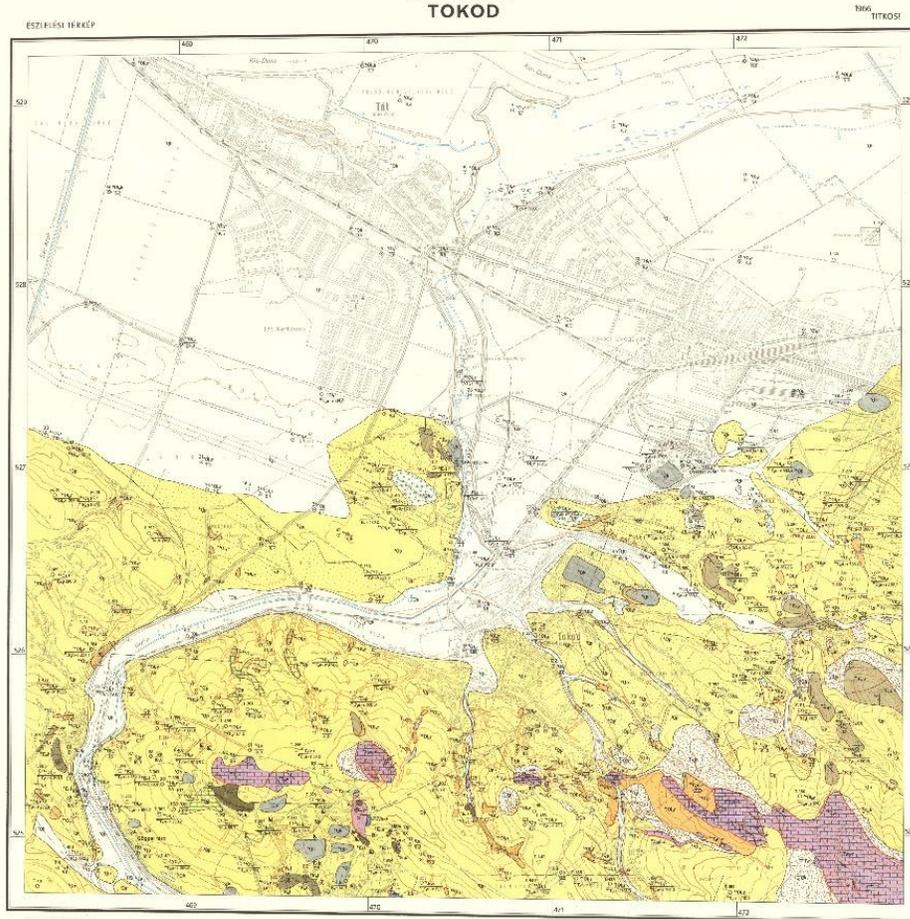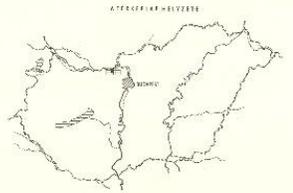
Previous geological map of the region

Gidai, L., Nagy, G., & Sipass, S. (1981). Geological map of the Dorog Basin 1: 25 000. [in Hungarian] Geological Institute of Hungary, Budapest.

Geological series of the Study Area

Legend

- ☐ Holocene
- ☐ Pleistocene
- ☐ Eocene
- ☐ Jurassic
- ☐ Triassic
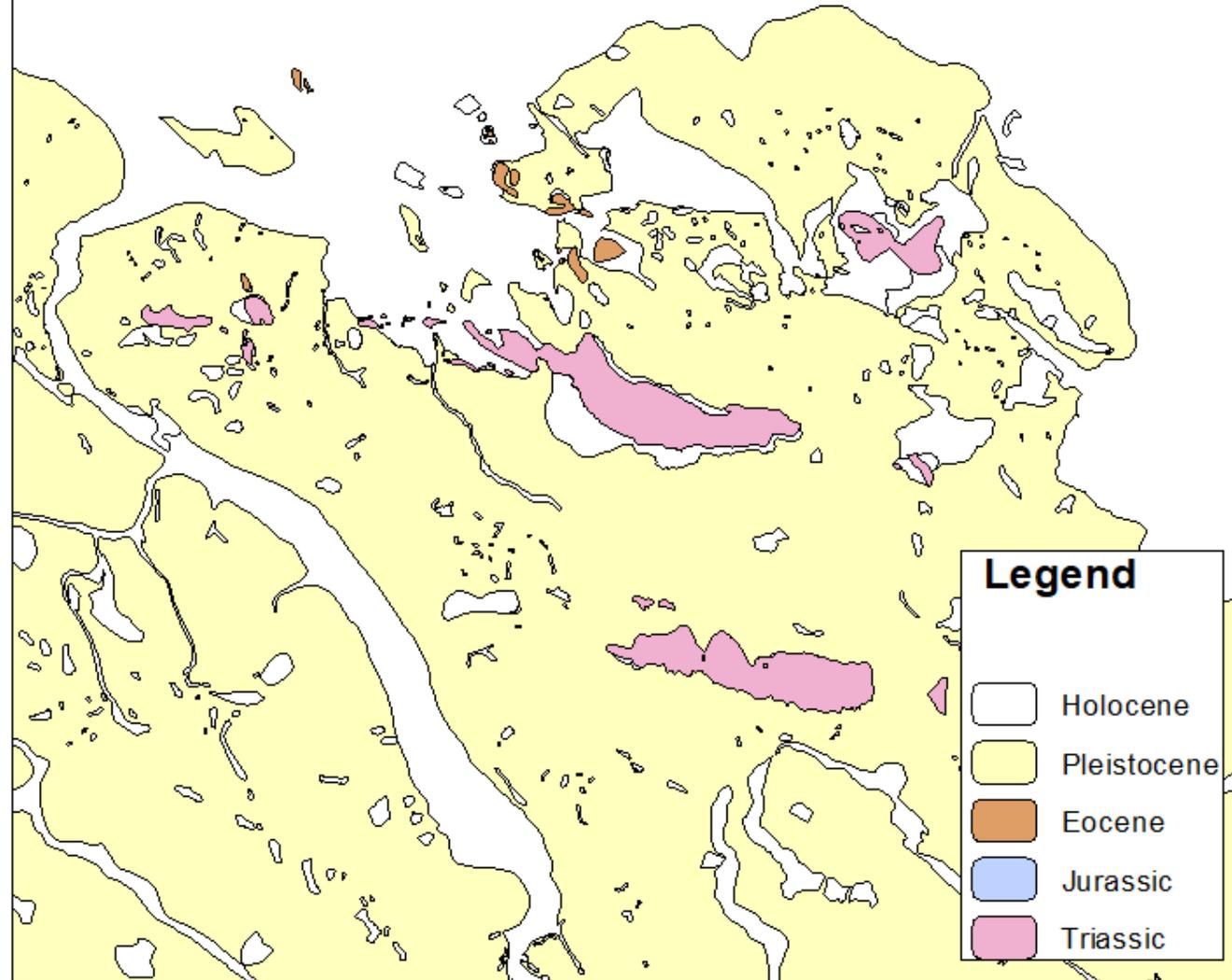
The study areas (red), and the classification of the existing geological map. White: pre-Quarternety, Black:Quarternery
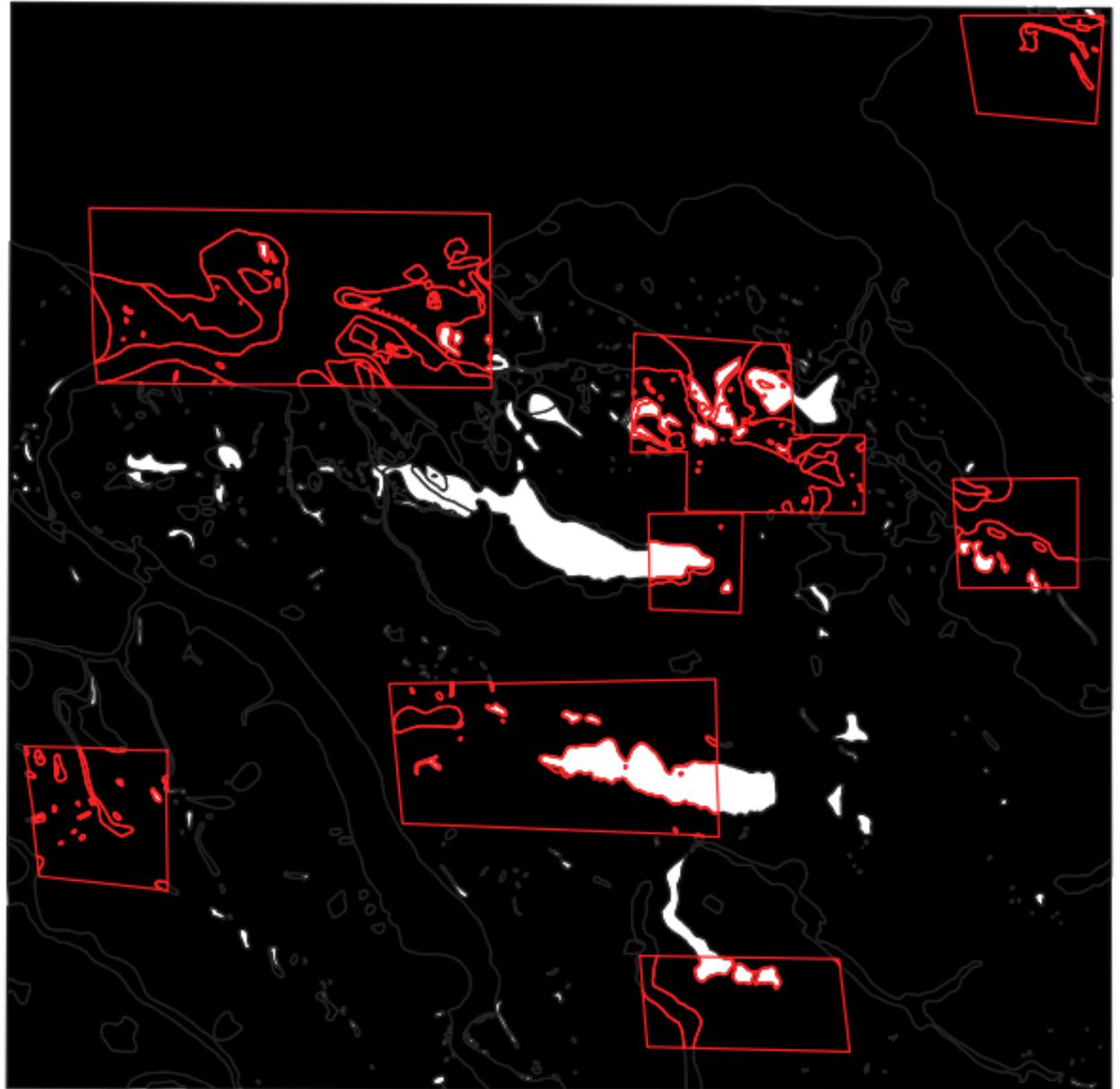
Methodology

- Random Forest is a multivariate classification, the result is from several decision trees, the overall result is the most popular one

- We created a raster with 5 bands (including aspect, curvature, height, slope and topographic wetness index - twi), generated from the DEM, and used r programming language's script for the modell. Also meaning that the „height of the trees" in the RF was 5.

- The forest population was 1000, training area was at the center containing max. 20% on the whole set.

- Final accuracy was derived from the difference of the whole digitalized maps raster form and the result raster.

Results and conclusion

- The lighter coloured pixels represent the pre-Quaternary formations, after the classification.
- Accuracy is ~85 %
- The main pre-Quaternary part is the Gete Hill at the center of the image, although some pre-Queternary predictions appear on the south-east part of it.
- Most of the Quaternary formations are from geomorphological processes, that indicates the good accuracy.
- The differences can not only be computation mistakes, but we also have to consider, that the scale of the original geological mapping was 1:10000, while the MERIT DEM has a much worse resolution. Also, the original map can have several inaccuracies especially regarding the Quaternary. Therefore field mapping, and validation is required. Furthermore, the predictions should be extended.
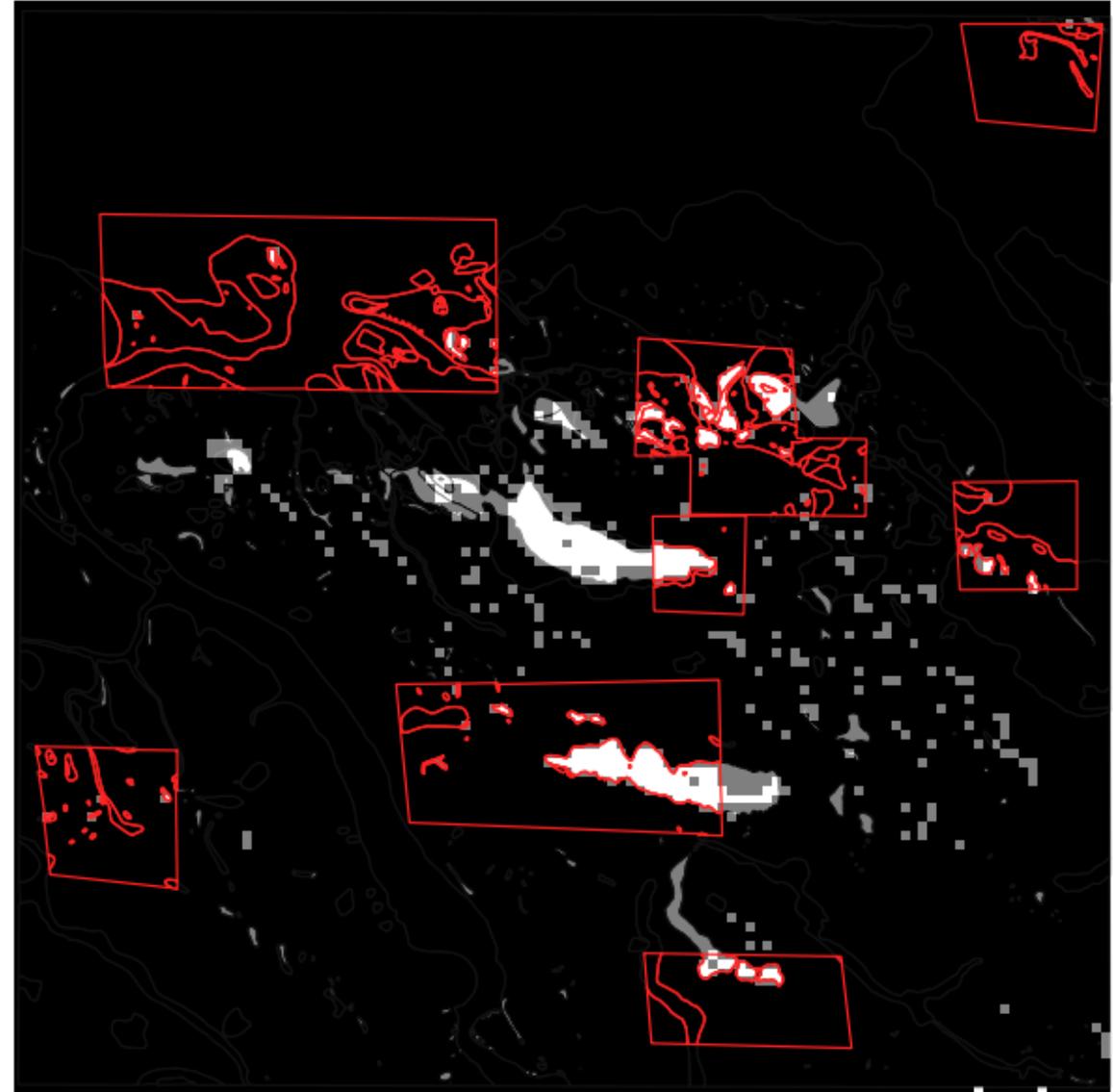
Acknowledgements

The beautiful landscape of the Gete Hill seen from the West (photo: G. Albert)