



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

Reka Pogacsas<sup>1</sup> and Gaspar Albert<sup>2</sup>

<sup>1</sup>Eötvös Loránd University, Department of Cartography and Geoinformatics, Budapest (pogireka@caesar.elte.hu)

<sup>2</sup>Eötvös Loránd University, Department of Cartography and Geoinformatics, Budapest (albert@ludens.elte.hu)

The Dorog Basin is a morphologically unique region of the Transdanubian Mountains revealing the combined work of tectonic forces and erosion. Overprinted by the forms of fluvial erosion, numerous NW-SE striking half-graben and horst structures are present. The surface is dominantly covered by loose 1–15 m thick Quaternary sediments (aeolian loess, and siliciclastic alluvial and coluvial formations), while the lithified bedrock consists of Mesozoic carbonates, Paleogene limestones, marls and sandstones and limnic coal sequences. The rheological difference of the Quaternary and pre-Quaternary formations is so pronounced that the morphological characteristics of the outcrops also differ significantly. The area was in the focus of geologists for many decades, due to its Eocene coal beds, and a renewal of the geological map of the region is in progress. The current research aims to assist the mapping with multivariate methods based on geomorphological attributes, such as slope angle, aspect, profile curvature, height, and topographic wetness index. We perform a random forest classification (RFC) using these variables, to predict the outcrops of pre-Quaternary formations in the study area.

Random forest is a powerful tool for multivariate classification that uses several decision trees, each one with a prediction, where the most popular one will be the overall result [1]. The reason why it is getting popular in spatial predictions is the high accuracy to classify raster-type objects [2]. We used raster-type spatial data as subject of RFC predicting a result for each pixel. The geology of the study area was known from previous geological mapping [3]. Morphological information was derived from the MERIT DEM.

Our model used a raster with multiple bands containing geomorphological variables, and training data from the digitalized geological map. The number of random samples of data was 2500. After testing several combinations of the bands, and several spacing of the study areas, the best prediction has cca. 80% accuracy. Model validation is based on the calculation of rates of well predicted pixels in the same rasterized geological map that was used for training. Our aim was to use exact data, which is completely true for remotely sensed images, but not for geological maps. That means the accuracy still can be improved by field perception, or from borehole data.

References:

- [1] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.
- [2] Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31.
- [3] Gidai, L., Nagy, G., & Siposs, Z. (1981). Geological map of the Dorog Basin 1: 25 000. [in Hungarian] Geological Institute of Hungary, Budapest.