ИНСТИТУТ ГЕОГРАФИИ
Российской академии наук
основан в 1918 году

Landslide susceptibility prediction by supervised Kohonen network on classic and spectral geomorphometric variables (case study of the Krasnaya Polyana resort, Russia)

Sergey Shvarev	Institute of Geography, RAS, Moscow	Institute of Earth's Physics, RAS, Moscow
Valentin Golosov	Lomonosov Moscow State University	Institute of Geography, RAS, Moscow
Sergey Kharchenko (speaker)	Lomonosov Moscow State University	Institute of Geography, RAS, Moscow

D1885 | EGU2020-20234

ИНСТИТУТ ГЕОГРАФИИ
Российской академии наук
основан в 1918 году

Introduction. **Landslide susceptibility prediction** – more inventories, more factors, more powerful techniques

- Landslide susceptibility statistical prediction – is very popular topic in geomorphology and natural hazards studies. Approximately there are 10^3 articles in the topic (only in the last review [Reichenbach, 2018] tells about 565 articles and 621 study areas).
- **Some questions** in landscape susceptibility prediction methodology:
  1. What factors are most effective for landslide prediction in the different geological and landscape features?
  2. Are equal “a factor” and “a predictor”?
  3. What objects do we predict? A scarps, a landslide’s bodies, a average landslide’s positions?
  4. Are local (pixel) variables sufficient? Do we need statistical distributions of local variables?
  5. How we can and must verify models?
  6. What is a “problem of false positive result”?

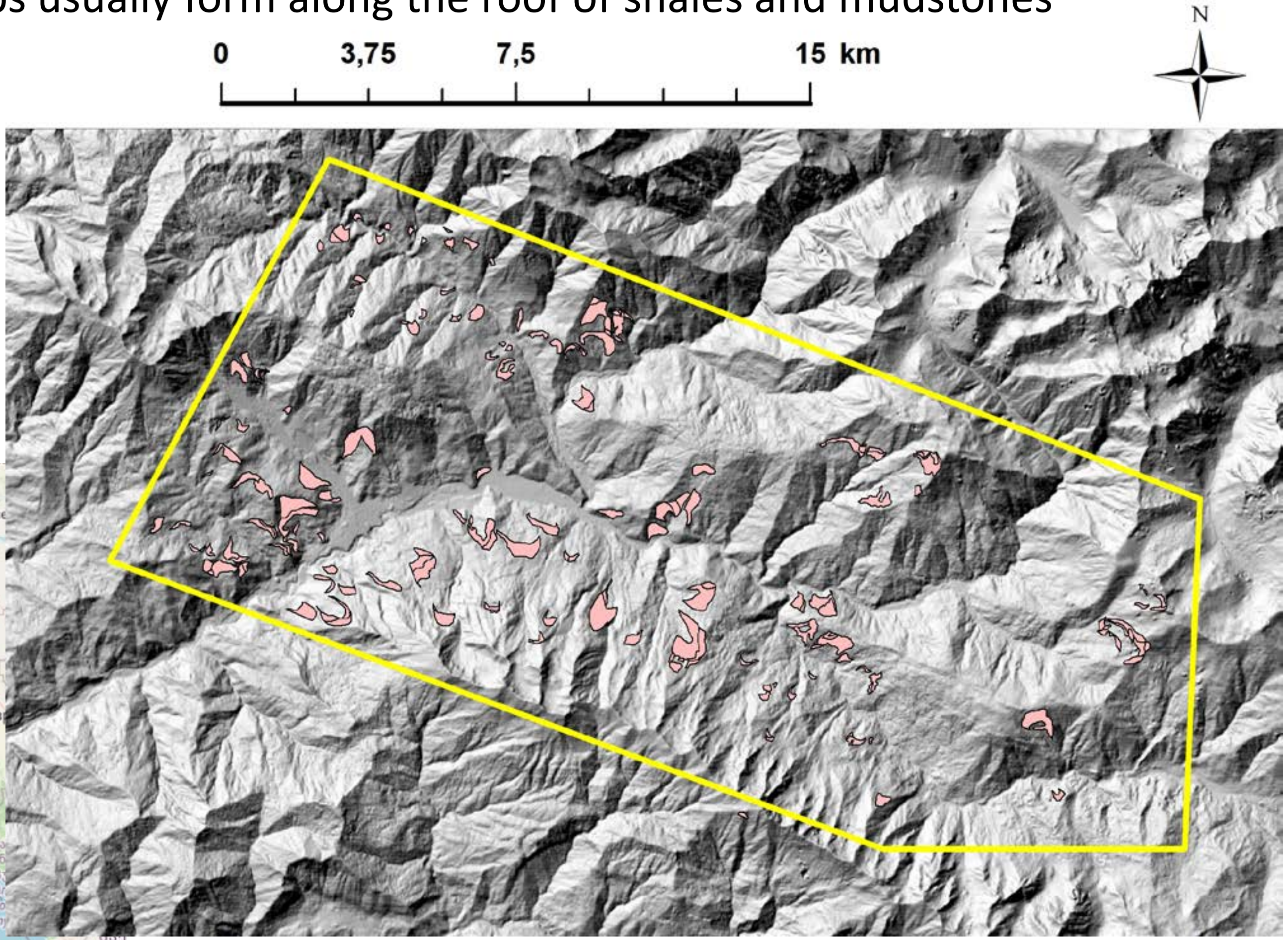
- And why aren’t any articles on landslide susceptibility prediction and mapping on the Russian territory ? ☺ (at least, on English)

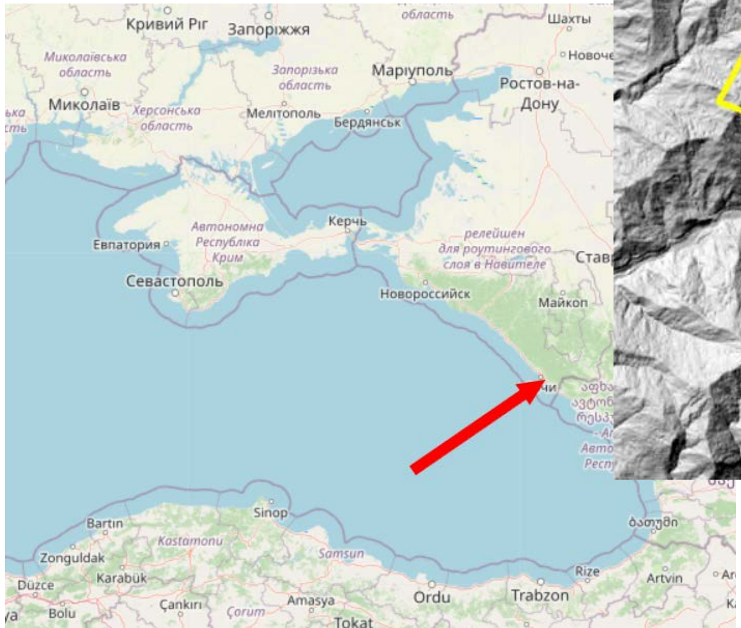
Study area

- The Krasnaya Polyana resort (Winter Olympics – 2014, mountain cluster)
- Elevation's range – 0-3335 m
- Geological features – the Lower and Middle Jurassic mudstones with thin layers of siltstones, sandstones, include pyrites and siderites, lie on the surface. Landslide escarps usually form along the roof of shales and mudstones

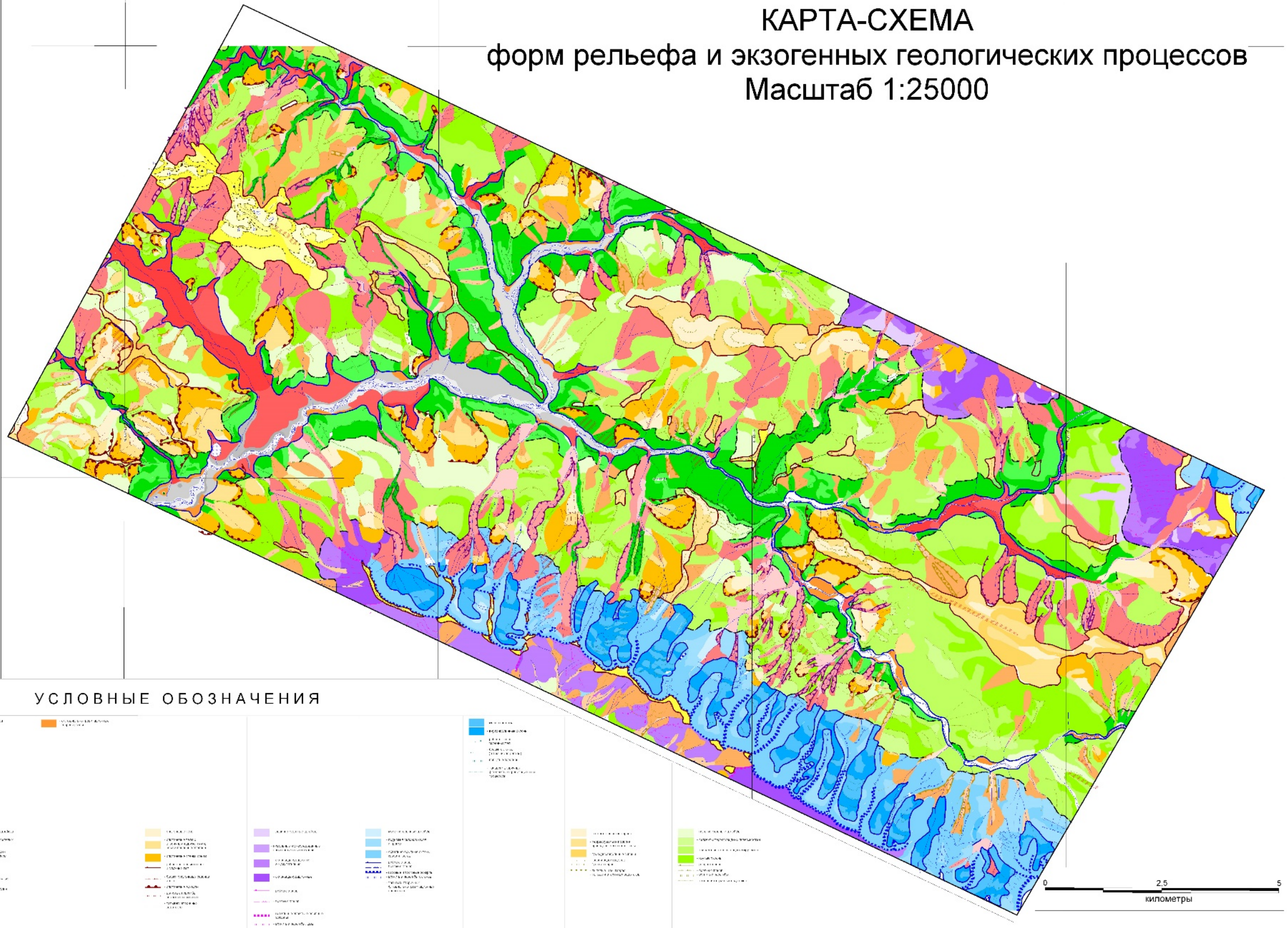
- Climate: subtropical wet
- Landscapes: deciduous forests, subalpine and alpine meadows
- Landforms: middle and high alpine mountains. Slopes modified by fluvial, slope and strong anthropogenic processes (and glacial processes in the past).

Source DEM:
Alos W3D
(resampled to 20 m)



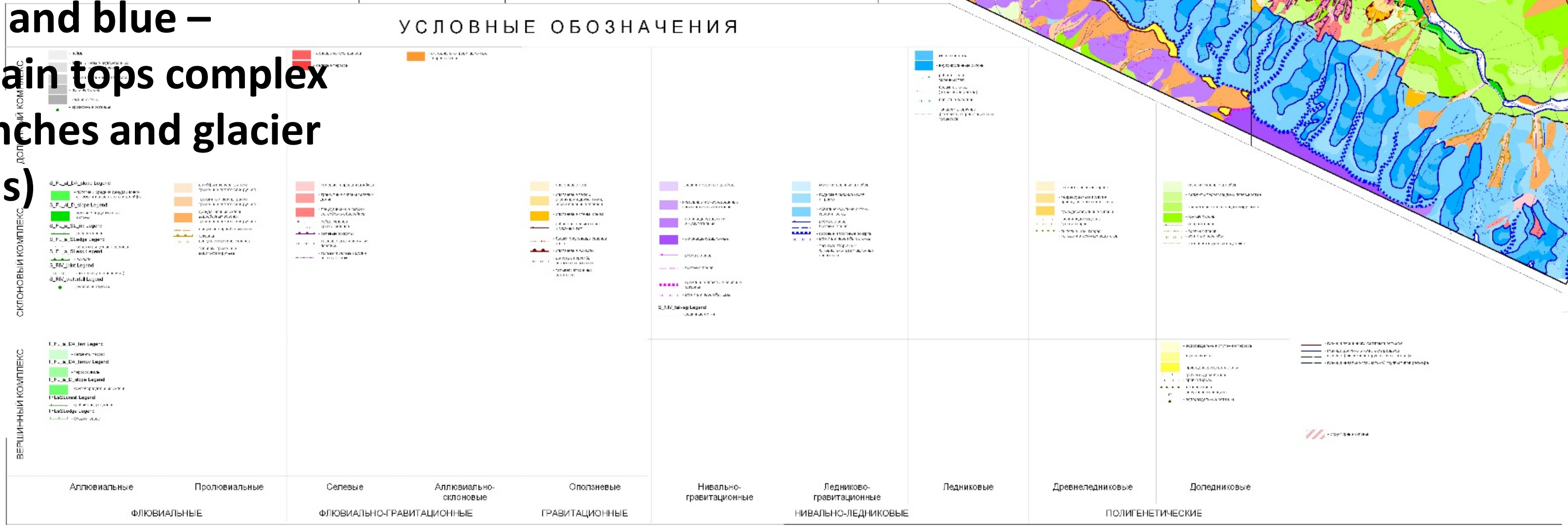


Overview geomorphological map created during Olympics-2014 preparation (by S.V.Shvarev, 2008, in Russian)



Legend (simplified)

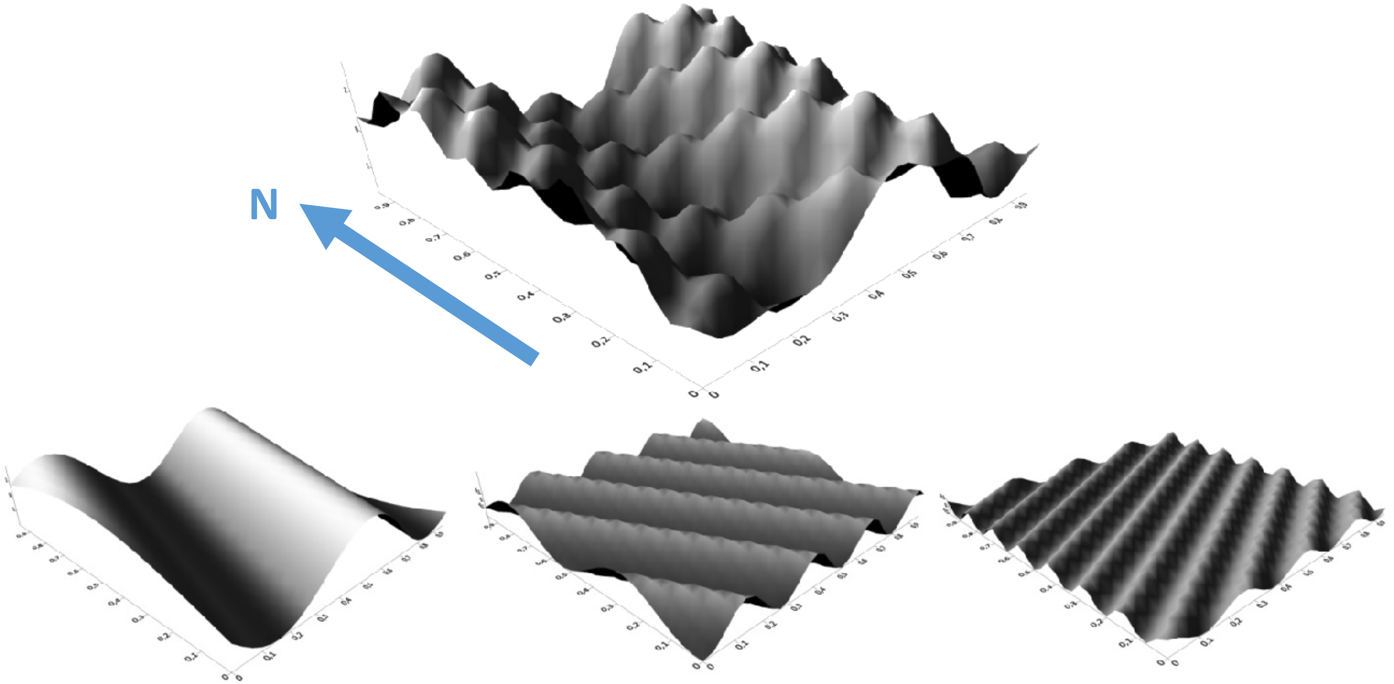
1) **Green – slope complex**
2) **Grey – valley's bottom complex**
3) **Red – debris flows complex**
4) **Brick – ephemeral waterway's complex**
5) **Orange – landslide's complex**
6) **Purple and blue – mountain tops complex (avalanches and glacier terrains)**



Question No. 1. **What factors are most effective for landslide prediction in the different geological and landscape features?**

- **We don't know what the factors are most effective for prediction**.
- Landslide danger affected by slope, aspect, distance to active faults, roads and waterways, vegetation type. If we use only standard parameters – **we risk to miss a hidden patterns**.
- **We need to test** a maximum number of geological, geomorphological, landscape characteristics for its predictive power's estimation.
- In this study we computed **103 classical geomorphometric variables** by SAGA GIS: hydrological (drainage), climatological, visibility, general geomorphometric and slope stability variables (only continuous , except different binary indices). We don't use geological (lithology, distance to faults and more), landscape (landuse, vegetation) or other variables.
- The **spectral geomorphometric variables** (8 variables in 6 scales each = 48) **were tested for predictive power**. This variables characterize pattern of surface dissection. Similar values of the variables reflect similar (visually) topographic texture.

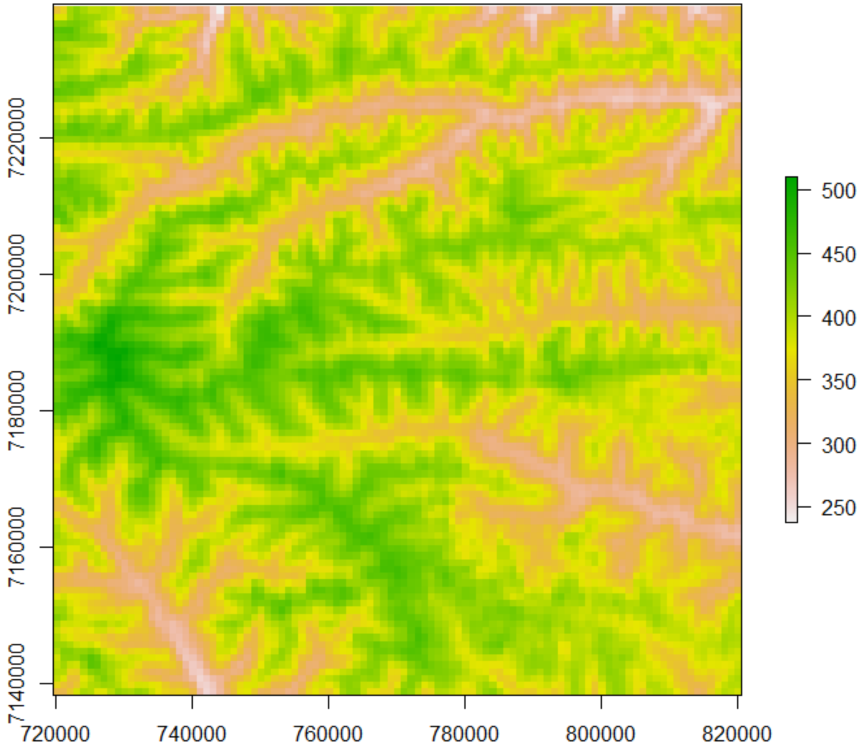
Artificial DEM decomposition into three components



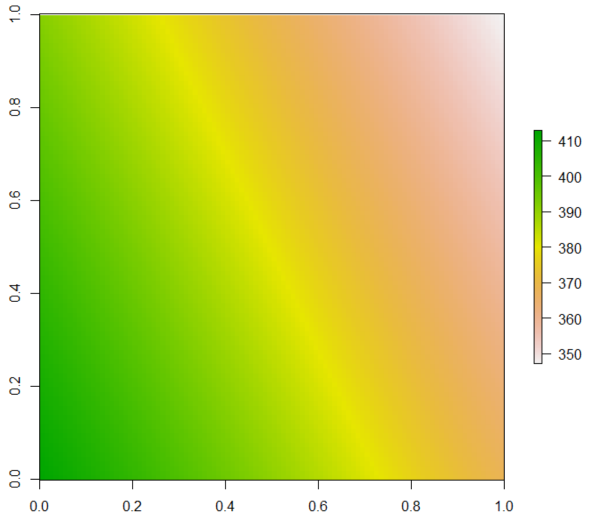
ИНСТИТУТ ГЕОГРАФИИ
Российской академии наук
основан в 1918 году

Spectral geomorphometric variables – SGV (from DEM decomposition into Fourier-series)

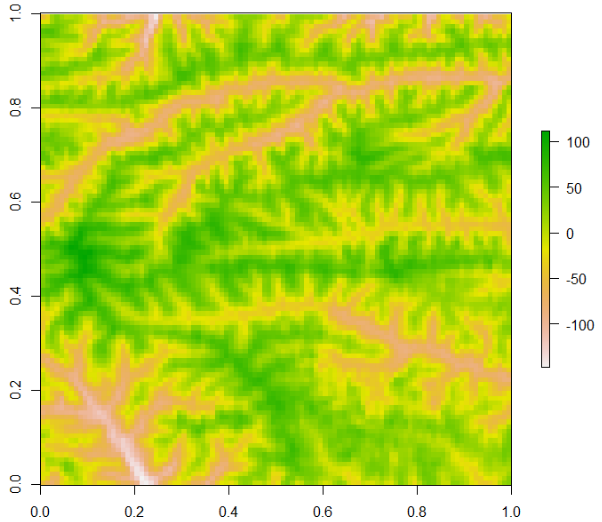
Source DEM



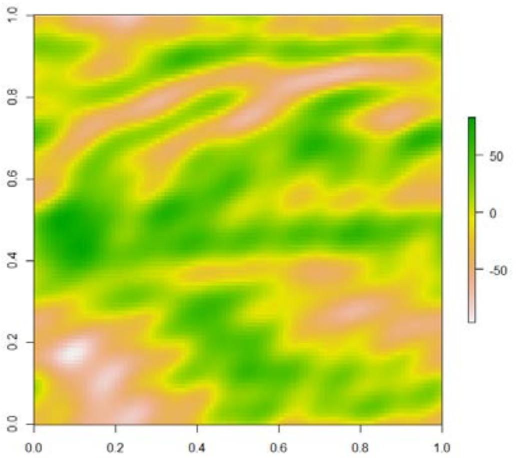
1) Fitting the surface by 2D linear trend

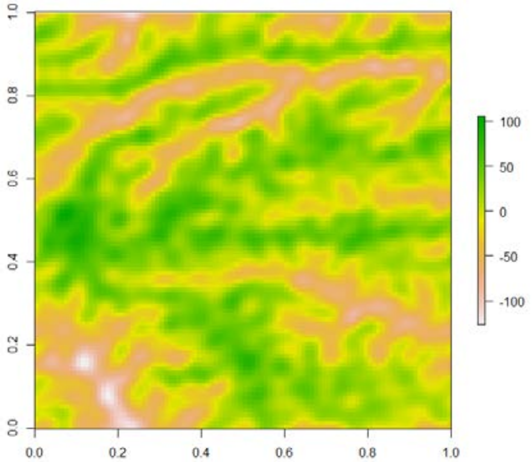


2) DEM detrending



3) Demonstration of reconstructed DEMs on the 5%, 20% shares of all waves





SGV describe elevation fluctuations:

1) Its magnitudes, 2) direction, 3) density and others patterns.

For more information about SGV see:

https://presentations.copernicus.org/EGU2020/EGU2020-19509_presentation.pdf

Methodology

1) DEM preprocessing and computing the classic and the spectral geomorphometric variables. Field-based landslide inventory (points or polygons landslides localization).

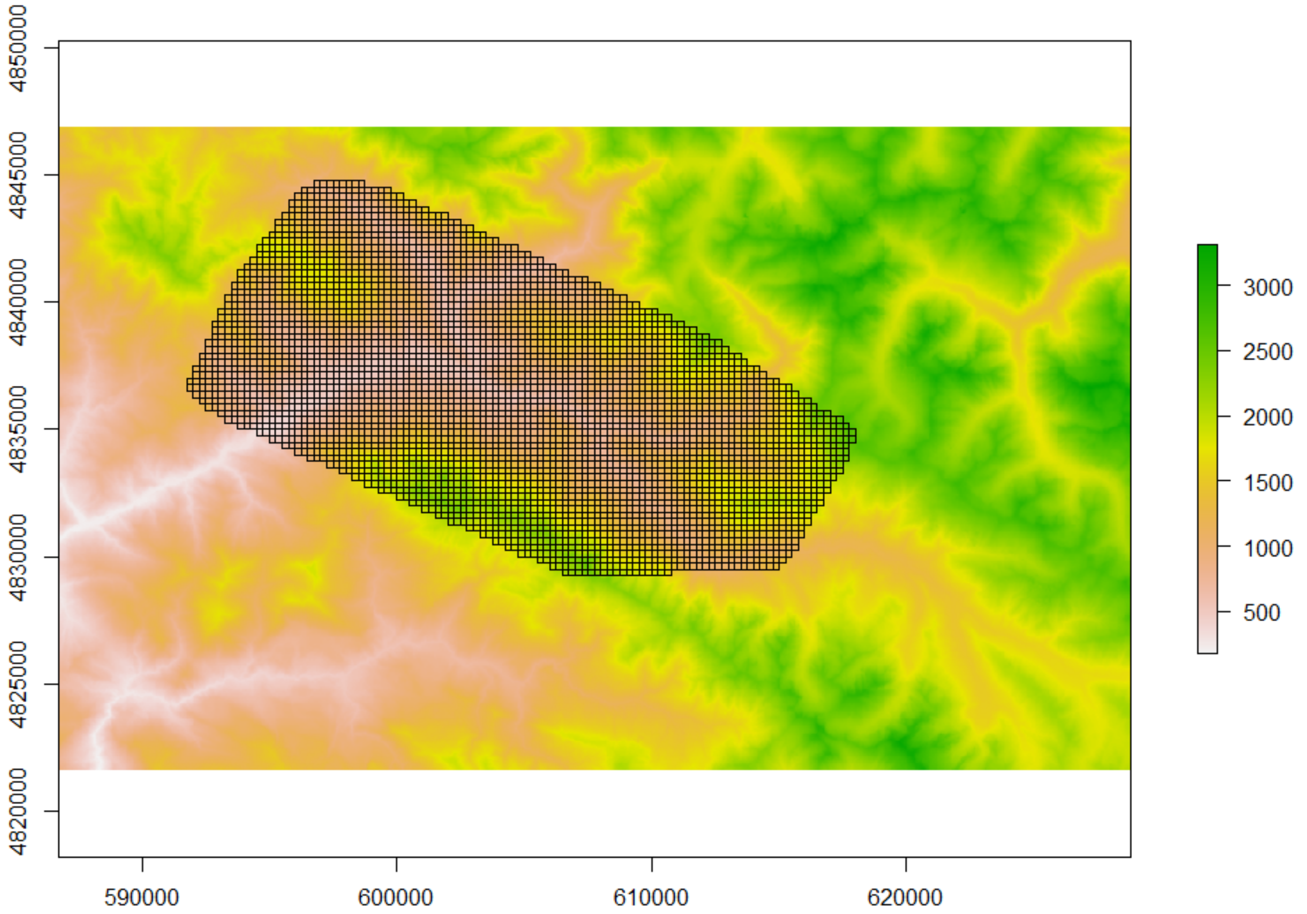
2) Defining the computational extents (network for averaging of the local variable). Computing of the statistics (mean, median, kurtosis, IQR and others) within network cells for all geomorphometric variables.

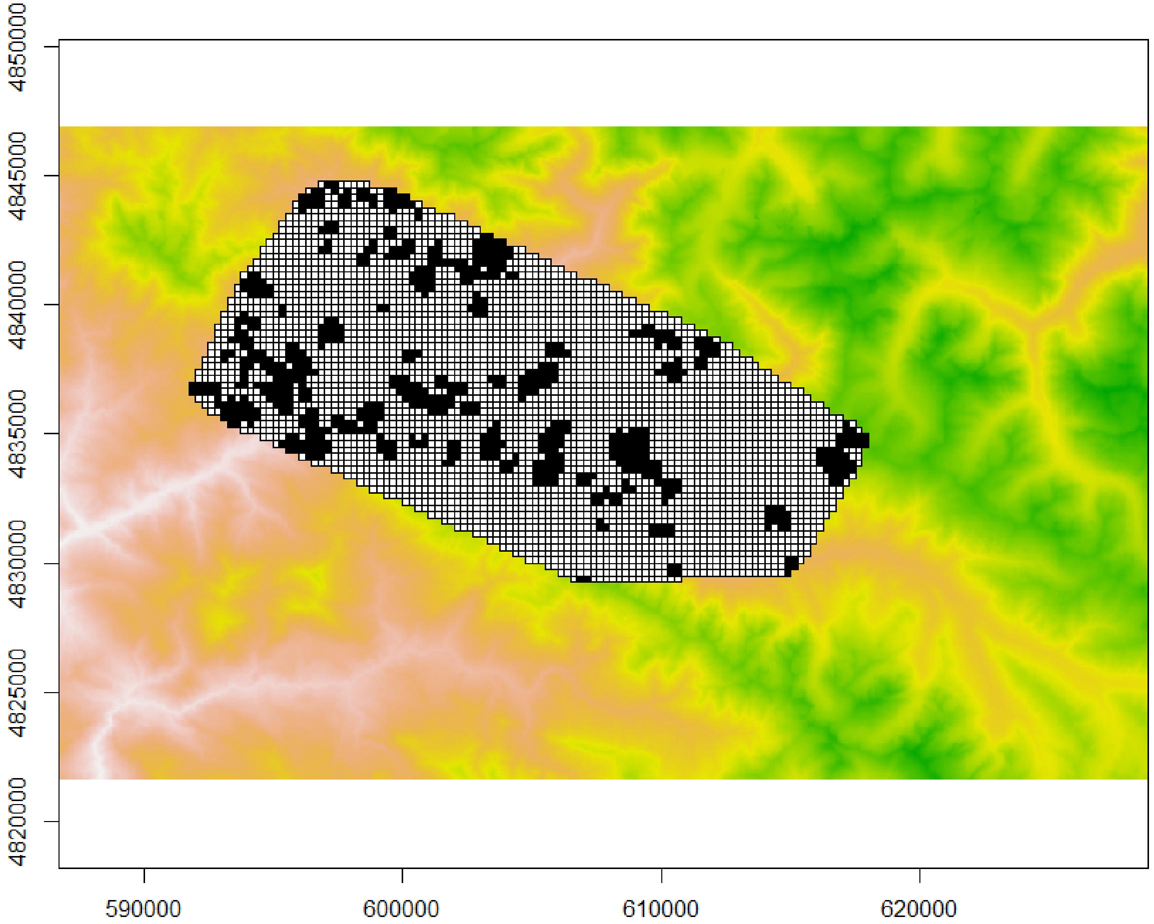
3) Preprocessing data (normalization with Box-Cox test, correspondence of the landslides and network cells, balancing of the unbalanced classes – landslide / non-landslide for model learning).

4) Multiple random variable's selection, model construction (supervised Kohonen's SOM) on the train sample and testing on the test sample. Choosing optimal variable's set and creation the final model.

5) Predicting and mapping landslide susceptibility on the study area.

Prediction for network cells. Each cell size (250 m) ≈ 10*10 DEM cells





White cells – non-landslide, black cells - landslide

Statistics computed over each geomorphometric variables: mean, min, max, IQR, skewness, kurtosis, sd

	mean_10_a_exp.tif	mean_10_dev_exp.tif	mean_10_l_exp.tif	mean_10_max.imp.tif	mean_10_max.mag.tif	mean_10_pr.deg.tif	mean_10_wavelen.tif	mean_12_a
1	13.015820	8.8185509	0.7527560	0.4037399	36.889603	0.2257363	195.8053	13.268280
2	11.077057	5.5004359	0.6989291	0.3552244	26.346991	0.2947961	184.9921	14.917460
3	6.269724	1.3759984	0.7261721	0.3278228	14.868256	0.1658772	197.7983	7.983963

Questions No. 5 and 6. How we can and must verify models? What is a "problem of false positive result"?

- **The problem of the class imbalance.** If our training sample have class imbalance and if we use "accuracy" metric for model quality assessment, we run the risk of obtaining fantastic accuracy by predicting a zero (major, non-landslide class).
- Standard metric for model quality assessment is accuracy* (objects recognized as true positive + true negative / number of all object). In case of landslides, the false positive cells may be potentially dangerous. But training process can give us the model for **excellent ignore potentially dangerous cells**!
- Random example on SGV data:

	Fact 0 (non landslides)	Fact 1 (landslides)
Pred 0	1563	371
Pred 1	7	1

Accuracy: 80.5%, but **we successfully recognized only 0.27% of landslide cells. High accuracy is consequence of major class perfect prediction.**

- So, **maybe is sensitivity the good choose** for quality measurement. It's ratio: true positive objects / all positive objects. In the our case sensitivity is 0.27%. High values of sensitivity we can combine with high accuracy, but **training target - must be to maximization the sensitivity** and not the accuracy. Without imbalance problem solving we can't get the effective predictive model. **"False positive" results – is not problem** and key for understanding future of the landslides on the studied area.

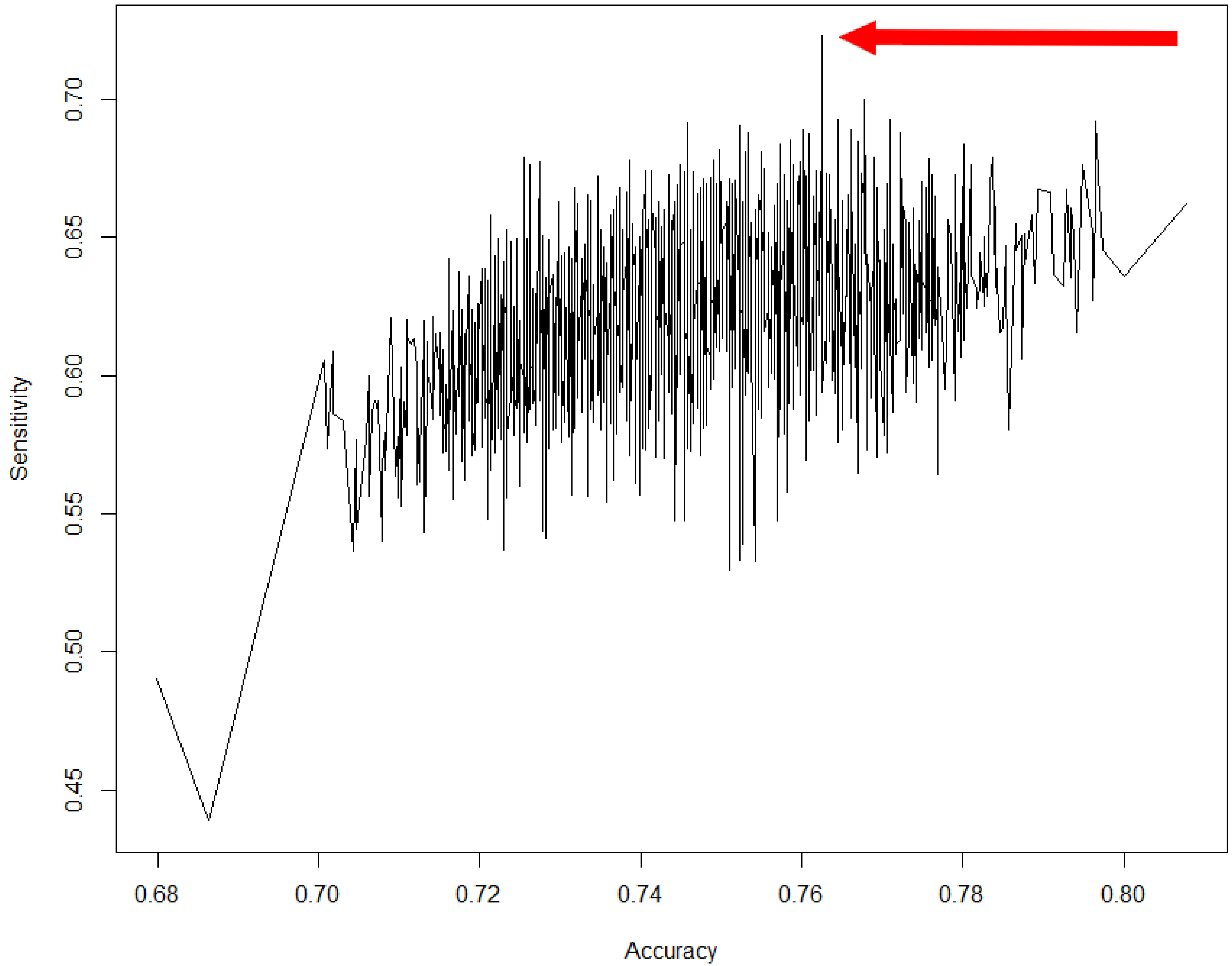
* - and ROC-curve, AUC criteria and others

Choosing optimal variable’s set

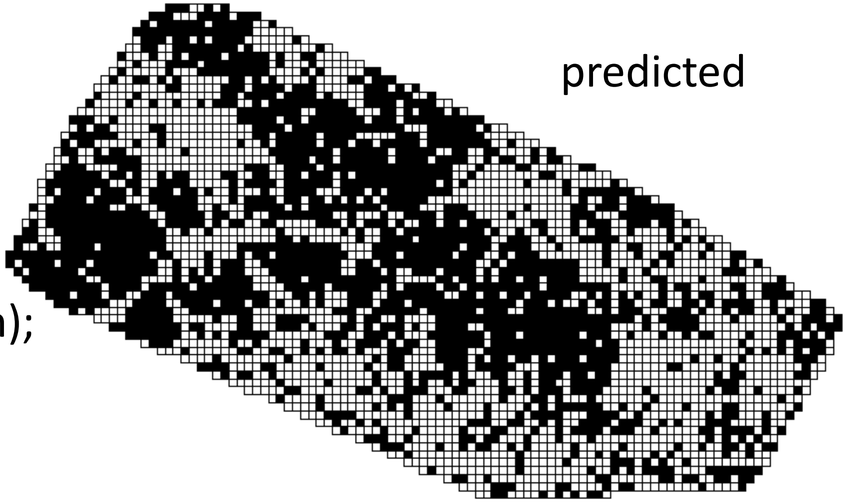
1. Recursive algorithms;
2. Genetic algorithms;
3. **Multiple random choosing and validation.**

- At the lot of variables the very big computational resources are necessary for algorithms from group 1 and 2. Random choosing (30 000 times) isn’t give the most effectiveness combination, but allows to get good results in a little time.

Firstly – maximization of sensitivity,
secondly – maximization of accuracy



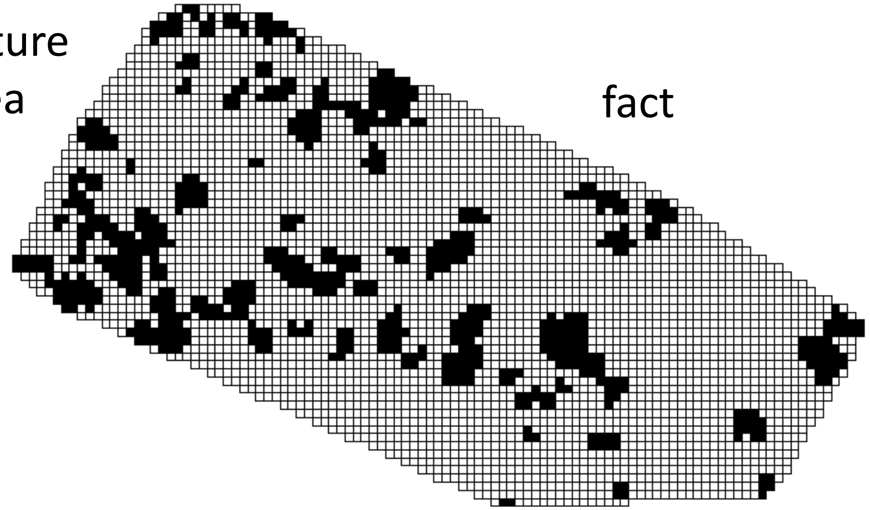
Classical geomorphometric variables (CGV)

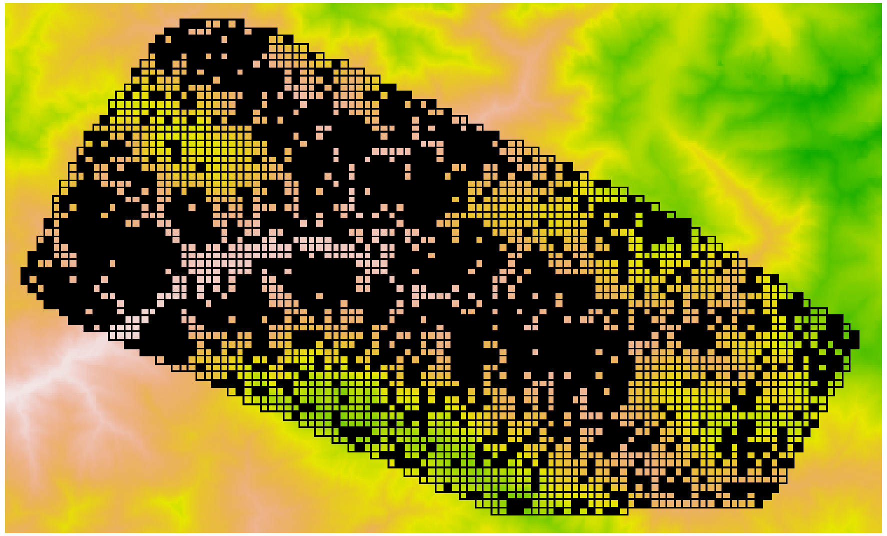


Optimal variables:

IQR, min, max, skewness and kurtosis – statistics within network cells (250*250 m);

- skewness_Topographic Wetness Index
- IQR_Convexity
- min_Convexity
- max_Maximum Height
- max_Local Downslope Curvature
- kurtosis_Total Catchment Area



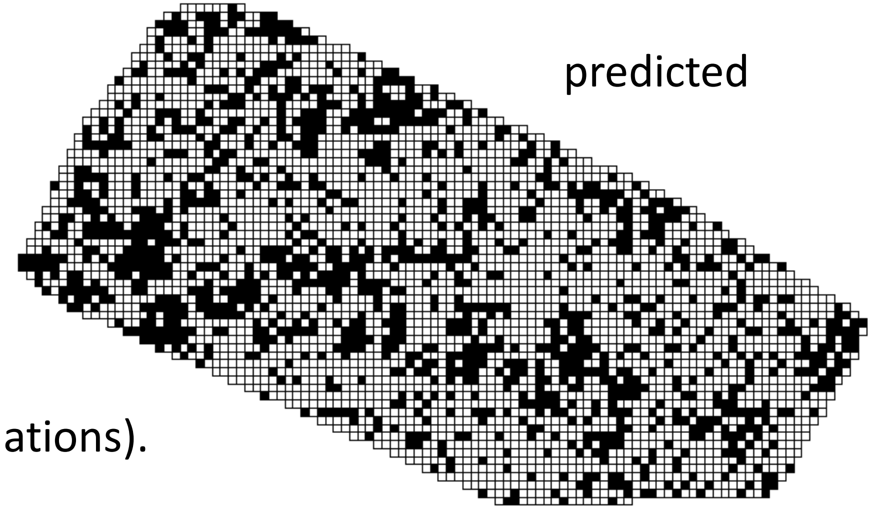


Balancing method:
undersampling

	Fact 0 (NLS)	Fact 1 (LS)
Pred 0	205	96
Pred 1	167	247

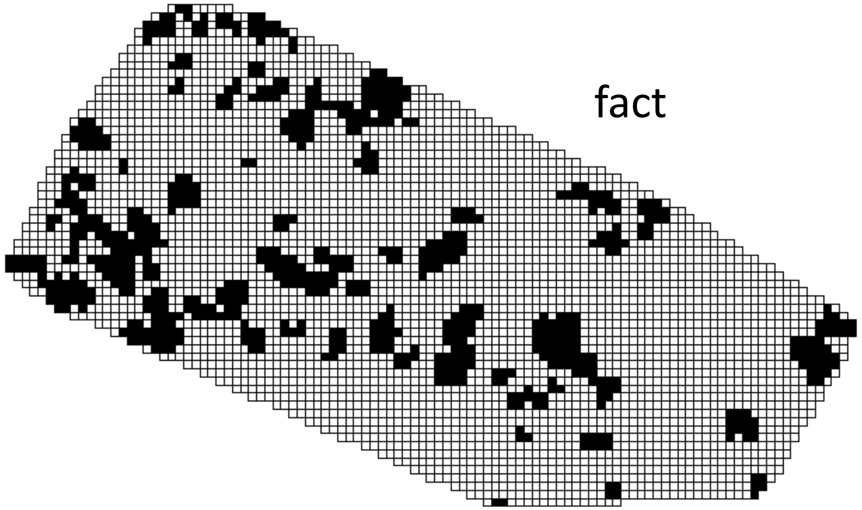
Sensitivity = 72%
Accuracy = 66.5%

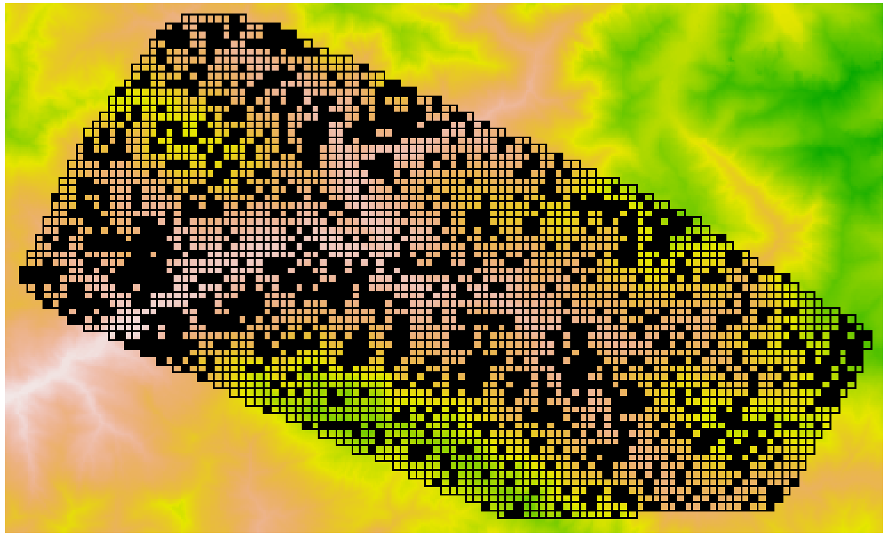
Spectral geomorphometric variables (SGV)



Optimal variables:
IQR, min, max, skewness and sd – statistics within network cells (250*250 m);
12, 14, 18, 20 – window sizes in cells of source DEM (1 cell – 20*20 m);
RD, A0, max.imp, wavelen, pr.dir (cos of general direction of the elevation's fluctuations).

- **IQR_20_RD**
- **min_18_A0**
- **skewness_14_max.imp**
- **sd_12_wavelen**
- **skewness_12_max.imp**
- **max_12_pr.dir**





Balancing method:
SMOTE (synthetic minority over-sampling technique)

	Fact 0 (NLS)	Fact 1 (LS)
Pred 0	1141	294
Pred 1	300	767

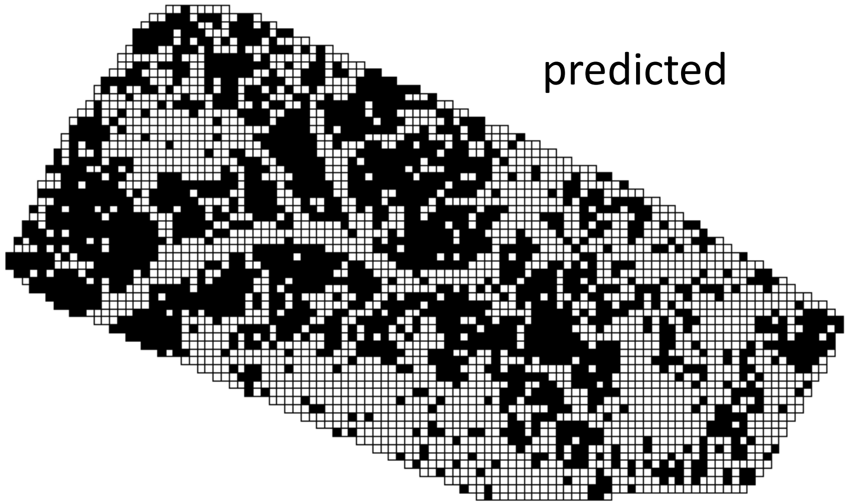
Sensitivity = 73%
Accuracy = 76.5%

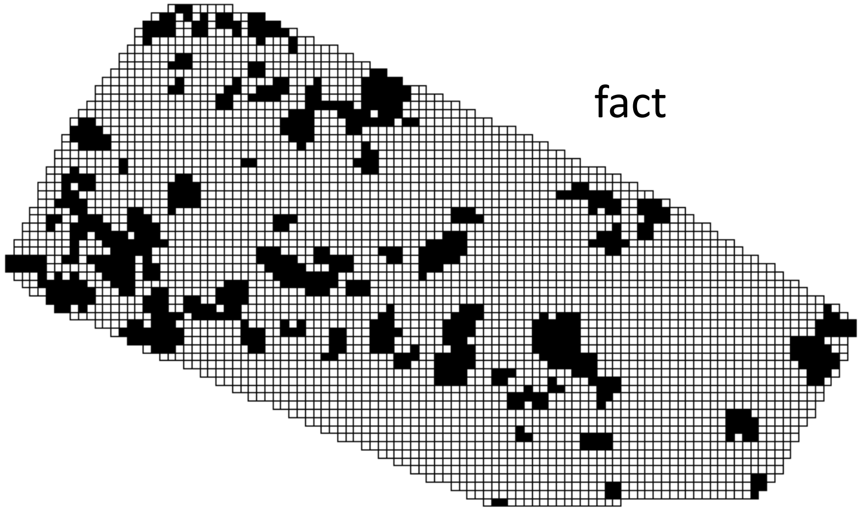
For more information about SGV see:
https://presentations.copernicus.org/EGU2020/EGU2020-19509_presentation.pdf

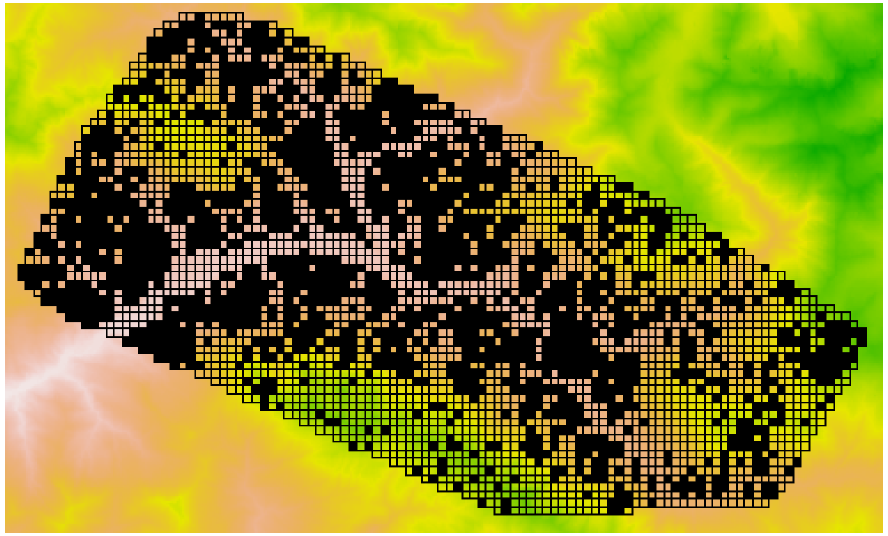
All geomorphometric variables

Optimal variables:
min, max and kurtosis – statistics within network cells (250*250 m);

- min_14_wavelen
- max_Aspect [Mean of Upslope Area]
- max_Maximum Height
- max_Topographic Wetness Index
- kurtosis_Protection Index
- min_14_pr.dir







Balancing method:
undersampling

	Fact 0 (NLS)	Fact 1 (LS)
Pred 0	213	99
Pred 1	152	251

Sensitivity = 72%
Accuracy = 65%

ИНСТИТУТ ГЕОГРАФИИ
Российской академии наук
основан в 1918 году

Conclusions

- The problems of landslide susceptibility prediction were summarized as a series of questions.
- The spectral geomorphometric variables proposed using in landslide susceptibility models.
- The methods of subsetting of the optimal variables set are very important. It is possible to use random sampling method.
- The model of landslide susceptibility prediction for the Krasnaya Polyana resort (Krasnodarskiy Kray, Russia) was created by three ways. Quality (“the sensitivity” and “the accuracy”) at the using classical variables only was 72/66.5%; at the using spectral variables only was 73/76.5%; at the using all variables 72/65%.

This study was funded by the Russian Science Foundation, project no. 19-17-00181.