Altitude as an indicator of biased sampling design in landslide susceptibility mapping ?

Andrei Dornik, Lucian Drăguț, Takashi Oguchi, Yuichi Hayakawa, Mihai Micu

EGU2020: Sharing Geoscience Online

OBJECTIVES

Main goal: to evaluate the sensitivity of landslide susceptibility modelling (LSM) to sampling in heterogeneous areas.

Hypotheses:

1. random sampling of the landslide absence data in such areas can lead to erroneous distributions of predictors.
2. altitude might points out issues in sampling design when appears as a main predictor in landslide modeling.

STUDY AREAS

Two study areas, with landslide inventories available:

1. In the Buzău County, Romania
2. In the Shizuoka Prefecture, Japan

Each study area was split in **distinct domains**, defined by **geological/ lithological** units that reflect homogeneous topographies.

The tests were conducted as following:

1. Within entire study area – **lithologically heterogeneous** (named B and J)

2. Within each **lithologically homogeneous domain** (named B1, B2, B3 and J1, J2)

METHODS

We train a **Random Forest (RF)** model with 14 terrain variables as predictors, in two sampling strategies:

1) **one random point** allocation within each landslide scarp polygon and the same number of points randomly created outside landslide scarp area, as absence data
2) Random sampling of **representative number of points** within landslide scarps (at least one point per scarp) and the same number of points randomly created outside landslide scarps
    - Representative number of samples is 10000 in Buzau and 6000 in Shizuoka, 50 % as presence and 50 % as absence.

The results are compared both quantitatively and relative to their geomorphic plausibility.

RESULTS

1) **one random point** allocation within each landslide scarp polygon

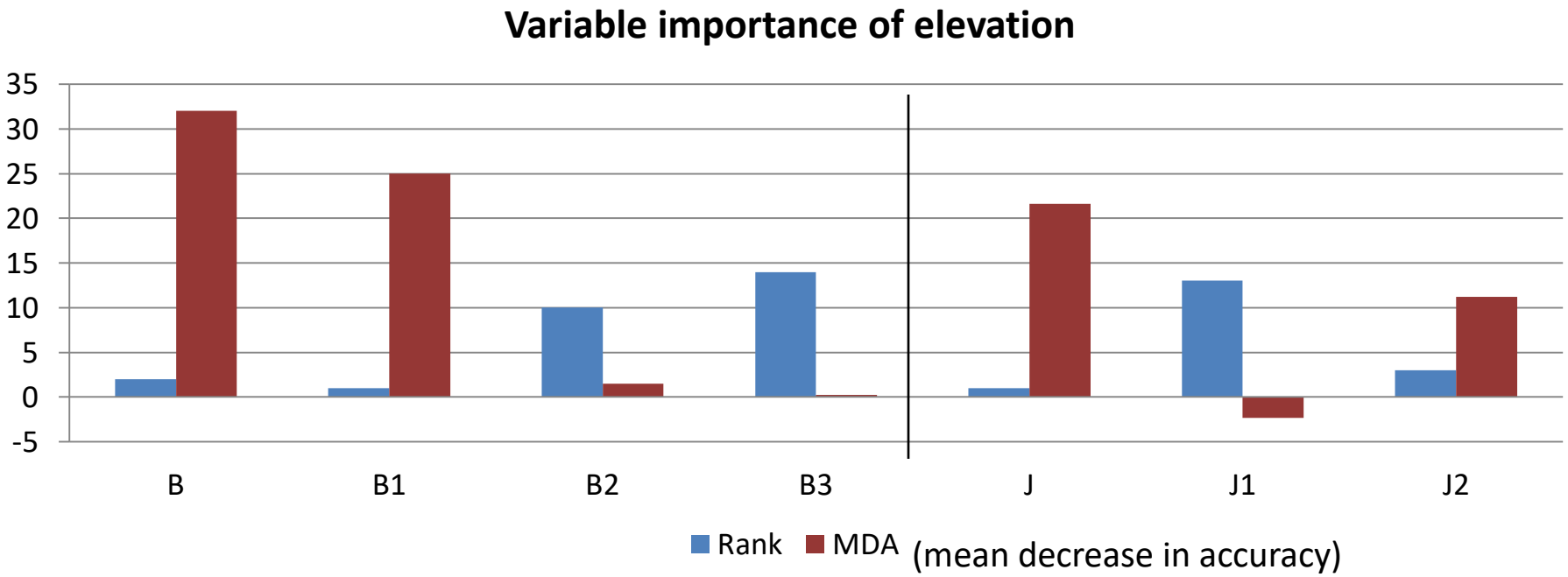
Study area	Number of samples	OOB	AUC	OA	Kappa
B	1154	0.30	0.77	0.73	0.45
B1	204	0.26	0.81	0.77	0.54
B2	220	0.36	0.74	0.70	0.39
B3	220	0.30	0.78	0.75	0.50
J	714	0.29	0.78	0.72	0.44
J1	220	0.32	0.84	0.75	0.50
J2	220	0.29	0.82	0.77	0.54

OOB - Out-of-bag error; AUC - Area Under The Curve; OA – overall accuracy; Kappa – kappa index of agreement

The results show that in four out of five cases the **lithologically stratified random sampling significantly improved the prediction**.

RESULTS

1) **one random point** allocation within each landslide scarp polygon



- variable importance analysis show that variable hierarchy changed significantly when using lithological stratified sampling
- Elevation decreased in importance in lithological domains
- **very high variation observed as repeating the procedure (100 times)**
    - Low confidence
    - Unrepresentative number of samples
- This result led to the second sampling strategy

RESULTS

2) Random sampling of **representative number of points**

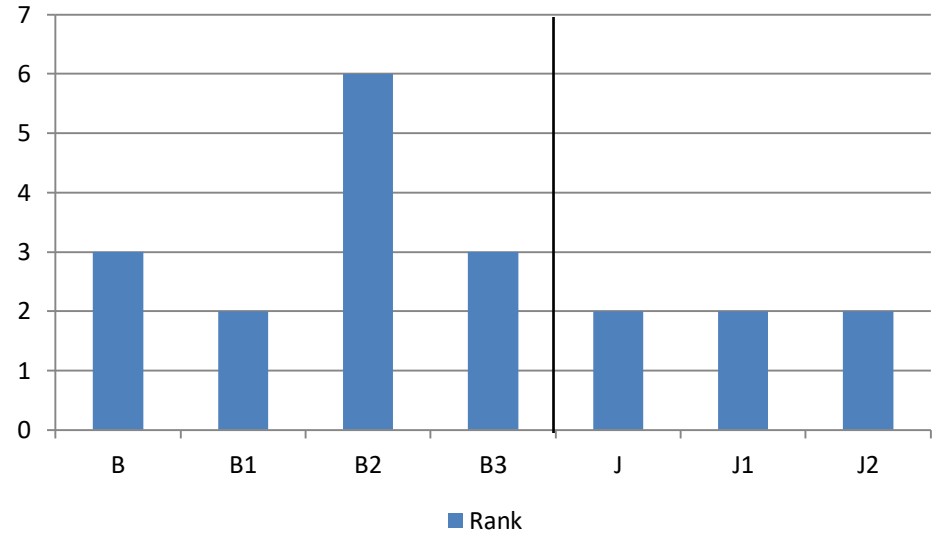
Study area	Number of samples	OOB	AUC	OA	Kappa
B	10000	0.26	0.82	0.75	0.49
B1	10000	0.20	0.90	0.81	0.62
B2	10000	0.26	0.83	0.75	0.50
B3	10000	0.18	0.91	0.83	0.66
J	6000	0.19	0.89	0.81	0.63
J1	6000	0.23	0.87	0.79	0.57
J2	6000	0.14	0.94	0.87	0.73

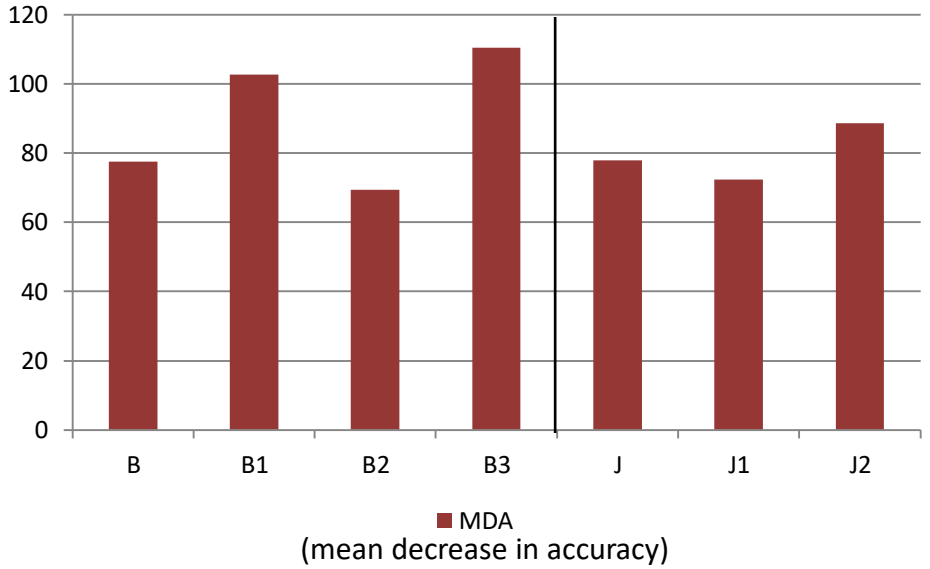
OOB - Out-of-bag error; AUC - Area Under The Curve; OA – overall accuracy; Kappa – kappa index of agreement

- in four out of five cases the **lithologically stratified random sampling of representative number of points improved the prediction**.

RESULTS

2) Random sampling of **representative number of points**





- variable hierarchy changed significantly when using lithological stratified sampling
- Elevation remains an important predictor, even more important in B1
- **Very low variation** was observed as repeating the procedure (20 times)
    - High confidence

CONCLUSIONS

- Landslide modeling is highly sensitive to sampling the absence data
- Accuracy measures improved when sampling in lithologically homogeneous domains, as compared to heterogeneous areas
- Altitude was still an important predictor although its geomorphic plausibility is questionable
- One point/scarp and equal number of absences is not recommended