CONSTANTINE THE PHILOSOPHER UNIVERSITY IN NITRA FACULTY OF NATURAL SCIENCES AND INFORMATICS

GeoKarto 2026
17 – 18 September 2026, Nitra

INSTITUTE OF GEOGRAPHY SLOVAK ACADEMY OF SCIENCES

PREDICTION OF FLUVIAL FLOOD HAZARD USING MACHINE LEARNING TRAINED ON HYDRAULICALLY DERIVED DATA

Authors:
Matej Vojtek[1,2], Dávid Držík[3], Jana Vojteková[1]

[1] Department of Geography, Geoinformatics and Regional Development, Constantine the Philosopher University in Nitra

[2] Institute of Geography, Slovak Academy of Sciences

[3] Department of Informatics, Constantine the Philosopher University in Nitra

Funded by the European Union
NextGenerationEU

RECOVERY AND RESILIENCE PLAN

INTRODUCTION

- Fluvial flooding → flash floods or riverine floods.
- Traditional fluvial flood mapping approach → hydraulic modeling → time consuming.
- Official flood maps (EU Floods Directive 2007) → only for critical river sections defined in the Preliminary Flood Risk Assessment (PFRA).
- Research question: How accurately can machine learning emulate hydraulically modeled fluvial flood extent and flow depth, and can it also reduce computational time requirements?

AIM OF THE STUDY

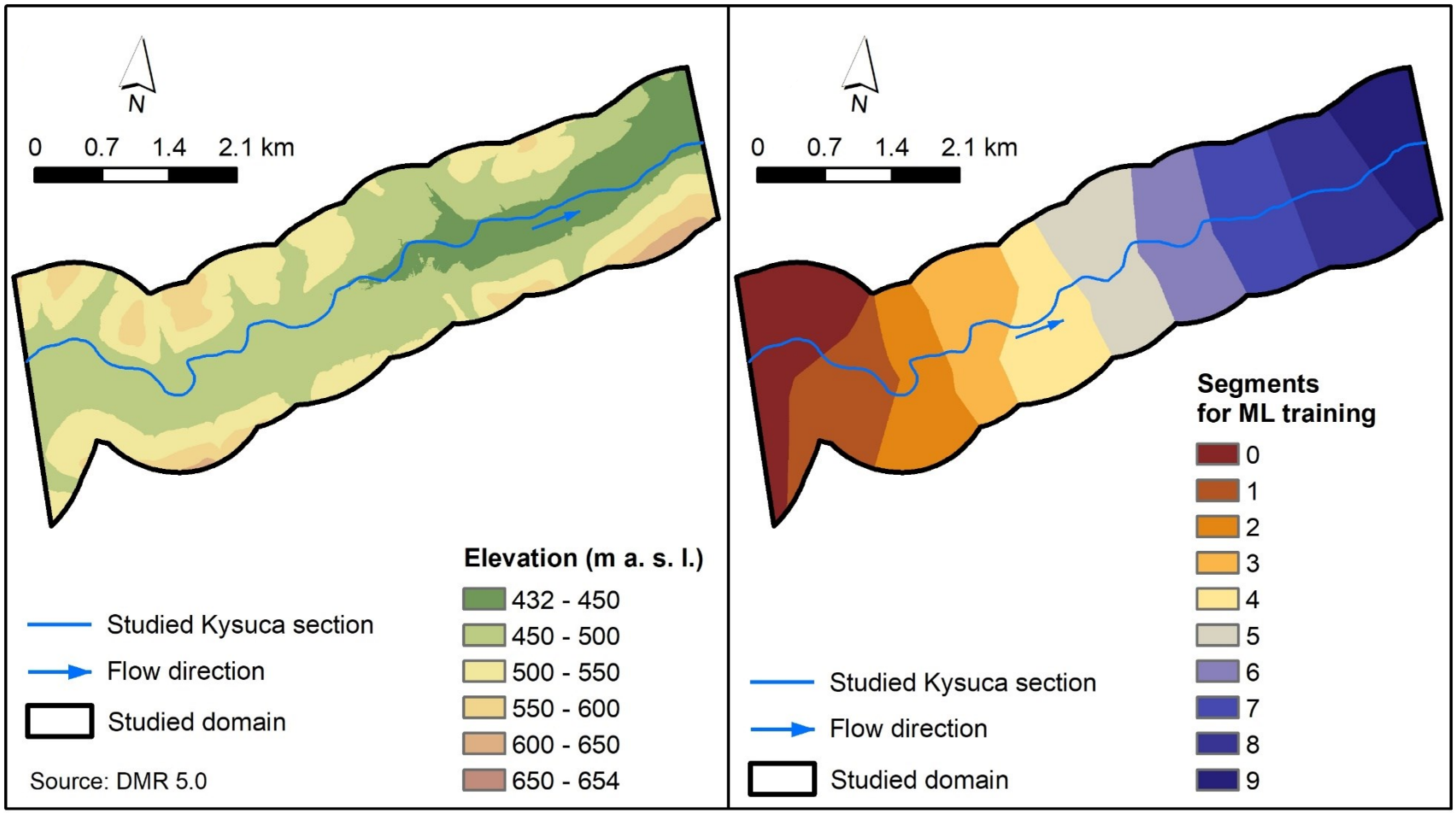
- Prediction of fluvial flood extent and depth based on machine learning (ML) eXtreme Gradient Boosting (XGBoost) model, hydraulically modeled flood maps used for training the ML-RF models, and seven predictors processed via GIS.

- Studied flood scenarios: Q_{1000}, Q_{100}, Q_{50}, Q_{20}, and Q_{10}.

- River segmentation approach (10 segments) was used and five combinations of training and testing segments 9/1, 8/2, 7/3, 6/4, and 5/5 were investigated.

STUDY AREA

- Kysuca River section (8.8 km); critical river section; bed slope = 0.004 m/m; 600 m buffer from river; studied domain = 13.6 km^2



DATA AND METHODS

- Predictors: slope, stream power index (SPI), topographic wetness index (TWI), height above the nearest drainage (HAND), distance from river, roughness, normalized difference vegetation index (NDVI) = 1 m resolution.

- Multicollinearity analyzed via Pearson correlation and Variance Inflation Factor (VIF) methods. Pearson below 0.7 and VIF below 5 were thresholds.

- Training/testing of ML models: flow depth maps for Q_{1000}, Q_{100}, Q_{50}, Q_{20}, and Q_{10} created by the 2D MIKE+ hydraulic modeling.

- 4 variants of training data splits: 9/1, 8/2, 7/3, 6/4, 5/5.

DATA AND METHODS

- XGBoost model: n_estimators = 500 class./600 regress., max_depth = 5, learn_rate = 0.01, min_child_weight = 50, subsample = 0.85, colsample_bytree = 0.85, gamma = 0.5.

- Performance of ML-RF regression assessed via RMSE, MAE, R^2, and Bias.

- Performance of ML-RF classification assessed via Recall, Precision, and F1-score metrics:

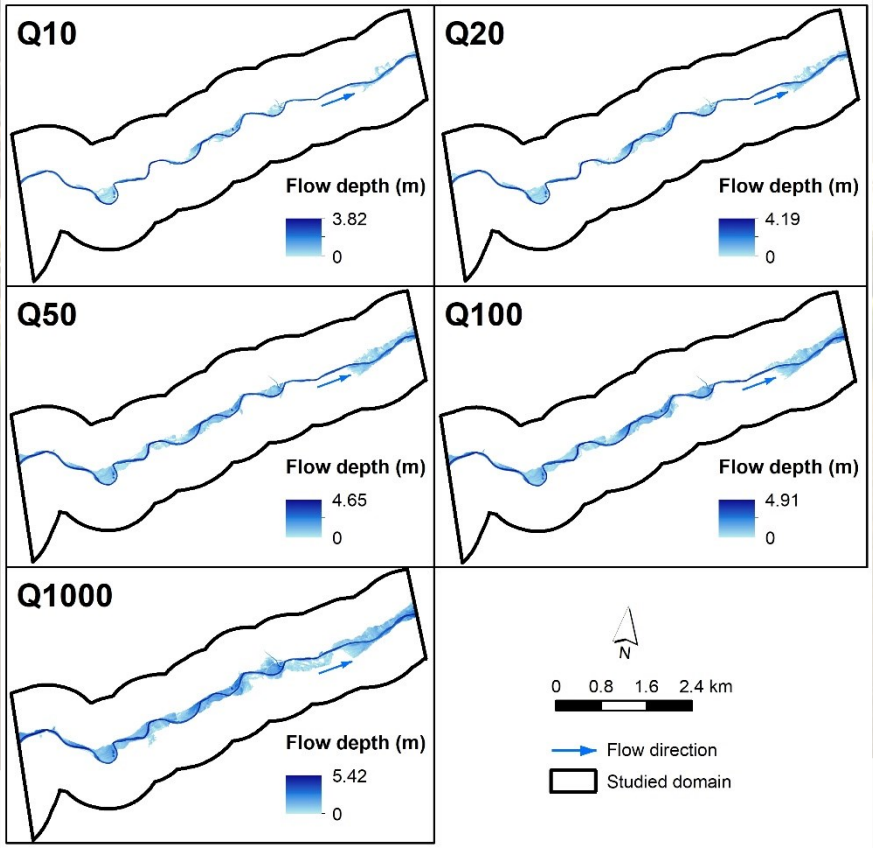
$$\text{Recall} = \frac{TP}{\text{TP} + \text{FN}} \qquad \text{Precision} = \frac{TP}{\text{TP} + \text{FP}}$$

$$\text{F1} - \text{score} = \frac{2 * Precision * Recall}{\text{Precision} + \text{Recall}}$$

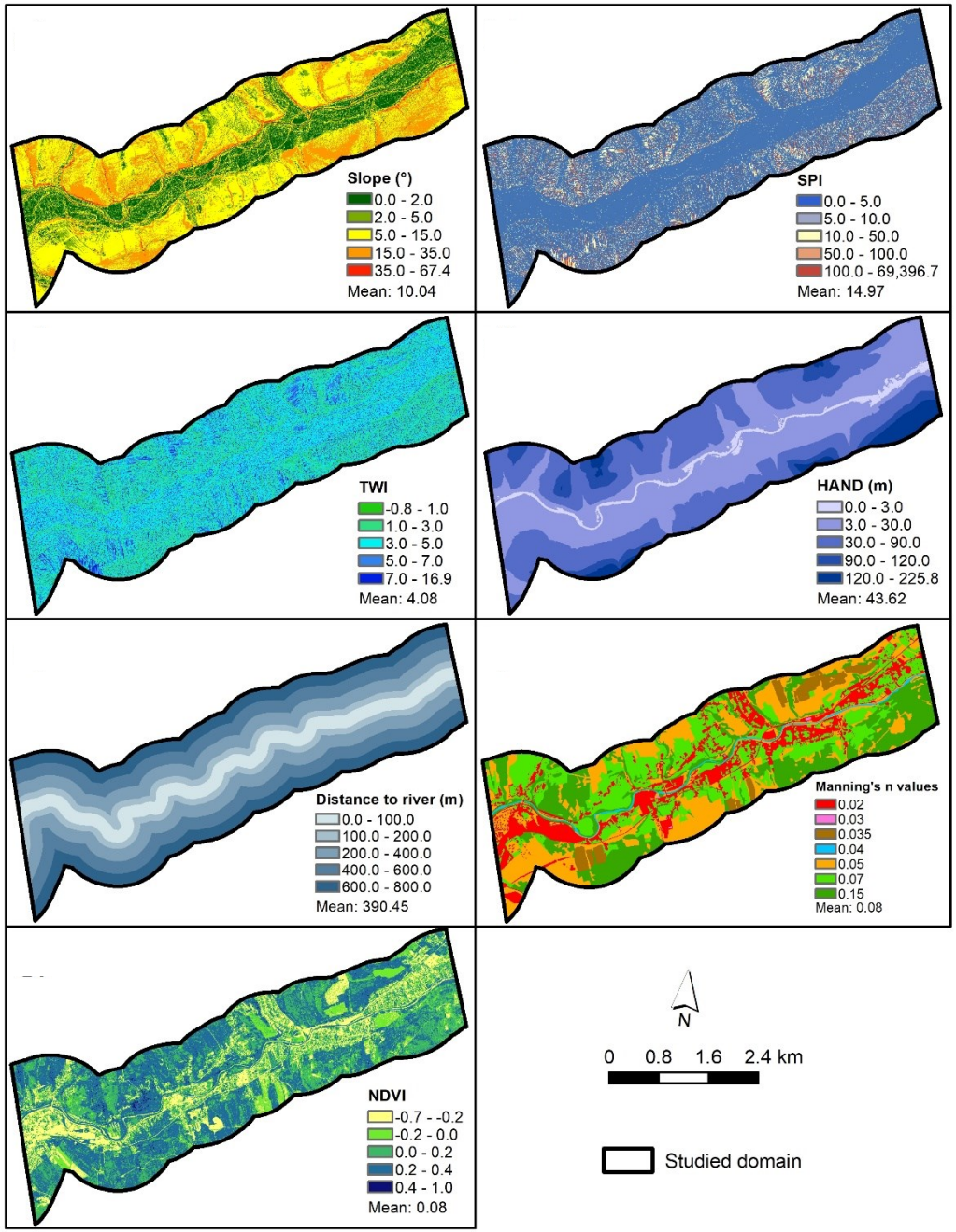
TP: true positive, FN: false negative, FP: false positive pixels

RESULTS: PREDICTORS

Hydraulically derived maps



GIS/RS-based predictors



RESULTS: MULTICOLINEARITY

Pearson correlation



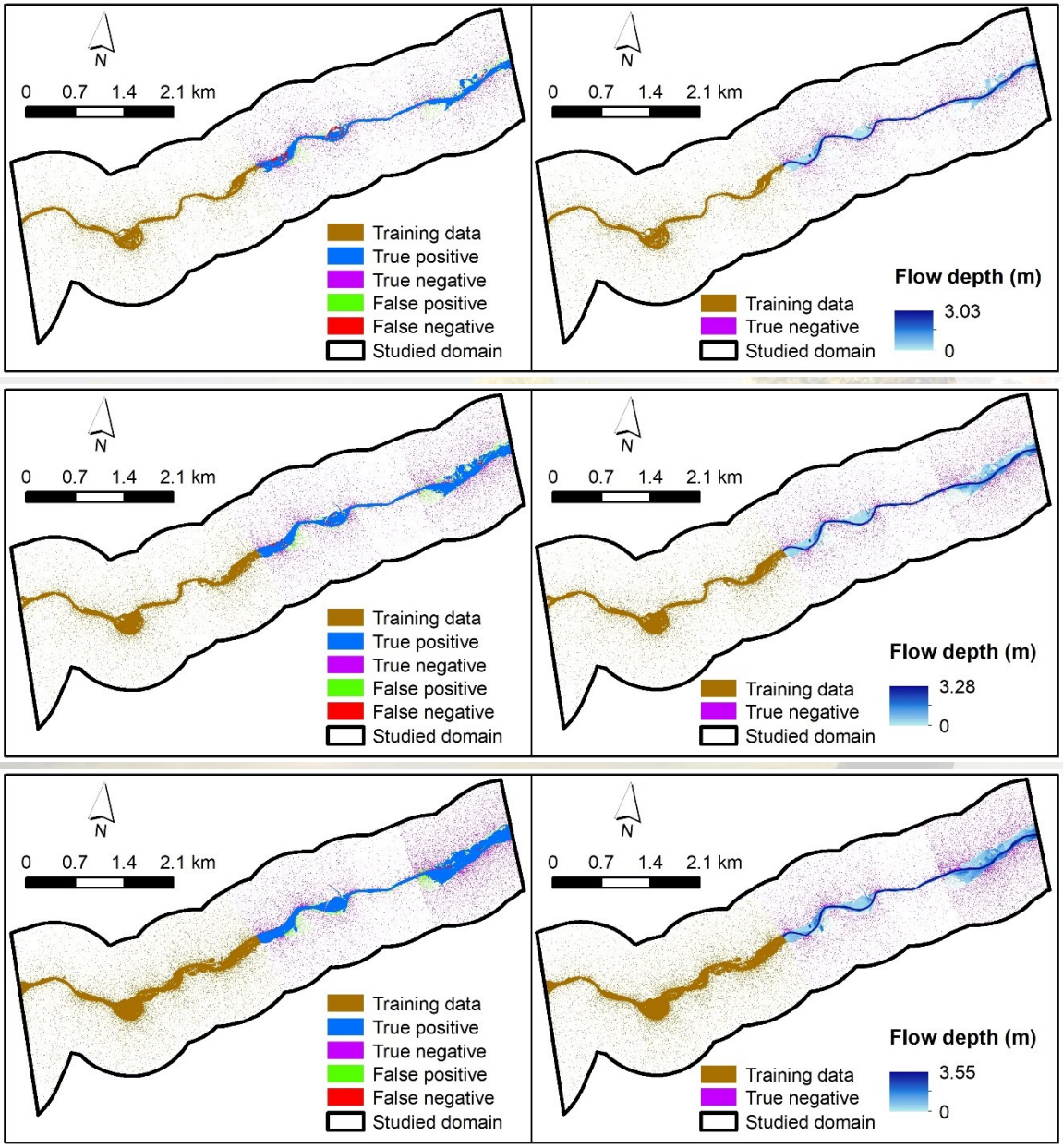
Predictor	VIF value
HAND	2.10
Distance from river	1.90
Slope	1.55
Roughness	1.50
TWI	1.23
NDVI	1.22
SPI	1.09

RESULTS: FLOOD INUNDATION MAPS BY XGBoost

Q_{10} (5/5 fold)

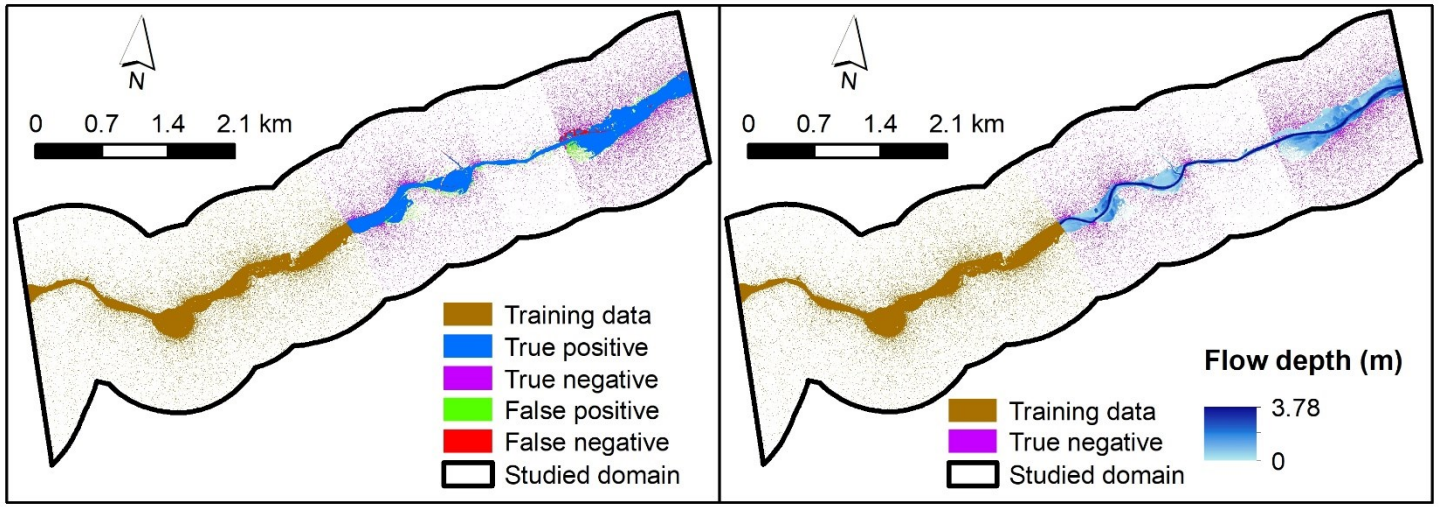
Q_{20} (5/5 fold)

Q_{50} (5/5 fold)

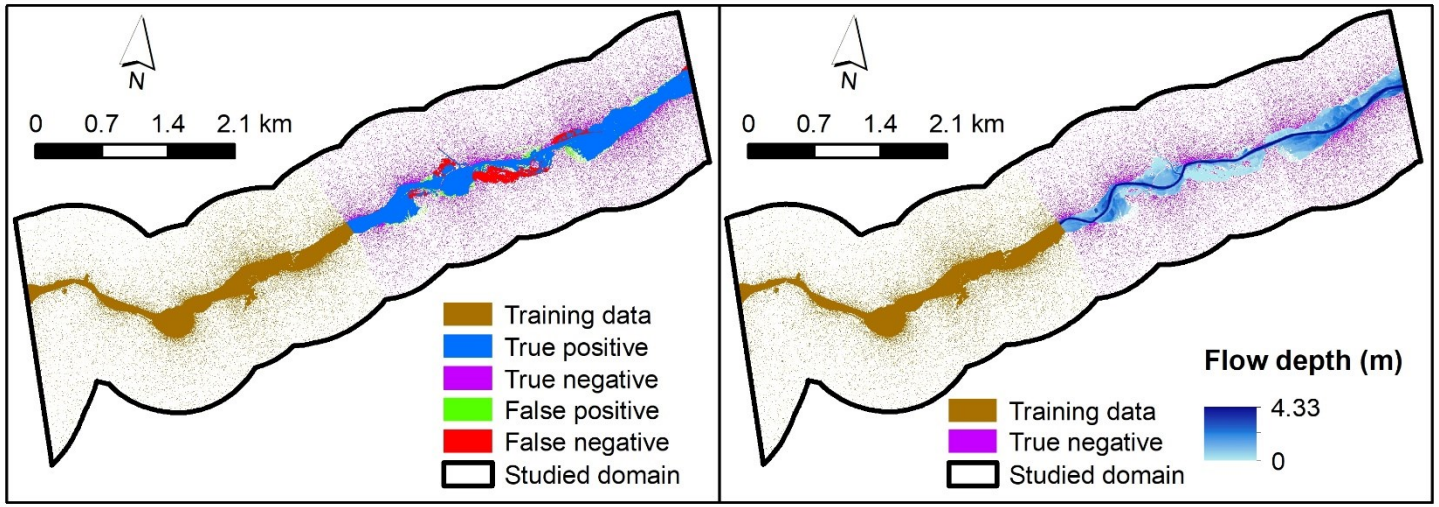


RESULTS: FLOOD INUNDATION MAPS BY XGBoost

Q_{100} (5/5 fold)



Q_{1000} (5/5 fold)



RESULTS: CLASSIFICATION METRICS

Flood scenario	Fold	Precision	Recall	F1-score	Training time (sec)	Testing time (sec)
Q_{10}	5_5	0.859	0.931	0.894	**4.8**	0.4
	6_4	0.848	0.945	0.894	5.4	0.3
	7_3	0.832	0.960	0.891	7.7	0.3
	8_2	**0.814**	0.951	0.877	9.2	0.2
	9_1	0.843	0.991	0.911	13.7	**0.1**
Q_{20}	5_5	0.865	0.980	0.919	6.1	0.6
	6_4	0.864	0.951	0.905	9.1	0.4
	7_3	0.862	0.963	0.910	11.4	0.3
	8_2	0.857	0.960	0.905	13.7	0.2
	9_1	0.896	0.983	0.938	15.6	0.1
Q_{50}	5_5	0.883	0.989	0.933	10.4	0.5
	6_4	0.891	0.968	0.928	7.9	0.4
	7_3	0.904	0.976	0.939	16.8	0.3
	8_2	0.909	0.972	0.940	12.8	0.3
	9_1	0.944	0.992	0.968	17.1	0.1
Q_{100}	5_5	0.914	0.971	0.941	14.3	0.8
	6_4	0.914	0.938	0.926	10.4	0.5
	7_3	0.927	0.936	0.931	19.3	0.4
	8_2	0.937	0.928	0.933	25.8	0.4
	9_1	**0.965**	0.976	**0.970**	25.0	0.2
Q_{1000}	5_5	0.935	0.852	0.892	13.5	**1.2**
	6_4	0.934	**0.806**	**0.865**	15.0	0.7
	7_3	0.929	0.838	0.881	26.1	0.8
	8_2	0.934	0.960	0.947	30.8	0.4
	9_1	0.923	**0.992**	0.956	**31.2**	0.2
Mean	-	**0.895**	**0.948**	**0.920**	**14.9**	**0.4**

RESULTS: REGRESSION METRICS

Flood scenario	Fold	RMSE	MAE	R^2	Bias	Training time (sec)	Testing time (sec)
Q_{10}	5_5	7.23%	4.52%	90.44%	2.68%	**6.0**	0.2
	6_4	7.14%	4.46%	90.83%	3.47%	6.7	0.2
	7_3	7.43%	**4.74%**	89.93%	**3.99%**	11.4	0.1
	8_2	7.27%	4.65%	88.82%	3.84%	8.9	0.1
	9_1	6.26%	3.81%	92.09%	2.66%	9.9	0.1
Q_{20}	5_5	6.95%	4.32%	89.69%	2.60%	7.5	0.4
	6_4	6.57%	4.11%	90.71%	2.97%	7.7	0.2
	7_3	6.77%	4.26%	89.85%	3.59%	8.5	0.2
	8_2	6.58%	4.12%	88.31%	3.45%	9.8	0.1
	9_1	5.20%	3.10%	93.40%	1.79%	13.2	0.1
Q_{50}	5_5	7.06%	4.37%	88.02%	2.68%	10.4	0.4
	6_4	6.29%	3.93%	90.22%	2.56%	13.6	0.2
	7_3	6.44%	4.00%	89.31%	3.24%	20.4	0.2
	8_2	6.33%	3.90%	87.08%	3.17%	17.8	0.2
	9_1	**4.72%**	**2.76%**	94.11%	0.95%	29.4	**0.1**
Q_{100}	5_5	6.92%	4.26%	88.00%	2.35%	14.9	0.3
	6_4	6.38%	3.91%	89.72%	2.01%	23.4	0.3
	7_3	6.39%	3.86%	89.16%	2.69%	25.7	0.2
	8_2	6.26%	3.76%	87.12%	2.57%	24.8	0.2
	9_1	4.95%	2.89%	94.07%	**-0.40%**	31.6	0.1
Q_{1000}	5_5	7.44%	4.48%	84.97%	2.02%	18.2	**0.6**
	6_4	7.19%	4.34%	85.00%	1.61%	22.2	0.4
	7_3	7.12%	4.31%	84.50%	2.69%	34.7	0.3
	8_2	**7.65%**	4.70%	**81.33%**	3.46%	33.8	0.3
	9_1	4.99%	3.07%	**94.52%**	0.13%	**48.3**	0.1
Mean	-	**6.54%**	**4.03%**	**89.25%**	**2.51%**	**18.4**	**0.2**

RESULTS: PREDICTOR IMPORTANCE

Flood scenario	Fold	XGBoost classification							XGBoost regression						
		HAND	Distance to river	Roughness	SPI	TWI	NDVI	Slope	HAND	Distance to river	Roughness	SPI	TWI	NDVI	Slope
Q_{10}	5_5	65.5%	15.5%	8.2%	1.5%	0.7%	4.7%	3.8%	70.5%	19.6%	5.9%	0.4%	1.1%	1.1%	1.3%
	6_4	66.4%	14.8%	7.2%	1.8%	1.0%	4.1%	4.6%	71.5%	19.3%	5.4%	0.4%	0.9%	1.1%	1.2%
	7_3	66.7%	14.9%	5.7%	2.1%	1.8%	3.9%	4.9%	71.9%	19.5%	5.2%	0.4%	0.8%	1.0%	1.2%
	8_2	66.2%	15.4%	6.2%	2.1%	1.5%	4.0%	4.6%	72.0%	19.7%	4.9%	0.4%	0.9%	1.0%	1.1%
	9_1	68.5%	14.8%	5.9%	1.7%	1.4%	3.4%	4.3%	73.5%	18.8%	4.6%	0.4%	0.7%	1.0%	1.0%
Q_{20}	5_5	67.5%	15.2%	6.9%	1.4%	1.3%	3.2%	4.6%	69.7%	20.3%	5.8%	0.5%	0.9%	1.1%	1.7%
	6_4	68.5%	13.9%	6.8%	1.8%	1.3%	3.0%	4.7%	71.0%	19.5%	5.3%	0.5%	0.8%	1.3%	1.5%
	7_3	68.2%	13.8%	5.9%	2.0%	2.0%	3.1%	5.0%	71.9%	19.4%	4.8%	0.5%	0.8%	1.1%	1.5%
	8_2	68.4%	13.8%	6.1%	2.0%	1.8%	3.1%	4.7%	72.3%	19.2%	4.7%	0.5%	0.8%	1.1%	1.4%
	9_1	69.9%	14.1%	5.8%	1.8%	1.6%	2.4%	4.4%	73.0%	18.9%	4.4%	0.4%	0.7%	1.2%	1.3%
Q_{50}	5_5	71.0%	12.6%	6.4%	1.5%	1.3%	2.0%	5.2%	71.0%	18.8%	5.9%	0.6%	0.8%	1.1%	1.8%
	6_4	70.9%	13.3%	5.4%	1.8%	1.2%	2.1%	5.4%	73.0%	17.3%	5.4%	0.7%	0.7%	1.3%	1.7%
	7_3	70.8%	13.0%	5.1%	2.1%	1.7%	2.0%	5.3%	73.8%	16.9%	5.1%	0.6%	0.7%	1.1%	1.8%
	8_2	71.3%	13.2%	5.4%	2.0%	1.5%	2.0%	4.6%	74.0%	17.0%	4.9%	0.6%	0.7%	1.0%	1.7%
	9_1	74.3%	12.3%	4.2%	2.0%	1.2%	1.6%	4.4%	74.8%	17.1%	4.3%	0.6%	0.6%	1.0%	1.6%
Q_{100}	5_5	72.1%	12.9%	5.5%	1.8%	1.2%	1.8%	4.7%	71.3%	18.4%	5.8%	0.6%	0.8%	1.2%	1.9%
	6_4	73.4%	12.4%	5.1%	1.6%	1.1%	1.7%	4.6%	72.8%	17.6%	5.0%	0.7%	0.7%	1.3%	1.9%
	7_3	74.5%	11.9%	4.9%	1.7%	1.2%	1.7%	4.1%	74.3%	16.7%	4.6%	0.6%	0.7%	1.1%	2.0%
	8_2	73.9%	12.5%	5.2%	1.8%	1.1%	1.8%	3.8%	74.6%	16.6%	4.6%	0.6%	0.7%	1.0%	1.8%
	9_1	76.4%	12.3%	3.6%	1.5%	0.9%	1.4%	3.9%	76.6%	15.7%	3.8%	0.7%	0.6%	1.0%	1.7%
Q_{1000}	5_5	76.3%	12.2%	4.1%	1.2%	0.8%	1.5%	3.8%	73.0%	17.4%	4.8%	0.7%	0.7%	1.3%	2.1%
	6_4	77.5%	11.6%	3.9%	1.2%	0.7%	1.3%	3.8%	75.0%	16.0%	4.4%	0.8%	0.6%	1.2%	2.0%
	7_3	76.4%	11.7%	3.5%	1.7%	0.9%	1.5%	4.3%	74.3%	16.0%	4.5%	1.0%	0.7%	1.1%	2.4%
	8_2	75.4%	11.9%	3.8%	1.5%	1.1%	1.8%	4.5%	75.9%	14.8%	4.4%	0.9%	0.7%	1.1%	2.2%
	9_1	77.0%	11.6%	3.0%	1.5%	1.0%	1.3%	4.6%	76.1%	15.5%	3.7%	0.8%	0.6%	1.3%	2.1%
Mean	-	**71.5%**	**13.3%**	**5.4%**	**1.7%**	**1.3%**	**2.4%**	**4.5%**	**73.1%**	**17.9%**	**4.9%**	**0.6%**	**0.7%**	**1.1%**	**1.7%**

CONCLUSION

- ML training on predictors specifically adapted for fluvial flood hazard.
- XGBoost-based approach proved to be computationally efficient to emulate traditional hydraulic modeling for flood extent and flow depth prediction.
- Short training and testing times of ML models compared to runtime of hydraulic models.
- Potential of ML-XGBoost models for near-real time flood mapping or early flood warning.

Financované Európskou úniou NextGenerationEU

PLÁN OBNOVY

ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU

VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA

MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY

VÝSKUMNÁ AGENTÚRA

Thank you for your attention

Funded by the European Union NextGenerationEU

RECOVERY AND RESILIENCE PLAN

Acknowledgment:
Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V03-00085.

CONSTANTINE THE PHILOSOPHER UNIVERSITY IN NITRA FACULTY OF NATURAL SCIENCES AND INFORMATICS

Matej Vojtek
Dávid Držík
Jana Vojteková

Faculty of Natural Sciences and Informatics
Constantine the Philosopher University in Nitra
Trieda A. Hlinku 1, 949 01 Nitra, Slovakia
E-mail: mvojtek@ukf.sk

INSTITUTE OF GEOGRAPHY SLOVAK ACADEMY OF SCIENCES

Matej Vojtek

Institute of Geography
Slovak Academy of Sciences
Štefánikova 49, 814 73 Bratislava
E-mail: matej.vojtek@savba.sk