SES 2026 · Twenty-second International Scientific Conference SPACE, ECOLOGY, SAFETY · 20–24 October 2026, Stara Zagora, Bulgaria

Comparative Evaluation of Validation Approaches for Water Body Mapping Using PlanetScope Imagery

Оценка на различни подходи за валидация при картиране на водни тела от изображения PlanetScope

Adrian Yordanov

University of Architecture, Civil Engineering and Geodesy · a.yordanov_fgs@uacg.bg

Same classifier, same samples, four validation designs. How the score is averaged changed IoU far more than which validation design was used.

Grouping unit: scene vs water body

<1 pt

94.1% vs 95.0% pooled IoU on identical samples

Scored rows: with vs without interior land

≈1 pt

93.0% vs 94.1% for leave-one-scene-out

Averaging: pooled vs per water body

>20 pt

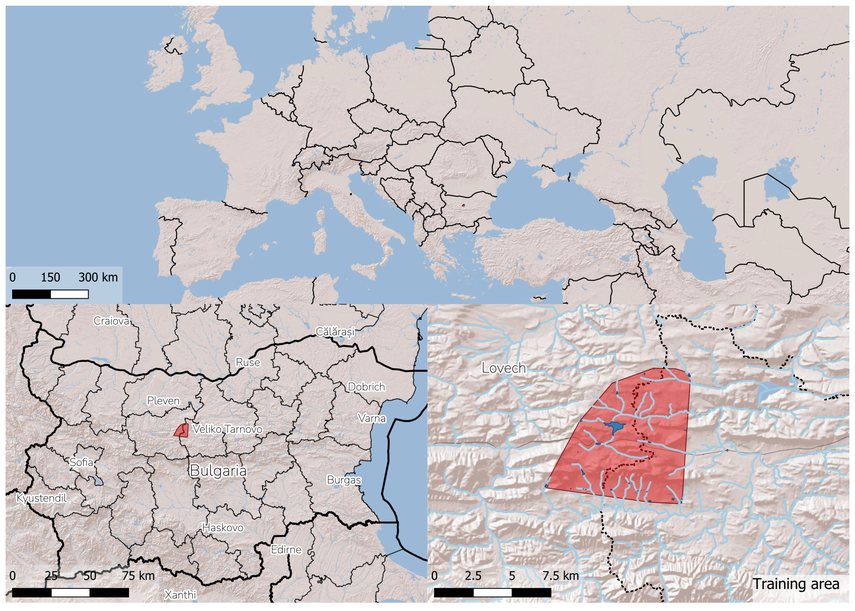
94.96% pooled vs 73.8% mean (median 85.7%)

1 Why it matters

Grouped validation holds out whole units so spatially correlated neighbors do not inflate accuracy. For water bodies the unit can be a scene or a water body. Two further choices are rarely reported: which land samples are scored and how unit scores are averaged.

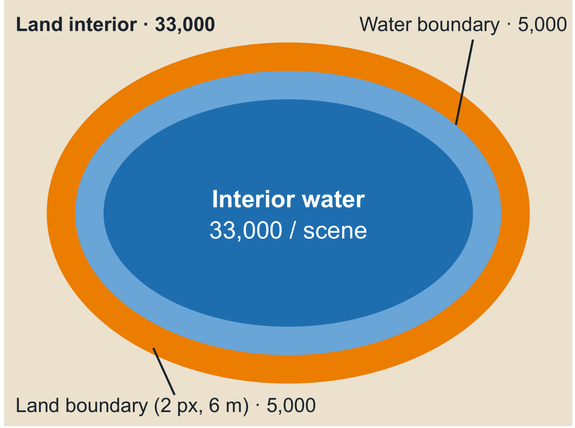
We do not rank validation designs. We ask how much the reported IoU depends on each choice.

2 Data and study area



- 14 PlanetScope images, 2023, 4 bands, 3 m, surface reflectance harmonized with Sentinel-2
- 31 water bodies: 1 reservoir, 5 large, 6 medium, 10 small, 9 very small ponds varying between acquisitions
- Polygons digitized per scene; persistent ID per water body

3 Method



Four classes from 2-pixel erosion and dilation of the reference water. 1,064,000 samples in total.

Random Forest, 500 trees, “√” features per split, Gini; 8 features selected inside each training fold.

Samples drawn at random with a minimum spacing.

Validation design	Folds / splits	Held out
Leave-one-scene-out	14	1 scene
Leave-one-water-body-out	31	1 water body
Group-shuffle by scene	10	4 scenes
Group-shuffle by water body	10	3 water bodies

IoU = TP / (TP + FP + FN), water positive. Pooled: counts summed over folds. Per unit: IoU per held-out unit, then mean / median.

4 Results

IoU, %	Pooled, all rows	Pooled, shared rows	Mean per unit	Median per unit	Min per unit
Leave-one-scene-out (14)	93.0	94.1	94.1	94.3	90.9
Leave-one-water-body-out (31)	n/a	95.0	**73.8**	85.7	**0.0**
Group-shuffle by scene (10)	92.8	–	–	–	–
Group-shuffle by water body (10)	–	89.0	–	–	–

Group-shuffle: pooled over 10 splits; split IoU mean ± SD 92.8 ± 1.0 (scene) and 88.4 ± 2.9 (water body). The largest water body was in none of the 10 water-body splits.

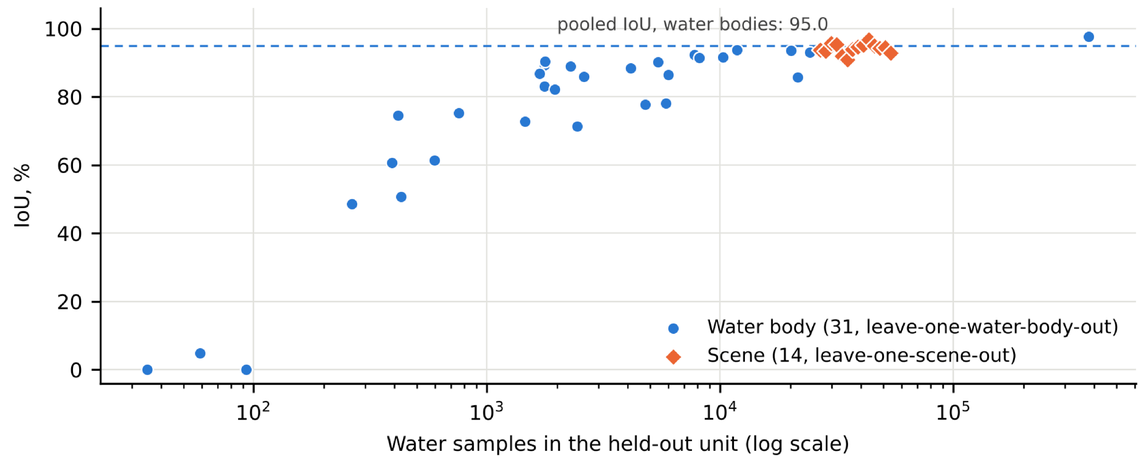
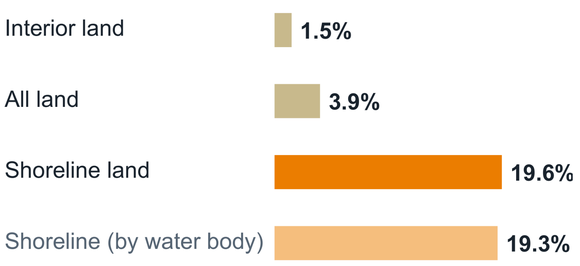


Fig. 2. IoU of each held-out unit against its water samples. Per-object IoU rises with size (Spearman ρ = 0.85); the three smallest water bodies score below 5%. The largest holds 72% of all water samples and dominates the pooled value.

False-water rate on land samples



Shoreline samples are 13% of land samples but produce 67% of false positives, under either grouping unit.

13 April scene: 85.1% IoU on all rows vs 92.0% on shared rows.

5 Report with every accuracy score

1. The grouping unit and the number of folds or splits
2. The scored rows and their class mix (here 532,000 water vs 70,000 land)
3. Pooled and per-unit scores side by side, with unit count and size range
4. A separate score for the shoreline zone

Limits: one area, one year, one classifier, manual reference polygons. Transfer to new areas is untested. The 6 m shoreline ring is close to digitizing error, so part of its error may be label noise.

1. Karasiak, N. et al. Machine Learning, 111, 2022, 2715–2740.
2. Maxwell, A. E. et al. Remote Sensing, 13(13), 2021, 2450.
3. Meyer, H., E. Pebesma. Methods Ecol. Evol., 12(9), 2021, 1620–1633.
4. Morales-Barquero, L. et al. Remote Sensing, 11(19), 2019, 2305.
5. Mullen, A. L. et al. Geophys. Res. Lett., 50, 2023, e2022GL102327.
6. Ploton, P. et al. Nature Communications, 11, 2020, 4540.
7. Roberts, D. R. et al. Ecography, 40(8), 2017, 913–929.
8. Stehman, S. V., G. M. Foody. Remote Sens. Environ., 231, 2019, 111199.
9. Yordanov, A., T. Atanasova-Evdenova. София: УАСГ, 2025, 843–864.