

Supplementary Materials

● S1. Per-Date Waterline Extraction Performance

Table S1 presents the full-scene waterline evaluation results for vanilla U-Net, HED-UNet, and TGB-UNet on the nine acquisition dates in the main test dataset, organized by the tidal-level groups defined in Section 4.4 of the main text. All models were evaluated over the same valid spatial extent using the methods described in Section 4.6. Results are reported as the mean and standard deviation across three random seeds for each date.

TGB-UNet achieved higher mean BF1 and lower mean ASSD and P90 than vanilla U-Net on all nine dates. Compared with HED-UNet, TGB-UNet achieved lower mean ASSD and P90 on all nine dates and higher mean BF1 on seven dates. These results indicate consistent reductions in positional error across the evaluated dates, whereas improvements in boundary matching relative to HED-UNet varied among dates.

Table S1. Per-date full-scene waterline extraction performance on the main test dataset.

Image Date	Tidal-Level Group	Method	BF1 (%)	ASSD (m)	P90 (m)
20200220	High	Vanilla U-Net	89.75±0.14	57.91±2.98	62.14±5.22
20200220		HED-UNet	90.97±1.74	28.84±2.97	58.58±5.76
20200220		TGB-UNet	91.64±0.94	26.98±2.23	49.65±4.24
20200222		Vanilla U-Net	88.02±0.29	69.23±5.56	66.01±5.08
20200222		HED-UNet	89.13±0.14	32.84±0.69	60.30±4.28
20200222		TGB-UNet	92.36±0.14	24.53±2.11	47.94±3.02
20260113		Vanilla U-Net	82.05±0.42	63.67±1.90	84.44±5.71
20260113		HED-UNet	83.19±0.19	33.09±0.19	76.32±4.69
20260113		TGB-UNet	86.72±0.40	24.71±0.52	61.57±3.17
20250109	Intermediate	Vanilla U-Net	83.45±0.15	76.76±0.11	79.51±0.11
20250109		HED-UNet	85.66±0.67	42.37±2.97	75.72±3.94
20250109		TGB-UNet	86.31±0.14	40.14±1.34	69.70±2.43
20260211		Vanilla U-Net	73.67±0.55	104.13±11.50	161.10±17.27
20260211		HED-UNet	72.73±2.15	84.31±17.03	157.76±8.92
20260211		TGB-UNet	76.09±1.32	76.68±21.41	136.24±7.34
20251121		Vanilla U-Ne	80.80±0.74	125.73±8.87	110.64±14.50
20251121		HED-UNet	84.63±0.47	78.09±1.76	77.47±6.54
20251121		TGB-UNet	83.31±0.34	71.29±5.30	66.59±4.99
20250320	Low	Vanilla U-Net	80.05±0.11	78.98±0.47	127.36±3.29
20250320		HED-UNet	80.42±1.63	51.77±0.76	104.79±8.47
20250320		TGB-UNet	81.96±1.53	47.89±3.14	86.59±6.26
20260221		Vanilla U-Net	62.85±0.78	166.01±10.04	358.09±23.67
20260221		HED-UNet	67.47±1.46	104.92±20.36	216.20±33.84
20260221		TGB-UNet	66.55±1.44	104.77±16.46	199.26±13.89
20260314		Vanilla U-Net	82.89±0.43	69.08±1.63	87.34±5.69
20260314		HED-UNet	84.41±1.43	40.48±6.86	74.35±7.95
20260314		TGB-UNet	86.76±0.05	34.15±1.17	53.54±3.17

● S2. Paired Date-Level Statistical Analysis of Waterline Metrics

To assess whether the performance improvements of TGB-UNet were consistent across the individual test dates, paired statistical comparisons were conducted using the nine test dates as statistical units. BF1, ASSD, and P90 were included because they characterize boundary matching, mean positional error, and relatively large positional deviations, respectively. Two-sided Wilcoxon signed-rank tests were used to compare TGB-UNet with Vanilla U-Net and HED-UNet. The 95% confidence intervals of the mean improvements were estimated by date-level bootstrap resampling, and the resulting p-values were adjusted for multiple comparisons using the Holm method. The complete results are presented in Table S2

Table S2. Paired date-level statistical comparisons of full-scene waterline metrics between TGB-UNet and the baseline models across the nine test dates.

Comparison	Metric	Mean Improvement	95% CI	Wilcoxon p	Holm-Adjusted p
TGB-UNet vs. Vanilla U-Net	BF1	3.13	[2.52, 3.80]	0.0039	0.0234
TGB-UNet vs. Vanilla U-Net	ASSD	40.04	[34.20, 48.62]	0.0039	0.0234
TGB-UNet vs. Vanilla U-Net	P90	40.62	[22.86, 89.72]	0.0039	0.0234
TGB-UNet vs. HED-UNet	BF1	1.45	[0.25, 2.50]	0.0742	0.0742
TGB-UNet vs. HED-UNet	ASSD	5.06	[3.00, 6.81]	0.0039	0.0234
TGB-UNet vs. HED-UNet	P90	14.49	[11.12, 17.67]	0.0039	0.0234

Note: The nine test dates were treated as paired statistical units ($n = 9$). Mean improvement was calculated as TGB-UNet minus the baseline for BF1 and as the baseline minus TGB-UNet for ASSD and P90; therefore, positive values favor TGB-UNet. BF1 differences are expressed in percentage points, whereas ASSD and P90 differences are expressed in meters. The 95% confidence intervals were estimated using bias-corrected and accelerated bootstrap resampling of the test dates. The p-values were obtained using two-sided Wilcoxon signed-rank tests and adjusted for multiple comparisons using the Holm method. Holm-adjusted p-values below 0.05 are shown in bold.

● S3. Overlap of Reference Waterlines with GLO-30 Water- and Shoreline-Related Processing Domains

The Edit Data Mask (EDM) and Water Body Mask (WBM) were used to quantify the spatial overlap between reference waterlines and GLO-30 processing domains for each test date. P11 denotes the proportion of reference-waterline length within the EDM shoreline class (value 11). Pedit denotes the proportion within a combined water-shoreline domain, defined by WBM values 1–3 or EDM values 8–11. Both proportions were normalized by the total reference-waterline length for the corresponding date. Overlapping areas within the combined domain were counted only once.

Table S3. Proportions of reference-waterline length overlapping the GLO-30 shoreline class and

combined water–shoreline domain for each test date.

Image Date	Total Reference-Waterline Length (m)	P11 (%)	Pedit (%)
20200220	348920.16	0.45	65.36
20200222	352950.75	0.44	79.05
20250109	358457.87	0.20	78.54
20250320	429969.65	0.16	82.35
20251121	382525.24	0.39	87.48
20260113	385718.86	0.27	80.68
20260211	390349.37	0.26	76.81
20260221	433596.26	0.36	82.44
20260314	359765.93	0.41	71.58

● S4. Terrain-Input Shift Sensitivity Analysis

To quantify the influence of potential co-registration errors in the terrain data on the model outputs, slope and RE330 were shifted simultaneously by one pixel (15 m) to the east, west, north, and south. The optical imagery, trained model weights, normalization parameters, and prediction post-processing procedure were kept unchanged, and inference was performed without additional training or fine-tuning. For a fair comparison, the unshifted and four shifted conditions were evaluated only within their common valid spatial extent. U-Net was also re-evaluated within the same extent. Because the common valid extent excluded invalid boundary pixels introduced by the shifts, the absolute metric values obtained in this sensitivity analysis differ from those of the main experiment and should not be compared directly.

The results are presented in Table S4. The 15 m northward shift produced the largest mean increase in positional error, although the overall changes remained limited. Under all four shifted conditions, TGB-UNet retained a higher mean BF1 and lower mean ASSD and P90 than U-Net. These results indicate that the model outputs were generally stable to a one-pixel co-registration offset in the terrain inputs.

Table S4. Mean full-scene waterline metrics under one-pixel terrain-input shifts across the nine test dates.

Terrain-input condition	BF1(%)	ASSD (m)	P90(m)
Vanilla U-Net	80.25	60.71	115.69
TGB-UNet, unshifted	83.24	46.76	98.31
TGB-UNet, 15 m eastward shift	83.28	46.69	97.72
TGB-UNet, 15 m westward shift	83.19	46.91	99.70
TGB-UNet, 15 m northward shift	83.18	46.03	100.03
TGB-UNet, 15 m southward shift	83.23	83.23	90.10