Flood Depth Mapping from SAR Imagery Using CS-Mamba with DEM Sensitivity Analysis

Zhongyuan Yang[1], Wei Yuan[1,*], Weihang Ran[2], Changqing Liu[1], Bruno Adriano[1], Ryosuke Shibasaki[3], Shunichi Koshimura[1]

[1]International Research Institute of Disaster Science (IRIDeS), Tohoku University, Sendai – Japan, wei.yuan@tohoku.ac.jp
[2]The University of Tokyo, Tokyo – Japan
[3]Reitaku University, Chiba – Japan

Keywords: CS-Mamba, SAR, Semantic segmentation, DEM, Flood depth

Abstract

Accurate flood extent and depth information are essential to emergency response, yet most existing studies treat these tasks separately. This work introduces an integrated SAR-to-depth framework that combines water body semantic segmentation with DEM-based geometric depth estimation to generate both flood-extent maps and pixel-wise depth products from Sentinel-1 imagery. For flood extent mapping, we propose a cross-scale Mamba with selective state-space blocks, which achieves a mean IoU of 79.8% across ten European flood events from the KuroSiwo benchmark, outperforming RSMamba by 7.4% and surpassing common CNN baselines. The experimental results demonstrate that the proposed model also generalizes well to unseen events, with test performance exceeding validation scores. When both CS-Mamba predictions and KuroSiwo reference masks are input to FLEXTH, the resulting depth estimates agree within $\pm 2\%$ across four global DEMs. Initial validation against ICESat-2 altimetry using MERIT DEM (19 matched points) shows RMSE of 4.60 m and Bias of -1.88 m, providing preliminary validation evidence with systematic underestimation. Systematic DEM comparison shows FLEXTH is robust across all four DEMs, with Copernicus and MERIT showing closest agreement with reference mask estimates. The framework produces three-class flood masks and pixel-wise depth maps, combining extent mapping with quantitative depth information for operational flood monitoring.

1. Introduction

Floods are among the most destructive natural hazards, causing severe socio-economic impacts and loss of life (Amitrano et al., 2024). Accurate flood mapping is critical for emergency response, while quantitative depth estimation supports damage assessment (Yuan et al., 2023b) and resource allocation (Bai et al., 2021). Synthetic Aperture Radar (SAR) enables large-scale flood monitoring with all-weather, day-night imaging capability. However, most operational services provide only binary flood extent maps that do not differentiate permanent water from transient floods, and lack quantitative depth information needed for disaster management (Cohen et al., 2019).

Deep learning has improved SAR-based flood detection significantly. CNNs (Jamali et al., 2024) and transformers (Sharma and Saharia, 2025) show strong results on benchmark datasets such as KuroSiwo (Bountos et al., 2023), which provides a standard testbed with 43 events covering 338 billion m². However, transformer-based methods face limitations: their $O(n^2)$ computational complexity restricts operational deployment for large-scale monitoring (Yuan et al., 2023a). Meanwhile, quantitative depth estimation remains largely unexplored. Geometric approaches (Cohen et al., 2019, Betterle and Salamon, 2024) provide faster alternatives to hydrodynamic models, but we lack systematic understanding of how segmentation quality and DEM selection affect depth estimation accuracy.

This study integrates flood extent mapping and depth quantification. These two research areas are typically studied separately. We make three contributions:

1. We develop an integrated SAR-to-depth pipeline that produces pixel-wise water depth maps from Sentinel-1 imagery and DEMs. The framework combines deep learning segmentation (CS-Mamba) with geometric depth estimation (FLEXTH (Betterle and Salamon, 2024)). FLEXTH's morphological processing and water level interpolation enable accurate depth estimation even with segmentation imperfections, demonstrating that automated segmentation can replace manual annotations for operational depth mapping.

2. We propose CS-Mamba (Cross-Scale Mamba), an architecture that brings multi-scale feature fusion into Mamba for SAR flood segmentation. Unlike existing Mamba-based approaches that operate at single scales, CS-Mamba integrates cross-scale information through hierarchical feature extraction and fusion mechanisms, so that the model captures both fine-grained flood boundaries and large-scale inundation patterns. CS-Mamba achieves 79.8% mIoU on KuroSiwo, outperforming RSMamba (72.4%) and CNN baselines while maintaining linear O(n) computational complexity.

3. We compare four global DEMs (SRTM, Copernicus, MERIT, FABDEM) and conduct initial validation using ICESat-2 altimetry with MERIT DEM. Depth estimates from CS-Mamba masks show mean differences within ±2% of reference masks across all DEMs, indicating FLEXTH's robustness to both automated and manual annotations. ICESat-2 validation with 19 matched points yields RMSE of 4.60 m and Bias of -1.88 m, providing preliminary validation evidence despite systematic underestimation and limited sample size that constrain statistical robustness.

* Corresponding author.

2. Related Work

2.1 SAR Flood Detection and Deep Learning Architectures

Deep learning has improved SAR-based flood mapping across different architectures. CNNs remain the dominant approach. U-Net (Ronneberger et al., 2015), an encoder-decoder with symmetric skip connections, is commonly used for Sentinel-1 flood segmentation (Jamali et al., 2024). Transfer learning with ImageNet-pretrained encoders (ResNet, VGG) works effectively for SAR applications despite domain differences (Saleh et al., 2024). DeepLabV3+ (Chen et al., 2018) uses Atrous Spatial Pyramid Pooling (ASPP) with parallel dilated convolutions for multi-scale context aggregation, benefiting flood feature extraction at different scales. Recent benchmarking on KuroSiwo (Bountos et al., 2023) (43 events, 338 billion m²) shows CNN models achieve 75–80flood segmentation.

Vision Transformers offer an alternative to CNNs through self-attention mechanisms for global context modeling. For flood detection, ViT architectures have been adapted to SAR imagery (Sharma and Saharia, 2025), where multi-head attention captures long-range spatial dependencies. However, transformers' quadratic $O(n^2)$ complexity limits scalability for high-resolution SAR imagery (e.g., 10m Sentinel-1), restricting operational deployment at continental scales (Yuan et al., 2023a).

Mamba (Zhu et al., 2024) achieves linear $O(n)$ complexity while preserving long-range modeling capability. Mamba has been applied to medical image segmentation (VM-UNet (Ruan and Xiang, 2024)) and remote sensing applications. RSMamba (Chen et al., 2024b), originally a scene classification model, uses patch-level predictions. RS3Mamba (Ma et al., 2024) targets semantic segmentation, while ChangeMamba (Chen et al., 2024a) targets change detection. Recent applications extend to building damage assessment, focusing on damage classification rather than flood mapping. SAR-based flood segmentation with Mamba is still limited. RSMamba's classification-oriented design lacks the hierarchical encoder and skip connections needed for pixel-level dense prediction. We propose CS-Mamba, a cross-scale Mamba architecture that introduces multi-scale feature fusion mechanisms. Unlike RSMamba's single-scale design, CS-Mamba integrates hierarchical feature extraction with cross-scale fusion, allowing the network to resolve both local boundaries and broad inundation areas for improved flood boundary delineation.

2.2 DEM-based Depth Estimation

Quantitative depth estimation involves a trade-off between accurate but expensive hydrodynamic models and faster geometric approaches. FwDET (Cohen et al., 2019) and FLEXTH (Betterle and Salamon, 2024) estimate water surfaces from flood boundaries and compute depths geometrically, enabling operational deployment. Validation of geometric depth methods uses paired ICESat-2 tracks from dry and wet periods (Betterle and Salamon, 2024), or single-track validation against DEM-derived depths. DEM quality affects results: Xu et al. (Xu et al., 2021) and Cohen et al. (Cohen et al., 2022) found that vertical errors transfer directly to depth estimates, and building/vegetation removal improves accuracy (Yuan et al., 2024). Global DEMs differ in preprocessing and vertical accuracy, but systematic comparisons using FLEXTH for European floods are limited. We address this gap by comparing four global DEMs (SRTM, Copernicus, MERIT, FABDEM) through controlled experiments and ICESat-2 validation.

3. Methodology

3.1 Overview

The framework (Fig. 1) combines deep learning flood segmentation with DEM-based depth estimation in a unified pipeline. CS-Mamba segments flood extent from multi-temporal Sentinel-1 SAR imagery, producing three-class masks (no water, permanent water, transient floods). FLEXTH (Betterle and Salamon, 2024) computes pixel-wise water depths from flood masks and DEMs using geometric methods. For DEM sensitivity analysis, FLEXTH processes both CS-Mamba predictions and reference masks with four DEMs using identical parameters, separating segmentation quality effects from DEM selection effects. Depth estimates are validated against ICESat-2 altimetry.

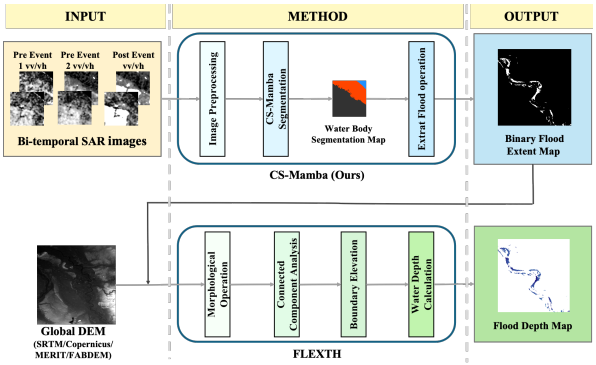


Figure 1. Integrated SAR to flood depth framework workflow. CS-Mamba processes multi-temporal Sentinel-1 SAR imagery (two pre-event + one post-event, VV/VH polarizations) for three-class flood segmentation. FLEXTH estimates pixel-wise water depths from flood masks and DEMs, with DEM sensitivity analysis conducted using four global products.

3.2 CS-Mamba Architecture

3.2.1 Network Design CS-Mamba is a cross-scale Mamba architecture that introduces multi-scale feature fusion into Mamba for efficient flood segmentation. Figure 2 illustrates the overall architecture and ConvRSMamba block design. The main design choice is to combine hierarchical feature extraction with cross-scale fusion, so that the network jointly captures fine-grained boundaries and large-scale patterns. Building on RSMamba (Chen et al., 2024b), we introduce three key components: multi-scale hierarchical encoder, symmetric decoder with cross-scale skip connections, and ConvRSMamba blocks that combine local and global features.

Encoder: The encoder has four hierarchical stages processing features at resolutions 56×56, 28×28, 14×14, and 7×7, with channel dimensions [96, 192, 384, 768]. Initial patch embedding uses stride-4 convolution ($224 \times 224 \rightarrow 56 \times 56$), then progressive downsampling through patch merging. Each stage contains [2, 2, 6, 2] ConvRSMambaBlocks (12 blocks total) with stochastic depth regularization.

ConvRSMamba Block: Each block combines depthwise-separable convolution (local features) with RSMamba layer (global context). The RSMamba component uses gated fusion across three scanning paths (forward, reverse, shuffled), enabling view-invariant feature learning. Features from convolutional and Mamba branches are fused through residual addi-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-493-2026 | © Author(s) 2026. CC BY 4.0 License.

tion, then passed through an FFN (expansion ratio 4). This hybrid design maintains linear $O(n)$ complexity while capturing fine-grained boundaries and large-scale patterns.

Decoder: The symmetric decoder has three upsampling stages ($7 \times 7 \rightarrow 56 \times 56$), matching the encoder structure. Each stage includes patch expanding ($2\times$ upsampling with channel halving), skip connection concatenation with encoder features, linear projection for dimension alignment, and ConvRSMambaBlocks for feature refinement. Final upsampling to 224×224 uses two transposed convolutions (stride 2) and a 1×1 convolution for three-class logits.

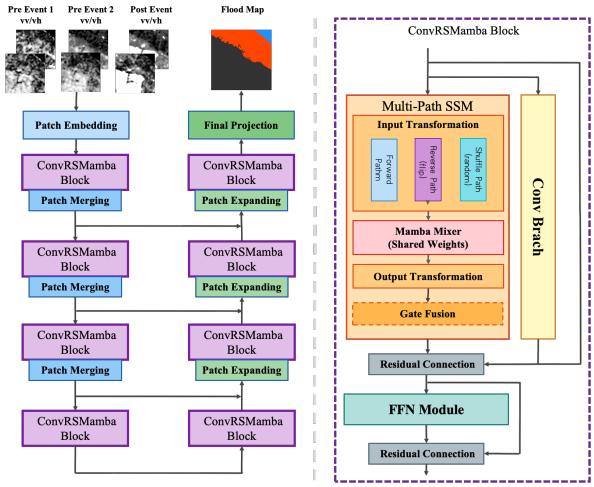


Figure 2. CS-Mamba architecture. Left: U-Net encoder-decoder structure with four-stage hierarchical design processing multi-temporal SAR inputs. Encoder and decoder stages are connected via skip connections. Right: ConvRSMamba Block internal structure integrating convolutional operations with multi-path Mamba scanning and gate fusion mechanisms.

3.2.2 Multi-Temporal Input Configuration The model processes six-channel input from three Sentinel-1 SAR acquisitions (VV and VH each): two pre-event images for baseline water conditions and one post-event image showing peak flooding. The channels are concatenated as [Pre-event 2, Pre-event 1, Post-event] with VV and VH for each timestamp, forming a $6 \times 224 \times 224$ input tensor. This multi-temporal setup distinguishes permanent water bodies (persistent across acquisitions) from transient floods (appearing only post-event), which helps prioritize newly flooded areas during emergency response.

3.2.3 Training Strategy To address class imbalance (floods ¡3% of pixels), we combine Focal Loss (weight 0.4, γ=3.5) for hard example mining with Dice Loss (weight 0.6) for IoU optimization. Class weights [0.2, 2.5, 3.5] prioritize rare flood and permanent water classes. We use AdamW optimizer (lr=10^{-4}, weight decay=0.08) with cosine annealing scheduler (5-epoch warmup, minimum lr 10^{-6}) for 80 epochs maximum with batch size 16. Data augmentation uses geometric transformations (flips, 90° rotations, affine) and intensity augmentations (blur, noise). Gradient clipping (max norm 1.0), exponential moving average (decay 0.999), and early stopping (patience 20 epochs) are applied for stability.

3.2.4 Evaluation Metrics Segmentation accuracy is quantified using four metrics: Recall, Precision, F1-score, and Intersection over Union (IoU). These metrics are computed per class and averaged to yield mean values.

IoU (Jaccard index) measures the spatial overlap between predicted and reference masks. For class c:

$$\text{IoU}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c + \text{FN}_c} \tag{1}$$

with TP_c, FP_c, and FN_c denoting true positives, false positives, and false negatives. Recall and Precision are defined as:

$$\text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c} \tag{2}$$

$$\text{Precision}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c} \tag{3}$$

F1-score represents the harmonic mean of Precision and Recall:

$$\text{F1}_c = \frac{2 \times \text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{4}$$

Mean IoU, computed as the arithmetic mean of class-wise IoU values, serves as the primary evaluation metric. Class-level analysis reveals performance variations among water types, important for applications requiring distinction between permanent water and flooding.

3.3 Depth Estimation and DEM Comparison

The FLEXTH framework (Betterle and Salamon, 2024) estimates pixel-wise water depths from flood masks and DEMs using geometric principles. Figure 3 summarizes the FLEXTH workflow used for depth estimation from flood masks and DEMs. To analyze DEM sensitivity, we apply FLEXTH to two mask sources— CS-Mamba predictions and KuroSiwo reference annotations—processing both with four DEMs using identical parameters. This design separates DEM effects from segmentation quality effects.

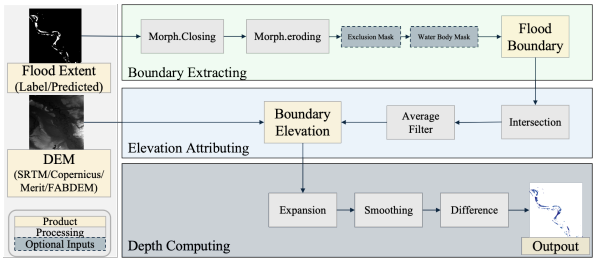


Figure 3. FLEXTH depth estimation workflow. The algorithm processes flood masks and DEMs through three stages: boundary extraction, elevation attribution, and depth computation. Required inputs are flood extent mask and DEM.

3.3.1 FLEXTH Algorithm FLEXTH estimates water depth in three steps: boundary extraction using morphological operations, water level assignment to boundary pixels from DEM elevations, and spatial interpolation across flooded regions. The water level at pixel (i, j) is computed as an inverse-distance weighted average of boundary elevations:

$$WL(i,j) = \frac{\sum_{k=1}^{N_{\max}} w_k \cdot DEM_k}{\sum_{k=1}^{N_{\max}} w_k}, \quad w_k = \frac{1}{d_k^{\alpha}} \tag{5}$$

where DEM_k is the elevation of the k-th boundary pixel, d_k is its distance to (i, j), and $\alpha = 2$ is the distance decay exponent. Water depth is computed as $WD(i, j) = WL(i, j) - DEM(i, j)$, with negative values set to zero. We use Method B (percentile-based water level estimation) with parameters: slope threshold 0.05, gap closing 0.05 km^2, border percentile 0.50, maximum 100 neighbors. The framework produces georeferenced flood masks and pixel-wise depth maps at 10m resolution, aligned with input SAR imagery.

3.3.2 ICESat-2 Validation Method Accuracy of depth estimates is assessed by comparing them with ICESat-2 laser altimetry. ICESat-2 ATLAS records elevation along ground tracks with sub-meter sampling (Neuenschwander and Pitts, 2019). ATL03 granule from Cycle 10 for Event 497 (Germany, October 2020) contains photons from three strong beams within the study area.

Photons are selected within 6.2°E–6.85°E, 51.15°N–51.75°N, with elevation between 0–35 m and water_conf ≥ 0. All data processing is conducted in WGS84 coordinate system to ensure direct compatibility between ICESat-2 photon coordinates (ellipsoid heights in WGS84) and DEM elevations. Water depth at photon locations is calculated as:

$$D_{\text{ICESat}} = h_{\text{ellipsoid}} - \text{DEM} \tag{6}$$

where DEM elevation is sampled from MERIT DEM at photon coordinates. MERIT is selected for validation due to its hydrological optimization and multi-error-removal processing, which minimize vertical artifacts in flood-prone terrain. The assumption is that ellipsoid height represents water surface elevation and DEM represents terrain surface.

FLEXTH water depth raster and MERIT DEM are reprojected to WGS84 for direct spatial correspondence with ICESat-2 photon coordinates. FLEXTH depths are sampled at photon locations using adaptive windows. Sampling starts with a 3×3 pixel window centered on photon coordinates. If no positive depth pixels exist (common near flood boundaries), the window is expanded up to 101×101 pixels (1 km) to find the nearest non-zero depth. This approach addresses spatial offset between ICESat-2 tracks (90 m beam separation) and flood boundaries.

Validation points are filtered by: photon-to-pixel distance ≤ 60 m, ICESat-2 depth ≥ 1.0 m, depth difference ≤ 30 m, and at least 4 valid pixels in the sampling window.

3.3.3 Depth Validation Metrics Depth estimation accuracy is assessed using three metrics: Bias, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). For n validation points:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(d_{\text{est},i} - d_{\text{ref},i}) \tag{7}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{\text{est},i} - d_{\text{ref},i})^2} \tag{8}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|d_{\text{est},i} - d_{\text{ref},i}| \tag{9}$$

where $d_{\text{est},i}$ and $d_{\text{ref},i}$ denote estimated and reference depths at point i. Bias indicates systematic overestimation (positive) or underestimation (negative), while RMSE and MAE quantify overall error magnitude. RMSE penalizes large errors more heavily than MAE.

4. Experiments and Results

4.1 Experimental Setup

4.1.1 Dataset and Evaluation Setup We use the European subset of KuroSiwo (Bountos et al., 2023), a benchmark dataset with three-class flood annotations (no water, permanent water, floods) covering 43 global events.

Flood segmentation dataset. We select 10 spatially and temporally independent European flood events (2016-2021) to evaluate cross-event generalization. The data partition (Table 1 and Figure 4) includes 7,882 labeled patches from 7 European countries across 2 climatic zones.

Table 1. Dataset partition across train/validation/test splits for 10 European flood events.

Event ID (Country)	Split	Samples	Climate Zone
118 (Spain)	Train	342	Zone 2+3
324 (France)	Train	548	Zone 3
411 (France)	Train	149	Zone 3
427 (Sweden)	Train	5,283	Zone 3
Training Subtotal		**6,322**	**4 events**
279 (Spain)	Val	346	Zone 2+3
417 (Portugal)	Val	162	Zone 3
445 (Romania)	Val	141	Zone 2
Validation Subtotal		**649**	**3 events**
421 (France)	Test	68	Zone 3
497 (Germany)	Test	398	Zone 3
502 (Ireland)	Test	445	Zone 3
Test Subtotal		**911**	**3 events**
Total		**7,882**	**10 events, 7 countries**

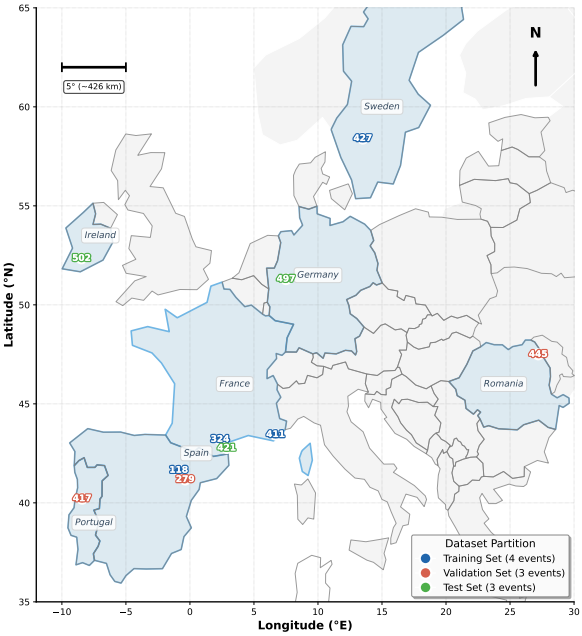


Figure 4. Geographic distribution of European flood events.

Depth estimation configuration. For depth estimation (Section 4.3), we use Event 497 (Germany, October 2020), which has the largest sample size (398 patches) and diverse terrain (urban, agricultural, forested floodplains). We select this event

for ICESat-2 validation due to temporal overlap with available altimetry data. This depth analysis is treated as a proof-of-concept for the integrated SAR-to-depth workflow. We evaluate four global DEMs: SRTM v3.0 (Farr et al., 2007) (30m), Copernicus DEM (European Space Agency, 2021) (30m, TanDEM-X), MERIT DEM (Yamazaki et al., 2017) (30m, hydrologically optimized), and FABDEM v1.2 (Hawker et al., 2022) (30m, vegetation-removed). All DEMs are harmonized to a common framework: resampled to 6.23m spacing using bilinear interpolation, reprojected to EPSG:3035, and aligned to 11,251×15,702 pixels. We preserve original vertical datums since FLEXTH uses relative elevation gradients.

4.1.2 Data Preprocessing and Implementation The KuroSiwo dataset includes preprocessed Sentinel-1 SAR imagery (VV and VH polarizations) at 10m resolution. Each sample contains multi-temporal acquisitions (two pre-flood + one flood-event) co-registered and cropped to 224×224 patches.

All experiments are implemented in PyTorch 2.5.1 with CUDA 12.8 on two NVIDIA A100 GPUs (80GB VRAM each), using random seed 42. SAR intensity values are clipped to ± 0.15 and normalized using training set statistics. Data augmentation includes geometric transformations (flips, rotations, affine) and intensity augmentations (blur, noise) via Albumentations (Buslaev et al., 2020). Test-time augmentation is applied during evaluation. CS-Mamba uses a hierarchical architecture with channel dimensions [96, 192, 384, 768] and [2, 2, 6, 2] blocks per stage. Training details are in Section 3.2.3.

4.2 Three-Class Flood Segmentation Performance

4.2.1 Baseline Comparison We compare CS-Mamba with other deep learning methods for flood segmentation: ResNet-based UNet and DeepLabV3+ (CNN architectures), FloodViT (transformer), and RSMamba (Mamba baseline). RSMamba (originally a classification model) is adapted for segmentation by replacing the classification head with a segmentation decoder and training from scratch on KuroSiwo. FloodViT is evaluated using pre-trained weights provided by KuroSiwo, achieving 43.5% mIoU, likely due to domain mismatch between pre-training and the target task. Table 2 presents the comparative performance on the test set.

Table 2. Segmentation performance comparison on test set.

Model	No Water IoU (%)	Permanent IoU (%)	Floods IoU (%)	mIoU (%)
FloodViT	87.3	9.8	33.4	43.5
RSMamba	95.7	67.5	54.1	72.4
DeepLabV3+	95.8	76.5	58.3	76.9
UNet-ResNet50	96.2	75.9	60.0	77.3
CS-Mamba (Ours)	**96.4**	**79.6**	**63.5**	**79.8**

CS-Mamba achieves the highest performance across all architectures, with 79.8% mIoU surpassing CNN methods (DeepLabV3+: 76.9%, UNet-ResNet50: 77.3%), RSMamba (72.4%), and FloodViT (43.5%). The 7.4 percentage point improvement over RSMamba stems from the architectural design: hierarchical encoder and symmetric decoder with multi-scale skip connections enable cross-scale feature fusion, combining fine-grained boundaries with semantic context. For permanent water, CS-Mamba (79.6% IoU) outperforms RSMamba (67.5%) and FloodViT (9.8%), effectively capturing temporal patterns through hierarchical encoding. For flood detection, CS-Mamba (63.5% IoU) outperforms all baselines, with 9.4% above RSMamba and 3.5-5.1% above CNN methods, suggesting that selective scanning mechanisms better capture long-range spatial patterns.

4.2.2 Detailed Performance Analysis Table 3 breaks down CS-Mamba's per-class performance. The 79.8% mean IoU exceeds validation by 3.8 percentage points. Performance varies across classes: no-water (96.4% IoU) is substantially easier than floods (63.5% IoU), reflecting the difficulty of detecting transient floods in SAR imagery where backscatter varies with water depth, surface roughness, and vegetation.

Table 3. CS-Mamba per-class performance on test set (Events 421, 497, 502; 911 samples). Metrics include Recall, Precision, F1, and IoU with test-time augmentation enabled.

Class	Recall (%)	Precision (%)	F1 (%)	IoU (%)
No Water	98.3	98.0	98.1	96.4
Permanent Water	89.4	87.9	88.6	79.6
Floods	74.0	81.7	77.7	63.5
Mean	87.2	89.2	88.1	**79.8**

Validation mIoU: 75.9% | Generalization gain: +3.8%

Recall and precision are balanced across all three classes (Table 3). Test performance (79.8%) exceeds validation (75.9%), suggesting that CS-Mamba generalizes across Continental, Oceanic, and Mediterranean climates rather than overfitting to training events.

4.2.3 Qualitative Assessment Figure 5 presents qualitative segmentation results from three test events (France, Germany, Ireland). Each row displays multi-temporal SAR inputs (Pre-event 1, Pre-event 2, Post-event in VV/VH polarizations), ground truth annotation, and CS-Mamba prediction. Predictions closely match ground truth across all three events. For Germany, CS-Mamba distinguishes permanent water (cyan) from floods (pink) with smooth flood boundaries. For Ireland, fine flood details are precisely captured. For France, accurate flood extent delineation is achieved. These visual results support the quantitative performance in Table 3.

4.3 Depth Estimation and DEM Sensitivity Analysis

4.3.1 Experimental Design To assess the effects of DEM selection and segmentation quality on depth estimation, FLEXTH is applied to Event 497 (Germany, October 2020), which contains 398 patches covering urban, agricultural, and forested terrain.

The 2×4 factorial design combines two mask sources (CS-Mamba predictions, reference annotations) with four harmonized DEMs (SRTM, Copernicus, MERIT, FABDEM) using identical FLEXTH parameters.

In the absence of ground-truth depth measurements, depth estimates from CS-Mamba masks are compared against those from reference masks. Comparisons across DEMs isolate DEM-specific effects.

Table 4. Depth estimation comparison for Event 497 (Germany): CS-Mamba predictions versus masks across four DEMs.

DEM	Pred. Mean (cm)	Ref. Mean (cm)	Diff. (cm)	Diff. (%)	Pred. Median (cm)	Ref. Median (cm)	Pred. Std (cm)
SRTM	1659.61	1686.08	-26.47	-1.57	2080.80	2015.10	953.90
MERIT	1608.05	1621.89	-13.84	-0.85	2011.63	1945.18	909.93
Copernicus	1548.25	1547.83	+0.42	+0.03	1964.57	1870.72	875.51
FABDEM	1545.64	1561.69	-16.05	-1.03	1967.05	1893.14	877.45

Pred. pixels: 1,135,276 (44.13 km²); Ref.: 1,358,115 (52.79 km²)

4.3.2 Quantitative Comparison Table 4 presents FLEXTH depth estimates across eight configurations (2 mask sources × 4 DEMs) for Event 497. Despite predicting fewer pixels (1.14M vs 1.36M, 16.4% under-prediction), CS-Mamba predictions

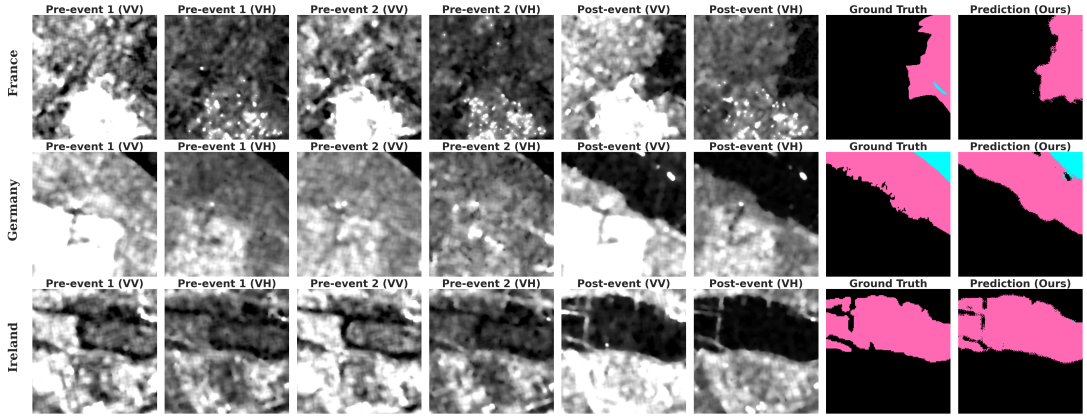


Figure 5. Qualitative flood segmentation results on test samples from three events (France, Germany, Ireland). Each row shows multi-temporal SAR inputs, ground truth annotation, and CS-Mamba prediction for three-class flood segmentation.

agree closely with reference masks, with mean depth differences within ±2% across all DEMs: SRTM (-1.57%), MERIT (-0.85%), Copernicus (+0.03%), FABDEM (-1.03%). FLEXTH also exhibits low sensitivity to DEM selection, with only 27 cm variation across products (mean depths: 1545–1686 cm). Deep learning predictions can therefore replace manual annotations for operational depth estimation.

Among the four DEMs, Copernicus exhibits near-perfect agreement (+0.03%), followed by MERIT (-0.85%), FABDEM (-1.03%), and SRTM (-1.57%). The narrow 27 cm range indicates that FLEXTH is robust to DEM preprocessing differences. Copernicus uses TanDEM-X with ±2 m vertical accuracy. MERIT's multi-error-removal algorithm, designed for hydrology, performs well despite coarser resolution. MERIT and FABDEM perform similarly, indicating that building/vegetation removal leads to smaller differences compared to reference mask estimates than unprocessed products like SRTM.

Statistical distributions show similar patterns across DEMs. Median depths (1870-2081 cm) exceed means (1548-1686 cm) by 15-25%, indicating right-skewed distributions. This pattern is typical for riverine floods, where shallow areas along river banks and margins lower the mean, while most flooded regions have moderate to deep water. Standard deviations (876-954 cm) reflect high depth variability within flooded regions. This underscores why pixel-wise depth maps are preferable to uniform depth values for flood response and damage assessment.

Figure 6 presents spatial depth patterns across DEMs. MERIT and FABDEM exhibit finer topographic detail in urban/vegetated areas due to building/vegetation removal preprocessing, while SRTM and Copernicus exhibit smoother depth gradients. Despite these visual differences, quantitative statistics remain tightly clustered within ±2%, indicating that FLEXTH's elevation averaging and interpolation mechanisms effectively normalize local variations. The algorithm's robustness to DEM texture differences supports use across different DEM products without re-tuning parameters.

4.3.3 Validation with ICESat-2 Altimetry Depth estimates are validated following the method described in Section 3.3.2 using MERIT DEM. After filtering, 19 matched points remain for Event 497. Table 5 presents validation statistics computed using metrics defined in Section 3.3.3. FLEXTH depth estimates (computed with MERIT DEM) yield RMSE of

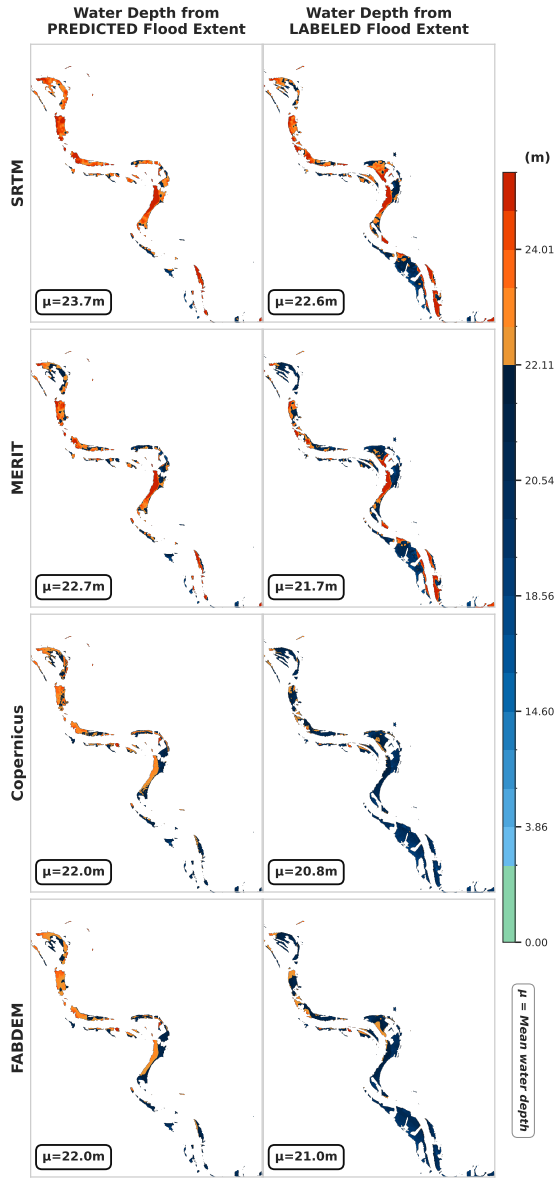


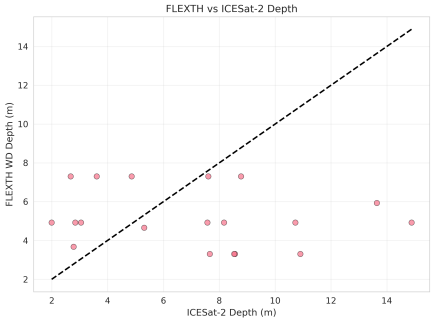
Figure 6. Flood depth maps for Event 497 (Germany, October 2020) across four global DEMs. Spatial distribution of pixel-wise depths with consistent color scales for CS-Mamba predictions and reference masks.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-493-2026 | © Author(s) 2026. CC BY 4.0 License.

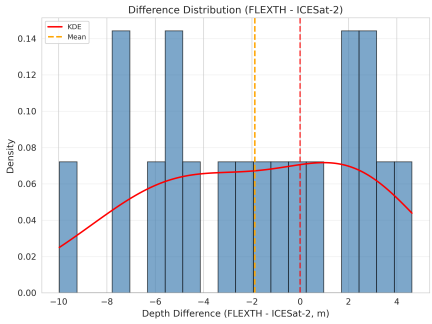
4.60 m and Bias of -1.88 m compared to ICESat-2 measurements. The negative Bias reflects systematic underestimation that arises from four factors: vertical errors in MERIT DEM (typically 2–5 m in floodplain terrain) propagate directly into ICESat-2 reference depths, FLEXTH’s inverse-distance interpolation tends to underestimate water levels near flood margins where boundary elevations are higher, and temporal offsets between ICESat-2 overpasses and SAR acquisition may introduce water level discrepancies, and ICESat-2 photon returns over water surfaces exhibit lower signal-to-noise ratios than land surfaces, with residual noise photons potentially biasing elevation measurements despite confidence filtering. While the limited sample size (19 points) constrains statistical robustness, these results provide initial independent validation for operational flood depth retrieval.

Table 5. ICESat-2 validation results for Event 497 (Germany).

Metric	Value	Units
Valid points	19	-
Bias	-1.88	m
RMSE	4.60	m
MAE	3.83	m



(a)



(b)

Figure 7. ICESat-2 validation results for Event 497 (Germany). (a) Scatter plot of FLEXTH versus ICESat-2 depths. (b) Distribution of depth differences (FLEXTH - ICESat-2).

The scatter plot (Figure 7(a)) exhibits substantial scatter around the 1:1 reference line, with most points falling below the line. ICESat-2 depths range from 2 to 14 m, while FLEXTH estimates span 3 to 7 m. The difference distribution (Figure 7(b)) exhibits a bimodal pattern with primary concentration in the negative range (-6 to -3 m) and a secondary peak around +2 m, consistent with the -1.88 m Bias. Large scatter likely arises from small sample size, spatial heterogeneity, and uncertainties in DEM elevations and FLEXTH interpolation near flood boundaries. Sample size is limited (19 points) due to spatial coverage, track geometry, and filtering requirements. ICESat-2 depth calculation depends on DEM accuracy, introducing systematic errors. Paired dry/wet ICESat-2 tracks (Betterle and Salamon, 2024) would eliminate DEM errors through differential analysis, but such pairs were unavailable for this event.

5. Discussion and Conclusion

This study presents a unified SAR-to-depth workflow combining CS-Mamba-based water body segmentation with FLEXTH geometric depth estimation. CS-Mamba achieves a mean IoU of 79.8% and demonstrates strong generalization across diverse flood events. Depth estimates derived from CS-Mamba predictions remain within ±2% of those derived from ground truth masks across four DEMs, confirming that automated segmentation is sufficiently accurate for operational applications. Initial ICESat-2 validation using MERIT DEM with 19 matched points yields RMSE of 4.60 m and Bias of -1.88 m, where the systematic underestimation is associated with DEM vertical uncertainty, boundary and interpolation effects, temporal mismatch between SAR and ICESat-2, and uncertainty in over-water photon measurements. The quantitative results support the practical feasibility of the proposed approach.

This work is subject to several constraints. ICESat-2 validation is constrained to Event 497 because the remaining two test events (France, Ireland) lack temporally coincident altimetry tracks within their respective flood periods, a restriction imposed by ICESat-2’s 91-day repeat cycle and narrow ground track spacing. Within Event 497, strict spatial and quality filtering further reduces the number of usable photon-to-pixel matches to 19 points, which limits the statistical power of the validation. ICESat-2 photon returns over water surfaces also exhibit lower confidence than over land, which adds uncertainty to reference depth estimates. Nevertheless, the core finding that CS-Mamba and reference masks produce depth estimates within $\pm 2\%$ across all four DEMs does not depend on ICESat-2. The ICESat-2 comparison provides additional independent support and is not the only basis for depth accuracy assessment. FLEXTH also assumes a continuous water surface under steady-state conditions, which may not fully capture transient hydraulic gradients or backwater effects in complex river networks. Incorporating in-situ water level measurements from gauging stations, where available, could constrain water surface interpolation and account for hydraulic gradients. For rapidly flowing channels requiring transient hydraulic modeling, coupling with physics-based hydrodynamic solvers remains necessary. Finally, segmentation and depth experiments are limited to European flood events; performance on tropical or arid-region floods with different land cover and terrain characteristics has not yet been tested.

Future work should extend validation to non-European regions using KuroSiwo’s 43 global events, expand ICESat-2 verification through paired dry/wet tracks that eliminate DEM-related errors, and explore end-to-end integration of geometric depth estimation into the segmentation pipeline to reduce error accumulation from the current two-stage design.

Acknowledgements

This study was supported by JSPS KAKENHI (21H05001, 23K13419), the Project Grant from the Co-creation Center for Disaster Resilience, IRIDeS, Tohoku University (ID: 1-DT020), the JAXA EO-RA4 Project (ID: ER4A2N09), JST SICORP (JPMJSC2311), JST CRONOS (JPMJCS25K5), and the SIP Program of CSTI (JPJ012289).

References

Amitrano, D., Di Martino, G., Di Simone, A., Imperatore, P., 2024. Flood Detection with SAR: A Review of Techniques and Datasets. *Remote Sensing*, 16(4), 656.

Bai, Y., Wu, W., Yang, Z., Yu, J., Zhao, B., Liu, X., Yang, H., Mas, E., Koshimura, S., 2021. Enhancement of Detecting Permanent Water and Temporary Water in Flood Disasters by Fusing Sentinel-1 and Sentinel-2 Imagery Using Deep Learning Algorithms: Demonstration of Sen1Floods11 Benchmark Datasets. *Remote Sensing*, 13(11), 2220.

Betterle, A., Salamon, P., 2024. Water depth estimate and flood extent enhancement for satellite-based inundation maps. *Natural Hazards and Earth System Sciences*, 24(8), 2817–2840.

Bountos, N. I., Sdraka, M., Zavras, A., Karasante, I., Karavias, A., Herekakis, T., Thanasou, A., Michail, D., Papoutsis, I., 2023. Kuro siwo: 33 billion m^2 under the water. a global multi-temporal satellite dataset for rapid flood mapping. *Advances in Neural Information Processing Systems*, 37, 38105–38121.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A. A., 2020. Albumentations: Fast and Flexible Image Augmentations. *Information*, 11(2), 125.

Chen, H., Song, J., Han, C., Xia, J., Yokoya, N., 2024a. ChangeMamba: Remote Sensing Change Detection With Spatiotemporal State Space Model. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–16.

Chen, K., Chen, B.-Y., Liu, C., Li, W., Zou, Z., Shi, Z., 2024b. RSMamba: Remote Sensing Image Classification With State Space Model. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *European Conference on Computer Vision (ECCV)*, Springer, 801–818.

Cohen, S., Peter, B. G., Haag, A., Munasinghe, D., Moragoda, N., Narayanan, A., May, S., 2022. Sensitivity of Remote Sensing Floodwater Depth Calculation to Boundary Filtering and Digital Elevation Model Selections. *Remote Sensing*, 14(21), 5313.

Cohen, S., Raney, A., Munasinghe, D., Loftis, J. D., Molthan, A., Bell, J., Rogers, L., Galantowicz, J., Brakenridge, G. R., Kettner, A. J. et al., 2019. The Floodwater Depth Estimation Tool (FwDET v2.0) for improved remote sensing analysis of coastal flooding. *Natural Hazards and Earth System Sciences*, 19(9), 2053–2065.

European Space Agency, 2021. Copernicus DEM: Global digital elevation model. Available at `https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model`.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L. et al., 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2).

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J., 2022. A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2), 024016.

Jamali, A., Roy, S. K., Beni, L. H., Pradhan, B., Li, J., Ghamisi, P., 2024. Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. *International Journal of Applied Earth Observation and Geoinformation*, 127, 103662.

Ma, X., Zhang, X., Pun, M.-O., 2024. RS3Mamba: Visual State Space Model for Remote Sensing Image Semantic Segmentation. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5.

Neuenschwander, A. L., Pitts, K. L., 2019. The ATLAS lidar on ICESat-2. *Remote Sensing of Environment*, 221, 247–259.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 234–241.

Ruan, J., Xiang, S., 2024. VM-UNet: Vision Mamba UNet for Medical Image Segmentation. *arXiv preprint arXiv:2402.02491*.

Saleh, T., Holail, S., Xiao, X., Xia, G.-S., 2024. High-precision flood detection and mapping via multi-temporal SAR change analysis with semantic token-based transformer. *International Journal of Applied Earth Observation and Geoinformation*, 131, 103991.

Sharma, N., Saharia, M., 2025. DeepSARFlood: Rapid and Automated SAR-based flood inundation mapping using Vision Transformer-based Deep Ensembles with uncertainty estimates. *Science of Remote Sensing*, 13, 100203.

Xu, K., Fang, J., Fang, Y., Sun, Q., Wu, C.-C., Liu, M., 2021. The Importance of Digital Elevation Model Selection in Flood Simulation and a Proposed Method to Reduce DEM Errors: A Case Study in Shanghai. *International Journal of Disaster Risk Science*, 12, 890–902.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O’Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., Bates, P. D., 2017. A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, 44(11), 5844–5853.

Yuan, W., Cai, Y., Li, J., 2024. Hybrid Network-Based Automatic Seamline Detection for Orthophoto Mosaicking. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–16.

Yuan, W., Ran, W., Shi, X., Shibasaki, R., 2023a. Multiconstraint Transformer-Based Automatic Building Extraction From High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 9763–9778.

Yuan, W., Yuan, X., Cai, Y., Shibasaki, R., 2023b. Fully Automatic DOM Generation Method Based on Optical Flow Field Dense Image Matching. *Geo-Spatial Information Science*, 26(2), 145–162.

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X., 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *International Conference on Machine Learning*.