IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 May 2026, accepted 31 May 2026, date of publication 8 June 2026, date of current version 11 June 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3701151

RESEARCH ARTICLE

XTR: An Explainable Transformer-Based Retracker for CryoSat-2 Altimetry

ALIREZA DEHGHANPOUR, DANIAL BARAZANDEH, AND GABRIEL ZACHMANN
Faculty of Mathematics and Computer Science, University of Bremen, 28359 Bremen, Germany

Corresponding author: Alireza Dehghanpour (alireza.dehghanpour@uni-bremen.de)

This Open Access publication has been made possible by funding from the publication fund of the State and University Library Bremen (SuUB). This work was supported by the Data Science Center of the University of Bremen (DSC@UB), funded by the State of Bremen.

ABSTRACT Satellite radar altimetry is essential for monitoring the mass balance of polar ice sheets and their contribution to sea-level rise. However, over the dry-snow facies of Greenland and Antarctica, Ku-band pulses penetrate the firn, undergo volume scattering, and create time-dependent elevation biases. Traditional physical-model and threshold retrackers only partially compensate for this effect and usually require empirical post-processing based on waveform parameters such as backscatter and leading-edge width (LEW). Recently, deep-learning (DL) retrackers—especially convolutional neural networks (CNNs) trained on large ensembles of simulated waveforms—have achieved state-of-the-art accuracy by mapping echoes directly to surface range, yet they provide limited intrinsic interpretability and only indirect physical insight. This study presents XTR, an explainable Transformer-based retracker for CryoSat-2 low-resolution mode (LRM) waveforms. XTR couples a lightweight convolutional stem with multi-head self-attention to capture both local leading-edge structure and long-range waveform context, and uses mild waveform augmentation during training to improve robustness. We evaluate XTR using simulated waveforms and CryoSat-2 LRM crossover statistics over interior East Antarctica from 2020 to 2025 under a common retracker-only processing framework. On simulated data, XTR reduces the mean squared error (MSE) by approximately 35% relative to the classical threshold first-maximum retracker algorithm (TFMRA), although its error remains higher than that of the augmentation-matched AWI-ICENet1 CNN baseline. On real CryoSat-2 data, XTR lowers the crossover standard deviation by approximately 40% relative to AWI-ICENet1, while maintaining modest, region-dependent median offsets. Explainability analyses based on self-attention, SmoothGrad saliency, Integrated Gradients, and occlusion-based sensitivity tests indicate that XTR anchors its range estimate near the surface-dominated leading edge while using trailing-waveform structure as contextual information. Overall, our findings suggest that attention-based retrackers complement CNN-based schemes by offering competitive accuracy, intrinsic interpretability, and improved crossover stability in this retracker-only evaluation.

INDEX TERMS CryoSat-2, radar altimetry, transformers, self-attention, retracking, explainable AI (XAI).

I. INTRODUCTION

Satellite radar altimetry has transformed our ability to monitor the mass balance of the Greenland and Antarctic ice sheets over the past three decades [1], [2], [3], [4]. By providing spatially extensive and temporally repeated measurements of surface-elevation change (SEC), satellite altimetry complements gravimetry and input–output mass-balance estimates and has become a key component of sea-level assessments and climate studies. The European Space Agency's (ESA) CryoSat-2 mission, launched in 2010, was designed for ice-sheet monitoring through a Ku-band radar altimeter (SIRAL) that operates in synthetic-aperture interferometric mode over ice margins and in pulse-limited low-resolution mode (LRM) over the interior. CryoSat-2 has since become a key mission for monitoring polar

The associate editor coordinating the review of this manuscript and approving it for publication was Xin Sun.

elevation change and ice-sheet mass imbalance [5], [6]. Over the dry-snow facies of the East Antarctic Plateau and high-elevation Greenland, however, the interaction between Ku-band pulses and the firn column remains a major source of uncertainty [7], [8]. Unlike laser altimeters, which sense the physical snow surface, Ku-band echoes include significant volume scattering; the effective scattering horizon can lie centimeters to meters below the surface, generating penetration-induced range biases. Seasonal and interannual variations in accumulation, temperature, and microstructure modulate this depth and can lead to elevation errors exceeding tens of centimeters [9]. These penetration effects complicate the interpretation of long-term SEC, especially during anomalous events such as the 2012 Greenland melt season [10], [11].

A. LIMITATIONS OF TRADITIONAL RETRACKERS

Retrieving surface range from the distorted radar waveform—a process known as *retracking*—is therefore a central problem in ice-sheet altimetry. Traditional retrackers can be broadly categorized into physical-model fits and empirical threshold approaches. Physical retrackers, such as ICE-2 and related model-based schemes, fit analytical or semi-empirical echo models, often based on or inspired by Brown-type formulations, to retrieve parameters such as surface range, backscatter, leading-edge width (LEW), and trailing-edge slope [9], [12], [13], [14], [15]. Threshold-type retrackers, such as the threshold first-maximum retracker algorithm (TFMRA) or offset center-of-gravity (OCOG) schemes, instead identify a characteristic threshold crossing on the leading edge relative to the waveform peak or use a thresholded center-of-gravity. These methods offer robustness and computational efficiency, but at the expense of limited physical realism in strongly penetrating firn conditions.

Over ice sheets, both classes of methods face well-known limitations. First, the assumed functional forms often cannot fully represent the combined surface and volume response of the firn, especially in regions with complex stratigraphy or mixed scattering regimes. Second, the retrieved elevations exhibit spurious temporal variability associated with changes in snowpack properties rather than true surface motion. For this reason, altimetric time series are typically subjected to extensive empirical post-processing, in which correlations between elevation residuals and waveform parameters such as backscatter (σ^0) or leading-edge width (LEW) are exploited to construct penetration corrections [10], [16], [17], [18]. Although successful in many applications, these corrections implicitly assume a stationary relationship between waveform shape and penetration depth and may break down during extreme or rapidly evolving events [10].

B. DEEP LEARNING RETRACKERS AND THE INTERPRETABILITY CHALLENGE

Deep learning (DL) provides an alternative paradigm in which the mapping from waveform to surface range is learned directly from data, without prescribing a particular functional form. In the context of ice-sheet altimetry, Helm et al. introduced the AWI-ICENet1 retracker, a deep convolutional neural network (CNN) trained on a large ensemble of simulated waveforms spanning a wide range of surface slopes, topographic settings, and attenuation conditions [19]. By optimizing the network to predict the true surface range in the forward model, they reported that AWI-ICENet1 implicitly corrects penetration-related biases, reduces the need for explicit LEW/backscatter-based post-processing, and performs strongly on both simulated and real data.

Optimizing a deep CNN on large ensembles of simulated waveforms, however, does not automatically guarantee uniformly optimal behavior across all real-data regimes. The forward model can only approximate the true measurement physics, and several sources of variability—for example residual timing jitter, non-Gaussian noise, or sub-footprint roughness—are difficult to capture perfectly. In our experiments we therefore treat AWI-ICENet1 as a reproducible baseline: we re-implement the published architecture and training setup and train it once on the publicly released simulation dataset [20]. As we show later using CryoSat-2 crossovers over interior regions such as Vostok and Queen Maud Land, this re-trained AWI-ICENet1 instance still exhibits substantial scatter in crossover height differences within our evaluation framework, and it remains unclear which waveform regions most strongly drive the CNN's range estimates. Despite these successes, the use of deep CNNs raises important questions about interpretability and trustworthiness in geoscientific applications. In particular, it is difficult to determine which parts of the waveform (e.g., the first samples on the leading edge versus the trailing volume-scattering tail) are most influential for the predicted range, or whether the network exploits artifacts specific to the simulation setup. Such limited intrinsic interpretability is increasingly viewed as problematic in high-stakes Earth system science, where scientific insight and robust extrapolation are as important as predictive performance [21], [22]. At the same time, convolutional models can also be analyzed with post-hoc attribution methods such as Integrated Gradients and Grad-CAM; however, these methods remain external explanation tools rather than intrinsic sequence-wise weighting mechanisms [23], [24]. Furthermore, standard CNNs rely on local convolutional receptive fields and pooling operations; although deeper layers expand the effective receptive field, they do not provide an explicit, physically interpretable weighting over the waveform, which makes it challenging to model long-range dependencies between distant waveform regions in a transparent and principled way.

C. TOWARDS EXPLAINABLE TRANSFORMER-BASED RETRACKING

Recent advances in sequence modeling have shown that Transformer architectures with self-attention can capture long-range dependencies in a data-driven manner [25]. In contrast to CNNs, which rely on local receptive fields and pooling operations, self-attention enables each position in a sequence to directly attend to all others, assigning data-dependent, learnable weights. This makes Transformers particularly attractive for radar altimetry waveforms, where local features such as the sharp rise of the leading edge coexist with broader structures associated with volume scattering. In this study, "local" refers to short-range waveform structure over neighboring bins around the onset and rise of the leading edge, where the echo shape is most sensitive to noise, speckle, and small alignment perturbations. In XTR, this information is first captured by the stacked convolutional stem. Since the stem consists of six Conv1D layers (kernel size 3, stride 1), its effective receptive field is approximately 13 bins (about 6.1 m). By contrast, "long-range" refers to dependencies between waveform regions separated by tens of bins, such as the relationship between the near-surface onset and the delayed trailing-waveform structure associated with volume scattering. These longer-range dependencies are modeled by self-attention, which has access to the full 128-bin waveform span (approximately 60 m).

In this paper, we introduce XTR, an explainable deep-learning retracker for CryoSat-2 LRM waveforms. Our main contributions are as follows:

- We propose XTR, a lightweight CNN-stem-plus-Transformer retracker for penetration-affected CryoSat-2 LRM waveforms that jointly captures local leading-edge structure and long-range waveform context.
- We develop an explainable AI (XAI) framework for radar altimetry retracking by combining gradient-based saliency, attention-based diagnostics, and occlusion-based sensitivity tests to relate retracking decisions to physically meaningful waveform regions.
- We provide a controlled comparison against classical and deep-learning retrackers on both simulated waveforms and real CryoSat-2 crossover data, showing that waveform-level regression accuracy and crossover consistency capture different aspects of retracker behavior.
- We perform ablation and augmentation-control experiments, including an augmentation-matched AWI-ICENet1 baseline, an XTR model trained without augmentation, and CNN-only and Transformer-only variants, to clarify the roles of augmentation and architectural design in the observed performance.

Unlike prior Transformer-based retrackers—such as TransR, which was developed for Sentinel-3 SRAL inland-water echoes [26]—XTR targets penetration-affected CryoSat-2 LRM waveforms and is evaluated in a retracker-only framework using a common correction set. To the best of our knowledge, this is among the first studies to develop a Transformer-based retracker specifically for CryoSat-2 LRM ice-sheet waveforms and to couple it with a joint gradient-, attention-, and perturbation-based explainability framework linked to crossover statistics over interior Antarctica. Beyond CryoSat-2, the same framework is transferable to future altimetry missions and to other acquisition modes where waveform distortions and penetration effects challenge traditional retrackers. We emphasize that our real-data evaluation compares the retrackers under the same geophysical corrections, quality filters, and height definition, so that the reported differences primarily reflect the retracked range estimate itself rather than downstream processing choices.

II. RELATED WORK

A. CLASSICAL RADAR ALTIMETRY RETRACKERS

Early radar altimetry processing relied on parametric waveform models and empirical criteria to estimate surface range from the received echo. Martin et al. [14] provided an early analysis of continental ice-sheet waveforms using Brown-type models to derive range and geophysical parameters. Building on this foundation, the ICE-2 retracking algorithm adapted Brown-type modeling concepts to inland-ice waveforms and was adopted in processing chains for missions such as Envisat and CryoSat [9], [15]. In parallel, TFMRA and OCOG schemes were developed as robust and computationally efficient alternatives to model fitting. Over the dry-snow facies, the common assumption of a single dominant surface return is often violated, because Ku-band signals can penetrate into the firn and produce substantial volume scattering [7], [8]. The resulting mixture of surface and subsurface contributions introduces elevation biases that vary with accumulation rate, temperature, and snow microstructure [9]. As a consequence, multi-year time series derived from TFMRA, OCOG, or related schemes often require empirical post-processing based on correlations between elevation residuals and waveform parameters such as backscatter (σ^0) and leading-edge width (LEW) [10], [16], [17], [18]. While effective in many cases, these corrections typically assume a stable relationship between waveform shape and penetration depth and may degrade during anomalous or rapidly evolving conditions [10]. Even after such tuning, crossover statistics over low-accumulation interior basins can still exhibit substantial residual scatter, suggesting that classical approaches do not always fully separate transient penetration effects from true surface-elevation change. Reviews of Antarctic radar altimetry similarly emphasize that waveform shape, surface slope, penetration, and scattering regime jointly control the reliability of elevation retrievals over ice sheets [27].

B. DEEP LEARNING FOR ALTIMETRY WAVEFORMS

Motivated by the limitations of fixed-form retrackers, several studies have explored neural networks for waveform classification and retracking. Mattes [28] demonstrated that feed-forward and convolutional networks can learn to distinguish waveform classes and infer retracking offsets using simulated and real altimetry data. Memarian Sorkhabi et al. [29] combined wavelet decomposition with a CNN to retrack inland-water waveforms and reconstruct lake levels over Lake Urmia, showing that data-driven models can outperform classical threshold approaches in challenging regimes.

Transformer-based architectures have also been introduced for waveform retracking. Tao et al. proposed TransR, a Transformer-based retracking network for Sentinel-3 SRAL lake-level monitoring and constructed a dedicated dataset for training and validation [26]. More generally, recent retracking reviews show that waveform retracker design remains strongly application-dependent across altimetry domains, which supports the need for domain-specific architectures rather than assuming that one retracking formulation transfers unchanged across inland-water, coastal, and ice-sheet waveform regimes [30]. Although Transformer-based retrackers such as TransR demonstrate the promise of attention mechanisms for waveform analysis, the present work addresses a substantially different problem setting. TransR was developed for Sentinel-3 SRAL lake-level retracking, whereas XTR targets CryoSat-2 LRM waveforms over ice sheets, where firn penetration and volume scattering strongly distort the trailing part of the waveform. In particular, unlike inland-water retracking settings, the present problem is dominated by penetration-induced distortion of the trailing part of the waveform over dry-snow firn. This motivates the use of a surface-focused attention analysis and an attenuation-stratified explainability evaluation. In other domains such as coastal altimetry, hybrid schemes that integrate machine-learning components into optimized waveform models have been shown to improve sea-surface-height retrieval under complex conditions [31]. These studies highlight the potential of sequence models to represent long-range waveform structure beyond what is readily captured by conventional model fits. More broadly, the success of Transformer-style architectures in vision and sequence modeling has motivated their adoption in remote sensing and geoscientific applications where long-range contextual information is important [32], [33]. Recent attention-based studies further illustrate their growing role in Earth-observation problems, including multimodal Earth observation [34], spatio-temporal prediction [35], and geographic representation learning under distribution shifts [36].

In the context of ice-sheet altimetry, Helm et al. introduced AWI-ICENet1, a deep CNN retracker trained on a large ensemble of simulated waveforms spanning a broad range of slopes, topography, and bulk attenuation regimes [19], [20]. Their results suggest that AWI-ICENet1 learns to discount delayed volume-scattering contributions and can reduce penetration-related artifacts in CryoSat-2 elevation time series. At the same time, model optimization on simulated ensembles does not automatically guarantee uniformly optimal behavior across all real-data regimes, and convolutional architectures provide limited interpretability regarding which waveform regions drive the predicted range.

C. EXPLAINABLE AI IN REMOTE SENSING AND GEOSCIENCES

The increasing use of deep learning in Earth observation and geosciences has motivated a growing interest in explainable AI (XAI). Recent reviews discuss methodological, operational, and domain-specific challenges for explaining remote-sensing models [37], [38], [39], [40], [41]. Reichstein et al. [21] argued that combining DL with process understanding is essential for trustworthy Earth system science, and Samek et al. [22] reviewed gradient-based and perturbation-based explanation techniques. Attention-based sequence models are particularly attractive for interpretability because they provide explicit, learnable weights over input positions, although these weights should be interpreted as diagnostic indicators rather than complete causal explanations [22], [25]. Gradient-based saliency methods such as SmoothGrad can further highlight which parts of the input most strongly influence the model output [22], [42]. However, explainability analyses that explicitly connect attention and saliency diagnostics to retracking behavior and crossover performance in radar altimetry remain limited. Existing DL retrackers for ice sheets, such as AWI-ICENet1, rely on convolutional architectures without explicit attention mechanisms [19], and most evaluations emphasize waveform-level errors and large-scale elevation-change products rather than interpretable retracker behavior in strongly penetrating interior basins. This motivates the present study: we adapt a Transformer encoder architecture [25] to CryoSat-2 LRM waveforms and exploit its self-attention, together with gradient-based saliency and occlusion-based sensitivity tests, to derive physically consistent explanations of retracking decisions and relate these diagnostics to crossover statistics over interior Antarctica.

III. METHODOLOGY

A. SIMULATED TRAINING DATASET AND PREPROCESSING

To enable a controlled comparison with a state-of-the-art retracker, we use the simulated waveform dataset released for AWI-ICENet1 [19], [20]. The dataset contains approximately 3.8×10^6 simulated waveforms generated using a refined Brown-type radar altimeter forward model with an additional depth-dependent attenuation term to represent firn volume scattering. Simulations were produced at 1000 randomly selected Antarctic locations using local topography from the REMA digital elevation model (DEM) [20], spanning nearly flat to moderately sloping surfaces and a broad range of attenuation conditions parameterized by L_A, with thermal and speckle noise consistent with the CryoSat-2 SIRAL instrument. Each sample consists of a 128-bin waveform $\mathbf{x}$ and a scalar label y representing the reference surface range bin. Before being fed to the network, each waveform is peak-normalized:

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\max(\mathbf{x})}. \quad (1)$$

This preserves the relative waveform shape while removing absolute amplitude scaling. We randomly split the full dataset into training, validation, and test subsets with proportions 68%, 12%, and 20%, respectively. This same split is used for XTR, AWI-ICENet1, and all ablation variants to ensure a controlled comparison. Prediction errors are initially computed in units of range bins and converted to meters using the CryoSat-2 LRM bin spacing $\Delta r = 0.46875$ m.

$$\mathrm{RMSE}_{\mathrm{m}} = \Delta r\, \mathrm{RMSE}_{\mathrm{bin}}, \qquad \mathrm{MSE}_{\mathrm{m}^2} = \Delta r^2\, \mathrm{MSE}_{\mathrm{bin}}. \quad (2)$$

1) WAVEFORM AUGMENTATION

During training, we apply mild, physically plausible augmentations to improve robustness to noise, small misalignments, and residual simulation–reality mismatch. Augmentations are applied on-the-fly to peak-normalized waveforms with probability $p_{\mathrm{aug}} = 0.95$. We apply (i) multiplicative amplitude jitter drawn uniformly from $\pm 8\%$ to emulate small gain/backscatter variations; (ii) additive Gaussian noise with standard deviation 0.015 (in normalized amplitude units); (iii) a circular shift of up to ± 2 bins to emulate residual timing jitter; and (iv) with probability 0.35, a short three-bin smoothing kernel implemented as a depthwise 1-D convolution to emulate mild low-pass broadening. After augmentation, amplitudes are clipped to [0, 1.2] to avoid unphysical excursions. These augmentations are conservative by design: they preserve leading-edge structure while exposing the model to plausible perturbations. The same augmentation policy is used for XTR, the AWI-ICENet1 baseline, and the ablation variants.

2) LOCAL FEATURE ENCODING (CNN STEM)

Radar altimetry waveforms exhibit strong local structure around the onset of the leading edge, which can be corrupted by speckle, thermal noise, and small alignment errors. As defined in Sect. I-C, XTR combines local leading-edge structure with long-range waveform context. To robustly capture local patterns prior to global sequence modeling, XTR applies a lightweight convolutional stem. The input sequence of length $T = 128$ is processed by six one-dimensional convolutional layers with filter sizes (64, 64, 96, 96, 128, 128), kernel size $k = 3$, unit stride, and same padding. Each layer is followed by batch normalization, a ReLU activation, and dropout with rate 0.05. The resulting feature sequence is mapped to the Transformer dimension d_{model} via a point-wise linear projection and passed to the self-attention encoder.

B. XTR ARCHITECTURE: A TRANSFORMER-BASED RETRACKER WITH LOCAL FEATURE ENCODING

Motivated by the need to capture long-range dependencies along the waveform, we adopt a Transformer encoder architecture [25] tailored to one-dimensional radar altimetry waveforms and refer to the resulting model as XTR (Explainable Transformer Retracker). The encoder consists of L stacked Transformer blocks with model dimension d_{model}. Let $\mathbf{z}_l \in \mathbb{R}^{T \times d_{\mathrm{model}}}$ denote the input to the l-th Transformer block, where $l = 1, \ldots, L$, $T = 128$ is the number of range bins, and d_{model} is the model dimension.

Each block follows a pre-norm structure with a multi-head self-attention (MHSA) sub-layer and a position-wise feed-forward network (FFN), both with residual connections:

$$\tilde{\mathbf{z}} = \mathbf{z}_l + \mathrm{MHSA}\big(\mathrm{LayerNorm}(\mathbf{z}_l)\big), \quad (3)$$

$$\mathbf{z}_{l+1} = \tilde{\mathbf{z}} + \mathrm{FFN}\big(\mathrm{LayerNorm}(\tilde{\mathbf{z}})\big). \quad (4)$$

After the CNN stem, the feature sequence is projected to d_{model} and combined with a learnable positional embedding of length $T = 128$ to encode range-bin position. The enriched sequence is then processed by a stack of Transformer encoder blocks. After this stack, we apply a final layer normalization and global average pooling over the range dimension to obtain a single d_{model}-dimensional representation, which is then passed to the regression head.

1) MULTI-HEAD SELF-ATTENTION

Given query, key, and value tensors $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$, scaled dot-product attention is defined as

$$\mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}}\right)\mathbf{V}, \quad (5)$$

where d_k is the key dimension. The resulting (per-head) attention-weight matrices $\mathbf{A} \in \mathbb{R}^{T \times T}$ provide explicit, learnable importance weights over waveform positions and form the basis of our attention-based analysis.

2) POSITION-WISE FEED-FORWARD NETWORK

Each block also contains a position-wise FFN that operates independently at each range bin:

$$\mathrm{FFN}(\mathbf{x}) = \mathbf{W}_2(\mathrm{Swish}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)) + \mathbf{b}_2, \quad (6)$$

with intermediate dimension $4d_{\mathrm{model}}$.

We use $L = 6$ Transformer encoder blocks, a model dimension of $d_{\mathrm{model}} = 128$, and $H = 8$ attention heads. We apply dropout (0.1) after the MHSA and FFN sub-layers, followed by a two-layer dense head (256 and 128 units, ReLU activations, dropout rates 0.1 and 0.05) and a final linear regression output.

C. TRAINING SETUP

XTR is implemented in Python 3.10 using TensorFlow 2.x/Keras. We train the model with the Adam optimizer using a base learning rate of 1×10^{-4} and apply global-norm gradient clipping with a threshold of 1.0 for stability. The training objective is the mean squared error (MSE) between the predicted and reference surface range, expressed in range-bin units. Training is performed for up to 30 epochs with a batch size of 128. We use a *ReduceLROnPlateau* scheduler to reduce the learning rate by a factor of 0.5 when the validation MSE plateaus (patience of 4 epochs), with a minimum learning rate of 1×10^{-5}. Early stopping monitors the validation MSE with a patience of 20 epochs, and the best-performing model is selected via a checkpoint callback based on the validation MSE. To quantify generalization, we report MSE, RMSE, and MAE on the test set in range-bin units, and convert errors to meters using Eq. (2).

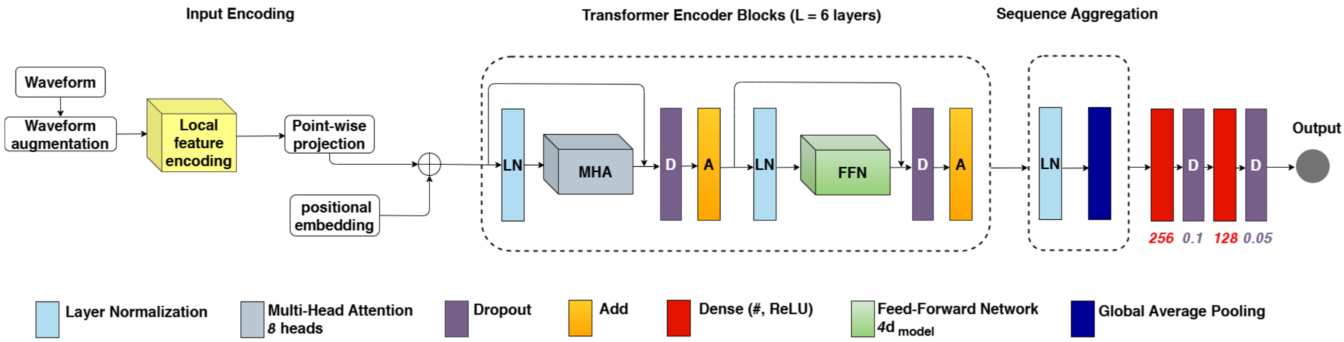


FIGURE 1. Overview of the proposed XTR architecture for CryoSat-2 LRM waveform retracking. A length-128 waveform is first processed by a local feature encoding stage (CNN stem) that enhances leading-edge structure while preserving temporal resolution. The resulting sequence is projected to a $d_{\text{model}} = 128$ embedding and combined with a learnable positional embedding. Six pre-norm Transformer encoder blocks, each with eight-head self-attention and a position-wise feed-forward network with $4d_{\text{model}}$, model global context along the waveform. A final layer normalization, global average pooling, and a two-layer dense head (256 and 128 units with ReLU activations and dropout rates 0.1/0.05) map the sequence representation to a single scalar surface-range estimate. During training only, mild waveform augmentations (additive noise, scale jitter, small bin shifts, and occasional smoothing) are applied to improve robustness.

TABLE 1. Main hyperparameters used for XTR training.

Hyperparameter	Value
Input length	128 bins
CNN stem filters	(64, 64, 96, 96, 128, 128)
CNN kernel size	3
Transformer blocks	6
Model dimension d_{model}	128
Attention heads	8
FFN expansion	$4d_{\text{model}}$
Optimizer	Adam
Initial learning rate	1×10^{-4}
Gradient clipping	1.0
Batch size	128
Maximum epochs	30
Output activation	Linear
LR scheduler	ReduceLROnPlateau
Early stopping patience	20
Augmentation probability	0.95
Amplitude jitter	±8%
Additive noise std	0.015
Shift range	±2 bins
Smoothing probability	0.35

D. AWI-ICENET1 BASELINE IMPLEMENTATION

To provide a reproducible deep learning baseline for comparison, we re-implement the AWI-ICENet1 convolutional neural network (CNN) retracker introduced by Helm et al. [19] using TensorFlow/Keras and train it from scratch on the same public simulation dataset [20]. Our implementation follows the published one-dimensional convolutional architecture, consisting of six convolutional blocks with increasing channel counts (64, 96, 128, 160, 196, 228). Each block comprises two or three Conv1D layers with kernel size 3, ReLU activations, batch normalization, and max pooling, followed by dropout. A fully connected regression head with 128 hidden units maps the extracted features to a scalar surface-range estimate. Input preprocessing is identical to that used for XTR: each 128-bin waveform is peak-normalized. We use the same random train/validation/test split (68%/12%/20%) as for XTR to ensure a controlled and directly comparable forward-model evaluation. The AWI-ICENet1 baseline is trained using the Adam optimizer with a learning rate of 1×10^{-4}, a batch size of 128, and mean squared error loss on the surface-range label for 30 epochs using the same waveform augmentation policy as XTR. We do not use the original AWI-trained weights and do not reproduce the K-fold model selection procedure described in [19]. Instead, we use this single trained instance as the consistent, re-trained AWI-ICENet1 baseline throughout this study. We emphasize that the goal of this study is not to reproduce the full original AWI operational processing chain, but to establish a controlled and reproducible baseline under the same input representation, training/validation/test split, and evaluation framework used for XTR. We further note that the original AWI-ICENet1 trained weights were not publicly released. For this reason, we use a single re-trained AWI-ICENet1 instance rather than the original published model-selection procedure or externally trained weights. This allows a more transparent comparison of retracker behavior under a common experimental protocol. Throughout this study, AWI-ICENet1 refers to our re-trained, augmentation-matched implementation used for controlled comparison with XTR.

E. APPLICATION TO CRYOSAT-2 LRM DATA AND CROSSOVER ANALYSIS

Beyond the simulated training experiment, we apply XTR to real CryoSat-2 LRM measurements and evaluate its performance using standard crossover statistics. For this purpose, we process the ESA Baseline-E CryoSat-2 LRM record from January 2020 to October 2025 over Antarctica using a modular, parallelized processing chain.

In a first step, we construct a compact Level-1B database by reading the stack-averaged 128-bin power waveforms together with their geolocation, acquisition time, orbit parameters, and the 1 Hz geophysical corrections, which are mapped to 20 Hz and combined into a single correction term C_{geo}. In a second step, these L1B records are collocated with the corresponding ESA Level-2I (L2I) ice products. This collocation provides the operational ESA retracker ranges (ICE-1 and ICE-2), their elevation estimates, and associated surface-type and quality flags. Temporal and spatial matching ensures that XTR and the ESA retrackers are always evaluated on the same set of footprints.

For each collocated footprint, the peak-normalized L1B waveform is passed through the trained XTR network to obtain a predicted surface range in units of range bins. This range is converted to meters using the CryoSat-2 LRM range-bin spacing of approximately 0.46875 m and transformed into surface height using the common convention given in (7). Applying this projection consistently to the ESA ICE-1 and ICE-2 ranges, to the AWI-ICENet1 retracker, and to XTR yields along-track elevation time series that share the same reference frame and geophysical correction set, differing only in the underlying retracking algorithm.

The real-data processing chain used in this study is intentionally limited to a self-consistent retracking comparison under a common geometric height definition and correction set. In particular, our pipeline does not reproduce the full operational processing chain of AWI-ICENet1, which includes additional steps such as footprint relocation, slope-related corrections, and further empirical conditioning tailored to ice-sheet elevation-change products. Instead, all retrackers considered here (ESA ICE-1/ICE-2, an AWI-ICENet1 reimplementation, and XTR) are evaluated using identical geophysical corrections, quality filtering, and height definitions, differing only in the retracked range term. As a result, the crossover statistics reported in this study quantify the internal consistency of the retracking algorithms themselves rather than the performance of fully optimized geophysical height products. Crossover statistics are computed by intersecting ascending and descending ground tracks within a prescribed spatial and temporal window and forming height differences between the two passes. Valid crossovers are obtained from line-segment intersections of ascending and descending tracks after common geometric filtering, including limits on temporal separation, endpoint proximity, and intersection angle; heights are then linearly interpolated to the crossover location on both tracks, and monthly statistics are computed after the common $|\mathrm{CPE}| \leq 5$ m editing threshold. We evaluate performance over the interior Antarctic LRM domain and two geodetically stable regions: Vostok and Queen Maud Land. For each retracker, we summarize the distribution of crossover height differences Δh by the standard deviation and the normalized median absolute deviation (NMAD) over January 2020–October 2025. These crossover statistics are later used in the results section to compare XTR against the ESA retrackers and AWI-ICENet1 and to assess the impact of the learned, attention-based retracking on large-scale elevation-change studies.

For the evaluation on real CryoSat-2 data we compute monthly crossover elevation differences between ascending and descending ground tracks over the two interior regions (Sect. IV-F). For every month and retracker we then derive robust summary statistics (number of crossovers, median and standard deviation of CPE), which serve as an internally consistent measure of along-track stability without relying on external reference data.

We define all self-consistent retracked elevations using a common height convention

$$h_R = \left(\mathrm{alt}_{20,\mathrm{Ku}} - C_{\mathrm{geo}}\right) - r_R, \tag{7}$$

where $\mathrm{alt}_{20,\mathrm{Ku}}$ is the raw satellite altitude at 20 Hz, C_{geo} is the combined 1 Hz geophysical correction mapped to 20 Hz, and r_R is the retracked range for retracker R (ICE-1, ICE-2, AWI-ICENet1, or XTR). We restrict the analysis to 20 Hz records that pass the ESA Level-2I quality flag filter and to crossover pairs that satisfy the common geometric and consistency filters applied in the processing chain. Throughout this study, we use the standard ESA Baseline-E geophysical correction set for all retrackers. Because all retrackers share the same correction set, the reported differences mainly reflect the retracked range term r_R, although the absolute crossover statistics may vary under alternative correction bundles.

F. EXPLAINABILITY TOOLS FOR XTR

To gain physical insight into the behavior of the Transformer-based retracker XTR, we complement the quantitative evaluation with gradient-based saliency analysis and an examination of the internal self-attention patterns. Rather than relying on a single explanation method, we combine three complementary explainability techniques: gradient-based attribution (SmoothGrad), intrinsic attention-based diagnostics (surface-query attention), and perturbation-based sanity checks (windowed occlusion). This combination allows us to contrast local sensitivity, contextual importance, and occlusion-based functional relevance within a unified analysis framework. For a given input waveform $\mathbf{x} \in \mathbb{R}^T$, XTR produces a scalar surface-range estimate $f(\mathbf{x})$ in units of range bins. We estimate the local sensitivity of this prediction to perturbations of individual waveform samples using the SmoothGrad method. Specifically, we compute

$$M(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\left|\nabla_{\mathbf{x}} f\left(\mathbf{x} + \mathcal{N}(0, \sigma^2)\right)\right|, \tag{8}$$

where $N = 16$ noisy realizations are generated by adding independent Gaussian noise with standard deviation $\sigma = 0.01$ to the input waveform. The resulting saliency profile $M(\mathbf{x})$ highlights those range bins for which small perturbations have the strongest influence on the predicted surface range. In addition to gradient-based saliency, we exploit the explicit self-attention mechanism of the Transformer encoder. Let $\mathbf{A}^{(L)} \in \mathbb{R}^{H\times T\times T}$ denote the attention tensor from the final encoder block $l = L$, where H is the number of attention heads. We average over heads to obtain a mean attention matrix $\bar{\mathbf{A}} \in \mathbb{R}^{T\times T}$. From this matrix, we derive two one-dimensional attention profiles along the range dimension:

- *Raw attention*, defined as

$$a_{\mathrm{raw}}(t) = \frac{1}{T}\sum_{q=1}^{T} \bar{A}_{q,t},$$

which reflects the overall tendency of the model to attend to each range bin when averaged over all query positions.
- *Surface-query attention*, which focuses on the attention associated with the predicted surface location. Let $\hat{b}$ denote the surface bin predicted by XTR and define a small query set

$$Q_{\mathrm{surf}} = \{q : |q - \hat{b}| \leq 2\}.$$

The surface-query attention is then given by

$$a_{\mathrm{surf}}(t) = \frac{1}{|Q_{\mathrm{surf}}|}\sum_{q\in Q_{\mathrm{surf}}} \bar{A}_{q,t}.$$

While the raw attention profile summarizes global context usage across the waveform, the surface-query attention identifies waveform samples most strongly attended by queries near the retracked surface range. Together with the SmoothGrad saliency profiles, these diagnostics provide complementary insight into how XTR combines local leading-edge information with broader contextual cues when estimating the surface location. For grouped explainability analysis, absolute-bin averaging is not physically meaningful because the simulated CryoSat-2 LRM waveforms are randomly shifted along the range-bin axis during dataset generation. We therefore align each waveform and its associated explanation profiles relative to the reference retracking bin before averaging. Specifically, for each sample we define a surface-centered coordinate Δr such that $\Delta r = 0$ at the reference retracking bin, and compute grouped mean waveform, SmoothGrad saliency, and surface-query attention profiles in this aligned coordinate system. In the grouped analysis reported here, the attention query window is centered on the reference retracking bin rather than the predicted bin, so that grouped averages remain directly comparable across attenuation subsets. For the grouped reference-aligned analysis, each waveform and its explanation profiles were evaluated in a local window extending 10 bins before and 50 bins after the reference retracking bin, corresponding to approximately -4.7 to $+23.5$ m. This asymmetric window was chosen to retain a short pre-surface region together with a broader trailing-waveform interval that captures the delayed waveform structure associated with volume scattering. Finally, to assess whether the identified regions are functionally relevant rather than merely correlated with the output, we perform windowed occlusion tests by perturbing selected range intervals (zeroing, noise-floor replacement, and within-window permutation) and measuring the resulting change in the retracked surface height. As an additional post-hoc comparison with an established attribution method, we further analyze Integrated Gradients (IG) for both XTR and the re-trained AWI-ICENet1 baseline on held-out simulated test waveforms. Because the simulated waveforms are randomly shifted along the range-bin axis during dataset generation, all IG profiles are aligned relative to the reference retracking bin before averaging. We further stratify this analysis by attenuation level (low, medium, and high attenuation groups based on terciles of the bulk attenuation parameter). In the present study, IG is used as a complementary diagnostic to compare the distribution of attribution mass between the Transformer-based and CNN-based retrackers, rather than as the primary basis for our attenuation-dependent interpretability claims.

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETUP

We evaluate the proposed XTR retracker using both simulated waveforms and CryoSat-2 measurements. The simulated experiments are based on the publicly available AWI-ICENet1 dataset described in Sect. III-A, which is split into training, validation, and test. All simulated-waveform errors are first computed in units of range bins and subsequently converted to metric units using the CryoSat-2 LRM range-bin spacing $\Delta r \approx 0.46875$ m (see Eq. (2)). We report mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) on the held-out test set. For real-data evaluation, we apply XTR, the re-trained AWI-ICENet1 baseline, and the ESA ICE-1 and ICE-2 retrackers to the same set of CryoSat-2 LRM measurements processed under a common correction set and height definition (Sect. III-E). Performance on real data is assessed using crossover point elevation (CPE) statistics over two geodetically stable regions (Vostok and Queen Maud Land) and over the interior Antarctic LRM domain. All retrackers are evaluated under identical quality-control criteria and editing thresholds (Sect. III-E). As a result, differences in the reported error and crossover statistics directly reflect differences in the retracking algorithms rather than downstream processing choices.

B. PERFORMANCE ON SIMULATED WAVEFORMS

Table 2 compares XTR against a classical threshold retracker (TFMRA) and the re-trained AWI-ICENet1 on the simulated CryoSat-2 LRM test set. XTR achieves an RMSE of 0.229 m and an MAE of 0.186 m, clearly outperforming TFMRA (RMSE 0.286 m, MAE 0.222 m). In terms of mean squared error, XTR reduces the MSE by approximately 35% relative to TFMRA, indicating that the Transformer-based model learns a substantially more accurate surface-range estimate than the classical threshold approach under a wide range of simulated penetration regimes.

As expected, the AWI-ICENet1 re-trained with the same augmentation policy as XTR still attains the lowest error on the simulated distribution (RMSE $\approx$ 0.129 m). This reflects the strong match between the convolutional architecture and the simulated training ensemble. In the following sections, we therefore use real-data crossover statistics to

TABLE 2. Retracker performance on the simulated CryoSat-2 LRM test set. Errors are reported in metric units using the LRM bin spacing $\Delta r \approx 0.46875$ m.

Method	MSE (m^2)	RMSE (m)	MAE (m)
XTR	0.052	0.229	0.186
TFMRA	0.081	0.286	0.222
AWI-ICENet1	0.016	0.129	0.092

assess how these waveform-level improvements translate into along-track elevation consistency under a common correction set.

C. ABLATION AND AUGMENTATION-CONTROL EXPERIMENTS

In this study, our primary model is XTR trained with data augmentation. To address concerns about fairness and architectural contribution, we additionally evaluate augmented and non-augmented versions of both XTR and AWI-ICENet1, and we train two ablation variants with the same augmentation as the main XTR: (i) a CNN-only network obtained by removing the Transformer encoder while retaining the convolutional front-end and regression head, and (ii) a Transformer-only network obtained by removing the convolutional stem while retaining the Transformer encoder and regression head. The results are summarized in Table 3. Three main observations can be made. First, AWI-ICENet1 without augmentation remains the strongest model on the purely simulated distribution, indicating that the CNN baseline is highly effective when the training and testing conditions closely match the simulated waveform regime. Second, augmentation does not improve waveform-level regression accuracy on the simulated dataset for either model family. In fact, both AWI-ICENet1 and XTR achieve lower simulated-data error without augmentation. This indicates that augmentation should not be interpreted as the main source of XTR's performance, but rather as a robustness-oriented strategy intended to improve generalization under distribution shift. Third, the ablation experiments show that the architectural composition of XTR is important. Removing the Transformer encoder causes a severe degradation on the simulated test set, whereas the Transformer-only variant clearly outperforms the CNN-only ablation in this controlled simulated evaluation but still remains inferior to the full model. This suggests that neither the convolutional stem nor the Transformer encoder alone is sufficient to recover the performance of the complete XTR architecture; the best performance arises from their combination.

D. DEPENDENCE ON BULK ATTENUATION AND PENETRATION BIAS

To assess how different retrackers behave across firn penetration regimes, we analyze the mean range error as a function of the bulk attenuation rate L_A. We define

$$\Delta R = R_{\mathrm{ref}} - R_{\mathrm{retr}},$$

such that negative values indicate that the retracked surface is placed deeper (i.e., at a larger range) than the reference surface. Figure 2 shows the mean ΔR for XTR, the retrained AWI-ICENet1 baseline, and the classical TFMRA threshold retracker. TFMRA exhibits a pronounced negative bias in weakly attenuating (strongly penetrating) firn, with mean errors approaching -1 m at low L_A and only gradually approaching zero as attenuation increases. This behavior is consistent with the tendency of simple threshold retrackers to lock onto delayed volume-scattering energy rather than the surface-dominated leading edge. In contrast, both XTR and AWI-ICENet1 maintain mean errors close to zero across the full attenuation range. XTR shows a small negative bias at the lowest L_A values and a slight positive bias of a few centimeters at high attenuation, whereas AWI-ICENet1 remains close to zero throughout, with a mild positive bias at both very low and high attenuation. Overall, the deep-learning retrackers learn to discount delayed volume-scattering contributions and to anchor the range estimate near the onset of the surface-dominated leading edge over a wide range of simulated firn conditions.

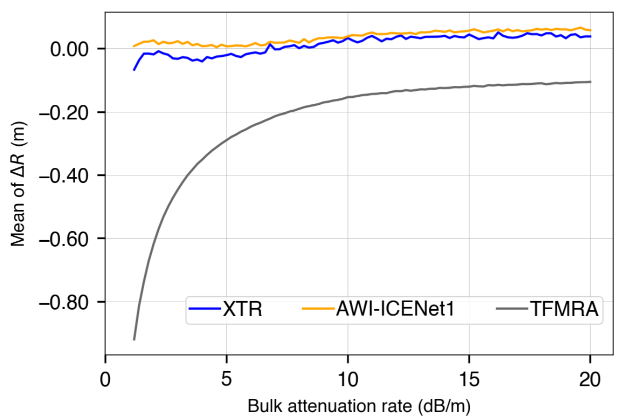


FIGURE 2. Mean range error $\Delta R = R_{\mathrm{ref}} - R_{\mathrm{retr}}$ as a function of bulk attenuation rate L_A on the simulated test set. Negative values indicate that the retracked range is placed deeper than the reference surface. TFMRA shows a strong penetration-related negative bias at low attenuation, whereas XTR and AWI-ICENet1 remain close to zero across most attenuation regimes.

E. SINGLE-WAVEFORM EXAMPLE: COMPARISON WITH CLASSICAL AND CNN RETRACKERS

To illustrate the behavior of the different retrackers at the waveform level, Fig. 3 shows a representative simulated CryoSat-2 LRM waveform generated under strongly attenuating (weakly penetrating) firn conditions. The black line depicts the peak-normalized received power as a function of range, while vertical lines indicate the reference surface range and the retracked surface locations obtained with XTR, AWI-ICENet1, and the TFMRA retracker. An inset highlights the

TABLE 3. Ablation and augmentation-control experiments on the simulated test set. Errors are converted from bin units to meters using the CryoSat-2 LRM bin spacing $\Delta r = 0.46875$ m.

Model	Architecture	Augmentation	MSE (m^2)	RMSE (m)	MAE (m)
AWI–ICENet1	CNN baseline	Yes	0.016	0.129	0.092
AWI–ICENet1 (no aug.)	CNN baseline	No	0.002	0.050	0.039
XTR	CNN + Transformer	Yes	0.052	0.229	0.186
XTR (no aug.)	CNN + Transformer	No	0.016	0.128	0.099
CNN-only ablation (with aug.)	CNN + head	Yes	0.536	0.732	0.394
Transformer-only ablation (with aug.)	Transformer + head	Yes	0.072	0.269	0.212

leading-edge region. For this waveform, the reference surface range is 3.173 m.

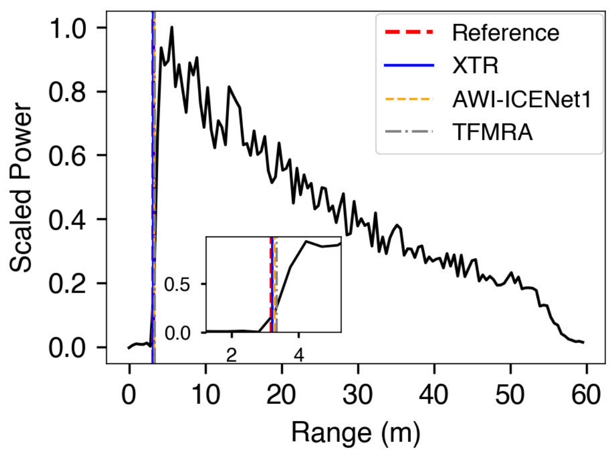


FIGURE 3. Representative simulated CryoSat-2 LRM waveform under strongly attenuating, weakly penetrating firn conditions. The black line shows peak-normalized power as a function of range. Vertical lines mark the reference surface bin (REF) and the surface bins estimated by XTR, AWI-ICENet1, and TFMRA.

XTR retrieves a surface range of 3.220 m, while the retrained AWI-ICENet1 baseline yields 3.281 m. In contrast, the classical TFMRA threshold retracker places the surface at 3.332 m. In terms of range error $\Delta R = R_{\mathrm{ref}} - R_{\mathrm{retr}}$, this corresponds to approximately -0.047 m for XTR, -0.108 m for AWI-ICENet1, and -0.159 m for TFMRA. XTR places the surface estimate within about 5 cm of the reference surface, while AWI-ICENet1 remains within about 0.11 m. In contrast, TFMRA places the surface more than 0.15 m deeper than the reference, indicating a larger positive range offset for this waveform. This example illustrates the characteristic failure mode of simple threshold retrackers and highlights how the deep-learning retrackers suppress delayed subsurface contributions.

F. CROSSOVER-BASED EVALUATION OVER INTERIOR ANTARCTICA

To assess how the retrackers generalize to real CryoSat-2 measurements, we perform a crossover-based evaluation over two regions in the interior of East Antarctica. In such regions, surface slopes are low and true surface-elevation changes over the CryoSat-2 period are minimal.

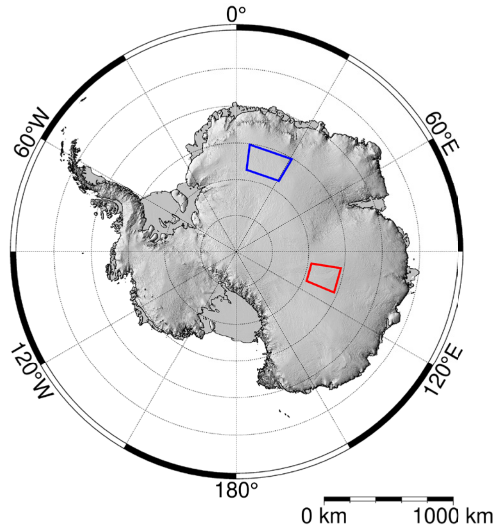


FIGURE 4. Interior Antarctic study areas used for crossover analysis. Red denotes the Vostok region and blue denotes the Queen Maud Land (QML) sector, shown on a shaded-relief map of Antarctic surface elevation.

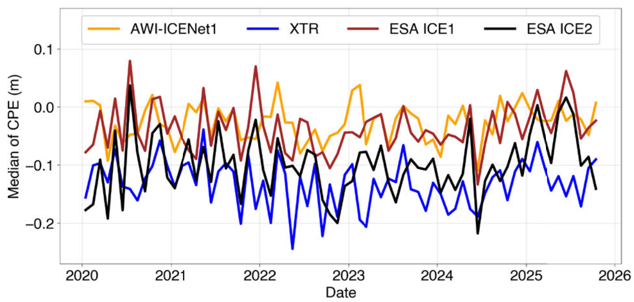
We focus on two interior study areas: the Vostok region and a sector in Queen Maud Land (QML), both characterized by low accumulation rates and weak dynamic thickness changes. These regions therefore provide a stringent test for retracker stability under strongly penetrating firn conditions. The locations of the study areas are shown in Fig. 4.

As described in Sect. III-E, all retrackers are evaluated under a common geometric height definition and correction set. No additional slope corrections, footprint relocation, or retracker-specific post-processing are applied. Consequently, the crossover statistics reported here quantify the intrinsic consistency of the retracking algorithms themselves rather than the performance of fully optimized geophysical height products. Crossover elevation differences are computed by intersecting ascending and descending ground tracks and forming height differences between the two passes. Such crossover and repeat-track approaches are widely used to assess altimetric elevation consistency and ice-sheet elevation change where independent ground truth is limited [3], [43]. For each retracker and month, we summarize the distribution of CPE differences using the median, standard deviation (STD), and the normalized median absolute deviation (NMAD). All statistics are computed after applying identical quality-control criteria and a common outlier threshold of $|CPE| \leq 5$ m to all retrackers.

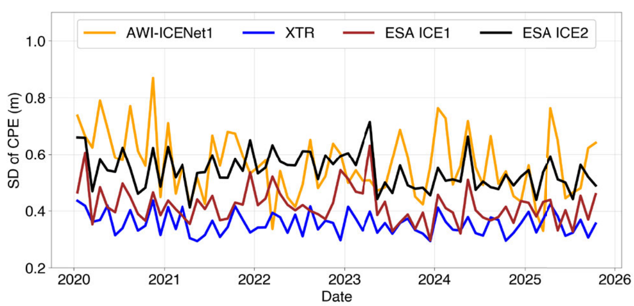
1) VOSTOK REGION

Figure 5 summarizes the crossover statistics for the Vostok region over the period 2020–2025. Panel 5a shows the monthly median crossover point elevation (CPE), which reflects the residual elevation bias of each retracker. The AWI-ICENet1 retracker keeps its median CPE generally close to zero throughout the period, with no clear long-term drift. XTR shows more negative median CPE values, typically between approximately -0.25 m and -0.05 m, corresponding to an offset of roughly 0.1–0.15 m relative to the AWI-ICENet1 baseline. The ESA ICE-1 retracker exhibits intermediate behavior, while ICE-2 generally shows larger negative offsets and stronger temporal variability.

The corresponding monthly standard deviations of the CPE distributions are shown in Fig. 5b. Here, XTR consistently yields the lowest scatter among all retrackers, with typical STD values in the range 0.30–0.45 m and only occasional excursions above this band. ICE-1 and ICE-2 show intermediate scatter, typically between 0.40 and 0.70 m, whereas AWI-ICENet1 exhibits the largest dispersion, frequently around 0.45–0.75 m and reaching values above 0.8 m in some months. Taken together, the Vostok results highlight a clear trade-off between bias and noise. AWI-ICENet1 provides the most bias-free solution in terms of median CPE, while XTR delivers a substantially quieter elevation record, reducing the crossover scatter by approximately 40–60% relative to AWI-ICENet1 in the Vostok region.

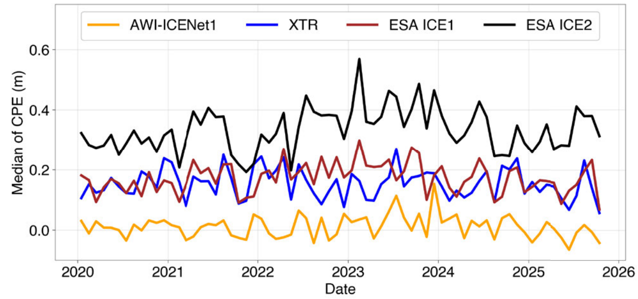


(a) Monthly median CPE.

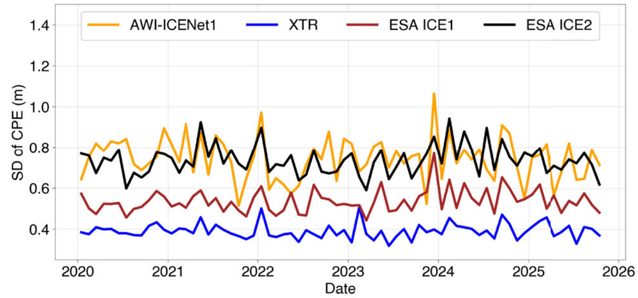


(b) Monthly standard deviation of CPE.

FIGURE 5. Monthly crossover elevation differences in the Vostok region from 2020 to 2025 for AWI-ICENet1, XTR, and ESA ICE-1/ICE-2 retrackers. (a) Median CPE. (b) Standard deviation of CPE.



(a) Monthly median CPE.



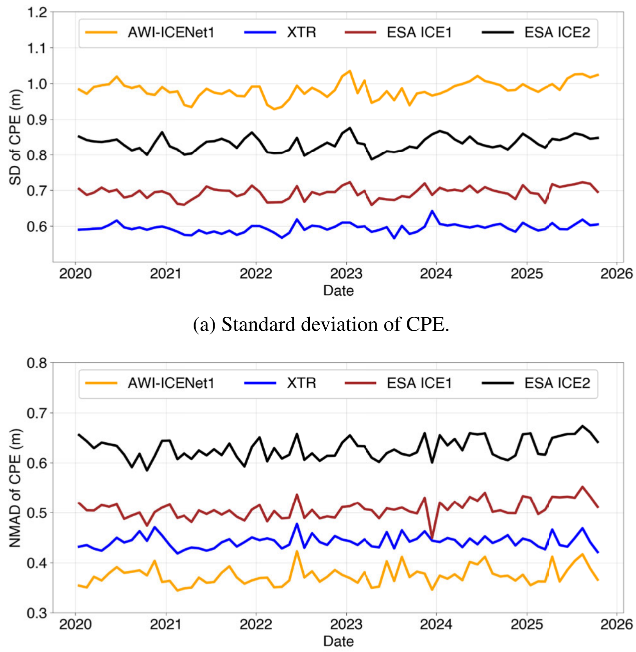
(b) Monthly standard deviation of CPE.

FIGURE 6. Monthly crossover elevation differences in the Queen Maud Land (QML) region from 2020 to 2025 for AWI-ICENet1, XTR, and ESA ICE-1/ICE-2 retrackers. (a) Monthly median CPE. (b) Monthly standard deviation of CPE.

2) QUEEN MAUD LAND (QML)

The QML sector provides an independent interior test with a different orbit geometry and slightly higher surface slopes. Figure 6 summarizes the corresponding crossover statistics. Panel 6a shows the monthly median CPE values. As in the Vostok region, the re-trained AWI-ICENet1 retracker exhibits the smallest median offsets overall, with medians generally remaining close to zero. XTR is consistently shifted toward positive median CPE values in QML, typically on the order of $+0.10$–0.25 m, indicating a modest regional bias relative to AWI-ICENet1. ESA ICE-1 and ICE-2 show larger positive offsets—ICE-1 usually ≈ 0.15–0.30 m, whereas ICE-2 often rises to 0.30–0.45 m and occasionally exceeds 0.50 m—with ICE-2 exhibiting the largest biases.

Panel 6b shows the monthly standard deviation of CPE differences. XTR again yields the lowest scatter, with typical STD values around 0.35–0.45 m. ICE-1 and ICE-2 exhibit intermediate scatter—ICE-1 about 0.50–0.65 m and ICE-2 about 0.60–0.9 m—while AWI-ICENet1 fluctuates between approximately 0.60 and 0.90 m, with occasional spikes above 1.0 m. The contrast between the low STD of XTR and the higher STD of AWI-ICENet1 is most pronounced during 2020–2023, where several months show differences in scatter of roughly 0.25–0.35. Overall, the QML results confirm the pattern observed over Vostok: AWI-ICENet1 provides the smallest median CPE (lowest bias), whereas XTR produces the most stable elevation differences in terms of crossover scatter, at the expense of a modest regional median offset.



(a) Standard deviation of CPE.

(b) Normalized median absolute deviation (NMAD) of CPE.

FIGURE 7. Monthly scatter of crossover elevation differences (CPE) over interior East Antarctica (LRM domain, 2020–2025) for AWI-ICENet1, XTR, and ESA ICE-1/ICE-2 retrackers. Panel (a) shows the standard deviation, which is sensitive to rare large outliers, whereas panel (b) shows the robust NMAD of the CPE distribution.

3) INTERIOR ANTARCTIC LRM DOMAIN

To extend the regional analysis across the interior Antarctic LRM domain, we aggregate all quality-controlled crossovers from the interior Antarctic over the period 2020–2025. The corresponding scatter statistics are summarized in Fig. 7. The monthly median CPE values (not shown) indicate small overall offsets for all retrackers. AWI-ICENet1 remains closest to zero on average, while XTR exhibits a slightly negative median offset of a few centimeters, and the ESA ICE-1/ICE-2 retrackers show small positive offsets. Figure 7a shows the monthly standard deviation of CPE differences.

Consistent with the regional analyses, XTR yields the lowest overall spread, with STDs typically around 0.58–0.62 m. ICE-1 and ICE-2 show intermediate dispersion—ICE-1 about 0.67–0.72 m and ICE-2 about 0.80–0.87 m—whereas AWI-ICENet1 exhibits the largest standard deviations, roughly 0.95–1.05 m, within our retracker-only evaluation under a common correction set. Overall, XTR reduces the standard deviation of CPE differences by roughly 35–45% relative to AWI-ICENet1 while maintaining small median offsets. However, the robust NMAD curves in Fig. 7b provide a complementary perspective. AWI-ICENet1 achieves the lowest NMAD over the full period, indicating the smallest scatter in the core of the crossover distribution. XTR exhibits larger NMAD values than AWI-ICENet1 but remains generally lower than the ESA ICE-1/ICE-2 retrackers. Taken together, these results indicate that AWI-ICENet1 is tighter in the central part of the CPE distribution, whereas XTR primarily reduces a subset of larger residuals that broaden the tails and inflate the standard deviation. In other words, these results suggest that XTR suppresses a subset of larger crossover outliers, while AWI-ICENet1 remains the most consistent retracker in the bulk of the distribution.

G. ADDITIONAL REAL-DATA CROSSOVER ANALYSIS

We further extended the crossover-based real-data analysis to additional model variants. The additional comparisons include: (i) AWI-ICENet1 without augmentation, (ii) XTR trained without augmentation, (iii) a CNN-only ablation, and (iv) a Transformer-only ablation without the convolutional stem. The full XTR model yields substantially lower monthly CPE standard deviations than AWI-ICENet1 on real CryoSat-2 data. XTR trained without augmentation remains close to the full XTR, indicating that the real-data advantage cannot be attributed solely to augmentation. Conversely, retraining AWI-ICENet1 with augmentation does not close the gap to XTR, and the re-trained AWI-ICENet1 baseline remains substantially noisier in the continent-scale crossover analysis. These findings also show that simulated waveform-level regression accuracy and real-data crossover consistency capture different aspects of retracker behavior: a model that best fits the simulated training distribution is not necessarily the one that yields the lowest crossover scatter under real observational conditions. Finally, although the CNN-only ablation yields competitive monthly CPE standard deviations in the present analysis, the CNN-only and Transformer-only ablations do not reproduce the full-model behavior as consistently, supporting the contribution of the combined convolutional–Transformer architecture. The corresponding continent-scale monthly CPE standard-deviation time series are provided in the supplementary material. To assess whether the observed crossover-scatter differences are statistically significant, we additionally performed paired Wilcoxon signed-rank tests on the monthly crossover statistics of XTR and AWI-ICENet1. The tests were applied to the monthly standard deviation series over Vostok, Queen Maud Land, and the interior Antarctic LRM domain, and to the monthly NMAD series over the interior Antarctic domain. These paired tests quantify whether the lower crossover scatter obtained by XTR is statistically robust at the time-series level rather than arising from isolated months. As summarized in Table 4, the paired Wilcoxon signed-rank tests confirm that the monthly STD differences between XTR and AWI-ICENet1 are statistically significant over Vostok, Queen Maud Land, and the interior Antarctic LRM domain. The NMAD comparison is also statistically significant, but in the opposite direction: the re-trained AWI-ICENet1 has lower monthly NMAD values, indicating a tighter core crossover distribution. These tests support the interpretation that XTR primarily reduces the tails of the crossover distribution, whereas AWI-ICENet1 remains stronger in the robust central spread.

TABLE 4. Two-sided paired Wilcoxon signed-rank tests comparing monthly crossover statistics of XTR and AWI-ICENet1.

Metric / Region	Comparison	p-value
STD (Vostok)	XTR vs AWI-ICENet1	4.42×10^{-13}
STD (QML)	XTR vs AWI-ICENet1	3.56×10^{-13}
STD (Antarctica)	XTR vs AWI-ICENet1	3.56×10^{-13}
NMAD (Antarctica)	XTR vs AWI-ICENet1	3.56×10^{-13}

V. EXPLAINABILITY ANALYSIS

Rather than relying on a single post-hoc explanation method, we deliberately combine three complementary explainability techniques: gradient-based attribution (SmoothGrad), intrinsic attention-based diagnostics (surface-query attention), and perturbation-based sanity checks (windowed occlusion). This combination allows us to compare local sensitivity, contextual importance, and occlusion-based functional relevance within a unified explainability framework.

A. SINGLE-WAVEFORM ATTENTION AND SALIENCY

Figure 8 illustrates the explainability analysis for a representative simulated CryoSat-2 waveform dominated by volume scattering. Panel (a) shows the peak-normalized waveform together with the reference surface range. The shaded band highlights the leading-edge area relevant for retracking.

Panel (b) shows the SmoothGrad saliency profile computed according to Eq. (8). The saliency magnitude is sharply concentrated at the onset of the leading edge, indicating that small perturbations in this region have the strongest influence on the predicted surface range. The pre-arrival noise floor and most of the delayed trailing tail contribute comparatively little, with a secondary response at far range.

Panel (c) presents the surface-query attention profile derived from the final Transformer encoder block. The attention exhibits a pronounced maximum within the leading-edge region and remains non-negligible across the trailing part of the waveform. The qualitative correspondence between the surface-query attention and the SmoothGrad saliency indicates that the retracked range is anchored near the surface-dominated return, while the broader distribution of attention suggests that trailing-waveform structure is additionally used as contextual information. Together, these diagnostics provide a physically consistent explanation of XTR's retracking behavior: the model anchors its estimate at the onset of the leading edge while using delayed waveform structure as supplementary context.

B. GROUPED EXPLAINABILITY ANALYSIS ACROSS ATTENUATION REGIMES

To bridge the gap between single-waveform case studies and dataset-level summaries, we perform a grouped explainability analysis across attenuation regimes on the simulated test set. The waveforms were partitioned into low, medium, and high attenuation subsets based on terciles of the bulk attenuation parameter, and 500 samples were used from each attenuation group. Because the simulated waveforms are randomly shifted along the range-bin axis, all waveform and explanation profiles were first aligned relative to the reference retracking bin before averaging.

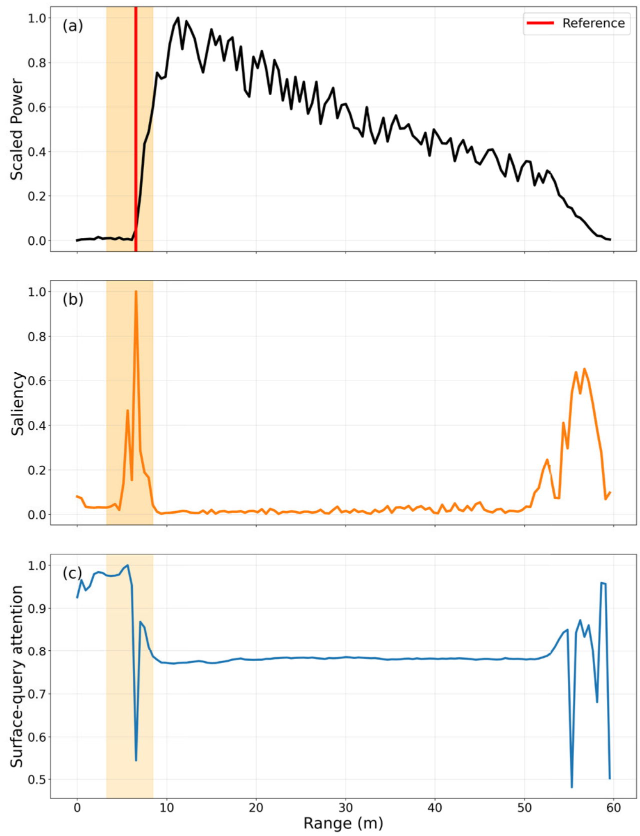


FIGURE 8. Single-waveform explainability analysis for a simulated CryoSat-2 LRM echo dominated by volume scattering. (a) Peak-normalized power as a function of range. The vertical red line marks the reference surface range from the simulated dataset; the shaded band indicates the leading-edge region used for interpretation. (b) SmoothGrad saliency profile as a function of range. Saliency is strongly concentrated on the leading edge, while the pre-arrival noise floor and the trailing tail exhibit much lower values, with a secondary peak at far range. (c) Attention-based importance profile from the last Transformer block (see Sect. III-F). The blue line shows the surface-query attention averaged over queries within ±2 bins of the XTR surface estimate. It peaks in the leading-edge region and remains non-negligible across the trailing part of the waveform, indicating that the surface-query representation is anchored at the surface return, while later samples provide additional contextual information.

Figure 9 shows the resulting reference-aligned mean waveform, mean SmoothGrad saliency, and mean surface-query attention for the three attenuation groups. After alignment, the grouped mean waveforms exhibit a consistent onset near $\Delta r = 0$ across all attenuation regimes. The leading-edge transition differs across attenuation groups, with the low-attenuation group exhibiting a broader leading-edge region than the higher-attenuation groups. This is consistent with stronger delayed subsurface-scattering contributions under low L_A, where weaker attenuation allows deeper radar penetration and enhances the volume-scattering influence. In contrast, the subsequent changes in SmoothGrad saliency

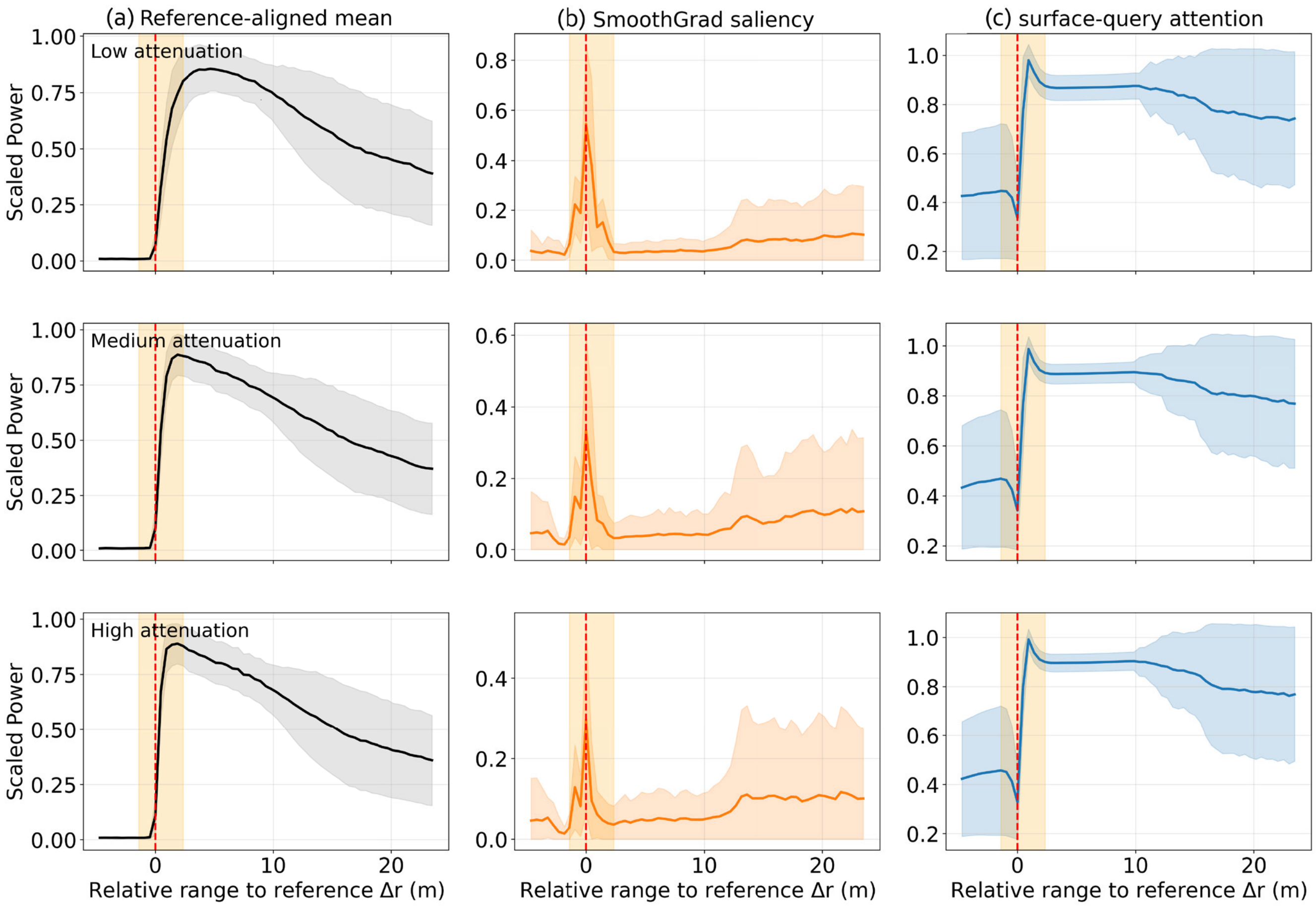


FIGURE 9. Grouped explainability analysis across attenuation regimes after reference alignment. The simulated test waveforms were partitioned into low-, medium-, and high-attenuation subsets, and each waveform was aligned relative to its reference retracking bin before averaging. The three columns show the reference-aligned mean waveform, mean SmoothGrad saliency, and mean surface-query attention, respectively. The dashed red line marks the aligned reference location ($\Delta r = 0$), and the shaded orange region denotes the near-surface analysis window. The plotted profiles are shown in a local window extending 10 bins before and 50 bins after the reference location, corresponding to approximately −4.7 to +23.5 m. This grouped analysis complements the single-waveform case study by showing that the explanation patterns persist across representative attenuation regimes despite the random range-bin shifts in the simulated dataset.

describe model sensitivity rather than a direct broadening of the mean waveform itself. The grouped SmoothGrad profiles retain their strongest peak in the near-surface region around the aligned reference location, indicating that XTR remains most locally sensitive to the surface-dominated onset of the echo. At the same time, the saliency profiles become more distributed toward the trailing part of the waveform in the higher-attenuation groups. As summarized in Table 5, which reports mean $\pm$ standard deviation of the relative saliency and attention fractions across samples, the near-surface saliency fraction decreases from 0.446 in the low-attenuation group to 0.298 and 0.228 in the medium and high-attenuation groups, respectively, while the delayed-tail saliency fraction increases from 0.122 to 0.158 and 0.181. This pattern indicates that XTR increasingly distributes its attribution toward the delayed trailing part of the waveform in the higher-attenuation groups, reflecting a change in model sensitivity rather than a broadening of the mean leading edge itself.

The corresponding surface-query attention profiles are broader than the saliency profiles and extend well into the trailing part of the waveform. This suggests that, in addition to its strongest local sensitivity near the aligned surface region, XTR uses a wider contextual representation of delayed waveform structure when forming its retracking decision. Thus, the grouped analysis shows that XTR combines strong near-surface sensitivity with broader contextual cues, rather than an attention pattern confined solely to the immediate leading edge.

For comparison, we additionally computed grouped reference-aligned SmoothGrad saliency profiles for the AWI-ICENet1 baseline using the same attenuation grouping, alignment procedure, and near-surface / delayed-tail windows. AWI-ICENet1 likewise peaks near the aligned surface, showing that both models are anchored to the surface-dominated onset of the echo. However, the attenuation dependence differs: compared with AWI-ICENet1, XTR exhibits a stronger broadening of the saliency profile across

TABLE 5. Summary of grouped reference-aligned explainability metrics across attenuation regimes. Values are mean ± standard deviation across samples. Near-surface and tail fractions denote the fraction of total SmoothGrad saliency (sal.) or surface-query attention (att.) mass contained in the aligned near-surface and delayed-tail windows, respectively.

Group	Near-surface sal.	Tail sal.	Near-surface att.	Tail att.
Low attenuation	0.446 ± 0.173	0.122 ± 0.055	0.117 ± 0.017	0.339 ± 0.072
Medium attenuation	0.298 ± 0.142	0.158 ± 0.065	0.114 ± 0.015	0.335 ± 0.070
High attenuation	0.228 ± 0.124	0.181 ± 0.066	0.114 ± 0.014	0.340 ± 0.072

attenuation groups, with a more pronounced redistribution of saliency from the near-surface region toward the delayed trailing part of the waveform. This suggests that XTR adapts more flexibly to attenuation-dependent changes in waveform structure, whereas the CNN baseline remains more locally concentrated. The corresponding grouped AWI-ICENet1 saliency figure and summary metrics are provided in the supplementary material. As complementary evidence, we also evaluate reference-aligned grouped Integrated Gradients (IG) profiles for XTR and AWI-ICENet1 in the same attenuation-stratified setting. The grouped IG results, provided in the supplementary material, show a consistent qualitative difference between the two models: AWI-ICENet1 remains more concentrated near the aligned surface region, whereas XTR assigns a larger fraction of attribution to the delayed trailing part of the waveform. This supports the interpretation that XTR uses broader waveform context than the CNN baseline. However, because the attenuation dependence in the grouped IG summaries is weaker than in the SmoothGrad analysis, we use IG here as complementary evidence rather than as the primary basis for attenuation-stratified interpretability claims.

C. AGGREGATE EXPLAINABILITY AND OCCLUSION-BASED SANITY CHECKS

To complement the single-waveform and grouped attenuation-based analyses, we perform a broader aggregate explainability analysis on a random subset of simulated test samples. As in the grouped analysis, waveform and explanation profiles are aligned before averaging to remove the effect of random range-bin shifts in the simulated dataset. This aggregate view is intended to summarize the overall distribution of explanation patterns across many waveforms, whereas the grouped analysis in Sect. V-B resolves how these patterns vary across attenuation regimes.

The aggregate view is qualitatively consistent with the grouped attenuation-based analysis: SmoothGrad saliency remains most concentrated near the aligned surface location, whereas the surface-query attention remains broader and extends farther into the trailing part of the waveform. Thus, the two diagnostics are complementary: saliency highlights where the output is most locally sensitive, while attention reflects a wider context integration pattern. To assess occlusion-based functional relevance beyond attribution maps, we conduct windowed occlusion tests by perturbing predefined range intervals and measuring the resulting absolute change in retracked surface height. As summarized in Table 6, occluding the core region around the surface leads to meter-scale deviations, whereas perturbations applied to the pre-arrival noise region have negligible effect. For the trailing tail, permutation-based occlusion yields only small changes, whereas zeroing or replacing the tail with the baseline level produces much larger deviations.

TABLE 6. Occlusion-based sanity check on 100 simulated test waveforms. Reported values are median absolute height changes $|\Delta h|$ (m) after perturbing different range windows relative to the XTR-predicted surface bin.

Window (relative to surface)	Zero	Baseline	Permute
Pre-leading $[-20, -5]$ m	0.008	0.001	0.000
Core $[-2, 2]$ m	1.866	1.869	0.371
Core $[-5, 5]$ m	4.676	4.680	0.494
Tail $[2, 8]$ m	5.852	5.622	0.015
Tail $[8, 20]$ m	12.658	12.486	0.046

This indicates that the detailed ordering of the tail samples is not itself critical, but that the presence of trailing-return energy contributes important contextual information for the surface estimate. Overall, the aggregate and occlusion-based analyses support the interpretation that XTR places most of its weight on the surface-dominated leading edge when estimating the surface range, while also using the trailing return as contextual input that helps stabilize retracking across different penetration regimes.

VI. DISCUSSION

This study introduced XTR, an explainable Transformer-based retracker for CryoSat-2 LRM radar waveforms, and evaluated its performance on both simulated data and satellite measurements. The results highlight a consistent trade-off between waveform-level accuracy, crossover scatter, and interpretability across different retracking approaches.

On the simulated dataset, XTR clearly outperforms the classical TFMRA threshold retracker, demonstrating substantially reduced range errors across a wide range of firn attenuation conditions. As expected, the re-trained AWI-ICENet1 achieves the lowest mean squared error on the simulated test set, reflecting the close alignment between its convolutional architecture and the simulated training ensemble. These results confirm that CNN-based retrackers remain highly effective when the training and test distributions are well matched. However, the real-data crossover analysis reveals a different perspective. Over two interior

Antarctic test regions, XTR consistently reduces the standard deviation of crossover elevation differences relative to AWI-ICENet1, in some cases by approximately 40–60%. This suggests that XTR suppresses a subset of large residuals that broaden the tails of the crossover distribution and inflate the overall scatter. In contrast, AWI-ICENet1 generally achieves the lowest NMAD, suggesting superior consistency in the core of the crossover distribution but increased sensitivity to large deviations. The ablation and augmentation-control experiments further refine this interpretation. On the simulated dataset, the re-trained AWI-ICENet1 baseline and XTR trained without augmentation show that augmentation alone does not explain the behavior of XTR. In fact, both model families achieve lower waveform-level error without augmentation than with augmentation on the simulated distribution. At the same time, the real-data crossover analysis shows that XTR remains highly competitive even without augmentation, whereas AWI-ICENet1 re-trained with augmentation still does not match the crossover stability of XTR. This indicates that augmentation should be interpreted primarily as a robustness-oriented training strategy rather than as the main source of performance improvement. The architectural ablations reveal a similarly nuanced picture. On the simulated dataset, removing the Transformer encoder causes a severe loss in accuracy, while the Transformer-only variant performs substantially better than the CNN-only ablation on the simulated test set but still remains inferior to the full XTR architecture. This supports the view that the strongest simulated-data performance arises from combining local convolutional feature extraction with global self-attention [25]. In the current real-data crossover analysis, however, the relative behavior of the ablated models appears more metric-dependent, reinforcing that waveform-level regression error and crossover stability quantify different aspects of retracker behavior.

The explainability analyses provide a physical interpretation for these differences. SmoothGrad saliency consistently indicates that XTR is most sensitive to the onset of the surface-dominated leading edge. The attention-based diagnostics, in turn, show that the model combines this strong near-surface sensitivity with a broader use of trailing-waveform context. The grouped reference-aligned analysis further demonstrates that this behavior persists across attenuation regimes, while becoming more distributed toward the trailing part of the waveform in some attenuation groups. A grouped saliency comparison with the AWI-ICENet1 baseline further shows that both models are anchored near the aligned surface region, but that XTR exhibits a stronger attenuation-dependent broadening of its saliency profile. This suggests that the Transformer-based retracker may adapt more flexibly to attenuation-dependent changes in waveform structure, whereas the CNN baseline remains more locally concentrated. Windowed occlusion experiments further indicate that the detailed ordering of the trailing waveform samples is not itself critical for the surface estimate, although the presence of trailing-return energy contributes contextual information that may help stabilize the retracking under varying penetration conditions. This behavior aligns well with established physical interpretations of radar altimetry waveforms and supports the use of attention-based architectures as physically interpretable retrackers.

Several limitations should be noted. First, the satellite-data evaluation presented here is intentionally restricted to a retracker-only comparison under a common correction set and height definition. Consequently, the reported crossover statistics do not represent fully optimized geophysical height products and should not be interpreted as a direct replacement for operational processing chains. Second, XTR is trained exclusively on simulated waveforms and may therefore be sensitive to residual simulation–reality mismatches, particularly in regions with complex surface roughness or unusual firn properties. Overall, the results suggest that attention-based retrackers such as XTR provide a valuable and complementary alternative to CNN-based schemes. By combining competitive retracking performance with explicit interpretability and improved crossover stability in this retracker-only evaluation, XTR provides a promising pathway toward more transparent and robust radar altimetry retracking over ice sheets.

VII. CONCLUSION

This paper presented XTR, an explainable Transformer-based retracker for CryoSat-2 LRM radar waveforms. Trained on the publicly available AWI-ICENet1 simulation ensemble, XTR outperforms a classical threshold retracker on simulated data, although it remains less accurate than the augmentation-matched AWI-ICENet1 CNN baseline in terms of waveform-level MSE on the same forward-model distribution. Additional ablation and augmentation-control experiments with the re-trained AWI-ICENet1 and an XTR model trained without augmentation show that augmentation alone does not explain the behavior of XTR. The architectural ablation experiments further indicate that the Transformer encoder contributes substantially to the simulated-data performance of XTR. A CNN-only ablation degrades severely, while a Transformer-only variant performs substantially better than the CNN-only model but still remains inferior to the full architecture. These results indicate that the best simulated-data performance arises from the combination of local convolutional feature extraction and global self-attention. On CryoSat-2 LRM measurements over interior Antarctica (2020–2025), evaluated under a common correction set and height definition, XTR generally reduces the standard deviation of crossover elevation differences relative to the re-trained AWI-ICENet1 baseline. At the same time, AWI-ICENet1 remains tighter in the core of the crossover distribution according to the NMAD. Taken together, these results indicate that simulated waveform-level regression accuracy and real-data crossover consistency capture different aspects of retracker performance.

Beyond performance, XTR provides explicit interpretability through self-attention and gradient-based saliency diagnostics. Reference-aligned Integrated Gradients analyses further support the interpretation that XTR uses broader trailing-waveform context than the CNN baseline. These analyses indicate that XTR tends to anchor the surface estimate near the onset of the surface-dominated leading edge, while also using broader trailing-waveform context that may help stabilize retracking under varying penetration regimes. The grouped reference-aligned comparison with the AWI-ICENet1 baseline further shows that both models are primarily driven by the near-surface waveform region, but that XTR exhibits a stronger attenuation-dependent broadening of its saliency profile, consistent with a potentially more flexible adaptation to attenuation-dependent waveform structure. Overall, XTR provides a competitive and interpretable alternative to CNN-based retrackers for CryoSat-2 altimetry.

ACKNOWLEDGMENT

The authors would like to thank V. Helm for making the simulated AWI-ICENet1 training dataset publicly available via PANGAEA [20]. This work was conducted at the Faculty of Mathematics and Computer Science, University of Bremen, Germany.

DATA AND SOFTWARE AVAILABILITY

All experiments were conducted in Python within a Conda environment. The main libraries included pandas, NumPy, SciPy, matplotlib, scikit-learn, scikit-image, and TensorFlow (v2.16.1). Computations were performed on a Linux-based virtual machine. The virtual machine was equipped with four virtual CPU cores, 32 GB RAM, and access to an NVIDIA Tesla P40 GPU (24 GB), with CUDA 12.2. CryoSat-2 Level-1B and Level-2 intermediate (L2I) altimetry products were obtained from ESA Earth Online, and the simulated dataset was derived from the publicly available AWI-ICENet1 simulations [20]. Geospatial data processing and visualization were carried out using QGIS (v3.40.7) and the Generic Mapping Tools (GMT) (v6.5.0). The XTR training and evaluation scripts are publicly available in a dedicated GitHub repository at https://github.com/Alireza-Dehghanpour

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Alireza Dehghanpour: Conceptualization; Methodology; Software; Data curation; Formal analysis; Investigation; Visualization; Writing–original draft. Danial Barazandeh: Conceptualization; Methodology; Visualization; Writing–review and editing. Gabriel Zachmann: Supervision; Writing–review and editing. All authors discussed the results and contributed to editing the article.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

DECLARATION OF GENERATIVE AI

During the preparation of this work, the authors used generative AI tools for text polishing. After using these tools, they carefully reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] A. Shepherd et al., "Mass balance of the Antarctic Ice Sheet from 1992 to 2017," *Nature*, vol. 558, no. 7709, pp. 219–222, 2018.

[2] I. N. Otosaka et al., "Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020," *Earth Syst. Sci. Data*, vol. 15, no. 4, pp. 1597–1616, 2023. [Online]. Available: https://essd.copernicus.org/articles/15/1597/2023/

[3] D. J. Wingham, A. J. Ridout, R. Scharroo, R. J. Arthern, and C. K. Shum, "Antarctic elevation change from 1992 to 1996," *Science*, vol. 282, no. 5388, pp. 456–458, Oct. 1998.

[4] A. Shepherd et al., "A reconciled estimate of ice-sheet mass balance," *Science*, vol. 338, no. 6111, pp. 1183–1189, 2012.

[5] D. Wingham, "CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields," *Adv. Space Res.*, vol. 37, no. 4, pp. 841–871, 2006.

[6] M. McMillan, A. Shepherd, A. Sundal, K. Briggs, A. Muir, A. Ridout, A. Hogg, and D. Wingham, "Increased ice losses from Antarctica detected by CryoSat-2," *Geophys. Res. Lett.*, vol. 41, no. 11, pp. 3899–3905, Jun. 2014.

[7] C. H. Davis and H. J. Zwally, "Geographic and seasonal variations in the surface properties of the ice sheets by satellite-radar altimetry," *J. Glaciology*, vol. 39, no. 133, pp. 687–697, 1993.

[8] F. Larue, G. Picard, J. Aublanc, L. Arnaud, A. Robledano-Perez, E. Le Meur, V. Favier, B. Jourdain, J. Savarino, and P. Thibaut, "Radar altimeter waveform simulations in Antarctica with the snow microwave radiative transfer model (SMRT)," *Remote Sens. Environ.*, vol. 263, Sep. 2021, Art. no. 112534.

[9] V. Helm, A. Humbert, and H. Miller, "Elevation and elevation change of Greenland and Antarctica derived from CryoSat-2," *Cryosphere*, vol. 8, no. 4, pp. 1539–1559, Aug. 2014.

[10] J. Nilsson, A. S. Gardner, and F. S. Paolo, "Elevation change of the Antarctic ice sheet: 1985 to 2020," *Earth Syst. Sci. Data*, vol. 14, no. 8, pp. 3573–3598, Aug. 2022.

[11] J. Nilsson, P. Vallelonga, S. B. Simonsen, L. S. Sørensen, R. Forsberg, D. Dahl-Jensen, M. Hirabayashi, K. Goto-Azuma, C. S. Hvidberg, H. A. Kjær, and K. Satow, "Greenland 2012 melt event effects on CryoSat-2 radar altimetry," *Geophys. Res. Lett.*, vol. 42, no. 10, pp. 3919–3926, May 2015.

[12] G. Brown, "The average impulse response of a rough surface and its applications," *IEEE Trans. Antennas Propag.*, vol. AP-25, no. 1, pp. 67–74, Jan. 1977.

[13] J. K. Ridley and K. C. Partington, "A model of satellite radar altimeter return from ice sheets," *Int. J. Remote Sens.*, vol. 9, no. 4, pp. 601–624, Apr. 1988.

[14] T. V. Martin, H. J. Zwally, A. C. Brenner, and R. A. Bindschadler, "Analysis and retracking of continental ice sheet radar altimeter waveforms," *J. Geophys. Res., Oceans*, vol. 88, no. 3, pp. 1608–1616, Feb. 1983.

[15] B. Legresy, F. Papa, F. Remy, G. Vinay, M. van den Bosch, and O.-Z. Zanife, "ENVISAT radar altimeter measurements over continental surfaces and ice caps using the ICE-2 retracking algorithm," *Remote Sens. Environ.*, vol. 95, no. 2, pp. 150–163, Mar. 2005.

[16] S. B. Simonsen and L. S. Sørensen, "Implications of changing scattering properties on Greenland ice sheet volume change from cryosat-2 altimetry," *Remote Sens. Environ.*, vol. 190, pp. 207–216, Mar. 2017.

[17] T. Slater, A. Shepherd, M. Mcmillan, T. W. K. Armitage, I. Otosaka, and R. J. Arthern, "Compensating changes in the penetration depth of pulse-limited radar altimetry over the Greenland ice sheet," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9633–9642, Dec. 2019.

[18] L. Schröder, M. Horwath, R. Dietrich, V. Helm, M. R. van den Broeke, and S. R. M. Ligtenberg, "Four decades of Antarctic surface elevation changes from multi-mission satellite altimetry," *Cryosphere*, vol. 13, no. 2, pp. 427–449, Feb. 2019. [Online]. Available: https://tc.copernicus.org/articles/13/427/2019/

[19] V. Helm, A. Dehghanpour, R. Hänsch, E. Loebel, M. Horwath, and A. Humbert, "AWI-ICENet1: A convolutional neural network retracker for ice altimetry," *Cryosphere*, vol. 18, no. 9, pp. 3933–3970, Sep. 2024. [Online]. Available: https://tc.copernicus.org/articles/18/3933/2024/

[20] V. Helm, "Convolutional neural network training dataset and results of AWI-ICENet1 retracker" [Dataset]. PANGAEA, Bremen/Bremerhaven, Germany, 2024, doi: 10.1594/PANGAEA.964596.

[21] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.

[22] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller, "Explaining deep neural networks and beyond: A review of methods and applications," *Proc. IEEE*, vol. 109, no. 3, pp. 247–278, Mar. 2021.

[23] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3319–3328.

[24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[26] C. Tao, L. Meng, J. Li, Z. Zhang, Z. Wang, X. Wang, W. Zhang, and L. Li, "TransR: A transformer-based satellite altimetry waveform retracking network for lake water level," *SSRN*, pp. 1–28, Mar. 2024, doi: 10.2139/ssrn.4767280.

[27] F. Rémy and S. Parouty, "Antarctic ice sheet and radar altimetry: A review," *Remote Sens.*, vol. 1, no. 4, pp. 1212–1239, Dec. 2009.

[28] D. F. Mattes, "Analysis of waveforms in the satellite altimetry by using neural networks," Master's thesis, Universität Stuttgart, Stuttgart, Germany, 2019.

[29] O. Memarian Sorkhabi, J. Asgari, and A. Amiri-Simkooei, "Wavelet decomposition and deep learning of altimetry waveform retracking for lake urmia water level survey," *Mar. Georesources Geotechnology*, vol. 40, no. 3, pp. 361–369, Mar. 2022.

[30] X. Deng, L. Ke, L. Jiang, K. Nielsen, X. Fan, J. Wang, and C. Song, "A review of altimetry waveform retracking for inland water levels," *Geodesy Geodynamics*, vol. 16, no. 4, pp. 488–508, Jul. 2025.

[31] J. Zheng, X. Xu, Y. Xu, and C. Guo, "Coastal waveform retracking for synthetic aperture altimeters using a multiple optimization parabolic cylinder algorithm," *Remote Sens.*, vol. 15, no. 19, p. 4665, 2023.

[32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–21.

[33] D. Tuia, K. Schindler, B. Demir, X. X. Zhu, M. Kochupillai, S. Džeroski, J. N. van Rijn, H. H. Hoos, F. D. Frate, M. Datcu, V. Markl, B. L. Saux, R. Schneider, and G. Camps-Valls, "Artificial intelligence to advance Earth observation: A review of models, recent trends, and pathways forward," *IEEE Geosci. Remote Sens. Mag.*, vol. 13, no. 4, pp. 119–141, Dec. 2025.

[34] R. Shao, C. Yang, Q. Li, L. Xu, X. Yang, X. Li, M. Li, Q. Zhu, Y. Zhang, Y. Li, Y. Liu, Y. Tang, D. Liu, S. Yang, and H. Li, "AllSpark: A multimodal spatiotemporal general intelligence model with ten modalities via language as a reference framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5606620, doi: 10.1109/TGRS.2025.3526725.

[35] S. He, P. Shen, P. Xu, Q. Luo, and H. Li, "STDCformer: A transformer-based model with a spatial–temporal causal de-confounding strategy for crowd flow prediction," *Inf. Fusion*, vol. 126, Feb. 2026, Art. no. 103645, doi: 10.1016/j.inffus.2025.103645.

[36] Y. Wang, S. He, Q. Luo, H. Yuan, L. Zhao, J. Zhu, and H. Li, "Causal invariant geographic network representations with feature and structural distribution shifts," *Future Gener. Comput. Syst.*, vol. 169, Aug. 2025, Art. no. 107814.

[37] E. Hasanpour Zaryabi, L. Moradi, B. Kalantar, N. Ueda, and A. A. Halin, "Unboxing the black box of attention mechanisms in remote sensing big data using XAI," *Remote Sens.*, vol. 14, no. 24, p. 6254, Dec. 2022.

[38] F. Huang, S. Jiang, L. Li, Y. Zhang, Y. Zhang, R. Zhang, Q. Li, D. Li, W. Shangguan, and Y. Dai, "Applications of explainable artificial intelligence in Earth system science," 2024, *arXiv:2406.11882*.

[39] A. Höhl, I. Obadic, M.-Á. Fernández-Torres, D. Oliveira, and X. X. Zhu, "Recent trends, challenges, and limitations of explainable AI in remote sensing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 8199–8205.

[40] O. Hall, M. Ohlsson, and T. Rögnvaldsson, "A review of explainable AI in the satellite data, deep machine learning, and human poverty domain," *Patterns*, vol. 3, no. 10, Oct. 2022, Art. no. 100600.

[41] P. Ghamisi, W. Yu, A. Marinoni, C. M. Gevaert, C. Persello, S. Selvakumaran, M. Girotto, B. P. Horton, P. Rufin, P. Hostert, F. Pacifici, and P. M. Atkinson, "Responsible artificial intelligence for Earth observation: Achievable and realistic paths to serve the collective good," *IEEE Geosci. Remote Sens. Mag.*, vol. 13, no. 3, pp. 72–96, Sep. 2025.

[42] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "Smooth-Grad: Removing noise by adding noise," 2017, *arXiv:1706.03825*.

[43] T. Flament and F. Rémy, "Dynamic thinning of Antarctic glaciers from along-track repeat radar altimetry," *J. Glaciology*, vol. 58, no. 211, pp. 830–840, 2012.

ALIREZA DEHGHANPOUR received the B.Sc. degree in physics and the dual M.Sc. degree in artificial intelligence and robotics and remote sensing. He is currently a Researcher in computer engineering with the University of Bremen, Germany, and a member of the Computer Graphics and Virtual Reality (CGVR) Laboratory. His research interests include artificial intelligence, machine learning, computer vision, point-cloud processing, 3D Gaussian splatting, and collision detection, with a focus on their applications in geospatial data analysis, large-scale scientific computing, and computer graphics. He has been a Researcher with the Alfred Wegener Institute (AWI), where he developed deep-learning pipelines for CryoSat-2 waveform retracking on high-performance computing (HPC) systems.

DANIAL BARAZANDEH received the M.Sc. degree in artificial intelligence and robotics. He is currently pursuing the Ph.D. degree in computer science with the University of Bremen, where he conducts research in computer graphics, virtual reality, and generative models, with a particular focus on diffusion-based systems for 3D environments. His work bridges academic research and large-scale software development, combining machine learning, computer vision, and interactive graphics. He has a strong background in building interpretable, high-performance AI applications.

GABRIEL ZACHMANN received the Ph.D. degree in computer science from Darmstadt University, in 2000. He is currently a Full Professor of computer graphics, visual computing, and virtual reality with the University of Bremen, Germany, where he has headed the Computer Graphics Laboratory, since 2012. He has published widely, served on numerous program committees, and contributed to the organization of major conferences, such as IEEE VR and EuroXR. His research interests include geometric algorithms for computer graphics, collision detection, virtual prototyping, haptics, GPU computing, and immersive visualization.