DiffWater: A Conditional Diffusion Model for Estimating Surface Water Fraction Using CyGNSS Data

Yuhan Chen, *Student Member, IEEE*, Qingyun Yan, *Member, IEEE*, Shuanggen Jin, *Senior Member, IEEE*, and Weiming Huang, *Senior Member, IEEE*

***Abstract*— Recent advances in cyclone global navigation satellite system (CyGNSS) data have significantly improved the extraction of monthly surface water fraction (SWF), with neural networks being widely used for large-scale water body mapping based on global navigation satellite system-reflectometry (GNSS-R) signals. However, inherent noise in CyGNSS signals, such as multipath effects and interference, presents substantial challenges to the accuracy of SWF estimation. Diffusion models, an emerging class of generative deep learning techniques, have shown remarkable capabilities in capturing complex data distributions. By leveraging an iterative process of noise addition and removal, these models demonstrate significant advantages in processing low signal-to-noise ratio data, offering a novel methodology for precise SWF estimation from CyGNSS data. This study introduces DiffWater, a framework designed to address the unique characteristics of CyGNSS data and systematically explore the applicability of conditional diffusion models for remote sensing tasks. Utilizing a composite reference dataset, which includes the global surface water (GSW) dataset and the global surface water dynamics (GLAD) dataset as training targets, DiffWater enhances the objectives of conditional diffusion models by integrating advanced conditional feature extractors and implementing multilevel fusion of conditional and temporal features, thereby achieving significant improvements in SWF estimation performance. Comprehensive experimental evaluations on the reference dataset demonstrate that DiffWater achieved the best performance, with a root-mean-squared error (RMSE) of 4.987% and a correlation coefficient (*R*) of 0.946. Compared to state-of-the-art SWF estimation methods, the proposed approach demonstrated significant improvements in both quantitative and qualitative results.**

***Index Terms*— Cyclone global navigation satellite system (CyGNSS), deep learning, diffusion models, global navigation satellite system-reflectometry (GNSS-R), surface water.**

Received 2 February 2025; revised 3 April 2025; accepted 24 April 2025. Date of publication 28 April 2025; date of current version 9 May 2025. This work was supported by the National Natural Science Foundation of China under Grant 42001362. *(Corresponding author: Qingyun Yan.)*

Yuhan Chen is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with Qingdao Innovation and Development Base (Centre), Harbin Engineering University, Qingdao 266000, China.

Qingyun Yan is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: 003257@nuist.edu.cn).

Shuanggen Jin is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China, also with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China, and also with Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China.

Weiming Huang is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada.

Digital Object Identifier 10.1109/TGRS.2025.3564612

I. Introduction

REGULAR, accurate, and widespread estimates of surface water fraction (SWF), particularly in pan-tropical regions, are crucial for applications such as hydrological modeling and analysis of the Earth's water cycle [1], [2]. Surface water is closely linked to greenhouse gas emissions, ecosystem biodiversity, and various life forms [3]. For instance, aquatic ecosystems are estimated to account for 41%–53% of global methane emissions, with rivers, lakes, and reservoirs collectively contributing to half [4]. However, surface water estimates are significantly affected by errors arising from uncertainties in the distribution of small water bodies [5]. Furthermore, the presence, extent, and quantity of surface water exhibit high variability in both space and time, rendering monitoring efforts a considerable challenge [6]. Therefore, accurate, effective, and timely monitoring of surface water and its spatiotemporal evolution has become a crucial and complex task.

Research on the use of remote sensing sensors to detect surface water has made significant progress. Remote sensing technology provides an effective means for continuously monitoring surface water data at regional and global scales, with data primarily being sourced from Moderate Resolution Imaging Spectroradiometer (MODIS) [7], [8], [9], the Landsat missions [10], [11], [12], and the Sentinel-1/2 missions [13], [14], [15], [16]. Notably, the Landsat mission, with its 30-m spatial resolution and 40 years of historical data, is crucial for global water resource monitoring [17]. Landsat imagery is vital for highlighting the significant impacts of climate change and human activities on global water resources. However, optical remote sensing is hindered by cloud cover and vegetation, resulting in a significantly lower number of cloud-free satellite images available during the rainy season compared to the dry season [11]. This leads to a lack of data, thus limiting its applicability to tropical regions, especially during the rainy season. Furthermore, the spectral characteristics of water surfaces are quite complex globally, as they vary with depth, dissolved substances, and chlorophyll content [18].

Different types of water bodies and pollutants can significantly affect reflectance, thereby impacting the interpretation of remote sensing data. The 30-m spatial resolution and 40 years of historical data from Landsat imagery are crucial for global water resource monitoring [19], [20]. The global surface water (GSW) dataset [11] provides a compilation of monthly water extent data globally from 1984 to 2021, while the global land analysis and discovery (GLAD) dataset [12] offers a compilation of monthly water extent data from 1999 to 2021. Reports indicate that the GLAD dataset outperforms the GSW dataset in identifying water bodies, and both can complement each other [18].

In contrast, microwave remote sensing is less susceptible to the aforementioned limitations and can provide all-weather monitoring [21], [22]. Microwave sensors can penetrate clouds and facilitate effective monitoring under various weather conditions. However, passive radiometers typically exhibit a coarse spatial resolution (25–50 km), which poses a challenge for high-resolution water system detection [23]. Conversely, active microwave platforms, such as synthetic aperture radar (SAR) and radar altimeters, possess suitable spatial resolution, but their revisit period is typically long [13]. For example, Sentinel-1 has a spatial resolution ranging from a few meters to tens of meters, depending on the operating mode, with a revisit time of six days at the equator and a longer time at higher latitudes [24], [25]. The surface water microwave product series (SWAMPS) dataset [26] represents one of the longest records (over 20 years) of global water coverage, obtained through the combination of passive and active microwave observations. The revisit time and spatial resolutions of this dataset are approximately three days and 25 km, respectively. Although it has been successfully applied in various contexts, its low level of detail renders it unsuitable for fine-scale or regional water mapping.

Global navigation satellite system reflectometry (GNSS-R), recognized as an efficient remote sensing tool, has garnered significant attention in recent years. GNSS-R operates in a multistatic mode, with the spaceborne system collecting forward-scattered signals from surface regions near specular point (SP) opportunistically. Due to the dynamic motion of both the transmitter and receiver orbits, the ground trajectories of satellites display a quasi-random distribution. GNSS-R collects information regarding surface characteristics by exploiting the unique properties of GNSS signals when reflected from the Earth's surface [18], [27], [28]. Compared to other remote sensing instruments, GNSS-R offers significant advantages in terms of cost-effectiveness and all-weather performance [29] and has been widely applied in ocean wind measurement [30], [31], altimetry [32], soil moisture estimation [33], [34], ice and snow monitoring [27], [35], and many other fields. Moreover, numerous advanced deep learning models have been applied for GNSS-R applications. For example, Chu et al. [36] proposed the heterogeneous multimodal deep learning (HMDL) method, which significantly improved the accuracy and robustness of GNSS-R sea surface wind speed estimation. Furthermore, Qiao and Huang [37] developed a Transformer-based WaveTransNet network, which captures long-range dependencies in the delay–Doppler map (DDM) through the Transformer encoder and incorporates an auxiliary parameter feature extraction branch, enhanced by an attention mechanism, to substantially improve GNSS-R data inversion performance for global significant wave height. A more comprehensive review of the applications of deep learning methods in the field of GNSS-R can be found in [38].

Furthermore, the substantial potential of GNSS-R for surface water mapping has been confirmed. The L-band signal employed by GNSS-R exhibits strong canopy penetration capability, as evidenced by the sensitivity of cyclone global navigation satellite system (CyGNSS) to the tributaries of the Amazon River. It can effectively detect water bodies even beneath dense vegetation cover [39]. Due to these characteristics, GNSS-R technology is extensively utilized in hydrological fields such as flood detection [40], wetland monitoring [41], [42], and inland water body mapping [18], [43], [44], [45], [46], [47], [48]. Recent studies have further improved the capability of GNSS-R for water detection by integrating multiple variables and applying advanced machine learning techniques. For instance, Yan et al. [18] integrated GNSS-R data with multisource auxiliary variables (e.g., soil moisture, vegetation optical depth, and geographic location), employed the bootstrap aggregation of regression trees (BARTs) method to estimate the monthly SWF at a spatial resolution of 0.025°, and validated the results with SWAMPS, GSW, GLAD, and ground measurement data. The results demonstrate that CyGNSS-based SWF estimation serves as an effective complement to existing microwave and optical data products, offering a broader spatial coverage and an improved accuracy. However, this method primarily relies on local information for estimation, making it challenging to capture the global contextual characteristics of water bodies. In contrast, semantic segmentation technology employs end-to-end learning to more effectively integrate neighborhood information, thereby more accurately capturing the spatial distribution of water bodies. In this regard, Chen and Yan [48] designed an enhanced U-Net architecture based on the Swin Transformer and context module, which effectively extracts water body distribution characteristics, further highlighting the potential and advantages of deep learning in SWF estimation tasks.

Although existing research demonstrates that combining GNSS-R data with multisource remote sensing data can significantly improve the accuracy of SWF estimation, substantial errors remain when relying solely on GNSS-R data [48]. The root cause of these errors lies in the fixed biases inherent in current model training methods. Traditional convolutional neural networks (CNNs) and Transformer-based models typically perform SWF estimation directly on GNSS-R data. However, these models exhibit notable deficiencies in handling nonstationary noise, which can lead to model instability or collapse, particularly in areas with strong noise or complex data distributions. While prior studies have somewhat mitigated this issue by incorporating multisource data (e.g., digital elevation models and vegetation indices), existing methods still struggle to meet practical needs when using only GNSS-R data. The inherent noise characteristics and observational uncertainties of GNSS-R data complicate the modeling process [34], [49], further limiting the model's generalization ability and stability. Fortunately, recent advances in denoising diffusion models show considerable promise in handling noisy data [50], [51], [52]. These models offer distinct advantages by simulating forward diffusion (adding noise) and reverse denoising (systematically removing noise), thereby capturing complex data distributions and mitigating noise effects. Consequently, we propose exploring the applicability of conditional diffusion models for SWF estimation using GNSS-R data.

Specifically, the diffusion-based model framework first introduces noise into the data through a forward diffusion process to simulate a noisy data distribution; subsequently, it removes the noise via a reverse diffusion process, ultimately generating high-precision SWF estimation results. To overcome the estimation challenges inherent in relying solely on GNSS-R data, we develop an SWF estimation framework named DiffWater in this study. The DiffWater framework decomposes the SWF estimation task into a series of forward and reverse diffusion steps. This method effectively learns a reasonable estimation process for surface water distribution, rather than directly modeling the complex and noisy GNSS-R data. With this decomposition strategy, DiffWater excels at balancing noise removal and data structure preservation. Simultaneously, we re-examine the training optimization objectives, offering an efficient and robust solution for the SWF estimation task that relies exclusively on GNSS-R data. Our main contribution is the proposal of a novel conditional diffusion model composed of conditional block and time noisy block, and details are given as follows.

1) To address the singularity problem commonly encountered by traditional conditional diffusion models in regression estimation tasks, we propose an improved framework called DiffWater. By reweighting the loss function, we effectively mitigate the singularity bias inherent in traditional noise prediction models, thereby enabling the conditional diffusion model to estimate SWF more accurately.
2) We designed a lightweight denoising model based on a dual-branch U-Net structure. This model extracts conditional features at multiple scales through the condition module and redesigns a lightweight time noisy module to jointly process noise and condition information across multiple levels. Finally, we explicitly fuse multiscale features, multiscale conditional embeddings, and implicit time embeddings of noisy images, significantly enhancing the model's inference efficiency and estimation accuracy.

II. Datasets

A. Calculation of CyGNSS Observables

This study utilized the CyGNSS Level 1 (L1) Version 3.2 dataset, which is accessible on NASA Earthdata (accessible at https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2832195379-POCLOUD/). The CyGNSS satellite constellation receives GPS L1 band signals reflected from the Earth's surface, generating a composite signal that comprises both coherent and incoherent scattering components, which are influenced by the surrounding surface roughness. Research by Ghasemigoudarzi et al. [53] indicates that water reflections are primarily coherent, particularly in areas with dense biomass surrounding the surface water. By assuming that coherent reflections dominate over land, it is possible to derive the surface reflectivity (Γ) from CyGNSS data as

$$\Gamma = \frac{\sigma(R_t + R_r)^2}{4\pi(R_t R_r)^2} \tag{1}$$

where σ is the bistatic radar cross section, and R_t and R_r are the distances from SP to the transmitter and the receiver, respectively, which are accessible in the abovementioned CyGNSS L1 dataset. For each CyGNSS σ observation (an 11×17 pixel image), we utilize the peak value from the image as the representative BRCS value for our analysis.

In addition to Γ, an indicator of CyGNSS DDM's spread [pixel number (PN)] that depicts surface roughness is employed as well [43], [54]. Yan and Huang [55] define the PN as the total number of pixels that are above a specified threshold (DDMthres) within a normalized DDM. Here, the value of DDMthres is set to 0.2.

The data quality control scheme is based on the methodology articulated in [33]. Data marked with the quality indicator "SP in the sidelobe" are omitted due to their low confidence in antenna gain. Based on the geographic locations of the SPs, the CyGNSS data are aggregated into monthly scales with a spatial resolution of 0.025° × 0.025°, encompassing the region between 37.5°S and 37.5°N.

B. Reference Data

GSW [11] and GLAD [12] both generated from Landsat optical imagery are among the most comprehensive sources of surface water data, spanning extended periods (GSW from 1984 to 2021 and GLAD from 1999 to 2021), with the spatial and temporal resolutions of 30 m and one month, respectively. The analysis focuses on data from August 2018 to December 2021 (a total of 41 months). Given the lack of consensus on the superior dataset, this study utilized both GSW and GLAD as references to derive surface water fraction and merged them into a unified reference dataset, the GSW fraction (GSWF). It is important to note that GSW contains only classification (water body/land) information, and the SWF for each 0.025° grid was calculated as the percentage of water pixels within the grid. GLAD provides SWF directly, and the average for each grid was computed. However, they may face challenges such as insufficient coverage due to cloud contamination or dense canopy. To secure the quality of GSWF, a 0.025° grid is regarded as valid only if it contains enough (over 90%) GLAD/GSW data. Otherwise, it is classified as an occluded area, and a gridded cloud cover mask is created. Consequently, the cloud-occluded portions of the gridded CyGNSS data are masked to prevent the model from accurately learning the mapping relationship between CyGNSS data and surface water.

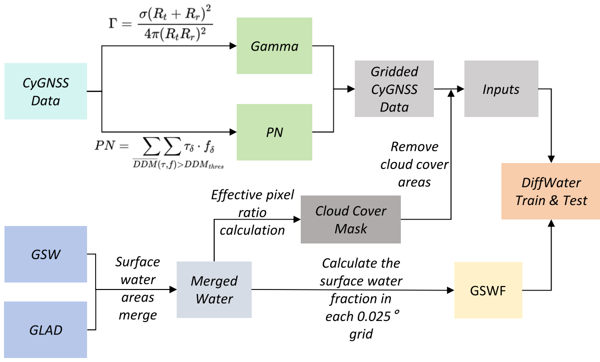

Fig. 1. Dataset processing flowchart consisting of CyGNSS data creating, reference data merging, and training/test splitting.

C. Data Collocation

The flowchart describing data preparation is presented in Fig. 1. This study aims to provide SWF inversion with a spatial resolution of 0.025° and a temporal step size of one month. Accordingly, GNSS-R observation data and SWF reference data were matched based on these specifications. Both GNSS-R data and SWF reference data were spatially averaged to grid cells of 0.025°, followed by uniform up/down sampling, normalization, and stitching. Finally, a cropping algorithm was employed to extract multiple image blocks of 256 × 256 pixels (6.4° × 6.4°). The dataset spans from August 2018 to December 2021, comprising 41 months of data, with the first 24 months used for training and the remaining 17 months reserved for testing.

III. Method

A. Overall Diffusion Framework

The training procedure comprises two phases: a forward diffusion phase and a backward denoising phase. The training and sampling procedure of the overall framework is illustrated in Fig. 2. During the training phase, the model progressively adds noise at each time step t and optimizes by minimizing the backward loss, enabling the model to learn the denoising function for each time step based on conditional information. This step-by-step learning approach ensures that the model accurately captures the underlying data structure while gradually reducing noise. In the sampling phase, the trained model generates clean samples through the reverse diffusion process. Starting with a noise prior (generate noisy images at a known time t), the model incrementally removes noise at each time step t and guides the generation process using conditional information (GNSS-R data). This iterative optimization allows the model to reconstruct high-fidelity samples consistent with the target distribution.

1) Forward Diffusion Process: We denote the reference SWF image as x and model the forward process of the diffusion model through a Markov chain. Specifically, the initial state x_0 is devoid of noise, while x_T represents pure noise. The state x_t at time step t is expressed as $x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{\beta_t}\epsilon_{t-1}$, where $\epsilon_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ represents the Gaussian noise and t is a scalar randomly sampled during the model training process. The parameter α_t denotes the weighting coefficients of the signal and noise components in x_t, satisfying $\alpha_t + \beta_t = 1$. Specifically, a cosine time schedule is employed to compute the noise content at each noise level t. The cumulative signal-to-noise ratio is defined as $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$, which can be derived to $\alpha_t = (\bar{\alpha}_t / \bar{\alpha}_{t-1})$. The noise schedule is constructed as $\bar{\alpha}_t = (f(t)/f(0))$, where $f(t) = \cos^2(((t/T + s)/(1+s)) \cdot (\pi/2))$, with s initialized to 8×10^{-3}. To ensure numerical stability and prevent singularities near $t = T$ at the conclusion of the diffusion process, $\alpha_t \geq 0.001$. Here, T denotes a hyperparameter that specifies the total number of noise levels. Therefore, the forward noise addition process is expressed as $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)\mathbf{I})$. Based on the reparameterization technique and the principles of Markov chain and normal distribution, the following formula can be derived:

$$\begin{aligned} x_t &= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1} \\ &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\epsilon_{t-1} \\ &= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}\alpha_{t-2}}\bar{\epsilon}_{t-2} \\ &= \cdots \\ &= \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\bar{\epsilon} \end{aligned} \tag{2}$$

where ϵ_t, $\epsilon_{t-1} \sim \mathcal{N}(0, I)$ and $\bar{\epsilon}_{t-2}$ is their merged result. According to the additivity of the independent Gaussian distribution, i.e., $\mathcal{N}(0, \sigma_1^2 I) + \mathcal{N}(0, \sigma_2^2 I) \sim \mathcal{N}(0, (\sigma_1^2 + \sigma_2^2)I)$, therefore, any noised image x_t satisfies

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I) \tag{3}$$

when $t \to T$, x_t can converge to the standard normal distribution $\mathcal{N}(0, I)$, consistent with the original design intention. For remote sensing tasks, effective noise scheduling can make the diffusion process more natural and efficient while accelerating model convergence and enhancing performance. Fig. 3 illustrates the visual effects of various noise scheduling strategies throughout the diffusion process. In Section IV-B, we examine and discuss the impact of various noise scheduling strategies on SWF estimation performance.

2) Backward Diffusion Process: This process aims to obtain the reversed transition probability $q(x_{t-1}|x_t)$, thereby gradually restoring the image $\hat{x_0}$ from the noise. Based on Bayes' theorem, the posterior distribution of the forward diffusion process, $q(x_{t-1}|x_t, x_0)$, is expressed in terms of $\tilde{\beta}_t$ and $\tilde{\mu}_t(x_t, x_0)$, defined as follows:

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{4}$$

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}x_t \tag{5}$$

$$\begin{aligned} q(x_{t-1}|x_t, x_0) &= \frac{q(x_t|x_{t-1}, x_0)q(x_{t-1}|x_0)}{q(x_t|x_0)} \\ &\propto \mathcal{N}\left(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I\right). \end{aligned} \tag{6}$$

Since $q(x_{t-1}|x_t, x_0)$ is challenging to compute explicitly, we approximate it using a neural network, denoted as $p_\theta(x_{t-1}|x_t)$, where θ represents the learnable parameters of

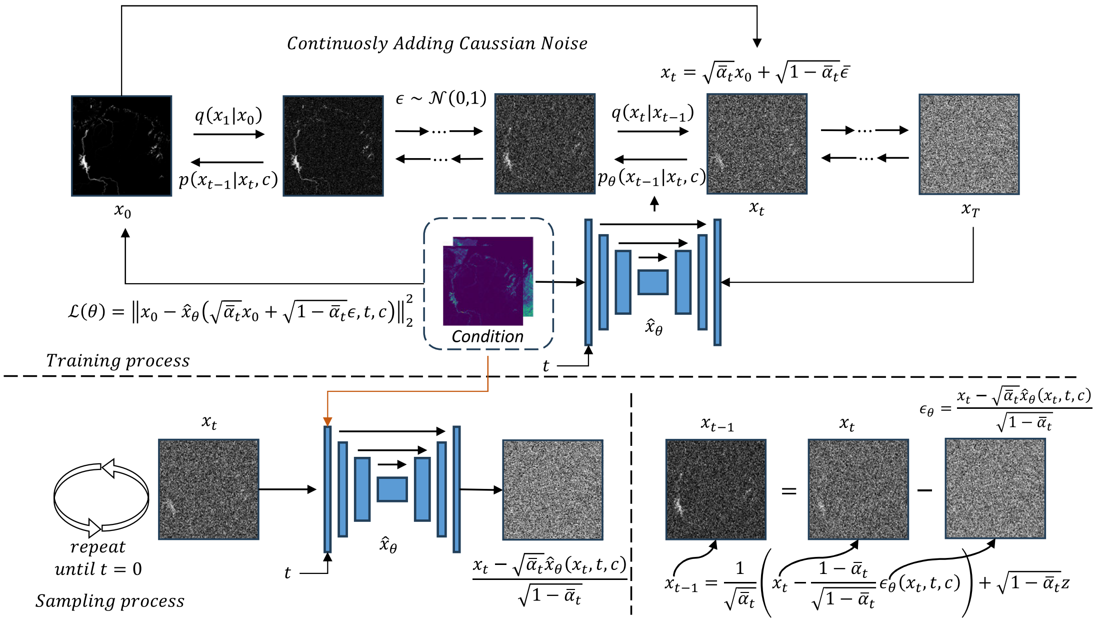

Fig. 2. Training and sampling procedure of DDPM.

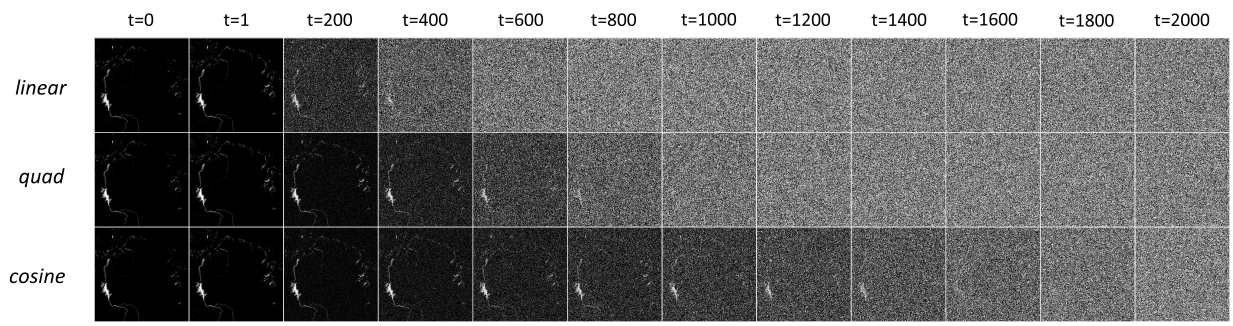

Fig. 3. Visualization of the noise adding process $T = 2000$ is defined, and the noise adding methods include linear, quad, and cosine.

the denoising model (Algorithm 1). The optimization objective is to minimize the discrepancy between $p_\theta(x_{t-1}|x_t)$ and $q(x_{t-1}|x_t)$. To achieve this, we derive $q(x_{t-1}|x_t, x_0)$ and set the variance of $p_\theta(x_{t-1}|x_t)$ equal to that of $q(x_{t-1}|x_t)$ at the same time step t.

Through the posterior probability distribution $q(x_{t-1} \mid x_t, x_0)$ of the forward process, we define the probability distribution of the reverse process as $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t, c), \Sigma_\theta(x_t, t))$, where $\mu_\theta(x_t, t, c)$ is a function of x_t, t, and the conditioning variable c. This definition is based on the assumption that for equivalent states in the forward and reverse processes, the variance of the reverse process is consistent with that of the forward process. Define $\mu_\theta(x_t, t, c)$ as a function of x_t, t, and the signal condition c. Similarly, express $\mu_\theta(x_t, t, c)$ in terms of $\tilde{\mu}_t(x_t)$ using $\mu_\theta(x_t, t, c) = (1/\sqrt{\alpha_t})(x_t - ((1-\alpha_t)/(1-\bar{\alpha}_t)^{1/2})\epsilon_\theta(x_t, t, c))$. The optimization objective function is then formulated using the Kullback–Leibler (KL) divergence as

$$
\begin{aligned}
\mathcal{L}(\theta) &= D_{\mathrm{KL}}(q(x_0, x_1, \ldots, x_T) \parallel p_\theta(x_0, x_1, \ldots, x_T)) \\
&= -\mathbb{E}_{q(x_0,x_1,\ldots,x_T)}[-\log p_\theta(x_0, x_1, \ldots, x_T)] + \tau \\
&= -\mathbb{E}_{q(x_0,x_1,\ldots,x_T)}\left[-\log p(x_T) - \sum_{t=1}^{T}\log\frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right] \\
&\quad + \tau
\end{aligned}
\tag{7}
$$

where τ denotes a constant independent of θ. To achieve the optimization goal, it suffices to minimize the following objective function:

$$
\begin{aligned}
&\arg\min_\theta D_{\mathrm{KL}}(q(x_t|x_{t-1}) \parallel p_\theta(x_{t-1}|x_t)) \\
&= \arg\min_\theta \frac{1}{2\sigma_q^2(t)}\left\|\tilde{\mu}_t(x_t, t) - \mu_\theta(x_t, t, c)\right\|_2^2 \\
&= \arg\min_\theta \frac{1}{2\sigma_q^2(t)} \cdot \frac{(1-\alpha_t)^2}{\alpha_t(1-\bar{\alpha}_t)}\|\epsilon - \epsilon_\theta(x_t, t, c)\|_2^2
\end{aligned}
\tag{8}
$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$. Existing unconditional and conditional diffusion models typically adopt a noise prediction approach, wherein a neural network is trained to estimate the Gaussian noise added at each step of the forward diffusion process. The loss function can be defined as $L_{\mathrm{simple}}(\theta) = \mathbb{E}_{x_0,c,\epsilon,t}[\|\epsilon -$

$\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, c, t)\|^2]$ and is used to optimize the model parameters θ for image generation tasks. For the remote sensing SWF estimation task, the conditional input information is highly correlated with the target SWF. As a result, adopting the noise prediction approach increases the complexity of training and may even lead to model instability or failure. Although ϵ and x_0 are mutually dependent and can be derived from $x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$, altering the regression target affects the scale of the loss function, subsequently influencing the training dynamics. Given the similarity in data distribution between the conditional GNSS-R inputs and SWF images, data prediction proves to be significantly simpler than noise prediction. This simplification enhances model convergence and improves the accuracy of SWF estimation. Therefore, we redefine the model as a data prediction model, denoted as $\hat{x}_\theta$, to replace the traditional noise prediction model ϵ_θ

$$\begin{aligned}
L_{\text{simple}}(\theta) &= \mathbb{E}_{x_0,c,\epsilon,t}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t, c\right)\right\|_2^2\right] \\
&= \mathbb{E}_{x_0,c,\epsilon,t}\left[\left\|\frac{1}{\sqrt{1-\bar{\alpha}_t}}\left(x_t - \sqrt{\bar{\alpha}_t}x_0\right)\right.\right. \\
&\qquad \left.\left. - \frac{1}{\sqrt{1-\bar{\alpha}_t}}\left(x_t - \sqrt{\bar{\alpha}_t}\hat{x}_\theta(x_t,t,c)\right)\right\|_2^2\right] \\
&= \mathbb{E}_{x_0,c,\epsilon,t}\left[\frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\left\|x_0 - \hat{x}_\theta(x_t,t,c)\right\|_2^2\right]. \qquad (9)
\end{aligned}$$

Here, we define $\text{SNR} = (\bar{\alpha}_t/(1-\bar{\alpha}_t))$ as the signal-to-noise ratio. From the perspective of the Fourier domain, the model initially removes the high-frequency components (texture features) of the target SWF during the noise addition process (forward process), followed by the removal of low-frequency components (overall image features). Consequently, during the inference phase, the model first reconstructs the low-frequency components (overall image features) of the SWF and subsequently generates the high-frequency components (texture features). Considering the aforementioned noise-weighted loss function, it becomes apparent that the model exhibits a singularity at $t = T$. As $t \to T^-$, we have $\lim_{t\to T^-}\text{SNR}(t) = \lim_{t\to T^-}(\bar{\alpha}_t/(1-\bar{\alpha}_t)) = 0$. The model's denoising ability is constrained at the initial moment $t = T$, and average brightness issues may arise at different sampling stages, which can be detrimental to the SWF estimate task. At $t = 0$, the model distribution degenerates into a singular distribution, specifically a Gaussian distribution with zero variance, $\lim_{t\to 0^+}\text{SNR}(t) = \lim_{t\to 0^+}(\bar{\alpha}_t/(1-\bar{\alpha}_t)) = +\infty$. This step is straightforward for the model to learn, and increasing in weight is unnecessary. The singularity at $t = 0$ represents an inherent characteristic of the diffusion model. Provided that appropriate sampling techniques are employed, this singularity does not need to be circumvented [56], [57]. Consequently, it is necessary to reduce the weight as $t \to 0^+$ and increase the weight as $t \to T^-$. Following this principle, we adopt a reweighted loss function to ensure that the model learns noise removal with equal emphasis across all time steps t.

By removing the coefficient $(\bar{\alpha}_t/(1-\bar{\alpha}_t))$, the loss function of the reweighted diffusion model can be reformulated as

$$L_{\text{simple}}(\theta) = \left\|x_0 - \hat{x}_\theta(x_t,t,c)\right\|_2^2. \qquad (10)$$

3) Sampling Process: In the inference process, which is also known as the sampling process (Algorithm 2). The new image x_0 can be generated from either Gaussian noise or a noisy image x_t by iteratively sampling x_{t-1} until $t = 1$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t,c)\right) + \sqrt{1-\alpha_t}z \qquad (11)$$

where $z \sim \mathcal{N}(0,\mathbf{I})$, and it is assumed that $z = 0$ when $t = 1$. According to the forward noise addition formula $x_t = \sqrt{\bar{\alpha}_t}\hat{x}_\theta + (1-\bar{\alpha}_t)^{1/2}\epsilon_\theta$, we can get $\epsilon_\theta = ((x_t - \sqrt{\bar{\alpha}_t}\hat{x}_\theta)/(\sqrt{1-\bar{\alpha}_t}))$. We use the original prediction model to replace the noise prediction model, and the corrected expression is

$$\begin{aligned}
x_{t-1} = &\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\cdot\frac{x_t - \sqrt{\bar{\alpha}_t}\hat{x}_\theta(x_t,t,c)}{\sqrt{1-\bar{\alpha}_t}}\right) \\
&+ \sqrt{1-\alpha_t}z. \qquad (12)
\end{aligned}$$

B. Conditional Denoising Model

The standard U-Net model has made significant progress in large-scale surface water extraction tasks. However, when applied to diffusion models, the conventional U-Net architecture often suffers from an excessive number of parameters and computational overhead caused by redundant floating-point operations. These inefficiencies can impede both the training and inference processes. To address this, we introduce an efficient conditional denoising model, DiffWater, which integrates a simplified conditional module and a time-noise block (TNB) module as core components, as illustrated in Fig. 4. The model employs a decoupled feature extraction mechanism that decomposes the original feature space into multiple independent factors. These factors represent distinct attributes of the data, enhancing the model's generalization ability and interpretability. The DiffWater architecture consists of two main components.

1) *Dual-Branch Conditional and Temporal Noise Encoder:* This module extracts the spatial distribution of SWF from the conditional input and captures the temporal features corresponding to the noise level t.
2) *Temporal Denoising Decoder:* The spatial and temporal features are concatenated and passed to the decoder, which performs step-by-step, noise-free SWF reconstruction.

Supervised by the restoration mean squared error (mse) loss, the model generates high-precision SWF estimation results. By reducing unnecessary computations and enhancing efficiency, DiffWater accelerates both the training and inference processes while ensuring reliable performance for surface water extraction tasks. This design ensures the model's efficiency while preserving performance, offering a robust solution for accurate SWF extraction from noisy data.

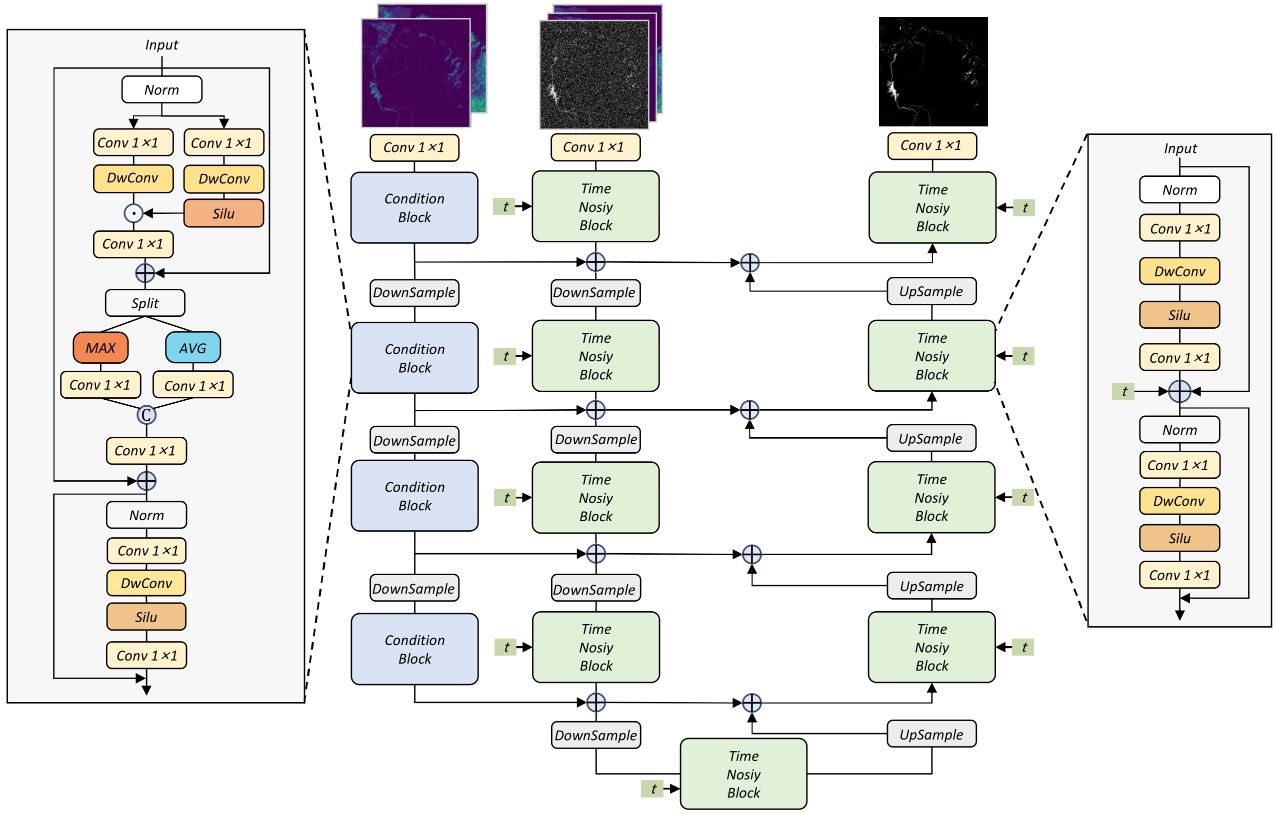

Fig. 4. Overall architecture of the DiffWater conditional denoising model. The model consists of three parts: 1) a conditional encoder; 2) a temporal noise encoder; and 3) a denoising decoder. The conditional encoder extracts the spatial features of GNSS-R conditional information, while the temporal noise encoder extracts temporal features corresponding to the noise level t and spatial features of the noisy image. These features are subsequently input into the denoising decoder, which estimates the data distribution of SWF under the supervision of the recovery loss.

1) Time Noisy Block: The training of diffusion models typically involves multiple iterative steps, necessitating significant computational resources. Consequently, we prioritized simplicity and efficiency in our design. Building upon the U-Net architecture, we employed depthwise separable convolution to replace the conventional 3×3 convolution module and selected the sigmoid linear unit (Silu) as the activation function. To incorporate temporal information t, we utilized sinusoidal encoding and implemented a multilayer perceptron (MLP) to facilitate the integration of this time information into the intermediate layers of the module. The complete TNB comprises two sets of repeated convolution modules, and its forward process is outlined as

$$F_m = W^c_{1\times1}\left(\sigma\left(W^d_{3\times3}\left(W^c_{1\times1}(\mathrm{Norm}(F_{\mathrm{in}}))\right)\right)\right) \tag{13}$$

$$F^f_m = F_{\mathrm{in}} + F_m + \mathrm{Pos}(\mathrm{MLP}(t)) \tag{14}$$

$$F_{\mathrm{out}} = W^c_{1\times1}\left(\sigma\left(W^d_{3\times3}\left(W^c_{1\times1}\left(F^f_m\right)\right)\right)\right) + F^f_m \tag{15}$$

where $W^c_{1\times1}$ denotes a 1×1 pointwise convolution, $W^d_{3\times3}$ indicates a 3×3 depthwise convolution (DwConv), Norm refers to the layer normalization, and σ represents the activation function, specifically the Silu. Furthermore, $F_{\mathrm{in}} \in \mathbb{R}^{H\times W\times C}$ represents the input features, $F_{\mathrm{out}} \in \mathbb{R}^{H\times W\times C}$ denotes the output features, and F_m and F^f_m are intermediate variables.

2) Conditional Block: We have developed a conditional module based on the TNB. This module is designed to accurately capture conditional information by considering both spatial and channel features. To enhance the model's ability to extract conditional information, we implemented efficient spatial attention (SA) and channel attention mechanisms. For SA, we utilize a 1×1 convolution to reproject channel information, followed by a 3×3 DwConv to aggregate information from adjacent pixels. This approach enables the model to effectively capture local features while preserving the details of the spatial structure, thereby improving its understanding of local context. Additionally, we incorporated a gating mechanism to further enhance the encoding capability of spatial information. This allows the model to extract and utilize spatial information more comprehensively, leading to improved overall performance. The forward process of SA can be expressed as follows:

$$\begin{aligned} F_{\mathrm{sa}} = {} & W^c_{1\times1}\left(\sigma\left(W^d_{3\times3}\left(W^c_{1\times1}(\mathrm{Norm}(F_{\mathrm{in}}))\right)\right)\right) \\ & \odot W^d_{3\times3}\left(W^c_{1\times1}(\mathrm{Norm}(F_{\mathrm{in}}))\right) + F_{\mathrm{in}}. \end{aligned} \tag{16}$$

Then, we introduce a lightweight channel attention mechanism to enhance the robustness of conditional information extraction. This mechanism comprises two branches: a global channel attention branch and a local channel attention branch. Initially, the input features are partitioned into two segments along the channel dimension. Each segment undergoes global average pooling (GAP) and global maximum pooling (GMP). Following this, a 1×1 convolution is applied to establish explicit correlations between channels within each segment. Ultimately, the features derived from both branches are concatenated along the channel dimension, resulting in robust features that integrate both global and local information. The average pooling emphasizes the mean representation of the extracted features, thereby preserving essential overall characteristics of the image, while the maximum pooling focuses on capturing detailed texture features. The significance of varying channels can be emphasized through GAP and GMP

$$F_{\max}, F_{\text{avg}} = \text{Split}(F_{\text{sa}}) \tag{17}$$

$$F_{\text{ca}} = W_{1\times1}^{c}\left(\Phi\left(W_{1\times1}^{c}(\text{GMP}(F_{\max})), W_{1\times1}^{c} \times \left(\text{GAP}(F_{\text{avg}})\right)\right)\right) \tag{18}$$

where $F_{\max} \in \mathbb{R}^{H\times W\times(C/2)}$ and $F_{\text{avg}} \in \mathbb{R}^{H\times W\times(C/2)}$ represent the results obtained from splitting the input features by channel using the Split operation, the operator Φ denotes the concatenation of features along the channel dimension, GAP refers to the global average pooling, and GMP denotes the global maximum pooling.

Subsequently, a series of lightweight convolution modules are employed to recalibrate the spatial and channel features of the conditional image. Each module consists of a normalization layer, two point-wise convolution layers, a DwConv, and an activation layer, as detailed in the following:

$$F_{\text{out}} = W_{1\times1}^{c}\left(\sigma\left(W_{3\times3}^{d}\left(W_{1\times1}^{c}(\text{Norm}(F_{\text{ca}}))\right)\right)\right) + F_{\text{ca}}. \tag{19}$$

3) Overview: The overall model is structured as a two-branch U-Net. Unlike previous studies, this model integrates conditional, noise, and temporal information while introducing new branches for the independent extraction of conditional information, thereby improving its capability to effectively eliminate noise. To maintain a lightweight design, our down-sampling module utilizes convolutions with a kernel size of 2×2 and a stride of 2, while the upsampling module employs convolutions with a kernel size of 1×1 in combination with PixelShuffle to minimize information loss. In alignment with the U-Net architecture, we incorporate skip connections; however, we replace the conventional channel concatenation approach with element-wise addition. This direct addition of skip connections introduces no additional parameters, whereas channel concatenation requires the learning of extra parameters to adjust feature channel dimensions. By directly adding low-level features to high-level features, we achieve more efficient fusion, as opposed to channel concatenation, which demands more complex transformations for feature integration. Additionally, prior research on GNSS-R has highlighted the advantages of direct fusion, further supporting our design choice.

Algorithm 1 Training DiffWater

Input: Conditional GNSS-R data $\boldsymbol{c}$, noisy water image $\boldsymbol{x}_t$, noise level t

Output: $\hat{x}_0$ estimated by conditional denoising model $\hat{x}_\theta$

1: **repeat**
2: $(c, x_0) \sim q(c, x)$
3: $t \sim \text{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take a gradient descent step on $\nabla_\theta \left\| \boldsymbol{x}_0 - \hat{\boldsymbol{x}}_\theta(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t, \boldsymbol{c}) \right\|_2^2$
6: **until** converged

Algorithm 2 Inference With Iterative Refinement

Input: Conditional GNSS-R data $\boldsymbol{c}$, Gaussian noise $\boldsymbol{x}_T$, noise level t

Output: $\boldsymbol{x}_0$ estimated by pre-trained conditional denoising model

1: $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \frac{\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t}\hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, c, t)}{\sqrt{1-\bar{\alpha}_t}}\right) + \sqrt{1-\alpha_t}\mathbf{z}$
5: **end for**
6: **return** $\boldsymbol{x}_0$

IV. Experiments

A. Experimental Platform Parameter Settings

All experiments were conducted on a workstation equipped with an AMD Ryzen 7 7800X3D 8-core processor, 64 GB of memory, and an NVIDIA GeForce RTX 4090 D GPU (24 GB of memory). The operating system used was Ubuntu 22.04, and all networks were implemented in PyTorch 2.4.0 with CUDA 11.8. The Adam optimizer was selected as the initial optimizer, with an initial learning rate set to $5e^{-4}$. During the training phase, the batch size was set to 8, and the training process involved approximately 2 million steps. Additionally, we initialized the model weights using the Kaiming normal distribution and applied the exponential moving average (EMA) during the training process, starting from the first iteration and updating with a decay rate of 0.999 after each iteration, which helps stabilize the training process and prevent overfitting. During the diffusion process, we set the maximum noise step to 2000. To quantitatively analyze and compare the effectiveness of the proposed method with other methods, five evaluation metrics were used: mean absolute error (MAE), root-mean-squared error (RMSE), structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and the correlation coefficient (R). Lower values of MAE and RMSE are preferred, while higher values of SSIM, PSNR, and R are desired.

B. Different Noise Adding Methods

Table I presents the impact of various noise schedules on the diffusion process. It can be observed that the cosine schedule yields the best SWF estimation performance, while the linear schedule exhibits the poorest performance. Simultaneously, Fig. 5 illustrates the variation in $\bar{\alpha}_t$ during the noise addition process. For data prediction, the noise level near $t = T$ must be considered. During the data prediction process, the loss function is reweighted. It is crucial that the prediction at each step is considered important for the model. In comparison to the cosine schedule, the linear and quadratic $\bar{\alpha}_t$ values decay to zero much more rapidly, resulting in faster information degradation than necessary, which is undesirable. The cosine schedule is designed to have a linear decrease in $\bar{\alpha}_t$ in the

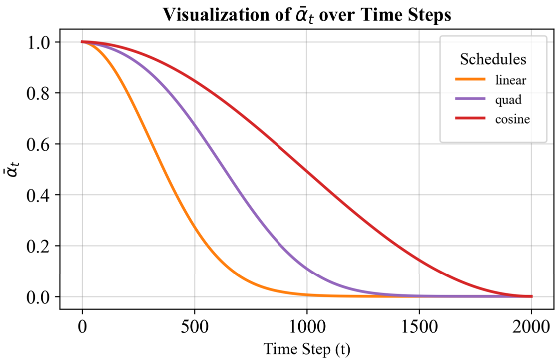

Fig. 5. Variation of $\bar{\alpha}_t$ across different noise addition methods: $\bar{\alpha}_t$ represents the proportion of the original image. A smoother curve in the changes of $\bar{\alpha}_t$ indicates a more balanced noise addition process over time.

TABLE I

QUANTITATIVE COMPARISON OF SWF ESTIMATION PERFORMANCE ACROSS VARIOUS NOISE SCHEDULES IN THE DIFFUSION PROCESS: MAXIMUM NOISE STEP SET TO 2000 FOR ALL SCHEDULES, WITH LINEAR AND QUADRATIC APPROACHES UTILIZING THE SAME DENOISING MODEL AS COSINE

Schedule	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑	R ↑
Linear	3.799%	7.337%	0.9800	47.7685	0.891
Quad	3.477%	6.502%	0.9804	47.8284	0.909
Cosine	2.852%	4.987%	0.9839	48.0647	0.946

TABLE II

QUANTITATIVE COMPARISON OF DIFFWATER'S NOISE PREDICTION AND DATA PREDICTION IN SWF REGRESSION PERFORMANCE: NOISE PREDICTION UTILIZES NOISE AS THE TRAINING TARGET FOR THE DENOISING MODEL, WHILE DATA PREDICTION EMPLOYS NOISE-FREE REFERENCE IMAGES. ALTHOUGH BOTH APPROACHES CAN BE EQUIVALENTLY TRANSFORMED, DATA PREDICTION EXCLUDES SNR CONSIDERATIONS COMPARED TO THE NOISE PREDICTION OBJECTIVE FUNCTION

Regression Target	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑	R ↑
Noise prediction	3.910%	9.594%	0.9819	47.8405	0.801
Data prediction	2.852%	4.987%	0.9839	48.0647	0.946

middle of the process, with minimal change near the extreme values of $t = 0$ and $t = T$ to prevent sudden fluctuations in the noise level. The experimental results demonstrate that the cosine schedule is the optimal solution.

C. Different Prediction Methods

Table II presents a quantitative comparison of the SWF regression performance for noise prediction and data prediction during model training. We experimentally assess the impact of the proposed data prediction and noise prediction on SWF regression performance. As anticipated from the training objective (described in Section III-A2), the reweighted data prediction outperforms noise prediction. This is because data prediction alleviates the singularity problem, ensuring that the model is as accurate as possible at each step of SWF estimation. Furthermore, it is easier to predict an outcome that closely resembles the input, whereas noise prediction is less accurate near the singular point, often leading to inaccurate SWF estimation.

TABLE III

QUANTITATIVE COMPARISON OF VARIOUS CONDITIONAL INFORMATION PROCESSING METHODS: ALL THREE APPROACHES USE COSINE SCHEDULE AS THE BENCHMARK FOR DATA PREDICTION

Condition Process	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑	R ↑
Concat	3.434%	6.422%	0.9794	47.4081	0.920
Decouple	2.974%	5.353%	0.9833	48.0216	0.938
Concat & Decouple	2.852%	4.987%	0.9839	48.0647	0.946

D. Conditional GNSS-R Data Processing

We quantitatively explored the relationship between noise and conditional information. Conventional diffusion models concatenate noise and conditional information and directly use temporal noise modules for condition-guided noise removal. However, this approach can not ensure that noise does not disrupt the conditional information. To address this, we examined three approaches for handling the interaction between noise and conditional information.

1) *Concat:* Conditional information is directly concatenated with noise and processed using a single module.
2) *Decouple:* Noise and conditional information are decoupled and processed separately using independent modules.
3) *Concat and Decouple:* Noise and conditional information are concatenated while also decoupling the conditional information for separate processing, combining the strengths of the first two methods.

The experimental results are presented in Table III. The decoupling method demonstrates a positive impact on model performance. Additionally, incorporating independent decoupling modules into the standard method further enhances the model performance. Compared to the Concat method, the Decouple method achieves a 1.069% reduction in RMSE. The Concat and Decouple methods achieve the best performance, reducing RMSE by 1.435% and significantly outperforming other methods.

E. Ablation Experiments

Table IV summarizes the results of ablation experiments conducted on various modules, where CA refers to the inclusion of the channel attention module in the condition module, SA represents the use of the SA module, and DwConv signifies the incorporation of a lightweight module. The results demonstrate that using the CA module enhances the accuracy of SWF estimation, as evidenced by a reduction in MAE from 3.562% to 3.318%, with only a marginal increase in the number of parameters (from 46.884 to 47.408 M) and floating-point operations (from 69.408 to 70.486 G). This improvement can be attributed to the CA module's ability to strengthen the robustness of the model in extracting conditional information. The SA module significantly improves the overall quality of SWF estimation, as indicated by an increase in SSIM from 0.9783 to 0.9804. This is due to the SA module's capacity in enhancing the model's understanding of local context. The combination of both modules yields a significant improvement in estimation accuracy, as reflected by a reduction in MAE to 3.289% and an increase in SSIM to 0.9809 while also enhancing overall image quality. The introduction of the lightweight

TABLE IV

Quantitative Comparison of Ablation Experiments Across Different Modules: CA Refers to Channel Attention, SA Refers to Spatial Attention, and DwConv Indicates the Use of Depthwise Separable Convolution in Place of Standard 3 × 3 Convolution

CA	SA	DwConv	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑	R ↑	Param(M) ↓	GFLOPS(G) ↓
			3.562%	6.996%	0.9783	47.3819	0.883	46.884	69.408
√			3.318%	6.277%	0.9795	47.8741	0.908	47.408	70.486
	√		3.367%	6.343%	0.9804	47.8654	0.904	47.951	72.771
√	√		3.289%	5.972%	0.9809	47.8911	0.918	48.475	73.849
		√	3.222%	5.823%	0.9829	48.0498	0.932	17.667	20.982
√	√	√	2.852%	4.987%	0.9839	48.0647	0.946	18.550	23.205

TABLE V

Quantitative Comparison of Different Methods for SWF Estimation: Average Overall Metrics Across All Regions. Training Period for All Regions: August 2018–July 2020. Testing Period for All Regions: August 2020–December 2021

Method	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑	R ↑	Param(M) ↓	GFLOPS(G) ↓
Unet	3.181%	7.358%	0.973	46.638	0.858	24.433	7.798
LinkNet	3.302%	7.748%	0.954	45.027	0.842	21.769	5.423
PspNet	4.704%	10.988%	0.950	43.916	0.682	1.498	2.307
Deeplabv3+	4.087%	9.481%	0.958	44.650	0.764	26.674	9.175
CRFT	2.809%	6.364%	0.977	46.912	0.894	15.406	5.902
Pix2Pix	3.438%	9.249%	0.976	47.013	0.775	11.779	49.784
TransGAN	2.614%	6.763%	0.987	47.954	0.880	32.499	9.383
DDPMBase	3.910%	9.594%	0.982	47.841	0.801	32.197	58.367
DiffWater	2.852%	4.987%	0.984	48.065	0.946	18.550	23.205

DwConv not only accelerates model convergence compared to traditional convolutions but also reduces the model's sensitivity to noisy data, thereby enhancing the robustness of the method. DwConv further enhances the model's overall denoising capability, enabling more accurate predictions even under high-noise conditions. The combination of CA, SA, and DwConv results in a substantial improvement in the model's ability to accurately estimate SWF, as evidenced by a reduction in MAE to 2.852% and an increase in SSIM to 0.9839 while simultaneously improving parameter efficiency (reducing the number of parameters to 18.550 M) and lowering computational complexity (reducing giga floating-point operations per second (GFLOPS) to 23.205 G).

F. Comparative Experiment

Fig. 6 presents the historical river system maps of two regions: the Congo Basin (a) and the Amazon Basin (b), alongside the interannual variations in these areas. It is important to note that cloud cover issues may lead to missing reference data in certain regions, as the reference data are derived from optical images. Despite these limitations, the seasonal changes in the river systems at these locations remain evident, with noticeable differences in river system distributions across various periods. Therefore, the design of SWF estimation models must focus on capturing monthly variations rather than learning a fixed, uniform pattern. Failure to account for signal changes during the inference stage may hinder the accurate representation of SWF distribution and lead to imprecise SWF estimations.

To comprehensively evaluate the performance of the model, we conducted a comparative analysis of DiffWater against widely adopted deep learning models, including semantic segmentation-based architectures such as U-Net [58], LinkNet [59], PspNet [60], and Deeplabv3+ [61]. Additionally, we compared DiffWater with the Transformer-based CFRT model [48], which is specifically designed for SWF estimation, as well as with classic GAN-based models such as Pix2Pix [62] and its variant TransGAN [63], and the noise prediction-based denoising diffusion probabilistic model (DDPM) [50]. The parameter configurations of these compared methods are detailed as follows and the global SWF retrieval results are summarized in Table V.

1) *U-Net:* It employs a symmetric encoder–decoder architecture with skip connections for efficient feature extraction and reconstruction. ResNet34 serves as the backbone and a Sigmoid activation function is applied for SWF estimation.
2) *LinkNet:* LinkNet integrates an encoder–decoder structure with residual connections to enhance feature extraction and segmentation accuracy, demonstrating robustness in surface water detection. ResNet34 is used as the backbone and a sigmoid activation function applied for SWF estimation.
3) *PspNet:* PspNet incorporates a pyramid pooling module to capture multiscale contextual information, improving image comprehension and segmentation accuracy. The first two layers of ResNet34 serve as the encoder, while the original decoder is retained. A Sigmoid activation function is applied for SWF estimation.
4) *Deeplabv3+:* This network combines dilated convolutions with a decoder structure to enhance multiscale feature extraction and detail preservation, making it suitable for complex scene parsing. ResNet34 is used as the encoder, and a sigmoid activation function is applied for SWF estimation.
5) *CRFT:* It is specifically designed for SWF regression using CyGNSS data, excluding multisource constraints.

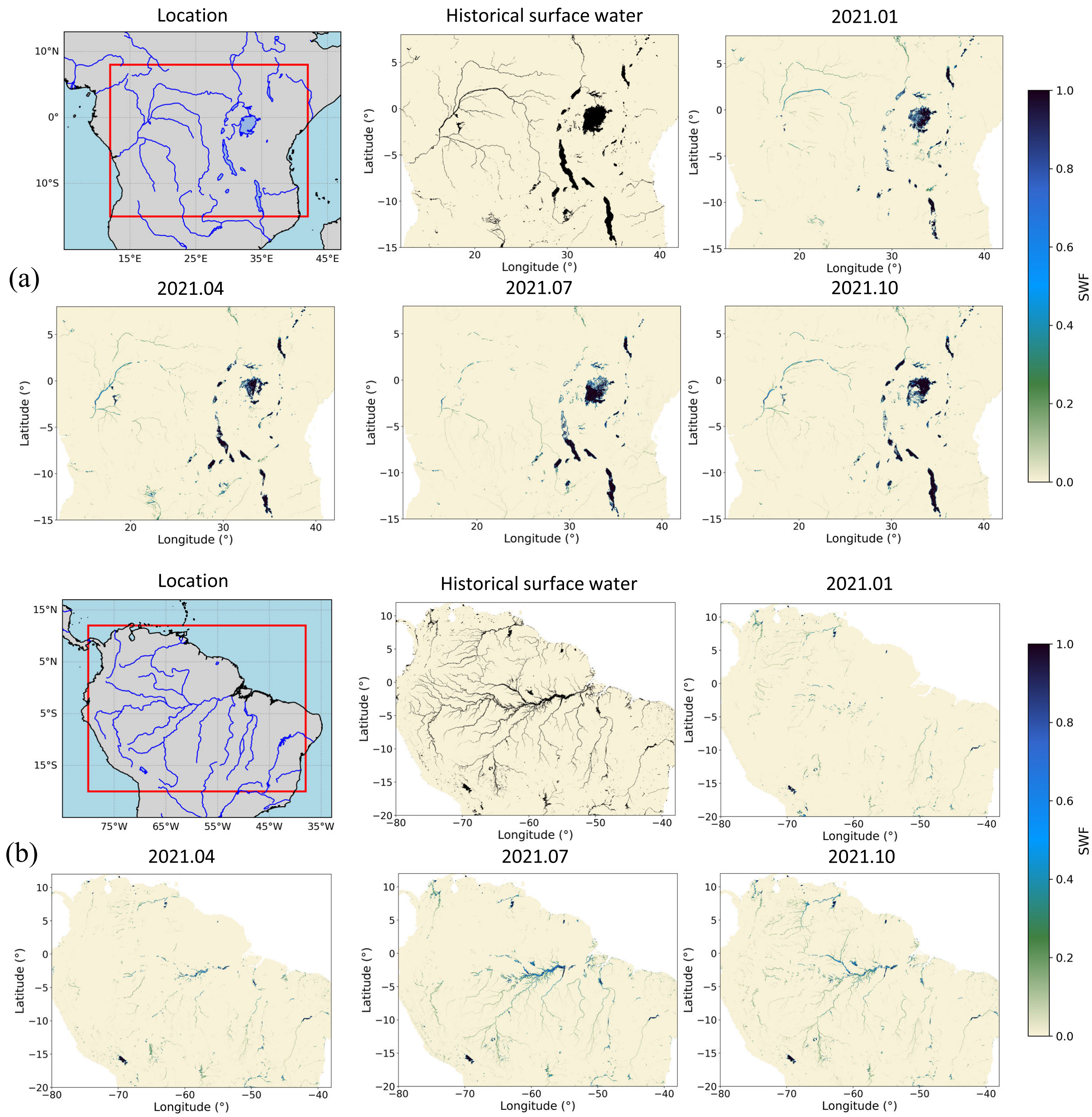

Fig. 6. Interannual variation of SWF in (a) Congo Basin and (b) Amazon Basin: seasonal reference data for January, April, July, and October, with cloud cover-induced data gaps filled with zero.

A sigmoid activation function is applied at the output for value range scaling.

6) *Pix2Pix:* It is a conditional generative adversarial network (cGAN) for image-to-image translation, generating target images from the inputs. U-Net serves as the generator, with a discriminator composed of four stacked convolutional layers.
7) *TransGAN:* It is a Transformer-based GAN that incorporates SSIM as a generator constraint, enhancing detail and accuracy in complex terrains. It is applied to SWF estimation while retaining the SSIM loss.
8) *DDPMBase:* DDPM is a denoising-based generative model that simulates reverse diffusion to produce high-quality samples. The original U-Net serves as the denoising model, with conditional information concatenated to the noisy image. A linear noise schedule is employed for noise prediction.

The results indicate that semantic segmentation methods struggle to accurately estimate SWF quantities, as reflected by their high RMSE values and lower SSIM and PSNR metrics. The CFRT model, designed specifically for SWF estimation, integrates the Swin Transformer and U-Net, employs an improved loss function, and achieves better results across multiple metrics. GAN-based generative models perform well in terms of SSIM and PSNR but exhibit significant limitations in accurately estimating SWF (see its RMSE). Diffusion-based models demonstrate strong performance in SSIM and PSNR, producing high-quality SWF estimation images. However, DDPMBase encounters challenges in accurately estimating SWF due to singularity issues, resulting in cumulative errors

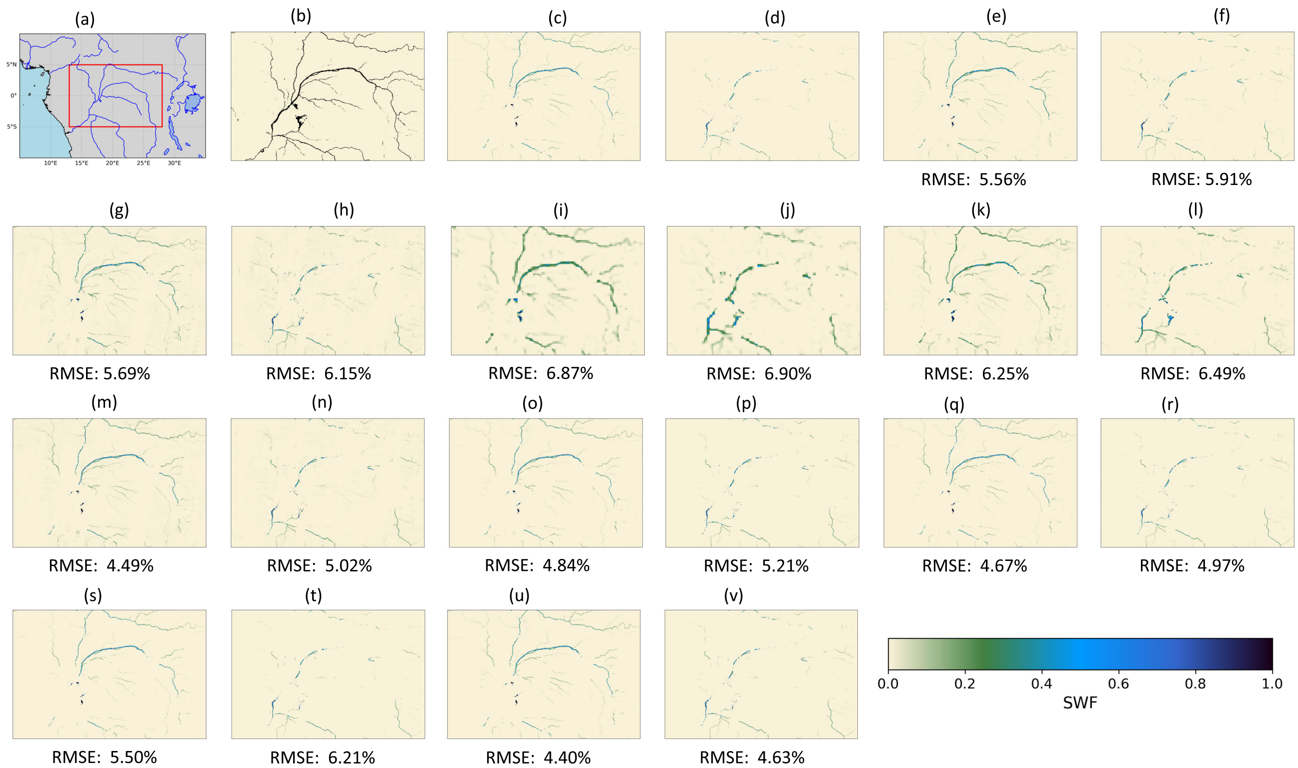

Fig. 7. Visualization of experimental results in the Congo Basin region. (a) Study area. (b) Historical surface water. (c) GWSF (January). (d) GWSF (July). (e) U-Net (January). (f) U-Net (July). (g) LinkNet (January). (h) LinkNet (July). (i) PspNet (January). (j) PspNet (July). (k) DeepLabV3+ (January). (l) DeepLabV3+ (July). (m) CRFT (January). (n) CRFT (July). (o) Pix2Pix (January). (p) Pix2Pix (July). (q) TransGAN (January). (r) TransGAN (July). (s) DDPM (January). (t) DDPM (July). (u) DiffWater (January). (v) DiffWater (July). Note: the numerical values below each method indicate the RMSE for the corresponding region.

during the sampling process, particularly near the $t \rightarrow T^{-}$ time step, and producing inaccurate estimates. In contrast, DiffWater resolves the singularity problem by applying equal-weight training at each time step, effectively reducing error accumulation during the sampling process and enabling accurate SWF estimation near the $t \rightarrow T^{-}$ time step. As a result, DiffWater achieves superior performance, with an RMSE of 4.987.

The SWF estimation results from various models in three representative regions (the Congo Basin, the Amazon Basin, and the middle and lower reaches of the Yangtze River) are presented in Figs. 7–9, illustrating conditions across different months. Additionally, the RMSE values of each method's results, along with the reference data for each region, are displayed below the corresponding figure. Visual analysis reveals that the seasonal variations of the water system are highly significant. Methods based on semantic segmentation exhibit notable errors in estimation. In certain areas, U-Net fails to detect surface water, resulting in biased SWF estimates. LinkNet tends to overestimate SWF and demonstrates excessive sensitivity to surface water in certain areas. In contrast, PspNet and Deeplabv3+ lack texture detail features in their surface water inversion results, leading to inaccuracies in estimating areas where surface water is present. CRFT considerably enhances performance in these aspects; however, omissions still occur in some certain areas. Pix2Pix demonstrates inaccuracies in SWF estimation. TransGAN exhibits excessive sensitivity to surface water, leading to the erroneous detection of surface water in areas where it is actually absent. DDPM demonstrates a lack of precision in SWF estimation, failing to detect the presence of surface water in certain areas. DiffWater prioritizes the accuracy of SWF estimation while also considering the presence of surface water. Although it may underestimate SWF, its overall results are the closest to the reference data.

G. Water Body Sensitivity Discussion

The sensitivity of surface water trajectory detection is crucial for accurately estimating SWF. Therefore, regions where surface water has previously occurred are assigned a value of 1, while regions without surface water are assigned a value of 0. The intersection over union (IoU) metric is employed to evaluate the overall surface water extraction performance of different models. Additionally, we accumulated the monthly occurrences of surface water in 2021 based on SWF estimation ratio intervals, generating a sensitivity statistics map of surface water, with the value range spanning from 0 to 12. Finally, by overlaying SWF estimations across different ratio intervals and calculating the normalized RMSE, we evaluated the sensitivity of various models to water bodies across multiple seasons and ratio ranges. This process facilitates the analysis of oversight phenomena in SWF estimation by different models under varying ratio conditions. Table VI presents a comparison of DiffWater with several representative models over the globe. DiffWater demonstrates significant advantages, achieving the best performance across multiple metrics. However, its overall

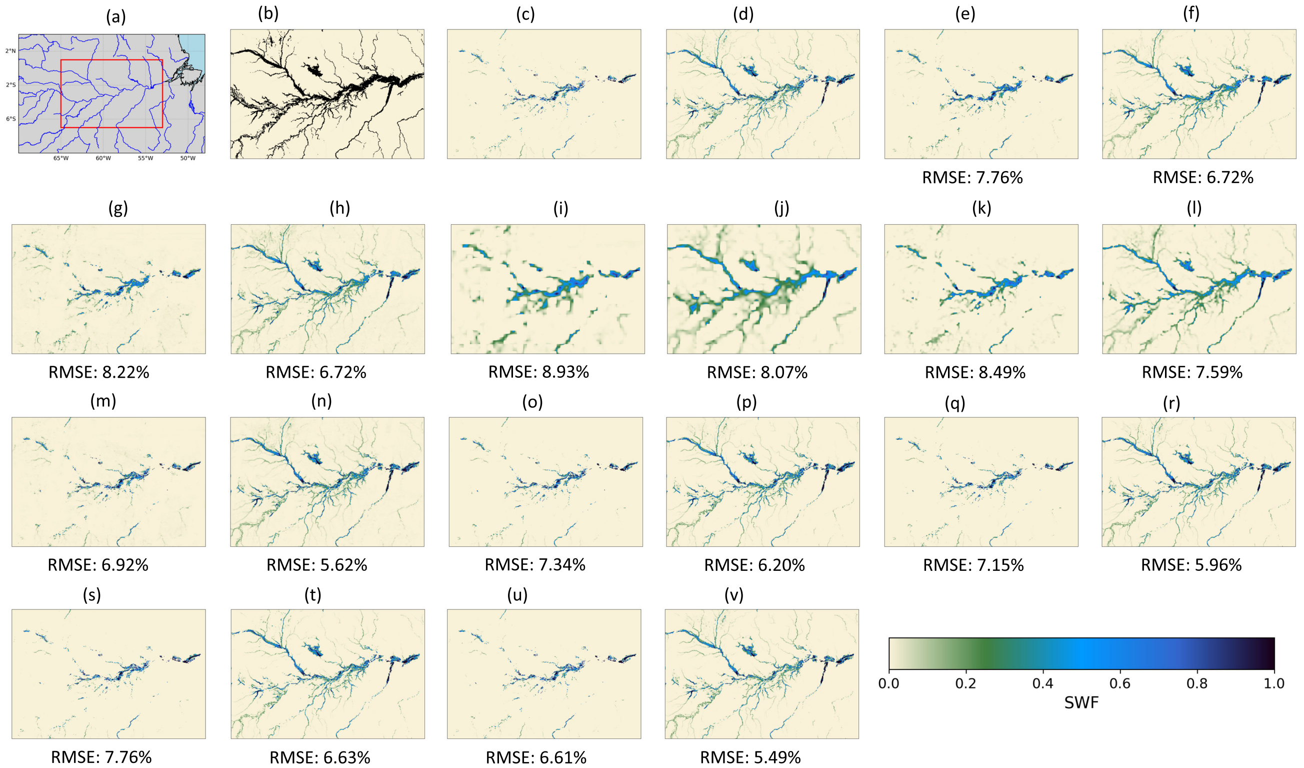

Fig. 8. Visualization of experimental results in the Amazon Basin region. (a) Study area. (b) Historical surface water. (c) GWSF (April). (d) GWSF (October). (e) U-Net (April). (f) U-Net (October). (g) LinkNet (April). (h) LinkNet (October). (i) PspNet (April). (j) PspNet (October). (k) DeepLabV3+ (April). (l) DeepLabV3+ (October). (m) CRFT (April). (n) CRFT (October). (o) Pix2Pix (April). (p) Pix2Pix (October). (q) TransGAN (April). (r) TransGAN (October). (s) DDPM (April). (t) DDPM (October). (u) DiffWater (April). (v) DiffWater (October). Note: the numerical values below each method represent the RMSE for the corresponding region.

TABLE VI

QUANTITATIVE COMPARISON OF SURFACE WATER SENSITIVITY USING DIFFERENT METHODS: FROM JANUARY TO DECEMBER 2021, THE RMSE WAS CALCULATED BASED ON THE MONTHLY CUMULATIVE NORMALIZATION OF SENSITIVITY, WHILE THE IOU WAS COMPUTED FOR AREAS WHERE SWF WAS PRESENT

Method	IOU ↑	RMSE (1%-10%) ↓	RMSE (10%-50%) ↓	RMSE (50%-100%) ↓	RMSE (1%-100%) ↓
Unet	0.547	7.61%	3.22%	1.46%	5.43%
LinkNet	0.486	11.25%	3.52%	1.54%	6.54%
PspNet	0.315	12.02%	5.70%	2.48%	9.35%
Deeplabv3+	0.395	11.09%	4.82%	2.09%	7.97%
CRFT	0.597	7.26%	2.93%	1.11%	5.11%
Pix2Pix	0.652	6.00%	2.49%	1.24%	4.10%
TransGAN	0.674	6.02%	2.38%	1.16%	4.04%
DDPMBase	0.616	5.20%	2.31%	1.10%	3.89%
DiffWater	0.667	4.92%	2.20%	0.65%	3.83%

sensitivity to surface water is slightly weaker than that of TransGAN. Specifically, DiffWater exhibits weaker sensitivity in detecting SWF in regions with a ratio of 10%–50% while achieving significant advantages in the 1%–10% and 50%–100% intervals. Fig. 10 illustrates the monthly cumulative surface water maps for 2021 generated by DiffWater, along with other models, using GSWF as reference data. The results indicate that U-Net and LinkNet fail to detect surface water in certain areas, while PspNet and Deeplabv3+ are unable to capture surface water details. In contrast, CRFT, Pix2Pix, and TransGAN demonstrate excessive sensitivity to surface water, leading to overdetection. Diffusion-based models, such as DDPM and DiffWater, exhibit strong detection capabilities for subtle surface water features, enabling accurate detection of surface water at varying ratios. However, due to differences in optimization objectives, DDPM, which focuses on noise estimation, struggles with SWF estimation accuracy and fails to provide precise judgments regarding SWF quantities. On the other hand, DiffWater not only achieves optimal accuracy in identifying water system locations but also provides accurate SWF estimations across multiple periods.

H. Time Series Analysis

To comprehensively evaluate the performance of multiple models, we conducted a detailed analysis of time series data for six case study regions. By averaging the SWF values within each region, we generated time series fitting graphs, as shown

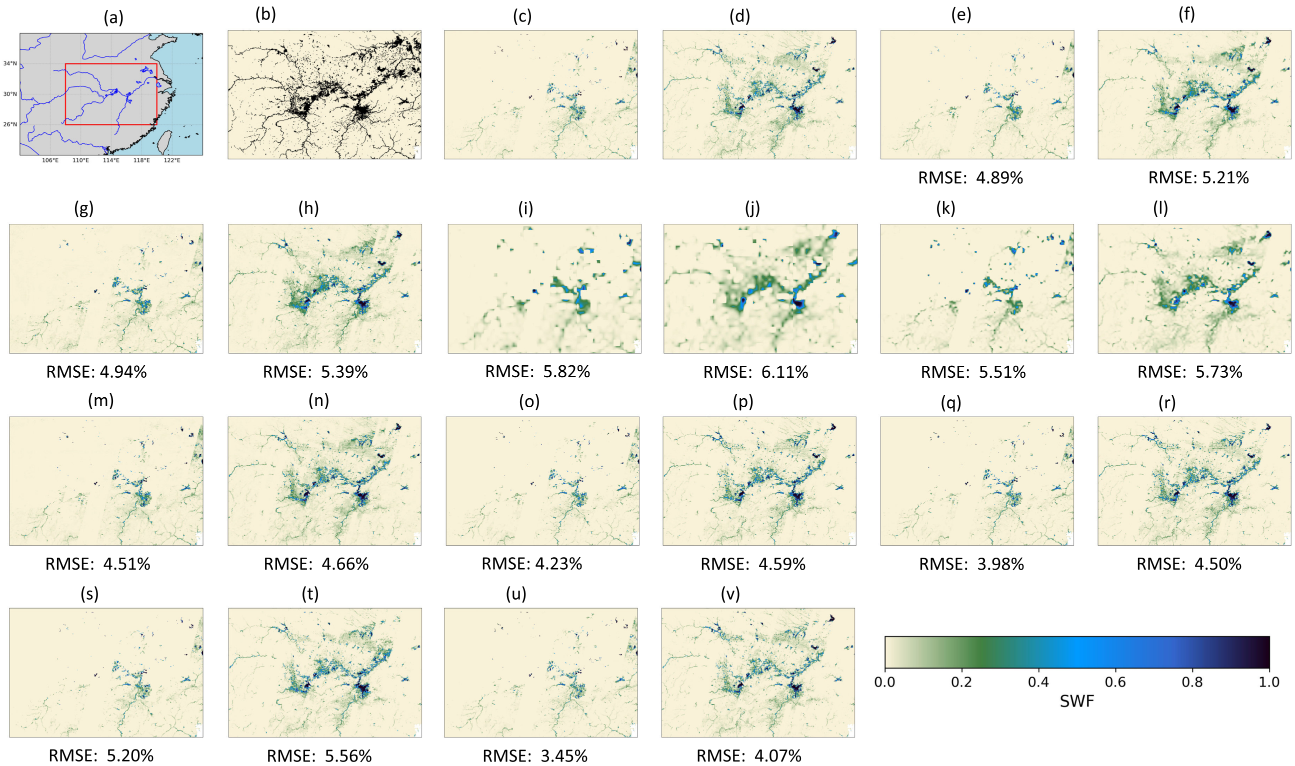

Fig. 9. Visualization of experimental results in the middle and lower reaches of the Yangtze River. (a) Study area. (b) Historical surface water. (c) GWSF (April). (d) GWSF (October). (e) U-Net (April). (f) U-Net (October). (g) LinkNet (April). (h) LinkNet (October). (i) PspNet (April). (j) PspNet (October). (k) DeepLabV3+ (April). (l) DeepLabV3+ (October). (m) CRFT (April). (n) CRFT (October). (o) Pix2Pix (April). (p) Pix2Pix (October). (q) TransGAN (April). (r) TransGAN (October). (s) DDPM (April). (t) DDPM (October). (u) DiffWater (April). (v) DiffWater (October). Note: the numerical values below each method represent the RMSE for the corresponding region.

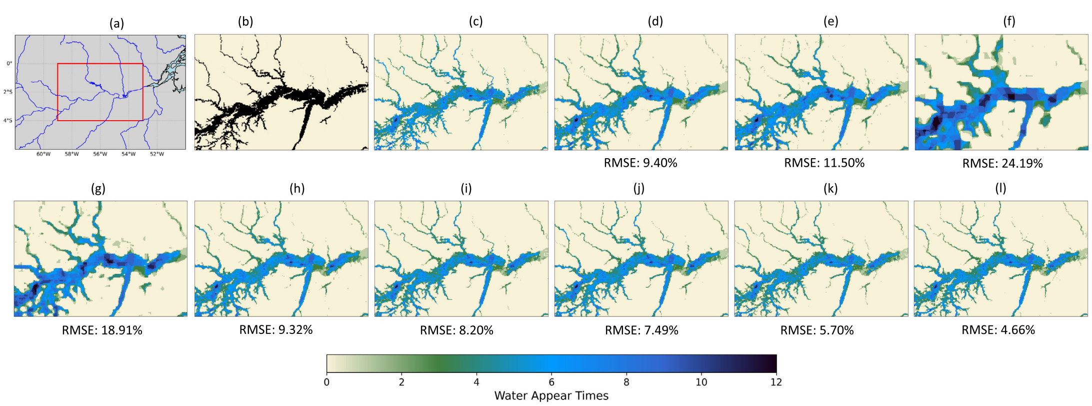

Fig. 10. 1%–100% surface water occurrence statistics for various methods. (a) Study area. (b) Historical surface water. (c) GWSF. (d) U-Net. (e) LinkNet. (f) PspNet. (g) DeepLabv3+. (h) CRFT. (i) Pix2Pix. (j) TransGAN. (k) DDPM. (l) DiffWater. Note: the numerical values below each method represent the RMSE for the corresponding region.

in Fig. 11. Representative models, including CRFT, DDPM, and DiffWater, were compared with the reference data GSWF. The results demonstrate that DiffWater effectively captures long-term time series features across all regions, exhibits high sensitivity to GNSS-R signal variations, and accurately estimates regional SWF content, closely aligning with optical reference data. However, both DDPM and DiffWater exhibit a consistent underestimation of regional SWF content, while CRFT, with its improved loss function, tends to overestimate SWF content. These discrepancies may be attributed to differences in optimization objectives. Future research could focus on refining the design of loss functions, as the reliance on a single mse loss function may pose a bottleneck for model performance.

I. Applications and Challenges of Regional Cross Validation

Our prior investigation employed temporal cross validation to assess interhemispheric generalization capacity and validate geographical transferability. Building on this foundation, we implemented a spatial cross-validation framework utilizing Northern Hemisphere data (August 2018–December 2021) for model training, followed by a systematic evaluation on

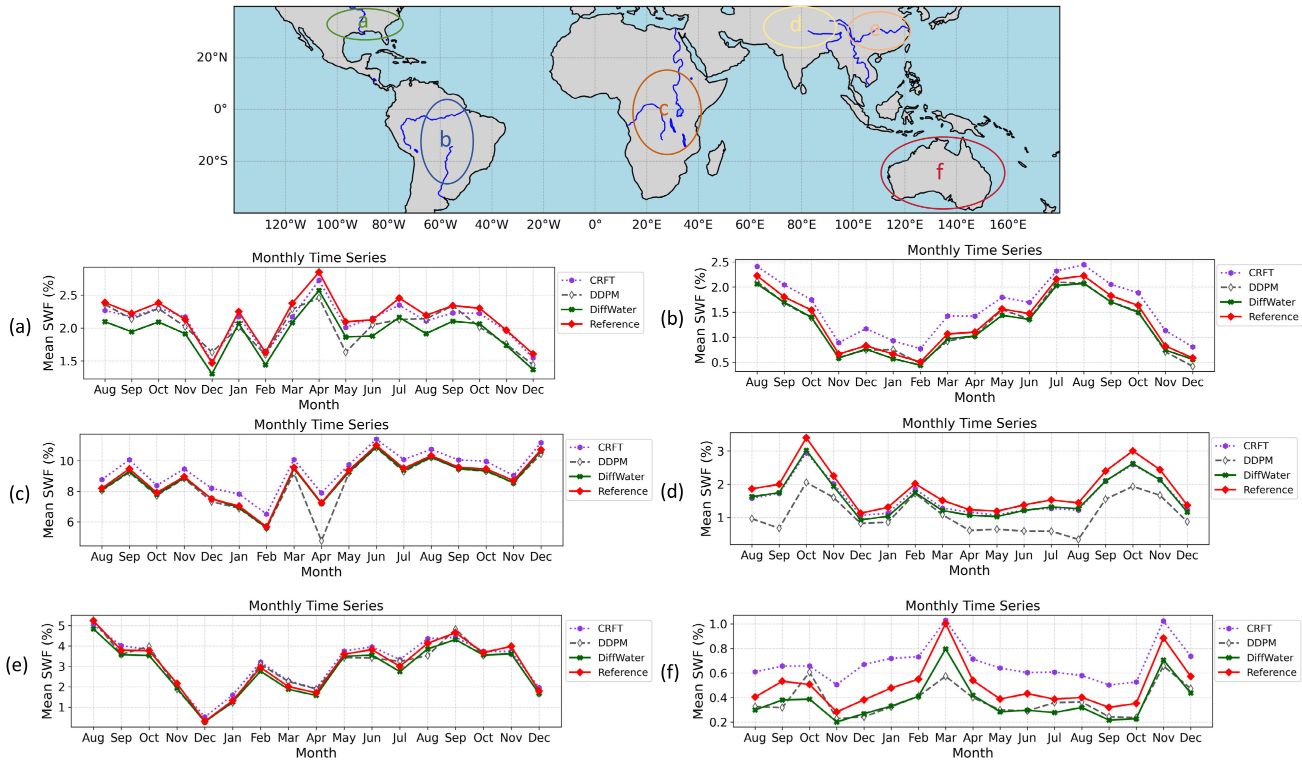

Fig. 11. Temporal analysis of SWF across six geographic regions (August 2020–December 2021): a comparative performance evaluation of three representative methods against reference data. Regions covered: (a) North America, (b) South America, (c) Africa, (d) Central Asia, (e) East Asia, and (f) Australia.

TABLE VII

QUANTITATIVE COMPARISON OF DIFFERENT SWF ESTIMATION METHODS: AVERAGE OVERALL METRIC FOR THE SOUTHERN HEMISPHERE. MODEL TRAINING WAS CONDUCTED USING DATA FROM THE NORTHERN HEMISPHERE FROM AUGUST 2018 TO DECEMBER 2021, WHILE TESTING WAS PERFORMED ON THE SOUTHERN HEMISPHERE REGION OVER THE SAME PERIOD

Method	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑	R ↑	Param(M) ↓	GFLOPS(G) ↓
Unet	4.529%	11.858%	0.969	46.860	0.682	24.433	7.798
LinkNet	4.853%	12.476%	0.967	46.572	0.651	21.769	5.423
PspNet	5.858%	14.565%	0.960	45.140	0.525	1.498	2.307
Deeplabv3+	5.279%	13.121%	0.966	46.504	0.614	26.674	9.175
CRFT	4.755%	12.671%	0.974	47.859	0.640	15.406	5.902
Pix2Pix	4.732%	12.698%	0.974	47.957	0.639	11.779	49.784
TransGAN	4.360%	11.176%	0.973	47.695	0.718	32.499	9.383
DDPMBase	4.930%	13.107%	0.972	47.495	0.612	32.197	58.367
DiffWater	4.176%	10.574%	0.974	47.700	0.747	18.550	23.205

Southern Hemisphere observations within the same temporal window. As listed in Table VII, the analysis reveals significant performance degradation (ΔRMSE = 15.8%–21.3%) across all architectures when applied to geographically distinct regions. Notably, encoder–decoder architectures (U-Net: RMSE $=$ 11.86% and $R = 0.682$; LinkNet: RMSE $=$ 12.48% and $R = 0.651$) exhibited superior cross-domain robustness. Generative approaches (Pix2Pix: $\Delta R = 0.032$; TransGAN: $\Delta R = 0.041$) demonstrated enhanced generalization capacity compared to discriminative architectures (CRFT: $\Delta R = 0.112$), consistent with their inherent ability to model data distributions under varying conditions. Despite employing robust U-Net foundations, DDPMBase exhibited convergence instability (RMSE $=$ 13.11%), potentially attributable to singularities in its diffusion process. While DiffWater achieved optimal performance (RMSE $=$ 10.57% and $R = 0.747$), it retained a 12.3% R-value reduction compared to the same-region temporal validation benchmarks.

We speculate that the substantial decline in performance across different models may stem from climate differences. The seasonal rainfall patterns in tropical regions of the Southern Hemisphere (such as the Amazon Basin) differ markedly from those in temperate regions of the Northern Hemisphere, leading to considerable variations in precipitation across different regions. Although the model can learn the correlation between CyGNSS data and surface water, cross-scenario tasks remain challenging. Regarding terrain complexity, the steep gradients in southern Africa and Australia present challenges for texture feature extraction. Complex terrain affects not only the spatial distribution of surface water but also increases the uncertainty of sensor data, further exacerbating the model's performance degradation. These findings indicate that further development of region-invariant representation learning techniques is essential for global-scale surface water monitoring applications.

V. Conclusion

In this study, we propose DiffWater, a conditional diffusion model designed for monthly scale GNSS-R signal-based SWF estimation. Existing regression methods based on semantic segmentation and GANs struggle to accurately estimate SWF in noisy environments. Diffusion models provide the potential for more precise SWF estimation under such conditions. However, traditional DDPMs encounter inaccuracies in SWF estimation due to singularity issues, and diffusion-based models generally require thousands of iterations to achieve high sampling quality. Consequently, balancing sampling speed and addressing singularity problems remains a significant challenge. DiffWater addresses these issues by improving the loss function objectives, significantly enhancing the estimation accuracy at singularity locations, and enabling precise SWF estimation. Additionally, it accelerates model training through the decoupling of conditional information and the introduction of lightweight modules. Compared to other deep learning methods, DiffWater achieves higher accuracy in SWF estimation tasks.

The accuracy of SWF estimations generated by DiffWater fundamentally depends on the quality and precision of the optical reference data used. Although this study integrates multiple sources of optical reference data and applies unified corrections to GNSS-R input data in regions lacking reference data, resulting in significantly improved estimation accuracy compared to previous studies, ensuring reliable and accurate reference data remains a critical challenge. Future work should investigate the incorporation of more suitable reference sources and utilize additional ground-truth data to further validate the model's performance. Simultaneously, when the reference data are appropriate, we will consider introducing a Transformer-based diffusion model for short-term SWF retrieval from large-scale GNSS-R data. In cloudy regions, incorporating downsampling methods in GNSS-R-based SWF estimation warrants further exploration. Furthermore, we observed that the performance of existing deep learning models varies significantly across different regions. The subsequent development of domain-invariant representation learning techniques is essential for global-scale surface water monitoring applications.

Acknowledgment

The authors are grateful to the reviewers for their constructive feedback and insightful recommendations on this article.

References

[1] M. Rast, J. Johannessen, and W. Mauser, "Review of understanding of earth's hydrological cycle: Observations, theory and modelling," *Surv. Geophys.*, vol. 35, pp. 491–513, Jan. 2014.

[2] X. Song, J. Zhang, C. Zhan, Y. Xuan, M. Ye, and C. Xu, "Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications," *J. Hydrol.*, vol. 523, pp. 739–757, Apr. 2015.

[3] B. R. Deemer et al., "Greenhouse gas emissions from reservoir water surfaces: A new global synthesis," *BioScience*, vol. 66, no. 11, pp. 949–964, Nov. 2016.

[4] J. A. Rosentreter et al., "Half of global methane emissions come from highly variable aquatic ecosystem sources," *Nature Geosci.*, vol. 14, no. 4, pp. 225–230, Apr. 2021.

[5] P. A. Herrera, M. A. Marazuela, and T. Hofmann, "Parameter estimation and uncertainty analysis in hydrological modeling," *WIREs Water*, vol. 9, no. 1, Jan. 2022, Art. no. e1569.

[6] Z. Xiao et al., "Unveiling the hidden dynamics of intermittent surface water: A remote sensing framework," *Remote Sens. Environ.*, vol. 311, Sep. 2024, Art. no. 114285.

[7] A. Khandelwal, A. Karpatne, M. E. Marlier, J. Kim, D. P. Lettenmaier, and V. Kumar, "An approach for global monitoring of surface water extent variations in reservoirs using MODIS data," *Remote Sens. Environ.*, vol. 202, pp. 113–128, Dec. 2017.

[8] F. Ling et al., "Monitoring surface water area variations of reservoirs using daily MODIS images by exploring sub-pixel information," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 141–152, Oct. 2020.

[9] J. Liang and D. Liu, "Automated estimation of daily surface water fraction from MODIS and Landsat images using Gaussian process regression," *Int. J. Remote Sens.*, vol. 42, no. 11, pp. 4261–4283, Jun. 2021.

[10] M. G. Tulbure and M. Broich, "Spatiotemporal dynamic of surface water bodies using Landsat time-series data from 1999 to 2011," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 44–52, May 2013.

[11] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, pp. 418–422, Dec. 2016.

[12] A. H. Pickens et al., "Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series," *Remote Sens. Environ.*, vol. 243, Jun. 2020, Art. no. 111792.

[13] W. Huang et al., "Automated extraction of surface water extent from Sentinel-1 data," *Remote Sens.*, vol. 10, no. 5, p. 797, May 2018.

[14] X. Yang et al., "Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data," *Remote Sens. Environ.*, vol. 244, Jul. 2020, Art. no. 111803.

[15] M. Wieland and S. Martinis, "Large-scale surface water change observed by Sentinel-2 during the 2018 drought in Germany," *Int. J. Remote Sens.*, vol. 41, no. 12, pp. 4742–4756, Jun. 2020.

[16] Z. Chen and S. Zhao, "Automatic monitoring of surface water dynamics using Sentinel-1 and Sentinel-2 data with Google Earth Engine," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, Sep. 2022, Art. no. 103010.

[17] T. R. Loveland and J. L. Dwyer, "Landsat: Building a strong future," *Remote Sens. Environ.*, vol. 122, pp. 22–29, Jul. 2012.

[18] Q. Yan, et al., "Mapping water fraction over the pan-tropical region using CYGNSS data,"*IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5800914, doi: 10.1109/TGRS.2024.3394744.

[19] Y. Li, G. Zhao, G. H. Allen, and H. Gao, "Diminishing storage returns of reservoir construction," *Nature Commun.*, vol. 14, no. 1, p. 3203, Jun. 2023.

[20] F. Yao et al., "Satellites reveal widespread decline in global lake water storage," *Science*, vol. 380, no. 6646, pp. 743–749, May 2023.

[21] S.-A. Boukabara, F. Weng, and Q. Liu, "Passive microwave remote sensing of extreme weather events using NOAA-18 AMSUA and MHS," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2228–2246, Jul. 2007.

[22] C. A. Varotsos and V. F. Krapivin, *Microwave Remote Sensing Tools in Environmental Science*. Cham, Switzerland: Springer, 2020.

[23] C. Derksen et al., "Retrieving landscape freeze/thaw state from soil moisture active passive (SMAP) radar and radiometer measurements," *Remote Sens. Environ.*, vol. 194, pp. 48–62, Jun. 2017.

[24] M. Davidson et al., "Sentinel-1 mission overview," in *Proc. 8th Eur. Conf. Synth. Aperture Radar*, Jun. 2010, pp. 1–4.

[25] R. Torres, M. Davidson, and D. Geudtner, "Copernicus sentinel mission at C- and L-band: Current status and future perspectives," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 4055–4058.

[26] K. Jensen and K. Mcdonald, "Surface water microwave product series version 3: A near-real time and 25-year historical global inundated area fraction time series from active and passive microwave remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1402–1406, Sep. 2019.

[27] Q. Yan and W. Huang, "Sea ice remote sensing using GNSS-R: A review," *Remote Sens.*, vol. 11, no. 21, p. 2565, Nov. 2019.

[28] S. Jin et al., "Remote sensing and its applications using GNSS reflected signals: Advances and prospects," *Satell. Navigat.*, vol. 5, no. 1, p. 19, Dec. 2024.

[29] N. Rodriguez-Alvarez, J. F. Munoz-Martin, and M. Morris, "Latest advances in the global navigation satellite system—Reflectometry (GNSS-R) field," *Remote Sens.*, vol. 15, no. 8, p. 2157, Apr. 2023.

[30] Z. Dong and S. Jin, "Evaluation of spaceborne GNSS-R retrieved ocean surface wind speed with multiple datasets," *Remote Sens.*, vol. 11, no. 23, p. 2747, Nov. 2019.

[31] C. Wang, K. Yu, F. Qu, J. Bu, S. Han, and K. Zhang, "Spaceborne GNSS-R wind speed retrieval using machine learning methods," *Remote Sens.*, vol. 14, no. 14, p. 3507, Jul. 2022.

[32] A. Rius et al., "Altimetry with GNSS-R interferometry: First proof of concept experiment," *GPS Solutions*, vol. 16, no. 2, pp. 231–241, Apr. 2012.

[33] Q. Yan, W. Huang, S. Jin, and Y. Jia, "Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111944.

[34] Y. Jia, S. Jin, P. Savi, Q. Yan, and W. Li, "Modeling and theoretical analysis of GNSS-R soil moisture retrieval based on the random forest and support vector machine learning approach," *Remote Sens.*, vol. 12, no. 22, p. 3679, Nov. 2020.

[35] Y. Hu, X. Yuan, W. Liu, Q. Hu, J. Wickert, and Z. Jiang, "An SVM-based snow detection algorithm for GNSS-R snow depth retrievals," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6046–6052, 2022.

[36] X. Chu et al., "Multimodal deep learning for heterogeneous GNSS-R data fusion and ocean wind speed retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5971–5981, 2020.

[37] X. Qiao and W. Huang, "WaveTransNet: A transformer-based network for global significant wave height retrieval from spaceborne GNSS-R data," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4208011, doi: 10.1109/TGRS.2024.3433397.

[38] J. Bu et al., "Machine learning methods for Earth observation and remote sensing using spaceborne GNSS reflectometry: Current status, challenges, and future prospects," *IEEE Geosci. Remote Sens. Mag.*, doi: 10.1109/MGRS.2025.3539983.

[39] C. S. Ruf et al., "A new paradigm in Earth environmental monitoring with the CYGNSS small satellite constellation," *Sci. Rep.*, vol. 8, no. 1, p. 8782, Jun. 2018.

[40] W. Yang et al., "Daily flood monitoring based on spaceborne GNSS-R data: A case study on henan, China," *Remote Sens.*, vol. 13, no. 22, p. 4561, Nov. 2021.

[41] J. Bu, Q. Wang, Z. Wang, S. Fan, X. Liu, and X. Zuo, "Land remote sensing applications using spaceborne GNSS reflectometry: A comprehensive overview," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 12811–12841, 2024.

[42] C. Gerlein-Safdi, A. A. Bloom, G. Plant, E. A. Kort, and C. S. Ruf, "Improving representation of tropical wetland methane emissions with CYGNSS inundation maps," *Global Biogeochem. Cycles*, vol. 35, no. 12, Dec. 2021, Art. no. e2020GB006890.

[43] Q. Yan et al., "Inland water mapping based on GA-LinkNet from CyGNSS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[44] J. Zhang, Y. J. Morton, Y. Wang and C. J. Roesler, "Mapping surface water extents using high-rate coherent spaceborne GNSS-R measurements," *IEEE Trans. Geosci. Remote Sens*, vol. 60, 2022, Art. no. 4211115, doi: 10.1109/TGRS.2022.3218254.

[45] B. Liu et al., "Statistical analysis of CyGNSS speckle and its applications to surface water mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5803915, doi: 10.1109/TGRS.2022.3196012.

[46] P. Zeiger, F. Frappart, J. Darrozes, C. Prigent, and C. Jiménez, "Analysis of CYGNSS coherent reflectivity over land for the characterization of pan-tropical inundation dynamics," *Remote Sens. Environ.*, vol. 282, Dec. 2022, Art. no. 113278.

[47] B. D. Chapman et al., "Comparison of SAR and CYGNSS surface water extent metrics," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3235–3245, 2022.

[48] Y. Chen and Q. Yan, "Unlocking the potential of CYGNSS for pan-tropical inland water mapping through multi-source data and transformer," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 133, Sep. 2024, Art. no. 104122.

[49] W. Li, E. Cardellach, S. Ribó, S. Oliveras, and A. Rius, "Exploration of multi-mission spaceborne GNSS-R raw IF data sets: Processing, data products and potential applications," *Remote Sens.*, vol. 14, no. 6, p. 1344, Mar. 2022.

[50] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[51] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8162–8171.

[52] S. Saxena et al., "The surprising effectiveness of diffusion models for optical flow and monocular depth estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 39443–39469.

[53] P. Ghasemigoudarzi, W. Huang, O. De Silva, Q. Yan, and D. Power, "A machine learning method for inland water detection using CYGNSS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[54] Q. Yan and W. Huang, "Sea ice sensing from GNSS-R data using convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1510–1514, Oct. 2018.

[55] Q. Yan and W. Huang, "Spaceborne GNSS-R sea ice detection using delay-Doppler maps: First results from the U.K. TechDemoSat-1 mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 10, pp. 4795–4801, Oct. 2016.

[56] P. Zhang, H. Yin, C. Li, and X. Xie, "Tackling the singularities at the endpoints of time intervals in diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 6945–6954.

[57] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," 2022, *arXiv:2202.00512*.

[58] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[59] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[60] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[61] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with Atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[62] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[63] C. Li, X. Liu and S. Li, "Transformer meets GAN: Cloud-free multispectral image reconstruction via multisensor data fusion in satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5408513, doi: 10.1109/TGRS.2023.3326545.