applied sciences

MDPI

Article

Systematic Evaluation of Machine Learning Models for Regression-Based Error Refinement in SAR-to-Optical Image Translation for Cloud Removal

Inseon Lee [1], Soyeon Park [1], Eui Ho Hwang [2] and No-Wook Park [1,*]

[1] Department of Geoinformatic Engineering, Inha University, Incheon 22212, Republic of Korea; inseonlee@inha.edu (I.L.); sypark531@inha.edu (S.P.)
[2] Water Resources Satellite Center, K-Water Research Institute, Daejeon 34045, Republic of Korea; ehhwang@kwater.or.kr
* Correspondence: nwpark@inha.ac.kr

Abstract

Generative deep learning-based synthetic aperture radar (SAR)-to-optical image translation (SOIT) has been widely employed for cloud removal. However, since cloud-contaminated regions reconstructed by SOIT inevitably contain prediction errors, an additional error refinement procedure is required to achieve reliable spectral reflectance reconstruction. In this study, three machine learning-based regression models, including Random Forest (RF), eXtreme Gradient Boosting (XGB), and Natural Gradient Boosting (NGB), are comprehensively evaluated for the error refinement of optical imagery initially reconstructed by SOIT. The factors influencing refinement performance are categorized into four components: (1) the sampling strategy of training pixels from cloud-free regions (random vs. quantile-based sampling); (2) the refinement target (actual spectral reflectance vs. residual between actual and initially reconstructed reflectance); (3) SAR features (pixel-level raw SAR features vs. local spatial SAR features); and (4) the cloud fraction in the scene of interest. A systematic sensitivity analysis of their effects on error refinement performance was conducted over cropland using PlanetScope optical imagery and COSMO-SkyMed SAR imagery. The results showed that cloud fraction had the greatest impact on refinement performance. Regarding SAR features for regression, the use of local spatial SAR features improved spectral similarity by up to approximately 4.6%p compared to raw SAR features. In terms of sampling strategy, quantile-based sampling yielded better refinement performance, whereas the effect of the refinement target was less pronounced. These results suggest that local spatial SAR features and quantile-based sampling strategies are the key determinants of regression-based refinement performance in SOIT-based cloud removal.

Keywords: cloud removal; SAR-to-optical image translation; generative deep learning; error refinement; machine learning; sensitivity analysis

Check for updates

Academic Editor: Andrea Prati

Received: 8 May 2026
Revised: 22 May 2026
Accepted: 22 May 2026
Published: 25 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Optical remote sensing imagery serves as a critical information source for various environmental monitoring applications [1–3]. For example, in crop monitoring, since the spectral characteristics of crops vary sensitively depending on their phenological stages, multi-temporal image analysis is generally preferred over single-date observations [4,5]. However, optical imagery is frequently contaminated by clouds, resulting in incomplete

time series and degraded accuracy in monitoring analyses [1,6]. To overcome this limitation, a variety of cloud removal techniques have been proposed to reconstruct cloud-contaminated regions, enhancing the utility of optical image time series [7–10].

In cloud removal, cloud-free images acquired over the same area on different dates can be used as reference data [2,7,11]. This approach is widely applied because spectral reflectance information across multiple bands can be directly utilized. However, accurate reconstruction requires cloud-free reference images acquired near the prediction date [7,11]. For example, larger temporal gaps between the images may increase prediction errors due to surface condition mismatches in croplands where surface reflectance changes rapidly [12,13].

Synthetic aperture radar (SAR) imagery can provide surface information regardless of weather conditions, unlike optical imagery. However, its interpretation is constrained by sensor-specific characteristics, including speckle noise and geometric distortions [14,15]. To exploit the complementary properties of optical and SAR imagery in terms of data availability, SAR-to-optical image translation (SOIT) has been proposed to generate optical imagery synthesized from SAR imagery [12,16,17]. SOIT learns the relationship between the two sensor domains using co-registered SAR and cloud-free optical image pairs from reference dates. It subsequently generates virtual optical imagery using SAR imagery acquired on the prediction date as input [18,19].

The key challenge in SOIT lies in effectively modeling the complex relationship between SAR and optical images. SAR imagery measures backscattering from the physical properties of surface targets, whereas optical imagery measures the spectral reflectance of surface targets. These fundamentally different imaging mechanisms give rise to a highly nonlinear relationship between the two image domains [18,20,21]. Deep learning models have been developed to model this relationship, with generative ones proving particularly effective due to their strong capability in image synthesis. Conditional generative adversarial networks (cGANs) have been widely applied to SOIT [16,21,22], and, more recently, diffusion models have been explored to further improve generation quality [23,24]. These generative models can produce synthetic optical imagery closely resembling actual optical observations by conditioning on input SAR imagery. Nevertheless, generative deep learning-based SOIT results remain subject to inherent limitations. When inter-sensor relationships are difficult to fully learn, spectral distortions or structural artifacts may occur in the reconstructed optical imagery [25].

To address the prediction errors inherent in SOIT outputs, two approaches can be considered. The first involves the use of advanced model architectures to enhance prediction performance [20,24,26]. While this approach represents the only viable option when the entire scene is covered by clouds, it has inherent limitations in that translation accuracy is fundamentally constrained by the limited information content of SAR imagery. In addition, when clouds only partially cover the scene, the approach does not exploit the actual reflectance information available in cloud-free regions. The second approach involves the regression-based error refinement of initially reconstructed optical imagery using information from cloud-free regions as a post-processing step, which is of considerable practical value as it can directly leverage actual reflectance observations in partially cloudy conditions. Regression-based approaches have indeed proven effective in cloud removal and error refinement tasks. Li et al. [27] modeled the error characteristics of initially reconstructed optical imagery under a linear assumption, and Kwak et al. [25] applied Random Forest (RF) to learn the relationship between cGAN-based initial predictions and actual reflectance in cloud-free regions for the error refinement of cloud-contaminated regions with the partial consideration of nonlinearity. Tahsin et al. [28] reconstructed cloud-contaminated vegetation indices using RF, and Wang et al. [29] further demonstrated the

validity of nonlinear regression-based approaches by applying a spatial–spectral RF for gap-filling in thick cloud regions.

However, the characteristics of prediction errors in SOIT outputs are affected by not only the relationship between initial predictions and actual reflectance values but also various experimental factors. For example, the sampling strategy of training pixels from cloud-free regions is critical because machine learning regression models are highly dependent on the quantity and distribution of training samples [30,31], and the extent to which the reflectance distribution of cloud-free regions represents that of cloud-contaminated ones critically determines generalization performance. The choice of refinement target is also important; directly predicting actual spectral reflectance and learning the residual between initial predictions and actual reflectance represent two fundamentally different strategies, and residual-based approaches may promote training stability by allowing the model to focus on error patterns rather than the full reflectance scale [9,27]. Furthermore, the SAR features used in the SOIT process can also serve as auxiliary input in the refinement stage, as they reflect the physical properties of surface targets and may provide useful information for modeling error patterns in initially reconstructed optical imagery [32,33]. In particular, local spatial SAR features derived from neighboring pixels may more effectively capture parcel-level structural patterns associated with prediction errors than pixel-level raw features. Finally, cloud fraction directly determines the quantity and distributional representativeness of training samples obtainable from cloud-free regions, and its effect is expected to interact with the aforementioned factors in a compound manner [34,35]. To the best of our knowledge, the effects of such factors on refinement performance have not yet been systematically evaluated. Without systematic evaluation across diverse experimental conditions and multiple regression models, it is difficult to determine which factors should be prioritized in practical applications or establish evidence-based guidelines for practical applications.

The objective of this work is to comprehensively evaluate machine learning models for the regression-based error refinement of optical imagery initially reconstructed by SOIT and systematically analyze the factors influencing refinement performance. To this end, the following three machine learning regression models are compared to evaluate their applicability for regression-based error refinement of SOIT outputs: RF, eXtreme Gradient Boosting (XGB), and Natural Gradient Boosting (NGB). Moreover, the effects of the following four key factors in error refinement performance are quantitatively analyzed through a systematic sensitivity analysis: the sampling strategy of training pixels from cloud-free regions, refinement target, SAR features, and cloud fraction. Based on the results, practical guidelines for the application of regression-based error refinement to SOIT-based cloud removal are derived.

2. Study Area and Data

The evaluation experiments were conducted over a portion of cropland located in Gimje, Republic of Korea, which is one of the major rice paddy cultivation regions in Korea requiring periodic large-scale cropland monitoring. The study area consists predominantly of agricultural fields, with a small number of roads and built-up areas also present, covering a total area of approximately 944 ha, corresponding to 1024 × 1024 pixels at a spatial resolution of 3 m. This area serves as the test area for generative deep learning-based SOIT and the target for error refinement. Figure 1 shows optical and SAR images of the study area along with the synthetic cloud masks used in the experiments. The cloud masks were synthesized at four cloud fraction levels (10%, 25%, 50%, and 75% of the total study area) to enable a quantitative evaluation of error refinement performance across a range of cloud conditions encountered in real-world scenarios. The synthetic cloud masks

were generated using a Gaussian blob-based random generation approach to replicate the irregular spatial patterns of real clouds. Specifically, multiple anisotropic Gaussian blobs were randomly placed and superimposed across the scene, and a quantile-based threshold was automatically determined to produce binary masks corresponding to the target cloud fractions. Post-processing steps, including boundary smoothing and removal of small cloud components, were subsequently applied. Both clouds and cloud shadows were treated as a single mask without distinction, as both represent obscured regions where surface reflectance information is unavailable, as illustrated in Figure 1c–f.

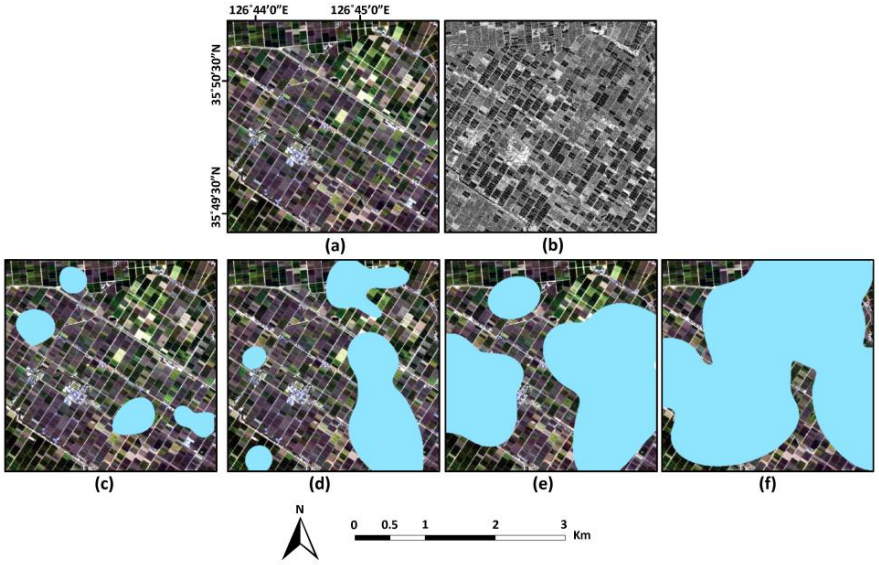

Figure 1. Optical and SAR imagery of the study area acquired on 24 May 2019: (**a**) PlanetScope (PS) true color composite imagery; (**b**) COSMO-SkyMed (CSK) SAR imagery in HH polarization; (**c**–**f**) synthetic cloud masks corresponding to cloud fractions of 10%, 25%, 50%, and 75%, respectively, overlaid on the PS imagery.

Table 1 summarizes the specifications and acquisition dates of the optical and SAR images used in this study. As for the optical imagery, the Level 3B product of PlanetScope (PS) imagery acquired by the Dove-R satellite in PS2.SD mode was used as the target optical imagery, providing four spectral bands: blue, green, red, and near-infrared (NIR) [36]. The Level 3B product provides atmospherically and geometrically corrected surface reflectance. All PS images were acquired under cloud-free conditions to enable a quantitative evaluation of the effect of error refinement in cloud-contaminated regions.

Table 1. List of PlanetScope and COSMO-SkyMed image pairs used in this study.

	PlanetScope	**COSMO-SkyMed**
Instrument/mode	Dove-R (PS2.SD)	Stripmap HIMAGE
Product	Level 3B surface reflectance	Level 1A single-look complex slant
Spectral bands/polarization (central wavelength/frequency)	Blue (490 nm), Green (565 nm), Red (665 nm), Near-infrared (865 nm)	X-band HH polarization (9.6 GHz)
Spatial resolution	3 m	3 m
Acquisition dates	24 May 2019 6 July 2019 30 September 2019	24 May 2019 10 July 2019 29 September 2019

COSMO-SkyMed (CSK) imagery was used as SAR imagery for both generating initially reconstructed optical imagery and serving as auxiliary input for refinement. CSK SAR imagery was acquired in Stripmap HIMAGE mode at the X-band (9.6 GHz) with HH single polarization, and the Level 1A single-look complex slant (SCS) format product was used [37]. The SCS format contains raw SAR data in complex form, and preprocessing was performed using the Sentinel Application Platform (SNAP) developed by the European Space Agency (ESA) in the following sequence [25,38,39]: (1) multi-looking, (2) terrain correction, (3) speckle filtering with a Refined Lee filter, and (4) conversion to the dB scale. The resulting ground range detected (GRD) image was resampled to a spatial resolution of 3 m to match the PS imagery, and both datasets were projected to the UTM zone 52 N coordinate system for analysis.

PS and CSK images were acquired at three phenological stages of rice growth, at which the variations in surface reflectance and backscattering coefficients were clearly distinguishable. Figure 2 illustrates the PS and CSK images across the three acquisition dates. In May, the transplanting of rice had just begun, with some parcels already showing established crops while others had not yet been planted. Early July corresponds to the mid-growth stage of rice, during which most parcels exhibit low visible-band reflectance. In the CSK imagery, many parcels appear with high backscattering coefficients owing to the structural characteristics of rice stems. September marks the rice maturation and harvest period; by the target date of late September, some parcels had already been harvested and were observed as bare soil. As shown in Table 1, PS and CSK image pairs were configured such that the acquisition dates were either identical or within a maximum gap of four days.

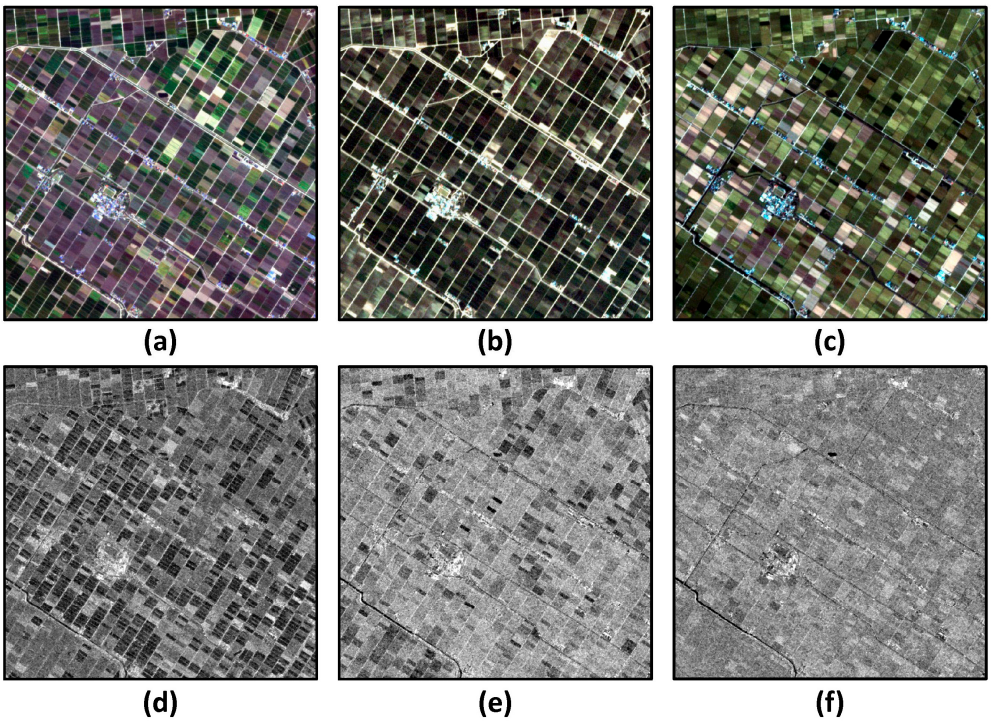

Figure 2. PS and CSK image pairs on three target dates. Columns from left to right correspond to 24 May, 6 July, and 30 September: (**a**–**c**) PS true color composite images and (**d**–**f**) CSK images in HH polarization.

3. Methods

Figure 3 presents the overall flowchart of the experiment designed in this study to compare refinement performance. The experiment began with initially reconstructed optical imagery generated through SOIT, which subsequently served as input to the refinement

procedure. Refinement was performed by training a regression model initially using reconstructed reflectance and SAR features from cloud-free regions as input variables and actual spectral reflectance as the target variable. The trained model was then applied to cloud-contaminated regions to perform error refinement. The objective of the experiments was not only to compare performance across machine learning models but also to evaluate refinement performance, with particular attention to four factors that may influence error refinement quality. These four factors, highlighted in yellow in Figure 3, were as follows: (1) sampling strategy (random vs. quantile-based sampling of training pixels from cloud-free regions), (2) refinement target (reflectance vs. residual), (3) SAR features (pixel-level raw vs. local spatial SAR features), and (4) cloud fraction (10%, 25%, 50%, and 75%).

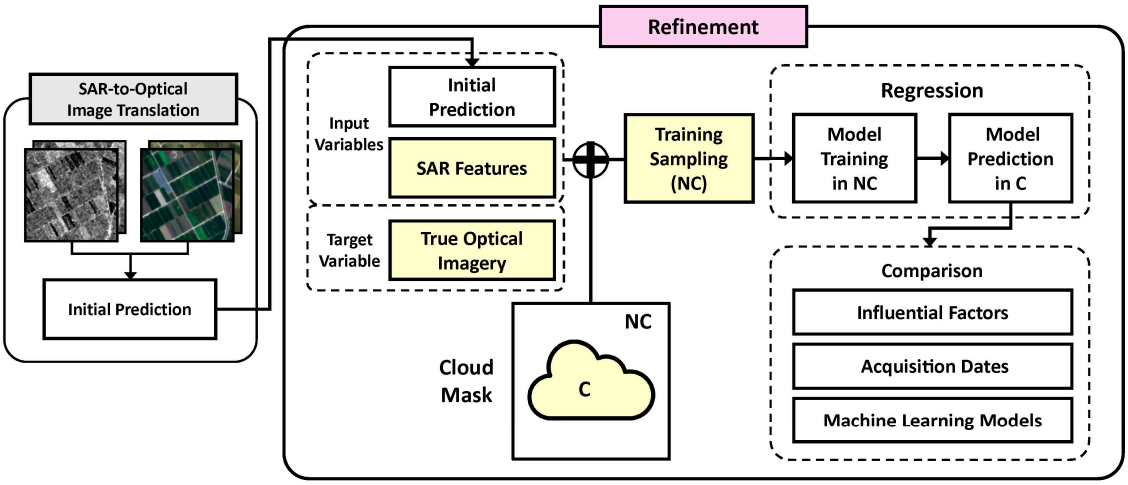

Figure 3. Workflow of regression-based error refinement experiment. NC denotes cloud-free regions and C denotes cloud-contaminated regions.

3.1. Initial Prediction via SOIT

The initially reconstructed optical imagery used as input for error refinement was generated using Pix2Pix, one of the cGAN models [40]. Pix2Pix is a model in which a generator and a discriminator are trained adversarially, generating optical imagery conditioned on SAR imagery. The generator and discriminator adopted U-Net and PatchGAN architectures, respectively, and the overall network architecture and hyperparameters followed the originally defined settings.

The model output had a value range of [0, 1] across four spectral bands, which is consistent with PS reflectance, while the input consisted of four bands comprising the CSK HH polarization band and spatial features extracted via the discrete wavelet transform (DWT) [41]. Through DWT decomposition, the approximation sub-band (LL) and the horizontal (LH) and vertical (HL) detail sub-bands were obtained. The diagonal sub-band (HH) was excluded as it contained high-frequency noise information. Following training, patch-level inference results were mosaicked to cover the full extent of the study area, thus yielding the initially reconstructed optical imagery of the full study area. Further details on the data and processing steps are available in the study by Park et al. [42].

3.2. Error Refinement Using Machine Learning Regression Models

Since the study area consisted predominantly of cropland, it was reasonable to assume that error patterns learned from cloud-free regions could be consistently applied to cloud-contaminated regions. However, prediction errors in the initially reconstructed optical imagery exhibited a systematic pattern, as shown in Figure 4, in which prediction errors decreased as actual reflectance values increased, and this pattern was consistently observed

across all three acquisition dates and spectral bands. This brightness-dependent error pattern was also considered to be associated with spatially structured patterns at the parcel level, thus suggesting that the reflectance distribution characteristics of training samples drawn from cloud-free regions may critically influence refinement performance.

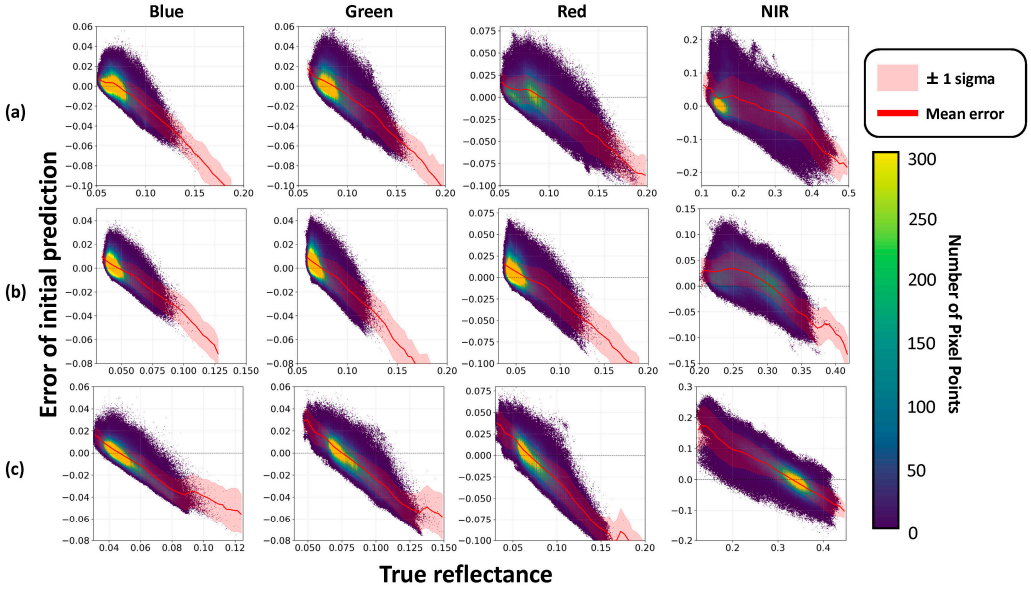

Figure 4. Scatter density plots between prediction errors in initially reconstructed imagery and actual PS reflectance values. (**a**), (**b**), and (**c**) represent images from 24 May, 6 July, and 30 September, respectively. The red line connects the mean errors for narrow reflectance intervals and the light red shading represents the range of the mean ± one standard deviation.

The study area was partitioned into cloud-free and -contaminated regions using synthetic cloud masks. A subset of pixels from the cloud-free region was used as training data, in which both actual PS reflectance and initially reconstructed reflectance were available, with the sampling strategy described in detail in the following section. The regression model was formulated as follows:

$$\hat{Y}_b = f_b(\mathrm{X}), \quad b = 1, 2, \cdots, B, \tag{1}$$

where $\mathrm{X} = [R_{init,1}, R_{init,2}, \cdots, R_{init,B}, F_{SAR}]$ denotes the input variables consisting of initially reconstructed reflectance across all spectral bands ($R_{init,b}$) and auxiliary SAR features (F_{SAR}). Y_b denotes the target variable, $\hat{Y}_b$ the corresponding model prediction, and f_b the machine learning regression model trained for band b. The trained model was then applied to pixels in the cloud-contaminated region to perform error refinement, and the final refined result was obtained based on the definition of the target variable.

In this study, tree-based machine learning regression models were selected for error refinement, comprising RF, XGB, and NGB. Such models are effective for modeling nonlinear relationships and enable pixel-level learning, thus making them applicable even when the number of available training samples is limited due to a high cloud fraction. In contrast, patch-based deep learning models are less appropriate in this context for two reasons. First, the availability of training patches is inherently limited as only cloud-free pixels can be used for training, and this limitation becomes more severe as cloud fraction increases. Second, patches constructed around cloud boundaries inevitably contain a mixture of cloud-free and -contaminated pixels, making it difficult to apply such models consistently during both training and inference.

RF is a representative bagging-based ensemble model in which the final prediction is determined by averaging the outputs of multiple independently trained decision trees [43]. By incorporating randomness in both training data sampling and feature selection for node splitting, RF achieves a complementary ensemble effect and has therefore been widely adopted for modeling nonlinear relationships [30,44]. XGB is an efficient ensemble model based on the gradient boosting framework [45]. Unlike RF, XGB iteratively learns the prediction errors of previous trees in subsequent trees, thus enabling more accurate regression predictions. However, this sequential learning structure may make XGB more susceptible to overfitting compared to RF. To mitigate this, the subsampling of both training data and features along with regularization are commonly recommended [45]. NGB is a probabilistic gradient boosting algorithm that treats the full conditional probability distribution of the output as the learning objective [46], thus enabling the simultaneous estimation of distribution parameters such as the mean and standard deviation. Unlike RF and XGB using conventional gradients, NGB employs natural gradients, offering the advantage of stable convergence even when the parameterization of the probability distribution changes. LogScore, based on negative log-likelihood, is used as the scoring rule [46]. This distinguishes NGB from simple point estimation approaches in that the goal is to align the predicted distribution with the true data distribution. Since the mean of the predicted distribution is used as the final inference output, comparisons with RF and XGB are made under equivalent conditions. However, the internal distribution learning process enables a more effective modeling of brightness-dependent error patterns.

The optimal hyperparameters for each model were determined via grid search [47] within predefined search ranges based on preliminary experiments. The hyperparameter search ranges and fixed settings for each model are summarized in Table 2. For RF, the number of trees was searched in the range of [800, 1200] and the maximum tree depth in [20, 40], with the remaining hyperparameters set to the default values of the scikit-learn package [48]. For XGB, the number of trees was searched in the range of [700, 1300], the learning rate in [0.03, 0.05], and the maximum tree depth in [5, 11], with subsampling ratios for training data and features set to 0.8 and 0.8, respectively, and the regularization parameter set to 1.0. For NGB, the number of trees was searched in the range of [700, 1000] and mini-batch fraction in [0.7, 0.9], with decision trees of maximum depth in the range of [3, 5] used as base learners.

Table 2. Major hyperparameter search ranges and fixed settings for RF, XGB, and NGB.

Hyperparameter	RF	XGB	NGB
Number of trees (step = 100)	[800, 1200]	[700, 1300]	[700, 1000]
Maximum tree depth	[20, 30, 40]	[5, 7, 9, 11]	[3, 4, 5]
Learning rate	-	0.03, 0.05	0.01
Subsampling ratio (data)	-	0.8 (fixed)	0.7, 0.8, 0.9
Subsampling ratio (features)	-	0.8 (fixed)	0.8 (fixed)
Regularization	-	1.0 (fixed)	-

3.3. Influential Factors

In this study, the sampling strategy, refinement target, and SAR features were considered key factors affecting error refinement performance. Each factor was divided into two options and compared, in which option (1) represented a more general or simpler baseline approach and option (2) was designed to improve the representativeness of training data or the model's ability to learn error patterns. In addition, cloud fraction was considered an external condition that exerted a compound effect on the influence of the three aforementioned factors.

3.3.1. Sampling Strategy

Although all pixels in the NC region were available as potential training data, using the entire NC region was computationally expensive. Therefore, a subset of pixels was sampled from the NC region as training data. However, the reflectance distribution of cloud-free regions was governed by the surface conditions of those areas and did not always align proportionally with the distribution of cloud-contaminated regions, which could critically affect error refinement performance. In particular, certain reflectance intervals could be insufficiently represented in the training data in environments with brightness-dependent error patterns, potentially degrading error refinement performance in those intervals. To address this issue, two training sample extraction strategies were compared: (1) random and (2) quantile-based sampling.

In the random sampling approach, samples were drawn randomly from all pixels in the NC region, resulting in training data that directly reflected the actual reflectance distribution of the NC region. In the quantile-based sampling approach, the actual PS reflectance values of each band were divided into five intervals using quantile boundaries at 20%, 40%, 60%, and 80%, and an equal number of pixels were drawn uniformly from each interval.

To ensure a fair comparison between the two strategies, the total number of training samples was fixed at 50,000 based on preliminary experiments, considering both computational efficiency and sample representativeness, corresponding to approximately 4.8% of all pixels in the study area. Previous studies have reported diminishing returns in model performance beyond a certain sample size in machine learning-based spatial prediction [31]. The results of a preliminary sensitivity analysis examining the effect of training sample size on refinement performance are provided in Appendix A (Figure A1).

3.3.2. Refinement Target

The definition of the refinement target, i.e., the target variable Y_b in Equation (1), also constitutes an important factor affecting error refinement performance. Two approaches were compared: (1) reflectance and (2) residual. In the reflectance approach, Y_b is defined as the actual PS reflectance of cloud-free pixels, and the model output $\hat{Y}_b$ directly serves as the refined reflectance for cloud-contaminated regions. However, the initially reconstructed reflectance already approximates the actual reflectance to some extent. Therefore, learning only the residual errors rather than the full reflectance scale may reduce task complexity and improve training stability. In the residual approach, Y_b was therefore defined as the residual between actual and initially reconstructed reflectance, i.e., $Y_b = R_{actual,b} - R_{init,b}$, and the final refined result was obtained by adding the model-predicted residual to the initially reconstructed reflectance:

$$\hat{R}_{refined,b} = \hat{Y}_b + R_{init,b} \quad (2)$$

3.3.3. SAR Features

As aforementioned, the SAR features used in the SOIT process were also incorporated as auxiliary input in the refinement stage. In this regard, whether to use pixel-level SAR features or local statistics that incorporated spatial context from neighboring pixels was an important consideration in the refinement stage. If prediction errors in the initially reconstructed optical imagery exhibited spatially structured patterns at the parcel level rather than random pixel-level variations, pixel-level SAR features did not sufficiently capture such spatial patterns. In particular, in areas with high within-parcel homogeneity, such as croplands, parcel-level textural and structural information could be more closely

associated with reflectance errors than with pixel-level SAR features. From this perspective, two options for SAR features were compared: (1) raw and (2) local spatial SAR features.

In the raw SAR feature approach, the same SAR configuration used for SOIT model training was adopted directly, comprising a four-band CSK image that included the original HH polarization band and the DWT-based LL, LH, and HL components. In the local spatial SAR feature approach, local statistics extracted from the HH polarization band and DWT components using a 7×7 window specified from preliminary experiments were used as input. Specifically, the local mean and standard deviation were derived from the HH polarization and LL component, while log-transformed local energy was computed for the LH and HL components, yielding a total of six SAR features. Accordingly, the raw SAR feature approach used eight input features, combining the initially reconstructed reflectance (four bands) and raw SAR features (four bands). In contrast, the local spatial SAR feature approach used ten input features, combining the initially reconstructed reflectance (four bands) and local spatial SAR features (six bands).

3.3.4. Cloud Fraction

Defined as the proportion of cloud coverage within the scene on the acquisition date, cloud fraction was an external condition that fundamentally affected the performance of regression-based error refinement. As cloud fraction increased, the spatial extent of cloud-free regions diminished, reducing the quantity and representativeness of available training samples. This, in turn, compounded the influence of the three factors described above. Therefore, systematic evaluation across a range of cloud fraction conditions was necessary to fully characterize the practical applicability of the proposed approach. To this end, four cloud fraction conditions were established, corresponding to approximately 10%, 25%, 50%, and 75% of the total study area. The 10% condition represented a state in which cloud-free regions were sufficiently available, thus making it the most favorable condition for refinement. The 25% and 50% conditions reflected partially cloudy situations that frequently occur during actual optical image acquisition. Finally, the 75% condition corresponded to a scenario in which cloud-free regions accounted for only 25% of the total area, substantially limiting the representativeness of training data. In the present study, this was treated as the practical upper bound for refinement applicability.

3.4. Experimental Design

Table 3 presents the eight experimental cases comprising all possible combinations of three factors: sampling strategy, refinement target, and SAR features. Each case was evaluated under four cloud fraction conditions, three acquisition dates, and three machine learning models, resulting in a total of 288 experiments. Case 1 is the baseline case in which option (1) was applied to all three factors, using random sampling, reflectance as the refinement target, and raw SAR features. Cases 2–4 correspond to configurations in which only one factor at a time was changed to option (2), and Cases 5–8 represent combinations in which option (2) was simultaneously applied to two or more factors.

The analysis of the experimental results was carried out in two stages. The first assessed both the individual and combined effects of each factor and was further divided into individual and combined factor analysis. Cloud fraction was analyzed separately, as it constituted an external condition with a fundamentally different nature from the three training-related factors. In the individual factor analysis, Cases 2, 3, and 4 were compared against the baseline, each representing a configuration in which one factor was individually changed to option (2) relative to Case 1. In the combined factor analysis, the compound effects of applying option (2) to additional factors beyond the most influential single factor were evaluated, and the best case was identified for each model, acquisition date, and cloud fraction condition. The second

stage analyzed changes in error refinement performance as a function of cloud fraction and acquisition date, restricted to the identified best cases.

Table 3. Experimental case configurations based on all possible combinations of options for three factors: sampling strategy, refinement target, and SAR features.

Case	Sampling Strategy	Refinement Target	SAR Features
1	Random	Reflectance	Raw
2	Quantile	Reflectance	Raw
3	Random	Residual	Raw
4	Random	Reflectance	Local
5	Random	Residual	Local
6	Quantile	Reflectance	Local
7	Quantile	Residual	Raw
8	Quantile	Residual	Local

3.5. Quantitative Evaluation

The quantitative performance of the experimental results was evaluated by comparing the refined reflectance with the actual PS reflectance in cloud-contaminated regions using four metrics.

To quantify the magnitude of refinement errors, the relative root mean square error (rRMSE) and relative mean absolute error (rMAE) were calculated by normalizing RMSE and MAE using the mean of the actual values, respectively. These metrics enable comparisons across spectral bands regardless of differences in reflectance range.

The structural similarity index measure (SSIM), which comprehensively quantifies the luminance, contrast, and structural similarity between two images [49], was calculated to assess spatial structural similarity as follows:

$$\mathrm{SSIM}\left(\hat{R}_{refined,b}, R_{actual,b}\right) = \frac{\left(2\mu_{\hat{R}_{refined,b}}\mu_{R_{actual,b}} + C_1\right)\left(2\sigma_{\hat{R}_{refined,b}R_{actual,b}} + C_2\right)}{\left(\mu^2_{\hat{R}_{refined,b}} + \mu^2_{R_{actual,b}} + C_1\right)\left(\sigma^2_{\hat{R}_{refined,b}} + \sigma^2_{R_{actual,b}} + C_2\right)}, \quad (3)$$

where $\hat{R}_{refined,b}$ and $R_{actual,b}$ denote the model output image and the actual PS image for each spectral band, respectively. μ and σ^2 denote the mean and variance of each image and $\sigma_{\hat{R}_{refined,b}R_{actual,b}}$ denotes the covariance between the two images. C_1 and C_2 are stabilizing constants to avoid division by zero.

The spectral angle mapper (SAM), which measures the angle between the multi-band spectral vectors of two images [50], was calculated to assess spectral similarity as follows:

$$\mathrm{SAM} = \cos^{-1}\left(\frac{\sum_{b=1}^{B} R_{actual,b}\hat{R}_{refined,b}}{\sqrt{\sum_{b=1}^{B} {R_{actual,b}}^2}\sqrt{\sum_{b=1}^{B} \hat{R}^2_{refined,b}}}\right) \times \frac{180}{\pi}, \quad (4)$$

where B denotes the total number of spectral bands.

The rRMSE, rMAE, and SSIM values ranged between 0 and 1, with rRMSE and rMAE indicating better performance as they approached 0 and SSIM indicating better performance as it approached 1. For the SAM, values closer to 0° indicated greater similarity between refined and actual reflectance.

To assess the degree of improvement achieved through error refinement for each metric, relative improvement (RI) was calculated as

$$RI_M(\%) = \frac{M_{init} - M_{refined}}{M_{init}} \times 100, \quad (5)$$

where M denotes the value of each evaluation metric. M_{init} denotes the metric value of the initially reconstructed optical imagery and $M_{refined}$ denotes the metric value after error refinement.

Finally, Kullback–Leibler divergence (D_{KL}) was used as supplementary information to aid in the interpretation of error refinement performance by quantifying the distributional difference between cloud-free and -contaminated regions [51], defined as

$$D_{KL}(P \parallel Q) = \sum_{x} P(x) \log \frac{P(x)}{Q(x)}, \quad (6)$$

where P and Q denote the reflectance distributions of cloud-free and -contaminated regions, respectively, and x denotes the reflectance value. A larger D_{KL} value indicates a greater distributional difference between the two regions and, consequently, a larger gap between the training and inference domains.

4. Results

4.1. Effect of Individual Factors

Table 4 presents the improvement in evaluation metrics with respect to spectral band before and after error refinement for Cases 1 through 4 across all three models. To enable the clear identification of differences across cases, the effects of cloud fraction and acquisition date were averaged across cloud fraction conditions and acquisition dates. For the rRMSE and rMAE, improvement values in Case 4 were approximately 2%p higher across all bands compared to Cases 1–3. For the SAM, Case 4 yielded improvements that were higher than those of Cases 1–3 by a minimum and maximum of approximately 1.7% and 4.6%p, respectively. For the SSIM in Cases 1–3, improvement values were negative across most configurations. However, in Case 4, this tendency was mitigated for RF and NGB, with positive improvement values observed across the blue, green, and red bands.

Two effects of local spatial SAR features can be observed in Figure 5. First, certain parcel boundaries that were not distinguishable without local spatial SAR features became identifiable in Case 4. Second, specific bright or dark tonal patterns appearing around cropland parcels were more effectively corrected using local spatial SAR features. For example, a vertically oriented parcel boundary visible in the center of image (f) is indistinct in (b–d) but becomes somewhat more clearly defined in (e), and a bright-toned structure on the lower left is more accurately reproduced in (e) to resemble the true PS image (f). In addition, a dark parcel visible on the left side of the true PS image (m) was initially reconstructed with anomalously high reflectance values (n), and this error was propagated into (i–k) without correction. However, in Case 4, corresponding to (l), the dark reflectance was reconstructed more accurately and more closely matched (m).

Table 4. Improvement in evaluation metrics for each spectral band in terms of the results of the three models in Cases 1–4. The greatest improvement among the four cases is in bold.

Metric	Model	Case 1				Case 2			
		Blue	Green	Red	NIR	Blue	Green	Red	NIR
rRMSE	RF	11.34	13.49	15.51	13.29	11.20	13.54	15.56	13.32
	XGB	11.75	13.94	16.01	13.67	11.71	13.93	16.05	13.72
	NGB	11.80	14.00	16.11	13.94	11.89	14.08	16.07	13.97
rMAE	RF	11.94	14.30	15.96	11.35	11.72	14.21	15.85	11.34
	XGB	12.87	15.13	16.79	11.84	12.61	14.98	16.70	11.88
	NGB	12.90	15.19	16.88	12.02	12.79	15.07	16.73	12.05

Table 4. *Cont.*

Metric	Model	Case 1				Case 2			
		Blue	**Green**	**Red**	**NIR**	**Blue**	**Green**	**Red**	**NIR**
SSIM	RF	−1.40	−2.19	−3.26	−9.89	−1.54	−2.05	−3.36	−9.71
	XGB	**0.98**	0.31	**−0.48**	−6.56	0.67	**0.39**	−0.52	**−6.39**
	NGB	1.53	0.78	0.17	−5.01	1.54	1.09	−0.09	−5.01
SAM	RF	10.94				10.90			
	XGB	11.78				11.77			
	NGB	11.87				11.87			
		Case 3				**Case 4**			
		Blue	**Green**	**Red**	**NIR**	**Blue**	**Green**	**Red**	**NIR**
rRMSE	RF	11.11	13.30	15.20	13.19	**13.79**	**15.32**	**17.77**	**15.39**
	XGB	11.74	13.96	15.85	13.73	**13.07**	**14.68**	**16.94**	**14.10**
	NGB	11.88	14.05	15.99	13.97	**13.86**	**15.56**	**17.87**	**15.60**
rMAE	RF	11.60	14.05	15.57	11.23	**14.86**	**16.78**	**18.84**	**14.39**
	XGB	12.75	15.06	16.61	11.85	**13.99**	**15.93**	**17.89**	**12.95**
	NGB	12.84	15.15	16.73	12.00	**15.37**	**17.33**	**19.29**	**14.58**
SSIM	RF	−2.04	−2.93	−4.45	−11.42	**1.86**	**1.36**	**1.32**	**−1.97**
	XGB	0.62	0.16	−0.88	−6.50	−0.71	−2.07	−2.72	−12.71
	NGB	1.28	0.62	−0.26	−5.44	**2.79**	**2.48**	**2.49**	**−1.87**
SAM	RF	10.60				**15.23**			
	XGB	11.62				**13.53**			
	NGB	11.68				**15.46**			

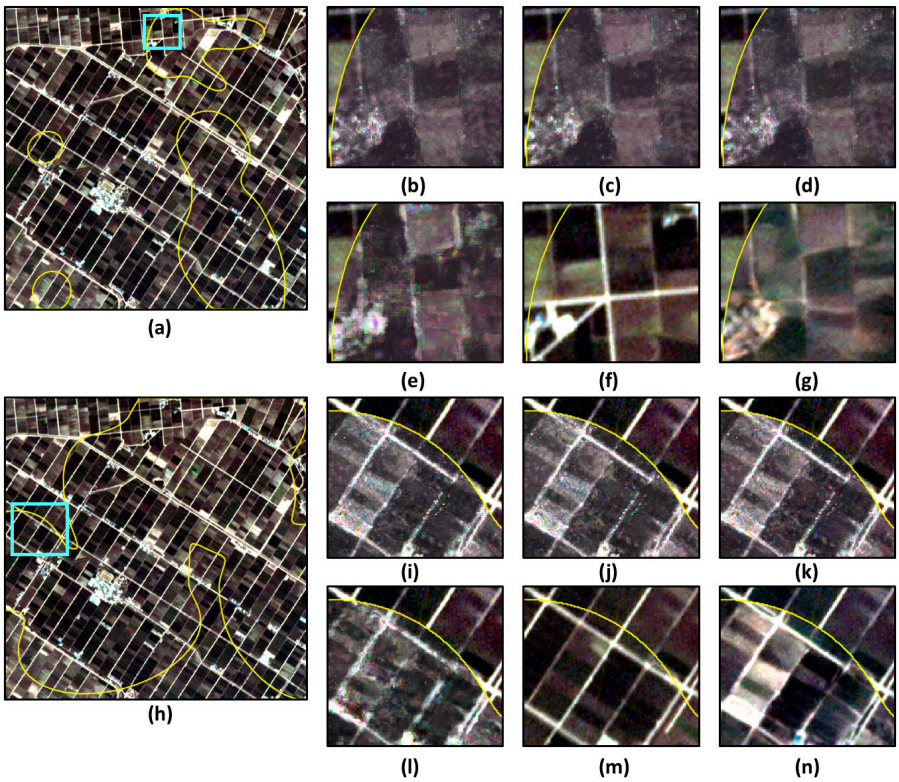

Figure 5. Experimental results for 6 July using 25% (**a**–**g**) and 75% (**h**–**n**) cloud fraction masks: (**a**,**h**) show the PS image and cloud masks (yellow lines) at each cloud fraction along with the zoomed-in area (cyan square), (**b**–**e**) and (**i**–**l**) correspond to Cases 1–4 in order, (**f**,**m**) show the zoomed-in true PS image, and (**g**,**n**) show the initial prediction overlaid on the cloud region.

4.2. Effect of Combined Factors

The best cases across all acquisition dates and models were narrowed down to Cases 4–6 and 8, all of which included local spatial SAR features (Table 5). This is consistent with the effect of local spatial SAR features confirmed in Section 4.1. Overall, Cases 4 and 6 were selected as the best with equal frequency, though differences were observed depending on the acquisition date. Case 4 demonstrated superior performance for the July and September images, whereas Case 6 was the most dominant for the May image.

Table 5. Summary of best cases with respect to the model selected based on each target date and cloud fraction.

Date	Model	Best Case *			
24 May	RF	8	6	6	8
	XGB	2	8	6	6
	NGB	6	6	6	8
6 July	RF	4	4	6	4
	XGB	4	5	6	4
	NGB	4	6	8	4
30 September	RF	4	4	4	6
	XGB	5	4	6	6
	NGB	4	6	4	4
Cloud fraction (%)		10	25	50	75

* Cases 4 and 6 ranked 1st with 14 and Case 8 ranked 3rd with 5.

Table 6 presents the improvement in evaluation metrics with respect to the spectral band for Cases 4–6 and 8. Here, Case 4 served as the reference configuration in which only the SAR features were changed to the local spatial approach relative to Case 1. Cases 5, 6, and 8 represented configurations in which the sampling strategy or refinement target was further modified to the quantile-based or residual approach. Overall, the highest improvement values were observed in Cases 4 and 6. Case 6 corresponded to the configuration in which training samples were drawn using the quantile-based approach and local spatial SAR features were used as input simultaneously. For the rRMSE, rMAE, and SSIM, Case 4 yielded slightly greater improvement in the visible bands, whereas Case 6 showed greater improvement in the NIR band. The SAM was higher in Case 4 than in Case 6. These results indicate that the quantile-based approach is particularly effective for the NIR band. Additional performance gains are expected through the combination of local spatial SAR features with quantile-based sampling during periods with a wide range of NIR values, such as in May.

Table 6. Improvement in evaluation metrics with respect to spectral band for the results of the three models in Cases 4–6 and 8. The greatest improvement among the four cases is in bold, and when two cases share the greatest improvement, both are underlined.

Metric	Model	Case 4				Case 5			
		Blue	Green	Red	NIR	Blue	Green	Red	NIR
rRMSE	RF	**13.79**	15.32	**17.77**	15.39	13.60	15.24	17.50	15.20
	XGB	**13.07**	<u>14.68</u>	16.94	14.10	13.16	14.56	16.81	13.87
	NGB	**13.86**	15.56	17.87	15.60	13.87	15.52	17.75	15.60
rMAE	RF	**14.86**	**16.78**	**18.82**	14.39	14.48	16.47	18.38	14.19
	XGB	**13.99**	**15.93**	**17.89**	12.95	14.04	15.81	17.77	12.86
	NGB	**15.37**	**17.33**	**19.29**	14.58	15.28	17.18	19.12	14.52

Table 6. *Cont.*

Metric	Model	Case 4				Case 5			
		Blue	**Green**	**Red**	**NIR**	**Blue**	**Green**	**Red**	**NIR**
SSIM	RF	**1.86**	1.37	**1.32**	−1.97	1.13	0.55	0.20	−4.48
	XGB	−0.71	−2.05	**−2.72**	−12.71	**−0.30**	−2.11	−2.82	−13.99
	NGB	2.79	**2.48**	**2.49**	−1.87	2.53	2.20	2.07	−2.05
SAM	RF			**15.23**				14.76	
	XGB			**13.53**				13.40	
	NGB			**15.46**				15.28	
		Case 6				**Case 8**			
		Blue	**Green**	**Red**	**NIR**	**Blue**	**Green**	**Red**	**NIR**
rRMSE	RF	13.64	**15.37**	17.50	**15.49**	13.60	15.26	17.13	15.24
	XGB	12.90	14.68	**16.98**	**14.12**	13.00	14.61	16.70	14.04
	NGB	13.79	**15.62**	**17.88**	**15.62**	13.71	15.55	17.71	15.61
rMAE	RF	14.70	16.76	18.61	**14.43**	14.43	16.42	18.12	14.17
	XGB	13.64	15.85	17.81	**13.06**	13.74	15.70	17.46	13.02
	NGB	15.27	17.24	19.15	**14.61**	15.09	17.11	18.95	14.54
SSIM	RF	1.83	**1.52**	1.18	**−1.65**	1.29	0.56	−0.12	−4.26
	XGB	−1.20	**−1.91**	−2.99	−12.65	−0.78	−2.26	−3.80	**−12.52**
	NGB	**3.00**	2.44	2.40	**−1.68**	2.68	2.13	1.97	−2.23
SAM	RF			15.16				14.68	
	XGB			13.48				13.25	
	NGB			15.40				15.17	

This pattern is further illustrated in Figure 6, which shows the variation in actual reflectance distributions across the three acquisition dates for the red and NIR bands. Distributional changes were more pronounced in the NIR band than in the visible bands owing to the characteristics of the study area. During the crop transplanting stage in May, the NIR reflectance distribution was broad and heterogeneous due to differences in parcel-level crop conditions, which explains the superior performance of the quantile-based approach during this period. In July and September, however, the NIR reflectance distribution converged to a unimodal shape as rice growth progressed and reflectance values concentrated in the higher range, thus reducing the additional benefit of quantile-based sampling.

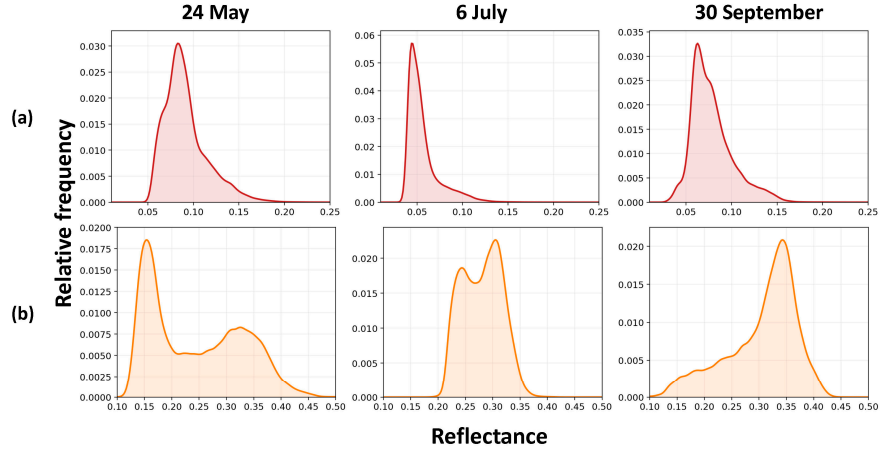

Figure 6. Distributions of PS reflectance values across three acquisition dates: (**a**) and (**b**) correspond to red and NIR bands, respectively.

4.3. Effect of Cloud Fraction

As shown in Figure 7, a consistent decrease in improvement across all metrics was observed as cloud fraction increased. As the cloud-contaminated area expanded, the spatial extent of cloud-free regions diminished, which fundamentally limited the number of extractable training samples. However, the improvement values decreased with increasing cloud fraction, even though the number of training samples was fixed at 50,000 in this study. This indicates that error refinement performance may be governed not by the absolute number of training samples but by the extent to which the reflectance distribution of cloud-free regions is representative of that of cloud-contaminated regions. With a limited pool of training samples, the range of the reflectance distribution that could be adequately represented also changed, and this effect was expected to be further amplified when the distributional discrepancy between cloud-free and -contaminated regions is large. Figure 7 also indirectly confirms the error refinement performance of each model. In terms of rRMSE, rMAE, and SAM improvement, NGB consistently outperformed RF and XGB in that order. Notably, negative SSIM improvement values were observed for RF and XGB under certain cloud fraction conditions, whereas NGB maintained positive SSIM improvement across all conditions. For XGB, SSIM improvement remained largely negative even in the best case.

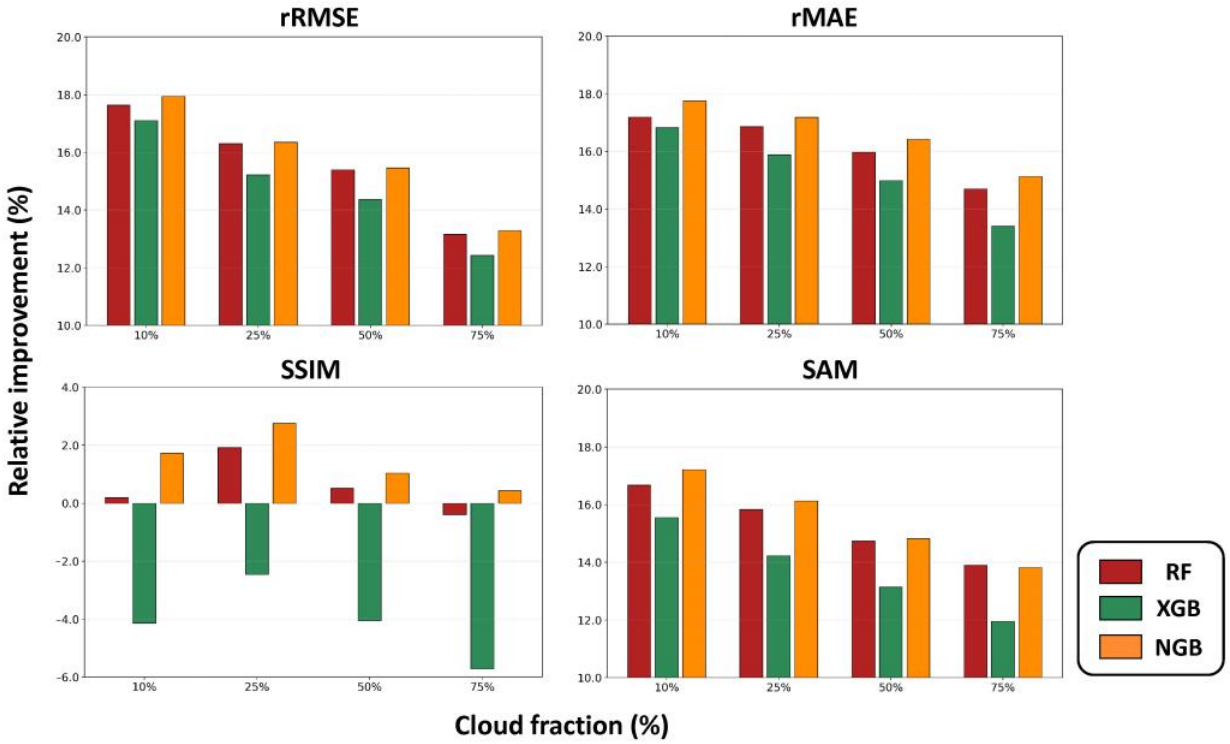

Figure 7. RI (%) of rRMSE, rMAE, SSIM, and SAM with respect to different cloud fractions.

Figure 8 shows the histogram distributions of cloud-free and -contaminated regions in the NIR band along with the corresponding D_KL values for each cloud fraction condition and acquisition date. The results confirm that as cloud fraction increased, the distributional discrepancy between the two regions grew, further limiting the representativeness of training samples. At the same time, it was confirmed that regression-based error refinement can improve the quality of initially reconstructed optical imagery under conditions in which cloud fraction is not excessively high.

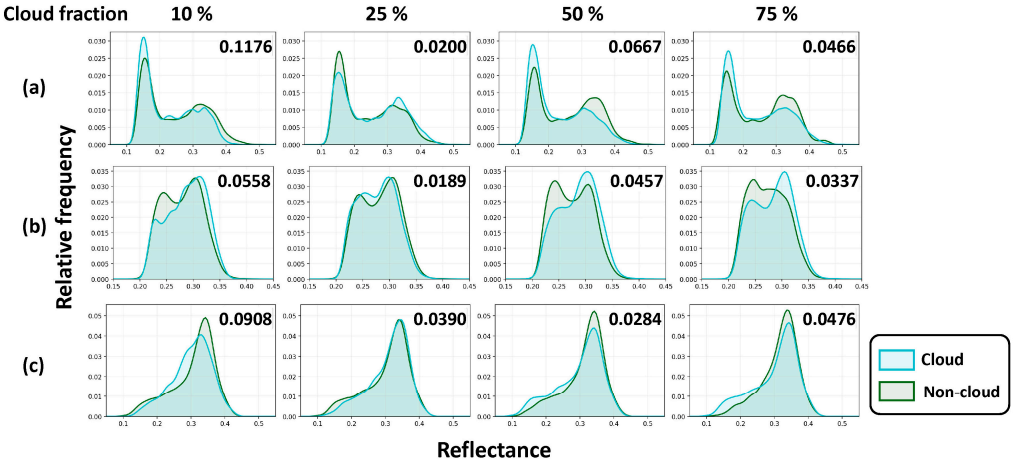

Figure 8. Histograms and D_{KL} values between cloud-free and -contaminated regions in the study area for each cloud fraction condition and acquisition date. The D_{KL} values are displayed in the upper right corner of each panel. (**a**), (**b**), and (**c**) correspond to 24 May, 6 July, and 30 September, respectively.

4.4. Effect of Acquisition Date

Figure 9 presents a comparison of the error refinement improvement in the best cases across all models with respect to acquisition date. In terms of the rRMSE and rMAE, improvement values were positive across all three acquisition dates. In particular, improvement values exceeding 20% were observed for all models in September, representing the greatest improvement among the three dates. In contrast, July and May showed relatively lower values of approximately 11–15%. All three models achieved approximately 18–20% improvement in SAM for the September results, indicating that the refinement effect in terms of spectral similarity was most pronounced in September. SSIM improvement, in contrast, exhibited a pattern inconsistent with the other metrics. In May, positive SSIM improvement was observed only for NGB, and in July, marginal positive values of approximately 2–3% were observed for RF and NGB, whereas all models showed negative SSIM improvement in September. Differences in model-level refinement performance were also observed in terms of the coefficient of determination. In Table 5, the best case for XGB was Case 4 or 6, similar to the other models, but Case 2 or 5 was selected under certain conditions.

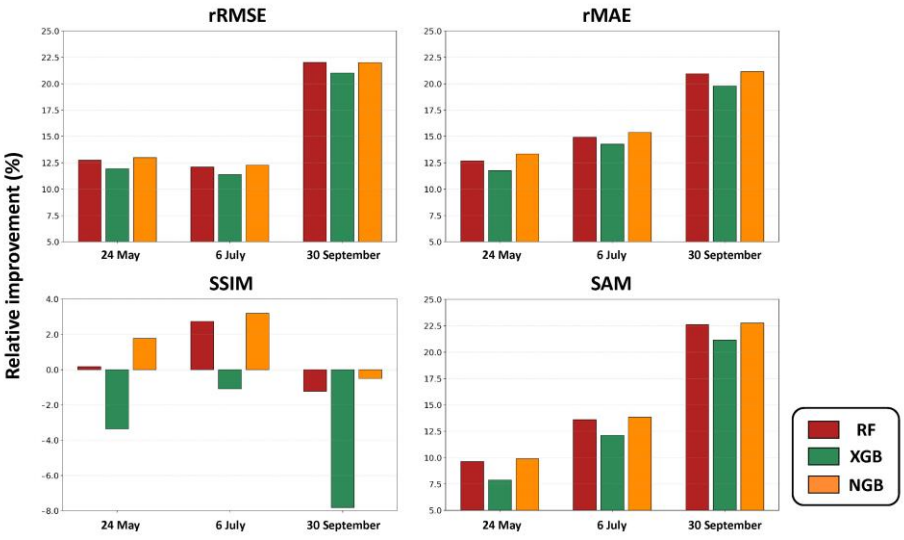

Figure 9. RI (%) of rRMSE, rMAE, SSIM, and SAM with respect to different acquisition dates.

The D_{KL} values quantifying the distributional discrepancy in NIR reflectance between cloud-free and -contaminated regions were 0.0627 for May, 0.0385 for July, and 0.0515 for September. Although all values are very small, they can be ranked in descending order as May, September, and July. A smaller D_{KL} value suggested a higher likelihood that the model could generalize from patterns learned in cloud-free regions to cloud-contaminated regions. The relatively larger D_{KL} value in May stemmed from the distributional characteristics of the NIR band described earlier. Because fields that had already been transplanted and those that had not yet been planted coexisted in May, the shape of the NIR reflectance distribution in cloud-free regions varied slightly depending on the location of the cloud mask. This distributional discrepancy between cloud-free and -contaminated regions provides an explanation for why May showed the least improvement.

For the image acquired in September, improvement was the greatest among the three acquisition dates, despite the distributional discrepancy between cloud-contaminated and cloud-free regions not being the smallest. This observation can be interpreted from two additional perspectives. One reason is that the magnitude of error in the initially reconstructed optical imagery differed across acquisition dates, as shown in Table 7. The band-averaged rRMSE and rMAE of the initially reconstructed optical imagery from September were 0.2895 and 0.2075, respectively, which were the largest among the three dates, while the SSIM and SAM were comparable between September and May at approximately 0.61 and 6.0°. July showed superior values across all error statistics compared to the other two dates. These results suggest that a larger magnitude of error in the initially reconstructed optical imagery left greater room for improvement through subsequent error refinement.

Table 7. Evaluation metrics of initially reconstructed optical imagery for each acquisition date and spectral band.

Date	Band	rRMSE	rMAE	SSIM	SAM (°)
24 May	Blue	0.1906	0.1282	0.6689	6.0171
	Green	0.1989	0.1397	0.6614	
	Red	0.2630	0.1869	0.6367	
	NIR	0.2716	0.1948	0.4690	
	Mean	0.2310	0.1624	0.6090	
6 July	Blue	0.1935	0.1284	0.7766	3.5796
	Green	0.1627	0.1114	0.7782	
	Red	0.3172	0.2220	0.7594	
	NIR	0.1347	0.1059	0.5384	
	Mean	0.2020	0.1419	0.7131	
30 September	Blue	0.2939	0.2032	0.6710	6.0068
	Green	0.2462	0.1805	0.6494	
	Red	0.3853	0.2770	0.5940	
	NIR	0.2327	0.1693	0.5278	
	Mean	0.2895	0.2075	0.6106	

The other reason for the abovementioned observation is that the degree to which SAR features captured the spatial structure of optical image reflectance varied across acquisition dates. As shown in Figure 10, as the season progressed from May (transplanting period) to July (mid-growth stage) and September (maturation and harvest period), the vegetation volume of rice increased, and the volume scattering component within the vegetation canopy became increasingly dominant in the SAR signal, thus causing it to more sensitively reflect vegetation structure rather than surface reflectance. The correspondence between local spatial SAR features and the spatial structure of reflectance was therefore the

weakest in September. As a result, while local spatial SAR features could still be useful for correcting pixel-level reflectance errors in September, they appeared to provide insufficient information for reconstructing the spatial structure evident in optical imagery, which was interpreted as the underlying cause of the pattern in which the rRMSE, rMAE, and SAM improved while the SSIM degraded in September. The quality of initially reconstructed optical imagery in July was the highest among the three acquisition dates, leaving the least room for error improvement. This is considered the primary reason why improvement in July was comparable to or only marginally greater than that in May, despite July showing the smallest distributional discrepancy between cloud-free and -contaminated regions.

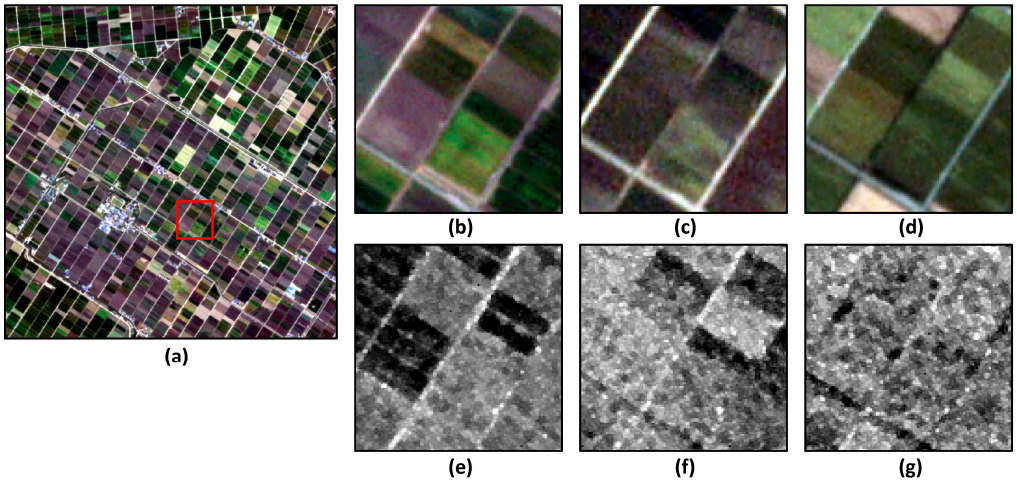

Figure 10. PS (**a**–**d**) and CSK (**e**–**g**) sub-images of a rice paddy area that varied across acquisition dates. The red square indicates the zoomed-in area. (**b**,**e**) correspond to 24 May, (**c**,**f**) to 6 July, and (**d**,**g**) to 30 September.

4.5. Comparison of Model Performance

Table 8 summarizes the coefficient of determination (R^2) computed for all three models in both the training and test stages. In the training stage, XGB achieved the highest mean R^2, while NGB showed the lowest, reflecting XGB's tendency to rapidly improve its fit to training data through its boosting-based structure. In the test stage, however, this trend was reversed: NGB achieved the highest mean R^2, RF performed at a nearly equivalent level, and XGB showed the lowest value.

Table 8. Summary of training and test R^2 values for each model.

Date	Training Data		
	RF	**XGB**	**NGB**
24 May	86.1% ± 0.02% p	81.9% ± 0.25% p	55.0% ± 0.03% p
6 July	81.7% ± 0.01% p	84.6% ± 0.15% p	47.8% ± 0.02% p
30 September	80.0% ± 0.03% p	90.7% ± 0.08% p	44.0% ± 0.05% p
Date	**Test Data**		
	RF	**XGB**	**NGB**
24 May	32.4% ± 0.02% p	31.2% ± 0.02% p	32.7% ± 0.03% p
6 July	27.3% ± 0.03% p	26.1% ± 0.03% p	27.5% ± 0.03% p
30 September	14.2% ± 0.02% p	12.1% ± 0.02% p	14.2% ± 0.03% p

These results can be explained by differences in the learning mechanisms of each model. XGB may develop an excessive dependence on specific features during the sequential training of multiple weak learners, which can lead to overfitting to the training data and

degraded generalization performance. Indeed, the standard deviation of the training stage R^2 was the largest for XGB, thus indicating high case-to-case performance variability within the same acquisition date condition. In contrast, NGB directly modeled the probability distribution of the output values, which could provide greater robustness to changes in the input data distribution and contribute to relatively stable generalization performance in the refinement setting where a distributional gap between the training and inference domains existed. Benefiting from a bagging-based ensemble structure, RF effectively suppressed overfitting while maintaining a relatively straightforward training process, achieving test performance comparable to NGB across all conditions.

5. Discussion

5.1. Interpretation of Results and Practical Implications

Existing studies have addressed error refinement either under linear assumptions [27], through limited nonlinear regression approaches [25], or in the broader contexts of cloud removal and gap-filling [28,29]. However, these studies did not systematically analyze the factors influencing refinement performance, limiting their ability to identify which factors should be prioritized in practical applications or provide evidence-based guidelines for designing regression-based error refinement procedures. One of the key contributions of this study is the experimental demonstration that refinement performance depends not only on the choice of regression model but also on the spatial representativeness of input features and the distributional representativeness of training samples. This systematic evaluation of multiple regression models across diverse experimental conditions, including different sampling strategies, refinement targets, SAR feature configurations, and cloud fraction levels, allowed us to quantify the relative importance of each factor and derive evidence-based guidelines that would be difficult to establish from single-model or single-condition analyses alone.

The major findings and their practical recommendations are summarized in Table 9. In this study, cloud fraction was identified as the most dominant factor, as its increase led to a fundamental degradation in the distributional representativeness of training samples that went beyond a simple reduction in sample count. This suggests that the feasibility of applying regression-based error refinement should be assessed in advance based on cloud fraction conditions in practical applications. The consistent superiority of local spatial SAR features suggests that neighborhood-level statistical information is more effective for modeling error patterns at the parcel level than pixel-level backscattering values. This finding is in line with the observation that the incorporation of spatial contextual information can improve reconstruction and gap-filling performance in remote sensing applications [29]. Quantile-based sampling is particularly effective when the reflectance distribution is heterogeneous, whereas random sampling provides sufficient representativeness when the distribution is homogeneous, and selective application based on distributional characteristics is therefore recommended. This observation is consistent with previous findings that the representativeness of training samples and the choice of sampling strategy are critical determinants of machine learning model performance [31]. Regarding refinement target, the performance difference between the two approaches was inconsistent across models and experimental conditions, and stable performance gains from the residual approach were not consistently observed. The relatively simpler reflectance approach is therefore considered a more practical baseline for regression-based error refinement. However, residual-based correction may still be effective for fine-scale error adjustment when the initially reconstructed reflectance closely approximates actual reflectance.

Table 9. Summary of key findings and practical recommendations for regression-based error refinement in SOIT-based cloud removal.

Factor	Key Finding	Practical Recommendation
Cloud fraction	Dominant factor affecting sample representativeness	Assess applicability based on cloud fraction
SAR features	Local spatial features outperform pixel-level features	Use local spatial SAR features
Sampling strategy	Quantile-based sampling effective for heterogeneous distributions	Select sampling strategy based on distribution characteristics
Refinement target	Residual refinement shows no consistent advantage	Use reflectance-based refinement as baseline

5.2. Limitations and Future Research Directions

While this study identified key factors influencing error refinement performance, substantial improvement in refinement accuracy was not achieved. This is attributed not to a limitation of the methodology itself but rather to the inherent characteristics of post-correction, which relies solely on SAR features and actual reflectance from cloud-free regions as conditional information, without incorporating any external data sources. As long as SAR features cannot fully explain the reflectance distribution of optical imagery, the scope of correction is inevitably limited, and this should be understood as the marginal limitation of such a conditionally constrained approach. The limitations of the current SAR feature-conditioned approach could potentially be mitigated by incorporating auxiliary optical imagery, such as cloud-free imagery acquired near the acquisition date or historical imagery from similar phenological periods [12], further improving refinement performance. Incorporating such auxiliary optical information into the regression-based refinement framework represents a promising direction for future research.

Although the combined factor analysis in Section 4.2 provided indirect evidence of interactions between factors, for example, the combination of quantile-based sampling and local spatial SAR features was particularly effective in May when the NIR reflectance distribution spanned a wide range, a systematic statistical analysis of interaction effects between two or three factors was not conducted. The present experimental design evaluated all possible combinations of factor options but did not explicitly quantify the magnitude or significance of pairwise or three-way interactions. Future studies should incorporate formal interaction analysis to more rigorously characterize the compound effects of these factors on error refinement performance.

This study was conducted using a single SOIT model (i.e., Pix2Pix); however, comparative experiments involving diverse SOIT models with varying initial prediction performance and error distribution characteristics are needed to systematically analyze how the effectiveness of regression-based error refinement depends on the characteristics of the specific SOIT model employed. Furthermore, since this study was conducted over spatially homogeneous cropland, the generalizability of the findings to heterogeneous environments remains a key limitation. In heterogeneous environments such as areas with diverse crop types, forests, or mountainous regions, the mapping relationship between SAR backscattering and optical reflectance is considerably more complex due to heterogeneous crop types, diverse scattering mechanisms, and geometric distortions. Under such conditions, the error patterns in initially reconstructed optical imagery are likely to be more spatially variable and less predictable from cloud-free region information alone, potentially reducing the effectiveness of the regression-based refinement approach. Moreover, the assumption that error patterns learned from cloud-free regions can be consistently applied to cloud-contaminated regions, which is reasonable for homogeneous cropland, may not

hold in heterogeneous environments where the reflectance distribution varies significantly across the scene. Validating the generalizability of the findings by extending the approach to heterogeneous environments with greater spatial complexity in reflectance distributions remains an important direction for future research.

In this study, synthetic cloud masks were used to control experimental conditions and enable a clear comparison of the effects of each factor. However, real clouds can exhibit diverse cloud types, translucent edges, and irregular boundary characteristics, and the binary mask representation used in this study does not fully reproduce these complexities. Future studies should further validate the generalizability of the findings by conducting experiments under more diverse cloud conditions with real cloud masks.

6. Conclusions

This study comprehensively evaluated machine learning regression models for the regression-based error refinement of optical imagery initially reconstructed by SOIT and quantitatively analyzed the effects of four key factors on refinement performance, comprising sampling strategy, refinement target, SAR features, and cloud fraction. The main findings are summarized as follows.

Cloud fraction was identified as the most dominant factor, as its increase leads to a fundamental degradation in the distributional representativeness of training samples drawn from cloud-free regions. Local spatial SAR features consistently outperformed pixel-level raw SAR features, thus confirming that spatial context at the parcel level is critical for effective error modeling. Quantile-based sampling was particularly effective when combined with local spatial SAR features during periods when the NIR reflectance distribution spanned a wide range. The effect of the refinement target was limited, and the reflectance approach is considered a more practical baseline. Among the three models, NGB demonstrated the most stable generalization performance. However, the performance differences compared with RF were often modest across experimental conditions, while NGB required a higher computational cost. In contrast, RF showed competitive performance with relatively lower computational requirements, suggesting that it may serve as a practical alternative depending on application requirements.

These findings demonstrate that the spatial representativeness of SAR input features and the distributional representativeness of training samples are the key determinants of regression-based error refinement performance in SOIT-based cloud removal. In practical applications, cloud fraction and the reflectance distribution characteristics of each acquisition date should therefore be jointly considered.

Author Contributions: Conceptualization, I.L., S.P. and N.-W.P.; methodology, I.L. and N.-W.P.; formal analysis, I.L.; data curation, I.L., S.P., N.-W.P. and E.H.H.; writing—original draft preparation, I.L. and S.P.; writing—review and editing, N.-W.P. and E.H.H.; supervision, N.-W.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (no. RS-2024-00397964, development of major regional analysis and realization intelligence technology based on micro satellite images).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. The PlanetScope optical imagery and COSMO-SkyMed SAR imagery were obtained from Planet Labs PBC (https://www.planet.com, accessed on 22 April 2026) and the Italian Space Agency (ASI) (https://www.asi.it, accessed on 22 April 2026), respectively. These data are available from the authors with the permission of the respective data providers.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

cGAN	conditional Generative Adversarial Network
CSK	COSMO-SkyMed
NGB	Natural Gradient Boosting
NIR	Near-Infrared
PS	PlanetScope
RF	Random Forest
RI	Relative Improvement
SAM	Spectral Angle Mapper
SAR	Synthetic Aperture Radar
SOIT	SAR-to-Optical Image Translation
SSIM	Structural Similarity Index Measure
XGB	eXtreme Gradient Boosting

Appendix A. Sensitivity Analysis of Training Sample Size

To evaluate the effect of training sample size on refinement performance, a preliminary sensitivity analysis was conducted using sample sizes ranging from 1000 to 100,000 pixels under a representative experimental condition (24 May acquisition date and 50% cloud fraction). The analysis was performed using NGB, which was selected because it required the longest training time among the three models evaluated in this study. The analysis was performed to examine whether increasing the number of training samples consistently improved refinement performance. The resulting performance metrics are presented in Figure A1. The results indicate that the NGB-based refinement performance improved with increasing sample size but became marginal beyond approximately 50,000 pixels, whereas computational cost continued to increase substantially.

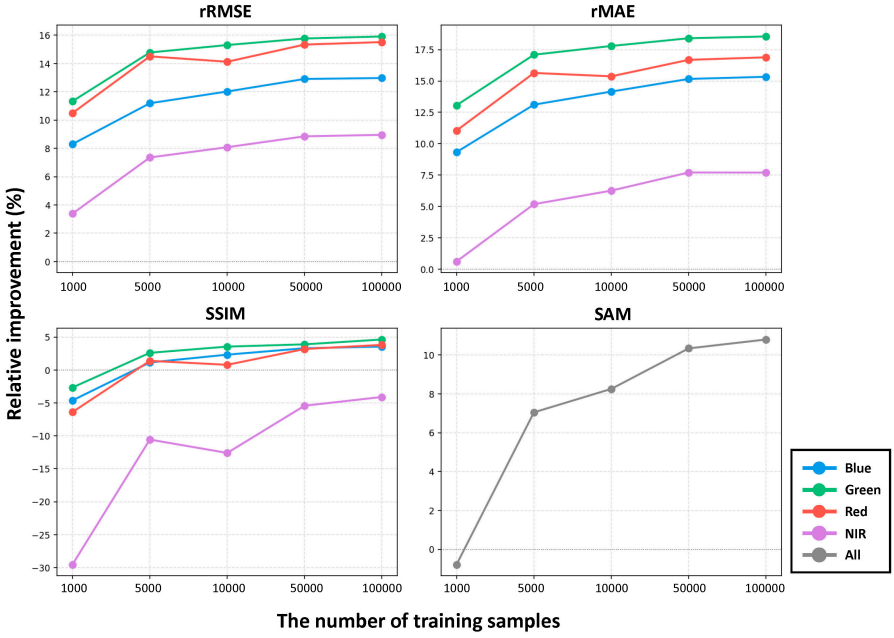

Figure A1. Effect of training sample size on NGB-based refinement performance for four spectral bands (blue, green, red, and NIR) in terms of rRMSE, rMAE, SSIM, and SAM under the representative experimental condition (24 May acquisition date and 50% cloud fraction).

References

1. Ju, J.; Roy, D.P. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens. Environ.* **2008**, *112*, 1196–1211. [CrossRef]
2. Shen, H.; Li, X.; Cheng, Q.; Zeng, C.; Yang, G.; Li, H.; Zhang, L. Missing information reconstruction of remote sensing data: A technical review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 61–85. [CrossRef]
3. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
4. Lee, S.-J.; Ryu, J.-H.; Kwak, G.-H.; Choi, L.-Y.; Jeon, D.; Park, M.; Lee, K.-D. Analysis of PlanetScope imagery-based NDVI changes for monitoring brown planthopper damage. *Korean J. Remote Sens.* **2025**, *41*, 717–726. [CrossRef]
5. Inglada, J.; Vincent, A.; Arias, M.; Marais-Sicre, C. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sens.* **2016**, *8*, 362. [CrossRef]
6. Park, S.; Park, N.-W. Combining Gaussian process regression with Poisson blending for seamless cloud removal from optical remote sensing imagery for cropland monitoring. *Agronomy* **2023**, *13*, 2789. [CrossRef]
7. Lin, C.H.; Tsai, P.H.; Lai, K.H.; Chen, J.Y. Cloud removal from multitemporal satellite images using information cloning. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 232–241. [CrossRef]
8. Cheng, Q.; Shen, H.; Zhang, L.; Yuan, Q.; Zeng, C. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 54–68. [CrossRef]
9. Meraner, A.; Ebel, P.; Zhu, X.X.; Schmitt, M. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 333–346. [CrossRef] [PubMed]
10. Xiong, Q.; Li, G.; Yao, X.; Zhang, X. SAR-to-optical image translation and cloud removal based on conditional generative adversarial networks: Literature survey, taxonomy, evaluation indicators, limits and future directions. *Remote Sens.* **2023**, *15*, 1137. [CrossRef]
11. Cao, R.; Chen, Y.; Chen, J.; Zhu, X.; Shen, M. Thick cloud removal in Landsat images based on autoregression of Landsat time-series data. *Remote Sens. Environ.* **2020**, *249*, 112001. [CrossRef]
12. He, W.; Yokoya, N. Multi-temporal Sentinel-1 and -2 data fusion for optical image simulation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 389. [CrossRef]
13. Whitcraft, A.K.; Vermote, E.F.; Becker-Reshef, I.; Justice, C.O. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sens. Environ.* **2015**, *156*, 438–447. [CrossRef]
14. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]
15. Kwak, G.-H.; Park, N.-W. Assessing the potential of multi-temporal conditional generative adversarial networks in SAR-to-optical image translation for early-stage crop monitoring. *Remote Sens.* **2024**, *16*, 1199. [CrossRef]
16. Fuentes Reyes, M.; Auer, S.; Merkle, N.; Henry, C.; Schmitt, M. SAR-to-optical image translation based on conditional generative adversarial networks—Optimization, opportunities and limits. *Remote Sens.* **2019**, *11*, 2067. [CrossRef]
17. Li, Y.; Fu, R.; Meng, X.; Jin, W.; Shao, F. A SAR-to-optical image translation method based on conditional generation adversarial network (cGAN). *IEEE Access* **2020**, *8*, 60338–60343. [CrossRef]
18. Bermudez, J.D.; Happ, P.N.; Oliveira, D.A.B.; Feitosa, R.Q. SAR to optical image synthesis for cloud removal with generative adversarial networks. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 5–11. [CrossRef]
19. Wang, Z.; Zhang, Z.; Shan, X.; Wei, H.A.; Tang, P. Generative models for SAR–optical image translation: A systematic review. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *146*, 105009. [CrossRef]
20. Yang, X.; Zhao, J.; Wei, Z.; Wang, N.; Gao, X. SAR-to-optical image translation based on improved CGAN. *Pattern Recognit.* **2022**, *121*, 108208. [CrossRef]
21. Park, N.-W.; Park, M.-G.; Kwak, G.-H.; Hong, S. Deep learning-based virtual optical image generation and its application to early crop mapping. *Appl. Sci.* **2023**, *13*, 1766. [CrossRef]
22. Grohnfeldt, C.; Schmitt, M.; Zhu, X. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1726–1729.
23. Bai, X.; Pu, X.; Xu, F. Conditional diffusion for SAR to optical image translation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4000605. [CrossRef]
24. Zou, X.; Li, K.; Xing, J.; Zhang, Y.; Wang, S.; Jin, L.; Tao, P. DiffCR: A fast conditional diffusion framework for cloud removal from optical satellite images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5612014. [CrossRef]
25. Kwak, G.-H.; Park, S.; Park, N.-W. Combining conditional generative adversarial network and regression-based calibration for cloud removal of optical imagery. *Korean J. Remote Sens.* **2022**, *38*, 1357–1369, (In Korean with English Abstract).
26. Wang, Z.; Ma, Y.; Zhang, Y. Hybrid cGAN: Coupling global and local features for SAR-to-optical image translation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5236016. [CrossRef]

27. Li, X.; Wang, L.; Cheng, Q.; Wu, P.; Gan, W.; Fang, L. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 103–113. [CrossRef]
28. Tahsin, S.; Medeiros, S.C.; Hooshyar, M.; Singh, A. Optical cloud pixel recovery via machine learning. *Remote Sens.* **2017**, *9*, 527. [CrossRef]
29. Wang, Q.; Wang, L.; Zhu, X.; Ge, Y.; Tong, X.; Atkinson, P.M. Remote sensing image gap filling based on spatial-spectral random forests. *Sci. Remote Sens.* **2022**, *5*, 100048. [CrossRef]
30. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
31. Bouasria, A.; Bouslihim, Y.; Gupta, S.; Taghizadeh-Mehrjardi, R.; Hengl, T. Predictive performance of machine learning model with varying sampling designs, sample sizes, and spatial extents. *Ecol. Inform.* **2023**, *78*, 102294. [CrossRef]
32. Li, H.; Gu, C.; Wu, D.; Cheng, G.; Guo, L.; Liu, H. Multiscale generative adversarial network based on wavelet feature learning for SAR-to-optical image translation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5236115. [CrossRef]
33. Wei, J.; Zou, H.; Sun, L.; Cao, X.; He, S.; Liu, S.; Zhang, Y. CFRWD-GAN for SAR-to-optical image translation. *Remote Sens.* **2023**, *15*, 2547. [CrossRef]
34. Zhang, Q.; Yuan, Q.; Li, J.; Li, Z.; Shen, H.; Zhang, L. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 148–160. [CrossRef]
35. Zhu, S.; Li, Z.; Shen, H.; Lin, D. A fast two-step algorithm for large-area thick cloud removal in high-resolution images. *Remote Sens. Lett.* **2023**, *14*, 1–9. [CrossRef]
36. Planet Labs PBC. *PlanetScope Scene Imagery Product Specifications*; Planet Labs PBC: San Francisco, CA, USA, 2025. Available online: https://assets.planet.com/docs/Planet_PSScene_Imagery_Product_Spec_letter_screen.pdf (accessed on 22 April 2026).
37. Agenzia Spaziale Italiana (ASI). *COSMO-SkyMed Mission and Products Description*; Doc. N°. ASI-CSM-PMG-NT-001, Rev. 3; ASI: Rome, Italy, 2019. Available online: https://www.asi.it/wp-content/uploads/2019/08/COSMO-SkyMed-Mission-and-Products-Description_rev3-2.pdf (accessed on 22 April 2026).
38. Filipponi, F. Sentinel-1 GRD preprocessing workflow. In Proceedings of the 3rd International Electronic Conference on Remote Sensing, Online, 22 May–5 June 2019; Volume 18, p. 11.
39. Konishi, T.; Suga, Y. Landslide detection using COSMO-SkyMed images: A case study of a landslide event on Kii Peninsula, Japan. *Eur. J. Remote Sens.* **2018**, *51*, 205–221. [CrossRef]
40. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
41. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]
42. Park, S.; Park, M.; Lee, I.; Hwang, E.H.; Park, N.-W. Spatiotemporal feature integration with wavelet decomposition and temporal encoding for SAR-to-optical image translation using generative deep learning. Inha University: Incheon, Republic of Korea, 2026, *manuscript in preparation*.
43. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
44. Cho, M.; Song, H.; Jeong, S.; Han, H. Tracking large tabular icebergs in Antarctica using AMSR2 observations and random forest: A case study of A23A. *Korean J. Remote Sens.* **2025**, *41*, 341–352. [CrossRef]
45. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
46. Duan, T.; Anand, A.; Ding, D.Y.; Thai, K.K.; Basu, S.; Ng, A.; Schuler, A. NGBoost: Natural gradient boosting for probabilistic prediction. In Proceedings of the 37th International Conference on Machine Learning (ICML), Virtual, 13–18 July 2020; pp. 2690–2700. Available online: https://proceedings.mlr.press/v119/duan20a.html (accessed on 9 April 2026).
47. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, É. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
48. Probst, P.; Wright, M.N.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
49. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
50. Dennison, P.E.; Halligan, K.Q.; Roberts, D.A. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sens. Environ.* **2004**, *93*, 359–367. [CrossRef]
51. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.