Reconstructing Sentinel-2 Rice NDVI Time Series for Cloudy Seasons in India

Zhiqing Huang[1], Meha Jain[1], Victor H. Prudente[1], Monish Desphande[1], Hantao Wang[1]
[1] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

Abstract

Persistent monsoon cloud cover creates long interruptions in optical satellite time series over rice-growing regions of India, often during the most informative stages of crop development. Because synthetic aperture radar (SAR) is much less affected by cloud cover than optical sensors, it provides an important complementary source of information under these conditions. Field-scale reconstruction of optical time series for crop phenology remains difficult, however, because observations are sparse and direct radar-to-optical learning is weakly constrained. Here we developed a baseline-guided framework for 502 smallholder rice fields in Chhattisgarh, India. A daily Optical Sentinel-2 NDVI baseline was first estimated with Whittaker smoothing, and a machine-learning model using past-only SAR Sentinel-1 features were then used to predict deviations from that baseline within the cloud-gap period. Evaluation was centered on an inside-gap benchmark rather than whole-season averages. In the Jul 1–Sep 28 window, RMSE decreased from 0.019 to 0.012 for direct-seeded fields and from 0.032 to 0.018 for transplanted fields relative to the baseline alone. A SAR-only ablation performed similarly to the broader full model, indicating that radar backscatter itself contributed useful information. A supplementary direct-prediction experiment without the baseline performed substantially worse (RMSE = 0.112 and 0.084 for direct-seeded and transplanted fields respectively), confirming that an optical scaffold is necessary for stable reconstruction under current data conditions. These results support baseline-guided reconstruction as a stable and defensible strategy for strengthening field-scale rice NDVI trajectories under severe monsoon cloud gaps.

Introduction

Optical satellite time series are fundamental to crop monitoring because they provide direct information on canopy condition, seasonal development, and spatial variability in agricultural systems[1,2-5]. This problem is especially acute in monsoon Asia, where long optical gaps often occur during the most dynamic stages of rice development[3,4]. In such settings, field-scale phenology analysis becomes difficult not simply because observations are missing, but because the missing period often overlaps with establishment and rapid vegetative growth.

To address this limitation, previous studies have used three main types of approaches. Optical-only studies have applied smoothing, fitting, and gap-filling methods to incomplete vegetation time series[2,5,6]. These methods can stabilize noisy or discontinuous trajectories, but their success depends strongly on gap structure, the number of available support points, and the type of information used to fill missing values[2,5,7,8]. A second line of work has used Sentinel-1 or other SAR observations for rice monitoring, including rice mapping, start-of-season detection, crop-growth tracking, and even direct NDVI estimation from radar backscatter9-13. Because SAR observations are much less affected by cloud cover than optical observations, they are particularly valuable in monsoon rice environments. These studies show that C-band radar is sensitive to flooding, canopy development, and management differences, making it useful for monitoring rice under persistent cloud conditions[10-12,14]. A third line of work combines optical and radar data in broader fusion and reconstruction frameworks, including machine-learning and deep-learning

approaches[5,13,15-17]. Together, these studies show that SAR can support optical reconstruction, but they do not resolve the specific problem addressed here, namely field-scale NDVI reconstruction under long contiguous monsoon gaps in a smallholder rice system.

This problem remains difficult for two reasons. First, the missing interval in the present case is not a routine sparse-pixel interpolation problem. It behaves more like a blocked observation window: the main canopy-rise stage occurs during a long cloud-dominated period, and clear Sentinel-2 observations inside Jul–Sep are typically fewer than three per field. Second, the predictive task is based on field-level temporal summaries rather than large image tensors, so model choice is part of the scientific design rather than only a computational preference.

This study was motivated by that methodological gap. The central question was not whether Sentinel-1 is correlated with NDVI in a general sense, but whether it can strengthen field-scale NDVI trajectories under long monsoon cloud gaps (Jul–Sep) in a way that remains phenologically meaningful and empirically stable. The final framework was shaped by several alternative routes explored during development, including direct radar-based prediction without an optical prior, multiple smoothing and curve-fitting options for baseline construction, downstream start-of-season extraction from reconstructed trajectories, and machine-learning alternatives with different grouping and feature-selection strategies. Separate modeling for direct-seeded and transplanted fields emerged as a particularly important design decision, reflecting the distinct early-season hydrological and structural conditions of the two establishment types and their different radar responses. Deep neural architectures were not prioritized in the final framework because the reconstruction task was formulated as a moderate-sized tabular regression problem with limited labeled field-date samples, a setting in which more parsimonious tree-based methods are often more suitable than data-hungry deep learning models[18,19].

The final framework therefore follows a two-stage logic. A daily baseline NDVI trajectory is first estimated from Sentinel-2 observations using Whittaker smoothing[6,7]. A machine-learning residual model then predicts deviations from that baseline using past-only Sentinel-1 summaries. This design reduces the burden placed on the radar model, because it no longer needs to learn the entire seasonal shape from sparse observations, and it also makes the role of Sentinel-1 more interpretable, because its contribution can be evaluated as incremental improvement over an explicit optical reference.

This study addresses three questions. First, does this baseline-guided framework improve NDVI reconstruction within the monsoon cloud-gap period relative to the optical baseline alone? Second, is any improvement attributable to Sentinel-1 backscatter itself rather than only to shortcut-like temporal features? Third, what do the supplementary experiments reveal about the conditions under which field-scale NDVI reconstruction is more stable than direct absolute prediction?

However, this study does not claim exact recovery of unobserved phenological detail. Instead, it shows that under severe monsoon cloud gaps, a baseline-guided reconstruction framework provides a more stable route than direct absolute prediction for strengthening field-scale rice NDVI trajectories. The framework also makes it possible to evaluate the contribution of Sentinel-1 explicitly through residual correction and ablation analysis. In this sense, the study contributes a practical reconstruction strategy for monsoon rice systems while also clarifying the limits imposed by sparse optical support, baseline smoothing, and radar temporal resolution.

Methods

Methodological pipeline

The framework comprises five modules (Fig. 1). (1) Data sources: field attributes (IRRI), Sentinel-2 optical time series (SCL=4 filtered), Sentinel-1 SAR backscatter (VV/VH/RVI/CR), all extracted via Google Earth Engine. (2) Layer A constructs a per-field seasonal optical scaffold using Whittaker smoothing. (3) Layer B predicts residuals from the baseline using past-only SAR feature windows (15/30/60 days) with a grouped ExtraTrees model. (4) Evaluation compares the full model, a SAR-only ablation, and a no-baseline control within the Jul–Sep Step B benchmark. (5) Downstream validation tests whether reconstructed trajectories support sowing-date prediction.

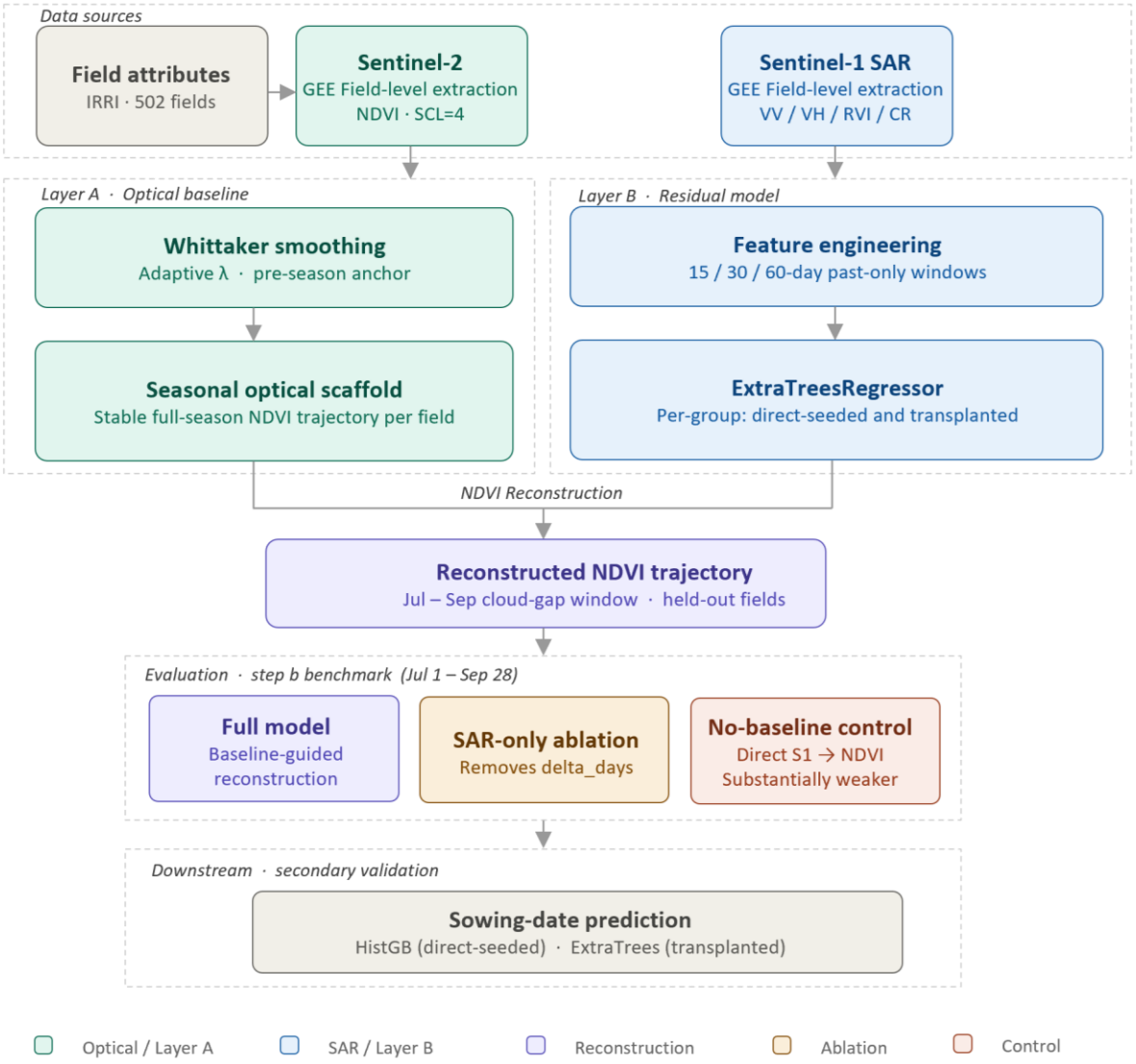


Figure 1. Methodological pipeline for Sentinel-2 NDVI reconstruction.

Study area

We performed our study in Chhattisgarh, a major rice-growing region in central India, which is located in the South Asian sub-monsoon domain, the Kharif rice production is connected with seasonal rain and cloud cover[3,4]. These conditions make Chhattisgarh a suitable setting for reconstruction analysis because the main rice development period overlaps with prolonged interruption of optical satellite observations. The agricultural landscape is dominated by smallholder fields. Rice is established through both direct seeding and transplanting. Together, persistent cloud cover, fragmented field structure, and multiple establishment pathways make Chhattisgarh a challenging but relevant testbed for field-scale NDVI reconstruction12–14 (Fig. 2)[9,10].

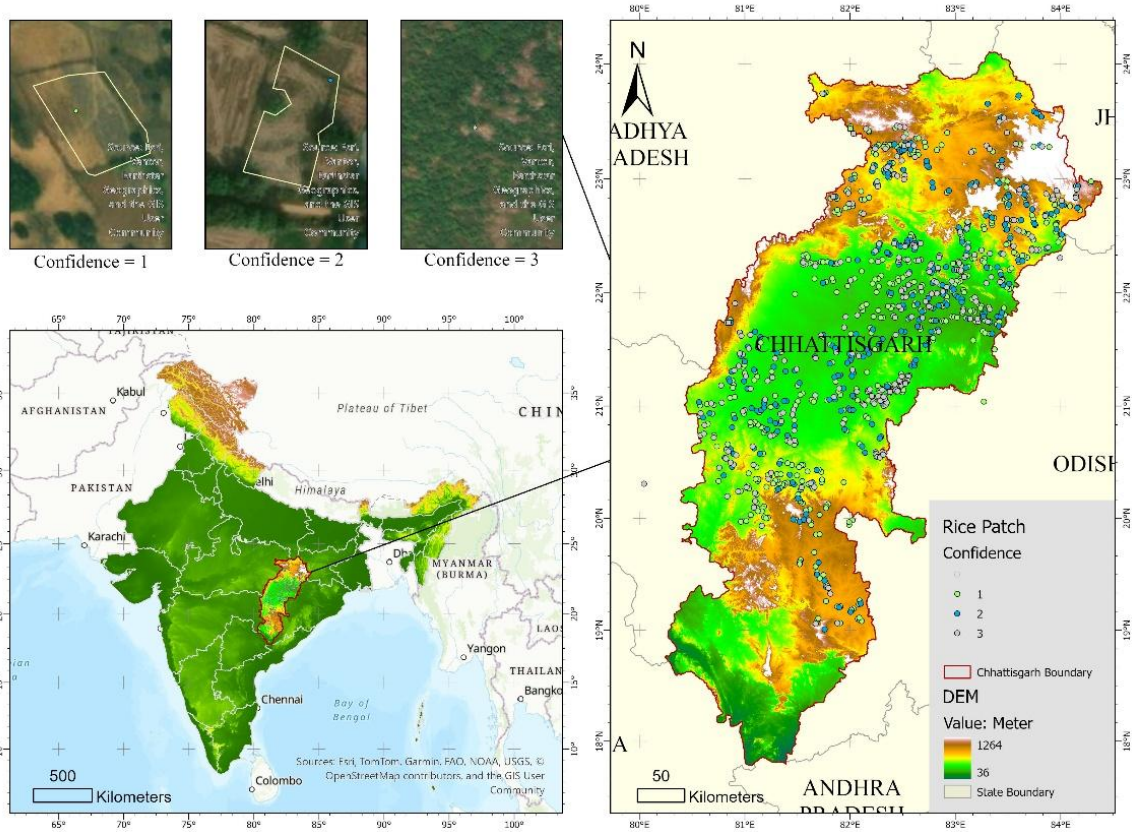


Figure 2. Study area and field selection. Location of the study area in Chhattisgarh, India (inset: India map; main panel: field distribution and district boundaries), together with the field-selection workflow from the initial inventory of 757 polygons to the 502-field monsoon analysis set.

From a remote-sensing perspective, the study area is a challenge not only because cloud gaps are severe, but also because it represents the kind of smallholder monsoon rice system in which reconstruction choices matter most. In larger and more homogeneous production systems, missing optical observations may be partly compensated by spatial aggregation. In fragmented field systems, however, field-level variability is part of the problem. The study area therefore provides an appropriate setting for evaluating whether a baseline-guided Sentinel-1 plus Sentinel-2 framework can improve NDVI trajectory continuity under prolonged cloud interruption and strong within-region management heterogeneity.

Data sources

The formal analysis combined field-level attribute data, field boundary polygons, Sentinel-2 optical observations, and Sentinel-1 SAR metrics. These datasets served complementary roles in defining the analysis sample, constructing the benchmark, and supporting field-level reconstruction and downstream validation. A summary of the main datasets and the final analytical sample is provided in Table 1.

Table 1. Data sources and analysis sample.

Dataset or sample	Source	Time range	Role in this study	Key quantity
Field attribute table	IRRI field information	2023 Kharif season	Establishment type, sowing date, field ID, confidence labels	757 fields
Field boundary polygons	Self-digitized shapefile	2023 Kharif season	Spatial analysis units for all field-level extractions	757 polygons
Final monsoon analysis set	Derived from field filtering	2023 Kharif season	Formal sample for all grouped analyses	502 fields
Direct-seeded fields	Derived from field filtering	2023 Kharif season	Group-specific modeling	145 fields
Transplanted fields	Derived from field filtering	2023 Kharif season	Group-specific modeling	357 fields
Sentinel-2 training/evaluation table	Google Earth Engine	2023-05-30 to 2023-12-31	Official NDVI source for Layer A support and Step B ground truth	6,083 rows raw; 2,150 rows after SCL = 4
Sentinel-2 visualization table	Google Earth Engine	2023-05-30 to 2023-12-31	Richer full-season plotting only	~13,785 rows
Sentinel-1 metrics table	Google Earth Engine	2023-05-01 to 2024-01-31	Radar predictors for Layer B	19,661 field-date records
Step B evaluation subset	Derived from S2 training table	Jul 1–Sep 28, 2023	Real within-gap reconstruction targets	~150 field-date pairs per group

Field data

Field attributes were obtained from IRRI (International Rice Research Institute) and included establishment type, sowing-date information, field identifiers, and confidence labels for the 2023 Kharif season. Self-digitized field polygons created in Google Earth Pro were used as the spatial units for all field-level extraction. The initial inventory contained 757 polygons. After restricting the analysis to fields with valid monsoon-relevant sowing-timing labels, 502 fields were retained for the formal analysis, including 145 direct-seeded fields and 357 transplanted fields. Observed sowing dates were used only for downstream validation and were not included as predictors in the primary reconstruction model.

Optical data

Field-level Sentinel-2 time series were extracted in Google Earth Engine from the COPERNICUS/S2_SR_HARMONIZED collection[1,20-24]. Restricting the formal pipeline to SCL = 4 therefore reduced the influence of exposed soil and other non-canopy surfaces on field-

mean NDVI support points. This was especially important because the baseline was fitted at field scale under sparse observations; including additional not-vegetated support points could shift the fitted trajectory toward surface-state variability rather than crop-canopy development.

SAR data

Field-level Sentinel-1 time series were compiled for the same agricultural season in Google Earth Engine from the COPERNICUS/S1_GRD collection[10-12,14,25,26]. The ratio-based descriptors were defined as

$$CR_{dB} = VV_{dB} - VH_{dB}$$

and

$$PR_{lin} = \frac{VH_{lin}}{VV_{lin}}$$

RVI was computed as

$$RVI = \frac{4 \cdot VH_{lin}}{VV_{lin} + VH_{lin}}$$

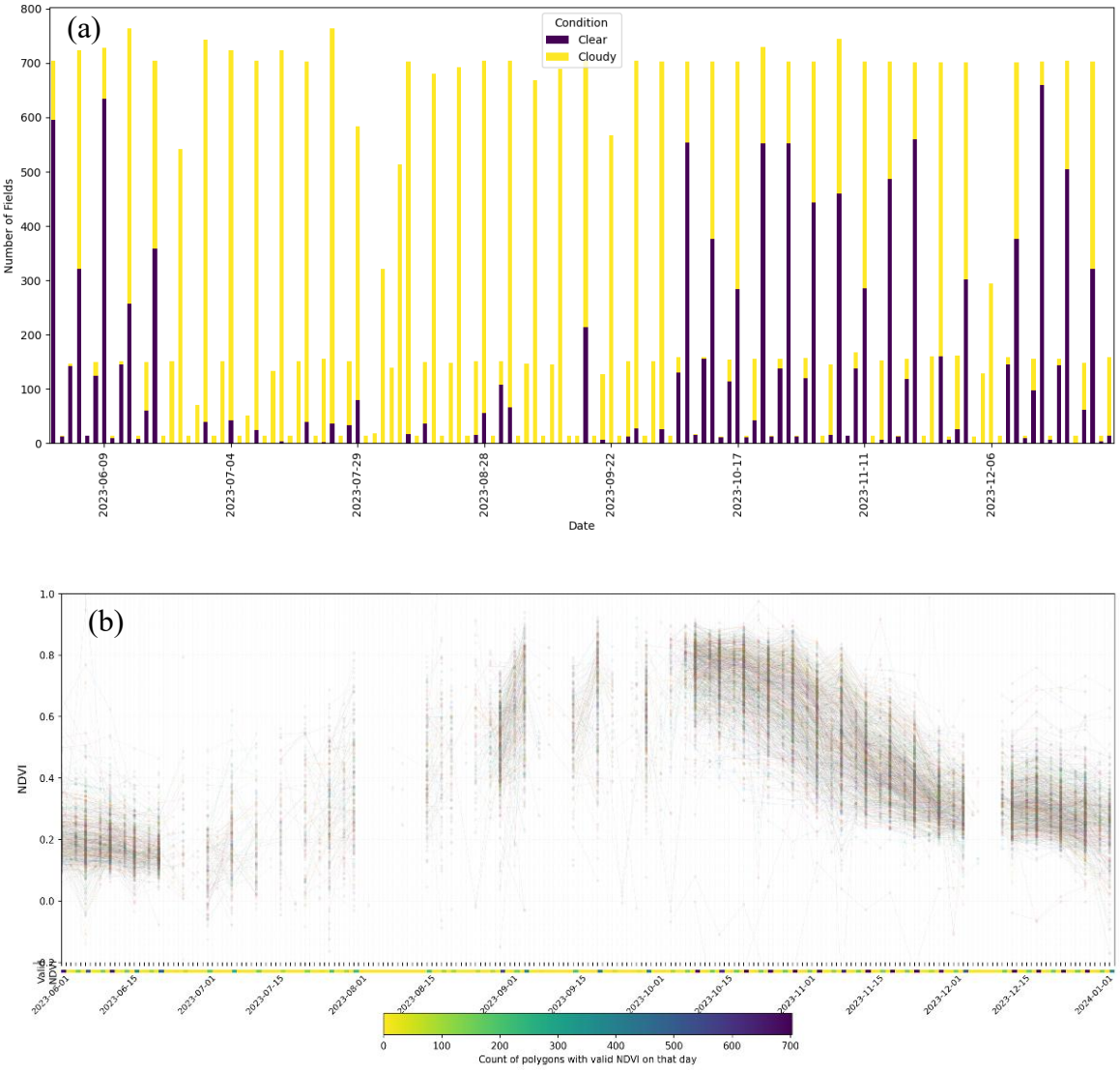
where VV_{lin} and VH_{lin}donate linear-scale backscatter. CR_{dB} and PR_{lin}were included as contrast features because ratio-type dual-polarization features highlight phenological change more clearly than either channel alone[10,12,27]. RVI was included as a vegetation-sensitive descriptor related to depolarization and volume scattering[26-28].

The multi-window design (15, 30, and 60 days) was adopted to capture SAR information at multiple temporal scales: the 15-day window reflects immediate local backscatter conditions, the 30-day window captures short-term seasonal transition dynamics, and the 60-day window integrates broader pre-gap trajectory context. This hierarchy allows the model to draw on both recent and cumulative Sentinel-1 signals without relying on explicit phenological timing variables, which were deliberately excluded to prevent shortcut learning from calendar proximity rather than from SAR-inferred vegetation state.

Benchmark

The primary evaluation was based on the Step B benchmark, a study-specific name used here to denote the official inside-gap reconstruction evaluation. In this benchmark,reconstructed NDVI was compared against real Sentinel-2 SCL = 4 observations that occurred within the cloud-gap window and on holdout test fields not seen during training. The broader cloud-gap period was defined at the field scale, where a field was classified as cloudy on a given date when at least half of its area was covered by cloudy or noisy pixels after optical quality filtering. Daily clear and cloudy field counts in Fig. 3 were then obtained by aggregating these field-level classifications through time. Based on this definition, the broader cloud-gap period extended from late June to mid-September, and the official evaluation window was Jul 1–Sep 28. This interval was chosen because it captures the core monsoon blocked-gap period while also covering the main canopy rise stage of the Kharif rice season. Full-season or global paired accuracy was deliberately not used as the primary metric, because observations outside the cloud-gap period are easier to reconstruct and would inflate apparent performance. In the present study area, July to September behaves less like routine sparse missingness and more like a prolonged optical blackout: clear Sentinel-2 observations inside this window were typically fewer than three per field, whereas reconstruction

targets every date in the interval. Restricting the official benchmark to Jul 1–Sep 28 therefore ensures that performance is assessed precisely where reconstruction is both most difficult and most



consequential forcrop monitoring.

Figure 3. Field-scale cloud distribution. (a) All-field Sentinel-2 NDVI observations through time, shown as daily field counts. (b) Daily clear versus cloudy field counts, highlighting the long contiguous observation gap during the main canopy-rise period.

Baseline estimation

A daily baseline NDVI trajectory was estimated for each field using Whittaker smoothing[6,7]. Sentinel-2 observations earlier than 2023-06-01 were excluded from the baseline fit. A single low-NDVI anchor (NDVI = 0.15, weight = 0.6) was added around 2023-06-15 to stabilize the pre-season portion of the curve. The baseline was obtained by solving the penalized least-squares problem:

$$\hat{z} = \arg\min_{z} \left\{ \sum_{t} w_t (y_t - z_t)^2 + \lambda \sum_{t} (\Delta z_t{}^2)^2 \right\}$$

where y_t is the observed Sentinel-2 NDVI, z_t is the smoothed value, w_t is the observation weight, Δ^2 is the second-difference operator, and λ is the smoothing parameter[6,18,19].

Ablation analysis

Two supplementary analyses were used to clarify model interpretation. The full residual model included Sentinel-1-based predictors as its dominant feature family alongside baseline-derived context variables, sparse Sentinel-2 helper features from outside the evaluation window, and temporal-staleness variables such as delta_days. First, a SAR-only ablation removed the temporal-staleness variables (delta_days) while retaining the Sentinel-1 backscatter features and the same overall evaluation framework, to assess whether improvement over the baseline was attributable to radar backscatter itself rather than to shortcut-like temporal proximity variables. Second, a direct-prediction experiment removed the optical baseline entirely, using the same SCL = 4 source, Step B window, and field-level split logic, to test whether Sentinel-1 features alone could reconstruct absolute NDVI without a prior seasonal scaffold.

Downstream validation

A secondary analysis evaluated whether the reconstructed trajectories retained useful timing information for sowing-date predictions. Multiple candidate approaches were tested, progressing from derivative-based start-of-season detection on smoothed curves to supervised regression using gap-period NDVI summaries, Sentinel-1 features, and baseline-derived context variables. Spearman rank correlation pre-filtering was applied to training features before model fitting, retaining only features with a stable monotonic relationship to observed sowing date. This more conservative selection step was appropriate for the downstream task because the sample size is substantially smaller than in the reconstruction step and the signal-to-noise ratio is lower, making high-dimensional feature input without pre-screening liable to overfit. No explicit calendar day-of-year variable was used as a predictor. The final downstream model used HistGradientBoosting for direct-seeded fields and ExtraTrees for transplanted fields, selected based on out-of-fold cross-validation performance.

Metrics

Main reconstruction performance was measured with the coefficient of determination (R^2) and root mean square error (RMSE) in Step B window. R^2 measures the proportion of variation in observed NDVI explained by the reconstruction, whereas RMSE measures the average magnitude of prediction error in NDVI units:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i} (y_i - \hat{y}_i)^2}$$

Downstream sowing-date prediction was additionally evaluated using mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i} |y_i - \hat{y}_i|$$

with dates represented in day-of-year units.

Results

Study sample and observation coverage

The initial field inventory contained 757 rice fields. After restricting the analysis to fields with valid monsoon-relevant sowing-timing labels, 502 fields were retained for the formal analysis, including 145 direct-seeded and 357 transplanted fields. Observed sowing dates had a mean day-of-year of 209 (July 28), a standard deviation of 16.5 days, and a range from June 4 to August 24. The official quantitative pipeline used a strict Sentinel-2 SCL = 4 vegetation filter, reducing field-level observations from 6,083 to 2,150. A richer Sentinel-2 source was retained only for visualization.

The reconstruction problem was concentrated in a long monsoon-period interruption of optical observations. Across the 502 analysis fields, clear Sentinel-2 SCL = 4 observations during Jul–Sep were typically fewer than three per field, whereas Sentinel-1 remained available at an approximately 12-day revisit interval throughout the same period. This structural asymmetry, in which the SAR record remains continuous but temporally coarse while optical observations are nearly absent during the main canopy-development phase, explains why the task is better understood as reconstruction under a blocked observation window rather than routine interpolation from a moderately sparse optical series. It also explains why inside-gap performance, rather than full-season accuracy, is the more appropriate primary benchmark (Fig. 3).

Reconstruction results

The official Step B results showed improvement over the Whittaker baseline in both establishment groups (Table 2; Fig. 4). Figure 4a summarizes the grouped RMSE comparison among the baseline, the full model, and the SAR-only ablation, whereas Figure 4b–c shows the corresponding predicted-versus-observed NDVI relationships. Here, the full model refers to the broader residual-learning setting that combines Sentinel-1 predictors with auxiliary context features, including temporal-staleness variables, whereas the SAR-only version removes the shortcut-like timing variables and retains the Sentinel-1 backscatter feature set. Together, these panels show that the reconstruction gain is not limited to a single summary metric but is also reflected in the pointwise agreement between reconstructed and observed NDVI. For direct-seeded fields, the baseline achieved $R^2 = 0.976$ with RMSE = 0.019, whereas the final reconstruction achieved $R^2 = 0.990$ with RMSE = 0.012 (∼37% reduction). For transplanted fields, the baseline achieved $R^2 = 0.881$ with RMSE = 0.032, whereas the final reconstruction achieved $R^2 = 0.961$ with RMSE = 0.018 (∼42–44% reduction). The contrast between the two groups reflects the agronomic differences addressed by the per-group modeling design. Direct-seeded fields started from a much stronger optical baseline ($R^2 = 0.976$), consistent with their more gradual and predictable early-season vegetative development.

Table 2. Main Step B results.

Group	Baseline R^2	Baseline RMSE	Full model R^2	Full model RMSE	SAR-only R^2	SAR-only RMSE
Direct-seeded	0.976	0.019	0.990	0.012	0.992	0.011
Transplanted	0.881	0.032	0.961	0.018	0.959	0.019

Transplanted fields are structurally harder to reconstruct: transplanting is associated with rapid field flooding and compressed canopy establishment, creating a steeper and shorter NDVI rise than in direct-seeded systems, and the pre-gap observations alone constrain the trajectory less tightly (baseline $R^2 = 0.881$). The larger relative improvement in the transplanted group (42–44% vs. 37% RMSE reduction) shows that the residual model contributed most where the optical scaffold was under greatest strain, and where Sentinel-1 flooding and early canopy signals were presumably most informative. This asymmetry is consistent with the agronomic contrast between the two groups: transplanted fields undergo a more abrupt early-season transition associated with flooding and transplanting that the Whittaker baseline cannot fully characterize from sparse observations, leaving more correctable signal for the Sentinel-1 residual model to exploit.

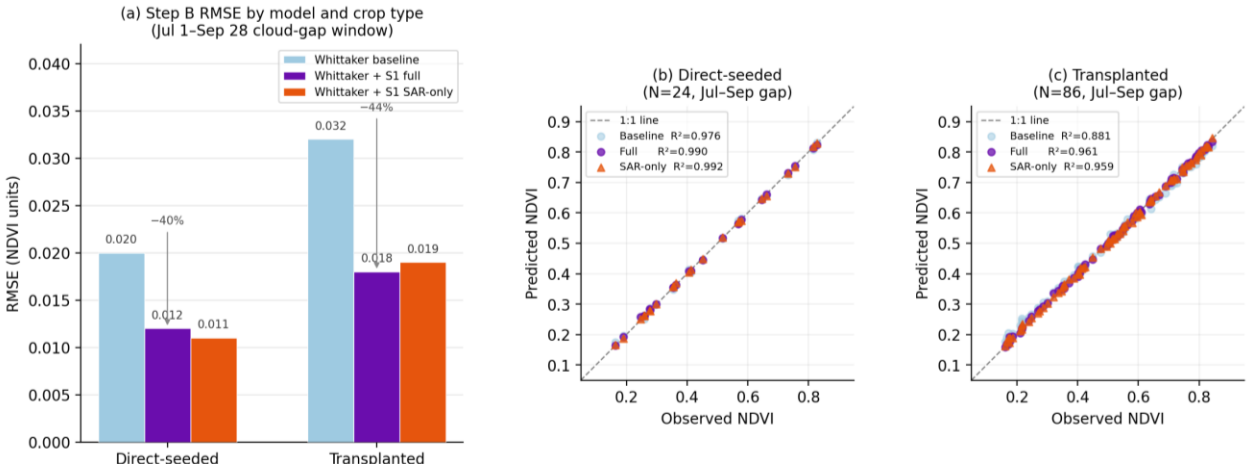


Figure 4. Step B cloud-gap reconstruction performance. (a) Grouped RMSE comparison for direct-seeded and transplanted fields. (b–c) Predicted-versus-observed NDVI relationships for each model version.

Representative field trajectories further show that the reconstructed series do more than simply improve aggregate fit statistics relative to the Whittaker baseline. In many fields, the baseline remains relatively smooth and conservative within the cloud-gap interval, reflecting its role as a stable seasonal scaffold but also its tendency to attain local dynamics. By contrast, the SAR-corrected trajectories recover modest but meaningful within-gap variability, making the canopy-rise segment more consistent with the observed trajectory on either side of the gap. This pattern suggests that Sentinel-1 contributes information that helps correct the under-responsive behavior of the baseline, rather than merely producing small numerical adjustments, while still preserving the overall full-season seasonal shape (Fig. 5).

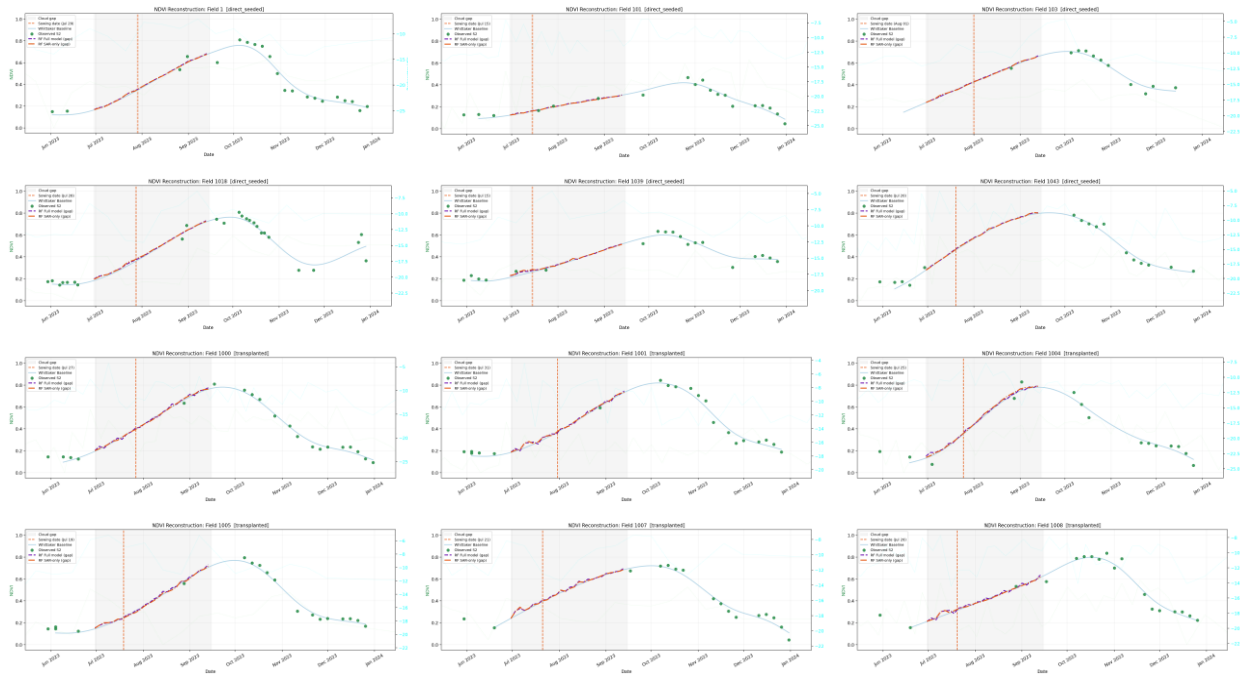


Figure 5. Representative reconstructed trajectories. Each panel shows one example field: Sentinel-2 SCL = 4 observations (circles), Whittaker baseline (dashed line), and reconstructed NDVI (solid line). The shaded region indicates the Jul–Sep cloud-gap evaluation window. Top two rows: direct-seeded fields; bottom two rows: transplanted fields.

SAR ablation

A SAR-only ablation removed temporal-staleness variables such as delta_days while preserving the same overall evaluation framework (Table 2; Fig. 4). In direct-seeded fields, the SAR-only model achieved $R^2 = 0.992$ with RMSE = 0.011. In transplanted fields, it achieved $R^2 = 0.959$ with RMSE = 0.019.

These results are close to those of the broader full model. It is worth noting that the full model included not only Sentinel-1 features but also baseline-derived context variables and temporal-staleness predictors such as delta_days; the SAR-only ablation removed these shortcut-like timing variables while retaining the Sentinel-1 backscatter features. The near-equivalence of the two settings therefore indicates that the gain over the baseline was not primarily driven by the temporal shortcut features. Instead, the ablation supports the interpretation that Sentinel-1 backscatter itself carried useful information for correcting within-gap NDVI. The direct-seeded group even showed slightly lower RMSE in the SAR-only setting, whereas the transplanted group showed near-equivalent performance.

Direct prediction without baseline

A supplementary direct-prediction experiment was conducted without the optical baseline (Table 3). To maintain consistency with the official evaluation framework and avoid leakage, the model used the same formal SCL = 4 data source, the same Step B window, and the same field-level train–test split; gap-period Sentinel-2 observations were excluded from the feature set to prevent reconstruction targets from appearing as predictors. Under these conditions, the direct-

prediction model achieved $R^2 = 0.231$ with RMSE = 0.112 for direct-seeded fields and $R^2 = 0.054$ with RMSE = 0.084 for transplanted fields.

Table 3. Supplementary performance comparisons.

Analysis	R^2	RMSE
Direct prediction without baseline, direct-seeded	0.231	0.112
Direct prediction without baseline, transplanted	0.054	0.084
Grouped sowing-date model	0.177	14.000

The magnitude of this performance drop is substantial: direct-prediction RMSE was approximately six times higher than the full reconstruction framework in direct-seeded fields and nearly five times higher in transplanted fields. Under the current data regime, Sentinel-1 features alone could not anchor the absolute NDVI level reliably, and in many fields the model regressed toward a flat, moderately elevated prediction rather than tracking the true low-to-high canopy rise through the monsoon period. This behavior is not surprising given that the model lacked any seasonal shape prior: without knowing where a field sits in its phenological trajectory, a learned mean response is often the safest prediction under high uncertainty. The optical scaffold resolves this by providing a physically grounded seasonal trajectory before SAR-based residual correction is applied.

The pattern was consistent across both establishment groups. In direct-seeded fields, removing the baseline produced RMSE = 0.112, roughly six times higher than the full reconstruction framework; in transplanted fields the gap was similarly large (RMSE = 0.084 vs. 0.018). In both cases the direct model tended to regress toward a flat, moderately elevated curve rather than tracking the actual low-to-high canopy rise through the monsoon period. This behavior indicates that under the current data regime, that sparse SCL = 4 observations, a long unobserved window, and field-level tabular features rather than image tensors, and Sentinel-1 predictors alone are not sufficient to anchor the absolute NDVI trajectory. The optical scaffold is what makes the residual-learning task tractable.

Sowing-date analysis

The reconstructed trajectories were also evaluated in a downstream sowing-date prediction task. Multiple approaches were tested before arriving at the final model. Derivative-based start-of-season methods applied to the reconstructed or baseline curves consistently failed, returning RMSE values of 23–26 days and negative R^2 in both groups. The primary reason is that Whittaker smoothing attenuates the sharp early-season green-up transitions that such methods rely on, and the cloud gap covers precisely the period where those transitions occur. A supervised machine-learning approach using gap-period NDVI and Sentinel-1 features performed better; the final grouped downstream model (HistGradientBoosting for direct-seeded fields and ExtraTrees for transplanted fields) with Spearman-based pre-filtering of training features achieved $R^2 = 0.177$, RMSE = 14.0 days, and MAE = 11.3 days, representing approximately 15% improvement over a naïve baseline that predicts the observed mean sowing date for all fields (Fig. 6). A structural constraint on this result is that the Sentinel-1 revisit interval of approximately 12 days is comparable to the natural spread of observed sowing dates (standard deviation 16.5 days), which limits the temporal resolution available for fine timing recovery regardless of model choice.

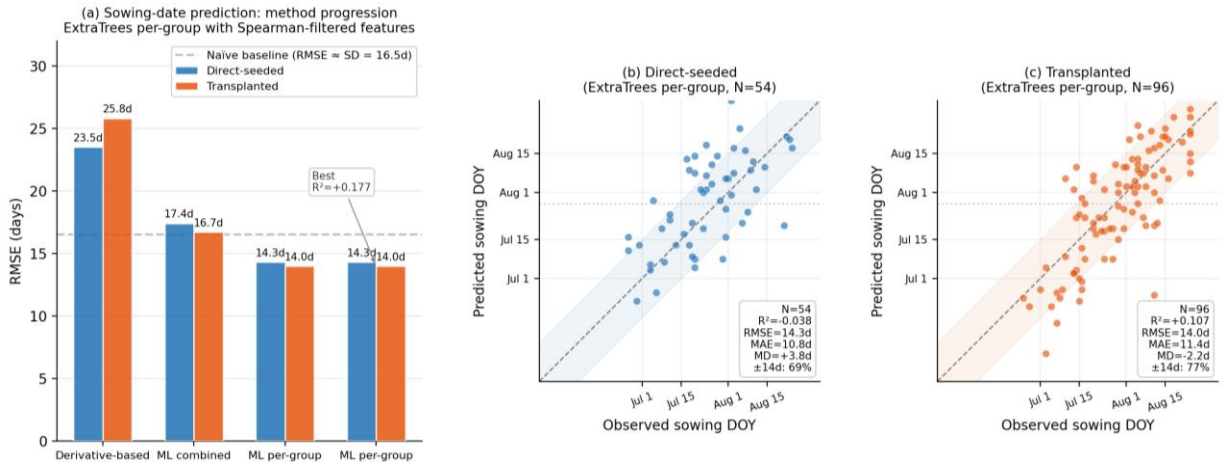


Figure 6. Downstream sowing-date validation. Predicted versus observed day-of-year for (a) direct-seeded and (b) transplanted fields under the final grouped downstream model. Panel (c) compares primary reconstruction RMSE (Step B) with sowing-date RMSE to illustrate the gap between continuity recovery and timing precision.

It is important to note that the overall $R^2 = 0.177$ reflects combined performance across both groups and should not be interpreted as equivalent within-group accuracy. Evaluated separately, transplanted fields achieved $R^2 = 0.107$, whereas direct-seeded fields showed $R^2 = -0.038$. This pattern suggests that the retained timing signal was stronger in transplanted fields, while direct-seeded sowing-date variation remained difficult to recover from the reconstructed trajectories.

Permutation importance analysis of the grouped downstream sowing-date model reveals distinct feature contributions across establishment types (Fig. 7). For direct-seeded fields, gap-period Sentinel-1 backscatter features carried the largest positive contribution, with VH and VV+VH intensity at the start of the gap window (s1_VH_dB_first, s1_VVplusVH_dB_first) accounting for the largest individual contributions (total S1 gap importance: 0.099). Reconstructed NDVI gap features (ndvi_gap_net_change, ndvi_gap_std) also contributed positively (0.033), while pre-gap S1 features showed a net negative importance (−0.139), suggesting that backscatter conditions recorded before the gap window carry little additional predictive signal once gap-period observations are included (Fig. 7). For transplanted fields, temporal coverage (s1_span_days) dominated overwhelmingly (importance ≈ 0.53), reflecting that the span of available Sentinel-1 observations within the gap period is the single strongest indicator of transplanting timing in this group. Total S1 gap importance reached 0.746, compared to 0.038 for reconstructed NDVI gap features and 0.022 for baseline/residual context. Pre-gap S1 features again showed negative importance (−0.240), consistent with the pattern observed for direct-seeded fields. These results suggest that for sowing date prediction, gap-period SAR temporal structure, particularly observation span and early-gap backscatter intensity, carries the most discriminative information for downstream timing inference, while pre-gap conditions and baseline context contribute minimally or negatively after gap-period signals are accounted for.

Permutation-importance summary for grouped feature families after removal of temporal-staleness variables. Results are shown separately for direct-seeded and transplanted fields. Gap-period Sentinel-1 features carried the largest positive contribution, especially in transplanted fields, while baseline/residual and sparse NDVI-gap features played smaller supporting roles.

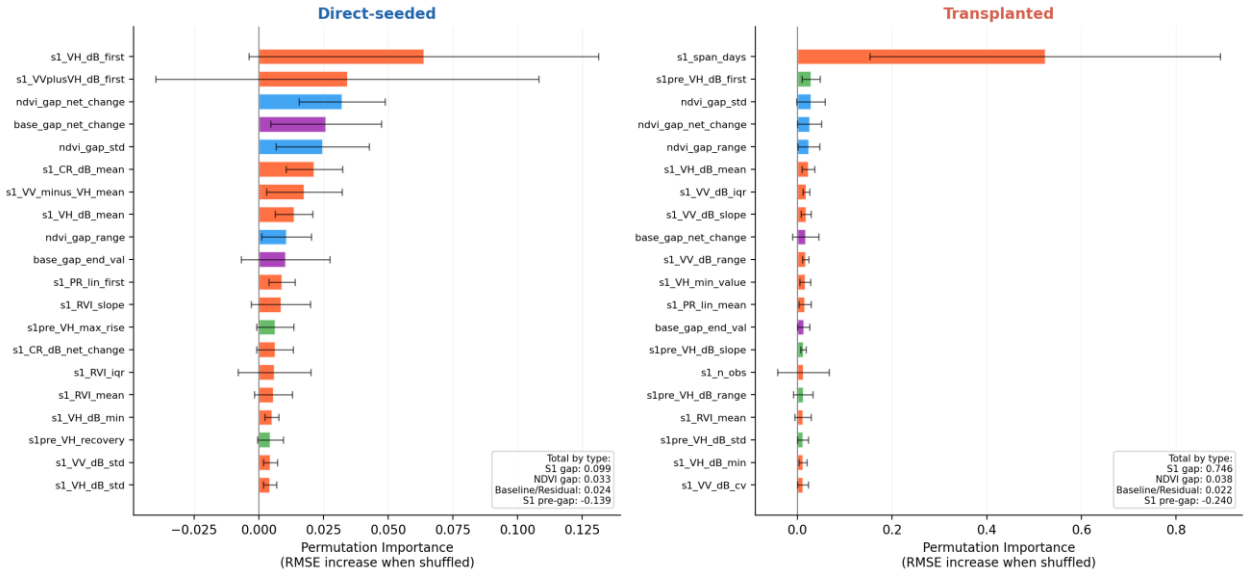


Figure 7. Feature-family contribution in the SAR-only setting.

Taken together, the downstream results indicate that the reconstructed curves retained some timing signal, but its precision is limited both by the smoothing inherent to the baseline scaffold and by the coarse temporal resolution of Sentinel-1 relative to the narrow sowing-date distribution. The gap between strong Step B reconstruction performance and modest downstream sowing-date performance reflects a fundamental distinction between improving curve continuity and recovering fine-grained timing events.

Discussion

Reconstruction stability

The central finding of this study is not that the hidden monsoon-season NDVI trajectory was fully recovered. A more defensible interpretation is that, under severe cloud-gap conditions, a baseline-guided reconstruction pathway is substantially more stable than direct absolute-NDVI prediction, while still permitting Sentinel-1 backscatter to contribute measurable within-gap improvement. This distinction matters because the reconstruction problem is fundamentally under-constrained: when the most dynamic part of the season falls within a prolonged optical blackout and labeled field-date samples are sparse, a highly flexible model has little to anchor itself to and is prone to collapse toward a flat average prediction. The supplementary no-Whittaker experiment demonstrated this directly. Providing a stable seasonal scaffold first, then correcting it with radar-based residual learning, makes the problem tractable in a way that unconstrained direct learning is not, at least under current data conditions. In that respect, the present findings align with recent gap-filling work emphasizing that reconstruction success depends strongly on gap structure, data support, and method class[5-7]. In the present implementation, Whittaker smoothing demonstrably attenuated part of the early-season phenology signal: derivative-based start-of-season methods applied to the reconstructed curves failed in both establishment groups, and the modest downstream sowing-date result (RMSE = 14.0 days) is partly a consequence of this smoothing. This trade-off is the price of stability. A smoother and more stable baseline is not necessarily a sharper one, and improved curve continuity does not guarantee preservation of all early-season transitions2,7.

Interpretation of SAR effects

The SAR-only ablation is central to interpreting the contribution of Sentinel-1. Prior studies have demonstrated direct S1-to-NDVI regression in rice systems, suggesting that C-band backscatter can function as an NDVI proxy. The present framework takes a different approach: SAR is used to correct residuals relative to an optical baseline rather than to recover the full seasonal NDVI shape from radar alone. This residual-learning formulation is more constrained, more interpretable, and as the no-Whittaker experiment confirms, is more stable under the data conditions of the present study. A persistent concern in fusion-based reconstruction is that apparent SAR contribution may be largely driven by temporal shortcut features. In particular, predictors encoding how recently the last clear observation occurred, rather than anything the radar has directly sensed about vegetation state. The ablation addresses this by removing delta_days and related temporal-staleness variables while retaining all Sentinel-1 backscatter features. The near-equivalence between the ablated and full models indicates that the improvement over the baseline was not primarily a temporal proximity artifact: radar backscatter itself carried genuine corrective information within the cloud-gap interval. This reading is consistent with prior work showing that dual-polarization C-band metrics are responsive to rice flooding, transplanting and establishment transitions, and canopy development across the growing season, without acting as simple one-to-one surrogates of optical greenness[10-12,14,26].

The grouped modeling strategy warrants explicit attention. Direct-seeded and transplanted fields differ in early-season hydrological and structural conditions, which influence both the optical NDVI trajectory and the Sentinel-1 radar response[10,14]. Per-group modeling was a key design decision during development, producing the largest single performance jump relative to a combined model: separate direct-seeded and transplanted models better capture the divergent early-season SAR-optical relationships in each system. The stronger baseline but smaller relative gain in direct-seeded fields, compared with the weaker baseline but larger relative gain in transplanted fields, is consistent with this agronomic contrast and with the known sensitivity of C-band backscatter to transplanting-associated flooding.

Limitations and future directions

Several limitations define the scope of the study. First, the official benchmark relied on the strict SCL = 4 Sentinel-2 source, which may have removed some early-season surface states relevant to rice establishment[22,23]. Broader S2 inclusion was explored earlier in model development but led to unstable baseline behavior in some fields; the decision to restrict the formal pipeline to SCL = 4 therefore reflects a deliberate trade-off that favors stability over completeness of optical support. Second, the Step B benchmark validates improvement on surviving within-gap observations rather than on every fully unobserved day, which means the evaluation is conservative relative to what a fully unobserved reconstruction target would require. Third, the framework is correctly characterized as retrospective reconstruction rather than real-time forecasting. The Whittaker baseline is fitted using all available S2 observations across the season, including those occurring after the cloud-gap window, which means the baseline already has access to the post-gap trajectory when it is used to define the training targets for Layer B. This is appropriate for the reconstruction objective but means the pipeline cannot be directly deployed in an operational near-real-time setting without modification. Fourth, the analysis covers one season and one region; performance under different cloud regimes, multi-year variability, or contrasting rice systems remains to be established. Fifth, the modest downstream sowing-date result should be understood as a joint product of smoothing-induced signal attenuation and the coarse temporal resolution of Sentinel-1

relative to the sowing-date distribution, not as evidence against the primary reconstruction contribution.

A note on the supplementary no-Whittaker experiment is also warranted. The comparison used the same formal pipeline as the main analysis, but a small number of fields were missing from the pre-saved baseline outputs, resulting in a modest reduction in the effective sample for that condition. This does not change the direction or magnitude of the finding, because direct prediction without the baseline was far worse in both groups, and means the no-Whittaker result should be treated as supporting evidence for the necessity of Layer A rather than as a directly comparable paired metric with the main results. Taken together, these limitations clarify the scope rather than undermine the central finding. The study demonstrates that a baseline-guided strategy can strengthen field-scale rice NDVI trajectories in a way that is empirically supported, comparatively stable, and more defensible than direct absolute reconstruction under the same data constraints.

Future work should test the framework across multiple years and regions to assess transferability across different cloud regimes and cropping calendars. Alternative baseline constructions that preserve more early-season phenological sharpness without sacrificing smoothing stability warrant investigation; constrained parametric models or hybrid approaches that incorporate phenological priors may offer a better trade-off than the current Whittaker smoother for timing-sensitive downstream tasks. Denser SAR acquisition, whether through multi-mission combinations or the shorter revisit intervals of future sensors, would also directly address the temporal resolution constraint identified in the downstream validation.

Conclusion

This study developed and evaluated a two-layer baseline-guided framework for reconstructing Sentinel-2 rice NDVI time series under severe monsoon cloud gaps in Chhattisgarh, India. A Whittaker-smoothed optical baseline (Layer A) provided a daily NDVI scaffold, and a residual model using past-only Sentinel-1 summaries (Layer B) corrected deviations from that baseline during the Jul–Sep gap window. Evaluated on 502 smallholder rice fields, separate models were fitted for direct-seeded and transplanted groups, which is a design choice that produced the largest single accuracy improvement during development by accounting for the distinct early-season hydrological dynamics and radar responses of the two establishment types. The framework achieved RMSE reductions of approximately 37% for direct-seeded and 42–44% for transplanted fields over the baseline alone in the Step B inside-gap benchmark. A SAR-only ablation confirmed that these gains were attributable to radar backscatter itself rather than to shortcut-like temporal proximity features. A supplementary direct-prediction experiment without the baseline collapsed toward flat average predictions with RMSE of 0.112 and 0.084, confirming that the optical scaffold is necessary for stable reconstruction under current data conditions. A downstream sowing-date analysis served as secondary validation: the reconstructed curves retained some timing signal ($R^2 = 0.177$, RMSE = 14.0 days), though the precision remained limited, reflecting the inherent tension between the smoothing required for stable reconstruction and the fine temporal resolution required for accurate phenological timing recovery. Together, these results support baseline-guided reconstruction as a stable and empirically defensible strategy for strengthening field-scale rice NDVI trajectories under conditions of fundamental observational incompleteness.

Data availability

Sentinel-1 and Sentinel-2 data used in this study are publicly available through the Copernicus program and were processed in Google Earth Engine. The release status of field boundaries, field attributes, and derived training tables depends on project-specific data-sharing permissions and should be finalized before submission. A public repository for non-restricted derived data products will be specified in the final manuscript.

Code availability

Code used for preprocessing, baseline construction, residual modeling, ablation analysis, and downstream validation will be deposited in a public repository upon manuscript finalization. Repository URL and archived release DOI will be added in the final submission.

References

1. Drusch, M. et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ*. **120**, 25–36 (2012). https://doi.org/10.1016/j.rse.2011.11.026
2. Zeng, L., Wardlow, B. D., Xiang, D., Hu, S. & Li, D. A review of vegetation phenological metrics extraction using time-series multispectral satellite data. *Remote Sens. Environ*. **237**, 111511 (2020). https://doi.org/10.1016/j.rse.2019.111511
3. Bowden, C., Foster, T. & Parkes, B. Identifying links between monsoon variability and rice production in India through machine learning. *Sci. Rep*. **13**, 2446 (2023). https://doi.org/10.1038/s41598-023-27752-8
4. Maiti, A. et al. Optimal rainfall threshold for monsoon rice production in India varies across space and time. *Commun. Earth Environ*. **5**, 302 (2024). https://doi.org/10.1038/s43247-024-01414-7
5. Grich, S., Elfarkh, J., Ouaadi, N. et al. Evaluating Sentinel-2 gap filling techniques for cloud removal and data reconstruction. *Sci. Rep*. **16**, 9464 (2026). https://doi.org/10.1038/s41598-026-39488-2
6. Eilers, P. H. C. A perfect smoother. *Anal. Chem*. **75**, 3631–3636 (2003). https://doi.org/10.1021/ac034173t
7. Atkinson, P. M. et al. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sens. Environ*. **123**, 400–417 (2012). https://doi.org/10.1016/j.rse.2012.04.001
8. Jönsson, P. & Eklundh, L. TIMESAT—a program for analyzing time-series of satellite sensor data. *Comput. Geosci*. 30, 833–845 (2004). https://doi.org/10.1016/j.cageo.2004.05.006
9. Dong, J. et al. Evolution of regional to global paddy rice mapping methods: A review. ISPRS J. Photogramm. *Remote Sens*. **119**, 214–227 (2016). https://doi.org/10.1016/j.isprsjprs.2016.05.010
10. Yang, Z. et al. An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data. *Remote Sens. Environ*. **195**, 184–201 (2017). https://doi.org/10.1016/j.rse.2017.04.016
11. Sellaperumal, P., Kaliaperumal, R., Dhanaraju, M. et al. Time series analysis of Sentinel 1A SAR data to retrieve annual rice area maps and long-term dynamics of start of season. *Sci. Rep*. **15**, 8202 (2025). https://doi.org/10.1038/s41598-025-91655-z

12. Casis, P. O., Tanael, E. N. & Rollan, A. Estimation of transplanting and harvest dates of rice crops in the Philippines using Sentinel-1 data. *Remote Sens. Appl. Soc. Environ*. **37**, 101435 (2025). https://doi.org/10.2139/ssrn.4939576
13. Roßberg, T. & Schmitt, M. A globally applicable method for NDVI estimation from Sentinel-1 SAR backscatter using a deep neural network and the SEN12TP dataset. *PFG* **91**, 171–188 (2023). https://doi.org/10.1007/s41064-023-00238-y
14. Fikriyah, V. N., Darvishzadeh, R., Laborte, A. G., Khan, N. & Nelson, A. D. Discriminating transplanted and direct seeded rice using Sentinel-1 intensity data. *Int. J. Appl. Earth Obs. Geoinf*. **76**, 143–153 (2019). https://doi.org/10.1016/j.jag.2018.11.007
15. Caballero, G. et al. Synergy of Sentinel-1 and Sentinel-2 time series for cloud-free vegetation water content mapping with multi-output Gaussian processes. *Remote Sens*. **15**, 1822 (2023). https://doi.org/10.3390/rs15071822
16. Chen, D. et al. Crop NDVI time series construction by fusing Sentinel-1, Sentinel-2, and environmental data with an ensemble-based framework. *Comput. Electron. Agric*. **215**, 108388 (2023). https://doi.org/10.1016/j.compag.2023.108388
17. Chen, Y. et al. A new deep learning-based model for reconstructing high-quality NDVI time-series data in heavily cloudy areas: fusion of Sentinel 1 and 2 data. *Int. J. Digit. Earth* **17**, 2407941 (2024). https://doi.org/10.1080/17538947.2024.2407941
18. Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst*. **35**, 507–520 (2022).
19. Safonova, A. et al. Ten deep learning techniques to address small data problems with remote sensing. *Int. J. Appl. Earth Obs. Geoinf*. **125**, 103569 (2023). https://doi.org/10.1016/j.jag.2023.103569
20. Gorelick, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ*. **202**, 18–27 (2017). https://doi.org/10.1016/j.rse.2017.06.031
21. Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*. **351**, 309 (1974).
22. European Space Agency. *Sentinel-2 Level-2A Algorithm Theoretical Basis Document*, issue **2**.10 (2021).
23. Hollstein, A., Segl, K., Guanter, L., Brell, M. & Enesco, M. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. *Remote Sens*. **8**, 666 (2016). https://doi.org/10.3390/rs8080666
24. European Space Agency. Copernicus Sentinel-2 Collection 1 MSI Level-2A Product Specification (2021).
25. Torres, R. et al. GMES Sentinel-1 mission. *Remote Sens. Environ*. **120**, 9–24 (2012). https://doi.org/10.1016/j.rse.2011.05.028
26. Filgueiras, R. et al. Crop NDVI monitoring based on Sentinel 1. *Remote Sens*. **11**, 1441 (2019). https://doi.org/10.3390/rs11121441
27. Mandal, D. et al. Dual polarimetric radar vegetation index for crop growth monitoring using Sentinel-1 SAR data. *Remote Sens. Environ*. **247**, 111954 (2020). https://doi.org/10.1016/j.rse.2020.111954
28. Zhang, X. et al. Dry biomass estimation of paddy rice with Sentinel-1A satellite data using machine learning regression algorithms. *Comput. Electron. Agric*. **176**, 105674 (2020). https://doi.org/10.1016/j.compag.2020.105674

Acknowledgements

This work was conducted as part of an M.S. thesis at the University of Michigan. We thank Meha Jain, Victor Prudente, Monish Desphande and collaborators for discussions on experimental design, benchmarking strategy, and interpretation. Funding details should be added in the final manuscript if required.

Author contributions

Z.H. conceived the study, processed the data, designed and implemented the workflow, conducted the experiments, and wrote the manuscript draft. Meha Jain, Victor Prudent, and Monish Desphande contributed to study design, supervision, interpretation, and manuscript revision. Hantao Wang processed the data. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Information

Supplementary Table S1. Feature-engineering design for the reconstruction model.

Feature family	Definition	Example variables
Base SAR variables	Field-level Sentinel-1 observables extracted for each acquisition date	VV_dB, VH_dB, CR_dB, PR_lin, RVI, VVplusVH_dB
Multi-window summaries	Past-only summaries computed over 15-, 30-, and 60-day windows	w15_*, w30_*, w60_*
Central tendency	Summary of within-window level	mean, median
Dispersion	Summary of within-window spread	std, iqr, cv
Boundary values	Extremes and edge values inside each window	min, max, first, last
Trend descriptors	Simple temporal behavior within a window	slope
Coverage descriptors	Observation support and temporal coverage	n_obs, span_days
Recency descriptors	Distance from last available support to target date	delta_days (removed in SAR-only ablation)
Cross-band descriptors	Features derived from relationships among VV, VH, and related series	VVplusVH, CR-derived statistics, RVI-derived statistics
Design-matrix size	Number of predictors before selection	255 in the full setting; 195–252 in ablation-specific settings
Feature selection	Final retained predictors before model fitting	SelectKBest, k = 60

Supplementary Table S2. Key fixed modeling settings.

Component	Setting
Sentinel-2 training/evaluation source	results_individual Sentinel-2 table, SCL = 4 only
Sentinel-2 visualization source	richer gee_metrics Sentinel-2 table, used for plotting only
Cloud-gap focus	Jun 30–Sep 15
Official evaluation window	Jul 1–Sep 28
Baseline method	Whittaker smoother
Adaptive lambda	100 (≤2 obs), 300 (≤5), 800 (≤10), 2000 (>10)
Pre-season anchor	2023-06-15, NDVI = 0.15, weight = 0.6
SAR base variables	VV_dB, VH_dB, CR_dB, PR_lin, RVI, VVplusVH_dB
Look-back windows	15, 30, and 60 days
Feature selection	SelectKBest, k = 60
Hyperparameter search iterations	80
Leakage control	GroupKFold by field; past-only Sentinel-1 windows
Grouping strategy	Separate direct-seeded and transplanted models
SAR-only ablation	exclude_delta_days = True

Supplementary Table S3. Model configuration and search space.

Component	Setting
Reconstruction learners compared	RF, XGBoost, Ridge, ExtraTrees
Final reconstruction learner	ExtraTreesRegressor
Downstream learners compared	XGBoost, LightGBM, CatBoost, ExtraTrees, HistGradientBoosting, stacking
Final direct-seeded downstream learner	HistGradientBoosting
Final transplanted downstream learner	ExtraTrees
Hyperparameter search	RandomizedSearchCV
Search iterations	80
n_estimators search range	100–800
max_features candidates	sqrt, log2, 0.3, 0.5, 0.7
min_samples_leaf candidates	1, 2, 5, 10
CV design	GroupKFold, k = 5
Train/test split	70/30 at field level
Phenology leakage control	No DOY, no future observations

Supplementary Table S4. Key feature-family findings.

Analysis	Main finding
SAR-only Step B ablation	Performance remained similar after removal of delta_days, supporting genuine SAR contribution
Full-model importance	w15_delta_days dominated importance before ablation
SAR-only grouped importance	Gap-period Sentinel-1 features carried the largest positive contribution, especially in transplanted fields
Downstream grouped modeling	Separate direct-seeded and transplanted models produced the largest single improvement
Downstream timing limit	Sentinel-1 revisit (~12 d) was close to sowing-date spread (16.5 d), constraining fine timing recovery