

Article

Support-Aware Dual-SWIR Downscaling from Sentinel-2 and Submetre Jilin-1 Imagery on Non-Nested Grids

Zixu Wang¹, Jiali Li¹, Hui Zhan¹, Hao Feng^{1,2}, Yin Ouyang¹ and Peng Guo^{1,*}

¹ College of Urban and Environmental Sciences, Shihezi University, 221 North 4th Road, Shihezi 832003, China; 20211018242@stu.shzu.edu.cn (Z.W.); lijiali1@stu.shzu.edu.cn (J.L.); zhanhui@stu.shzu.edu.cn (H.Z.); fenghao@stu.shzu.edu.cn (H.F.); ouyangyin@shzu.edu.cn (Y.O.)

² Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

* Correspondence: guopeng@shzu.edu.cn

Highlights

What are the main findings?

- Exact area overlap provides the main accuracy gain on non-nested grids.
- Continuous correction yields a smaller additional improvement.
- The geometry advantage persists under blur, registration offsets and scale ratios.

What are the implications of the main findings?

- The method outperforms two adapted baselines under a matched observation budget.
- Real 0.75 m products satisfy coarse-observation consistency.

Abstract

Fine-scale shortwave-infrared (SWIR) imagery can support detailed land-surface mapping, but submetre satellites generally acquire only blue, green, red and near-infrared (RGBN) bands. Combining these data with Sentinel-2 SWIR requires an observation model that accounts for non-nested pixel supports. We developed a support-aware inverse-allocation framework using an RGBN-conditioned prior, exact polygon-overlap weights and a continuous correction field. The framework was evaluated in four regions of Xinjiang, China, using retained 20 m B11/B12 observations as reference. In a strictly matched control, replacing centre ownership with exact area overlap reduced mean joint root-mean-square error (RMSE) by 0.00174–0.00217 in reflectance units. The reduction occurred in all 15 non-nested phases and all 16 spatial blocks of each region. Continuous correction provided a further reduction of 0.000066–0.000166. Under a matched single-observation budget, the method achieved lower mean joint RMSE than adapted area-to-point regression kriging and DSen2 baselines in all four regions. Exact area overlap retained its advantage under controlled blur, ± 10 m guide-image offsets and scale ratios of 2, 3 and 6, although the advantage narrowed at the coarsest ratio. Full-scene application to 0.75 m Jilin-1 RGBN and 20 m Sentinel-2 SWIR produced observation-consistent conditional SWIR allocations, with accuracy evaluated at the controlled 20 m reference resolution.

Keywords: area overlap; cross-sensor fusion; Jilin-1; non-nested grids; observation-constrained inverse allocation; Sentinel-2; shortwave infrared



Academic Editor: Zhenming Peng

Received: 1 September 2026

Revised: 17 October 2026

Accepted: 19 October 2026

Published: 20 October 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

1.1. Motivation and Application Context

Shortwave infrared (SWIR) observations are informative for vegetation water status, soil condition, mineral composition and artificial materials, and are therefore important to agricultural monitoring and detailed land-surface mapping. Sentinel-2 and Landsat provide stable SWIR measurements, but their 20–30 m resolutions mix field boundaries, narrow roads and fragmented cover. Submetre satellites such as Jilin-1 (JL1) provide much finer blue, green, red and near-infrared (RGBN) observations but generally carry no SWIR band at the same scale, so extending their spectral dimensionality without losing spatial resolution is of practical interest [1,2].

Fine-scale SWIR estimation from RGBN involves spectral ambiguity, an underdetermined scale transformation and differences in pixel support. Materials with similar RGBN responses can have different SWIR responses, limiting reconstruction from broadband measurements [3–5]. Each coarse SWIR observation also integrates many fine-scale values. Georegistration establishes their coordinate relationship but does not ensure coincident pixel boundaries or an integer scale ratio [6]. Regular-window averaging and centre ownership can therefore misrepresent the observation relationship when the grids differ in phase.

1.2. Related Work and the Unresolved Gap

Four families of methods are relevant, and recent reviews map the wider hyperspectral–multispectral fusion literature they belong to [7,8].

Model-driven sharpening exploits structures shared by high-resolution auxiliary bands and low-resolution targets, restricting the solution set through subspaces, geometric degradation models or self-similarity. Representative examples are band-independent geometry, self-similarity priors, reduced-rank sharpening and the area-to-point regression kriging family [6,9–13]. These methods emphasise the observation model or the internal scene structure, but they are generally developed under predefined scale relationships or regular downsampling.

In learning-based Sentinel-2 sharpening, residual networks and consistency constraints recover coarser-resolution bands. DSen2 established an influential residual-learning baseline [14] and SEN2SR constrains low-frequency radiometric and spatial consistency in the super-resolved output [15]. Reduced-rank learning and systematic benchmarks show that network capacity does not remove the effect of degradation protocols and scene distributions on method rankings [16,17], and cross-sensor learning has also been applied to Sentinel-2–Landsat fusion [18].

High-resolution visible and near-infrared data have been used to enhance coarser Sentinel-2 bands for water detection and agricultural analysis [1,2,19,20], which shows that such data carry useful spatial information for representing coarse SWIR at a finer scale. When two georegistered images have non-nested pixel boundaries, however, the area-support relationship between the coarse SWIR observations and the fine variables must still be specified explicitly: a fine pixel crossing a coarse-pixel boundary contributes to several neighbouring observations and cannot be represented faithfully by simple centre ownership.

Spatial statistics has long treated transformations among supports as change-of-support problems. Area-to-point regression kriging, area-to-point cokriging and mass-preserving interpolation connect coarse observations to fine variables through areal integration or regional-total constraints [9,13,21], and geographically weighted area-to-point methods let the relationship vary in space [22]; together they show that scale transformation should not rely on pixel centres or integer ratios alone. Observation consistency, in which a fine-scale result must reproduce the coarse observation after a prescribed aggregation, is used in Sentinel-2 super-resolution and model-driven multisource fusion [15,18]. Continuous or implicit representations express the target as a coordinate-dependent

function, supporting reconstruction or queries at different scales: OTIAS, implicit transformers, FeINFN, arbitrary-resolution self-supervised representations and INF3 extend hyperspectral–multispectral fusion through sampling, implicit features or continuous-scale modelling [23–27], and generalised inverse optimisation has been applied to semi-blind fusion [28]. Methods for unregistered or residually misaligned inputs, such as IR&ArF and QRCODE, estimate, avoid or absorb unknown geometric relationships [29,30].

When independently positioned grids are georeferenced but not nested, two effects appear together and existing evaluations do not separate them: centre-based ownership replaces the true area–support relationship, and schemes that correct each coarse cell separately can assign conflicting values to fine pixels that cross cell boundaries. Non-nested work has concentrated on unknown registration [29,30] or on regular downsampling with a predefined ratio [16,17], and change-of-support studies have established the principle of areal integration without quantifying how much error the centre-ownership approximation introduces on controlled non-nested grids for which fine-scale truth is retained. The contribution of a continuous correction basis is likewise usually reported together with the geometry change, so its separate size is unknown. The evidence below addresses both points.

1.3. Contributions and Scope

We formulate dual-SWIR synthesis as observation-constrained inverse allocation on non-nested grids. The study examines observation geometry separately from the correction representation.

1. **Exact support geometry in the observation model.** Exact polygon overlaps encode the native coarse-to-fine relationship without imposing an integer scale ratio or a common grid origin. Section 4.3 reports a strictly matched control in which the prior, the lifting basis, the solver and the constraints are identical between the two configurations and only the observation matrix changes, which isolates the geometry effect from the correction parameterisation.
2. **Separation of geometry from later processing.** Section 4.3 separates observation geometry from variable organisation in the same controlled benchmark and shows that the additional gain attributable to the continuous correction subspace is an order of magnitude smaller than the geometry effect.
3. **Controlled robustness and deployment boundaries.** Controlled tests of blur, residual registration offset and scale ratio surround the benchmark, and full-scene deployment demonstrates operation on real 0.75 m JL1 and 20 m Sentinel-2 grids. Section 5.4 states what these evidence layers can and cannot establish about the deployed product.

The framework builds on areal weighting [9,21], shared fine-pixel variables and coarse-observation consistency [15,18]. Its contribution is to quantify the relative effects of support geometry and continuous correction on grids with a known coordinate relationship.

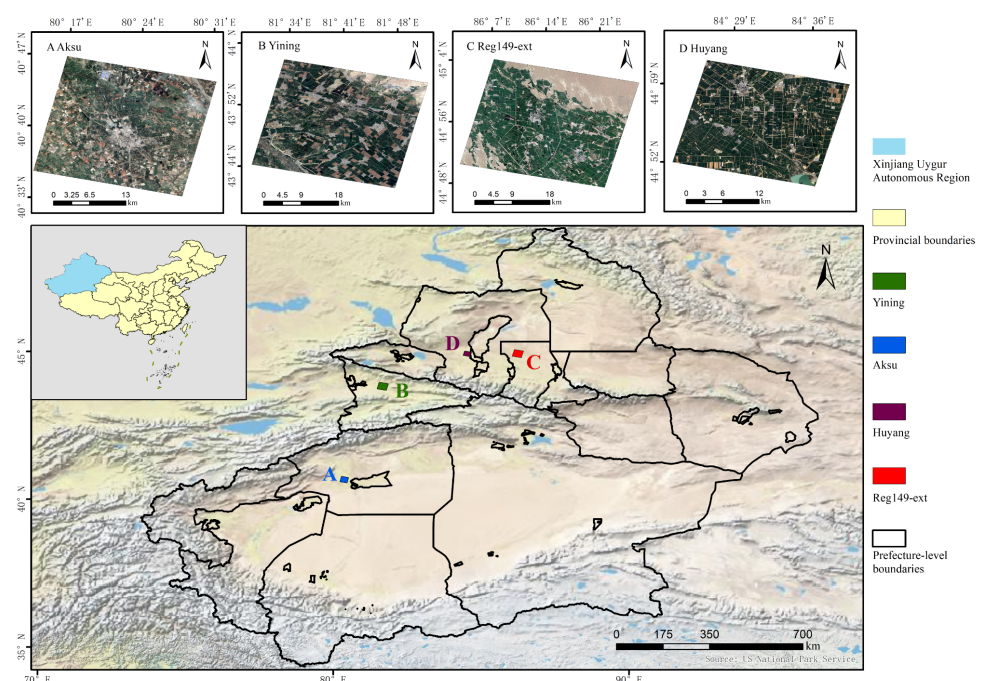
2. Study Areas, Data, and Preprocessing

2.1. Study Areas and Cross-Sensor Acquisitions

The study covers Aksu, Yining, an extended area around the 149th Regiment (Reg149-ext) and Huyang River in Xinjiang, China. All four contain oasis agriculture, urban land or their interfaces, but the JL1 payloads and acquisition times differ: GF02B/PMS1 for Aksu, KF01A/PMS01 for Yining and Reg149-ext, and GF03D27/PMS for Huyang River. JL1 and Sentinel-2 were acquired on the same day for Aksu and Yining and about one day apart for the other two, so geography, date, payload and land-surface composition co-vary among the external regions; the external results characterise transfer performance but cannot attribute differences to any one factor (Section 5.5). Table 1 summarises the acquisitions and the experimental roles, and Figure 1 shows the study locations.

Table 1. Payloads, acquisition dates and experimental roles in four regions.

Region	JL1 Payload	JL1 Acquisition (UTC+8)	S2 Date	Δt (d)	JL1 RGBN (m)	Role and Target Bands
Aksu	GF02B/PMS1	2023-07-03 12:37	2023-07-03	0	0.75	development (source); B2,B3,B4,B8 → B11,B12
Yining	KF01A/PMS01	2023-07-03 13:08	2023-07-03	0	0.75	external target; same bands
Reg149-ext	KF01A/PMS01	2023-07-13 12:53	2023-07-14	1	0.75	near-date target; same bands
Huyang River	GF03D27/PMS	2023-07-24 13:08	2023-07-25	1	0.75	cross-payload target; same bands

**Figure 1.** Study regions and Jilin-1 image coverage in Xinjiang, China. Insets A–D show Aksu, Yining, Reg149-ext and Huyang River; colours identify their regional footprints.

2.2. Radiometric and Geometric Preprocessing

Sentinel-2 Level-2A products provided B2, B3, B4 and B8 as the blue, green, red and near-infrared inputs and B11 and B12 as the SWIR-1 and SWIR-2 targets; JL1 Level-3C pan-sharpened RGBN products had a nominal resolution of about 0.75 m.

The archived Sentinel-2 and JL1 integer arrays use a 10^4 reporting scale. The prior, local ridge regressions, observation operators and inverse-allocation solve operate on these scaled values. Thus, $\lambda = 1$ in Equation (8) applies to scaled input. The adapted DSen2 network uses the separate checkpoint normalisation constant 65536; its outputs are converted back to the shared scale before evaluation. Unless specified otherwise, errors are divided by 10^4 and reported in reflectance units. This describes the numerical processing chain. The archived product metadata are insufficient to independently verify the conversion to physical surface reflectance (Section 5.6).

JL1 Level-3C RGBN values were used as delivered by the provider, without independent absolute calibration or cross-calibration to Sentinel-2. The prior uses RGBN deviations

from a local mean (Equation (12)), so a common additive offset cancels in the deviation term. Gain differences and band-dependent offsets remain relevant, and the full prior also depends on the retrieval state through the coarse anchor. Field or pseudo-invariant calibration targets were unavailable.

All images were transformed to a common projected coordinate system and their common valid support was identified. JL1 pixels were used on their native grid, without an integer-nested Sentinel-2 subgrid: true pixel polygons were generated from the original affine transforms, and coarse-to-fine relationships were calculated from polygon-overlap areas. The coarse grid is defined at 60 m, three 20 m reference cells on a side, and the virtual non-nested conditions are produced by shifting this coarse grid in phase, which preserves the known geospatial relationship while letting the pixel boundaries differ. Virtual coarse observations are formed by area integration of the native 20 m SWIR reference over the shifted cells. The source array and resampling history of the controlled 20 m RGBN guide were not retained.

Invalid and nonfinite values and quality masks were processed before sample construction, incomplete edge supports and complete interior supports were recorded separately, and the coverage fraction c_j of Equation (4) was retained for every coarse observation. The controlled benchmark evaluates only fine pixels inside a fixed interior mask obtained by eroding the valid region by two 20 m cells, so every evaluated pixel is fully covered in every phase. The common-scale experiment evaluates the child pixels of parent blocks whose 60 m supports are fully valid.

Provider affine transforms established the cross-sensor coordinate relationship, which was refined by phase correlation over the overlap. Applied translations ranged from about 0.9 to 1.7 m in columns and 0.4 to 1.4 m in rows. Control-point deviations about the translation had median absolute values of 0.94–1.52 m and root-mean-square values of 1.31–1.91 m. These are descriptive dispersions of the fitted correspondences. Independent registration accuracy was not measured because no disjoint check-point set was available. The ± 10 m offsets in Section 4.4 test a larger displacement rather than estimate operational misregistration.

2.3. Conditional Prior, Data Flow and Information Isolation

The RGBN-conditioned prior combines the current coarse SWIR anchor with fine-scale auxiliary-band variation. Figure 2 and Supplementary Table S8.7 distinguish two observation budgets. The *original budget* includes an additional nested coarse SWIR image derived from the native 20 m data. Its anchor is the mean of the nine native 20 m cells inside each fixed 60 m parent block, and the prior is phase independent. The *single-observation budget* uses only the current-phase coarse observation. Its anchor is the area-weighted mean of coarse cells overlapping each parent block, with nearest-cell fallback for empty overlap, so the anchor and prior vary with phase. Both budgets retain the source library, RGBN standardisation and fixed-parent-grid lifting operator L . The correction basis B is constructed for the current coarse grid and shared within each matched phase.

The response library was fitted in Aksu. Each source block r stores an RGBN state vector, a local response matrix J_r (Equation (8)) and a quality weight q_r (Equation (10)). Spatial splits use the normalised source-block column index q : development $q < 0.52$, selection $0.57 \leq q < 0.73$ and internal confirmation $0.78 \leq q < 0.99$, separated by buffer bands. Selection used the detail-skill criterion $1 - \sum dt / \sum m_0$. The configuration was fixed before evaluation of the external regions. Applying the frozen split rule to the retained per-block index separates the 172,873 Aksu source blocks into 88,669 development blocks ($q < 0.52$), 27,977 selection blocks ($0.57 \leq q < 0.73$) and 37,274 internal confirmation blocks ($0.78 \leq q < 0.99$), with buffer bands between them. The retained response library holds 88,669 entries, the same count as the development part. The library stores no block

identifier, so its entries are not matched to the retained index one by one, and a partition count is recovered from the rule and that index rather than from an archived membership list.

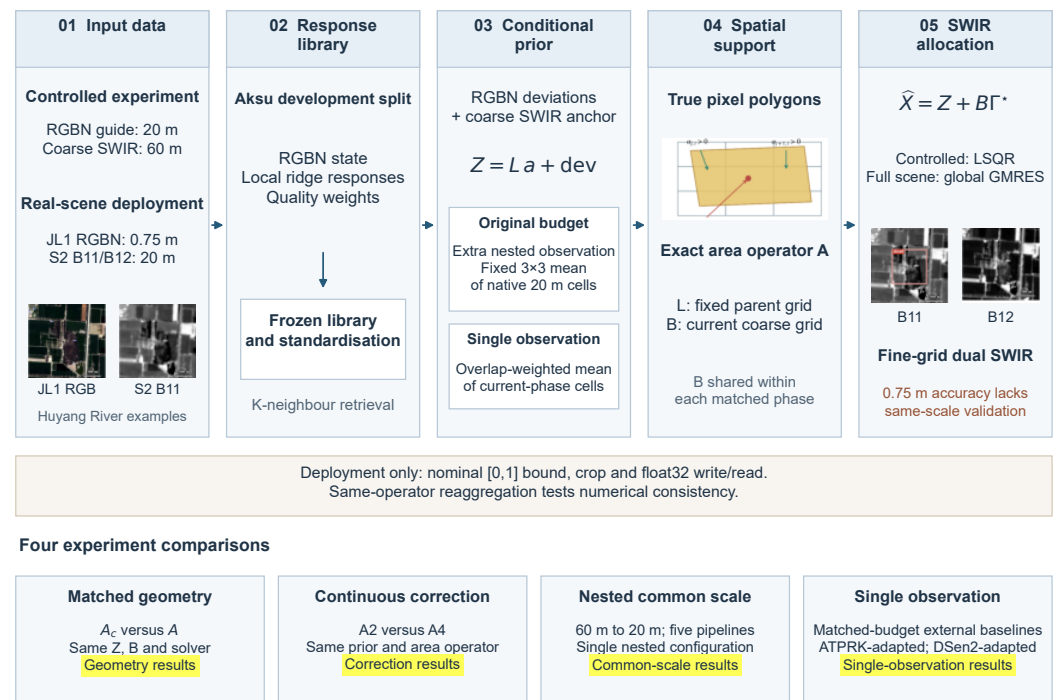


Figure 2. Five-stage workflow and four experiment comparisons. Controlled experiments use a 20 m RGBN guide and virtual 60 m SWIR observations; the guide’s source array and resampling are not retained. The original-budget anchor is the phase-independent mean of nine native 20 m cells within a fixed 60 m parent block. The single-observation anchor uses current-phase overlap weights, with nearest-cell fallback for empty overlap. L acts on the fixed parent grid; B follows the current coarse grid and is shared within a matched phase. Target-region fine-scale SWIR labels remain outside prediction. Controlled outputs are scored without bounds or post-processing. Only deployment adds a nominal [0, 1] bound, crop and float32 write/read closure. The superscript star in Γ^* denotes the optimised correction coefficients. Scene thumbnails show Huyang River inputs and outputs; the schematic illustrates centre ownership and exact area overlap. Same-operator closure does not validate 0.75 m accuracy.

Native 20 m B11/B12 data generate the controlled coarse observations before prediction and provide reference labels after prediction. Aksu development and selection pairs additionally support library fitting and configuration selection. In deployment, the coarse input is native Sentinel-2 B11/B12 at 20 m; in the controlled single-observation experiment, it is the phase-specific non-nested 60 m observation. Target-region fine-scale labels are excluded from prior construction, the solve, parameter selection and fill-in. Thus, controlled coarse inputs are derived from the reference, while reference labels remain outside prediction. Internal matched configurations share a prior within each phase; external baselines use their own predictors.

2.4. Validation Hierarchy

The evidence chain comprises three levels that use different evaluation masks, different observation budgets and different reference quantities, and metrics from different levels are never averaged together. The first is a regular 60-to-20 m common-scale experiment, in which native 20 m B11/B12 serve as fine-scale truth and nine 20 m pixels form one 60 m parent block; because the grids are strictly nested it cannot isolate area geometry, but it tests whether the conditional prior and the complete pipelines recover meaningful fine-scale

SWIR variation under a common scale relationship. The second is a controlled non-nested benchmark, in which native 20 m SWIR remains the truth while a virtual 60 m grid is shifted along both axes by the prespecified phases

$$d_x, d_y \in \{0, 5, 10, 15\} \text{ m.} \quad (1)$$

This produces 16 phases; the (0,0) phase is the nested control and the other 15 are non-nested. Every virtual coarse observation is generated from the 20 m truth by the true-area matrix, so the observation operator used to generate the data and the operator used in reconstruction are the same object at the reference configuration. This is an idealised condition, and Sections 4.4 and 5.6 quantify what happens when it is relaxed. Within this level the interior mask, the conditional prior and the phase set are fixed, and only the observation geometry and the variable organisation change. Grid phases are deterministic sensitivity conditions, not independent statistical replicates.

The third level deploys the method on real cross-sensor grids using native 0.75 m JL1 RGBN and 20 m Sentinel-2 SWIR, without rebuilding an integer-nested grid. No 0.75 m SWIR reference exists for these scenes, so this level evaluates full-scene operation, output validity and closure to the observed 20 m SWIR. To prevent regional information leakage, method design, library construction and hyperparameter selection were completed in Aksu with spatial buffers separating the development, selection and confirmation samples; the three external regions were evaluated only after all parameters had been frozen. Their native 20 m SWIR was used to generate the controlled coarse observations and to provide the evaluation labels, while the target-region fine-scale labels were withheld from prediction, parameter selection and fill-in.

3. Methods

3.1. Problem Formulation

Let $X \in \mathbb{R}^{n \times 2}$ denote the target fine-scale dual-SWIR field, $Y \in \mathbb{R}^{m \times 2}$ the coarse dual-SWIR observations and $Z \in \mathbb{R}^{n \times 2}$ the conditional prior built from high-resolution RGBN; the complete workflow is summarised in Figure 2. Considering pixel support alone, the coarse-to-fine observation relationship is

$$Y = AX, \quad (2)$$

where $A \in \mathbb{R}^{m \times n}$ is a sparse matrix of support weights. Typically $m \ll n$, so the coarse observations cannot uniquely determine the fine-scale field, and the RGBN-to-SWIR mapping is nonunique as well. An RGBN regression output is therefore not treated as final SWIR: RGBN generates the conditional prior Z , whose spatial allocation is corrected through A using the actual observations Y . The prior supplies RGBN-conditioned fine-scale structure, while A specifies which fine pixels contribute to each coarse observation and with what weights.

3.2. Exact Area-Overlap Observation Operator

In the observation model of Section 3.1, the sparse matrix $A \in \mathbb{R}^{m \times n}$ links the coarse dual-SWIR observations to the fine dual-SWIR field. To construct this matrix on a non-nested grid, let $\mathcal{P}_j$ be the spatial support of coarse pixel j and $\mathcal{Q}_i$ the support of fine pixel i , where $j = 1, \dots, m$ and $i = 1, \dots, n$. Their true overlap area is

$$\alpha_{ji} = \text{area}(\mathcal{P}_j \cap \mathcal{Q}_i), \quad (3)$$

where $\text{area}(\cdot)$ gives the area of a spatial support. If $\alpha_{ji} = 0$, fine pixel i does not contribute to coarse observation j . With $\mathcal{V} \subseteq \{1, \dots, n\}$ the index set of currently valid fine pixels, the valid coverage of coarse pixel j is

$$c_j = \frac{\sum_{i \in \mathcal{V}} \alpha_{ji}}{\text{area}(\mathcal{P}_j)}, \quad (4)$$

where $c_j \in [0, 1]$ is the fraction of the support of coarse pixel j covered by valid fine pixels, and constraints are built only for coarse pixels that pass this criterion. For those pixels the area weights are normalised over the valid support, yielding

$$A_{ji} = \begin{cases} \frac{\alpha_{ji}}{\sum_{k \in \mathcal{V}} \alpha_{jk}}, & i \in \mathcal{V}, \\ 0, & i \notin \mathcal{V}. \end{cases} \quad (5)$$

The entry A_{ji} is therefore the normalised area contribution of fine pixel i to coarse pixel j , and reaggregation of the fine field gives the aggregated coarse response; in that response, $X_{i,\cdot}$ holds B11 and B12 of fine pixel i . The definition requires neither an integer scale ratio nor coincident grid origins or phases.

On a non-nested grid, one fine pixel may overlap several coarse pixels, and we quantify this sharing by the number of coarse observation constraints involving fine pixel i :

$$n_{\text{support},i} = |\{j \mid A_{ji} > 0\}|, \quad (6)$$

where $|\cdot|$ denotes the set cardinality. A fine pixel is called *shared* if $n_{\text{support},i} \geq 2$ and *interior-owned* if $n_{\text{support},i} = 1$.

Centre ownership assigns a fine pixel exclusively to the coarse cell containing its centre and ignores the portion crossing a coarse boundary. Its matrix $A_c \in \mathbb{R}^{m \times n}$ has one nonzero entry per covered fine pixel after row normalisation. The entire difference between the two geometries is the choice between A and A_c : both are sparse, both are row-normalised, and in the matched control both are applied to the same prior with the same correction machinery. Matrix A is the geometric component of the observation model; optical effects such as the point-spread function, the modulation transfer function and spatially varying blur are treated separately in Sections 4.4 and 5.6.

Controlled ablations and matched controls are evaluated as unconstrained least-squares solutions. In deployment, a nominal $[0, 1]$ bound is applied after the solve to the written product. The closure test in Section 4.5 uses this stored product.

3.3. RGBN-Conditioned Local Prior

The observation matrix A and the coarse dual-SWIR observations Y do not uniquely determine the fine-scale field. High-resolution RGBN information is therefore used to construct the conditional prior Z . To distinguish source-region training samples from the fine-scale field X and the coarse observations Y , G_r and S_r denote the RGBN and dual-SWIR samples in source block r .

For source block r , $G_r \in \mathbb{R}^{N_r \times 4}$ and $S_r \in \mathbb{R}^{N_r \times 2}$ are the RGBN and dual-SWIR sample matrices over the N_r valid samples used for estimation, with mean vectors $\bar{g}_r$ and $\bar{s}_r$. Removing the local means gives the centered variations

$$\Delta G_r = G_r - \mathbf{1}_{N_r} \bar{g}_r^T, \quad \Delta S_r = S_r - \mathbf{1}_{N_r} \bar{s}_r^T, \quad (7)$$

where $\mathbf{1}_{N_r} \in \mathbb{R}^{N_r}$ is an all-ones column vector. The local response matrix $J_r \in \mathbb{R}^{4 \times 2}$ is then estimated for source block r by ridge regression without an intercept:

$$J_r = \arg \min_{J \in \mathbb{R}^{4 \times 2}} \|\Delta S_r - \Delta G_r J\|_F^2 + \lambda \|J\|_F^2, \quad \lambda = 1, \quad (8)$$

where $\|\cdot\|_F$ is the Frobenius norm and $\lambda = 1$ for scaled input (Section 2.2). The matrix J_r characterises the local empirical response between relative RGBN variation and relative dual-SWIR variation inside the data support of block r .

The response matrices J_r are fitted on the Aksu development split, and the retrieval hyperparameters are fixed on the Aksu selection split; no target-region data enter either. For a target coarse block c with standardised state u_c and a source block r with state u_r , the state distance is

$$d_{cr} = \|u_c - u_r\|_2, \quad (9)$$

where $\|\cdot\|_2$ is the Euclidean norm. The $K = 40$ nearest source blocks form the neighbourhood $\mathcal{N}_c$, and with R_r^2 the coefficient of determination of the local no-intercept ridge regression, the quality weight is $q_r = \text{clip}(R_r^2, 0, 1)$ and the unnormalised weight is

$$\tilde{w}_{cr} = \frac{q_r}{d_{cr} + \varepsilon}, \quad \varepsilon = 10^{-12}, \quad (10)$$

normalised within $\mathcal{N}_c$ as $w_{cr} = \tilde{w}_{cr} / \sum_{k \in \mathcal{N}_c} \tilde{w}_{ck}$. The conditional local response matrix of the target block is the weighted combination

$$J_c = \sum_{r \in \mathcal{N}_c} w_{cr} J_r. \quad (11)$$

After the target blocks are processed, the coarse dual-SWIR anchors, the coarse RGBN means and the local response matrices are bilinearly lifted to fine-pixel locations. For fine pixel i , bilinear lifting provides the local RGBN mean $\mu_i^G \in \mathbb{R}^4$, the dual-SWIR anchor $\mu_i^S \in \mathbb{R}^2$ obtained under the observation budget in Section 2.3, and the local response matrix $J_i \in \mathbb{R}^{4 \times 2}$. The conditional prior at this pixel is

$$z_i = \mu_i^S + J_i^T (g_i - \mu_i^G), \quad (12)$$

where $z_i \in \mathbb{R}^2$ is the conditional prior of fine pixel i and stacking the z_i^T in pixel order yields $Z \in \mathbb{R}^{n \times 2}$. Here μ_i^S fixes the local absolute dual-SWIR state, $g_i - \mu_i^G$ describes fine-scale RGBN variation relative to that local state, and J_i maps that relative variation to a conditional dual-SWIR offset. Because the RGBN contribution is the centred term, the prior depends on relative rather than absolute RGBN radiometry, which is the property relied upon in Section 2.2 for provider-delivered JL1 values.

3.4. Hyperparameters: Provenance, Selection and Reported Sensitivity

The hyperparameters are neighbourhood size K , ridge parameter λ and retrieval distance floor ε . The implementation records the candidate grids and the selection procedure, and the frozen values were fixed before the external regions were read. Selection on the Aksu split used the detail-skill criterion and yielded $K = 40$, $\lambda = 1.0$ and distance $\times$ quality weighting (Supplementary Table S2.1). These settings were retained for all target regions. Section 4.4 evaluates $K \in \{20, 40, 80\}$ and finds lower errors at $K = 80$ in all four regions; this sensitivity analysis did not alter the selected configuration.

The distance floor $\varepsilon = 10^{-12}$ in Equation (10) keeps retrieval weights finite as $d_{cr} \rightarrow 0$. Geometric intersection and parent-block classification use fixed tolerances, which the sensitivity scan does not vary. Controlled solves use separate LSQR tolerances, $atol = btol = 10^{-10}$, with at most 2000 iterations. The ridge parameter is added to $\Delta G^T \Delta G$ in scaled units, so $\lambda = 1$ depends on the input scale. Across source blocks, $\lambda / \lambda_{\min}$ spans about four orders of magnitude, and 99.9989% satisfy $\lambda / \text{tr}(\Delta G^T \Delta G) < 10^{-3}$.

3.5. Observation-Constrained Inverse Allocation in a Unified Variable Field

If the fine-scale result were corrected independently within each coarse pixel, a fine pixel crossing a coarse boundary could receive several conflicting candidates. To guarantee one final value per native fine pixel, inverse allocation is performed in a unified fine-scale variable field.

Let $B \in \mathbb{R}^{n \times q}$ map coefficients at q valid coarse-grid control nodes to the n fine-pixel locations. Bilinear weights connect an interior fine pixel to at most four nodes. At edges, available weights are renormalised. This gives each fine pixel one correction influenced by neighbouring supports. The prior lifting operator L is defined on the fixed parent grid. In contrast, B follows the current coarse grid and is rebuilt for each region-phase in the non-nested benchmark. It is built once for the single nested common-scale configuration. Both operators are geometry-defined, and coverage classes are shared across the configurations of each phase.

The conditional prior Z is first reaggregated by the same area-observation matrix A , and its residual relative to the actual coarse dual-SWIR observations Y is

$$R = Y - AZ, \quad (13)$$

where $R \in \mathbb{R}^{m \times 2}$ is the coarse dual-SWIR residual and $AB \in \mathbb{R}^{m \times q}$ maps control coefficients to coarse residuals. If R is not contained exactly in the column space of AB , the solution is its least-squares approximation within this correction subspace.

$$\Gamma^* = \arg \min_{\Gamma \in \mathbb{R}^{q \times 2}} \|AB\Gamma - R\|_F^2. \quad (14)$$

The final fine-scale dual-SWIR field is

$$\hat{X} = Z + B\Gamma^*, \quad (15)$$

where $\hat{X} \in \mathbb{R}^{n \times 2}$ is the final observation-constrained conditional solution and $B\Gamma^*$ is the continuous correction applied to the prior. Because all fine pixels are represented jointly in one $\hat{X}$, a shared fine pixel occupies a single row of variables and its contributions to several coarse observations are carried by the nonzero area weights in its column of A ; it can therefore respond to neighbouring constraints without receiving several final values.

The retained full-scene implementation separates three operations. Chunks of 16×16 coarse cells assemble observation rows and coarse fields and check their targets against the Sentinel-2 input. Sparse observation rows address the global fine array by linear pixel indices; a pixel may appear in several rows but those indices read the same stored value. One global sparse coarse-coefficient system is solved per band using GMRES, after which the correction is lifted and written in disjoint 1024×1024 fine-pixel windows. These windows also build the prior and checkpoint its memory-mapped float64 arrays; they are not independent inverse problems whose outputs are averaged. The controlled benchmark instead uses the LSQR settings below. The archival reference agrees with the retained production source on this structure but does not by itself establish the provenance of every executed stage; peak memory was not recorded.

The controlled benchmark solves each band separately using `scipy.sparse.linalg.lsqr`, with $atol = btol = 10^{-10}$ and at most 2000 iterations. The implementation forms no dense inverse and does not require AB to have full row rank. Basis B limits residual allocation to a shared, low-capacity correction subspace. Joint RMSE (Equation (16)) combines errors after the two independent band solves.

3.6. Implementation Details and Pseudocode

Algorithm 1 summarises the controlled procedure; Supplementary Table S8.1 gives additional implementation details and the stored-product closure check. The solve domain is the union of the evaluation mask and the column support of A . Pixels excluded from scoring can therefore remain in the solve when they contribute to valid coarse observations. Additional solve pixels range from zero at the reference configuration to about 0.3 % of the evaluation mask at $\sigma = 1$.

Algorithm 1 Observation-constrained inverse allocation in the unified fine-scale variable field (controlled LSQR procedure; full-scene GMRES scope: Section 3.5). Additional implementation details and the stored-product closure check are given in Supplementary Table S8.1.

- 1: **input** Y (current coarse dual-SWIR), target RGBN grid, geometry, frozen K, λ, ε ; the additional nested coarse observation for the original budget.
 - 2: **build library** from Aksu only: for each source block solve Equation (8) for J_r ; store the state u_r and the weight q_r .
 - 3: **freeze** K, λ and the weighting form on the Aksu selection split.
 - 4: **retrieve** the K nearest blocks (Equations (9) and (10)) and form J_c .
 - 5: **anchor**: use the fixed nested mean for the original budget or current-phase overlap weights for the single-observation budget (Section 2.3); classify coverage.
 - 6: **lift** anchors, coarse RGBN means and J_c to form Z (Equation (12)).
 - 7: **operators**: sparse exact-area A (Equations (3)–(6)) and correction basis B .
 - 8: **solve per band**: LSQR of $\min_{\Gamma} \|AB\Gamma - (Y - AZ)\|$ with $atol = btol = 10^{-10}$, ≤ 2000 iterations.
 - 9: **output** $\hat{X} = Z + B\Gamma^*$; in deployment only, bound, write float32, then reread and reaggregate with the same A .
-

Operators A, A_c and B , their products, the anchor and coverage classes are constructed per phase and reused across the parameter scan. The ε scan varies only the retrieval distance denominator. A separate arithmetic diagnostic compares boundary indices at 400 positions and four offsets. It records 358 changed decisions at 10^{-14} and none at 10^{-10} . These are index decisions rather than scene-pixel coverage changes. Its anchor loop uses identical tolerance arguments, leaving geometric-tolerance sensitivity untested.

Incomplete supports are normalised over their observed area (Equation (5)); the resulting constraint represents that partial-area mean. Coarse-cell inclusion uses the coverage fraction c_j (Equation (4)). The single-observation anchor uses a fixed 10^{-9} tolerance on the difference from nominal overlap.

Runtime per non-nested phase was about 0.2–1.6 s per region, compared with 59–126 s for one-off prior-asset construction. Prior construction dominated the reported runtime. Peak memory was not recorded.

4. Experiments and Results

This section has five subsections. Section 4.1 states the experimental design and the evaluation protocol: the data and their splits, the compared methods, the ablation settings as internal controls, and the metric and statistical treatment. Section 4.2 reports the common-scale reconstruction performance of the complete pipelines. Section 4.3 measures the effects of the observation geometry and of the correction representation, including the strictly matched control, the ablation contrast and the phase and boundary behaviour. Section 4.4 compares against the external baselines and examines the robustness to blur, residual registration offset, scale ratio and hyperparameter choice, and Section 4.5 reports the real-scene deployment and the observation-consistency diagnostics.

4.1. Experimental Design and Evaluation Protocol

Data and splits. The experiments use the four regional datasets described in Section 2.2 in two roles, and Table 2 lists for every partition that carries a result its role, its counting unit, its count and where the count is verified. Aksu is the source domain: its JL1 RGBN and Sentinel-2 SWIR pairs fit the response library, select the configuration and form the internal confirmation part of the split. Yining, Reg149-ext and Huyang River are target regions: they are read only after all parameters are frozen, and their native 20 m SWIR is used as the evaluation reference. Inside Aksu the source blocks are ordered by the normalised column index q and separated into development ($q < 0.52$), selection ($0.57 \leq q < 0.73$) and internal confirmation ($0.78 \leq q < 0.99$) parts by buffer bands. The partition counts were recovered by applying the frozen split rules to the retained per-block index (Table 2); the original membership lists were not archived.

Table 2. Roles, counting units and counts of the data partitions used in the experiments. Region order is Aksu, Yining, Reg149-ext and Huyang River. The pixel counts are evaluation extents, not independent sample sizes; Supplementary Table S1.1 lists the masks and the coverage behind them.

Partition	Role in the Experiments	Unit	Count	Verified from
Aksu development part	partition of the Aksu source blocks that the response library is fitted on	source block	88,669 by the frozen rule ($q < 0.52$); the retained library holds the same number of entries	retained per-block index with the frozen rule; retained library (it stores no block identifier, so its entries are not matched to the index one by one)
Aksu selection and internal confirmation	configuration selection and internal confirmation	source block	selection 27,977 ($0.57 \leq q < 0.73$); internal confirmation 37,274 ($0.78 \leq q < 0.99$); tail $q \geq 0.99$ 1658; buffer bands 8794 and 8501; Aksu total 172,873	retained per-block index with the frozen rule (no archived split table; the library stores no block identifier)
Adapted DSen2 training	training of the learned baseline	32×32 crop, stride 16	2623 training crops and 774 selection crops, all from Aksu; 180-epoch schedule, epoch 178 selected, batch 64, patience 25, seed 20260810	training metadata and training history
Adapted ATPRK regression stage	fitting of the geostatistical baseline	coarse–fine pair	fitted on the Aksu development data; the variogram is estimated from the evaluated phase; the number of fitting pairs is not retained	method provenance and settings record
Proposed method	response-library fit and inverse allocation	RGBN–SWIR source-block pair, coarse observation	block-wise closed-form ridge fitting using paired RGBN–SWIR observations from the Aksu development data; the total number of fitting pairs is not retained	Equation (8); Section 2.3
Evaluated 20 m pixels, controlled benchmark	internal and external accuracy	20 m pixel, interior mask	342,063/2,407,095/2,453,994/1,233,776	Table S1.1; block sufficient statistics
Evaluated 20 m child pixels, common scale	common-scale accuracy	20 m child pixel	350,388/2,419,983/2,466,630/1,243,593	Table S1.1; common-mask summary
Spatial blocks, controlled benchmark	resampling units and denominators	4×4 block	16 per region	Table S2.3
Grid phases	phase sensitivity; regional error is defined over the 15 non-nested phases	phase	16 per region: 15 non-nested and the (0,0) nested control	Equation (1); Table S2.3
Macroblocks, common-scale comparison	intervals of the common-scale paired difference	macroblock	10/19/17/20	paired macroblock bootstrap record

Two derived evaluation sets carry the accuracy claims. The controlled non-nested benchmark treats the native 20 m B11/B12 of a region as the fine reference and forms

virtual 60 m observations from it in the 16 phases of Equation (1); each region contributes 16 spatial blocks and the interior-mask pixel count of Table 2, and the same mask is used in every phase. The common-scale experiment uses the strictly nested 60-to-20 m relationship, in which nine 20 m child pixels form one 60 m parent block, and its paired intervals are formed on the 10, 19, 17 and 20 macroblocks of the four regions.

No predictor is fitted on fine-scale labels from a target region. The response library of the proposed method is obtained by the per-block closed-form ridge solve of Equation (8) on paired RGBN–SWIR observations of the Aksu development split, so the method fits data but performs no iterative optimisation; the number of fitting pairs is not retained. The retrieval hyperparameters and the RGBN state standardisation are fixed from the Aksu development and selection sets (Section 2.3). The ATPRK adaptation fits its regression stage on the Aksu development data and estimates its variogram from the evaluated phase; the number of fitting pairs is not retained. The DSen2 adaptation is the learned predictor of the study and is trained on Aksu development crops with epoch selection on the Aksu selection split: 2623 training crops of 32×32 pixels with stride 16 and 774 selection crops are retained, and the reported configuration is the epoch-178 checkpoint of a 180-epoch schedule with batch 64, patience 25 and seed 20260810.

Payload, acquisition date and land-surface composition differ among the four regions and co-vary, so the external results measure transfer across the sampled conditions rather than the effect of any single factor (Section 5.5). The improvement from exact-area geometry was observed in all four evaluated regions. The response library, the RGBN state standardisation and the retrieval hyperparameters were fixed at Aksu and reused unchanged, so a change of study area changes the input scenes and the evaluation masks rather than the parameter set. Because geography, date, payload and land-surface composition co-vary in these data, the experiments do not separate their individual effects. Absolute accuracy at 20 m outside the four regions tested is not established by these experiments.

Comparison methods. Three methods are run under the same single-observation budget, the same evaluation mask, the same true-area observation operator and the same metrics.

- *Our_single*: the proposed method in its single-observation-budget configuration.
- *ATPRK-adapted*: an adaptation of area-to-point regression kriging [9,13]. Area-to-point regression kriging is an established change-of-support approach: a regression of the coarse target band on fine auxiliary bands models the scale trend, and area-to-point kriging of the regression residuals restores fine-scale detail while the prediction reproduces the observed coarse support [9]. The family has since been extended to area-to-point cokriging and to multiple-point geostatistical prediction [13]. Here the regression stage is fitted on the Aksu development data and the area-to-point stage uses the true polygon overlaps of the evaluated phase, so the baseline receives the same observation geometry and the same auxiliary bands as the proposed method.
- *DSen2-adapted-single*: the public DSen2 residual-network backbone [14]. This is a widely used learned sharpening baseline: a residual convolutional network predicts the high-frequency detail of the Sentinel-2 bands that are coarser than the 10 m guide bands and is trained on simulated low-resolution inputs, so one model covers both the 20 m and the 60 m targets [14], and open implementations of it are included in systematic sharpening benchmarks [17]. The network is adapted to the threefold scale task here, receiving four-channel 20 m RGBN and two-channel 60 m SWIR replicated onto the 20 m grid and outputting dual SWIR at 20 m.
- *Coarse_replication*: each coarse dual-SWIR observation is assigned to its fine pixels without any correction; it is an internal no-refinement reference and is not counted as an external method.

The DSen2 adaptation was trained on the Aksu development split, with epoch selection on the Aksu selection split, and no external region supplied training or selection data (Section 4.1 and Supplementary Table S2.2). The original-budget A4 configuration receives an additional coarse observation and is evaluated separately from these single-observation baselines. The comparison covers one established geostatistical family and one learned residual-network baseline under matched input budgets, geometry and metrics; implicit-representation and generalised-inverse methods discussed in Section 1.2 are outside its scope.

Ablation settings and evaluation metrics (internal controls). The ablations are internal controls rather than external comparisons, and they are defined before the results. Five ablations, A0–A4, complement the full-pipeline comparison inside the controlled benchmark. A0 keeps only the RGBN-conditioned prior and applies no coarse correction. A1 locates each fine pixel's coarse cell from its centre and corrects in pixel space. A2 uses true area overlap and a minimum-norm pixel-space correction. A3 takes the coarse residual from true area geometry but distributes it piecewise by centre ownership. A4 combines true-area observations with a unified continuous correction subspace and is the complete non-nested formulation of this paper. The designs are tabulated in Supplementary Table S8.6.

The difference between A1 and A2 changes the observation relationship and the correction form at the same time, so it is reported as descriptive rather than as a single isolated factor. A2 and A4 share the area–observation relationship and the same prior, so their difference isolates the correction form alone: A2 corrects in pixel space, whereas A4 expresses the correction on a continuous basis defined at the coarse control nodes. Their contrast therefore measures the additional gain from the continuous correction basis and is not a geometry contrast, while A3 and A4 contrast piecewise correction with a unified variable field. The strictly matched pair is the only contrast in which the prior, the basis, the optimisation and the constraints are identical and only the observation matrix differs; it is reported in Section 4.3.

The evaluation metrics and the statistical treatment complete the settings of the controlled benchmark.

Single-band errors are evaluated by root-mean-square error (RMSE), mean absolute error (MAE) and bias. Overall dual-SWIR error is defined as

$$\text{RMSE}_{\text{joint}} = \sqrt{\frac{\text{RMSE}_{B11}^2 + \text{RMSE}_{B12}^2}{2}}. \quad (16)$$

Regional error is the mean of phase-wise regional joint RMSE over the 15 non-nested phases. Each phase-wise RMSE pools the evaluated pixels of that phase. Paired gains use either the difference of regional means or the regional mean of spatial-macroblock differences, as specified for each result.

In addition to absolute error we calculate the spectral angle mapper (SAM) and examine the relative structure of B11 and B12. Because a direct ratio becomes unstable when B12 approaches zero, the primary comparison uses a robust ratio with a 0.01 reflectance floor,

$$R_f = \frac{B11}{\max(B12, 0.01)}, \quad (17)$$

the log ratio

$$R_{\log} = \log \frac{\max(B11, 0) + 0.01}{\max(B12, 0) + 0.01}, \quad (18)$$

and the normalised SWIR contrast

$$C = \frac{B11 - B12}{B11 + B12 + 0.01}. \quad (19)$$

For the spectral diagnostics of Section 4.4 the dual-SWIR angle is defined on the $(B11, B12)$ vector in scaled units and is undefined when either norm is below 1 scaled unit, as is the normalised difference when $|B11 + B12|$ is below 1 scaled unit. These rules apply only to the spectral diagnostics and are not propagated into the RMSE metrics, which use the common evaluation mask.

Spatial blocks, rather than pixels or grid phases, are the resampling units. The controlled benchmark divides each evaluation extent into four segments along each axis, giving at most 16 phase-independent blocks. The single nested common-scale configuration uses 10, 19, 17 and 20 macroblocks in Aksu, Yining, Reg149-ext and Huyang River, respectively. These are the denominators of the improvement fractions in Section 4.2. Block resampling retains part of the spatial structure but does not establish complete independence.

For the common-scale comparison, the point estimate is the paired difference of the two complete pipelines on that configuration's macroblock partition. Conditional intervals use 5000 macroblock resamples (`paired_spatial_macroblock_bootstrap.csv`); the seed was not recorded. This statistic differs from a difference of region-pooled RMSE values. The per-phase mechanism analysis forms, for each region and phase separately, the mean of the paired spatial-block differences between the two ablation configurations and resamples those blocks within that region and phase; its draw count and seed are not recorded, and its intervals are reported in Supplementary Table S4.1. Its sampling settings are therefore stated separately from those of the four region-level comparisons below and are not transferred to them.

The four region-level parameter and external comparisons use 16 spatial blocks, 2000 resamples and seed 20261002. Each draw is reused across methods and all 15 non-nested phases. Within each draw, phase-wise joint RMSE is reconstructed from block sufficient statistics as $\sqrt{\sum_b SSE_b / \sum_b N_b}$, then averaged over the fixed phase set. Regions and phases are not pooled into a single RMSE. Supplementary Table S2.3 specifies each experiment's point estimate, spatial unit, draw count, seed and reuse. These intervals condition on the observed scenes and fixed model; their scope is discussed in Section 5.4.

4.2. Common-Scale Reconstruction Performance

The coarse and fine grids are strictly nested here, so this experiment involves no non-nested area geometry; native 20 m B11/B12 nevertheless serve directly as truth, so the level tests whether each pipeline recovers meaningful fine-scale SWIR variation under a common scale relationship.

Table 3 and Figure 3 show that V2 attains the lowest B11 RMSE, B12 RMSE, joint RMSE and joint MAE in all four regions. Its joint RMSE is 0.008977, 0.012093, 0.008727 and 0.011434 in Aksu, Yining, Reg149-ext and Huyang River, respectively. Relative to task-adapted DSen2 these values represent reductions of 28.80%, 8.90%, 19.93% and 16.49%, where a reduction is defined as $(RMSE_{DSen2} - RMSE_{V2}) / RMSE_{DSen2}$ computed from the unrounded values of Table 3. The paired spatial-macroblock intervals of the joint-RMSE difference lie entirely below zero in all four regions, and the fractions of macroblocks showing improvement are 10/10, 17/19, 17/17 and 19/20, i.e., 100%, 89.5%, 100% and 95%.

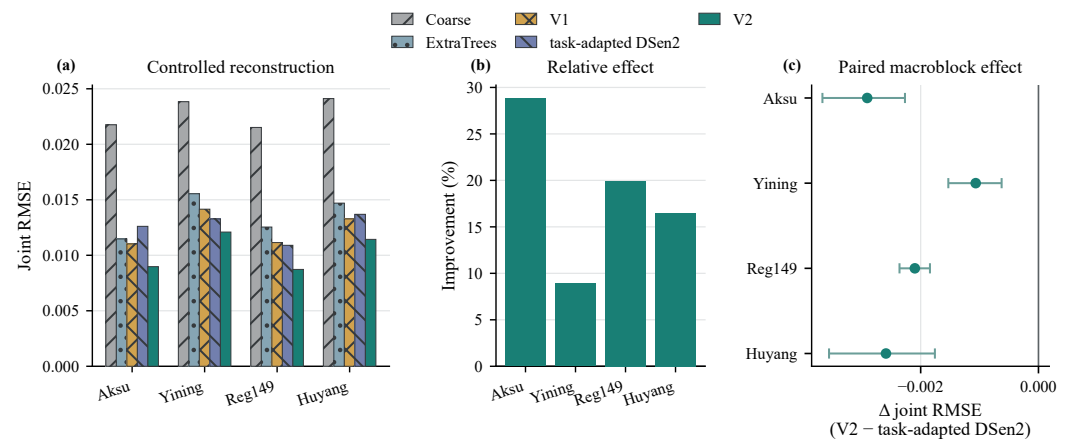


Figure 3. Common-scale performance: (a) joint RMSE; (b) V2 reduction relative to task-adapted DSen2; (c) paired macroblock differences (V2 minus DSen2) with 95% conditional resampling intervals. Aksu, Yining, Reg149-ext and Huyang River use 10, 19, 17 and 20 macroblocks, respectively. Negative differences favour V2.

Table 3. Common-scale accuracy of five complete pipelines on the nested 60-to-20 m scene. RMSE and MAE are pooled over valid child pixels and reported in reflectance units; SAM is in radians.

Region	Method	B11 RMSE	B12 RMSE	Joint RMSE	Joint MAE	SAM (rad)
Aksu	Coarse replication	0.020683	0.022781	0.021757	0.014149	0.019585
	ExtraTrees	0.010946	0.011974	0.011472	0.007354	0.010932
	V1	0.010652	0.011388	0.011026	0.006643	0.009972
	Task-adapted DSen2	0.012636	0.012580	0.012608	0.007225	0.009979
	V2	0.008511	0.009420	0.008977	0.005649	0.009169
Yining	Coarse replication	0.022279	0.025299	0.023837	0.014463	0.022087
	ExtraTrees	0.015127	0.015951	0.015544	0.009515	0.015607
	V1	0.013307	0.014952	0.014153	0.008013	0.013593
	Task-adapted DSen2	0.013048	0.013496	0.013274	0.007940	0.012771
	V2	0.011346	0.012797	0.012093	0.006926	0.012587
Reg149-ext	Coarse replication	0.017434	0.024943	0.021518	0.012348	0.020916
	ExtraTrees	0.011652	0.013377	0.012545	0.007866	0.011016
	V1	0.010271	0.011954	0.011144	0.006537	0.009973
	Task-adapted DSen2	0.010511	0.011274	0.010899	0.006664	0.009555
	V2	0.008038	0.009365	0.008727	0.005316	0.008731
Huyang	Coarse replication	0.020982	0.026889	0.024117	0.014265	0.030546
	ExtraTrees	0.013772	0.015572	0.014700	0.008693	0.016736
	V1	0.012833	0.013712	0.013280	0.007749	0.014989
	Task-adapted DSen2	0.014280	0.013076	0.013691	0.008457	0.018356
	V2	0.010719	0.012106	0.011434	0.006915	0.014401

Bold indicates the lowest value per region and metric.

V1, V2 and task-adapted DSen2 are complete pipelines with different configurations. V1 applies a pixel-space exact right inverse to its prior, while V2 uses continuous control-node correction. Their priors differ by an RMS of 67–80 scaled units (6.7×10^{-3} – 8.0×10^{-3} in reflectance units), so their contrast includes both prior and correction changes. V2 also differs from the non-nested A4 configuration in its prior and evaluation protocol. Section 4.3 uses identical priors to isolate geometry from correction.

Complete-pipeline comparison. This comparison uses the non-nested protocol and is independent of the common-scale Table 3 and Figure 3. The advantage of V2 in the complete non-nested pipeline is not confined to a particular grid phase. The median joint RMSE of V1 and V2 across the 15 non-nested phases is 0.010544 and 0.008479 in Aksu, 0.013489 and 0.011585 in Yining, 0.010651 and 0.008246 in Reg149-ext, and 0.012725 and 0.010981 in Huyang River. V2 is lower not only in every region but at every one of the 60 region–phase combinations, the two distributions being separated without overlap. Because the two configurations are differently configured pipelines, this is a comparison of two complete systems under a common evaluation protocol.

Relative dual-SWIR structure. These metrics use the nested common-scale configuration and therefore compare the same configurations as Table 3. All three are evaluated on the valid child pixels of the nested scene with the same code for the five pipelines: the 0.01-floor ratio of Equation (17), the log ratio of Equation (18) and the normalised contrast of Equation (19). The 0.01 constant enters a different expression in each case. The RMSE of the 0.01-floor ratio, the log ratio and the normalised contrast is 0.0486/0.0292/0.0150 for task-adapted DSen2 and 0.0454/0.0275/0.0141 for V2 in Aksu; 0.0717/0.0331/0.0159 and 0.0679/0.0306/0.0146 in Reg149-ext; and 0.1241/0.0515/0.3054 and 0.1066/0.0478/0.0230 in Huyang River. In Huyang River the normalised contrast of the DSen2 adaptation is an order of magnitude larger than that of V2, by 13.3 times, and the same configuration leaves the most B12 values at or below the floor of the five pipelines, 4845 of the 1,243,593 child pixels. V2 improves all three measures in these three regions. Yining does not follow the same trend: the three metrics are 0.0767, 0.0389 and 0.0194 for task-adapted DSen2 and 0.0800, 0.0386 and 0.0192 for V2, so the floor-ratio RMSE is worse for V2 while the log ratio and the normalised contrast are marginally better. The regression is therefore confined to the floor-ratio metric, and the controlled diagnostic of Section 4.4 shows that the dual-SWIR spectral angle improves as well. Lower absolute B11 and B12 error therefore does not guarantee an improvement of every relative metric in every region. This regional result is retained and is examined further in Section 4.4. The raw ratio without a floor is extremely sensitive to B12 near zero and is kept only as a denominator-sensitivity diagnostic, not for ranking.

4.3. Effects of Observation Geometry and Correction Representation

To isolate observation geometry, we compared two configurations sharing the conditional prior Z , correction basis B , solver, coverage classification and constraints. The residual is matched through center-ownership matrix A_c in one configuration and exact-area matrix A in the other:

$$\gamma_{\text{center}}^* = \arg \min_{\gamma} \|A_c(z + B\gamma) - y\|_2, \quad \gamma_{\text{overlap}}^* = \arg \min_{\gamma} \|A(z + B\gamma) - y\|_2. \quad (20)$$

The configurations share the same B array, prior and solver settings and use unconstrained outputs. The matched-overlap result reproduces archived A4 within 10^{-6} scaled units. Both configurations were computed in one process to share the arrays. Table 4 includes the A0 prior-only reference.

Table 4. Strictly matched geometry control over 15 non-nested phases. Only the observation matrix differs; A0 uses the prior alone. Values are mean joint RMSE in reflectance units. Δ is the mean per-phase difference (center minus overlap); positive values favour overlap. The last columns count improving phases and 4×4 spatial blocks.

Region	A0 Prior Only	Matched Centre	Matched Overlap	Δ (Centre – Overlap)	Δ Relative to Centre	Phases Overlap Better	Blocks Overlap Better
Aksu	0.009857	0.010556	0.008474	0.002082	19.72%	15/15	16/16
Yining	0.012995	0.013292	0.011555	0.001738	13.07%	15/15	16/16
Reg149-ext	0.009910	0.010395	0.008222	0.002173	20.90%	15/15	16/16
Huyang River	0.012413	0.012931	0.010956	0.001974	15.27%	15/15	16/16

Errors follow the reporting convention in Section 2.2. Supplementary Table S8.7 summarises the coarse inputs, anchors and prior-sharing rules of each experiment class.

Two summaries of a paired gain are in principle available: the region mean of the per-phase differences, $P^{-1} \sum_p (r_p - c_p)$ with $P = 15$, and the difference of the two region means. Both configurations use the same phase set and the same evaluation mask, so the two coincide algebraically and a single value is reported for each gain. Every relative gain in this paper, including those of Section 4.4, is a ratio of region means.

Exact area overlap reduced mean joint RMSE by 0.00174–0.00217, or 13.1–20.9% relative to center ownership. The reduction occurred in all four regions, all 15 non-nested phases and all 16 spatial blocks of each region. Because correction parameterisation was identical, the contrast isolates the observation matrix. Relative to A0, overlap reduced error by 0.00138–0.00169, whereas center ownership increased error in every region. These results use the original observation budget; Section 4.4 evaluates the single-observation budget separately.

The matched-control archive contains point estimates, per-block differences and improving-phase and improving-block counts, but no bootstrap interval. Table 4 therefore reports these quantities without an interval. The source is `E_R1_A0_geometry/matched_geometry_v1`, including `metrics_by_region_phase.csv` and the block table. Resampling settings for experiments that provide intervals are listed in Supplementary Table S2.3.

The contrast between observation geometry and variable organisation is the second part of this control.

Table 5 summarises A1–A4 over the 15 non-nested phases. Figure 4 shows phase-level errors, their distributions and differences by pixel ownership.

Table 5. Mean joint RMSE across 15 non-nested phases, in reflectance units. A0 (prior only) is 0.009857, 0.012995, 0.009910 and 0.012413 in Aksu, Yining, Reg149-ext and Huyang River, respectively.

Region	A1: Hard Centre Ownership	A2: True-Area Pixel Correction	A3: Piecewise Centre Ownership	A4: True-Area Continuous Correction
Aksu	0.010572	0.008603	0.008907	0.008474
Yining	0.013379	0.011635	0.011981	0.011555
Reg149-ext	0.010507	0.008388	0.008726	0.008222
Huyang River	0.012977	0.011022	0.011338	0.010956

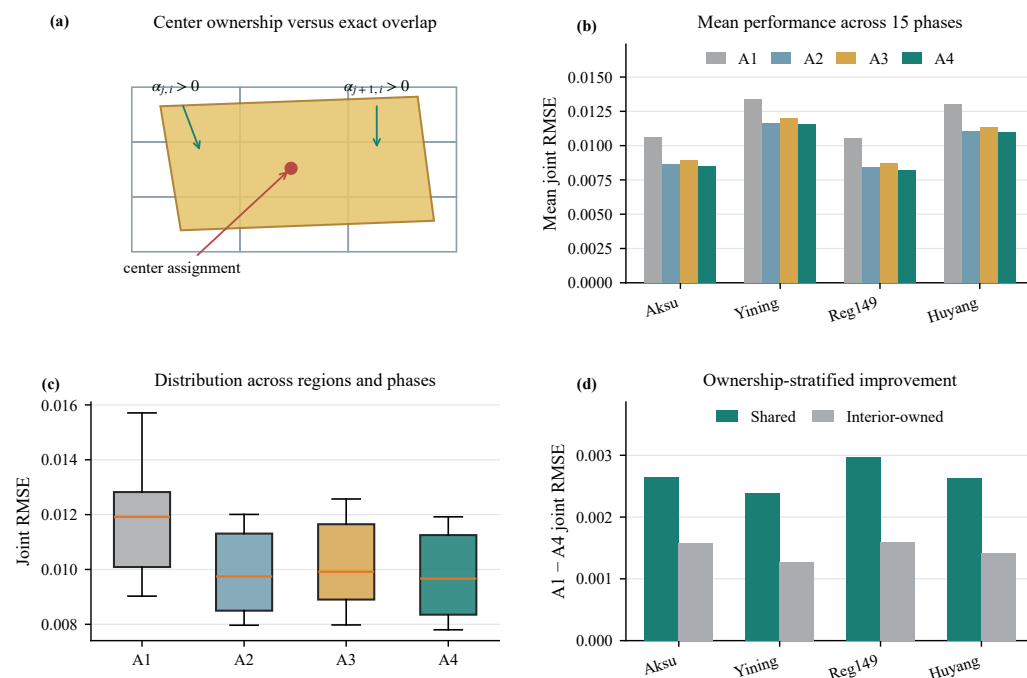


Figure 4. Non-nested ablations: (a) centre ownership and area overlap at a coarse boundary; (b) mean joint RMSE over 15 phases; (c) regional and phase-wise error distributions; (d) reductions for shared and interior-owned pixels. A1–A2 changes both geometry and correction; A2–A4 isolates correction. The separate geometry control is in Section 4.3.

Two summaries describe the ablation gains. Differences of regional mean errors are $A2-A4 = 0.0000657-0.0001660$ and $A1-A4 = 0.0018237-0.0022844$. Regional means of paired spatial-macroblock differences over the 15 phases are $0.000072678-0.000162756$ and $0.001813423-0.002231001$, respectively. Their different aggregation orders account for the numerical difference. The paired differences are written $A4-A2$ and $A4-A1$, so a negative value favours A4; their intervals lie entirely below zero in 58/60 region-phase combinations for the first and 60/60 for the second. The two exceptions are Yining at (15, 15) and Huyang River at (0, 15) (Supplementary Table S4.1).

The $A2 \rightarrow A4$ reduction is approximately $0.000066-0.000166$ in reflectance units, or 0.6–2.0% of the corresponding A2 error. Region-level paired-block intervals exclude zero in all four regions. Continuous correction thus adds a consistent but small improvement after the area relationship is specified.

For the difference of regional mean errors (A2 minus A4), the 95% conditional spatial-block intervals are $[0.00008950, 0.00017333]$ in Aksu, $[0.00005814, 0.00010244]$ in Yining, $[0.00013863, 0.00019384]$ in Reg149-ext and $[0.00003831, 0.00009500]$ in Huyang River, in reflectance units. These region-level intervals use the E_R5 statistic specified in Supplementary Table S2.3, row 7; they differ from the per-phase paired-block intervals of Table S4.1. The strictly matched control isolates the larger geometry effect (Section 4.3); the A2–A4 contrast isolates correction representation.

Phase and boundary behaviour. A1 had the largest phase sensitivity, with regional joint-RMSE ranges of $0.003590-0.004065$ across phases. For A4, the ranges narrowed to $0.000748-0.000829$, approximately a fivefold contraction. A2 and A4 had similar phase variability, indicating that area overlap accounts for most of this stability. Some non-nested phases had lower median RMSE than the nested (0, 0) condition. A shift changes the alignment between observation supports and scene structure, so phase sensitivity is evaluated by the cross-phase range, variability and worst phase.

Pixel-ownership types show the expected differences: shared fine pixels cross coarse boundaries, their observation weights are the ones most affected by the centre-ownership approximation, and their absolute-error reduction from A1 to A4 is generally larger than that of interior-owned pixels. Improvement is not confined to shared pixels, however, and interior pixels also show reduced error, because the area–observation relationship propagates through the whole constraint system; the full A1-to-A4 difference therefore cannot be attributed to boundary pixels alone.

The boundary-gradient metric is a complementary spatial diagnostic. A4 moves the ratio closer to one and reduces boundary imprints across grid phases, although every ablation configuration stays below the prespecified anomalous-boundary threshold of 1.25 [Figure 5a], and in the complete pipelines V2 has a lower ratio than V1 and task-adapted DSen2 [Figure 5b]. The paired controls in Figure 5c show a lower joint RMSE for V2 than for V1 and for the no-RGBN control under their common protocol. The no-RGBN control shares the coarse SWIR observations, aggregation matrix, continuous basis, evaluation samples and spatial bootstrap of V2, starts from a continuous interpolation of the coarse SWIR anchors, and omits the RGBN-conditioned fine-detail response; the paired controls are evaluated unconstrained on the nested common-scale configuration of Table 3. As noted in Section 4.2, V1 and V2 do not share a prior, so this comparison involves two configured pipelines rather than isolating the correction.

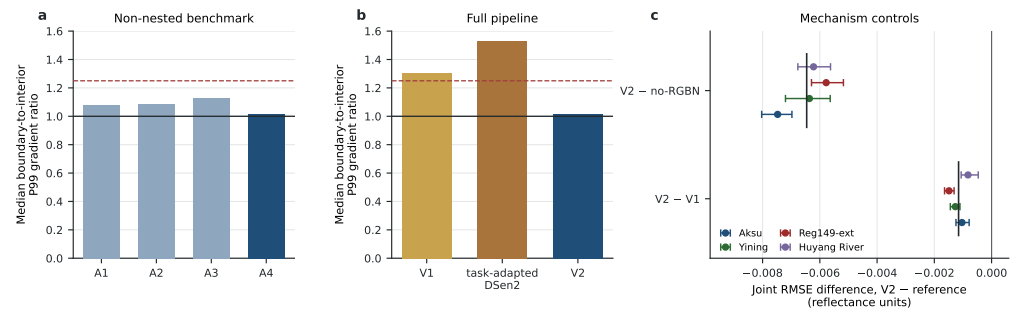


Figure 5. Boundary behaviour and paired controls. Panels (a,b) show 99th-percentile boundary-gradient ratios for A1–A4 and the complete pipelines. Lines mark equal boundary/interior gradients (1.0) and the anomaly threshold (1.25). Panel (c) shows V2-minus-control joint-RMSE differences with regional 95% paired-macroblock intervals. The short bar is the unweighted regional mean, without an interval; negative values favour V2.

4.4. Benchmark Comparison and Robustness Analysis

External baselines. The comparison above is internal. To compare against established method families under identical conditions, the external methods of Section 4.1 were run under the same single-observation budget, the same evaluation mask, the same true polygon-overlap operator and the same metrics. Table 6 reports the result; Coarse_replication is shown as the internal no-refinement reference. The paired differences δ in this table are the exception to the reflectance-unit reporting convention and are labelled in scaled units.

Table 6. Single-observation comparison over 15 non-nested phases under a shared evaluation mask and observation operator. Joint RMSE is in reflectance units; coarse replication is an internal reference. $\delta = \text{RMSE}_{\text{Our}} - \text{RMSE}_{\text{cand}}$ is the regional mean paired difference, with its 95% conditional spatial-block interval, in scaled units (10^{-4} reflectance). Negative δ favours the proposed method.

Region	Our_SINGLE	ATPRK-Adapted	DSen2-Adapted-Single	Coarse_Replication	δ vs ATPRK [Interval]	δ vs DSen2 [Interval]
Aksu	0.009216	0.017823	0.018697	0.033559	−86.07 [−98.69, −73.51]	−94.80 [−126.96, −61.65]
Yining	0.012209	0.021499	0.014866	0.027989	−92.90 [−102.58, −84.27]	−26.57 [−45.92, −11.72]
Reg149-ext	0.008813	0.018915	0.014191	0.027229	−101.02 [−115.67, −83.44]	−53.78 [−70.73, −34.33]
Huyang River	0.011607	0.017422	0.014305	0.027685	−58.15 [−67.66, −48.83]	−26.97 [−35.96, −18.52]

Bold indicates the lowest mean joint RMSE in each region.

The proposed method achieved the lowest mean joint RMSE in all four regions. Reductions relative to DSen2-adapted-single were 0.00266–0.00948 in reflectance units, or 17.9–50.7%; reductions relative to ATPRK-adapted were 0.00582–0.01010, or 33.4–53.4%. Every conditional block-resampling interval excluded zero. These single-observation results use the corrected DSen2 prediction archive and normalisation documented in Supplementary Section S5. The common-scale DSen2 comparison in Section 4.2 uses a separate configuration and is unchanged.

Because all three methods use the same observation geometry, Table 6 compares complete predictors within the single-observation budget. The geometry effect is evaluated by the separate matched control in Section 4.3. Broader baseline coverage is discussed in Section 5.6.

Robustness to blur, registration offset and scale ratio. The controlled benchmark generates observations with the same operator that reconstruction assumes. Three controlled groups test how the conclusions behave when that assumption is relaxed: a Gaussian blur applied to the observation-generation operator, a translation applied to the RGBN guide image, and a change of the coarse-cell size. Within each group and phase, all perturbation conditions are evaluated on the same pixels. Within one condition the matched geometry

pair of Equation (20) shares the prior Z , the basis B and the solver, and the perturbation enters through the observation-generation operator, the guide image or the coarse grid. The prior is re-derived for each condition from that condition's own inputs: the anchor from the condition's coarse observation and the RGBN-dependent term from the condition's guide image. The evaluation sets differ between groups, because each follows from the perturbation it has to accommodate: in Aksu they contain 333,941 pixels for the blur group and 336,921 for the offset group, and between 332,136 and 334,050 for the scale group, which builds its common set separately for each phase, against 342,063 for the matched control of Table 4. All three groups also use the single-observation budget, in which the anchor is estimated from the current coarse observation rather than from the fixed 60 m parent-block mean of native 20 m SWIR. Their unperturbed reference levels, 0.009064–0.009741 at Aksu, are therefore higher than the 0.008474 of the overlap configuration in Table 4. The comparisons below are within-group comparisons; reference levels are not interchangeable between groups or with Table 4, and no statement here compares configurations that do not share a prior.

The discrete kernel is characterised by σ in units of the 20 m reference cell (Supplementary Table S8.2). The $\sigma = 0.25$ condition is an almost-identity kernel with central weight 0.99933 and effective width 0.026 cells, about 0.5 m, so no robustness conclusion is drawn from it. At $\sigma = 0.5$ the mean joint RMSE of the overlap configuration rises by only 0.90–1.79% relative to the $\sigma = 0$ reference of the group, and at $\sigma = 1$ by 17.4–28.1%, with the ordering of the two geometries preserved in every condition and region.

The residuals against the two candidate operators are informative in their own right. Against the *assumed* operator the residual is 4×10^{-12} to 7×10^{-12} in every blur condition, that is numerically closed, whereas against the *true generating* operator it grows from 4.8×10^{-6} at $\sigma = 0.25$ to 1.4×10^{-3} at $\sigma = 0.5$ and 4.0×10^{-3} at $\sigma = 1$. Closure against an assumed operator therefore does not certify the observation process, and a small closure residual is not an accuracy measure. A separate exact-operator control, in which the true generating operator is used in the solve, gives 0.009674, 0.012294, 0.008951 and 0.011661 for the four regions at $\sigma = 0.5$, close to but not equal to the no-blur reference of 0.009741, 0.012361, 0.009002 and 0.011727. Because the observation dimension is far below the number of unknowns, this near-recovery demonstrates neither uniqueness of the fine-scale solution nor reversibility of a real mis-specified deployment.

Translating the RGBN guide by ± 10 m in rows or columns increased error while preserving the overlap advantage (Supplementary Table S8.3). The observation operator and the group's evaluation set are unchanged across the five offsets. Relative to the unshifted condition, errors increased by 14.3–27.3% for column offsets and 9.8–26.2% for row offsets. The relative advantage of overlap narrowed from 11.2–17.8% to 7.7–13.5%. Overlap retained lower error in all 20 region-condition combinations, each evaluated over the same 15 non-nested phases.

Changing the coarse-cell size from 60 m to 40 m (ratio 2) or 120 m (ratio 6) changes both the size of the coarse supports and the number of coarse constraints, and the two effects are not separated here (Supplementary Table S8.4). Within that limitation the pattern is clear: at ratio 2 the absolute error of the overlap configuration drops by 25–30% relative to ratio 3 and the relative advantage over centre ownership grows to 28.7–39.2%, whereas at ratio 6 the absolute error rises by 83–137% and the relative advantage narrows to 0.93–1.22%. Under these controlled conditions, and for this scene family, the benefit of exact area overlap becomes small when the coarse support is large relative to the fine structure; this is not a general scaling law and not a statement that the geometric advantage disappears.

Across all 48 region-condition combinations of the three perturbation groups, that is 4 regions $\times$ 12 conditions, there is no combination in which the overlap configuration fails

to improve on the prior-only reference and none in which the centre configuration beats the overlap configuration. The ordering is therefore preserved throughout the tested envelope, while the error level and the relative size of the geometric advantage vary substantially.

Hyperparameter sensitivity and error diagnostics. Supplementary Table S3.1 gives one-factor sensitivity relative to the fixed default on the same phase set and evaluation mask. Additional wiring checks confirmed that the parameters affect their intended components. Setting $K = 5$ changed the prior by 89.3%, setting $\lambda = 10^{-3}$ changed response matrices by 5.6%, and setting $\varepsilon = 10^{-2}$ changed the prior by 3.5%.

Error decreased monotonically over $K \in \{20, 40, 80\}$, with $K = 80$ performing best in all four regions. The original selection value $K = 40$ was retained; values above 80 were not tested. Changing λ affected regional RMSE by roughly $10^{-2}\%$ – $10^{-1}\%$, consistent with its small size relative to the source-block normal-matrix eigenvalues. The ε sweep changed RMSE by about $10^{-12}\%$, at floating-point rounding level. It therefore had no measurable effect over the tested retrieval-distance range. The scan reused phase geometry, anchors and coverage classes. The separate 358-decision arithmetic diagnostic and its fixed-argument anchor loop do not test sensitivity of overlap construction to geometric tolerance.

The controlled benchmark was further evaluated through error distributions and dual-SWIR spectral diagnostics.

Supplementary Table S8.5 summarises five methods over 15 non-nested phases. Across the 20 region–method pairs, signed joint bias ranged from about 3×10^{-5} to 9×10^{-2} of joint RMSE. Opposite bandwise offsets can cancel in a joint statistic, so the machine-readable tables also report per-band bias. The 95th-percentile absolute error was 1.30–2.10 times joint RMSE. Within-phase spatial-block variation was comparable to or greater than phase-to-phase variation, supporting spatial-block resampling rather than phase-wise inference alone. Pixel counts refer to distinct fine pixels in the evaluation mask. Supplementary Table S6.1 lists the per-phase and per-block tables together with the number of phases each one covers; repeated pixels across phases do not constitute independent observations.

On the controlled benchmark, where the same evaluation mask and the same phases apply, the dual-SWIR diagnostics improve from V1 to V2 in all four regions: the RMSE of the difference $D = B11 - B12$ falls from 0.004817 to 0.004238 in Aksu, from 0.006997 to 0.006335 in Yining, from 0.005457 to 0.004743 in Reg149-ext and from 0.007176 to 0.006500 in Huyang River; the dual-SWIR spectral angle falls from 0.562° to 0.508° , 0.765° to 0.693° , 0.560° to 0.477° and 0.848° to 0.795° ; and the RMSE of the normalised dual-SWIR difference falls from 0.0278 to 0.0145, 0.0715 to 0.0318, 0.0192 to 0.0183 and 0.1068 to 0.0627. Block-level analysis identifies no block in which the D RMSE increases from V1 to V2: across 240 block-phase cells in each region, the count of cells with an increase is zero.

These controlled V1–V2 diagnostics use different definitions and a different protocol from the relative-structure metrics of Section 4.2: they use the difference $D = B11 - B12$ in reflectance units, a dual-band angle in degrees and a normalised difference whose denominator is $B11 + B12$ under the validity rules stated in Section 4.1, not the 0.01-stabilised ratio, log ratio and contrast. The Yining floor-ratio reversal in Section 4.2 therefore remains a regional transfer limitation. A dual-band angle evaluates interband structure rather than the full spectrum, and normalised dual-SWIR difference is not a measured-moisture accuracy metric. Controlled predictions retain negative and out-of-range values, with a maximum out-of-range fraction of 2×10^{-5} . Deployment applies the output bound described in Section 3.2.

4.5. Real-Scene Deployment and Observation Consistency

The framework was applied to 0.75 m JL1 RGBN and 20 m Sentinel-2 B11/B12 over Huyang River. Each output GeoTIFF contains $28,268 \times 28,399$ pixels in WGS 84 / UTM

Zone 45N (EPSG:32645). Full-scene computation uses 5041 coarse-constraint chunks, while GeoTIFF storage uses 3136 internal 512×512 blocks per band. The fixed model settings were retained throughout prediction, overlap construction, solution and writing. All finite output pixels lie within the nominal $[0, 1]$ bound. With no same-scale SWIR reference, product evaluation uses output validity and coarse-observation closure.

The final float32 products were reread from disk, converted to float64 and reaggregated with the solution's area-observation operator. Closure RMSE was 2.16×10^{-10} for B11 and 2.70×10^{-10} for B12 in the record's reported reflectance units. Maximum absolute residuals were 2.52×10^{-8} and 7.79×10^{-8} , and 99th-percentile residuals were 6.22×10^{-10} and 5.10×10^{-10} . The closure target was read independently from the original Sentinel-2 SWIR input and matched the stored chunk targets. The archive labels the reference storage convention as "scaled reflectance integer (x10000)" and the comparison unit as "physical reflectance". These labels describe the recorded numerical comparison, without independently verifying physical calibration. Supplementary Section S5 documents the scale conversion and prediction-archive correction. Same-operator closure measures numerical consistency; it does not evaluate fine-scale accuracy.

Figure 6c–e illustrate the fine-scale spatial allocation produced on the native JL1 grid. Together with the closure test, the deployment confirms full-scene computation on actual non-nested JL1 and Sentinel-2 supports. The products are reported as observation-consistent conditional SWIR allocations, consistent with the validation hierarchy of Section 2.4.

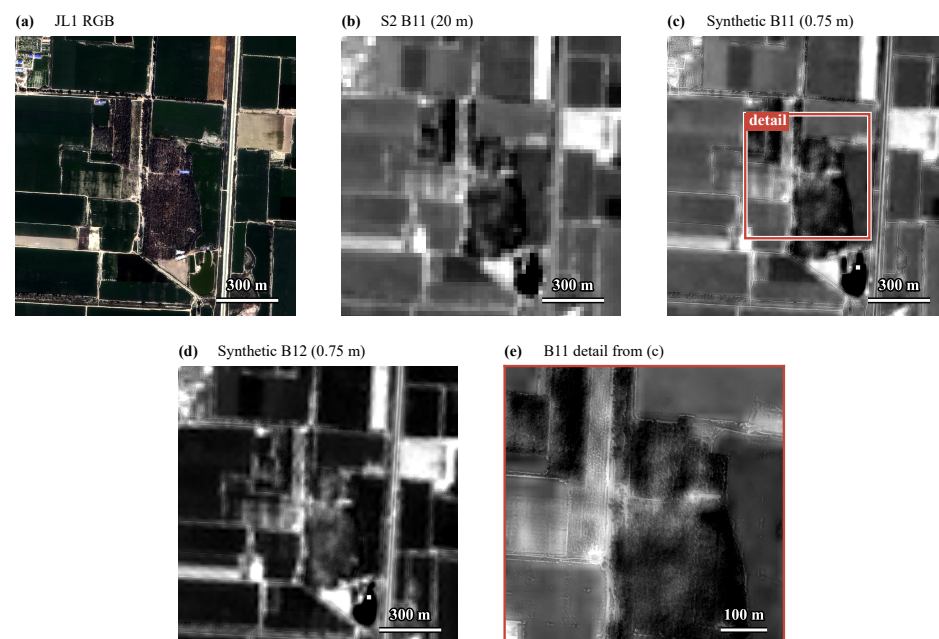


Figure 6. Deployment over Huyang River: (a) JL1 RGB; (b) Sentinel-2 20 m SWIR-1; (c,d) synthetic 0.75 m SWIR-1 and SWIR-2; (e) detail of (c). These observation-constrained allocations lack same-scale validation.

5. Discussion

5.1. What the Geometry Control Isolates, and What It Does Not

The matched control identifies observation geometry as the main source of improvement. Holding the prior, correction basis, solver, coverage and constraints fixed, exact area overlap reduces RMSE by 0.00174–0.00217. The improvement is consistent across every region, non-nested phase and spatial block. Centre ownership changes both the fine vari-

ables contributing to a coarse observation and their weights, thereby altering the constraint equations. Its correction can consequently increase error relative to the prior alone.

The experiment uses known shifts and observations generated by the reference overlap operator. It quantifies geometric approximation error under these controlled conditions. Registration perturbations extend the test to displaced guides, while unknown and non-rigid geometric errors remain outside the evaluated design.

5.2. The RGBN Prior and the Underdetermined Solution Set

The observation matrix constrains coarse means, while the RGBN-conditioned prior supplies fine-scale structure. Local response matrices describe empirical source-domain relationships. Metamerism and training-domain dependence limit their transfer across surfaces and acquisition conditions [3–5,31]. Centring the guide about a local coarse state cancels a common additive offset, but leaves gain differences and band-dependent offsets; the full prior also depends on the retrieval state through the coarse anchor. Independent cross-calibration remains necessary for an absolute radiometric assessment.

Yining illustrates the distinction between bandwise error and interband structure. V2 reduces absolute error but has a slightly worse floor-ratio RMSE than task-adapted DSen2 in the common-scale experiment. The controlled V1–V2 diagnostics improve in all four regions, including Yining. Together, these results indicate that lower absolute error does not ensure the same ranking for relative spectral metrics across protocols and transferred regions.

5.3. Relation to Change-of-Support, Area-to-Point, Sharpening and Observation-Consistent Methods

Areal integration is established in change-of-support and area-to-point methods [9,13,21,22]. Coarse-observation consistency also underlies Sentinel-2 super-resolution and model-driven multisource fusion [15,18]. The present framework combines these principles with a continuous correction field on native non-nested supports. A single fine-pixel variable carries contributions to neighbouring coarse constraints.

The matched controls distinguish two effects that are mixed in the A1–A2 ablation. Exact support geometry reduces error by 0.00174–0.00217, whereas A2–A4 continuous correction adds approximately 0.000066–0.000166. The roughly tenfold difference indicates that correct observation support matters more than correction representation in this benchmark. The single-observation comparison further shows lower errors than the two adapted baselines under shared geometry, budget and evaluation conditions. This supports the complete method within the tested protocol, with baseline coverage limited as described in Section 5.6.

5.4. Uncertainty Scope

The evaluation separates four forms of evidence:

1. **Controlled error at 20 m.** Reference-based error distributions in Section 4.4 describe empirical accuracy at the retained reference resolution.
2. **Conditional resampling intervals.** Spatial-block intervals in Sections 4.2–4.4 condition on the fixed model, observations and phase set. They assume exchangeable blocks and exclude training and model-selection uncertainty.
3. **Configuration and perturbation sensitivity.** Hyperparameter, blur, registration and scale tests describe responses to the evaluated conditions (Section 4.4).
4. **Real-product uncertainty at 0.75 m.** Same-scale SWIR reference data and a validated, probability-calibrated predictive distribution are unavailable.

The first three forms support reference-resolution comparisons within the controlled 60-to-20 m experiments. They do not provide pixel-level error bars for the real 0.75 m product, and no accuracy figure of the form $\pm X$ at 0.75 m is quoted for it. Real scenes

introduce registration residuals, optical-response differences, bidirectional-reflectance and terrain effects, and fine-scale heterogeneity beyond the controlled design.

Coarse observations constrain integrated SWIR responses, while RGBN imagery guides the allocation within those supports. Because the RGBN-to-SWIR relationship is nonunique, the prior selects one conditional allocation among fields compatible with the coarse observations. Fine-scale texture is therefore guided by auxiliary measurements but is not itself a measured SWIR field.

5.5. Date, Payload, Geography and the Limits of Attribution

Geography, acquisition date, JL1 payload and land-surface composition vary together across the four regions. With Aksu as the sole development region, external performance reflects their combined domain shifts. The 15 non-nested phases vary geometry on the same pixels and do not add independent dates or payloads. Separating these factors would require repeated acquisitions with controlled payload and date combinations, which the archive does not contain.

5.6. Limitations

Reference-based accuracy is limited to the controlled 60-to-20 m experiments. Real 20-to-0.75 m products are assessed through operation and coarse-observation closure; direct validation of their texture requires same-scale SWIR measurements.

The observation matrix models pixel-area support but omits the full point-spread function, modulation transfer function and spatially varying blur. Blur experiments show that closure to an assumed operator can coexist with substantial reference error. Optical-response modelling is therefore a useful extension, especially because degradation assumptions can affect method rankings [32,33].

Real-scene registration uncertainty lacks independent check points. Applied translations were of order 1–2 m, while control-point median absolute deviations ranged from 0.94–1.52 m. The ± 10 m guide perturbation is a stress test rather than an operational error estimate. The source-dependent library, standardisation and hyperparameters also remain fixed from Aksu. Yining's floor-ratio reversal illustrates a transfer limitation whose frequency and seasonal dependence cannot be determined from four regions.

The radiometric archive lacks Sentinel-2 product identifiers and processing metadata needed to verify quantification values and band-dependent offsets. JL1 inputs lack independent absolute and cross-sensor calibration. Reported errors consequently use the shared scaled-reflectance convention, without verification of absolute surface-reflectance accuracy. The retrieval-distance scan varies the distance denominator only, so geometric-tolerance sensitivity stays untested. Peak memory and the complete execution-stage provenance are also unavailable.

The external comparison establishes performance against two established baselines, adapted area-to-point regression kriging and DSen2, within the matched protocol.

6. Conclusions

We developed a support-aware inverse-allocation framework for combining high-resolution RGBN imagery with coarse dual-SWIR observations on non-nested grids. The strictly matched control identifies exact area overlap as the main source of improvement, reducing mean joint RMSE by 0.00174–0.00217 in reflectance units. Continuous correction adds approximately 0.000066–0.000166 after the area relationship is specified. These effects are separated by the strictly matched control and the A2–A4 correction contrast.

Under a matched single-observation budget, the method achieved lower mean joint RMSE than adapted ATPRK and DSen2 in all four regions, with conditional block-resampling intervals excluding zero. The area-overlap advantage persisted under tested

blur, ± 10 m guide offsets and scale ratios of 2, 3 and 6, narrowing at the largest ratio. External regions test combined geographic, date and payload shifts rather than their separate effects.

Full-scene deployment on real 0.75 m JL1 and 20 m Sentinel-2 grids produced conditional SWIR allocations with verified coarse-observation closure after writing and rereading. The results support explicit support geometry when fusing georegistered, non-nested data. Exact area overlap is therefore a design variable rather than an implementation detail whenever dual-SWIR observations are fused onto non-nested high-resolution grids, and the framework provides the accuracy accounting needed to place such a product in an application.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18203610/s1>, Section S1, extended study-area and acquisition metadata; Section S2, complete implementation settings; Section S3, full hyperparameter sensitivity tables; Section S4, full robustness tables; Section S5, prediction-archive diagnosis and correction; Section S6, per-phase metric tables; Section S7, reproducibility record; Section S8, full tables moved from the main text.

Author Contributions: Conceptualisation, Z.W., Y.O. and P.G.; methodology, Z.W.; software, Z.W.; validation, Z.W., J.L., H.Z. and H.F.; formal analysis, Z.W.; investigation, Z.W.; resources, Y.O. and P.G.; data curation, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, J.L., H.Z., H.F., Y.O. and P.G.; visualisation, Z.W.; supervision, Y.O. and P.G.; project administration, Y.O. and P.G.; funding acquisition, P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Tianshan Talent Science and Technology Innovation Team Program, grant number 2024TSYCTD0019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Sentinel-2 and Jilin-1 imagery are subject to provider licensing and are not redistributed. Retained derived products, masks, per-phase metrics, figure source values and analysis scripts are available from the corresponding author on reasonable request. The summary records support regeneration of the reported tables and figures. Full re-execution from original imagery is limited by the unrecorded items listed in Supplementary Sections S2 and S5.

Acknowledgments: The authors thank the providers of the Sentinel-2 and Jilin-1 imagery used in this study.

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

References

1. Li, Z.; Zhang, H.K.; Roy, D.P.; Yan, L.; Huang, H. Sharpening the Sentinel-2 10 and 20 m Bands to Planetscope-0 3 m Resolution. *Remote Sens.* **2020**, *12*, 2406. <https://doi.org/10.3390/rs12152406>.
2. Tayer, T.C.; Douglas, M.M.; Cordeiro, M.C.R.; Tayer, A.D.N.; Callow, J.N.; Beesley, L.; McFarlane, D. Improving the accuracy of the Water Detect algorithm using Sentinel-2, Planetscope and sharpened imagery: A case study in an intermittent river. *GISci. Remote Sens.* **2023**, *60*, 2168676. <https://doi.org/10.1080/15481603.2023.2168676>.
3. Zhang, J.; Su, R.; Ren, W.; Fu, Q.; Heide, F.; Nie, Y. A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Sci. Rep.* **2022**, *12*, 11905. <https://doi.org/10.1038/s41598-022-16223-1>.
4. Ma, Y.; Lin, S.; Wen, B. Uncertainty-Aware Hyperspectral Image Reconstruction From RGB Measurements Using Unrolled Sparse Coding. *IEEE J. Sel. Topics Signal Process.* **2025**, *19*, 1861–1872. <https://doi.org/10.1109/JSTSP.2025.3631405>.
5. Lin, Y.T.; Finlayson, G.D. Physically Plausible Spectral Reconstruction. *Sensors* **2020**, *20*, 6399. <https://doi.org/10.3390/s20216399>.
6. Brodu, N. Super-Resolving Multiresolution Images with Band-Independent Geometry of Multispectral Pixels. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4610–4617. <https://doi.org/10.1109/TGRS.2017.2694881>.
7. Yan, H.F.; Zhao, Y.Q.; Chan, J.C.W.; Kong, S.G.; Ei-Bendary, N.; Reda, M. Hyperspectral and multispectral image fusion: When model-driven meet data-driven strategies. *Inf. Fusion* **2025**, *116*, 102803. <https://doi.org/10.1016/j.inffus.2024.102803>.
8. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. <https://doi.org/10.1109/MGRS.2016.2637824>.

9. Wang, Q.; Shi, W.; Atkinson, P.M.; Zhao, Y. Downscaling MODIS images with area-to-point regression kriging. *Remote Sens. Environ.* **2015**, *166*, 191–204. <https://doi.org/10.1016/j.rse.2015.06.003>.
10. Lanaras, C.; Bioucas-Dias, J.; Baltasavias, E.; Schindler, K. Super-Resolution of Multispectral Multiresolution Images from a Single Sensor. In *Proceedings of the CVPR Workshops*; IEEE: New York, NY, USA, 2017; pp. 1505–1513. <https://doi.org/10.1109/CVPRW.2017.194>.
11. Lin, C.H.; Bioucas-Dias, J.M. An Explicit and Scene-Adapted Definition of Convex Self-Similarity Prior with Application to Unsupervised Sentinel-2 Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3352–3365. <https://doi.org/10.1109/TGRS.2019.2953808>.
12. Ulfarsson, M.O.; Palsson, F.; Dalla Mura, M.; Sveinsson, J.R. Sentinel-2 Sharpening Using a Reduced-Rank Method. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6408–6420. <https://doi.org/10.1109/TGRS.2019.2906048>.
13. Tang, Y.; Atkinson, P.M.; Zhang, J. Downscaling remotely sensed imagery using area-to-point cokriging and multiple-point geostatistical simulation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 174–185. <https://doi.org/10.1016/j.isprsjprs.2014.12.016>.
14. Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltasavias, E.; Schindler, K. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 305–319. <https://doi.org/10.1016/j.isprsjprs.2018.09.018>.
15. Aybar, C.; Contreras, J.; Donike, S.; Portalés-Julià, E.; Mateo-García, G.; Gómez-Chova, L. A radiometrically and spatially consistent super-resolution framework for Sentinel-2. *Remote Sens. Environ.* **2026**, *334*, 115222. <https://doi.org/10.1016/j.rse.2025.115222>.
16. Armannsson, S.E.; Ulfarsson, M.O.; Sigurdsson, J. A Learned Reduced-Rank Sharpening Method for Multiresolution Satellite Imagery. *Remote Sens.* **2025**, *17*, 432. <https://doi.org/10.3390/rs17030432>.
17. Ciotola, M.; Guarino, G.; Mazza, A.; Poggi, G.; Scarpa, G. A Comprehensive Benchmarking Framework for Sentinel-2 Sharpening: Methods, Dataset, and Evaluation Metrics. *Remote Sens.* **2025**, *17*, 1983. <https://doi.org/10.3390/rs17121983>.
18. Sigurdsson, J.; Armannsson, S.E.; Ulfarsson, M.O.; Sveinsson, J.R. Fusing Sentinel-2 and Landsat 8 Satellite Images Using a Model-Based Method. *Remote Sens.* **2022**, *14*, 3224. <https://doi.org/10.3390/rs14133224>.
19. Zhao, Y.; Liu, D. A robust and adaptive spatial-spectral fusion model for PlanetScope and Sentinel-2 imagery. *GISci. Remote Sens.* **2022**, *59*, 520–546. <https://doi.org/10.1080/15481603.2022.2036054>.
20. Sarkar, S.; Sagan, V.; Bhadra, S.; Fritsch, F.B. Spectral enhancement of PlanetScope using Sentinel-2 images to estimate soybean yield and seed composition. *Sci. Rep.* **2024**, *14*, 15063. <https://doi.org/10.1038/s41598-024-63650-3>.
21. Tobler, W.R. Smooth Pycnophylactic Interpolation for Geographical Regions. *J. Amer. Stat. Assoc.* **1979**, *74*, 519–530. <https://doi.org/10.1080/01621459.1979.10481647>.
22. Jin, Y.; Ge, Y.; Wang, J.; Heuvelink, G.B.M.; Wang, L. Geographically Weighted Area-to-Point Regression Kriging for Spatial Downscaling in Remote Sensing. *Remote Sens.* **2018**, *10*, 579. <https://doi.org/10.3390/rs10040579>.
23. Deng, S.; Ma, J.; Deng, L.J.; Wei, P. OTIAS: OcTree Implicit Adaptive Sampling for Multispectral and Hyperspectral Image Fusion. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 2708–2716. <https://doi.org/10.1609/aaai.v39i3.32275>.
24. Zhu, C.; Zhang, T.; Wu, Q.; Li, Y.; Zhong, Q. An Implicit Transformer-based Fusion Method for Hyperspectral and Multispectral Remote Sensing Image. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103955. <https://doi.org/10.1016/j.jag.2024.103955>.
25. Liang, Y.J.; Cao, Z.; Deng, S.; Dou, H.X.; Deng, L.J. Fourier-enhanced Implicit Neural Fusion Network for Multispectral and Hyperspectral Image Fusion. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc. (NeurIPS): San Diego, CA, USA, 2024; Volume 37, pp. 63441–63465. <https://doi.org/10.52202/079017-2026>.
26. Wang, T.; Yan, Z.; Li, J.; Zhao, X.; Wang, C.; Ng, M. Hyperspectral and Multispectral Image Fusion with Arbitrary Resolution Through Self-Supervised Representations. *Int. J. Comput. Vis.* **2025**, *133*, 7515–7535. <https://doi.org/10.1007/s11263-025-02540-1>.
27. Wu, R.C.; Deng, S.; Ran, R.; Dou, H.X.; Deng, L.J. INF3: Implicit Neural Feature Fusion Function for Multispectral and Hyperspectral Image Fusion. *IEEE Trans. Comput. Imaging* **2024**, *10*, 1547–1558. <https://doi.org/10.1109/TCI.2024.3488569>.
28. Lin, H.; Huang, D.; Li, J.; Zhou, T.; Long, J.; Peng, Y. Semi-blind hyperspectral and multispectral image fusion via generalized inverse matrix optimization. *Infrared Phys. Technol.* **2024**, *139*, 105316. <https://doi.org/10.1016/j.infrared.2024.105316>.
29. Qu, J.; Wu, X.; Dong, W.; Cui, J.; Li, Y. IR&ArF: Toward Deep Interpretable Arbitrary Resolution Fusion of Unregistered Hyperspectral and Multispectral Images. *IEEE Trans. Image Process.* **2025**, *34*, 1934–1949. <https://doi.org/10.1109/TIP.2025.3551531>.
30. Lin, C.H.; Hsu, C.C.; Young, S.S.; Hsieh, C.Y.; Tai, S.C. QRCODE: Quasi-Residual Convex Deep Network for Fusing Misaligned Hyperspectral and Multispectral Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5512215. <https://doi.org/10.1109/TGRS.2024.3378849>.
31. Xu, H.; Yang, J.; Lin, T.; Liu, J.; Liu, F.; Xiao, L. Hyperspectral Reconstruction From RGB Images via Physically Guided Graph Deep Prior Learning. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5533614. <https://doi.org/10.1109/TGRS.2024.3471616>.

32. Albanwan, H.; Qin, R.; Tang, Y. Image Fusion in Remote Sensing: An Overview and Meta-Analysis. *Photogramm. Eng. Remote Sens.* **2024**, *90*, 755–775. <https://doi.org/10.14358/PERS.24-00110R1>.
33. Armannsson, S.E.; Ulfarsson, M.O.; Sigurdsson, J.; Nguyen, H.V.; Sveinsson, J.R. A Comparison of Optimized Sentinel-2 Super-Resolution Methods Using Wald’s Protocol and Bayesian Optimization. *Remote Sens.* **2021**, *13*, 2192. <https://doi.org/10.3390/rs13112192>.

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.