

Article

Preserving Spatial Fidelity in Heterogeneous Landscapes: A Biophysical Index-Guided 1×1 CNN for Multi-Source Remote Sensing Fusion

Yanru Pei ^{1,2} and Pengchong Wang ^{2,3,*}¹ College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China² College of Forestry, Beijing Forestry University, Beijing 100083, China³ Beijing Linmiao Ecological Environment Technology Co., Ltd., Beijing 100083, China

* Correspondence: 13897923135@163.com

Highlights

What are the main findings?

- A physics-constrained 1×1 CNN reconstructs NPP on a standardized baseline grid while avoiding cross-pixel spatial smoothing.
- The point-wise architecture improves local boundary fidelity relative to RF and standard 3×3 CNN baselines in fragmented open-pit mining landscapes.

What are the implications of the main findings?

- Channel-wise fusion of VMSI, AVEI, SAVEI, AAI-VHI, and WET provides an interpretable GeoAI pathway for long-term ecological monitoring.
- Boundary-preserving NPP reconstruction can support ecological restoration assessment in highly heterogeneous mining regions without treating the task as spatial downscaling.

Abstract

Continuous spatial reconstruction in highly fragmented landscapes remains a persistent challenge in remote sensing and geographic information systems. Standardized products such as MODIS Net Primary Productivity (NPP) provide temporally consistent ecological baselines, but their moderate spatial resolution can obscure abrupt transitions in disturbed environments. This study develops a biophysical index-guided 1×1 Convolutional Neural Network (CNN) framework for NPP reconstruction over a mega-scale open-pit mining landscape. The framework uses five Landsat-derived biophysical tensors—Vegetation Moisture Stress Index (VMSI), Anti-Vegetation Environment Index (AVEI), Soil Adjusted Vegetation Environment Index (SAVEI), AAI-VHI, and Tasseled Cap Wetness (WET)—as point-wise predictors aligned to the MODIS NPP baseline grid. By restricting convolutional kernels to 1×1 , the model performs nonlinear channel-wise mapping without incorporating neighboring grid cells, thereby reducing boundary mixing that can occur in conventional multi-pixel CNNs. Benchmark comparisons with Random Forest and a standard 3×3 CNN showed that the 3×3 CNN achieved slightly higher global accuracy, whereas the 1×1 CNN provided stronger gradient correspondence and lower full-domain spatial error in the 2020 spatial fidelity assessment. The results indicate that point-wise convolution guided by physically interpretable indices provides a conservative and

Academic Editor: Parth Sarathi Roy

Received: 28 May 2026

Revised: 3 July 2026

Accepted: 6 July 2026

Published: 18 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

interpretable option for standardized ecological reconstruction at the MODIS baseline grid scale.

Keywords: Convolutional Neural Network (CNN); Landsat; MODIS NPP; Net Primary Productivity (NPP); open-pit coal mine

1. Introduction

Continuous spatiotemporal monitoring of the Earth's surface is a fundamental requirement in geographic information systems (GISs), remote sensing, and environmental modeling [1]. Landsat Collection 2 surface reflectance archives provide radiometrically corrected multispectral inputs for constructing annual biophysical predictors [2,3]. Global standardized products such as the MODIS Net Primary Productivity (NPP) dataset provide temporally calibrated ecological baselines, but their moderate spatial resolution can smooth local transitions in fragmented landscapes [4–7]. Rather than reconstructing a pre-MODIS record, this study focuses on reconstructing standardized annual NPP patterns during the MODIS-era observation period while using Landsat-derived biophysical information to improve spatial representation within a common baseline grid.

Previous NPP estimation and monitoring studies provide useful guidance for interpreting this problem. Comparisons between MODIS-derived NPP and inventory- or region-based estimates show that standardized products are valuable for temporal consistency, but their spatial support and land-cover mixture can introduce discrepancies that should be made explicit during model interpretation [4]. Regional NPP studies further indicate that water availability, anthropogenic activity, and disturbance intensity can dominate vegetation productivity change in dryland or strongly managed landscapes [5,6]. In mining areas, recent NPP and ecological quality assessments show that excavation, overburden dumping, reclamation, and vegetation recovery produce heterogeneous productivity responses rather than a single smooth regional trend [7–27]. These studies suggest three methodological requirements for NPP estimation in open-pit mining landscapes: the target grid scale must be stated clearly, predictors should represent vegetation condition, moisture, exposed-surface disturbance, and restoration status, and model evaluation should include spatial fidelity diagnostics in addition to global accuracy metrics.

In extremely fragmented environments, transitions among active mining pits, bare overburden dumps, restored vegetation patches, and undisturbed reference areas are abrupt and nonlinear. Global and regional remote sensing studies have shown that surface coal-mining disturbance and reclamation can be mapped from Landsat and other satellite records at annual to multi-decadal scales [8–12]. Ecological cumulative effect, reclamation, and remote sensing ecological index studies further indicate that vegetation recovery and environmental quality vary strongly across mining mosaics [13–16]. Traditional empirical models and machine learning algorithms, such as Random Forest (RF), are widely used for spatial regression, but they may produce patchy predictions when the input features do not adequately represent underlying biophysical gradients. This limitation is especially important in boundary zones, where spectrally similar surfaces may have contrasting ecological functions.

Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have gained traction in remote sensing and geospatial data fusion [28–37]. Standard CNNs employ multi-pixel receptive fields (e.g., 3×3 kernels) to leverage spatial context and often improve global statistical metrics; this interpretation draws on spatial autocorrelation theory and boundary-aware segmentation [38–40]. This neighborhood aggrega-

tion is valuable for spatially continuous processes, but in highly fragmented open-pit mining landscapes it can also mix adjacent grid cells that represent active pits, overburden dumps, restored vegetation, and undisturbed reference areas. In this manuscript, 'spurious spatial autocorrelation' refers to this model-induced neighborhood smoothing when it increases local similarity across abrupt functional boundaries without corresponding evidence in the target grid. Multi-source spatiotemporal fusion studies have also demonstrated the value of survey-based synthesis and hybrid deep learning frameworks [41–44], with subsequent multitask, attention, transformer, geological remote sensing, and decade-level review studies extending this literature [45–50].

To address this issue, this study proposes a biophysical index-guided 1×1 CNN framework for standardized NPP reconstruction in a highly disturbed open-pit mining landscape. Recent automated disturbance-mapping and exposed-coal index studies further demonstrate that coal mine surface units can be identified from Landsat-based spectra and autonomous sample generation workflows, but these advances do not directly solve standardized NPP reconstruction [17–19]. The model restricts the convolutional receptive field to 1×1 and therefore performs point-wise nonlinear mapping across engineered biophysical tensor channels. The objectives are: (1) to evaluate the predictive performance of the 1×1 CNN against Random Forest and a standard 3×3 CNN at a unified MODIS baseline grid resolution; (2) to quantify whether point-wise convolution reduces boundary zone error and preserves gradient structure in fragmented mining landscapes; and (3) to examine residual patterns and tensor contributions using spatial fidelity metrics and permutation-based attribution [38–40].

2. Study Area and Data

2.1. The Mega-Scale Open-Pit Mine as a Spatial Testbed

The empirical testbed comprises the Heidaigou and Haerwusu open-pit mining complexes in Ordos, Inner Mongolia, China. The area lies in a temperate continental semi-arid climate zone where water limitation, bare overburden, excavation pits, and staged vegetation restoration generate sharp ecological gradients. The site was selected because these abrupt transitions provide a demanding testbed for evaluating spatial fidelity at a standardized NPP baseline grid scale. Comparable mining area studies have emphasized remote sensing indicators for reclamation and vegetation recovery [20–24], while ecological quality and landscape-risk studies highlight the need to represent spatially heterogeneous disturbance effects [25–27].

The modeling domain covers approximately 869 km² of valid MODIS grid cells within the broader analysis extent of about 1201 km². The landscape includes active mining pits, overburden dumps, restored vegetation patches, and surrounding reference areas. The ROI boundary defines the complete standardized analysis domain used for model training and evaluation, whereas the mining complex ROI identifies the focused disturbance window used for later local spatial fidelity visualization. The local window was selected algorithmically from high-gradient 2020 reference NPP cells to reduce subjective window selection.

Figure 1 presents the geographic setting of the study area using an administrative boundary base map. The ROI boundary delineates the standardized analysis domain surrounding the Heidaigou–Haerwusu open-pit mining complex, while the mining complex ROI marks the focused disturbance window used for local spatial fidelity visualization.

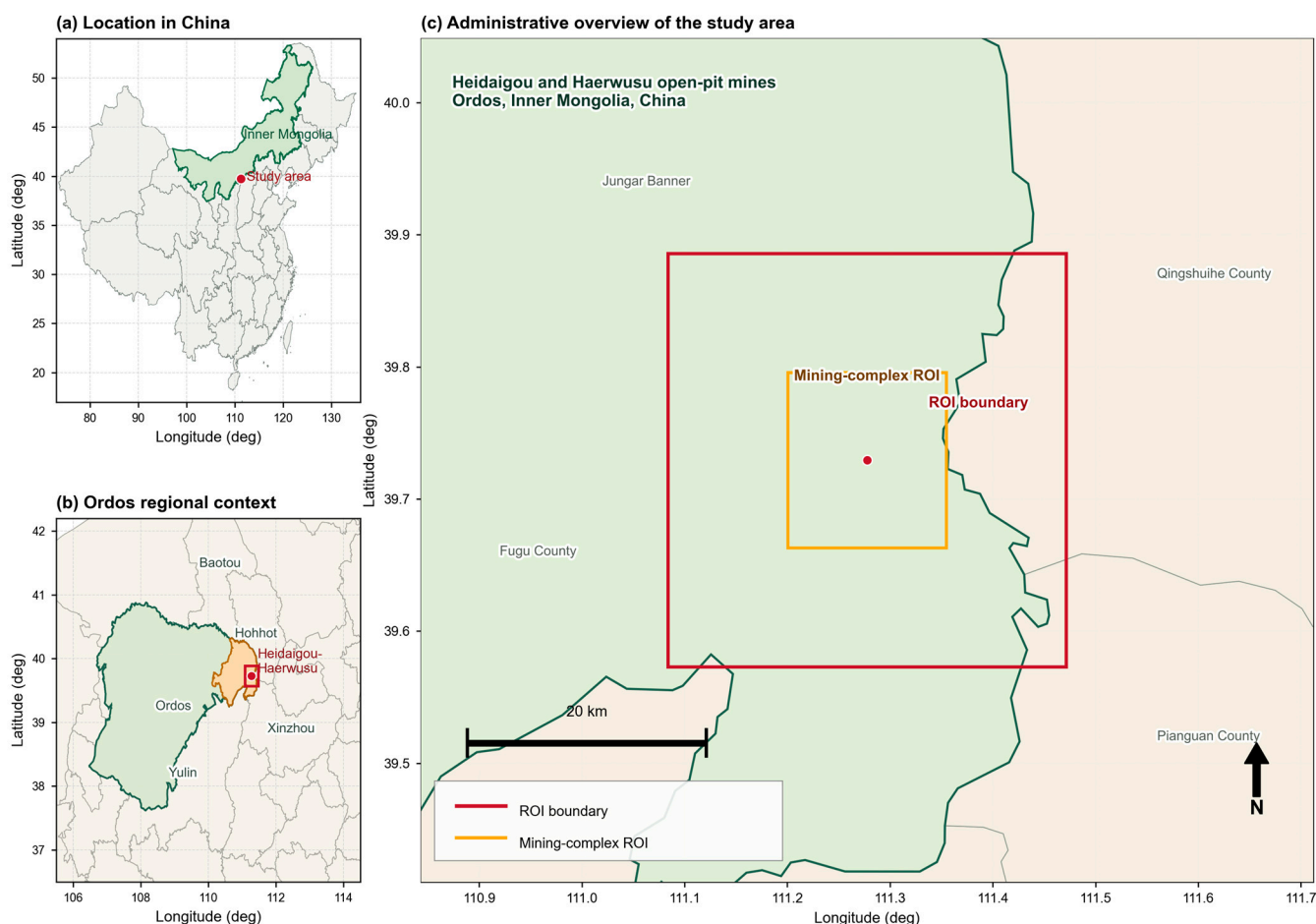


Figure 1. Study area overview. (a) Location of the study area within China; (b) regional context within Ordos, Inner Mongolia; and (c) administrative boundary overview of the Heidaigou–Haerwusu open-pit mining complex. The red rectangle indicates the ROI boundary, and the orange rectangle indicates the mining complex ROI.

2.2. Multi-Source Geospatial Data Acquisition and Radiometric Rigor

The empirical implementation assembled annual Landsat-derived biophysical tensors and MODIS NPP targets for 21 available years from 2000 to 2021, excluding 2012. The omission of 2012 was applied consistently because the Landsat 5-to-Landsat 8 transition created a temporal discontinuity in the annual Landsat composite sequence used by this study; Landsat 7 ETM+ was not used as a primary data stream for model training. Across the 21 years, the final modeling stack contained 94,878 valid point-wise samples for the 1×1 CNN and 86,856 valid 3×3 patch samples for the neighborhood baseline after edge exclusion.

(1) Fine-Scale Spectral Primaries and Atmospheric Correction

Cloud-reduced multispectral surface reflectance matrices were ingested from Landsat 5 Thematic Mapper (TM) for 2000–2011 and Landsat 8 Operational Land Imager (OLI) for 2013–2021. All scenes were sourced from the Landsat Collection 2, Tier 1, Level-2 archive, which provides systematically corrected surface reflectance products. We acknowledge that Collection 2 processing improves radiometric consistency but does not remove all cross-sensor spectral response differences between TM and OLI; this uncertainty is explicitly discussed as a limitation.

Atmospheric artifacts were suppressed using the specialized Land Surface Reflectance Code (LaSRC) algorithm for Landsat OLI, and the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) model for Landsat TM. These algorithms execute

rigorous gaseous absorption and aerosol scattering corrections based on auxiliary atmospheric characterization datasets [2,3].

Pixels flagged as cloud, cloud shadow, or high-confidence cirrus by the QA_PIXEL bitmask were excluded, and annual median compositing during the peak growing season was used to generate cloud-reduced spectral baselines. The year 2012 was omitted from all model training, validation, and temporal display sequences to maintain consistency during the Landsat 5-to-Landsat 8 sensor transition interval.

(2) Coarse Temporal Baseline Target

The MODIS annual Net Primary Productivity product (MOD17A3HGF, Version 6.1) served as the standardized baseline target. MOD17A3HGF provides annual NPP at a nominal 500 m sinusoidal grid scale; after reprojection and clipping in this workflow, the common modeling grid used EPSG:4326 with a pixel size of 0.004491576 degrees, corresponding to approximately 500 m at the latitude of the study area. The target should therefore be interpreted as a MODIS-scale standardized NPP baseline rather than a fine-resolution ground NPP.

2.3. Coordinate Homogenization and Grid Alignment Protocol

All multispectral bands, engineered biophysical tensors, and MODIS NPP targets were co-registered to the common MODIS baseline grid used in the released raster stack. The final grid is stored in EPSG:4326 with 88 columns and 71 rows for the 2020 clipped study extent, a pixel size of 0.004491576 degrees, and bounds of 111.0812–111.4764E and 39.5708–39.8897N.

Fine-scale Landsat-derived tensor layers were resampled or aggregated to the target NPP grid using bilinear resampling when dimensions differed from the MODIS grid. This procedure ensures that every predictor vector and NPP target share the same spatial support. Consequently, all boundary fidelity results in this paper refer to local gradients and boundary zones within the MODIS baseline grid; the framework is not presented as sub-MODIS downscaling or super-resolution.

3. Materials and Methods

The geospatial data fusion framework consists of two components: (1) engineering and grid alignment of biophysical index tensor predictors that describe localized vegetation, moisture, disturbance, and wetness conditions, and (2) a point-wise 1×1 CNN designed to map these predictors to the standardized NPP target while limiting cross-boundary information mixing.

3.1. Physical Tensor Engineering for Spatial Constraints

Five tensors were engineered from multispectral and thermal/atmospheric inputs before grid alignment. Prior Landsat-based exposed-coal and opencast-coal indices show that spectral and thermal combinations can help separate coal-bearing or mining-disturbed surfaces from surrounding vegetation and bare soils [18,19]. The tensor design also follows the broader use of Tasseled Cap Wetness [51], soil-adjusted vegetation response [52], and vegetation–temperature or drought–health indices as interpretable remote sensing predictors [53–57]. In this study, the engineered tensors are physically interpretable predictors rather than hard physical constraints in the loss function. Thus, the framework is described as biophysical index-guided or physics-aware, not as a physics-constrained neural network in the strict sense used for conservation-law or physics-consistency-loss models [28–32].

(1) Vegetation Moisture Stress Index (VMSI) Tensor

In semi-arid disturbed landscapes, water availability is the primary limiting factor causing sharp spatial gradients. We formulated the VMSI tensor to capture these acute

moisture stress transitions mathematically. The VMSI is computed using a multi-band spectral transformation (Equation (1)):

$$VMSI = 4 \times (\rho_{Green} - \rho_{NIR}) - (0.25 \times \rho_{Red} + 2.75 \times \rho_{SWIR1}), \quad (1)$$

$\rho_{Green}\rho_{NIR}\rho_{Red}\rho_{SWIR1}$, where Green, Red, NIR, and SWIR1 denote surface reflectance in the green, red, near-infrared, and first short-wave infrared channels of the custom GEE Landsat stack. The coefficients are not adopted directly from a published index; they were retained from exploratory AWEI-sh-inspired band-combination trials as empirical weights that enhanced the contrast between green/NIR vegetation response and red/SWIR1 soil moisture or exposed-surface response in the study area. The VMSI tensor is therefore empirical and unitless; it is standardized by Z-score normalization before model fitting, and its coefficients should be recalibrated before transfer to other landscapes.

(2) Anti-Vegetation Environment Index (AVEI) Tensor

AVEI was used to represent environmental stress by coupling vegetation vigor with thermal/atmospheric adjustment terms (Equation (2)). NDVI is defined as $(NIR - Red)/(NIR + Red)$, VAF is the vegetation adjustment factor calculated as $0.004 \times NDVI + 0.986$, TAF is the thermal/atmospheric adjustment factor calculated as $STATRAN \times 0.0001 + 0.1$, and $AVEI = TAF/[1 + TAF \times \ln(VAF)]$. In this workflow, STATRAN denotes the Landsat atmospheric transmittance adjustment band extracted in the custom GEE Landsat stack; the 0.0001 factor applies the scale conversion used in the stack, and the 0.1 offset keeps TAF positive for logarithmic transformation. Values of $VAF \leq 0$ were replaced by a small positive constant before the logarithm to avoid invalid operations.

$$AVEI = \frac{TAF}{1 + TAF \times \ln(VAF)}. \quad (2)$$

Together, VMSI and AVEI provide the neural network with disturbance- and moisture-sensitive predictors that help separate functionally different land covers within the standardized grid. They do not impose a separate physical constraint term; rather, they guide the point-wise mapping through interpretable input channels.

(3) Soil Adjusted Vegetation Environment Index (SAVEI) Tensor [51,52]

SAVEI is not the classical SAVI. In this study, SAVEI denotes a normalized stress ratio tensor derived from AVEI and TAF, calculated as $SAVEI = (AVEI - TAF)/(AVEI + TAF)$ (Equation (3)). This ratio places environmental stress information on an approximately bounded normalized scale and reduces the dominance of raw thermal/atmospheric magnitude during neural network optimization.

$$SAVEI = \frac{AVEI - TAF}{AVEI + TAF}. \quad (3)$$

Because SAVEI is a normalized difference form, it is unitless and approximately bounded near $[-1, 1]$ when the denominator is positive. Denominators equal to zero were handled by safe division. The index is used only as a model predictor and should not be confused with Huete's soil-adjusted vegetation index (SAVI) [52].

(4) Comprehensive Vegetation Health Index (AAI-VHI) Tensor

AAI-VHI fuses vegetation condition and environmental stress. VCI is calculated as $100 \times (NDVI - NDVI_{min})/(NDVI_{max} - NDVI_{min})$, while the stress component is calculated as $100 \times (SAVEI_{max} - SAVEI)/(SAVEI_{max} - SAVEI_{min})$. AAI-VHI is the sum of these two normalized components and is therefore approximately bounded in $[0, 200]$ before standardization. In this implementation, min-max bounds were computed within the available raster stack used for tensor generation.

(5) Tasseled Cap Wetness (WET) Tensor [51]

Tasseled Cap Wetness (WET) was integrated as a complementary indicator of soil and canopy moisture content [51]. Together, VMSI, AVEI, SAVEI, AAI-VHI, and WET form the five-channel predictor space used by the 1×1 CNN implementation. All tensors

were transformed to the common target grid and Z-score-standardized using training split statistics before model fitting.

3.2. The Biophysical Index-Guided 1×1 CNN Architecture

The methodological role of the 1×1 convolution is to perform nonlinear channel fusion independently at each grid cell. A 3×3 convolution integrates the eight neighboring grid cells around the target cell, which can be beneficial when the response is spatially smooth. In fragmented open-pit mining landscapes, however, this neighborhood operation may blend functionally distinct adjacent cells. The 1×1 model was therefore used as a deliberately point-wise architecture for testing whether excluding neighboring cells improves local spatial fidelity at the MODIS baseline grid scale, drawing on CNN-based remote sensing modeling [33–37] and spatial boundary analysis concepts [38–40].

To reduce spatial blurring caused by cross-pixel convolution, the proposed architecture restricts the convolutional receptive field to 1×1 . Let the input feature tensor for a given scene at the unified baseline resolution be defined as $\mathbf{X} \in \mathbf{R}^{H \times W \times C}$, where H and W represent grid height and width, and $C = 5$ denotes the five engineered tensor channels: VMSI, AVEI, SAVEI, AAI-VHI, and WET. The input array was normalized using Z-score standardization based on the training data to improve numerical stability and gradient scaling (Equation (4)):

$$\hat{\mathbf{X}}_c = \frac{\mathbf{X}_c - \mu_c}{\sigma_c}, \quad (4)$$

where μ_c and σ_c represent the empirical mean and standard deviation calculated across the training split for channel c .

The network consists of a sequential arrangement of 1×1 2D convolutional layers. The core mathematical transformation within the first hidden layer is formulated as Equation (5):

$$\mathbf{H}_{x,y,k}^{(1)} = \text{ReLU} \left(\sum_{c=1}^C \mathbf{W}_{1,1,c,k}^{(1)} \cdot \hat{\mathbf{X}}_{x,y,c} + b_k^{(1)} \right), \quad (5)$$

where $\mathbf{W}_{1,1,c,k}^{(1)}$ represents the weights of the k -th filter mapping the c -th input channel to a 1×1 spatial extent, $b_k^{(1)}$ is the localized bias parameter, and $\{\text{ReLU}\}(z) = \max(0, z)$ is the Rectified Linear Activation function applied element-wise to break linearity.

$(x \pm \Delta x, y \pm \Delta y)$. With 1×1 kernels, the network does not draw information from neighboring spatial coordinates during convolution. Instead, it performs nonlinear mapping along the channel dimension for each grid cell, which is functionally similar to applying a shared multilayer perceptron to every grid cell feature vector. The advantage of the CNN implementation is weight sharing over raster tensors, direct compatibility with raster inference workflows, and a controlled comparison against the 3×3 spatial CNN baseline. Future work should compare this design with standalone MLP, XGBoost, SVR, and GAM baselines.

Figure 2 summarizes the methodological framework and the internal structure of the proposed point-wise 1×1 CNN. The five engineered biophysical tensor channels are organized as a channel vector for each standardized grid cell. The neural pathway then applies 1×1 convolutional transformations, dropout regularization, a dense nonlinear mapping, and a final linear output layer to estimate NPP for the same grid cell. Because the convolutional receptive field remains 1×1 throughout the spatial pathway, no neighboring grid cells are aggregated during feature extraction.

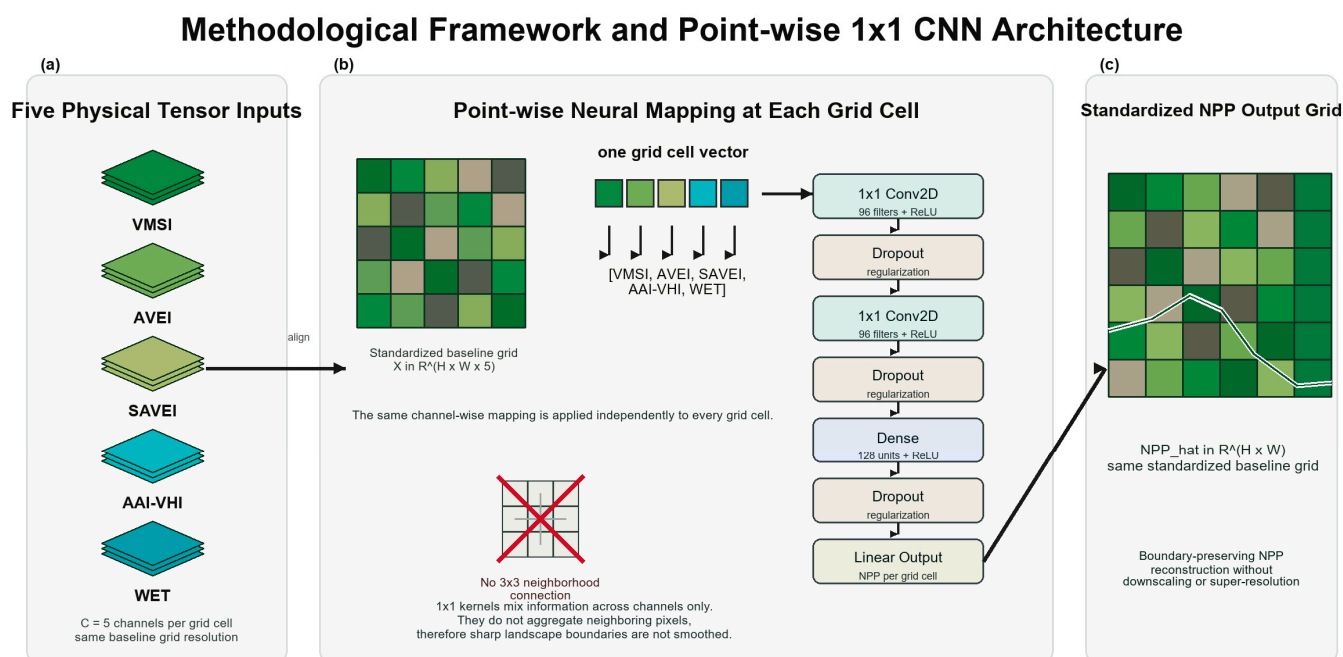


Figure 2. Methodological framework and network architecture of the biophysical index-guided 1×1 CNN. (a) Five physical tensor inputs (VMSI, AVEI, SAVEI, AAI-VHI, and WET) are aligned to the same standardized baseline grid; (b) point-wise neural mapping applies the same channel-wise 1×1 Conv2D, dropout, dense, and linear output operations independently to each grid cell; and (c) the output is a standardized NPP grid. The arrows indicate data flow from tensor inputs through the neural mapping pathway to the NPP output, and the crossed 3×3 grid denotes that no neighboring pixels are aggregated during feature extraction.

3.3. Algorithmic Optimization and Hyperparameter Configuration

The 1×1 CNN architecture was optimized using Bayesian hyperparameter tuning implemented with Keras 3.5.0 and the Keras Tuner BayesianOptimization tuner. The search objective minimized validation MSE using an internal validation split during training. The search space included filters in {32, 64, 96, 128} for each 1×1 Conv2D layer, dense units in {32, 64, 96, 128}, dropout rates from 0.2 to 0.5 at 0.1 intervals, a maximum of 10 Bayesian trials, 50 epochs during tuning, and an 80/20 validation split within the training data.

The selected 1×1 CNN contained two 1×1 Conv2D layers with 96 filters each, 1×1 max-pooling layers that do not expand the receptive field, dropout rates of 0.2 and 0.4 after the convolutional blocks, a dense layer with 128 units, a final dropout rate of 0.2, and a linear output layer. The model was trained with Adam, MSE loss, a batch size of 32 under the original Keras default unless otherwise specified, early stopping patience of 10 epochs, restore_best_weights = True, and random seed 42 for the train–test split. The final model contains 22,433 trainable parameters.

The objective loss function selected for network backpropagation was the Mean Squared Error (MSE), defined mathematically as Equation (6):

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (6)$$

Model optimization was conducted using the Adam (Adaptive Moment Estimation) optimizer. To guarantee generalization stability and prevent computational over-fitting, the training runtime was constrained by an early stopping mechanism. The patience threshold was strictly set to 10 epochs (patience = 10); if the validation loss failed to exhibit

a structural decrease over 10 consecutive iterations, the training loop was automatically terminated, and the optimal model weights were restored (restore_best_weights = True).

For benchmarking, the proposed model was compared with two baseline algorithms under the released supplementary implementation. The dataset was split into training and testing subsets using an 80/20 random split with a fixed random seed of 42. Model selection used an internal validation split and early stopping. Because random splitting can overestimate performance for spatiotemporal raster data with repeated locations, this design is treated as a baseline evaluation rather than a fully spatially independent validation; this limitation is discussed explicitly in Section 5.3.

Baseline 1 (Random Forest): A pure machine learning regressor parameterized with 100 independent decision trees ($n_{\text{estimators}} = 100$) and restricted to a maximum depth of 15 ($\text{max_depth} = 15$) to control tree complexity [58].

Baseline 2 (Standard 3×3 CNN): A conventional spatial deep learning network featuring multi-pixel receptive fields. This architecture was designed with two 3×3 Conv2D layers (64 filters each) and dual dropout layers (rate = 0.3), and was optimized over 100 epochs using a batch size of 256 [33–37,40].

The benchmarking protocol isolates the impact of the spatial receptive field (1×1 vs. 3×3) while keeping the same MODIS-scale target. The comparison therefore evaluates the trade-off between global statistical convergence and local boundary fidelity, rather than claiming universal superiority of one architecture.

4. Results

4.1. Global Statistical Performance and Accuracy Assessment

The overall predictive accuracy of the biophysical index-guided 1×1 CNN was evaluated against Random Forest (RF) and a standard 3×3 CNN baseline under the same train–test protocol. Table 1 summarizes the global performance metrics and computational characteristics of the three models. The standard 3×3 CNN achieved the highest global statistical accuracy ($R^2 = 0.8963$, RMSE = 176.42), closely followed by RF ($R^2 = 0.8938$, RMSE = 178.48). The 1×1 CNN yielded a slightly lower global coefficient of determination ($R^2 = 0.8766$) and a higher RMSE (191.90), while using fewer trainable parameters than the 3×3 CNN. These results indicate that global error minimization and local spatial fidelity should be interpreted as complementary objectives [58].

Table 1. Global performance metrics of RF, standard 3×3 CNN, and biophysical index-guided 1×1 CNN.

Model	R^2	RMSE	MAE	Params	Inference Time (s)
Random Forest	0.8938	178.48	132.29	N.A.	0.06
Standard 3×3 CNN	0.8963	176.42	135.13	44,097	0.79
Biophysical index-guided 1×1 CNN	0.8766	191.90	144.73	22,433	1.38

Note: RF and 3×3 CNN MAE values were recovered by re-evaluating the archived baseline models with the original 3×3 patch construction, 80/20 random split, and random seed 42. The 1×1 CNN inference time was newly timed over five runs on the saved model and scaler; archived RF and 3×3 inference times are retained from the original baseline report.

To further diagnose model behavior, predicted NPP values were plotted against reference NPP values for the three algorithms (Figure 3). The scatter distributions broadly follow the 1:1 reference line, indicating that none of the models exhibits severe systematic overestimation or underestimation. However, scatter plots and global metrics do not explicitly reveal whether abrupt geomorphological and ecological boundaries are preserved.

Therefore, the subsequent analysis focuses on spatial fidelity and residual organization in highly fragmented mining landscapes [38–40,59–62].

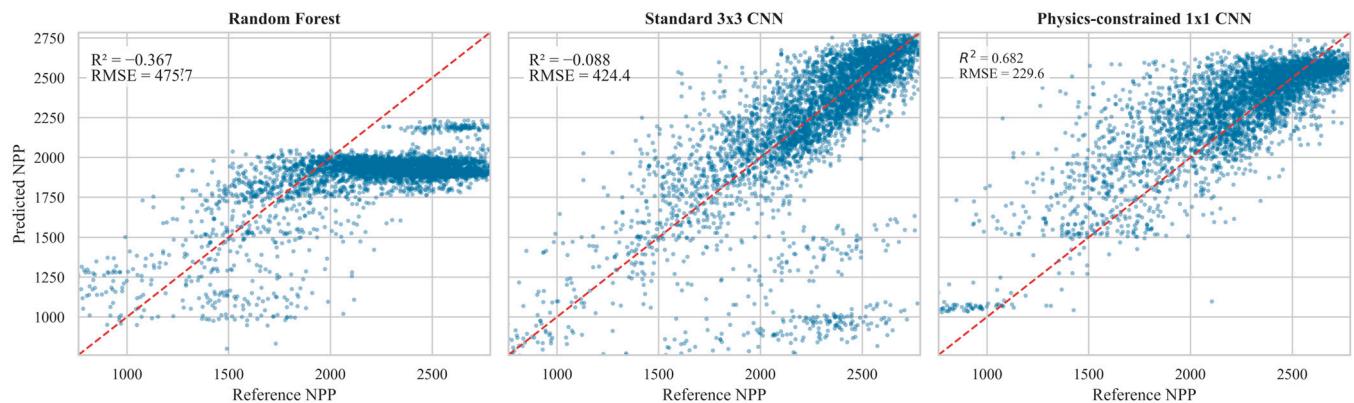


Figure 3. Accuracy assessment scatter plots comparing predicted and reference NPP values for RF, standard 3×3 CNN, and the biophysical index-guided 1×1 CNN. The dashed red line represents the 1:1 reference line.

4.2. Assessment of Local Spatial Fidelity

To assess spatial blurring and the preservation of geomorphological boundaries, a local subset with extreme spatial heterogeneity was extracted for visual comparison (Figure 4). The window was selected from the 2020 reference NPP field using maximum local gradient magnitude, thereby focusing on an abrupt transition zone without manually choosing a visually favorable example. The comparison includes the reference NPP field, RF predictions, standard 3×3 CNN predictions, and 1×1 CNN predictions over the same spatial window [38–40,59–62].

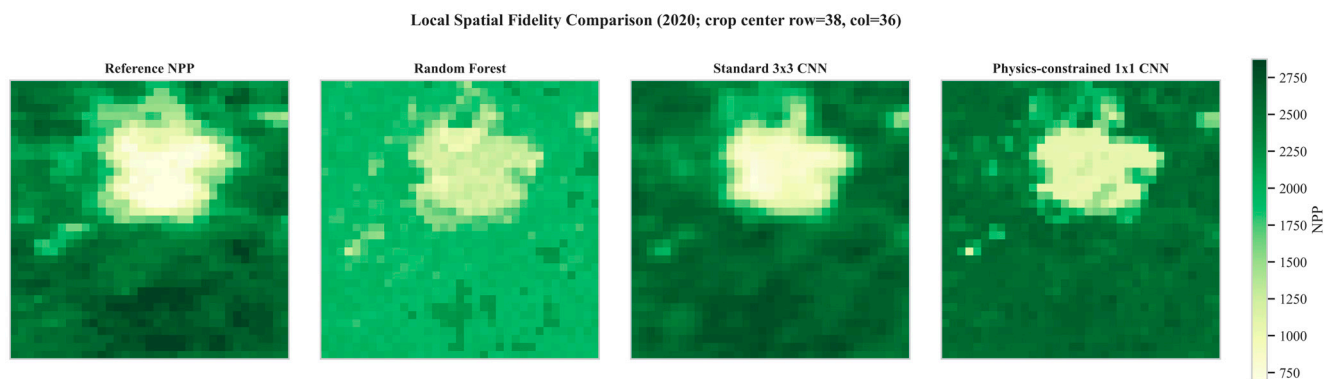


Figure 4. Local spatial fidelity comparison for the 2020 reference NPP field and predictions from RF, standard 3×3 CNN, and biophysical index-guided 1×1 CNN. The selected crop highlights abrupt transitions between disturbed and restored grid cells.

The visual comparison reveals distinct algorithmic behaviors. RF predictions exhibit a patchy, pixelated distribution, reflecting its cell-wise decision structure. The standard 3×3 CNN produces smoother transitions across the local boundary zone; this behavior can improve global error statistics but may also blend sharply contrasting land-cover units into intermediate values. In contrast, the 1×1 CNN reduces cross-boundary smoothing by mapping each grid cell through its own five-channel tensor vector. This result should be interpreted at the MODIS baseline grid scale, not as evidence of fine-resolution surface boundary reconstruction.

4.3. Spatial Distribution of Prediction Residuals

Spatial residual heatmaps were generated as the difference between predicted and reference NPP values to examine whether prediction errors were randomly distributed or structurally concentrated along landscape boundaries (Figure 5). In addition, quantitative spatial fidelity metrics were calculated for the 2020 full-domain raster comparison (Table 2). Boundary zones were defined as the top 10% of reference NPP gradient magnitude among valid 2020 cells, producing 452 boundary zone cells.

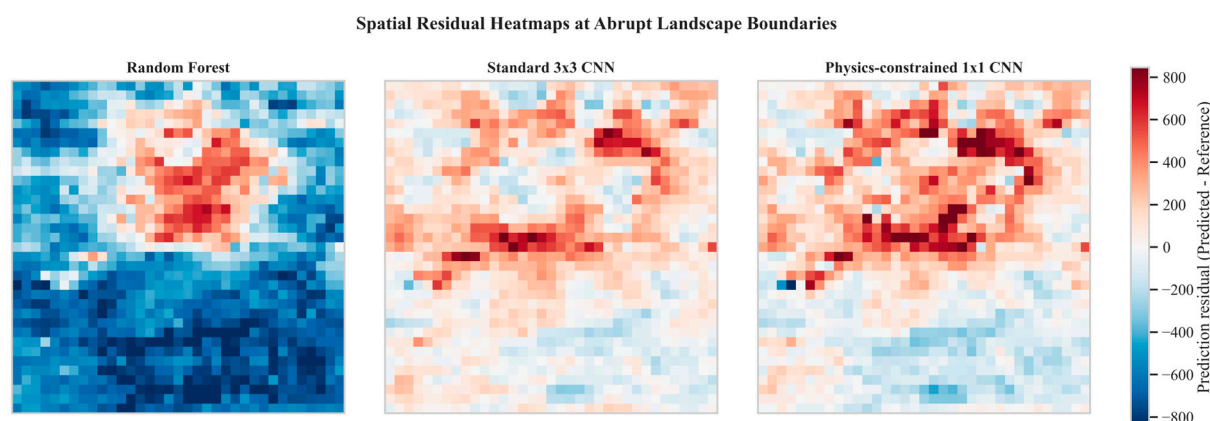


Figure 5. Spatial residual heatmaps for RF, standard 3×3 CNN, and biophysical index-guided 1×1 CNN. Residuals are calculated as predicted NPP minus reference NPP for the same local window used in Figure 4.

The residual maps show that the standard 3×3 CNN tends to produce organized residual structures around abrupt boundary zones, consistent with the smoothing effect introduced by multi-pixel receptive fields. The full-domain 2020 spatial fidelity metrics further show that the 1×1 CNN had the lowest full-domain RMSE (229.62) and MAE (169.32), the highest gradient correlation with the reference NPP field (0.611), and lower boundary zone error than the 3×3 CNN (boundary RMSE 376.07 versus 515.40; boundary MAE 297.35 versus 356.62). These results support the interpretation that point-wise mapping can improve local spatial fidelity even though the 3×3 CNN achieved slightly better random-split global accuracy.

Table 2. Quantitative spatial fidelity metrics for the 2020 full-domain raster comparison. Boundary zone cells are the top 10% of reference NPP gradient magnitude among valid cells.

Model	Full RMSE	Full MAE	Boundary RMSE	Boundary MAE	<i>n</i>	Gradient Corr.	Residual Moran's I
Random Forest	475.70	416.32	339.95	274.08	452	0.503	0.842
Standard 3×3 CNN	424.38	231.22	515.40	356.62	452	0.299	0.523
Biophysical index-guided 1×1 CNN	229.62	169.32	376.07	297.35	452	0.611	0.702

4.4. Long-Term Temporal Continuity of Reconstructed NPP

Because the objective of this study is not only annual prediction but also long-term standardized NPP reconstruction, representative grid cell trajectories were extracted from the multi-decadal reconstructed NPP stack (Figure 6). The selected grids cover high-change, low-productivity disturbed, high-productivity restored, and median-change conditions, thereby illustrating whether the reconstructed series remains temporally coherent across contrasting ecological states.

The reconstructed trajectories exhibit continuous multi-year dynamics rather than isolated annual noise. Disturbed grids maintain lower NPP levels or delayed recovery, whereas restored or high-productivity grids show higher and more stable NPP trajectories. High-change grids capture stronger temporal fluctuations, consistent with disturbance–reclamation dynamics in open-pit mining regions. These temporal patterns demonstrate that the proposed framework can support ecological interpretation of multi-decadal NPP dynamics at the standardized baseline grid scale; they should not be interpreted as independent validation against ground-measured NPP.

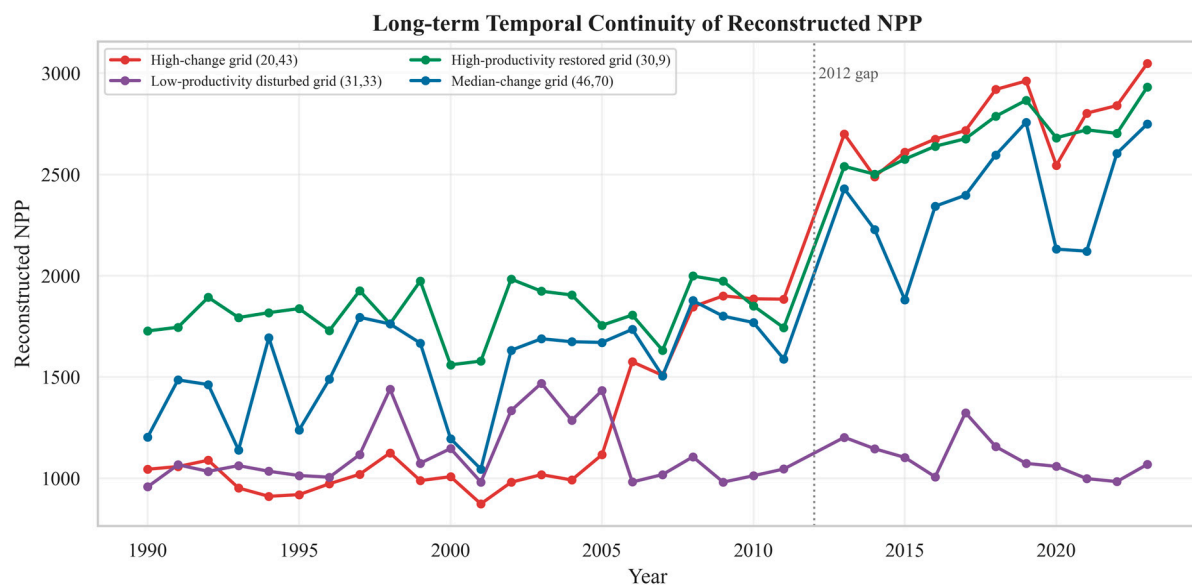


Figure 6. Long-term temporal continuity of reconstructed NPP for representative grid cells. The vertical dotted line marks the 2012 data gap associated with the Landsat 5-to-Landsat 8 transition and the exclusion of 2012 from the reconstruction sequence.

4.5. Computational Complexity and Efficiency

Beyond spatial accuracy, computational efficiency is important for multi-year geospatial reconstruction. The 3×3 CNN requires multi-pixel spatial filters, whereas the 1×1 CNN reduces convolutional operations by applying point-wise kernels. The final 1×1 model contains 22,433 trainable parameters compared with 44,097 parameters in the 3×3 CNN. A newly timed inference test on the saved 1×1 model and scaler yielded an average of 1.38 s across five runs for 18,976 test samples. Because hardware, batch size, and input construction affect timing, efficiency results are interpreted as indicative rather than definitive.

Taken together, the results identify three main outcomes. First, the standard 3×3 CNN achieved the best random-split global accuracy, whereas the 1×1 CNN remained competitive while using a point-wise receptive field. Second, the 1×1 CNN better preserved local NPP gradients and reduced full-domain and boundary zone errors in the 2020 spatial fidelity assessment. Third, the multi-year trajectories and tensor attribution analysis indicate that the reconstructed NPP series is temporally coherent and mainly driven by stress-, disturbance-, and moisture-sensitive biophysical indices.

5. Discussion

5.1. The Trade-Off Between Global Metrics and Local Spatial Fidelity

A pervasive challenge in geospatial deep learning is the over-reliance on global statistical metrics (e.g., R^2 and RMSE) during model optimization. As demonstrated in Table

1, the standard 3×3 CNN marginally outperformed the 1×1 CNN in random-split global accuracy. However, in highly fragmented landscapes such as open-pit mines, optimizing purely for global agreement can encourage neighborhood smoothing. Multi-pixel convolutions act as local filters that reduce variance by absorbing surrounding pixel values, which can increase agreement with a moderate-resolution target while degrading boundary zone fidelity [38–40,58–62].

This study therefore evaluates both global predictive accuracy and local spatial fidelity. The 3×3 CNN achieved slightly higher random-split accuracy, but its neighborhood-based receptive field produced higher boundary zone RMSE and lower gradient correspondence in the 2020 spatial fidelity assessment. The 1×1 CNN intentionally sacrifices access to neighboring pixels and learns channel-wise relationships at each grid cell. This design is appropriate when the research objective is standardized grid cell reconstruction rather than spatial downscaling or super-resolution.

5.2. Interpretability via Biophysical Tensor Attribution

Permutation-based feature attribution was used as an explainable AI diagnostic for the final 1×1 CNN (Figure 7) [63]. The analysis was conducted on the independent 20% test split. Each tensor channel was permuted 10 times while all other channels were held fixed, and importance was measured as the mean increase in RMSE relative to the unpermuted baseline prediction. Larger RMSE increases indicate stronger model sensitivity to the corresponding biophysical tensor.

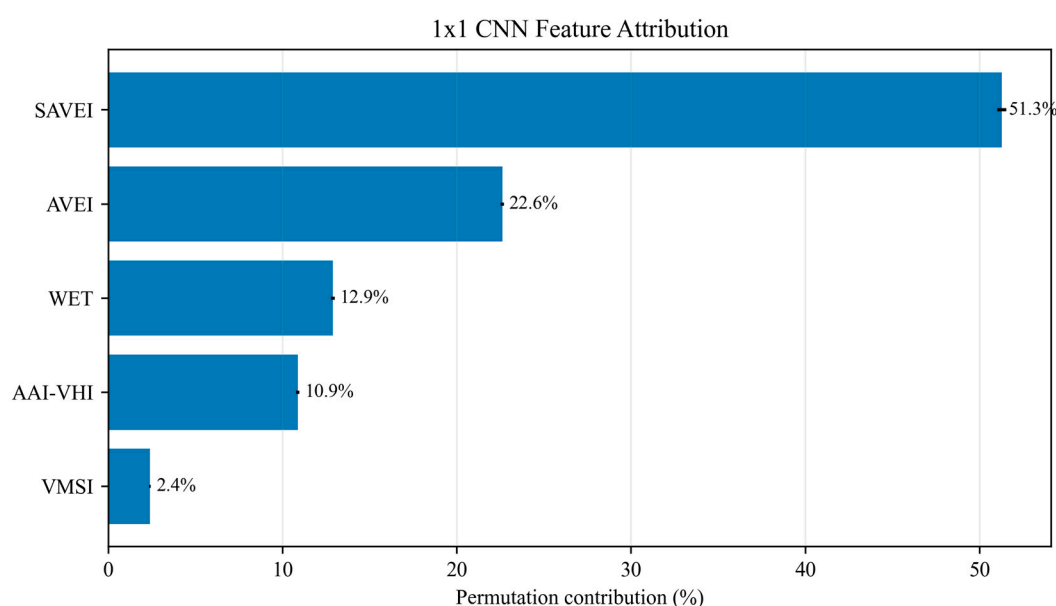


Figure 7. Permutation-based feature attribution of the five biophysical tensor channels for the final 1×1 CNN. Contributions are calculated from 10 repeated permutations on the independent test split using RMSE increase as the evaluation metric.

The attribution results show that SAVEI produced the largest relative contribution (51.3%), followed by AVEI (22.6%), WET (12.9%), AAI-VHI (10.9%), and VMSI (2.4%). This pattern suggests that normalized environmental stress, atmospheric/thermal adjustment, and wetness information were central to the point-wise NPP mapping. The attribution should be interpreted as model sensitivity under the trained feature distribution, not as causal proof of ecological drivers.

5.3. Applicability and Limitations

The proposed framework offers a scalable option for reconstructing standardized geospatial records at the MODIS NPP baseline grid. It is especially suitable for landscapes where adjacent grid cells may represent functionally distinct ecological states, such as active pits, overburden dumps, restored vegetation patches, and undisturbed reference areas. This landscape framing is consistent with surface coal-mining monitoring studies that map disturbance and reclamation at global or regional scales [8–12]. It also aligns with automated disturbance or exposed-coal mapping [17–19], reclamation and vegetation recovery assessments [13–16,20–24], and ecological quality or landscape-risk studies in mining areas [25–27]. The main added value relative to the MODIS NPP product is not the creation of sub-MODIS ground truth, but the use of Landsat-derived biophysical indices to generate a spatially conservative, interpretable reconstruction within the same standardized baseline support.

Several limitations should be acknowledged. First, the target variable is MODIS NPP at a moderate baseline grid scale, so the model cannot verify fine-scale real surface boundaries below the MODIS support. Second, the random 80/20 split may be affected by spatial and temporal autocorrelation because the same locations recur across years; future work should include spatial block, year block, leave-one-year-out, or mining/restoration zone holdout validation. Third, no independent ground NPP measurements were available for external validation. Fourth, cross-sensor consistency between Landsat 5 TM and Landsat 8 OLI remains a source of uncertainty even when Collection 2 Level-2 products are used. Fifth, the customized tensors are empirical indices whose coefficients and bounds may require recalibration in other landscapes. Finally, although the local visualization window was selected by a gradient criterion, broader multi-window and multi-year boundary assessments would further strengthen the evidence.

6. Conclusions

This study developed a biophysical index-guided 1×1 Convolutional Neural Network for standardized NPP reconstruction in the Heidaigou–Haerwusu open-pit mining landscape. The framework integrates Landsat-derived biophysical tensors with the MODIS annual NPP baseline while deliberately restricting the convolutional receptive field to each grid cell. The main contribution is therefore not sub-MODIS super-resolution, but a spatially conservative and interpretable reconstruction strategy for fragmented disturbance–restoration mosaics.

With respect to Objective 1, the benchmark comparison showed that the standard 3×3 CNN achieved the highest random-split global accuracy ($R^2 = 0.8963$, RMSE = 176.42), followed closely by Random Forest ($R^2 = 0.8938$, RMSE = 178.48). The proposed 1×1 CNN obtained a slightly lower global accuracy ($R^2 = 0.8766$, RMSE = 191.90), confirming that global statistical fit and local spatial fidelity should be treated as complementary evaluation criteria rather than a single performance target.

With respect to Objective 2, the 2020 full-domain spatial fidelity assessment showed that the 1×1 CNN reduced spatial error and preserved local gradients more effectively than the neighborhood-based 3×3 CNN. The 1×1 CNN produced the lowest full-domain RMSE (229.62) and MAE (169.32), the highest gradient correlation (0.611), and lower boundary zone error than the 3×3 CNN (boundary RMSE 376.07 versus 515.40; boundary MAE 297.35 versus 356.62). These results support the use of point-wise convolution when adjacent MODIS-scale cells represent functionally distinct mining, dumping, restored, and reference conditions.

With respect to Objective 3, residual maps, long-term trajectories, and permutation-based attribution clarified how the model behaves. The residual analysis showed less

cross-boundary smoothing for the 1×1 CNN, the reconstructed trajectories captured coherent multi-year differences among disturbed and restored grid cells, and the attribution results identified SAVEI (51.3%), AVEI (22.6%), WET (12.9%), AAI-VHI (10.9%), and VMSI (2.4%) as the relative tensor contributions. These findings indicate that normalized environmental stress, atmospheric/thermal adjustment, and wetness information are important for MODIS-scale NPP reconstruction in this mining landscape.

Future work should strengthen the framework through independent ground or flux-based NPP validation, spatial block and year block cross-validation, leave-one-year-out and mining/restoration zone holdout tests, and multi-site transfer experiments across different mining and reclamation settings. Additional comparisons with MLP, XGBoost, SVR, GAM, and physics-informed or process-guided models would further clarify the advantages and limits of point-wise convolution. Future applications should also recalibrate empirical tensor coefficients, test broader multi-window boundary diagnostics, and explore higher-resolution sensor fusion where independent fine-scale productivity observations are available.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18142395/s1>.

Author Contributions: Conceptualization, Y.P. and P.W.; methodology, Y.P.; software, Y.P.; validation, Y.P. and P.W.; formal analysis, Y.P.; data curation, Y.P.; writing—original draft preparation, Y.P.; writing—review and editing, P.W.; visualization, Y.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data, code, and model artifacts supporting this study are provided in the Supplementary Materials. The revision analysis outputs include recovered baseline MAE values, 1×1 CNN inference timing, 2020 spatial fidelity metrics, and 1×1 CNN permutation-attribution tables generated from the archived data and models. The saved 1×1 CNN model metadata report Keras 3.5.0.

Acknowledgments: The authors extend their sincere gratitude to the Google Earth Engine (GEE) platform for providing the robust cloud computing infrastructure and seamless access to the multi-decadal Landsat and MODIS data archives, which were essential for the execution of this research. During the preparation of this manuscript, the authors used Google Gemini strictly for improving English language readability and academic phrasing; after using this service, the authors reviewed and edited the content and take full responsibility for the final manuscript.

Conflicts of Interest: Author Pengchong Wang was employed by the company Beijing Linmiao Ecological Environment Technology Co., Ltd. All remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
2. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.K. A Landsat Surface Reflectance Dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 68–72. <https://doi.org/10.1109/lgrs.2005.857030>.
3. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
4. Park, J.H.; Gan, J.; Park, C. Discrepancies between global forest net primary productivity estimates derived from MODIS and forest inventory data and underlying factors. *Remote Sens.* **2021**, *13*, 1441. <https://doi.org/10.3390/rs13081441>.

5. Zhang, W.; Zhao, X.; Li, H.; Fang, Y.; Shi, W.; Zhao, S.; Guo, Y. Dynamic analysis and risk assessment of vegetation net primary productivity in Xinjiang, China. *Remote Sens.* **2024**, *16*, 3604. <https://doi.org/10.3390/rs16193604>.
6. Xie, C.; Wu, S.; Zhuang, Q.; Zhang, Z.; Hou, G.; Luo, G.; Hu, Z. Where anthropogenic activity occurs, anthropogenic activity dominates vegetation net primary productivity change. *Remote Sens.* **2022**, *14*, 1092. <https://doi.org/10.3390/rs14051092>.
7. Tian, H.; Liu, S.; Zhu, W.; Zhang, J.; Zheng, Y.; Shi, J.; Bi, R. Deciphering the drivers of net primary productivity of vegetation in mining areas. *Remote Sens.* **2022**, *14*, 4177. <https://doi.org/10.3390/rs14174177>.
8. Yang, Z.; Tian, W.; He, T.; Ren, H.; Wang, X.; Jiang, H.; Zhao, L.; Zhao, Y.; Chen, Y. A global perspective: Quantifying disturbance and reclamation of surface coal mining through remote sensing innovations. *Glob. Environ. Change* **2026**, *98*, 103138. <https://doi.org/10.1016/j.gloenvcha.2026.103138>.
9. Pericak, A.A.; Thomas, C.J.; Kroodsmas, D.A.; Wasson, M.F.; Ross, M.R.V.; Clinton, N.E.; Campagna, D.J.; Franklin, Y.; Bernhardt, E.S.; Amos, J.F. Mapping the yearly extent of surface coal mining in Central Appalachia using Landsat and Google Earth Engine. *PLoS ONE* **2018**, *13*, e0197758. <https://doi.org/10.1371/journal.pone.0197758>.
10. Townsend, P.A.; Helmers, D.P.; Kingdon, C.C.; McNeil, B.E.; de Beurs, K.M.; Eshleman, K.N. Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976–2006 Landsat time series. *Remote Sens. Environ.* **2009**, *113*, 62–72. <https://doi.org/10.1016/j.rse.2008.08.012>.
11. Ali, N.; Fu, X.; Ashraf, U.; Chen, J.; Thanh, H.V.; Anees, A.; Riaz, M.S.; Fida, M.; Hussain, M.A.; Hussain, S.; et al. Remote sensing for surface coal mining and reclamation monitoring in the Central Salt Range, Punjab, Pakistan. *Sustainability* **2022**, *14*, 9835. <https://doi.org/10.3390/su14169835>.
12. Vorovencii, I. Changes detected in the extent of surface mining and reclamation using multitemporal Landsat imagery: A case study of Jiu Valley, Romania. *Environ. Monit. Assess.* **2021**, *193*, 30. <https://doi.org/10.1007/s10661-020-08834-w>.
13. Li, Q.; Guo, J.; Wang, F.; Song, Z. Monitoring the characteristics of ecological cumulative effect due to mining disturbance utilizing remote sensing. *Remote Sens.* **2021**, *13*, 5034. <https://doi.org/10.3390/rs13245034>.
14. Karan, S.K.; Samadder, S.R.; Maiti, S.K. Assessment of the capability of remote sensing and GIS techniques for monitoring reclamation success in coal mine degraded lands. *J. Environ. Manag.* **2016**, *182*, 272–283. <https://doi.org/10.1016/j.jenvman.2016.07.070>.
15. Wang, S.; Ma, C.; Ma, Y.; Li, T. Monitoring and evaluation of ecological restoration in open-pit coal mine using remote sensing data based on a OM-RSEI model. *Int. J. Min. Reclam. Environ.* **2025**, *39*, 600–622. <https://doi.org/10.1080/17480930.2025.2458012>.
16. Zhu, D.; Chen, T.; Zhen, N.; Niu, R. Monitoring the effects of open-pit mining on the eco-environment using a moving window-based remote sensing ecological index. *Environ. Sci. Pollut. Res.* **2020**, *27*, 15716–15728. <https://doi.org/10.1007/s11356-020-08054-2>.
17. Yang, Z.; Niu, Y.; He, T.; Zhang, J.; Tian, W.; Zhao, Y. An automated method of identifying surface coal mining disturbances by autonomous sample generation. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 4403914. <https://doi.org/10.1109/TGRS.2026.3665648>.
18. Yang, Z.; He, T.; Zhang, J.; Zhao, Y. A novel index for exposed coal mapping using Landsat imagery. *Ecol. Indic.* **2024**, *166*, 112395. <https://doi.org/10.1016/j.ecolind.2024.112395>.
19. Mukherjee, J.; Mukherjee, J.; Chakravarty, D.; Aikat, S. A novel index to detect opencast coal mine areas from Landsat 8 OLI/TIRS. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 891–897. <https://doi.org/10.1109/jstars.2019.2896842>.
20. Chen, Y.; Suo, Z.; Lu, H.; Cheng, H.; Li, Q. Active-passive remote sensing evaluation of ecological environment quality in Juye Mining Area, China. *Remote Sens.* **2023**, *15*, 5750. <https://doi.org/10.3390/rs15245750>.
21. Hu, J.; Ye, B.; Bai, Z.; Feng, Y. Remote sensing monitoring of vegetation reclamation in the Antaibao Open-Pit Mine. *Remote Sens.* **2022**, *14*, 5634. <https://doi.org/10.3390/rs14225634>.
22. Zhang, C.; Li, F.; Li, J.; Zhang, K.; Ran, W.; Du, M.; Guo, J.; Hou, G. Assessing the effect, attribution, and potential of vegetation restoration in open-pit coal mines' dumping sites during 2003–2020 utilizing remote sensing. *Ecol. Indic.* **2023**, *155*, 111003. <https://doi.org/10.1016/j.ecolind.2023.111003>.
23. Nie, X.; Hu, Z.; Ruan, M.; Zhu, Q.; Sun, H. Remote-sensing evaluation and temporal and spatial change detection of ecological environment quality in coal-mining areas. *Remote Sens.* **2022**, *14*, 345. <https://doi.org/10.3390/rs14020345>.
24. Hong, F.; He, G.; Wang, G.; Zhang, Z.; Peng, Y. Monitoring of land cover and vegetation changes in Juhugeng Coal Mining Area based on multi-source remote sensing data. *Remote Sens.* **2023**, *15*, 3439. <https://doi.org/10.3390/rs15133439>.
25. Li, Y.; Wang, L.; Cheng, L.; Niu, Z.; Su, X.; Feng, H.; He, Q. Advancing ecological restoration in a mining city: Insights from ecological quality dynamics and driving mechanisms. *Remote Sens.* **2026**, *18*, 558. <https://doi.org/10.3390/rs18040558>.

26. Zhu, Z.; Cao, H.; Yang, J.; Shang, H.; Ma, J. Ecological environment quality assessment and spatial autocorrelation of northern Shaanxi mining area in China based-on improved remote sensing ecological index. *Front. Environ. Sci.* **2024**, *12*, 1325516. <https://doi.org/10.3389/fenvs.2024.1325516>.
27. Zhu, X.; Li, P.; Wang, B.; Zhao, S.; Zhang, T.; Yao, Q. Construction of landscape eco-geological risk assessment framework in coal mining area using multi-source remote sensing data. *Ecol. Inform.* **2024**, *81*, 102635. <https://doi.org/10.1016/j.ecoinf.2024.102635>.
28. Hsu, C.-Y.; Li, W.; Wang, S. Knowledge-driven GeoAI: Integrating spatial knowledge into multi-scale deep learning for Mars crater detection. *Remote Sens.* **2021**, *13*, 2116. <https://doi.org/10.3390/rs13112116>.
29. Camps-Valls, G.; Svendsen, D.H.; Cortes-Andres, J.; Marenno-Martinez, A.; Perez-Suay, A.; Adsuar, J.; Martin, I.; Piles, M.; Munoz-Mari, J.; Martino, L. Physics-aware machine learning for geosciences and remote sensing. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021*; IEEE: New York, NY, USA, 2021; pp. 2086–2089. <https://doi.org/10.1109/igarss47720.2021.9554521>.
30. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv.* **2022**, *55*, 1–37. <https://doi.org/10.1145/3514228>.
31. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. <https://doi.org/10.1038/s41586-019-0912-1>.
32. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
33. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>.
34. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep learning-based change detection in remote sensing images: A review. *Remote Sens.* **2022**, *14*, 871. <https://doi.org/10.3390/rs14040871>.
35. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. <https://doi.org/10.1109/tgrs.2021.3095166>.
36. Liu, Y.; Zhang, Y.; Wang, Y.; Mei, S. Rethinking transformers for semantic segmentation of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. <https://doi.org/10.1109/tgrs.2023.3302024>.
37. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015*; IEEE: New York, NY, USA, 2015; pp. 1–9. <https://doi.org/10.1109/cvpr.2015.7298594>.
38. Moran, P.A.P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. <https://doi.org/10.1093/biomet/37.1-2.17>.
39. Anselin, L. Local indicators of spatial association-LISA. *Geogr. Anal.* **1995**, *27*, 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
40. Xu, Y.; Jiang, J. High-resolution boundary-constrained and context-enhanced network for remote sensing image segmentation. *Remote Sens.* **2022**, *14*, 1859. <https://doi.org/10.3390/rs14081859>.
41. Zhu, X.; Cai, F.; Tian, J.; Williams, T. Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. *Remote Sens.* **2018**, *10*, 527. <https://doi.org/10.3390/rs10040527>.
42. Shi, W.; Guo, D.; Zhang, H. A reliable and adaptive spatiotemporal data fusion method for blending multi-spatiotemporal-resolution satellite images. *Remote Sens. Environ.* **2022**, *268*, 112770. <https://doi.org/10.1016/j.rse.2021.112770>.
43. Jia, D.; Cheng, C.; Song, C.; Shen, S.; Ning, L.; Zhang, T. A hybrid deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions. *Remote Sens.* **2021**, *13*, 645. <https://doi.org/10.3390/rs13040645>.
44. Yang, Z.; Diao, C.; Li, B. A robust hybrid deep learning model for spatiotemporal image fusion. *Remote Sens.* **2021**, *13*, 5005. <https://doi.org/10.3390/rs13245005>.
45. Jia, D.; Cheng, C.; Shen, S.; Ning, L. Multitask deep learning framework for spatiotemporal fusion of NDVI. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. <https://doi.org/10.1109/tgrs.2021.3140144>.
46. Li, W.; Cao, D.; Peng, Y.; Yang, C. MSNet: A multi-stream fusion network for remote sensing spatiotemporal fusion based on transformer and convolution. *Remote Sens.* **2021**, *13*, 3724. <https://doi.org/10.3390/rs13183724>.
47. Cao, H.; Luo, X.; Peng, Y.; Xie, T. MANet: A network architecture for remote sensing spatiotemporal fusion based on multiscale and attention mechanisms. *Remote Sens.* **2022**, *14*, 4600. <https://doi.org/10.3390/rs14184600>.
48. Cheng, Q.; Xie, R.; Wu, J.; Ye, F. Deep learning-based spatiotemporal fusion architecture of Landsat 8 and Sentinel-2 data for 10 m series imagery. *Remote Sens.* **2024**, *16*, 1033. <https://doi.org/10.3390/rs16061033>.

49. Sun, E.; Cui, Y.; Liu, P.; Yan, J. A decade of deep learning for remote sensing spatiotemporal fusion: Advances, challenges, and opportunities. *Inf. Fusion* **2026**, *126*, 103675. <https://doi.org/10.1016/j.inffus.2025.103675>.
50. Han, W.; Li, J.; Wang, S.; Zhang, X.; Dong, Y.; Fan, R.; Zhang, X.; Wang, L. Geological remote sensing interpretation using deep learning feature and an adaptive multisource data fusion network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. <https://doi.org/10.1109/tgrs.2022.3183080>.
51. Zhai, Y.; Roy, D.P.; Martins, V.S.; Zhang, H.K.; Yan, L.; Li, Z. Conterminous United States Landsat-8 top of atmosphere and surface reflectance tasseled cap transformation coefficients. *Remote Sens. Environ.* **2022**, *274*, 112992. <https://doi.org/10.1016/j.rse.2022.112992>.
52. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).
53. Kogan, F.N. Application of vegetation index and brightness temperature for drought detection. *Adv. Space Res.* **1995**, *15*, 91–100. [https://doi.org/10.1016/0273-1177\(95\)00079-T](https://doi.org/10.1016/0273-1177(95)00079-T).
54. Lee, S.-J.; Kim, N.; Lee, Y. Development of integrated crop drought index by combining rainfall, land surface temperature, evapotranspiration, soil moisture, and vegetation index for agricultural drought monitoring. *Remote Sens.* **2021**, *13*, 1778. <https://doi.org/10.3390/rs13091778>.
55. Chang, S.; Chen, H.; Wu, B.; Nasanbat, E.; Yan, N.; Davdai, B. A practical satellite-derived vegetation drought index for arid and semi-arid grassland drought monitoring. *Remote Sens.* **2021**, *13*, 414. <https://doi.org/10.3390/rs13030414>.
56. Hang, Q.; Guo, H.; Meng, X.; Wang, W.; Cao, Y.; Liu, R.; De Maeyer, P.; Wang, Y. Optimizing the vegetation health index for agricultural drought monitoring: Evaluation and application in the Yellow River Basin. *Remote Sens.* **2024**, *16*, 4507. <https://doi.org/10.3390/rs16234507>.
57. Xu, Z.; Sun, H.; Zhang, T.; Xu, H.; Wu, D.; Gao, J. The high spatial resolution Drought Response Index (HiDRI): An integrated framework for monitoring vegetation drought with remote sensing, deep learning, and spatiotemporal fusion. *Remote Sens. Environ.* **2024**, *312*, 114324. <https://doi.org/10.1016/j.rse.2024.114324>.
58. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
59. Meyer, H.; Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **2021**, *12*, 1620–1633. <https://doi.org/10.1111/2041-210X.13650>.
60. Xu, Q.; Li, B.; McRoberts, R.E.; Li, Z.; Hou, Z. Harnessing data assimilation and spatial autocorrelation for forest inventory. *Remote Sens. Environ.* **2023**, *288*, 113488. <https://doi.org/10.1016/j.rse.2023.113488>.
61. Holloway-Brown, J.; Helmstedt, K.J.; Mengersen, K.L. Spatial Random Forest (S-RF): A random forest approach for spatially interpolating missing land-cover data with multiple classes. *Int. J. Remote Sens.* **2021**, *42*, 3756–3776. <https://doi.org/10.1080/01431161.2021.1881183>.
62. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schroder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. <https://doi.org/10.1111/ecog.02881>.
63. Liu, P.; Wang, L.; Li, J. Unlocking the potential of explainable artificial intelligence in remote sensing big data. *Remote Sens.* **2023**, *15*, 5448. <https://doi.org/10.3390/rs15235448>.

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.