remote sensing

MDPI

Article

Optimizing Spatial Representativeness of LULC Samples over Complex Karst Terrain Using Remote Sensing Phenology and Landform-Constrained Joint Stratification

Ya Li [1,2], Zhongfa Zhou [1,2,3,4,*], Denghong Huang [1,2], Huanhuan Lu [1,2], Ruiqi Fan [3,4], Qingqing Dai [1,2], Ying Luo [1,2], Changyan Huang [1,2] and Yuexing Yu [1,2]

1 School of Karst Science, Guizhou Normal University, Guiyang 550001, China; 242100170588@gznu.edu.cn (Y.L.); hdh@gznu.edu.cn (D.H.); 242100170590@gznu.edu.cn (H.L.); 232100170576@gznu.edu.cn (Q.D.); 252100170607@gznu.edu.cn (Y.L.); 252100170602@gznu.edu.cn (C.H.); 252100170601@gznu.edu.cn (Y.Y.)
2 Guizhou Provincial Key Laboratory of Intelligent Processing and Application of Remote Sensing Big Data, Guiyang 550001, China
3 School of Geography & Environmental Science, Guizhou Normal University, Guiyang 550001, China; 242100090370@gznu.edu.cn
4 Anshun Agricultural Environment Field Observation and Research Station, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Anshun 551400, China
* Correspondence: fa6897@gznu.edu.cn

Highlights

What are the main findings?

- The Beipanjiang River Basin (Guizhou section) was divided into six spatially contiguous phenological subregions; LOS was used as a simplified indicator to characterize long-term vegetation phenological gradients, while the independent roles of SOS, OM, and EOS should be acknowledged.
- Under the current experimental setting, the landform-constrained phenological stratification and dual-weighted sample allocation scheme increased OA from 71.33% to 77.55% and Kappa from 0.43 to 0.62 compared with simple random sampling.

What are the implications of the main findings?

- The joint use of remote sensing phenological background and landform heterogeneity provides a potential sample-organization strategy for reducing insufficient spatial representativeness over fragmented karst landscapes.
- The proposed framework may serve as a methodological reference for LULC sample optimization in ecologically fragile karst regions, but its transferability still requires validation in additional years and regions.

Check for updates

Academic Editors: Shicheng Li, Fan Yang, Mengmeng Wang and Prabuddh Kumar Mishra

Received: 28 April 2026
Revised: 28 May 2026
Accepted: 29 May 2026
Published: 10 June 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.

Abstract

Karst regions are characterized by fragmented topography and significant micro-relief mosaics, leading to prominent spectral aliasing of land features, which can result in insufficient spatial representativeness of remote sensing samples for Land Use and Land Cover (LULC). The accuracy of LULC data directly affects the scientific basis of decision-making for rocky desertification control and ecological conservation. This study selected the Beipanjiang River Basin in Guizhou Province, a typical karst region, as the study area. The study selected the SOS, LOS, OM, and EOS indices from the 2001–2020 MODIS MCD12Q2 phenological dataset, combined with topographic zoning data. This study developed a sample spatial optimization scheme for complex karst terrain by integrating Spearman's correlation analysis, SKATER spatially constrained clustering, statistical tests, adaptive

stratified sampling, and Random Forest classification. The scheme was designed to test a phenology–landform joint stratification strategy for spatial sample allocation. The results indicate that (1) the study area was divided into six phenological pattern subregions, with significant spatial differentiation observed among them; (2) the "phenology–landform joint stratification + dual-weighted sample allocation" method was associated with improved sample representativeness and greater internal homogeneity within sample strata under the current experimental setting; and (3) compared to simple random sampling, the remote sensing phenological pattern-driven spatial optimization scheme improved overall accuracy from 71.33% to 77.55% and increased the Kappa coefficient from 0.43 to 0.62. These results suggest that, under the current study-area, sample-size, and validation settings, the phenology–landform joint stratification and dual-weighted allocation scheme can improve the spatial organization of training samples and classification performance over complex karst terrain, although weakly vegetated or bare classes remain difficult to separate.

Keywords: complex karst surfaces; Land Use and Land Cover (LULC); remote sensing phenology; adaptive stratified sampling; Google Earth Engine

1. Introduction

Dynamically updated global LULC datasets provide foundational data support for understanding the status, trends, and pressures of human activities on the carbon cycle, biodiversity, and other natural and anthropogenic processes [1], and play an irreplaceable role in fields such as ecological and environmental assessment, resource management, urban planning, and climate change research [2]. Despite continuous technological advancements, improvements in the accuracy of global LULC remote sensing interpretation still face significant bottlenecks over complex terrain [3–6]. Karst regions are typical ecologically fragile areas characterized by complex topography [7]. Due to intensive karst development, pronounced terrain fragmentation, and frequent micro-topographic mosaics, these regions exhibit strong spatial heterogeneity [8]. This heterogeneity often leads to severe spectral confusion due to intra-class spectral variability and inter-class spectral similarity. Consequently, the classification uncertainty of existing medium-to-high-resolution LULC products in this region is significantly higher than the global average [9], making karst regions a critical bottleneck for improving global LULC mapping accuracy. The accuracy of LULC data directly impacts the scientific soundness of major decisions regarding rock desertification control and ecological conservation [10].

Although previous studies have addressed LULC mapping challenges in complex karst terrain, existing LULC products still struggle to overcome classification accuracy limitations in these areas due to factors such as fragmented topography and severe spectral confusion [9]. To improve mapping accuracy, some studies have attempted to incorporate phenological characteristics or stratified sampling methods, such as using NDVI time series to capture vegetation growth dynamics and employing stratified sampling to ensure the representativeness of non-dominant land cover classes [11,12]. However, most studies have not fully considered the effects of topographic heterogeneity and phenological rhythms in karst regions, and the spatial distribution of training samples has not been precisely matched to surface complexity and spectral heterogeneity. For example, the same tree species in karst regions exhibits different spectral responses at different elevations [13], yet traditional sampling methods often overlook these topography-driven phenological differences, resulting in insufficient representativeness of samples in both feature space and geographic space [14]. At the same time, the phenomenon of "same object, different

spectra; different objects, same spectra" is prominent in this region [9]. Relying solely on spectral data from a single time period or generic sampling strategies makes it difficult to effectively distinguish easily confused land classes, and this mismatch between samples and complex surface features further exacerbates classification errors [15]. The key reason why LULC classification accuracy in karst regions remains difficult to improve lies in the mismatch between the selection of training samples and complex surface features. Various supervised classification algorithms serve as the primary technical means for automated land cover mapping at regional to global scales [16]. The core prerequisite for achieving high-precision mapping is the acquisition of training samples with sufficient spatial representativeness and intra-class consistency [17]. With advancements in sample collection technology, traditional simple random sampling or systematic sampling methods struggle to cover highly fragmented landscapes. This leads to samples being geographically concentrated in flat terrain or omitting areas with phenological transitions and complex microtopography, ultimately resulting in samples that fail to adequately represent different landscape components [18]. Sample selection based on phenological spectrum similarity can improve classification performance for classes with low inter-class similarity; this can be combined with multi-model validation to uncover patterns of model adaptability in the samples [19]. Spatial autocorrelation is an inherent property of remote sensing data [20]; intra-class consistency is closely related to phenological differences caused by spatial partitioning and is a major factor influencing LULC training samples [21].

Phenology serves as a key mediating variable linking climate drivers, canopy structure, and ecological processes; its rhythmic variations directly influence surface reflectance and spectral trajectories in time series, and have been widely utilized to enhance the remote sensing distinguishability of easily confused land cover classes such as cropland, grassland, and forest. Climate and topographic characteristics jointly determine vegetation phenological patterns [8,9], while the rhythmic changes in phenology can effectively characterize temporal differences in vegetation growth [22], further leading to opposite trends in phenological changes at different elevations [23]. Vegetation phenology in karst regions is influenced by the combined effects of climate and topography, which increases the complexity of remote sensing imagery for land cover classification, enhances feature diversity, and consequently leads to insufficient sample characteristics and geographic spatial representativeness, thereby limiting the accuracy of LULC classification.

The spatial representativeness and intra-class consistency of training samples also determine the accuracy of LULC mapping [24]; high-quality training samples are a prerequisite for ensuring that supervised classification algorithms effectively learn the spectral characteristics of land features [25]. Compared to general sampling methods, stratified sampling is widely recommended for land cover mapping because it allows for more effective sampling of non-dominant land cover types [12], while reducing the variance among samples within a stratum [26]. Adjacent pixels are more similar than distant ones [20]; by dividing the study area into several sub-regions, the spatial stratification method can significantly improve spatial sampling efficiency [14].

However, existing studies on LULC classification for complex land surfaces, although they have gradually incorporated strategies such as multi-temporal remote sensing, phenological characteristics, and stratified sampling [15], still face two key unresolved issues in karst regions: First, most studies focus on using phenological information to enhance the distinguishability of land cover classes but rarely analyze how phenological differentiation patterns contribute to the spatial organization of training samples [27]; second, existing sampling designs often rely primarily on area proportions or empirical partitioning, and have not yet integrated landform fragmentation, phenological heterogeneity, and sample allocation mechanisms [28]. For complex karst surfaces, the key bottleneck limiting

classification accuracy lies not only in the classification algorithms themselves but also in whether the training samples possess sufficient spatial representativeness and intra-class consistency to match the heterogeneity of the surface [29].

Based on this, taking the Beipanjiang River Basin (Guizhou section) in the Southern Karst of China as the study area, we selected the Start of Growing Season (SOS), Onset of Maturity (OM), end of growing season (EOS), and growing season length (LOS) from the MODIS MCD12Q2 phenology dataset for 2001–2020, along with landform classification data, to construct a sample optimization framework based on "remote sensing phenology zoning—landform joint stratification—dual-weighted sample allocation." The focus of this study is not merely to compare the merits of different classification algorithms but to explore whether, within the complex karst terrain, the long-term stable spatial patterns of phenological differentiation combined with topographic background information can jointly guide the configuration of training samples. This approach aims to enhance the samples' ability to characterize complex terrain and improve LULC classification performance under the current experimental setting. In karst terrain, landform affects LULC patterns indirectly by regulating elevation, relief, slope, aspect, water-thermal redistribution, soil development, rock exposure, erosion risk, and human accessibility. These controls influence vegetation structure, agricultural suitability, built-up expansion, grassland/bare-rock mosaics, and the spectral trajectories captured by Sentinel-2. Therefore, landform is used here as a background proxy for karst environmental heterogeneity rather than as a direct substitute for LULC.

The study aims to achieve the following three objectives: (1) Identify remote sensing phenological sub-regions within the study area that exhibit spatial continuity and relative internal homogeneity; (2) establish an adaptive sample allocation mechanism that balances the representativeness of sub-region areas with internal phenological complexity; and (3) quantitatively compare the proposed sampling scheme with simple random sampling and additional ablation controls under the same sample size, input features, and Random Forest configuration. This provides a methodological reference for LULC sample optimization and mapping in ecologically fragile karst regions of Southwest China and other highly heterogeneous landscapes.

2. Materials and Methods

2.1. Study Area

The Beipanjiang River Basin (Guizhou section) is located on the eastern edge of the Yungui Plateau (Figure 1), with a catchment area of approximately 21,852 km^2 and significant topographic variation (maximum elevation difference of about 2600 m) [30]. This region lies at the core of the world's largest contiguous karst landscape [8]. The region has a humid subtropical monsoon climate, with an average annual temperature of approximately 13–18 °C and annual precipitation of about 1100–1400 mm. The concurrent occurrence of rainfall and heat is conducive to vegetation growth but also facilitates slope erosion [30]. Additionally, the widespread exposure of carbonate rocks, combined with shallow and spatially discontinuous soil layers, makes the region more vulnerable to inappropriate cultivation and engineering disturbances [31]. Moreover, the difficulty of LULC classification in this karst landscape is further amplified by fine-scale mosaics of shallow soils, exposed carbonate rocks, shrubland/grassland, abandoned or degraded cropland, forest patches, terrain shadows, and soil-moisture and lithological heterogeneity. Therefore, the broad landform categories used in this study should be regarded as a proxy for karst surface complexity rather than a complete representation of all karst-related controlling factors. Since the early 21st century, Chinese governments at all levels have launched ecological projects such as the Grain-for-Green Program, forest conservation through mountain closure, and

ecological resettlement to alleviate poverty and restore degraded environments [8]. Benefiting from rapid recovery following the implementation of these ecological projects and the study area's unique topography and water-heat distribution characteristics, phenological indices within the region are significantly higher than those in non-karst areas at the same latitude, forming a distinct phenological gradient.

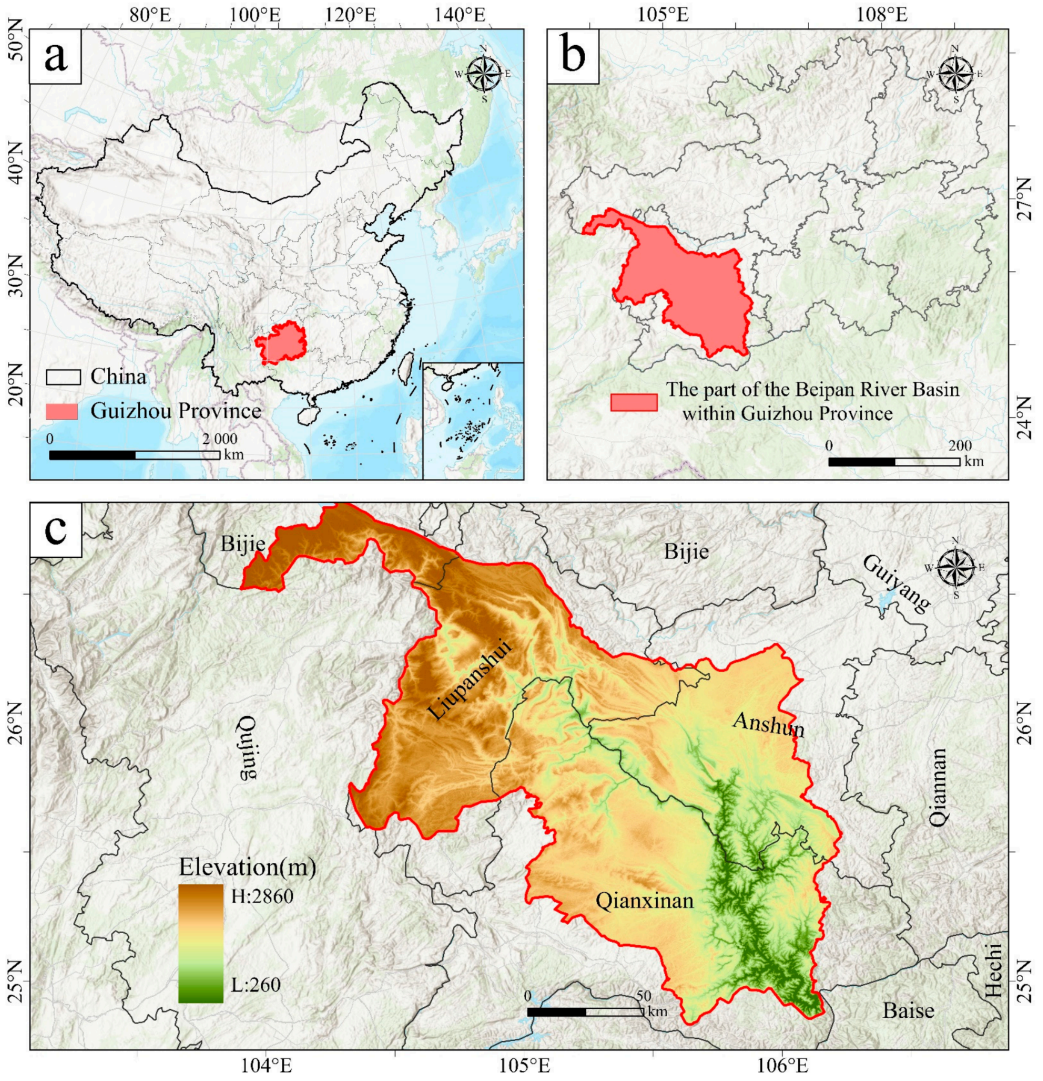


Figure 1. Overview map of the study area. (**a**) Location of Guizhou Province; (**b**) Location of the Beipanjiang River Basin (Guizhou section); (**c**) Significant elevation differences within the study area.

2.2. Data Sources and Pre-Processing

2.2.1. Phenology Data

The phenology data are sourced from the MODIS MCD12Q2 phenology dataset available on the NASA Level-1 and Atmosphere Archive & Distribution System (LAADS) platform (https://ladsweb.modaps.eosdis.nasa.gov) (accessed on 28 May 2026). MCD12Q2 is derived from MODIS EVI2 time-series trajectories and provides annual land-surface phenological transition dates, including the Start of Growing Season (SOS), Onset of Greenness Maximum (OM), End of Growing Season (EOS), and Length of Growing Season (LOS). It should be noted that MCD12Q2 is not a land-cover map and was not used as a 10 m pixel-level predictor in the Sentinel-2-based LULC classification. Instead, it was used as a 500 m background stratification layer to characterize regional phenological gradients and guide sample allocation. In practice, each 10 m Sentinel-2 sample point was assigned

the categorical phenological-zone label at its location by spatial overlay; continuous MODIS phenological metrics were not downscaled or used as 10 m spectral predictors.

The 2001–2020 mean values of SOS, OM, EOS, and LOS were calculated to represent the long-term spatial phenological background shaped by climate, terrain, and vegetation dynamics, rather than to replace the LULC condition of a specific year. The target LULC labels correspond to the 2021 classification year and were interpreted mainly using Sentinel-2 seasonal composite imagery and 2021 GF-2 high-resolution composite imagery. Field surveys conducted during 2025–2026 and manual visual inspection from July to December 2025 were used only as auxiliary checks for temporally stable locations and for excluding samples with obvious land-cover change or ambiguous interpretation, thereby reducing labeling uncertainty.

2.2.2. Remote Sensing Image Data

ESA Copernicus Program Sentinel-2 Level-2A surface reflectance imagery was used as the primary remote sensing input for LULC classification (https://dataspace.copernicus.eu/) (accessed on 28 May 2026). In this study, Sentinel-2 SR Harmonized images acquired from April to October 2021 were selected to generate the seasonal composite image. Images with cloud coverage higher than 10% were excluded. Cloud and cirrus pixels were further masked using the QA60 band, in which bit 10 and bit 11 represent cloud and cirrus contamination, respectively. The reflectance values were scaled by dividing by 10,000, and a median composite was then generated to reduce residual cloud contamination and short-term atmospheric noise.

The original Sentinel-2 bands used in the classification feature stack included B2, B3, B4, and B8, corresponding to the blue, green, red, and near-infrared bands. Based on these bands and B11, several spectral indices were calculated, including NDVI, SAVI, EVI, NDWI, MNDWI, and NDBI. In addition, elevation, slope, and aspect derived from the SRTM DEM were added as terrain features. The 10 m bands were used directly, while the 20 m B11 band was incorporated in the calculation of MNDWI and NDBI and matched to the 10 m classification scale during feature extraction in Google Earth Engine. The final feature stack was sampled at 10 m and used as the pixel-level input for Random Forest-based LULC classification.

2.2.3. Landform Classification Data

The landform classification data are sourced from the "2023 Global Basic Landform Type Unit Dataset" (http://geodata.nnu.edu.cn) (accessed on 28 May 2026) [32] released by the research group led by Tang Guoan at Nanjing Normal University. The dataset classifies land areas first by relief amplitude and then by elevation, forming six primary landform categories and 23 secondary landform types. In this study, the landform dataset was used as a spatial proxy for terrain-related environmental gradients in the karst region. The secondary landform types were overlaid with remote sensing phenological units to define stratification units for sample allocation. Seven secondary landform types were identified in the study area, including low- and mid-elevation plains, low- and mid-elevation hills, gently and moderately undulating low mountains, gently and moderately undulating middle mountains, and highly undulating middle mountains (Figure 2). It should be noted that landform types mainly describe relief and elevation conditions. They cannot fully represent other karst-related factors, such as lithology, soil depth, rock exposure, terrain shadow, or soil-moisture heterogeneity.

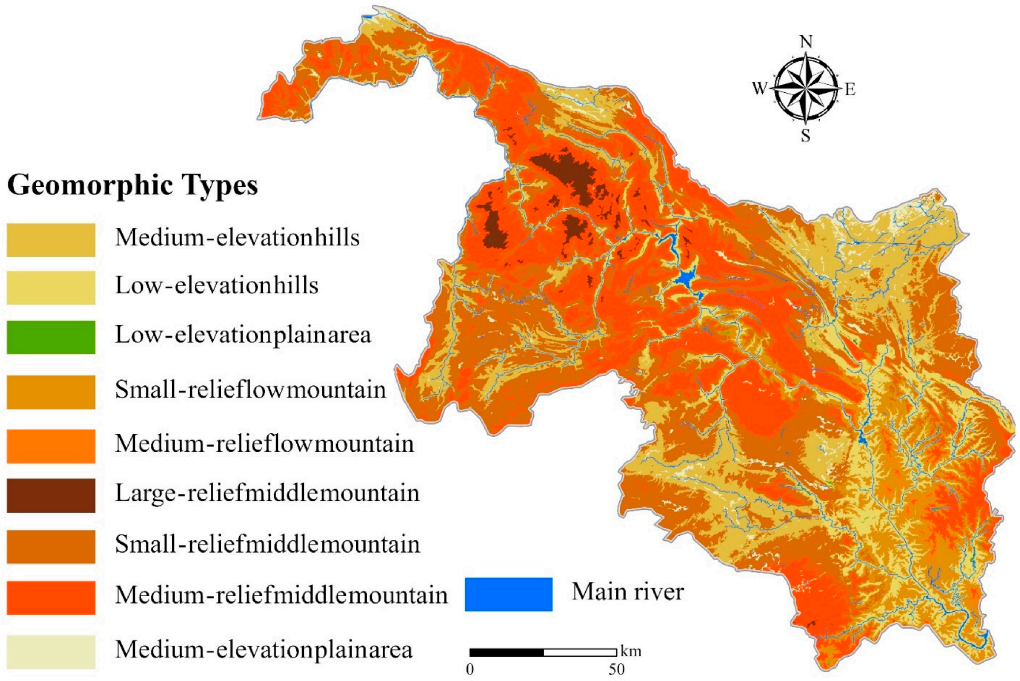


Figure 2. Landforms of the study area.

2.2.4. Sample Point Data

To ensure the authenticity, representativeness, and interpretability of the training samples, this study utilized 0.8 m high-resolution remote sensing imagery from Gaofen-2 (data source: National Remote Sensing Data and Application Service Platform (https://www.cpeos.org.cn) (accessed on 28 May 2026) as the primary basis for interpretation. This was combined with the current status of the study area and the "Classification of Land Use Status" (GB/T 21010-2017) [33] for visual interpretation and sample annotation using ArcGIS Pro 3.0, supplemented by field surveys to verify selected sample points. Based on this, and considering the approximate distribution of land cover types in the study area as well as the balance of rare land cover types, a raw sample database was constructed for the study area, yielding a total of 100,000 sample points, including 20,000 points of cropland, 40,000 points of forest land, 10,000 points of grassland, 5000 points of water bodies, 15,000 points of built-up land, and 10,000 points of unutilized land.

Considering that the spatial resolution of phenological data is 500 m and that remote sensing samples generally exhibit spatial autocorrelation, a minimum spacing of 500 m between sample points was set during the construction of the original sample database to reduce spatial redundancy and prevent multiple sample points from falling within the same phenological pixel. Finally, using the Random Point tool in ArcGIS Pro 3.0, random sampling was performed within the sample set under the 500 m minimum-distance constraint. The specific parameters of the above data are shown in Table 1.

Before training and validation splitting, sample quality control was conducted. Samples with uncertain labels caused by cloud or shadow contamination, inconsistent image dates, mixed land-cover boundaries, or disagreement during visual interpretation were treated as label-quality-control cases. These samples were not removed simply because they were spectrally difficult or located in complex terrain. A total of 1017 uncertain samples were excluded before model training, and the remaining samples were used for subsequent sample selection and classification experiments. For Random Forest classification, the selected samples were randomly divided into training and validation sets at a ratio of 7:3. The validation samples were independent of the training samples, and the 500 m minimum-distance rule was retained to reduce spatial redundancy between

samples. However, strict spatial-block validation was not implemented in this study, which may still leave some influence of spatial autocorrelation and is therefore acknowledged as a limitation.

Table 1. Temporal information, spatial resolution, and roles of datasets used in this study.

Dataset	Year/Period to Report	Spatial Resolution	Role in This Study
MODIS MCD12Q2 phenology	2001–2020 annual metrics; multi-year mean	500 m	Long-term phenological background for stratification/sample allocation
Sentinel-2 Level-2A	2021	10 m	Target-year LULC classification features
GF-2 reference imagery	2021	0.8 m	Target-year visual interpretation and label verification
LULC sample labels	2021 target year; auxiliary checks in 2025–2026 for stable sites only	Point samples	Training/validation labels for 2021 classification; temporally unstable or ambiguous samples excluded
Global Basic Landform Type Unit Dataset	2023	1:200,000-scale vector geomorphological dataset	Landform background stratification

2.3. Research Methods

2.3.1. Research Approach

Technical Approach (Figure 3): (a) Prepare the 2001–2020 mean MODIS phenological data (SOS/OM/EOS/LOS), 2021 Sentinel-2 image composites, landform classification, and DEM data, and perform preprocessing steps such as reprojection, resampling, and smoothing; (b) Calculate multi-year averages and screen for core indicators (LOS) using Spearman's correlation analysis; perform spatial clustering using the SKATER spatial constraint clustering algorithm; (c) Determine six phenological sub-zones using the pseudo-F value and the elbow rule, spatially overlay them with secondary landform types, assign each 10 m sample point to the corresponding stratification unit to generate 43 homogeneous stratified units, calculate the coefficient of variation (CV) and area for each unit and normalize them, allocate samples using a double-weighted model, and construct a high-quality sample set through visual interpretation and field verification; (d) Based on the GEE platform, the Random Forest algorithm was employed to perform LULC classification using stratified sampling. Accuracy was evaluated using metrics such as Overall Accuracy (OA) and the Kappa coefficient, and the effectiveness of the optimized scheme was evaluated by comparing it with simple random sampling and diagnostic ablation schemes.

2.3.2. Development of a Remote Sensing Phenological Zoning Scheme

There is a close association between vegetation phenology and land cover, and multiple biological indicators exist within the vegetation growth cycle (Figure 4). Due to spatial and temporal variations as well as differences in water and heat conditions, these indicators exhibit distinct characteristics. By dividing regions based on key phenological periods observed via remote sensing, the biophysical and spectral characteristics of vegetation within each region become more consistent, thereby reducing classification errors caused by "same object, different spectrum" and improving the accuracy of remote sensing classification [34]. This study selected four key growth stages—Start of Growing Season (SOS), Onset of Greenness Maximum (OM), End of Growing Season (EOS), and Length of the Growing Season (LOS)—that comprehensively reflect the vegetation growth cycle to achieve a complete representation of vegetation phenology. This study does not assume

that phenology directly explains non-vegetated categories; rather, phenological zones are used to organize samples within comparable ecological backgrounds.

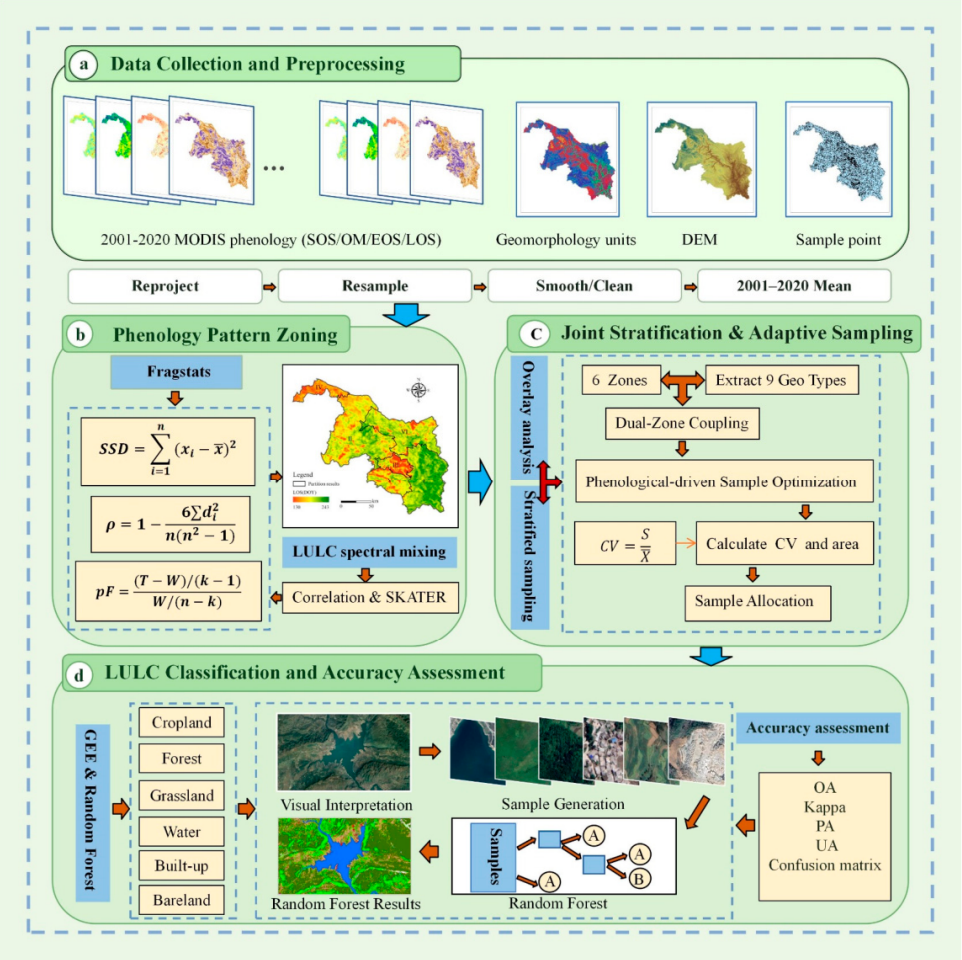


Figure 3. Technology roadmap.

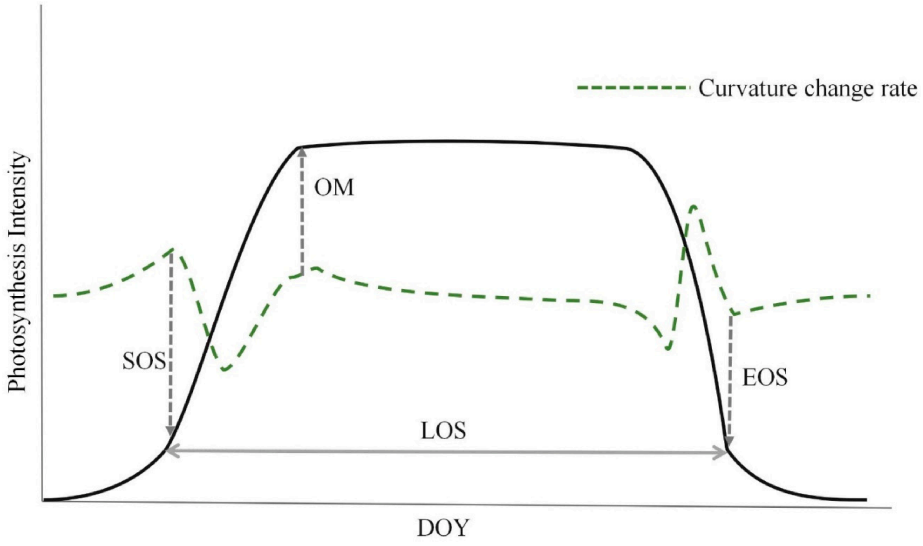


Figure 4. Schematic diagram of phenological indices [34].

Since phenological indicators are dominated by climatic conditions and exhibit consistent rhythms, Spearman's correlation coefficient was used to quantify the multicollinearity

among the indicators. This allowed for the selection of core indicators capable of representing information from multiple indicators, thereby achieving a balance between the simplification and representativeness of the indicator set [7]. Based on the results of the indicator screening, the SKATER clustering algorithm was employed to partition phenological patterns, laying the foundation for the development of a sample optimization strategy. This study uses individual pixels of remote-sensing surface phenological data as the clustering units. Because high correlation alone does not prove classification optimality, LOS was used here as a simplified background-stratification indicator, while the independent roles of SOS, OM, and EOS are discussed as a limitation.

The main calculation formulas involved are as follows:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1}$$

Spearman's correlation analysis is a non-parametric statistical method used to assess the correlation between two variables. Here, d_i represents the rank difference between each pair of data samples, and n denotes the total number of samples. The Spearman correlation coefficient is insensitive to outliers; since large-scale regions such as the Beipanjiang River Basin (Guizhou section) inevitably contain outliers, it serves as a highly useful measure of correlation when outliers are present in the data.

$$SSD = \sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{2}$$

The Sum of Squared Deviations (SSD) is a metric used in the SKATER algorithm to measure intra-cluster variability. x_i represents the attribute value of observation i, and $\overline{x}$ is the mean of the cluster. The objective of the SKATER algorithm is to maximize inter-cluster SSD while minimizing intra-cluster SSD by removing edges from the minimum spanning tree. The core distinction of the SKATER algorithm from other algorithms lies in the introduction of spatial constraints, which ensure that the resulting clusters remain geographically contiguous. This clustering algorithm primarily relies on spatial constraints, the sum of squared deviations (SSD), the minimum spanning tree (MST), and edge removal to ensure that the observations within each cluster are geographically adjacent.

$$pF = \frac{(T-W)/(k-1)}{W/(n-k)} \tag{3}$$

The pseudo-F statistic (pF) evaluates clustering performance by calculating its value for different numbers of clusters. In the formula, T represents the total sum of squares, W represents the intra-cluster sum of squares, k is the number of clusters, and n is the total number of observations. The pseudo-F statistic represents the between-cluster variance to within-cluster variance. The optimal number of clusters is determined using the elbow rule. As the number of clusters increases, the rate at which the pseudo-F value decreases slows significantly or even shows a slight increase; the inflection point is the elbow, and the corresponding number of clusters is the optimal number.

2.3.3. An Optimized Adaptive Stratified Sampling Scheme for Multidimensional Heterogeneity

To accurately reflect the complexity of surface processes in the study area and ensure a sample set representative of land cover types, this study combines the landscape framework provided by "landform zoning" with "phenological feature clustering" to generate "minimum hierarchical units" with both geographical continuity and environmental relevance. Based on this, a basic unit for sample allocation is constructed, driven by the dual factors of

"spatial representativeness" and "internal variability." Using ArcGIS Pro tools, with the aforementioned basic units as the stratification units, the coefficient of variation and area index of phenological indices were calculated and subsequently normalized to quantify the phenological complexity of each stratification unit. The sample allocation for each stratification unit was then determined by calculating the proportional allocation.

The relevant calculation formulas are as follows:

$$CV = \frac{S}{\overline{X}} \tag{4}$$

The coefficient of variation (CV) is a statistical measure of the relative dispersion of data, where S represents the standard deviation of the phenological indices within each hierarchical unit, and $\overline{X}$ represents the mean.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{5}$$

The normalized value (X_{norm}) is a data processing method that facilitates the quantification of the weighting relationship between the area size of a stratification unit and the dispersion of phenological indices. X_{min} is the minimum value in a dataset, and X_{max} is the maximum value in a dataset.

2.3.4. Land Use/Cover Classification

1. Sampling

The performance of sampling methods in karst regions is summarized in Table 2. This study selected stratified random sampling as the sampling method for optimizing the sample space. After dividing the study area into a certain number of internally homogeneous but externally heterogeneous sub-areas using remote sensing phenological patterns, the study combined these with landform sub-areas to obtain multiple sub-areas with consistent phenological-background and landform characteristics. Stratified random sampling is suitable when there are significant differences among the various sub-regions of the study area, and when individual sites possess characteristics that are distinctly different from those of other regions.

Table 2. The applicability of different sampling methods in karst regions [35].

Sampling Design	Principle	Applicability in Karst Areas
Simple Random Sampling	Samples are selected completely randomly across the entire study area.	Use with caution. It cannot ensure sufficient representativeness of all important feature types in the fragmented karst landscape.
Systematic Sampling	Sampling is conducted at fixed spatial intervals (grid) within the study area.	Use with caution. It is acceptable for macro-pattern analysis but may fail to capture the patch randomness under karst microtopography.
Stratified Random Sampling	Taking classification maps or prior knowledge as "strata", random sampling is performed (proportionally or with a fixed quantity) within each category (stratum).	Highly recommended. It is the most commonly used and scientific method for land cover classification accuracy assessment and can effectively address the uneven area distribution of feature types in karst areas.
Spatial Balanced Sampling	Specific algorithms (e.g., GRTS) are adopted to achieve uniform spatial distribution of samples while maintaining randomness.	High potential. It is particularly suitable for in-depth studies requiring spatial statistical analysis and uncertainty modeling, as it can more objectively reflect karst spatial heterogeneity.

2. Feature Parameter Selection

The effectiveness of Random Forests in land cover classification has been well-established, outperforming other algorithms such as SVM and CART [35,36]. It is widely used globally and is suitable for various applications, including urban expansion monitoring, long-term land cover dynamics analysis, and ecological and environmental assessments [19–21]. Sample points were selected through a combination of visual interpretation and historical imagery, covering the major land cover types within the study area, and were randomly divided into training and validation sets for model training and classification accuracy evaluation. In the model parameter settings, the Random Forest classifier was implemented using ee.Classifier.smileRandom Forest in Google Earth Engine. The number of decision trees was set to 100, while the remaining key parameters were kept as the default settings of Google Earth Engine. The same Random Forest configuration was used for all sampling schemes to ensure that the accuracy comparison mainly reflected differences in sample organization rather than classifier parameterization. The key parameter settings are listed in Table 3.

Table 3. Key Random Forest parameter settings used in this study.

Parameter	Setting
Number Of Trees	100
Variables Per Split	Sqrt (number of input variables)
Min Leaf Population	1
Bag Fraction	0.5
Max Nodes	no limit
seed	0

In the Random Forest classification, the selection and processing of feature parameters have a significant impact on model performance. Appropriate feature selection can improve classification accuracy, reduce the risk of overfitting, and enhance the model's interpretability. Additionally, mathematical calculation methods for feature parameters—such as the calculation of vegetation indices and water indices—can enhance the model's ability to distinguish between different land cover types. The feature parameters selected for this study include raw spectral bands, vegetation indices, water indices, building indices, and terrain features. Each parameter is derived directly from remote sensing data or calculated from spectral bands; the specific calculation methods are as follows:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{6}$$

$$EVI = 2.5 \times \frac{NIR - Red}{NIR + 6 \times Red - 7.5 \times Blue + 1} \tag{7}$$

The Normalized Difference Vegetation Index ($NDVI$) is a widely used vegetation index that effectively reflects vegetation growth conditions. The Enhanced Vegetation Index (EVI), by introducing additional coefficients, can more accurately reflect vegetation photosynthesis and chlorophyll content. NIR denotes the near-infrared band (B8), Red denotes the red band (B4), and Blue denotes the blue band (B2).

$$SAVI = \frac{(1 + L) \times (NIR - Red)}{NIR + Red + L} \tag{8}$$

The soil-adjusted vegetation index ($SAVI$) incorporates a soil brightness correction factor to better account for the influence of the soil background; L is the soil brightness correction factor, typically set to 0.5.

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{9}$$

$$MNDWI = \frac{Green - SWIR1}{Green + SWIR1} \tag{10}$$

The Normalized Difference Water Index ($NDWI$) effectively distinguishes between water and non-water areas; by incorporating the shortwave infrared band, $MNDWI$ can identify water bodies more accurately. $Green$ denotes the green band (B3). $SWIR1$ denotes the shortwave infrared band (B11).

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR} \tag{11}$$

The Normalized Building Index ($NDBI$) can effectively distinguish built-up areas from other land cover types.

2.3.5. Accuracy Validation

Constructing a confusion matrix for classification results is a commonly used method for evaluating the accuracy of land cover data. A confusion matrix is an n × n matrix, where n is the number of classification categories. The rows of the matrix represent the actual categories, and the columns represent the classification results. Using the confusion matrix, the following accuracy metrics were calculated: overall accuracy, the Kappa coefficient, class-level user's accuracy, and producer's accuracy:

$$OA = \frac{\sum_{i=1}^{k} N_{ii}}{N} \tag{12}$$

$$UA = \frac{N_{ii}}{N_{i+}} \tag{13}$$

$$PA = \frac{N_{ii}}{N_{+i}} \tag{14}$$

$$Kappa = \frac{N\sum_{i=1}^{n} N_{ii} - \sum_{i=1}^{n}(N_{i+}N_{+i})}{N^2 - \sum_{i=1}^{n} N_{i+}N_{+i}} \tag{15}$$

Overall Accuracy (OA): This metric directly reflects the consistency between classification results and reference data. A higher overall accuracy indicates that the classification results are closer to the actual situation. User's Accuracy (UA): This metric indicates the proportion of pixels or samples classified as a given category that are correctly classified. A higher UA suggests greater reliability of that category in the classification results. Producer's Accuracy (PA): This metric indicates the proportion of reference samples that actually belong to a given category and are correctly classified. A higher PA suggests better classification performance for that category in the reference data. Kappa Coefficient ($Kappa$): The Kappa coefficient accounts for the difference between classification results and random classifications, enabling a more objective assessment of classification consistency. The relationship between the Kappa coefficient and classification quality is shown in Table 4. A higher Kappa coefficient indicates that the classification results possess greater accuracy and reliability.

Table 4. The relationship between the Kappa coefficient and classification quality.

Kappa Value Range	Interpretation
Kappa < 0	Poor agreement
0 ≤ Kappa < 0.2	Slight agreement
0.2 ≤ Kappa < 0.4	Fair agreement
0.4 ≤ Kappa < 0.6	Moderate agreement
0.6 ≤ Kappa < 0.8	Substantial agreement
0.8 ≤ Kappa < 1	Almost perfect agreement

3. Results

3.1. Remote-Sensing Phenological Pattern Zoning and Evaluation

Remote sensing phenology in the study area exhibits significant spatiotemporal variation. The results of the correlation analysis of the four phenological indices (Figure 5) indicate that the length of the growing season (LOS) is significantly and relatively evenly correlated with the other three indices. LOS is significantly positively correlated with the end of the growing season (EOS) (r = 0.678), meaning that the longer the vegetation growing season, the later the growing season ends; however, it is significantly negatively correlated with the start of the growing season (SOS) and the onset of maturity (OM), with correlation coefficients of −0.669 and −0.516, respectively, indicating that, when SOS and OM are expressed as day of year, a longer growing season is generally associated with earlier growth onset and earlier maturity timing. At the watershed scale, LOS broadly summarizes the spatial variations in vegetation phenology within the study area and can serve as a representative indicator for characterizing the patterns of phenological differentiation in this region.

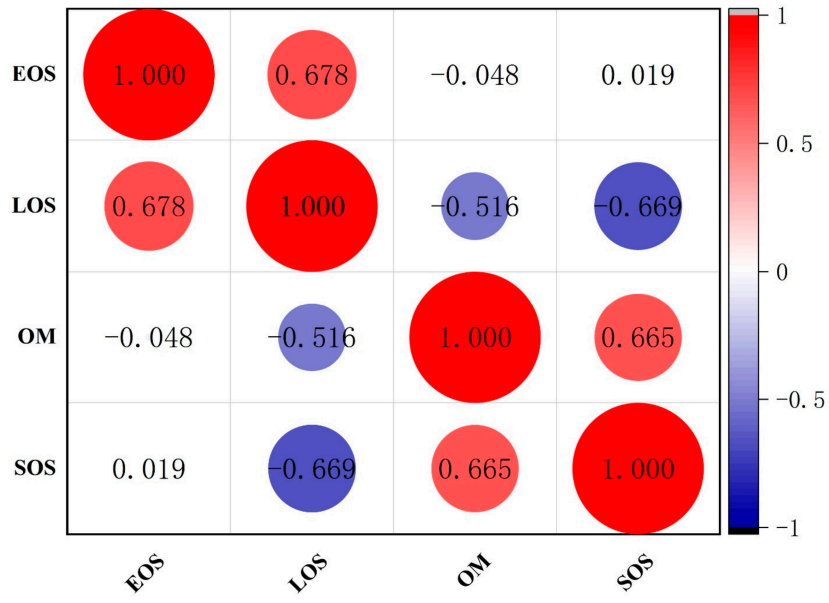


Figure 5. Spearman correlation test results for four phenological indicators within the region ($p < 0.05$).

The study area was divided into six subregions based on the LOS index (Figure 6). In the spatially constrained clustering algorithm, the pseudo-F value is a metric that measures the ratio of between-class variance to within-class variance; a higher value indicates greater between-class differences, smaller within-class differences, and better clustering results. However, in practical applications, the elbow rule is more applicable. As shown in Figure 7, the pseudo-F-score peaks when the number of partitions is 2, gradually decreases, and then

reaches an inflection point at 5. Although the pseudo-F-score is highest with 2 partitions, the SKATER algorithm essentially prunes along the minimum spanning tree in descending order of edge weights, which artificially inflates the error gradient of early partitions; when the number of partitions increases to 6, the standardized error gradient falls below the "degrees of freedom threshold" of $1/(n-2)$ for the first time, indicating that the marginal information gain has become lower than the cost of model complexity, forming an elbow point in the statistical-interpretability trade-off. Therefore, six partitions are selected as the optimal solution that balances internal homogeneity and policy operability.

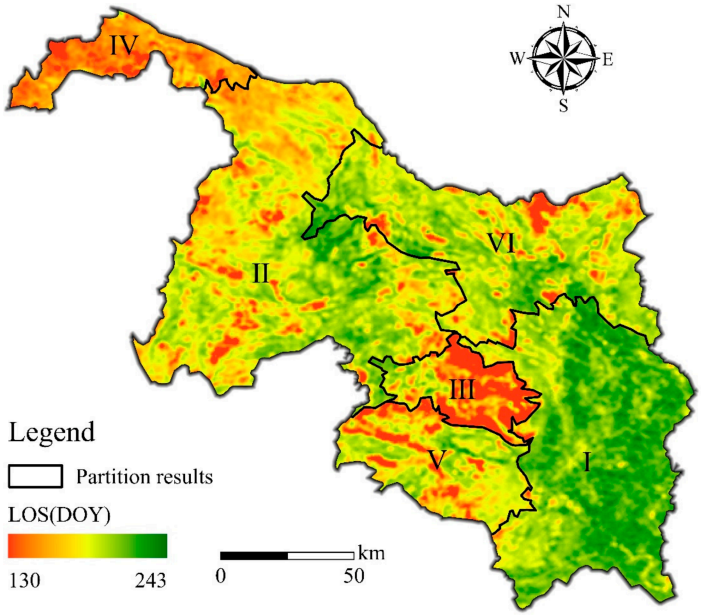


Figure 6. Results of remote sensing phenological pattern partitioning.

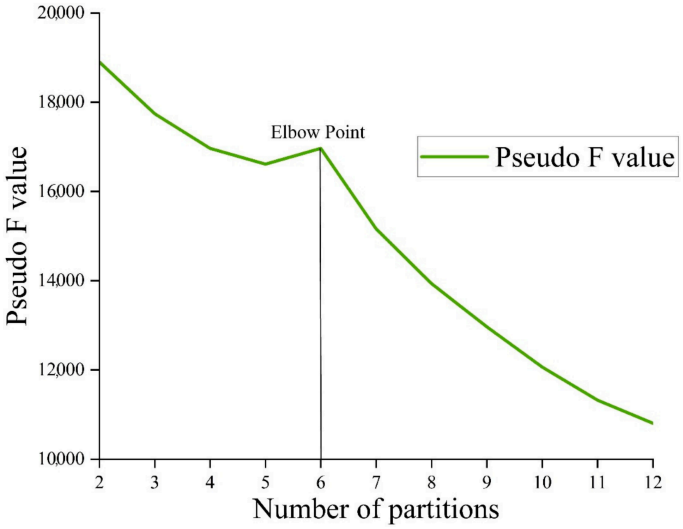


Figure 7. Number of partitions and pseudo-F-value variable.

The zoning results were evaluated based on statistical tests, and box plots were used to describe the distribution of growing season lengths across the six zones, as shown in Figure 8. The results indicate that there are statistically clear differences between the different zones. Within each subregion, vegetation is concentrated within relatively fixed time ranges; subregions II and V, which have similar time ranges, also exhibit significant spatial differences. Therefore, the phenological characteristics of vegetation within each subregion show a high degree of similarity; this result supports the use of the zones as

background strata for sample allocation. It can, to a certain extent, describe the spatial differentiation patterns of surface vegetation phenology remote sensing models.

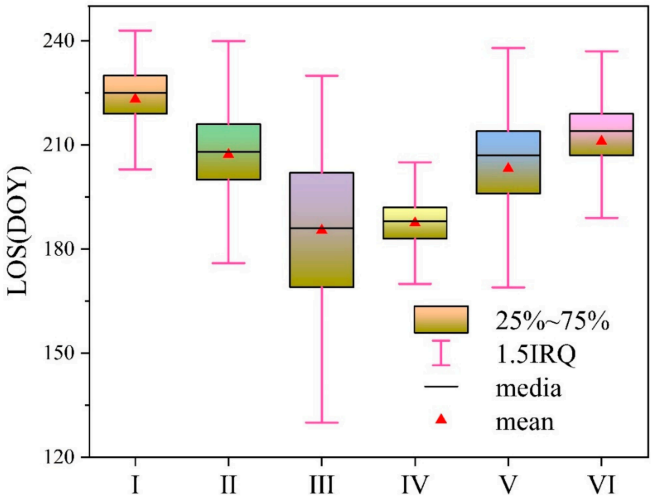


Figure 8. Boxplot of statistical test results for remote sensing phenological model zones.

The length of the growing season generally exhibits a gradual increase from south to north. Subdivision III, which deviates from this overall gradient pattern, serves as the central point of five spatially adjacent subdivisions (excluding Subdivision IV). Its growing season length is shorter than that of other subdivisions, at 186 ± 16 days. In contrast, the adjacent Subdivision I has the longest growing season among the six subdivisions, reaching 225 ± 15 days; the growing seasons of Sub-regions II, V, and VI—which are spatially adjacent to Sub-region III—are 208 ± 8 days, 205 ± 9 days, and 213 ± 6 days, respectively; and Sub-region IV, which is spatially adjacent only to Sub-region II, has a smaller area and relatively uniform vegetation phenological characteristics, resulting in a growing season length of 188 ± 4 days. The growing seasons of Subregion II and Subregion V are relatively consistent; however, although these two subregions are spatially adjacent, the shorter growing seasons in parts of Subregion III and the northern part of Subregion V serve as a demarcation line separating Subregion II and Subregion V.

3.2. Stratified Results of Phenology–Landform Joint Stratification and Spatial Distribution of Samples

By overlaying Tang Guo'an's landform zoning map with the surface phenological constraint-based clustering zones (Figure 9), we generated basic hierarchical units that integrate static landform backgrounds with dynamic phenological responses. The overlay resulted in a total of 43 hierarchical units, ranging in area from a minimum of 0.59 km^2 to a maximum of 3038.88 km^2. The CV coefficients of these units ranged from a minimum of 0 to a maximum of 0.13. The CV coefficients generally range from low to moderately low, indicating that as the basic stratification units were established through the integration of phenological zoning and topographic classification, the internal phenological complexity of each unit was reduced. This may enhance the spatial representativeness of the sample, and the results suggest the practical value of the joint stratification strategy: It preserves the macro-geographic consistency of topographic units while capturing the local heterogeneity of phenological dynamics.

By integrating the area weights and complexity weights of phenological characteristics within each hierarchical unit, a composite score was obtained, and sample allocation was subsequently carried out. After ranking the area weights from largest to smallest and summing them for each stratification unit, the sample allocation for the stratification units

was determined. Among the top 20% of stratification units, the cumulative area weights accounted for 68.37% of the total; however, after introducing the phenological complexity weighting index, the cumulative curve of the composite scores for the same top 20% of stratification units decreased to account for only 32.29% of the total (Figure 10). The curve also approaches a diagonal line more closely. This result quantitatively reveals that the LOS coefficient of variation can dilute the dominance of the 'large area = high weight' principle whilst maintaining spatial representation.

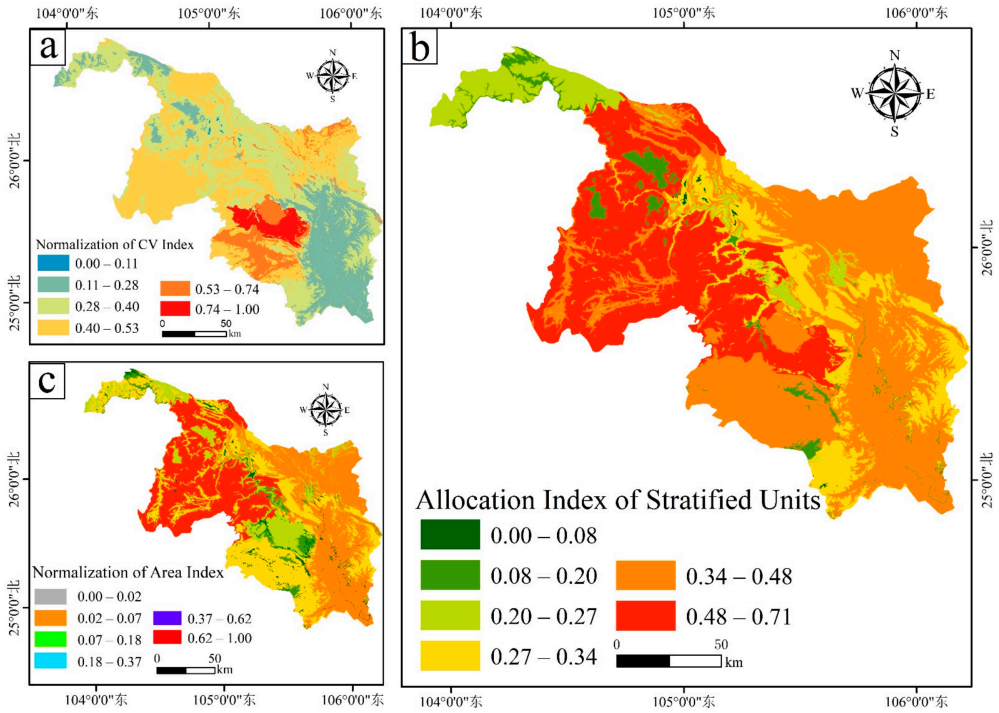


Figure 9. Superimposition of phenological zones and basic topographic zoning units. (**a**) Distribution of the coefficient of variation for hierarchical units in the study area; (**b**) Distribution of the allocation index for hierarchical units in the study area; (**c**) Dual-weighted allocation scores for hierarchical units.

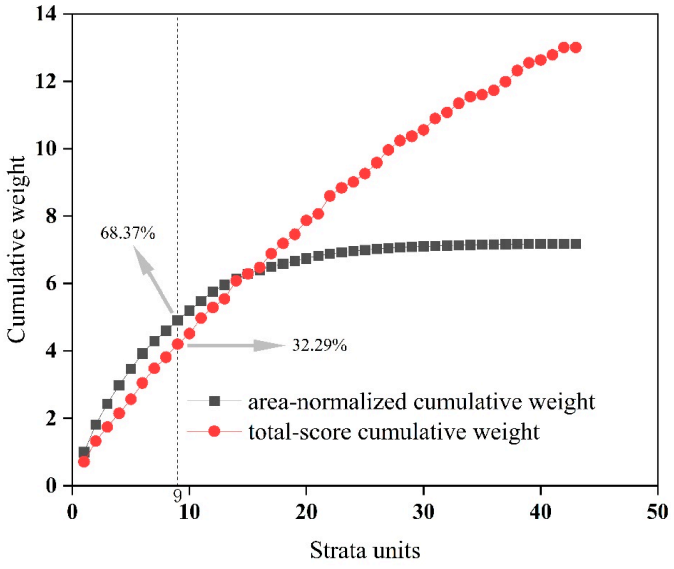


Figure 10. Cumulative weight distribution for two categories within stratification units and the weight dilution effect on the coefficient of variation of LOS.

Combined with the achieved effective dilution of the area effect, a dual-weighted sample allocation scheme was applied to the stratification unit differentiation coefficient and area index. The primary calculation method involved multiplying the number of allocated samples—derived by comparing the normalized indices of each stratification unit against the total and then multiplying by the total sample size—by the spatial distribution of samples across each partition (Figure 11), with sample counts represented by pie charts.

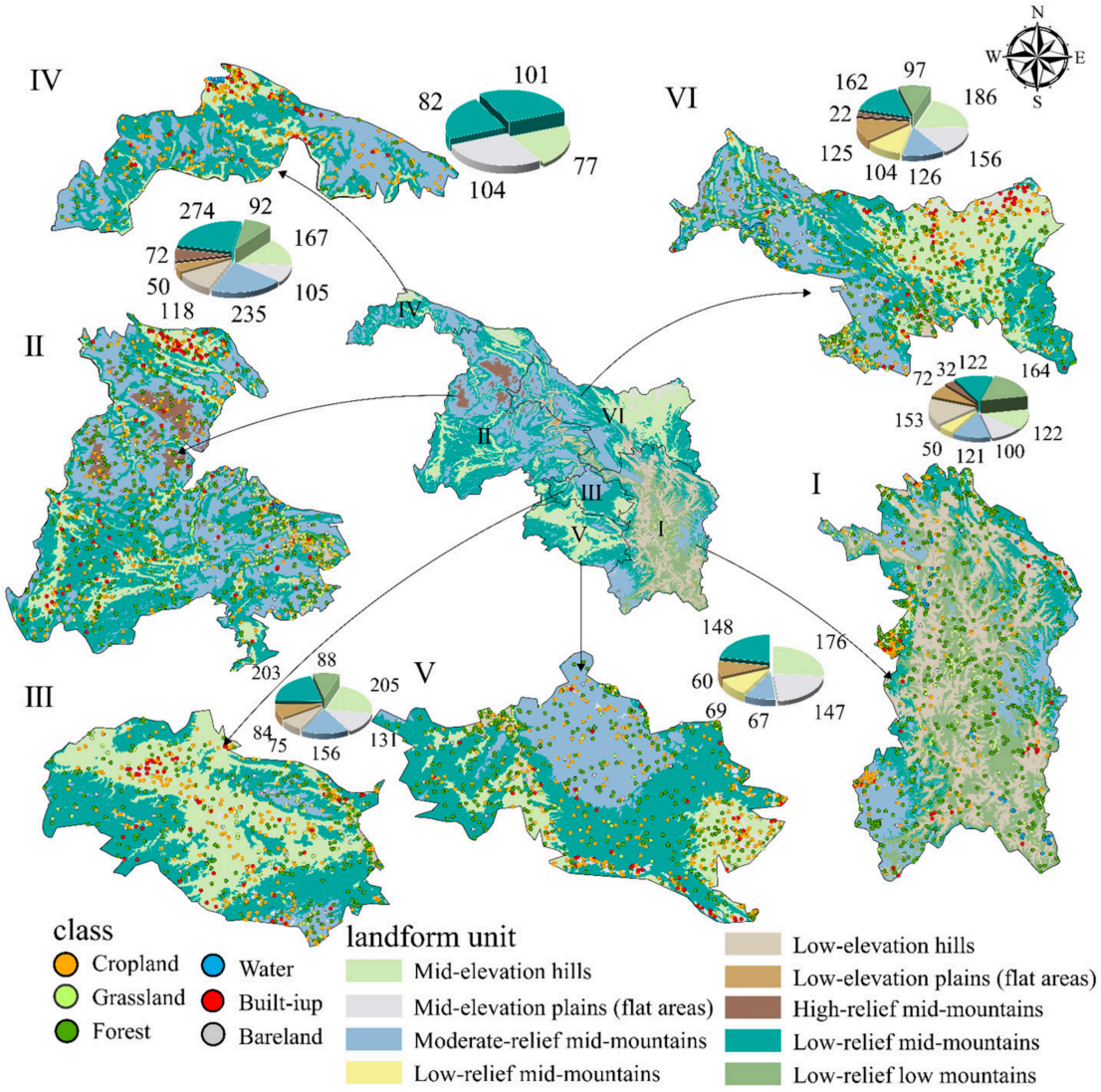


Figure 11. Sample space distribution for optimized stratified random sampling.

Simple random sampling was selected as the control scheme (Figure 12). Ultimately, simple random sampling and sample optimization were compared; both schemes were based on the sample set, with 5000 sample points selected according to their respective methods. Under the simple random sampling scheme, sample points are uniformly and randomly distributed across the study area. In contrast, the sample optimization scheme selects a larger number of samples in regions with complex phenological variation and a smaller number in regions with mild phenological variation (Figure 11). This approach is intended to help capture more surface features in complex regions for Random Forest learning.

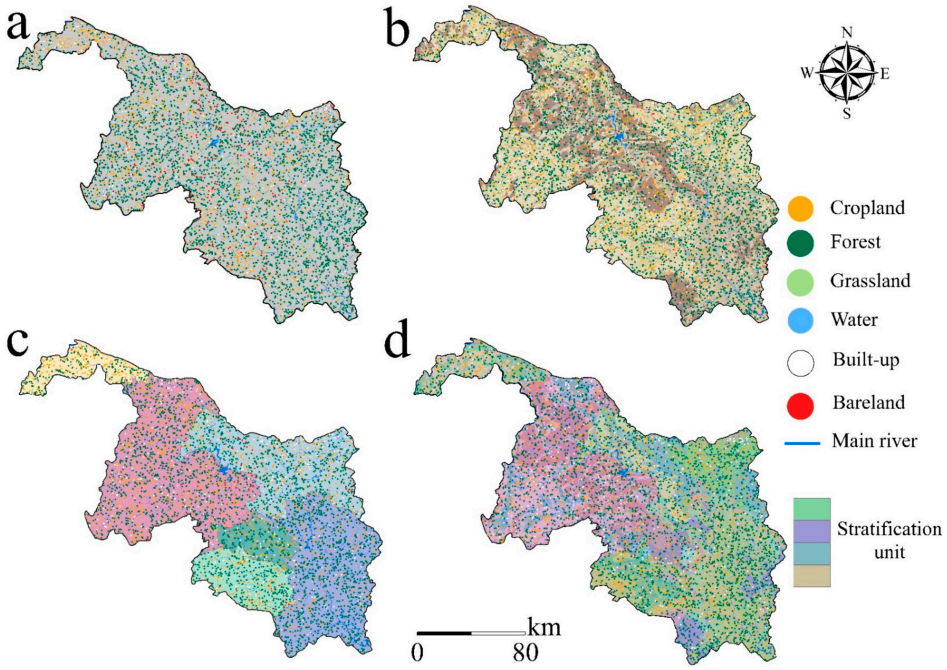


Figure 12. Spatial distribution of samples under the four comparison sampling schemes. (**a**) Simple random sampling; (**b**) landform-based sampling; (**c**) phenology-based sampling; (**d**) phenology–landform joint stratified sampling.

3.3. Response of the Sample Optimization Scheme to LULC Remote Sensing Interpretation

3.3.1. Ablation Analysis of Phenology, Landform, and Dual-Weighted Allocation

To identify the contribution of each component in the sample optimization framework, five sampling schemes were compared under the same total sample size, input features, and Random Forest configuration. These schemes included simple random sampling (T1), phenology-based area allocation (T2), landform-based area allocation (T3), phenology–landform joint stratification with area-based allocation (T4), and phenology–landform joint stratification with dual-weighted allocation (T5) (Table 5).

Table 5. Comparison of classification accuracy among ablation sampling schemes.

Scheme	Strategy	*OA* (%)	Δ*OA* (%)	*Kappa*	Δ*Kappa*
T1	Simple random	71.33	—	0.43	—
T2	Phenology-area	73.74	+2.41	0.49	+0.06
T3	Landform-area	73.21	+1.88	0.48	+0.05
T4	Joint-area	75.86	+4.53	0.56	+0.13
T5	Joint-dual	77.55	+6.22	0.62	+0.19

The results show a gradual increase in classification accuracy from T1 to T5. Simple random sampling achieved an OA of 71.33% and a Kappa coefficient of 0.43. After introducing phenological stratification alone, OA increased to 73.74% and Kappa increased to 0.49. The landform-only scheme also improved the result, with an OA of 73.21% and a Kappa coefficient of 0.48. These two results indicate that both long-term phenological background and landform heterogeneity provided useful information for sample organization. The joint stratification scheme further improved OA to 75.86% and Kappa to 0.56. The final dual-weighted allocation scheme achieved the highest accuracy, with an OA of 77.55% and a Kappa coefficient of 0.62. Compared with simple random sampling, OA increased by 6.22 percentage points and Kappa increased by 0.19. Compared with the joint-area scheme, the dual-weighted allocation further increased OA by 1.69 percentage points. This

suggests that the final scheme improved sample allocation not only by considering the area of each stratum but also by increasing the sampling weight of strata with greater internal LOS variability.

3.3.2. Class-Level Accuracy Response of the Optimized Sampling Scheme

To further examine the classification response of different LULC categories, the producer's accuracy (PA) and user's accuracy (UA) of simple random sampling and the final optimized scheme were compared (Table 6). The optimized scheme improved at least one class-level accuracy indicator for all six classes, but the improvement magnitude differed among categories.

Table 6. Class-level PA and UA comparison between simple random sampling and the optimized scheme.

Class	T1		T5		Change	
	PA	*UA*	*PA*	*UA*	Δ*PA*	Δ*UA*
Cropland	80.47	61.23	88.92	66.74	+8.45	+5.51
Forest	85.55	82.27	90.68	84.96	+5.13	+2.69
Grassland	20.73	40.58	29.14	49.37	+8.41	+8.79
Water	79.62	76.47	93.18	83.52	+13.56	+7.05
Built-up	43.86	69.81	55.24	78.73	+11.38	+8.92
Bareland	3.92	30.77	6.11	44.44	+2.19	+13.67

For vegetation-related classes, the optimized scheme showed clear improvements. The PA of cropland increased from 80.47% to 88.92%, and the PA of forest increased from 85.55% to 90.68%. Grassland remained a relatively difficult class, but its PA increased from 20.73% to 29.14%, and its UA increased from 40.58% to 49.37%. This indicates that the optimized sampling scheme improved the representation of fragmented grassland patches to some extent. For non-vegetated or weakly vegetated classes, water and built-up land also showed obvious improvements. The PA of water increased from 79.62% to 93.18%, and the PA of built-up land increased from 43.86% to 55.24%. The improvement in unutilized land was limited in PA, increasing from 3.92% to 6.11%, although its UA increased from 30.77% to 44.44%. This suggests that unutilized land was still strongly affected by confusion with grassland, exposed rock, and dry cropland in the karst landscape.

3.3.3. Local Comparison of LULC Classification Results

Based on the results of remote sensing interpretation for three local areas using different sampling methods (Figure 13), the optimized hierarchical unit random sampling LULC results are closer to the reference interpretation for shallow grassland (a). In contrast, in the simple random sampling sample (a1) for the same plot, the grassland plot was misclassified as unutilized land. Similarly, comparing (b) and (b1), the stratified unit random sampling LULC results provide more accurate identification of cropland in mixed cropland-forest areas. Additionally, for small, scattered riverbed areas along rivers, the results correctly identify them as grassland rather than the "built-up land" classification produced by simple random sampling. In (c) and (c1), for contiguous grassland–forest mixed cover types, driven by phenological patterns, the optimized method provides more accurate identification of contiguous grasslands.

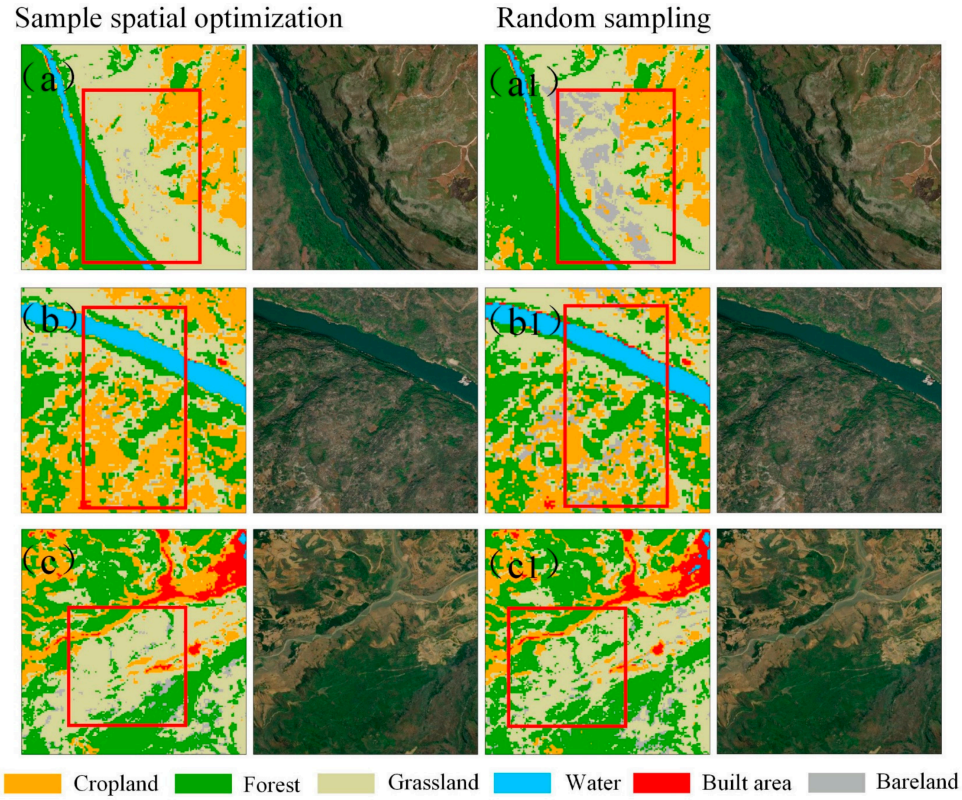


Figure 13. Comparison of local inversion results. (**a**,**a1**) The optimized method better identifies shallow grassland, while simple random sampling misclassifies part of it as bare land; (**b**,**b1**) The optimized method more accurately identifies cropland and scattered riverbed grassland, while simple random sampling shows confusion with forest and built-up land; (**c**,**c1**) The optimized method better distinguishes contiguous grassland from forest.

4. Discussion

4.1. Impact of Remote Sensing Phenological Model-Dominated Zones on Karst LULC Samples

Based on the SKATER spatial constraint clustering algorithm and MODIS phenological datasets, this study identified that the optimal number of phenological pattern zones for the Beipanjiang River Basin (Guizhou section)—a typical karst landscape with complex topography—is six, refining the conclusions of Liu's study at this scale in China [34]. The resulting zones were then used for stratified sampling and sample allocation. By incorporating phenological information with a temporal dimension, the study helps the hierarchical information to better reflect differences in vegetation growth cycles, offering a complement to traditional static map-based stratification. Using ArcGIS Pro spatial analysis tools and guided by the designed sample distribution schemes, random samples were selected for training and classification in the Random Forest model; through visual interpretation, samples with uncertain labels caused by cloud/shadow, temporal inconsistency, ambiguous boundaries, or inconsistent interpretation were treated as label-quality-control cases; difficult but interpretable mixed or boundary samples were retained to avoid overestimating classification accuracy.

Previous studies have shown that the accuracy of publicly available LULC datasets in this region is generally low: common LULC products such as GlobeLand30, CCI-LC, and CGLS-LC have accuracies of approximately 40–52% in the Southern Karst region, and the overall LULC accuracy in karst areas (36.9%) is significantly lower than in non-karst areas (51.2%) [29]. At the same time, differences in sample representativeness stem from the combined effects of natural environments and human activities [37]. This study theo-

retically expands the perspective of remote sensing classification sample design methods by incorporating phenological information into sample stratification, thereby providing a sampling-design option for improving sample representativeness in complex surface environments. The paper automatically generates a more balanced sample configuration through phenological zoning, which is consistent with previous conclusions regarding sampling design efficiency [38]. In practice, this study suggests that targeted sample collection, tailored to the complex topography and phenological characteristics of karst terrain, can reduce error variance, improve estimation accuracy, and bring the results closer to the reference labels under the current validation setting. Phenological zoning can also be integrated with training sample selection to help improve classification accuracy. For example, training samples can be selected evenly within each zone based on phenological zoning results, avoiding oversampling of dominant classes and thereby enhancing the classification model's ability to distinguish minor classes. The role of phenology should be interpreted as direct for vegetation-related classes and indirect for non-vegetated classes through surrounding ecological background and mixed-boundary constraints.

4.2. Analysis of Sample Optimization Gain Sources

Based on the "phenology–landform joint stratification + dual-weight sample allocation" method, the study results suggest: The first source of gain is "reduction of intra-class variance": Differences in landform positions cause spectral and textural dispersion within the same land cover class. Hierarchical grouping by landform units decomposes the number of samples mixed across landform positions into more homogeneous subspaces, thereby reducing intra-class noise and enhancing inter-class distinguishability [39]. The study area was divided into 43 hierarchical units based on six optimal phenological patterns and landform classifications; these hierarchical units, established by comprehensively considering phenological indices and landform types, were intended to reduce heterogeneity within training strata, thereby contributing to the observed classification gains.

The second source of gain may be related to the "rehomogenization" of the temporal dimension through phenology: in the context of big data mapping, optimizing the selection of training data and implementing localization are considered effective ways to improve mapping results, while dynamic phenological information can provide an additional discriminatory dimension for distinguishing between categories with similar seasonal phases [40]. The study converted remote sensing phenological data from 2001 to 2020 into Days of the Year (DOY) data to guide sample stratification. This approach allows for the construction of a more appropriate stratification framework using prior information such as NDVI to address surface heterogeneity, ensuring that stratification is not only spatially consistent but also more consistent in terms of vegetation activity intensity [41]. By coupling the LOS coefficient of variation with topographic zoning and remote sensing phenology model zones, spectral conflation areas resulting from phenological changes received a higher allocation of samples. This helped reduce the spatial bias caused by random sampling and provided a more representative feature training set for the study area's complex karst terrain, including deeply incised karst topography and the transition zone between cropland and forest, suggesting the practical value of the remote sensing phenology-driven sampling scheme under the current study-area and validation settings.

The third type of gain stems from dual-weight allocation, with its core benefit lying in the alignment of statistical optimality and learning difficulty. First is the allocation of variance within strata: theoretically, more samples should be allocated to strata with higher intra-stratum variance to improve estimation efficiency under a limited sample size budget [42]. Empirically, stratified sampling can improve representation of non-dominant classes and alleviate class imbalance, preventing accuracy variations from being confined to

dominant land cover types [12]. Next is the allocation of area within strata: in fragmented landscapes, area-weighted sampling ensures overall representativeness, while complexity-weighted sampling directs the budget toward transition zones and highly variable strata, aligning with the empirical recommendation that "areas with complex patterns should be prioritized for sampling [43]. Given the coupled effects of geographic heterogeneity and spatial autocorrelation in karst regions, sampling schemes that balance both factors may be more effective in reducing the risk of overfitting caused by spatial clustering of sampling points. Future research could focus on developing a unified optimization framework tailored to different complex land surfaces to systematically evaluate the applicability and mechanistic differences in this stratification scheme across scales and regions.

4.3. Limitations and Outlook

The results of this study indicate that, under the same total sample size, an optimized sampling strategy based on phenology–landform joint stratification was associated with improved Random Forest classification performance in complex karst terrain. However, the study still has limitations:

(1) Although this study utilized 20-year-scale phenological change data and landform classification data to cover the overall spatial scale of the study area as much as possible, the data-driven approach of spatially constrained clustering of the study area is dependent on the quality of the input data and the input parameters. However, the phenological data in this study come from a single source, and there are still limitations in handling the nonlinear synergistic mechanisms of multiple factors (such as slope, rock type, and variance of phenological indicators). Future research could combine conventional machine learning algorithms with geographically weighted regression to further explicitly quantify the local contributions of each factor to spectral aliasing, thereby enhancing the extrapolation capability and interpretability of the sample allocation scheme.

(2) Although the Random Forest algorithm performs well in large-scale LULC mapping, as well as in terms of tolerance to sample noise and computational efficiency, the use of a single algorithm makes it difficult to distinguish the differences in adaptability among various algorithms under complex karst terrain and to mitigate systematic errors through modeling. Future research could concurrently introduce more deep learning ensemble models, comprehensively compare and selectively integrate the spectral-phenological discrimination strengths of each model, thereby improving the stability and interpretability of complex land cover classification.

(3) The 500 m MODIS phenological zones and 10 m Sentinel-2 classification data involve an unavoidable scale mismatch, so the phenological zones should be interpreted as background strata rather than pixel-level predictors. The 2001–2020 average phenology represents a long-term spatial gradient and may not fully match the LULC condition of the target mapping year, especially in areas affected by ecological restoration, cropland abandonment, or construction expansion. The current random split with spacing constraints reduces but does not fully eliminate spatial autocorrelation; future work should adopt spatial block or buffered validation and report uncertainty estimates and class-level confidence intervals.

(4) Although a 500 m minimum-distance constraint was used to reduce sample redundancy, random splitting may still leave residual spatial autocorrelation between training and validation samples. Future work should use spatial-block validation, buffered validation, or independent probability-based validation samples to further test the transferability and robustness of the proposed sampling strategy. In addition, although class-level PA and UA were reported, complete error matrices, repeated sampling uncertainty, and con-

fidence intervals should be further provided in future work to more rigorously quantify classification uncertainty.

5. Conclusions

This study focuses on the Beipanjiang River Basin (Guizhou section) in southern China's karst region. Addressing the issue of insufficient spatial representativeness of training samples in LULC classification of complex karst surfaces, we established a sample optimization framework comprising "remote sensing phenological zoning—phenology–landform joint stratification—dual-weighted sample allocation" and evaluated it using the Google Earth Engine platform and a Random Forest classifier. The main conclusions are as follows:

(1) Among the selected phenological indices—SOS, OM, EOS, and LOS—the length of the growing season (LOS) broadly represents the spatial variation in vegetation phenology within the study area. Combined with the results of spatially constrained clustering, the study area was ultimately divided into six phenological zones with good spatial continuity and relative internal homogeneity. This use of LOS is a simplification for stratification and does not exclude the independent roles of SOS, OM, and EOS.

(2) The dual-weighted allocation method, which integrates area weighting and phenological complexity weighting, increased sample coverage in highly heterogeneous areas while maintaining overall spatial representativeness.

(3) With the total training sample size kept constant at 5000, the optimized sampling scheme improved the overall accuracy of LULC classification from 71.33% to 77.55% and increased the Kappa coefficient from 0.43 to 0.62 compared to simple random sampling, suggesting that, under the current study-area, sample-size, and model settings, the spatial configuration of samples can influence the classification results of complex karst land surfaces.

Author Contributions: All authors contributed to the manuscript. Conceptualization, Y.L. (Ya Li) and D.H.; methodology, Y.L. (Ya Li); software, Q.D., R.F. and H.L.; validation, D.H. and C.H.; formal analysis, Y.L. (Ya Li); data curation, Y.L. (Ya Li); writing—original draft preparation, Y.L. (Ya Li); writing—review and editing, Y.L. (Ya Li) and D.H.; visualization, Y.L. (Ya Li), Y.L. (Ying Luo) and Y.Y.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Geography Group-Based Higher Education Development Provincial Subsidy Fund (210-C426002); Guizhou Provincial 2025 Central Government Guided Local Science and Technology Development Fund Project (Qian Ke He Zhong Yin Di [2025] 031); Guizhou Provincial Key Technology R&D Program (Qiankehe [2023]; General No. 211); Guizhou Provincial Key Laboratory Construction Project (Qian Ke He Ping Tai [2025] 014).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
2. Wang, Y.; Sun, Y.; Cao, X.; Wang, Y.; Zhang, W.; Cheng, X. A Review of Regional and Global Scale Land Use/Land Cover (LULC) Mapping Products Generated from Satellite Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2023**, *206*, 311–334. [CrossRef]

3. Naboureh, A.; Li, A.; Bian, J.; Moharrami, M.; Ebrahimy, H.; Lei, G.; Nan, X.; Zhang, Z.; Feizizadeh, B.; Dabove, P.; et al. Accuracies, Discrepancies, and Challenges of the 10 m Global Land Cover Products in Mountains. *GIScience Remote Sens.* **2025**, *62*, 2556064. [CrossRef]
4. Sovann, C.; Olin, S.; Mansourian, A.; Sakhoeun, S.; Prey, S.; Kok, S.; Tagesson, T. Importance of Spectral Information, Seasonality, and Topography on Land Cover Classification of Tropical Land Cover Mapping. *Remote Sens.* **2025**, *17*, 1551.
5. Schug, F.; Pfoch, K.A.; Pham, V.-D.; Van Der Linden, S.; Okujeni, A.; Frantz, D.; Radeloff, V.C. Land Cover Fraction Mapping across Global Biomes with Landsat Data, Spatially Generalized Regression Models and Spectral-Temporal Metrics. *Remote Sens. Environ.* **2024**, *311*, 114260.
6. Li, Y.; Sun, B.; Gao, Z.; Su, W.; Wang, B.; Yan, Z.; Gao, T. Extraction of Rocky Desertification Information in Karst Area by Using Different Multispectral Sensor Data and Multiple Endmember Spectral Mixture Analysis Method. *Front. Environ. Sci.* **2022**, *10*, 996708. [CrossRef]
7. Xu, A.; Wang, F.; Li, L. Vegetation Information Extraction in Karst Area Based on UAV Remote Sensing in Visible Light Band. *Optik* **2023**, *272*, 170355. [CrossRef]
8. Wang, K.; Zhang, C.; Chen, H.; Yue, Y.; Zhang, W.; Zhang, M.; Qi, X.; Fu, Z. Karst Landscapes of China: Patterns, Ecosystem Processes and Services. *Landsc. Ecol.* **2019**, *34*, 2743–2763. [CrossRef]
9. Huang, D.; Zhou, Z.; Zhang, Z.; Dai, Q.; Lu, H.; Li, Y.; Huang, Y. Land Use/Land Cover Remote Sensing Classification in Complex Subtropical Karst Environments: Challenges, Methodological Review, and Research Frontiers. *Appl. Sci.* **2025**, *15*, 9641. [CrossRef]
10. Wang, J.; Zhou, Z.; Zhu, M.; Wan, J.; Wu, X.; Liu, R.; Zheng, J. Study on the Response of Net Primary Productivity to Vegetation Phenology and Its Influencing Factors in Karst Ecologically Fragile Regions. *Atmosphere* **2024**, *15*, 1464. [CrossRef]
11. Feng, S.; Jiang, S.; Liu, X.; Zhang, L.; Gan, Y.; Xia, N.; Wu, W.; Zhou, C. Extraction of Abandoned Cropland Using Multisource Remote Sensing Images in Suburban Regions: A Case Study of Zengcheng, Guangdong Province. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 18055–18067. [CrossRef]
12. Zeng, H.; Wu, B.; Wang, S.; Musakwa, W.; Tian, F.; Mashimbye, Z.E.; Poona, N.; Syndey, M. A Synthesizing Land-Cover Classification Method Based on Google Earth Engine: A Case Study in Nzhelele and Levhuvu Catchments, South Africa. *Chin. Geogr. Sci.* **2020**, *30*, 397–409. [CrossRef]
13. Liu, X.; Frey, J.; Munteanu, C.; Still, N.; Koch, B. Mapping Tree Species Diversity in Temperate Montane Forests Using Sentinel-1 and Sentinel-2 Imagery and Topography Data. *Remote Sens. Environ.* **2023**, *292*, 113576. [CrossRef]
14. Dong, S.; Guo, H.; Chen, Z.; Pan, Y.; Gao, B. Spatial Stratification Method for the Sampling Design of LULC Classification Accuracy Assessment: A Case Study in Beijing, China. *Remote Sens.* **2022**, *14*, 865.
15. Shetty, S.; Gupta, P.K.; Belgiu, M.; Srivastav, S.K. Assessing the Effect of Training Sampling Design on the Performance of Machine Learning Classifiers for Land Cover Mapping Using Multi-Temporal Remote Sensing Data and Google Earth Engine. *Remote Sens.* **2021**, *13*, 1433. [CrossRef]
16. Foody, G.M.; Arora, M.K. An Evaluation of Some Factors Affecting the Accuracy of Classification by an Artificial Neural Network. *Int. J. Remote Sens.* **1997**, *18*, 799–810. [CrossRef]
17. Shao, Q.; Huang, C.; Xiao, Y.; Liu, L.; Liu, W.; Huang, R.; Zhou, C.; Weng, W.; Huang, J. Selecting of Global Phenological Field Observations for Validating Coarse AVHRR-Derived Forest Phenology Products Based on Spatial Heterogeneity and Temporal Consistency. *Ecol. Inform.* **2025**, *90*, 103216. [CrossRef]
18. Zhang, Q.; Zhang, Z.; Xu, N.; Li, Y. Fully Automatic Training Sample Collection for Detecting Multi-Decadal Inland/Seaward Urban Sprawl. *Remote Sens. Environ.* **2023**, *298*, 113801.
19. Belgiu, M.; Bijker, W.; Csillik, O.; Stein, A. Phenology-Based Sample Generation for Supervised Crop Type Classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102264. [CrossRef]
20. Karasiak, N.; Dejoux, J.-F.; Monteil, C.; Sheeren, D. Spatial Dependence between Training and Test Sets: Another Pitfall of Classification Accuracy Assessment in Remote Sensing. *Mach. Learn.* **2022**, *111*, 2715–2740. [CrossRef]
21. Suman, S.; Rawat, A.; Kumar, A.; Pant, N. Study of Training Parameters Effect in Noise Clustering Classifier for Handling Heterogeneity Within the Class for LULC Classification. *J. Indian Soc. Remote Sens.* **2025**, *53*, 1183–1196. [CrossRef]
22. Wang, J.; Li, M.; Yu, C.; Fu, G. The Change in Environmental Variables Linked to Climate Change Has a Stronger Effect on Aboveground Net Primary Productivity Than Does Phenological Change in Alpine Grasslands. *Front. Plant Sci.* **2022**, *12*, 798633. [CrossRef]
23. Peng, H.; Xia, H.; Chen, H.; Zhi, P.; Xu, Z. Spatial Variation Characteristics of Vegetation Phenology and Its Influencing Factors in the Subtropical Monsoon Climate Region of Southern China. *PLoS ONE* **2021**, *16*, e0250825. [CrossRef]
24. Li, C.; Ma, Z.; Wang, L.; Yu, W.; Tan, D.; Gao, B.; Feng, Q.; Guo, H.; Zhao, Y. Improving the Accuracy of Land Cover Mapping by Distributing Training Samples. *Remote Sens.* **2021**, *13*, 4594. [CrossRef]
25. Ahmed, S.A. Land Use and Land Cover Classification Using Machine Learning Algorithms in Google Earth Engine. *Earth Sci. Inform.* **2023**, *16*, 3057–3073.

26. Shao, S.; Zhang, H.; Fan, M.; Su, B.; Wu, J.; Zhang, M.; Yang, L.; Gao, C. Spatial Variability-Based Sample Size Allocation for Stratified Sampling. *Catena* **2021**, *206*, 105509. [CrossRef]
27. Xie, H.; Wang, F.; Gong, Y.; Tong, X.; Jin, Y.; Zhao, A.; Wei, C.; Zhang, X.; Liao, S. Spatially Balanced Sampling for Validation of GlobeLand30 Using Landscape Pattern-Based Inclusion Probability. *Sustainability* **2022**, *14*, 2479. [CrossRef]
28. Ye, N.; Morgenroth, J.; Xu, C.; Cai, Z. Improving Neural Network Classification of Indigenous Forest in New Zealand with Phenological Features. *J. Environ. Manag.* **2022**, *314*, 115134. [CrossRef]
29. Yang, Y.; Zhou, Z.; Huang, D.; Zhang, F.; Deng, F.; Du, S. A Comparative Analysis of the Applicability of Typical Land Cover Datasets in the Karst Strongly Heterogeneous Terrain in Southern China. *Geocarto Int.* **2024**, *39*, 2395318. [CrossRef]
30. Liu, Z.; Zhang, Y. Vegetation Cover Change and Its Response to Human Activities in the Southwestern Karst Region of China. *Front. Ecol. Evol.* **2024**, *12*, 1326601. [CrossRef]
31. Li, Y.; Ke, Q.; Zhang, Z. Millennial Evolution of a Karst Socio-Ecological System: A Case Study of Guizhou Province, Southwest China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15151. [CrossRef]
32. Tang, G.; Yang, X.; Li, F.; Xiong, L.; Li, S. Global Basic Landform Units Datasets. Version 1. Yangtze River Delta Science Data Center, National Earth System Science Data Sharing Infrastructure Nanjing, China, National Science & Technology Infrastructure of China, Nanjing, China. 2023. Available online: http://geodata.nnu.edu.cn/ (accessed on 28 May 2026).
33. *GB/T 21010-2017*; General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Current Land Use Classification. Standards Press of China: Beijing, China, 2017.
34. Liu, X.; Wang, Z.; Yang, X.; Cheng, W.; Zhang, J.; Liu, Y.; Liu, B.; Meng, D.; Zeng, X. Remotely-sensed phenology pattern regionalization for land cover classification of natural scenes: A case study in China. *Acta Geogr. Sin.* **2024**, *79*, 2206–2229.
35. Huang, X.; Zhou, Z.; Zhao, X.; Wu, G.; Long, Y.; Chen, J. Information Extraction and Characteristic Analysis of Cultivated Land Abandonment in Karst Rocky Desertification Mountainous Areas Based on Time-Series Vegetation Index. *Sci. Rep.* **2025**, *15*, 12554. [CrossRef]
36. Amini, S.; Saber, M.; Rabiei-Dastjerdi, H.; Homayouni, S. Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series. *Remote Sens.* **2022**, *14*, 2654.
37. Ahmed, R.; Zafor, M.A.; Trachte, K. Land-Use and Land-Cover Changes in Cottbus City and Spree-Neisse District, Germany, in the Last Two Decades: A Study Using Remote Sensing Data and Google Earth Engine. *Remote Sens.* **2024**, *16*, 2773.
38. Mu, X.; Hu, M.; Song, W.; Ruan, G.; Ge, Y.; Wang, J.; Huang, S.; Yan, G. Evaluation of Sampling Methods for Validation of Remotely Sensed Fractional Vegetation Cover. *Remote Sens.* **2015**, *7*, 16164–16182. [CrossRef]
39. Lehmkuhl, F.; Römer, W. Geomorphological Processes and Landforms in a Global Scale—Previous Concepts and Future Challenges from a German Perspective. *Z. Geomorphol.* **2022**, *64*, 53–71. [CrossRef]
40. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C. Land Cover Classification in an Era of Big and Open Data: Optimizing Localized Implementation and Training Data Selection to Improve Mapping Outcomes. *Remote Sens. Environ.* **2022**, *268*, 112780. [CrossRef]
41. Lv, T.; Zhou, X.; Tao, Z.; Sun, X.; Wang, J.; Li, R.; Xie, F. Remote Sensing-Guided Spatial Sampling Strategy over Heterogeneous Surface Ground for Validation of Vegetation Indices Products with Medium and High Spatial Resolution. *Remote Sens.* **2021**, *13*, 2674. [CrossRef]
42. Shao, Z.; Cheng, T.; Fu, H.; Li, D.; Huang, X. Emerging Issues in Mapping Urban Impervious Surfaces Using High-Resolution Remote Sensing Images. *Remote Sens.* **2023**, *15*, 2562.
43. Su, M.; Guo, R.; Chen, B.; Hong, W.; Wang, J.; Feng, Y.; Xu, B. Sampling Strategy for Detailed Urban Land Use Classification: A Systematic Analysis in Shenzhen. *Remote Sens.* **2020**, *12*, 1497. [CrossRef]

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