remote sensing

MDPI

Article

A Scene-Response-Driven Adaptive Segmentation Method for Burned Areas from Forest and Grassland Fires

Xinyu Hu [1,2], Xianlin Qin [1,2,*], Yakui Shao [1,2], Linfeng Yu [1,2], Shenghao Li [1,2], Feng Jiang [1,2] and Shuisheng Huang [1,2]

[1] Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China; hxyuinu@163.com (X.H.)
[2] Laboratory of Remote Sensing for Forestry and Grassland Monitoring and Evaluation, National Forestry and Grassland Administration, Beijing 100091, China
* Correspondence: qxl@ifrit.ac.cn; Tel.: +86-10-62888847

Highlights

What are the main findings?

- A scene-response-driven method integrates background constraints, response normalization, and core support for adaptive burned-area segmentation.
- BNCS achieved macro F1 scores of 91.83% and 89.56% on Sentinel-2 and Landsat 8/9 imagery, outperforming the best conventional thresholding method by 5.55 and 6.70 percentage points, respectively.

What are the implications of the main findings?

- Scene-adaptive thresholding and core-supported region selection provide an interpretable alternative to fixed-threshold burned-area delineation.
- BNCS maintained comparatively high segmentation accuracy as the analysis extent increased, supporting its application across heterogeneous fire scenes.

Abstract

Accurate burned-area delineation is essential for wildfire impact assessment and post-fire ecosystem monitoring, yet scene-dependent spectral responses limit the transferability of conventional threshold-based methods. This study proposes a background-constrained, response-normalized, and core-supported (BNCS) adaptive segmentation method that links scene-relative spectral response analysis with spatially supported region selection. Pre- and post-fire spectral constraints first define a candidate domain to reduce background interference. Scene-wise quantile normalization provides a relative response scale, and a geometric elbow of the area-response curve determines an adaptive low threshold. A root-mean-square-based high threshold then identifies burned cores, which support the retention of connected low-response regions while excluding unsupported patches. BNCS was evaluated across 12 fire events in northeastern, central, and southwestern China using multispectral imagery from Sentinel-2 and Landsat 8/9. It achieved macro F1 scores of 91.83% and 89.56%, respectively, exceeding the best conventional automatic thresholding method by 5.55 and 6.70 percentage points (pp). Across four threshold-transfer settings, BNCS exceeded training-derived fixed thresholds by 3.06–3.40 pp without reference-based calibration on test scenes. As the analysis extent increased from three to nine times the reference bounding-rectangle area, macro F1 decreased by 2.28 and 3.23 pp for Sentinel-2 and Landsat 8/9, respectively, while BNCS maintained comparatively high accuracy across background extents, demonstrating a favorable balance between segmentation accuracy and stability. Ablation experiments identified core support as the clearest conditional contributor to accuracy, while the effects of other components depended on their configuration.

Check for updates

Academic Editor: Ioannis Gitas

Received: 20 October 2026
Revised: 18 November 2026
Accepted: 25 November 2026
Published: 28 November 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.

Comparisons with two existing annual mapping products across 11 matched events showed more complete event-level coverage by BNCS, subject to differences in product objectives and temporal definitions. These findings support BNCS as an interpretable and transferable approach to event-specific burned-area mapping using multispectral satellite imagery across heterogeneous regions and observation conditions.

Keywords: dNBR; Sentinel-2; Landsat 8/9; scene heterogeneity; adaptive segmentation; threshold transferability

1. Introduction

Accurate delineation of burned areas is essential for wildfire impact assessment, burn severity analysis, carbon emission estimation, and post-fire vegetation recovery monitoring. Satellite remote sensing has become a fundamental tool for burned-area mapping owing to its extensive spatial coverage, repeated observations, and long-term data availability [1,2]. In particular, multispectral imagery from Sentinel-2 and Landsat 8/9 provides relatively fine spatial resolution, enabling detailed characterization of fire-induced land-surface changes. Advances in satellite observation systems and multisource remote sensing have expanded the capabilities of wildfire detection, monitoring, and ecological impact assessment [3–5]. The integration of complementary satellite observations has further supported global burned-area mapping [6], while cloud computing and automated algorithms have facilitated the transition from event-specific analysis to large-scale operational mapping [7,8]. These developments have increased the demand for automated segmentation methods capable of accurately delineating burned areas across heterogeneous vegetation, terrain, and background conditions without requiring reference labels from each target scene.

Fires alter vegetation structure, canopy moisture, and surface material composition, producing measurable changes in near-infrared (NIR) and shortwave-infrared (SWIR) reflectance. These changes provide the spectral basis for burned-area mapping using the Normalized Burn Ratio (NBR) and its multitemporal or relativized derivatives, including the differenced Normalized Burn Ratio (dNBR), Relativized differenced Normalized Burn Ratio (RdNBR), and Relativized Burn Ratio (RBR) [9,10]. Multispectral and hyperspectral studies have further demonstrated the utility of burn-sensitive spectral indices for burned-area delineation and burn severity assessment [11,12]. However, spectral changes associated with fire do not necessarily correspond directly to fire intensity or ecological burn severity [13]. Their magnitude is influenced by pre-fire vegetation conditions [14,15], observation timing and phenological differences [16,17], and sensor characteristics [18]. Moreover, burned surfaces may exhibit spectral responses similar to those of shadows, water bodies, exposed soil, and other land-surface disturbances. Consequently, burned and unburned pixels do not necessarily exhibit consistent spectral response ranges across scenes, complicating the selection of transferable segmentation thresholds.

Spectral-index thresholding remains widely used in burned-area mapping owing to its computational simplicity and interpretability. However, fixed thresholds often require regional calibration and may exhibit limited transferability across heterogeneous scenes [19]. Conventional automatic thresholding methods, including Otsu, Kapur, Kittler, Triangle, IsoData, and Gaussian mixture modeling (GMM), offer alternatives based on histogram characteristics, statistical optimization, and probabilistic modeling [20–22]. Their applicability to spectral-index-based burned-area detection has also been systematically investigated [20]. Nevertheless, threshold estimation may remain sensitive to background composition and spectral response distributions. To improve adaptability, burned-area-specific approaches

have integrated background masking, adaptive threshold selection, burned-core identification, and region growing [23,24], explored pixel- and object-based strategies across different biogeographical conditions [25], and employed maximum-curvature criteria and NIR–SWIR spectral contrasts for single-image detection [26]. Other automated mapping approaches have incorporated multitemporal spectral analysis, active-fire information, and object-based classification to improve burned-area identification [27,28]. These developments demonstrate the importance of adaptive spectral decisions and spatial refinement, while reflecting differences in threshold estimation and region selection strategies.

In parallel, machine learning and deep learning have further advanced automated burned-area mapping through multispectral feature integration and spectral–spatial representation learning. Conventional machine learning methods, particularly Random Forest, have been widely applied to burn severity assessment and burned-area classification using spectral bands and indices [29,30]. Deep learning approaches further exploit spatial context and multitemporal information, with transfer learning and attention-enhanced networks supporting high-resolution burned-area delineation [31,32]. Recent advances have extended these techniques to global near-real-time monitoring and foundation-model-based change detection [33,34]. Pretrained segmentation models incorporating active-fire information [35] and super-resolution-assisted cross-sensor transfer learning [36] have further expanded their applicability. Nevertheless, the deployment of learning-based methods may depend on representative training data, pretrained models, auxiliary observations, and adaptation to different observation conditions. Training-free adaptive methods with explicit spectral decision rules therefore remain a complementary approach to burned-area mapping.

Despite these advances, threshold-based burned-area segmentation still faces three interrelated challenges under heterogeneous scene conditions. First, spectral confusion between burned surfaces and non-fire backgrounds, such as shadows, water bodies, and exposed soil, can interfere with threshold estimation and increase false-positive detections [23,24]. Second, variations in pre-fire vegetation, burning conditions, and observation timing produce scene-dependent spectral response distributions, limiting the transferability of fixed thresholds [14–17]. Although relativized burn indices can mitigate the influence of initial vegetation conditions, they do not necessarily eliminate differences in response distributions across scenes [9,10]. Third, spatial heterogeneity in fire-induced changes produces both strongly affected surfaces and weak-response burned regions, making it difficult to preserve burned-area boundaries while suppressing background-related false positives. Validation studies of burned-area products have further identified fire size, land-cover characteristics, and spatial resolution as important factors influencing mapping errors [37,38]. Together, these challenges highlight the need to consider background composition, scene-specific response distributions, and spatial support relationships in adaptive burned-area segmentation.

Although background masking, adaptive thresholding, and core-guided region expansion have been incorporated into existing burned-area mapping methods [23–25], their implementations differ in response-domain definition, threshold determination, and spatial region selection. Accordingly, a key methodological consideration is how to integrate background-constrained response analysis, scene-adaptive threshold estimation, and spatial support into a coherent segmentation procedure applicable across heterogeneous fire events and optical satellite sensors.

To address these challenges, this study proposes a background-constrained, normalized-response, and core-supported (BNCS) adaptive segmentation method for burned-area delineation. The method integrates background constraints (B), response normalization (N), and core support (CS) into a scene-driven decision framework. Specifically, pre- and

post-fire spectral constraints define the candidate response domain, while scene-wise quantile normalization establishes a relative scale for spectral change responses. An adaptive low threshold is then determined from the area-response curve using a geometric elbow criterion. Subsequently, a high-response threshold derived from the root mean square (RMS) is used to identify burned cores, and spatial connectivity is employed to retain low-threshold candidate regions supported by these cores. This procedure enables scene-specific threshold estimation and burned-area delineation without reference-based threshold calibration for the target fire event.

The main contributions of this study are threefold: (1) a scene-response-driven adaptive thresholding framework that integrates background-constrained candidate selection, scene-relative response normalization, and area-response analysis to determine segmentation thresholds without target-scene reference labels; (2) a core-supported region selection strategy that combines high-response spectral evidence with spatial connectivity to preserve plausible weak-response burned regions while suppressing unsupported candidate patches; and (3) a systematic evaluation across 12 fire events and two optical satellite sensor datasets to assess segmentation accuracy, threshold transferability, component contributions, parameter sensitivity, and agreement with existing burned-area mapping products. These analyses aim to evaluate the effectiveness, adaptability, and limitations of BNCS under heterogeneous scene and sensor conditions.

2. Materials

2.1. Study Areas and Fire Events

Twelve representative forest and grassland fire events that occurred between 2018 and 2025 were selected from northeastern (NE), central (CT), and southwestern (SW) China, covering Liaoning, Inner Mongolia, Shanxi, Sichuan, and Yunnan. These three regions differ considerably in climatic conditions, topography, vegetation composition, and land-cover characteristics, providing diverse environmental settings for evaluating burned-area segmentation. Northeastern China contains extensive forest and grassland ecosystems, including landscapes characterized by forest–grassland mosaics. Seasonal drought, strong winds, and fuel accumulation contribute to elevated fire risk in certain areas, while lightning is an important natural ignition source in some high-latitude forests. Central China is characterized by mountainous and hilly terrain with heterogeneous landscapes comprising forests, shrublands, croplands, and exposed surfaces. Fire occurrence and spread in this region are influenced by both human activities and environmental conditions. Southwestern China features complex mountainous terrain, extensive forest cover, diverse vegetation types, and pronounced seasonal variations in moisture availability. Prolonged dry conditions, terrain-influenced winds, and topographic complexity can contribute to spatially heterogeneous fire patterns and irregular burned-area boundaries.

Overall, the selected regions encompass substantial variation in vegetation backgrounds, terrain conditions, fire drivers, and fire perimeter characteristics, providing a diverse set of cases for assessing the adaptability and stability of burned-area segmentation under heterogeneous conditions. The geographical distribution of the selected fire events is shown in Figure 1, and the corresponding event information and satellite observation dates are summarized in Table 1.

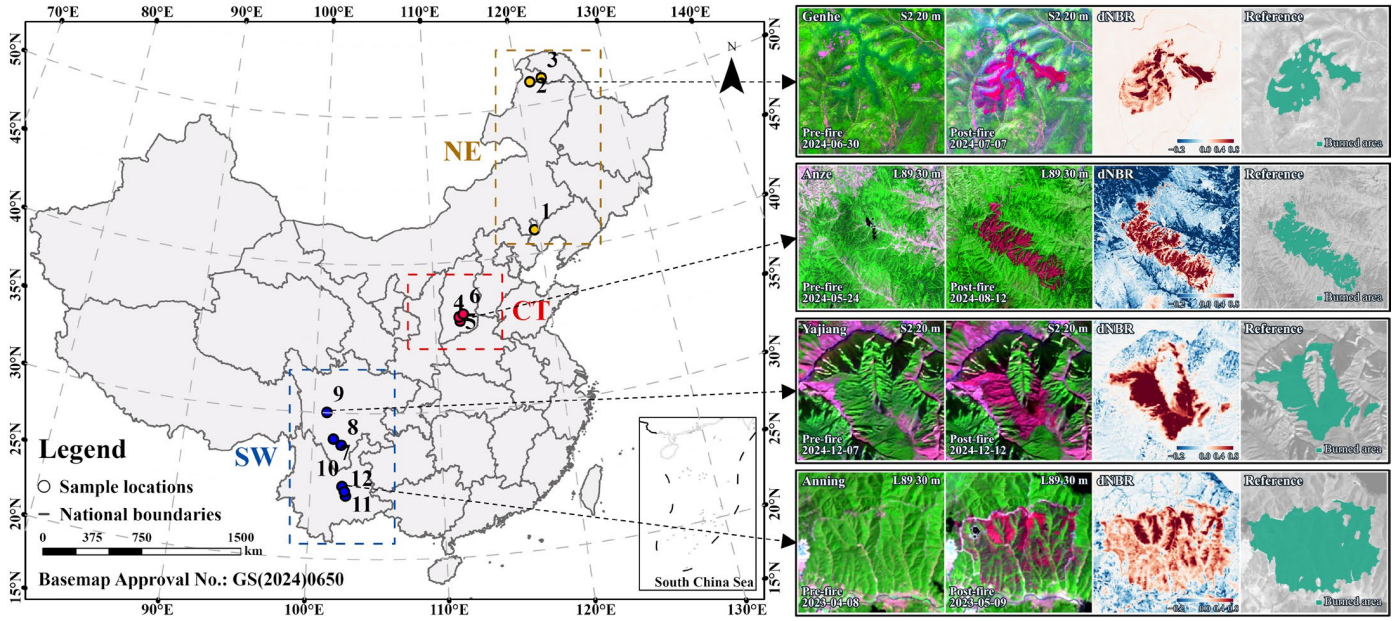

Figure 1. Spatial distribution of the study areas and representative fire events. Numbers 1–12 correspond to the fire events listed in Table 1. NE, CT, and SW represent northeastern, central, and southwestern China, respectively, and are indicated in orange, red, and blue, respectively. The pre-fire and post-fire images are displayed as false-color composites using the SWIR2, NIR, and red bands assigned to the red, green, and blue channels, respectively (Sentinel-2: B12/B8/B4; Landsat 8/9: B7/B5/B4). In the dNBR images, red and blue indicate positive and negative values, respectively. The turquoise areas in the reference images represent the reference burned areas.

Table 1. Fire events and corresponding satellite imagery information.

ID	Region	Fire Event	Fire Period	Sentinel-2		Landsat 8/9	
				Pre-Fire	Post-Fire	Pre-Fire	Post-Fire
1	NE	Jianping, Liaoning	2023/06/30–07/01	2023/6/25	2023/7/3	2023/6/27	2023/7/5
2	NE	Genhe, Inner Mongolia	2024/07/02–07/06	2024/6/30	2024/7/7	2024/6/20	2024/8/14
3	NE	Hanma, Inner Mongolia	2018/06/01–06/06	2018/5/30	2018/9/17	2018/5/3	2018/9/24
4	CT	Anze, Shanxi	2024/06/12–06/16	2024/03/14–06/11 *	2024/06/17–09/14 *	2024/5/24	2024/8/12
5	CT	Guxian, Shanxi	2024/06/26–07/01	2024/6/23	2024/7/13	2024/6/25	2024/7/11
6	CT	Qinyuan, Shanxi	2025/06/08–06/14	2025/5/24	2025/8/7	2025/6/4	2025/8/7
7	SW	Muli, Sichuan	2023/05/27–06/03	2023/5/9	2023/7/23	2023/5/9	2023/9/5
8	SW	Xichang, Sichuan	2020/03/30–04/02	2020/3/25	2020/4/4	2020/3/13	2020/6/1
9	SW	Yajiang, Sichuan	2024/12/09–12/11	2024/12/7	2024/12/12	2024/12/4	2024/12/12
10	SW	Anning, Yunnan	2023/04/13–04/18	2023/4/11	2023/4/19	2023/4/8	2023/5/9
11	SW	Jiangchuan, Yunnan	2023/04/11–04/15	2023/4/6	2023/4/21	2023/3/31	2023/04/16–07/14 *
12	SW	Jinning, Yunnan	2024/04/12–04/15	2024/4/10	2024/04/16–07/14 *	2024/4/10	2024/4/18

Note: An asterisk (*) indicates a median composite, with the listed dates representing the compositing period; otherwise, dates refer to individual satellite acquisitions.

2.2. Satellite Data and Preprocessing

Sentinel-2 Multispectral Instrument (MSI) imagery (hereafter referred to as S2) and Landsat 8 Operational Land Imager (OLI) and Landsat 9 Operational Land Imager 2 (OLI-2) imagery (collectively referred to as L89). S2 imagery was retrieved from the harmonized surface reflectance collection (COPERNICUS/S2_SR_HARMONIZED), supplemented with cloud probability data (COPERNICUS/S2_CLOUD_PROBABILITY). Images with scene-level cloud cover exceeding 60% were excluded. Invalid, shadow-contaminated, cloudy,

and snow-covered pixels were masked using the Scene Classification Layer (SCL; classes 0, 1, 3, 8, 9, 10, and 11), together with a pixel-level cloud probability threshold of 40%. L89 imagery was obtained from the Landsat Collection 2 Tier 1 Level 2 products, with scene-level cloud cover similarly restricted to 60%. Cloud-, shadow-, and snow-contaminated pixels were masked using QA_PIXEL, while radiometrically saturated pixels were excluded using QA_RADSAT. The blue, green, red, NIR, and SWIR2 bands were retained for both sensors, with surface reflectance derived using the corresponding product scaling factors. S2 and L89 imagery were processed at spatial resolutions of 20 and 30 m, respectively.

To prioritize observations representative of surface conditions before and after each fire, a single-scene-first strategy with composite fallback was adopted. For each fire event, image quality was assessed within its predefined area of interest (AOI) and reference burned area. Pre-fire imagery was retrieved backward from the fire start date, whereas post-fire imagery was retrieved forward from the day following the fire end date. Individual acquisitions were screened sequentially within progressively expanded 30-, 60-, 90-, and 120-day windows, requiring more than 95% valid-pixel coverage within both the AOI and reference burned area. When multiple acquisitions satisfied these criteria, the observation closest to the fire event was selected. If no qualifying single acquisition was identified, median compositing was attempted sequentially over the same 30-, 60-, 90-, and 120-day windows. A composite was accepted only if it achieved more than 95% valid-pixel coverage within both the AOI and reference burned area, incorporated at least two acquisitions, and had a median valid observation count of at least two. Following image selection, the intersection of valid pre- and post-fire pixels was determined. Only image pairs with more than 95% common valid-pixel coverage within both the AOI and reference burned area were retained, and fire-induced spectral changes were calculated exclusively within this common valid domain.

2.3. Reference Burned-Area Delineation

Field investigations were conducted at representative fire sites, including Hanma in September 2021, Anning in June 2023, Jianping in July 2024, and Jiangchuan and Jinning in July 2024. Unmanned aerial vehicle (UAV) imagery was also acquired over selected areas. These observations provided supporting information on burned characteristics, spatial patterns, and boundary features, informing the criteria for visual interpretation.

Reference burned-area boundaries for all 12 fire events were manually delineated using pre- and post-fire S2 imagery as the primary interpretation data, supplemented by high-resolution satellite imagery from GF-2 (0.8 m), GF-7 (0.65 m), GF-1 (2 m), and GF-6 (2 m). The high-resolution images were acquired close to the corresponding experimental observation dates and spatially aligned with the Sentinel-2 imagery to support the identification of burned surfaces and refinement of ambiguous boundaries. Interpretation considered fire-induced spectral changes, spatial continuity, and differences between burned surfaces and surrounding backgrounds. The resulting reference polygons were delineated using the S2 spatial reference system and rasterized onto the corresponding S2 (20 m) and L89 (30 m) analysis grids for pixel-level accuracy assessment.

3. Methods

3.1. Method Development

The proposed BNCS method first calculates NBR from pre- and post-fire multispectral imagery and derives dNBR to characterize fire-induced spectral changes. A background constrained candidate domain is then constructed based on pre-fire vegetation conditions, pre-fire spectral characteristics associated with burnable surfaces, and post-fire water responses to reduce interference from unburned background pixels. Within this domain,

robust percentiles of the dNBR distribution are used to normalize spectral changes across scenes, transforming absolute response magnitudes into a common relative scale. An adaptive low threshold is subsequently determined from the area-decline characteristics of the normalized responses under progressive thresholding, while high-response burned cores are used to screen low-threshold connected regions and obtain the final burned-area. The BNCS methodological framework is illustrated in Figure 2.

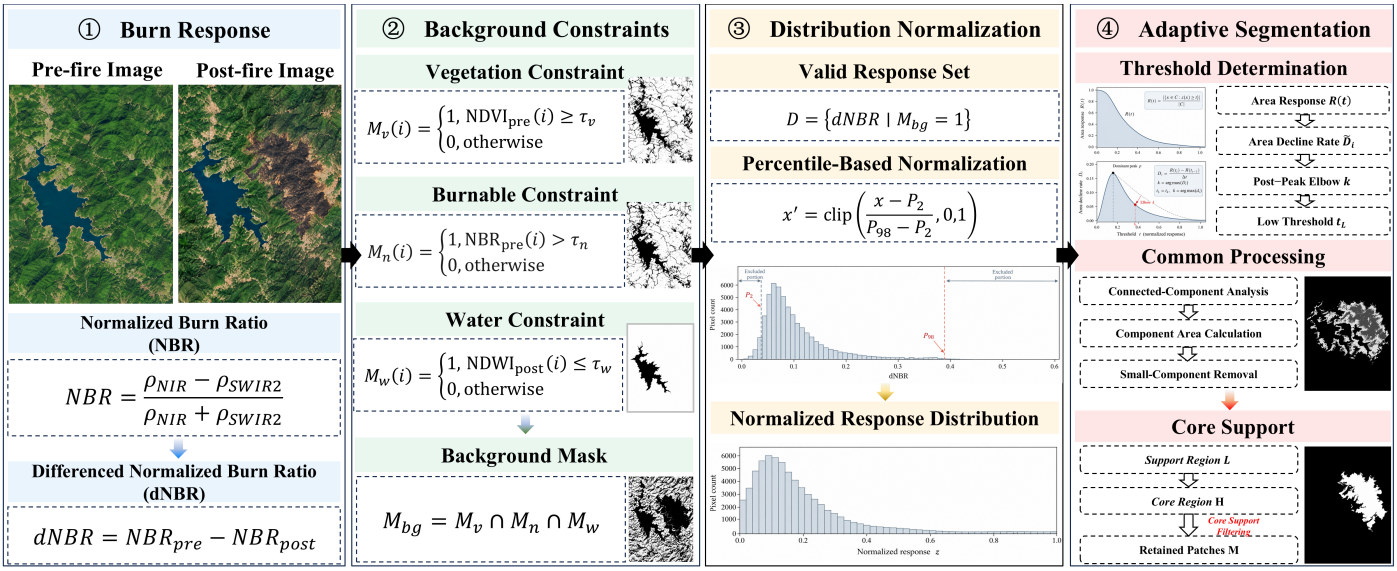

Figure 2. BNCS methodological framework.

NBR exploits the contrasting spectral responses of vegetation and moisture content in the NIR and SWIR2 bands to distinguish burned surfaces from unburned backgrounds. Fire-induced spectral changes were further quantified using dNBR, calculated from the difference between pre- and post-fire NBR values. In addition, RdNBR and RBR were considered to assess the applicability of BNCS to alternative burn indices. The four indices were calculated using Equations (1)–(4):

$$\mathrm{NBR} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{SWIR2}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{SWIR2}}} \quad (1)$$

$$\mathrm{dNBR} = \mathrm{NBR}_{\mathrm{pre}} - \mathrm{NBR}_{\mathrm{post}} \quad (2)$$

$$\mathrm{RdNBR} = \frac{\mathrm{dNBR}}{\sqrt{|\mathrm{NBR}_{\mathrm{pre}}|}} \quad (3)$$

$$\mathrm{RBR} = \frac{\mathrm{dNBR}}{\mathrm{NBR}_{\mathrm{pre}} + 1.001} \quad (4)$$

where the ρ_{NIR} and ρ_{SWIR2} denote the NIR and SWIR2 reflectances, respectively, and $\mathrm{NBR}_{\mathrm{pre}}$ and $\mathrm{NBR}_{\mathrm{post}}$ represent the pre- and post-fire NBR values, respectively. dNBR was adopted as the primary spectral response index, while RdNBR and RBR were used to examine the applicability of BNCS across different burn-index formulations. To reduce interference from unburned pixels exhibiting similar spectral changes, BNCS applies spectrally informed background constraints to define the candidate domain for adaptive threshold estimation.

The candidate domain is defined using three complementary spectral constraints based on pre-fire vegetation conditions, pre-fire spectral characteristics of potentially burnable surfaces, and post-fire water responses. Specifically, $\mathrm{NDVI}_{\mathrm{pre}}$ restricts the candidate domain to pixels exhibiting a sufficient pre-fire vegetation signal, $\mathrm{NBR}_{\mathrm{pre}}$ provides an additional

constraint on the initial vegetation background, and $NDWI_{post}$ excludes pixels with water-like spectral responses. The indices are calculated using Equations (5) and (6):

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \tag{5}$$

$$NDWI = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}} \tag{6}$$

where the ρ_{Red} and ρ_{Green} denote the reflectances in the red and green bands, respectively. Based on these indices, the background constrained candidate domain C is defined by Equation (7) as the set of valid pixels i that simultaneously satisfy all conditions:

$$C = \left\{ i \mid NDVI_{pre}(i) \geq \tau_v,\ NBR_{pre}(i) > \tau_n,\ NDWI_{post}(i) \leq \tau_w \right\} \tag{7}$$

where τ_v, τ_n, and τ_w denote the thresholds for pre-fire vegetation cover, pre-fire burnable background, and post-fire water conditions, respectively, and are set to 0.25, 0, and 0. These constraints define the candidate domain for subsequent response characterization and threshold estimation rather than directly classifying pixels as burned or unburned.

Following candidate-domain construction, within-scene distribution normalization is applied to account for variations in dNBR response ranges across fire events. The 2nd and 98th percentiles of dNBR within the candidate domain C, denoted as P_2 and P_{98}, respectively, are used as robust lower and upper bounds for response scaling. The dNBR values are then normalized using Equation (8):

$$z(i) = \mathrm{clip}\left[\frac{\hat{d}NBR(i) - P_2}{P_{98} - P_2}, 0, 1 \right] \tag{8}$$

where $z(i)$ denotes the normalized burn-related change response of pixel i, and $\mathrm{clip}(\cdot)$ constrains the normalized values to the interval $[0, 1]$. Both percentile bounds are estimated independently for each scene from its candidate-domain dNBR distribution.

Within the normalized response domain, thresholds ranging from 0 to 1 are applied at intervals of 0.01 to quantify the proportion of candidate pixels retained at each threshold. The resulting area-response function is defined by Equation (9):

$$R(t) = \frac{|\{i \in C : z(i) \geq t\}|}{|C|} \tag{9}$$

As the threshold increases, the proportion of retained candidate pixels decreases. The corresponding area-decline rate between consecutive thresholds is calculated using Equation (10):

$$D_i = \frac{R(t_i) - R(t_{i+1})}{t_{i+1} - t_i} \tag{10}$$

where $R(t_i)$ denotes the proportion of candidate pixels with normalized responses at or above threshold t_i, and D_i denotes the area-decline rate between consecutive thresholds t_i and t_{i+1}.

To suppress local fluctuations introduced by discrete threshold sampling, the area-decline rate D_i is smoothed using a centered three-point moving average, yielding $\tilde{D}_i$. The dominant peak p is then identified within the valid interior search interval of the smoothed curve, indicating where the candidate area is most sensitive to threshold changes. The post-peak segment is subsequently extracted, with its horizontal axis normalized to $[0, 1]$ and its decline-rate values scaled by the peak magnitude. A chord is constructed between the endpoints of this segment, and the perpendicular distance from each interior point to the chord is calculated. The maximum-distance point k is identified as the post-peak elbow, and its corresponding value is selected as the adaptive low threshold t_L. This criterion

follows the geometric principle of maximum-distance-based threshold selection [39,40] and is related to established knee-point detection approaches [41]. Unlike curvature-based approaches that estimate local curvature, it identifies the elbow from the overall geometry of the smoothed post-peak segment without additional numerical differentiation.

Although the adaptive low threshold t_L facilitates the retention of weak-response burned regions and their boundaries, it may also introduce unburned patches with similar spectral responses. BNCS therefore incorporates a core-support mechanism to assess the spatial reliability of low-threshold candidate regions. Candidate regions are first extracted within the background-constrained domain using t_L. Eight-connected components smaller than 0.5 ha are removed to obtain the low-threshold support region L. The core threshold t_H is then calculated using Equation (11) as the root mean square of the normalized responses of candidate pixels at or above the low threshold:

$$t_H = \sqrt{\frac{1}{N_T}\sum_{i\in T} z(i)^2} \tag{11}$$

where T denotes the set of candidate pixels with normalized responses at or above t_L, and N_T is the number of pixels in this set. High-response candidates are then extracted from L using t_H, followed by eight-connected component filtering to remove components smaller than 0.5 ha, yielding the valid high-response cores H. The final burned-area mask is obtained by retaining only those low-threshold connected components L_j that spatially intersect at least one valid core, as expressed in Equation (12):

$$M = \bigcup \{ L_j \mid L_j \cap H \neq \emptyset \} \tag{12}$$

where M denotes the final burned-area segmentation mask. Rather than applying t_H as a second threshold to truncate the low-threshold mask, the core-support mechanism uses high-response cores as spatial evidence of patch reliability. Components supported by valid cores are retained, while unsupported components are discarded.

3.2. Experimental Design

To systematically evaluate BNCS under heterogeneous scene and sensor conditions, five groups of experiments were conducted to examine spectral response variability, threshold dependence and transferability, segmentation performance, background-extent sensitivity, and agreement with existing burned-area mapping products. The experiments involved 12 fire events distributed across northeastern, central, and southwestern China, using S2 and L89 imagery. Pre- and post-fire observations were selected and processed following the procedures described in Section 2.2. The two datasets represented the same fire events but differed in spatial resolution and sensor-specific observation characteristics. A consistent accuracy assessment protocol was adopted. Reference labels were used for accuracy assessment, explicitly defined reference-based threshold optimization, and the construction of controlled experimental regions, but not for BNCS threshold estimation.

(1) Spectral response variability across fire events and regions. To characterize the heterogeneity of fire-induced spectral changes, pre- and post-fire NIR and SWIR2 reflectances, together with NBR and dNBR, were analyzed within the reference burned areas. The analysis considered three levels of variability: spectral variations within individual fire events, differences among events within the same region, and differences across the three geographic regions. Event-level medians and interquartile ranges were calculated separately to characterize response distributions. Regional comparisons used arithmetic means of event-level medians and interquartile ranges rather than treating individual pixels as independent observations.

(2) Scene dependence and transferability of segmentation thresholds. Fixed-threshold transferability was evaluated using a common dNBR threshold search space under four settings: leave-one-event-out, leave-one-sensor-out, leave-one-region-out, and cross-sensor transfer to unseen events. In each setting, a fixed threshold was optimized exclusively using the reference labels of the training scenes and subsequently applied to the held-out test scenes without recalibration. A scene-specific Oracle threshold was also determined retrospectively by selecting the threshold that maximized F1 for each test scene within the same search space and specified single-threshold segmentation procedure. The Oracle therefore represented the optimum of this procedure rather than an upper bound for all segmentation methods. Comparisons among BNCS, transferred fixed thresholds, and scene-specific Oracle thresholds assessed scene dependence and transferability.

(3) Comparative evaluation of segmentation performance. BNCS was compared with a fixed-threshold baseline and six conventional automatic thresholding methods: Otsu, Kapur, Kittler, Triangle, IsoData, and GMM. Component ablation experiments were conducted to assess the contributions of B, N, and CS, while conventional methods were further evaluated under the same component configurations to examine their effects across different thresholding approaches. Parameter sensitivity was assessed by varying background constraint thresholds, percentile bounds, minimum patch areas, and core statistics. The applicability of BNCS to different burn indices was examined using dNBR, RdNBR, and RBR. Statistical significance was assessed using paired two-sided exact sign-flip tests across 12 fire events per sensor (4096 sign assignments), with Holm correction for seven method comparisons. Computational efficiency was evaluated through runtime comparisons under a consistent single-threaded execution environment.

(4) Sensitivity to background extent. To evaluate the sensitivity of segmentation performance to changes in analysis extent, analysis regions were constructed for all 12 fire events based on the minimum bounding rectangle of each reference burned area. These regions were expanded to 3, 6, and 9 times the area of the original rectangle while preserving its center and aspect ratio. For each extent, response statistics and segmentation thresholds were recalculated, and segmentation accuracy was evaluated within the corresponding valid region. The three background extents for each event–sensor combination were evaluated using the same pre- and post-fire imagery. Where necessary, image selection differed from the main comparison to ensure adequate coverage of the expanded analysis domain.

(5) Comparison with existing burned-area mapping products. To assess the agreement between BNCS delineations and publicly available mapping products, the results were compared with the Global Annual Burned Area Map (GABAM) [42] and Global Forest Loss Due to Fire (GLAD GFL) [43] datasets. Product layers corresponding to the fire-event years were aligned with the reference data and evaluated over common valid spatial domains on the S2 and L89 grids. Comparisons accounted for differences in product availability, temporal definitions, and mapping objectives. To further characterize omission patterns, the dNBR distributions of reference burned pixels detected and omitted by each product were examined. Omission rates were also calculated for the lowest and highest event-specific dNBR quartiles. These comparisons assessed event-level spatial agreement rather than providing a direct ranking of products with different mapping objectives.

3.3. Evaluation Metrics

Four evaluation metrics were used to assess burned-area segmentation performance: precision (P), recall (R), F1 score (F1), and intersection over union (IoU). P quantifies the proportion of correctly identified burned pixels among all pixels classified as burned, whereas R measures the proportion of reference burned pixels successfully detected. F1 combines precision and recall into a single measure, while IoU quantifies the spatial overlap between the predicted area and the reference. These metrics are calculated using Equations (13)–(16):

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{14}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{15}$$

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{16}$$

where TP, FP, and FN denote the numbers of true-positive, false-positive, and false-negative pixels, respectively. All four metrics range from 0 to 1, with higher values indicating better segmentation performance.

For multi-event evaluation, each metric was first calculated independently for each fire event and sensor. Scene-level macro-averaging (e.g., macro F1) was then applied using Equation (17) to obtain the overall performance for each sensor:

$$\mathrm{Macro}\ S_j = \frac{1}{N}\sum_{j=1}^{N} S_j \tag{17}$$

where the macro-averaged metric is calculated as the arithmetic mean of the corresponding metric across all evaluated fire events. This averaging strategy assigns equal weight to each event, preventing larger burned areas or scenes containing more valid pixels from disproportionately influencing the overall evaluation.

4. Results

4.1. Spectral Response Variability Across Fire Events and Regions

The spectral characteristics of burned areas varied substantially across the 12 fire events (Figure 3). Within the reference burned areas, the median NIR reflectance decreased from pre- to post-fire observations in all events for both S2 and L89. In contrast, the median SWIR2 reflectance increased in all S2 events and 11 of the 12 L89 events, with Muli being the exception. The median pre-fire NBR varied from 0.277 to 0.678 for S2 and from 0.350 to 0.666 for L89, indicating appreciable differences in the initial spectral conditions of the evaluated burned areas. Post-fire NBR also varied substantially, with event-level medians ranging from −0.303 to 0.298 for S2 and −0.154 to 0.414 for L89.

Considerable spectral heterogeneity was also observed within individual fire events and among events located in the same region. The interquartile range (IQR) of dNBR within individual burned areas ranged from 0.222 to 0.606 for S2 and from 0.232 to 0.564 for L89, indicating that the fire-induced spectral changes were not uniform within event boundaries. For example, the Yajiang event exhibited a broad dNBR distribution, with an IQR of 0.606 for S2 and 0.564 for L89. Differences among events within the same region were similarly pronounced. In southwestern China, the event-level median dNBR ranged from 0.240 at Muli to 0.657 at Yajiang for S2, and from 0.046 to 0.653 for L89. In northeastern China, the corresponding S2 medians ranged from 0.351 at Genhe to 0.756 at Hanma. These variations indicate that geographic proximity or regional classification alone does not imply similar spectral response distributions among fire events.

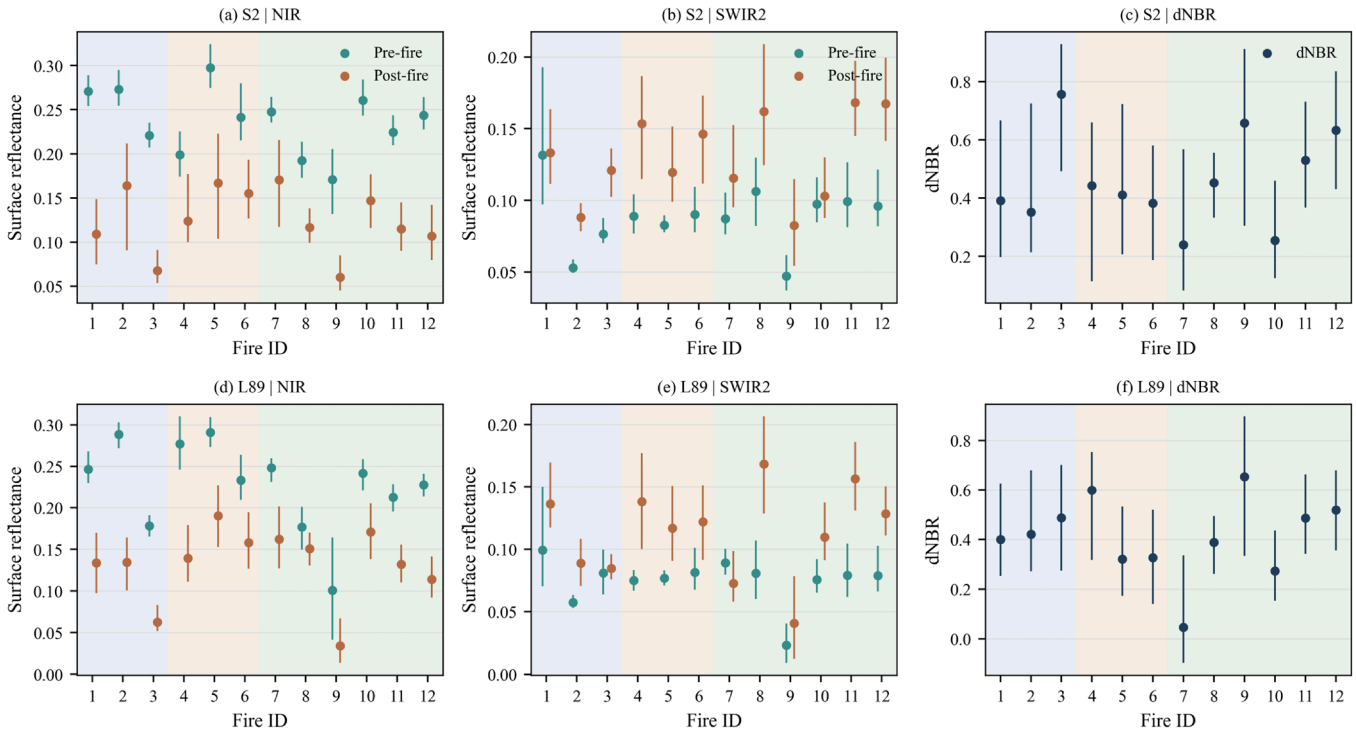

Figure 3. Spectral response variability across 12 fire events and three geographic regions. (**a**–**c**) NIR reflectance, SWIR2 reflectance, and dNBR derived from S2 MSI imagery; (**d**–**f**) corresponding results from L89 OLI imagery. Teal and orange markers indicate pre- and post-fire reflectance, respectively, while dark blue markers represent dNBR. Points denote event-level medians of reference burned pixels, and vertical lines indicate the 25th–75th percentile ranges. Light blue, beige, and pale green shading distinguish northeastern (NE; IDs 1–3), central (CT; IDs 4–6), and southwestern (SW; IDs 7–12) China, respectively. Fire event details are provided in Table 1.

At the regional scale, event-level dNBR statistics further revealed differences in both response magnitude and variability (Table 2). For S2, the mean of event-level median dNBR values was 0.499 in northeastern China, 0.412 in central China, and 0.461 in southwestern China. The corresponding values for L89 were 0.436, 0.415, and 0.394, respectively. However, the ranges of event-level medians overlapped across regions, with considerable between-event variability within regions, particularly in northeastern China for S2 and southwestern China for L89. Taken together, these results reveal substantial spectral response variability within and among fire events, as well as across regions.

Table 2. Regional spectral characteristics of burned areas from S2 and L89 imagery.

Sensor	Region	NIR		SWIR2		NBR		dNBR		
		Pre	Post	Pre	Post	Pre	Post	Median	Range	IQR
S2	NE	0.255	0.114	0.087	0.114	0.501	−0.056	0.499	0.351–0.756	0.472
S2	CT	0.246	0.149	0.087	0.140	0.457	0.007	0.412	0.382–0.442	0.484
S2	SW	0.223	0.119	0.089	0.133	0.436	−0.084	0.461	0.240–0.657	0.402

Table 2. *Cont.*

Sensor	Region	NIR		SWIR2		NBR		dNBR		
		Pre	Post	Pre	Post	Pre	Post	Median	Range	IQR
L89	NE	0.238	0.110	0.079	0.103	0.483	−0.008	0.436	0.400–0.487	0.402
L89	CT	0.267	0.163	0.078	0.126	0.549	0.101	0.415	0.321–0.598	0.392
L89	SW	0.201	0.127	0.071	0.113	0.485	0.048	0.394	0.046–0.653	0.359

Note: Pre and post denote pre- and post-fire observations. The spectral values are arithmetic means of event-level medians. dNBR range denotes the range of event-level medians, and dNBR IQR is the mean within-event interquartile range. All events are equally weighted.

4.2. Scene Dependence and Transferability of Segmentation Thresholds

To establish a common fixed-threshold baseline, the dNBR threshold was retrospectively optimized across all fire events and both sensors by maximizing macro F1. The optimal threshold was 0.12, which was adopted for the Fixed method. As shown in Figure 4, scene-specific optimal thresholds varied considerably, ranging from 0.02 to 0.24 for S2 and from −0.11 to 0.19 for L89. At the common threshold of 0.12, Fixed achieved macro F1 scores of 88.64% and 86.71% for S2 and L89, respectively, compared with 90.81% and 89.93% obtained using scene-specific Oracle thresholds. These results indicate that although a common fixed threshold can achieve relatively high overall accuracy, the optimal threshold varies substantially across fire events and sensors.

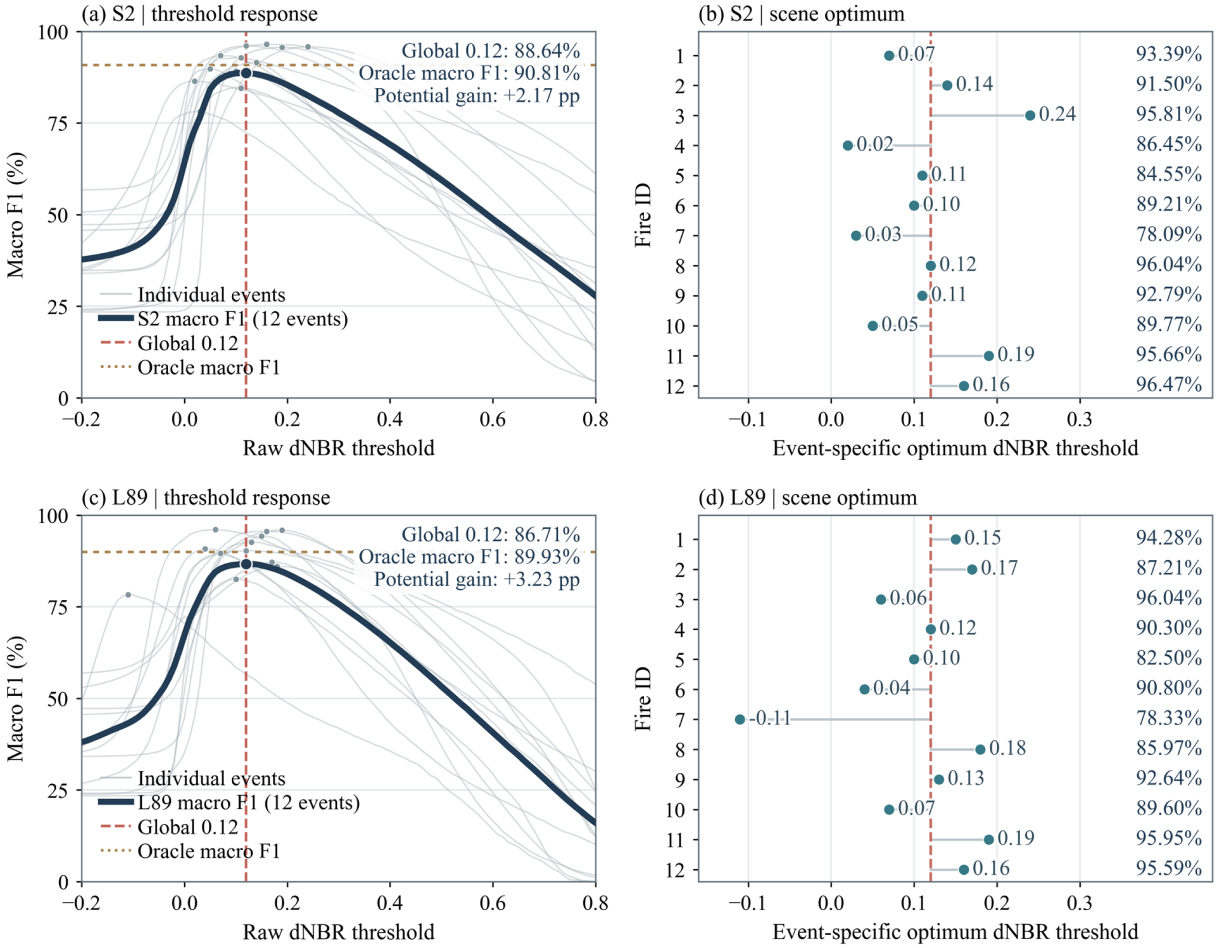

Figure 4. Fixed-threshold response curves and distribution of scene-specific optimal thresholds. (**a**) Event-specific dNBR threshold-response curves and macro F1 for S2; (**b**) scene-specific optimal dNBR thresholds and corresponding maximum F1 scores for S2; (**c**) event-specific dNBR threshold-

response curves and macro F1 for L89; (**d**) scene-specific optimal dNBR thresholds and corresponding maximum F1 scores for L89. Dashed lines indicate the common fixed threshold of 0.12, while dash-dotted lines indicate the mean scene-specific Oracle F1. The circles in (**a**,**c**) indicate the maximum F1 scores and their corresponding dNBR thresholds for individual fire events, whereas those in (**b**,**d**) indicate the event-specific optimal dNBR thresholds.

The threshold-transfer experiments further revealed the limitations of fixed-threshold selection under unseen conditions (Table 3). Across the four transfer settings, training-derived optimal thresholds ranged from 0.10 to 0.13. The median thresholds were 0.12 for leave-one-event-out, leave-one-region-out, and cross-sensor transfer to unseen events, and 0.115 for leave-one-sensor-out. Despite this relatively narrow range, the transferred thresholds yielded mean F1 scores of 87.29–87.64% on held-out scenes, consistently below the 90.70% achieved by BNCS. The performance gap ranged from 3.06 percentage points (pp) under leave-one-sensor-out transfer to 3.40 pp under leave-one-event-out transfer. The transferred thresholds also underperformed scene-specific Oracle thresholds by 2.73–3.08 pp. These results show that BNCS maintained higher average segmentation accuracy than transferred fixed thresholds across the four evaluation settings.

Table 3. Cross-scene transfer performance of training-derived fixed thresholds.

Transfer Setting	Threshold Range	Fixed Macro F1 (%)	BNCS Macro F1 (%)	ΔF1 (pp)	Oracle Gap (pp)
Event	0.11–0.13	87.29	90.70	3.40	3.08
Sensor	0.11–0.12	87.64	90.70	3.06	2.73
Region	0.10–0.13	87.31	90.70	3.38	3.06
Cross	0.10–0.13	87.54	90.70	3.16	2.83

Note: Event, Sensor, Region, and Cross represent leave-one-event-out, leave-one-sensor-out, leave-one-region-out, and cross-sensor transfer to unseen events, respectively. ΔF1 = BNCS − Fixed. Oracle gap represents the difference between scene-specific Oracle and transferred Fixed F1 scores. All F1 differences are expressed in percentage points (pp).

Overall, the retrospectively optimized common threshold achieved relatively high accuracy across the full dataset but showed reduced performance under training–test transfer conditions. In contrast, BNCS consistently achieved higher mean F1 scores than training-derived fixed thresholds across the evaluated settings, indicating greater adaptability to scene-dependent spectral response variations without requiring reference-label-based threshold calibration for the test scenes.

4.3. Comparison of Segmentation Performance Across Thresholding Methods

The segmentation performance of BNCS was compared with that of Fixed, Otsu, Kapur, Kittler, Triangle, IsoData, and GMM across 12 fire events under a consistent experimental protocol. As shown in Table 4 and Figures 5 and 6, BNCS achieved the highest macro F1 and macro IoU among the evaluated methods excluding the retrospective Oracle reference for both sensors. For S2, its macro P, R, F1, and IoU were 91.98%, 92.14%, 91.83%, and 85.20%, respectively, compared with 88.54%, 91.92%, 89.56%, and 81.67% for L89. Among the conventional automatic thresholding methods, Triangle performed best, with macro F1 scores of 86.28% for S2 and 82.87% for L89, respectively, 5.55 and 6.70 pp below BNCS. Fixed achieved macro F1 scores of 88.64% and 86.71%, remaining 3.19 and 2.85 pp below BNCS, respectively. Scene-level accuracy metrics for all 12 fire events and segmentation maps for the remaining 9 events are provided in Table S1 and Figures S3–S11 of the Supplementary Materials, respectively.

Table 4. Overall segmentation accuracy of different methods across fire events for S2 and L89.

Methods	S2				L89			
	Macro P (%)	**Macro R (%)**	**Macro F1 (%)**	**Macro IoU (%)**	**Macro P (%)**	**Macro R (%)**	**Macro F1 (%)**	**Macro IoU (%)**
Otsu	96.42	66.64	77.84	65.33	95.88	71.83	81.76	69.72
Kapur	90.07	69.11	74.28	61.67	87.63	63.63	71.60	61.09
Kittler	91.29	82.46	85.23	76.17	82.15	77.14	77.20	68.40
Triangle	82.58	93.40	86.28	77.42	78.37	93.35	82.87	73.28
IsoData	96.38	67.06	78.11	65.70	89.89	76.02	79.89	68.11
GMM	85.17	85.31	83.27	72.74	82.77	86.82	82.15	71.07
Fixed	*90.21*	*88.00*	*88.64*	*80.15*	*89.72*	*86.45*	*86.71*	*77.73*
Oracle	*90.85*	*90.97*	*90.81*	*83.59*	*89.51*	*90.53*	*89.93*	*82.13*
BNCS	**91.98**	**92.14**	**91.83**	**85.20**	**88.54**	**91.92**	**89.56**	**81.67**

Note: All metrics are averaged equally across 12 fire events per sensor. Fixed uses a common dNBR threshold of 0.12, retrospectively optimized across all evaluation scenes. Fixed and Oracle are reference methods and are shown in italics. Oracle represents the scene-specific optimum within the predefined single-threshold procedure and is not a theoretical upper bound for BNCS. BNCS is the proposed method, with its results highlighted in bold.

The differences in segmentation performance were accompanied by distinct precision–recall trade-offs. Otsu, Kapur, and IsoData generally exhibited higher macro P (87.63–96.42%) than macro R (63.63–76.02%) across both sensors, indicating a tendency toward burned-pixel omission. In contrast, Triangle achieved macro R values of 93.40% for S2 and 93.35% for L89, accompanied by lower macro P values of 82.58% and 78.37%, respectively. Kittler and GMM showed variable performance, with macro F1 scores ranging from 77.20% to 85.23%. Compared with Triangle, BNCS achieved higher macro P by 9.40 pp for S2 and 10.17 pp for L89, while its macro R was lower by 1.26 and 1.44 pp, respectively. Thus, the advantage of BNCS over the strongest conventional automatic thresholding method was primarily reflected in higher macro P, with a relatively small reduction in macro R. Paired two-sided exact sign-flip tests across 12 fire events confirmed higher F1 scores for BNCS than Triangle on both sensors after Holm correction for seven comparisons per sensor (Holm $p = 0.0088$ for S2 and 0.0034 for L89; Table S3).

The performance differences between sensors were not uniform across methods. BNCS achieved a 2.27-percentage-point higher macro F1 for S2 than for L89, whereas the corresponding difference for Triangle was 3.41 pp. Triangle also exhibited lower macro P on L89 than on S2, with a difference of 4.22 pp, compared with 3.44 pp for BNCS. Conversely, Otsu and IsoData achieved higher macro F1 scores for L89 than for S2, accompanied by increases in macro R. These contrasting results reveal method-dependent differences in the balance between omission and commission errors across the two sensor datasets.

The scene-specific Oracle achieved macro F1 scores of 90.81% for S2 and 89.93% for L89. BNCS exceeded the Oracle by 1.02 pp for S2 but remained 0.37 pp below it for L89. Notably, the Oracle represents retrospective optimization within a predefined single-threshold segmentation framework, whereas BNCS incorporates adaptive threshold determination and core-supported region selection. Accordingly, the Oracle provides a reference for scene-specific threshold optimization rather than an upper performance bound for BNCS.

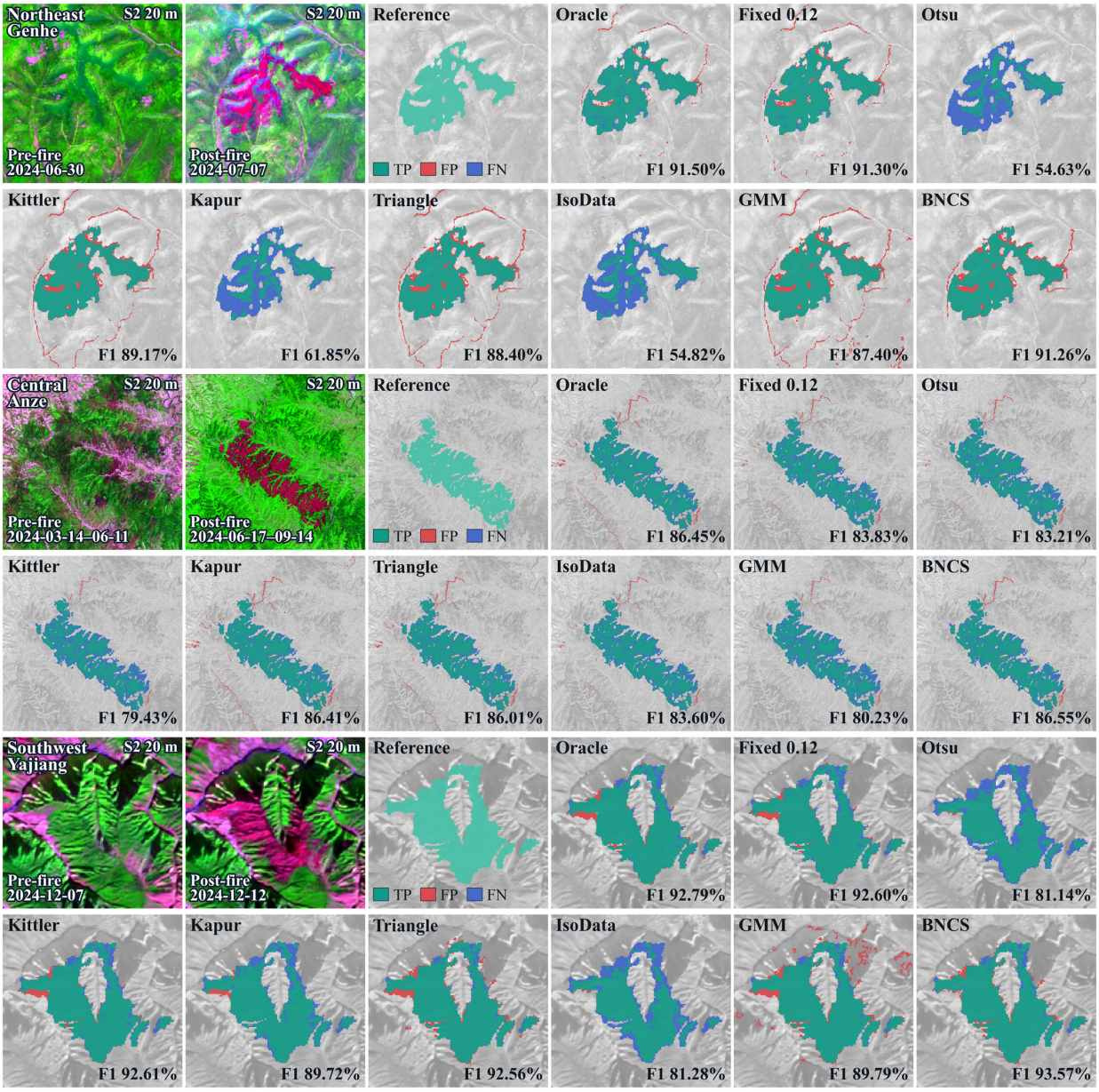

Figure 5. Comparison of burned-area segmentation results obtained using different methods for representative fire events across regions (S2). Pre-fire and post-fire images are displayed as SWIR2/NIR/red false-color composites. In the segmentation maps, turquoise, red, and blue indicate true positives (TP), false positives (FP), and false negatives (FN), respectively.

To examine the contributions of individual BNCS components, controlled ablation experiments were conducted using a common adaptive thresholding procedure as the baseline (Table 5). The baseline retained area-response-curve-based threshold selection and common postprocessing operations but excluded background constraints (B), response normalization (N), and core support (CS). It achieved macro F1 scores of 90.04% for S2 and 87.52% for L89. Introducing B alone improved F1 by 0.36 and 1.01 pp, respectively, indicating a positive but sensor-dependent contribution from background constraints. Adding N alone resulted in relatively small changes (−0.14 and +0.14 pp), indicating no consistent independent improvement in macro F1 across the two sensors. In contrast, CS improved F1 by 1.06 and 0.85 pp when applied individually, and by a further 1.49 and 1.32 pp when added after B and N. The complete BNCS framework achieved macro F1 scores of 91.83% and 89.56%, exceeding the baseline by 1.79 and 2.04 pp, respectively. To further

evaluate the effects of these components across different thresholding methods, controlled comparisons were conducted under the same B/N/CS configurations (Table S2). The results showed that component contributions varied across methods and configurations, suggesting that their effectiveness depended on the underlying threshold-selection approach. With all three components enabled, BNCS achieved the highest macro F1 scores for both sensors among the evaluated methods. Runtime comparisons further revealed differences in computational efficiency among the evaluated methods (Table S2). The complete BNCS framework required average inference times of 20.575 ms for S2 and 8.317 ms for L89. The additional B/N/CS operations increased runtime, although BNCS remained faster than GMM but slower than simpler thresholding methods.

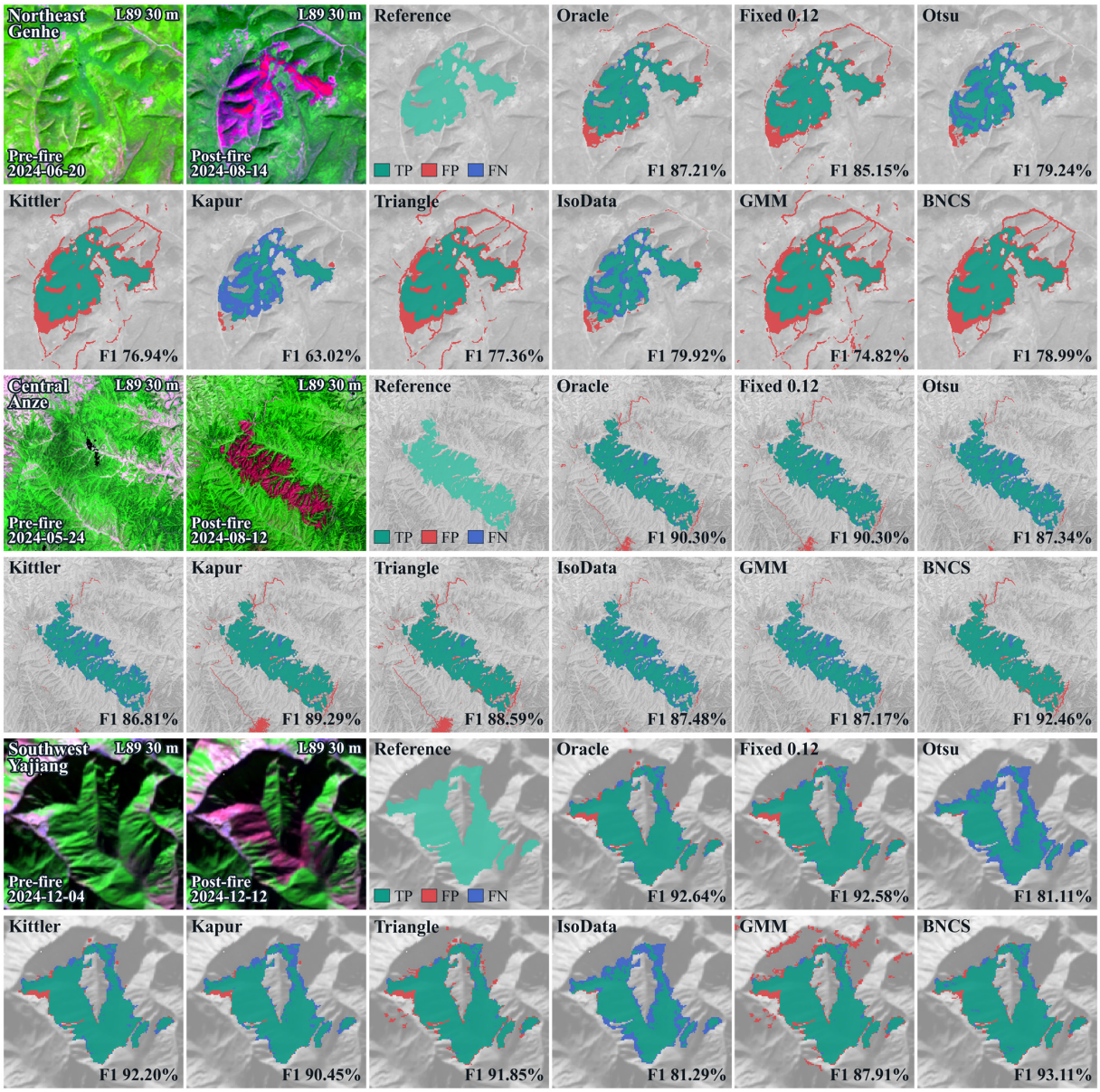

Figure 6. Comparison of burned-area segmentation results obtained using different methods for representative fire events across regions (L89). Pre-fire and post-fire images are displayed as SWIR2/NIR/red false-color composites. In the segmentation maps, turquoise, red, and blue indicate TP, FP, and FN, respectively.

Table 5. Ablation analysis of BNCS components for burned-area segmentation.

Configuration	Components			S2		L89	
	B	N	CS	Macro F1 (%)	ΔF1 (pp)	Macro F1 (%)	ΔF1 (pp)
Baseline	–	–	–	90.04	—	87.52	—
Baseline + B	✓	–	–	90.40	0.36	88.53	1.01
Baseline + N	–	✓	–	89.90	−0.14	87.66	0.14
Baseline + CS	–	–	✓	91.10	1.06	88.37	0.85
Baseline + B + N	✓	✓	–	90.34	0.30	88.24	0.72
Baseline + B + CS	✓	–	✓	91.34	1.30	89.62	2.10
Baseline + N + CS	–	✓	✓	91.59	1.54	89.40	1.88
Full BNCS	✓	✓	✓	**91.83**	**1.79**	**89.56**	**2.04**

Note: B, N, and CS denote background constraints, response normalization, and core support, respectively. The baseline retains area-response-curve-based adaptive threshold selection and common postprocessing. Checkmarks indicate enabled components, while dashes indicate disabled components. ΔF1 represents the difference in macro F1 relative to the baseline for each sensor, expressed in percentage points (pp). All metrics are averaged equally across 12 fire events per sensor. The full BNCS configuration and its corresponding results are highlighted in bold.

The applicability of BNCS to different burn indices was further evaluated using dNBR, RdNBR, and RBR (Table 6). The corresponding macro F1 scores were 91.83%, 91.62%, and 91.64% for S2, and 89.56%, 89.18%, and 89.44% for L89, respectively. The maximum F1 difference among the three indices was 0.21 pp for S2 and 0.38 pp for L89. The corresponding macro IoU values also showed relatively small differences, with dNBR achieving the highest F1 and IoU for both sensors. These results indicate that BNCS maintained comparable segmentation performance across the three burn indices within the evaluated fire events and sensor datasets.

Table 6. Segmentation performance of BNCS using different burn indices across S2 and L89 imagery.

Burn Index	S2				L89			
	Macro P (%)	Macro R (%)	Macro F1 (%)	Macro IoU (%)	Macro P (%)	Macro R (%)	Macro F1 (%)	Macro IoU (%)
RdNBR	91.94	91.76	91.62	84.83	88.25	91.68	89.18	81.06
RBR	91.94	91.86	91.64	84.87	88.45	91.88	89.44	81.49
dNBR	**91.98**	**92.14**	**91.83**	**85.20**	**88.54**	**91.92**	**89.56**	**81.67**

Note: All metrics are averaged equally across 12 fire events per sensor. The same BNCS framework and evaluation protocol were applied to all three burn indices. The highest macro F1 and macro IoU values for each sensor are highlighted in bold.

4.4. Sensitivity of Segmentation Performance to Background Extent

To evaluate the sensitivity of segmentation performance to background extent, all 12 fire events from northeastern, central, and southwestern China were included. For each event, three nested analysis regions were constructed with areas 3, 6, and 9 times that of the minimum bounding rectangle of the reference burned area. Segmentation performance was evaluated separately using S2 and L89 imagery, with image selection windows adjusted where necessary to ensure sufficient coverage of the expanded regions.

As shown in Figure 7 and Table 7, the evaluated methods exhibited different degrees of sensitivity to background expansion. BNCS maintained relatively high segmentation accuracy as the analysis area increased from 3× to 9×. Its macro F1 decreased from 91.88% to 90.72% and 89.60% for S2, corresponding to a reduction of 2.28 pp. For L89, the corresponding values were 89.86%, 87.97%, and 86.63%, representing a reduction of 3.23 pp. In comparison, Triangle showed larger reductions of 10.29 and 17.14 pp, while GMM de-

creased by 22.70 and 20.44 pp for S2 and L89, respectively. IsoData and Otsu also exhibited substantial performance declines under expanded background conditions.

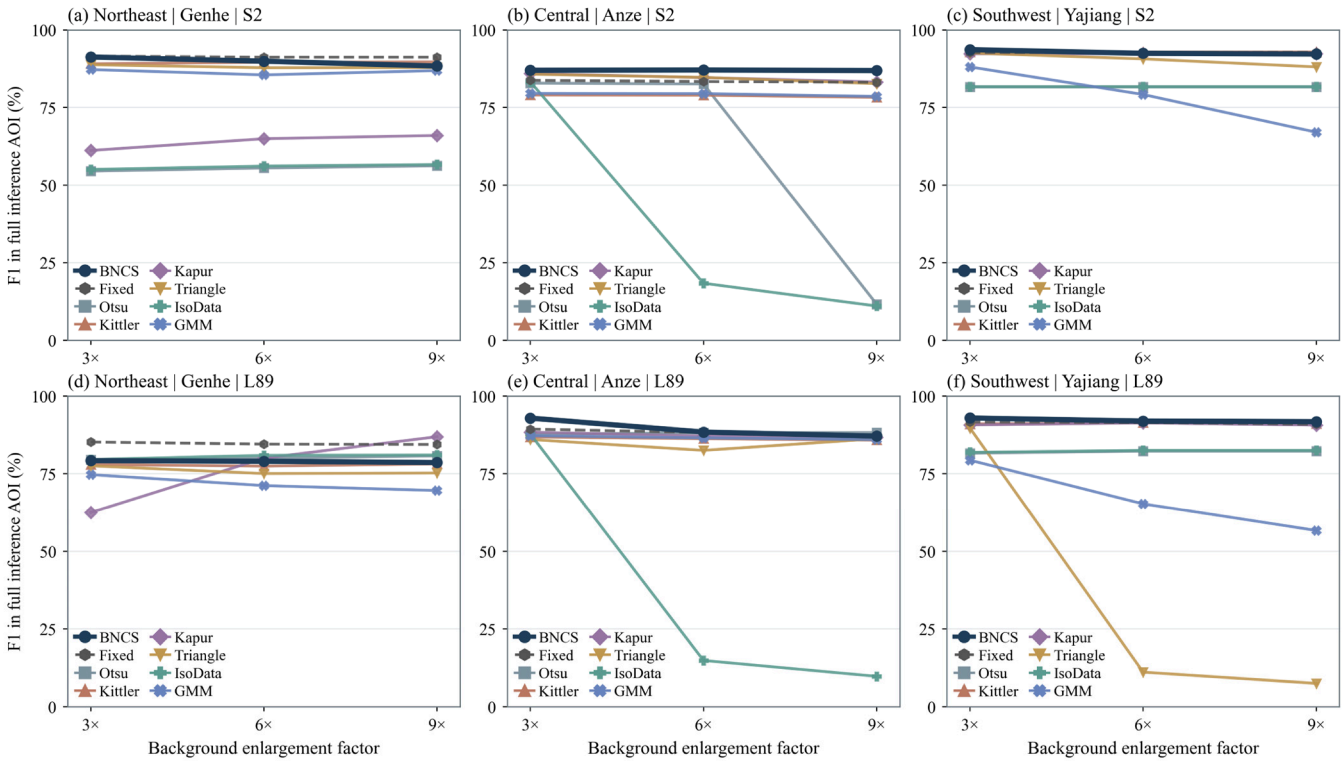

Figure 7. Sensitivity of segmentation accuracy to background extent across different methods.

Table 7. Sensitivity of segmentation performance to background extent across 12 fire events.

Methods	S2				L89			
	Macro F1 3× (%)	**Macro F1 6× (%)**	**Macro F1 9× (%)**	**ΔF1 (pp)**	**Macro F1 3× (%)**	**Macro F1 6× (%)**	**Macro F1 9× (%)**	**ΔF1 (pp)**
Otsu	77.87	75.14	56.99	−20.88	82.23	69.49	61.07	−21.16
Kapur	74.36	74.13	73.78	−0.58	75.30	72.93	73.65	−1.65
Kittler	78.71	73.19	70.72	−7.99	78.28	73.32	70.88	−7.41
Triangle	82.53	75.68	72.24	−10.29	76.04	62.13	58.91	−17.14
IsoData	70.55	54.32	50.41	−20.14	73.80	57.66	48.52	−25.28
GMM	79.78	56.48	57.08	−22.70	78.56	66.53	58.12	−20.44
Fixed	*82.62*	*76.32*	*72.64*	*−9.98*	*84.21*	*78.50*	*75.31*	*−8.90*
BNCS	**91.88**	**90.72**	**89.60**	**−2.28**	**89.86**	**87.97**	**86.63**	**−3.23**

Note: All F1 values are macro-averaged across 12 fire events for each sensor. The 3×, 6×, and 9× settings denote nested analysis regions with areas equal to 3, 6, and 9 times the area of the minimum bounding rectangle of the reference burned area, respectively. ΔF1 = F1(9×) − F1(3×), expressed in percentage points (pp), where negative values indicate a decrease in F1 as the analysis region expands. Fixed is a reference method and is shown in italics. BNCS is the proposed method, with its results highlighted in bold.

Although the magnitude of performance decline varied among methods, all evaluated methods showed lower macro F1 at 9× than at 3×. Kapur exhibited the smallest reductions among the conventional methods, at 0.58 and 1.65 pp for S2 and L89, respectively, but its absolute accuracy remained below that of BNCS. Across all three background extents, BNCS maintained the highest macro F1 for both sensors, with smaller performance declines than most comparison methods (Table 7). Spatial comparisons for three representative events—Genhe, Anze, and Yajiang—are presented in Figure 8, with results for the remaining nine events provided in Figures S12–S20.

4.5. Comparison with Existing Burned-Area Mapping Products

To further evaluate the applicability of BNCS, its segmentation results were compared with two existing mapping products, the Global Annual Burned Area Map (GABAM) and the GLAD Global Forest Loss due to Fire (GLAD GFL) dataset. The comparison was conducted using 11 fire events for which both products were available, with all results evaluated against the same reference burned areas on the corresponding S2 and L89 grids. As shown in Table 8, BNCS achieved macro F1 scores of 91.91% and 89.39% for S2 and L89, respectively, compared with 67.97% and 67.85% for GABAM and 60.46% and 60.33% for GLAD GFL. Although GABAM exhibited higher macro P (94.95% and 94.80%) than BNCS (91.73% and 87.87%), its macro R was substantially lower (55.72% and 55.59%, compared with 92.57% and 92.33% for BNCS). GLAD GFL showed a similar precision–recall imbalance, with macro R values of 49.00% and 48.87%. These results indicate that the existing products omitted a larger proportion of the reference burned areas in the evaluated events, while maintaining relatively high precision.

Table 8. Comparison of BNCS with existing burned-area products across 11 matched fire events.

Product	S2				L89			
	Macro P (%)	**Macro R (%)**	**Macro F1 (%)**	**Macro IoU (%)**	**Macro P (%)**	**Macro R (%)**	**Macro F1 (%)**	**Macro IoU (%)**
GABAM	94.95	55.72	67.97	54.56	94.8	55.59	67.85	54.36
GLAD GFL	91.06	49	60.46	47.9	91.06	48.87	60.33	47.73
BNCS	**91.73**	**92.57**	**91.91**	**85.37**	**87.87**	**92.33**	**89.39**	**81.44**

Note: All metrics are averaged equally across 11 matched fire events for each sensor. The Qinyuan 2025 event was excluded because the corresponding GABAM annual product was unavailable. GABAM represents annual burned-area mapping, while GLAD GFL identifies fire-attributed forest cover loss. All products were assessed against the same event-specific reference masks on the corresponding sensor grids. The highest value for each metric is highlighted in bold.

The spectral characteristics of detected and omitted burned pixels were further examined using event-specific dNBR values for all available events of each product(Table 9). Within the reference burned areas, pixels detected by GABAM or GLAD GFL generally exhibited higher median dNBR than omitted pixels across the evaluated events. For GABAM, the mean event-level median dNBR of detected and omitted pixels was 0.718 and 0.247 for S2, respectively, and 0.617 and 0.243 for L89. The corresponding values for GLAD GFL across its 12 available events were 0.650 and 0.258 for S2, and 0.574 and 0.239 for L89. When reference burned pixels were divided into lower and upper dNBR quartile groups within each event, GABAM omitted 81.26% and 5.95% of the two groups for S2, respectively, compared with 78.42% and 6.40% for L89. The corresponding omission rates for GLAD GFL were 80.40% and 18.23% for S2, and 80.22% and 18.35% for L89.

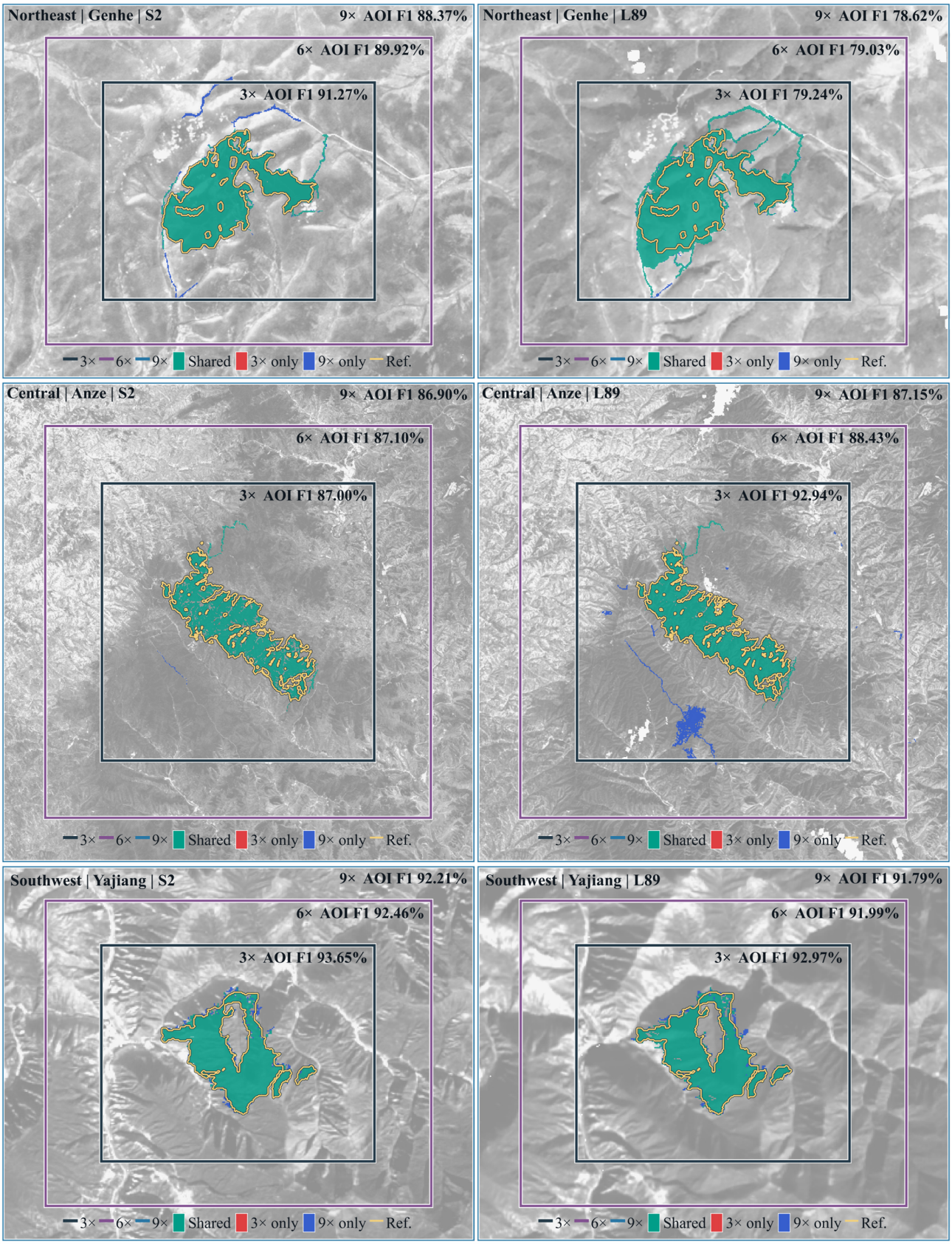

Figure 8. BNCS segmentation results for representative events under different background extents.

Spatial comparisons for representative events are presented in Figure 9, with additional results provided in Figures S21–S23. The two existing products generally showed more incomplete coverage of the reference burned areas than BNCS, although their omission patterns varied among events. These differences should be interpreted in relation to the products' mapping objectives and temporal definitions. GABAM represents annual burned areas, whereas GLAD GFL identifies fire-attributed forest cover loss rather than all burned surfaces. In particular, part of the late-2024 Yajiang event was assigned to 2025 in GLAD GFL, contributing to its low agreement with the event-specific reference when only the 2024 layer was evaluated. Overall, BNCS provided more complete delineation of the evaluated fire-event footprints, while the dNBR analysis revealed an association between weaker spectral responses and higher omission rates in the existing products.

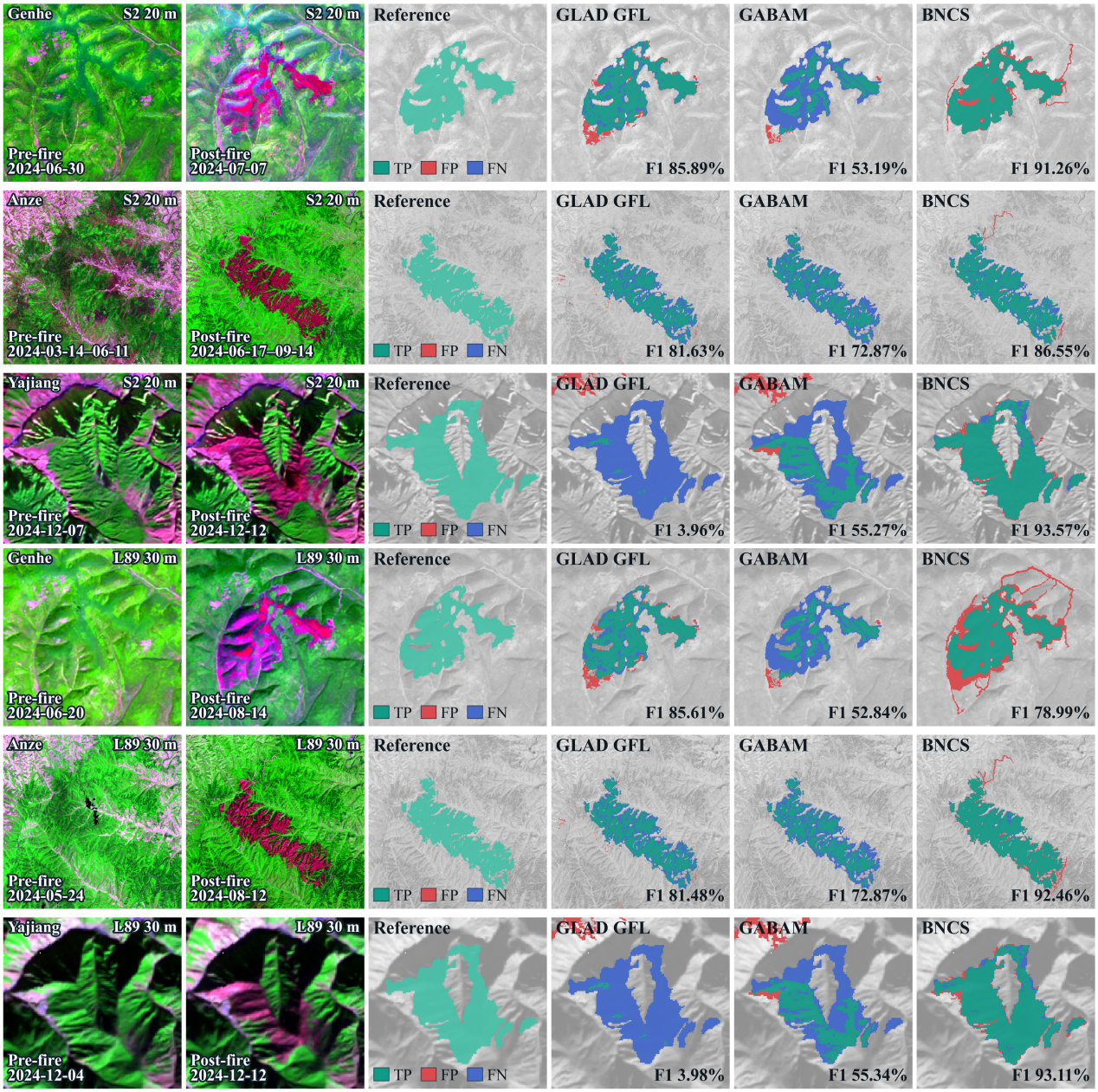

Figure 9. Comparison of BNCS with existing burned-area products for representative fire events. Pre-fire and post-fire images are displayed as SWIR2/NIR/red false-color composites. In the segmentation maps, turquoise, red, and blue indicate TP, FP, and FN, respectively.

Table 9. dNBR characteristics and omission rates of existing burned-area products.

Product	Sensor	Events	dNBR (det.)	dNBR (omit.)	ORlow (%)	ORhigh (%)
GABAM	S2	11	0.718	0.247	81.26	5.95
GABAM	L89	11	0.617	0.243	78.42	6.40
GLAD GFL	S2	12	0.650	0.258	80.40	18.23
GLAD GFL	L89	12	0.574	0.239	80.22	18.35

Note: det. and omit. denote detected and omitted reference burned pixels, respectively. dNBR values are arithmetic means of event-level medians. ORlow and ORhigh are mean omission rates for the lowest and highest event-specific dNBR quartiles. GABAM covers 11 events, whereas GLAD GFL covers 12 events.

5. Discussion

5.1. Scene Dependence of Burned-Area Spectral Responses and Threshold Transferability

Fixed-threshold methods are simple, efficient, and interpretable, and can achieve high segmentation accuracy when appropriately calibrated. The retrospectively optimized common threshold of 0.12 achieved relatively high macro F1 scores for S2 and L89. However, the spectral response analysis revealed substantial variations in NIR, SWIR2, and dNBR among fire events, including those within the same geographic region. These differences were accompanied by considerable variation in scene-specific optimal thresholds across fire events and sensors. Even though the training-derived optimal thresholds were concentrated within a narrow range of 0.10–0.13, their application to held-out scenes resulted in reduced segmentation accuracy. This suggests that the principal limitation of fixed thresholding lies not in the discriminative capability of a single threshold, but in the inconsistency of its decision position across scene-dependent response distributions.

These findings highlight that adaptive segmentation should focus on identifying an appropriate decision threshold from the response characteristics of the target scene rather than seeking a universally applicable fixed value. BNCS addresses this issue by constraining the candidate domain used for threshold estimation, representing dNBR responses on a scene-relative scale, and determining the low threshold from the structural characteristics of the area-response curve. In particular, the post-peak elbow of the area-decline curve provides a scene-specific criterion for threshold selection without using target-scene reference labels. This reduces dependence on thresholds calibrated from other scenes while allowing scene-specific threshold estimation. Nevertheless, background composition still influences threshold estimation, as shown by the background-extent experiments.

5.2. Mechanisms Underlying the Sensitivity of Threshold Estimation to Background Variation

The ablation experiments clarify the contributions of BNCS components. The common baseline, which retained area-response-curve-based adaptive threshold selection but excluded background constraints, response normalization, and core support, achieved macro F1 scores of 90.04% and 87.52% for S2 and L89, respectively. This indicates that the underlying adaptive thresholding procedure already provides substantial segmentation capability. The additional components serve different functions rather than producing equivalent or additive accuracy improvements. Background constraints restrict the candidate domain using pre-fire vegetation characteristics and post-fire water responses, reducing the influence of spectrally irrelevant surfaces on threshold estimation. Introducing background constraints alone improved macro F1 by 0.36 and 1.01 pp for S2 and L89, respectively, although their contributions varied across component configurations.

Within-scene distribution normalization maps spectral change responses onto a common numerical scale for threshold scanning and core-threshold calculation. However, the ablation experiments did not show a consistent independent accuracy improvement from normalization. In contrast, core support produced the clearest conditional F1 gains by preserving weaker-response regions connected to high-response cores while excluding unsupported patches. BNCS improved macro F1 by 1.79 and 2.04 pp over the common baseline for S2 and L89, respectively. These findings suggest that its effectiveness arises from integrating adaptive threshold estimation, candidate-domain constraints, and spatial support, with component contributions varying across configurations. Controlled comparisons also showed that Triangle combined with background constraints and core support achieved performance close to BNCS, whereas adding normalization did not consistently improve conventional methods. This indicates that component effectiveness depends on the underlying threshold selection criterion.

The background-extent experiments further demonstrate that threshold estimation is influenced not only by fire-induced spectral responses but also by the composition of pixels included in the analysis domain. As the analysis area expanded from 3× to 9×, changes in background composition could alter the spectral response distribution, even though the same pre- and post-fire imagery was used across the three extents within each event–sensor combination. For automatic thresholding methods based on histogram characteristics, between-class statistics, probability models, or entropy criteria, such changes may affect their optimization objectives and estimated thresholds, leading to different patterns of segmentation performance. For instance, IsoData, GMM, and Otsu exhibited substantial F1 reductions under background expansion, whereas Kapur showed smaller declines but maintained lower absolute accuracy than BNCS. These contrasting responses indicate that background sensitivity depends on how changes in pixel composition affect each method's threshold selection criterion.

Fixed thresholds remain numerically unchanged as the analysis domain expands, but segmentation accuracy may still vary with changes in the spatial composition of the evaluation domain. As the analysis extent increased from 3× to 9×, Fixed exhibited macro F1 reductions of 9.98 and 8.90 pp for S2 and L89, respectively, whereas the corresponding reductions for BNCS were 2.28 and 3.23 pp. BNCS also maintained the highest macro F1 across all three background extents for both sensors. However, because the threshold-estimation domain and evaluation domain expanded simultaneously, the observed performance changes reflect their combined effects and cannot be attributed exclusively to threshold estimation. These findings demonstrate the comparatively high accuracy and overall stability of BNCS under background expansion, while acknowledging its sensitivity to analysis extent.

The parameter sensitivity experiments provide additional evidence concerning the use of common parameter settings. The predefined background constraint thresholds maintained relatively high segmentation performance within the tested ranges, without requiring event-specific optimization. Nevertheless, this local stability should not be interpreted as evidence that the selected parameters are universally optimal. Their suitability may depend on vegetation conditions, background composition, and the distribution of spectral responses in other regions or sensor datasets.

5.3. Comparison with Existing Burned-Area Products and Implications for Mapping

Product comparisons highlighted differences between event-specific segmentation and annual mapping. Across the 11 matched fire events, BNCS achieved higher macro F1 and recall than GABAM and GLAD GFL, whereas GABAM exhibited higher precision. This contrast indicates that BNCS provided more complete coverage of the reference burned areas in the evaluated events, while the existing products generally adopted a more conservative

mapping outcome. The spectral analysis further showed that, within the reference burned areas, pixels detected by GABAM and GLAD GFL consistently exhibited higher median dNBR values than omitted pixels. The concentration of omissions among lower-response pixels suggests that incomplete coverage is associated, at least partly, with weaker fire-related spectral changes. However, lower dNBR values do not necessarily indicate low burn severity, and omission patterns cannot be attributed to a specific processing step without further evidence.

Importantly, the observed differences must also be considered in relation to the mapping objectives and temporal definitions of the products. GABAM represents annual burned areas, whereas GLAD GFL identifies fire-attributed forest cover loss rather than all burned surfaces. Consequently, differences in mapped coverage may reflect land-cover eligibility, annual attribution, and product-specific mapping criteria in addition to spectral detectability. For example, part of the late-2024 Yajiang event was attributed to 2025 in GLAD GFL, reducing its agreement with the event-specific reference when the 2024 layer was evaluated. Therefore, the higher agreement achieved by BNCS should be interpreted within the event-level burned-area delineation task considered here, rather than as evidence of general superiority over the broader operational objectives of these annual products. Nevertheless, the comparison highlights the value of event-specific adaptive segmentation for identifying burned areas incompletely represented in existing products.

5.4. Applicability, Limitations, and Future Research

BNCS is developed within a bi-temporal optical change-detection framework, and its applicability therefore depends on the quality of the input imagery and the reliability of fire-induced spectral changes. Non-fire disturbances, phenological variations, changes in soil moisture, and agricultural or forestry activities may produce dNBR responses resembling those associated with burning. In addition, prolonged intervals between fire occurrence and post-fire observation may weaken burn-related spectral evidence as vegetation recovers or surface conditions change. Although temporal search windows and image quality-control procedures were implemented to improve observation reliability, the effects of observation intervals on threshold estimation and burned-area boundary delineation were not systematically quantified in this study.

The present evaluation was also limited to 12 representative fire events in China and two optical multispectral sensor datasets, S2 and L89, which do not encompass the full diversity of ecological conditions and fire characteristics. Differences in segmentation performance between the two datasets may be associated with their spatial resolutions, spectral response characteristics, observation dates, and scene-specific environmental conditions. Because these factors were not independently controlled, their individual contributions cannot be isolated from the present comparison. For some event–sensor combinations, image acquisition dates or compositing windows differed between the main comparison and background-extent experiments. However, the three background extents within each event–sensor combination were evaluated using the same inputs, allowing their performance differences to be assessed under consistent observation conditions. In addition, uncertainties associated with manually delineated reference boundaries, particularly in weak-response and mixed-pixel areas, may introduce additional uncertainty into the accuracy assessment.

Furthermore, the core support mechanism assumes that a burned region contains sufficiently strong spectral responses to form a detectable high-confidence core. This assumption may not hold for extensive weak-response burns or scenes in which high-response signals have substantially diminished. Although core support provided the most consistent conditional improvement in the current experiments, its effectiveness under these conditions requires further investigation. Future research should therefore incorporate

independent fire events spanning broader climatic regions, vegetation types, and burn severity levels, while systematically examining the effects of observation intervals, sensor characteristics, and parameter transferability. The integration of higher-temporal-resolution optical observations, synthetic aperture radar (SAR), and complementary change-detection information may further improve burned-area delineation under complex background and weak-response conditions.

Overall, the findings indicate that cross-scene variability in burned-area segmentation arises from differences in background composition, fire-induced spectral response distributions, and the difficulty of simultaneously preserving weak-response burned regions and suppressing background-related false positives. BNCS addresses these challenges through scene-specific adaptive threshold estimation, background-constrained candidate selection, standardized response representation, and core-supported region selection. The results highlight the importance of adaptive threshold selection and core-supported region selection, while indicating that the contributions of individual components vary across configurations and datasets. The proposed framework therefore provides an interpretable approach to event-specific burned-area delineation without requiring a universally calibrated absolute threshold.

6. Conclusions

This study proposed BNCS, a scene-response-driven adaptive segmentation method for burned areas from forest and grassland fires that integrates background constraints, response normalization, and core support. The method aims to address three challenges in threshold-based burned-area delineation: heterogeneous background composition, scene-dependent variations in spectral change responses, and the difficulty of preserving weak-response burned regions while suppressing background-related false positives. By incorporating scene-specific response characteristics into adaptive threshold estimation and region selection, BNCS reduces dependence on predefined absolute thresholds. Background constraints restrict the candidate domain, distribution normalization provides a common scale for scene-relative response analysis, and core support retains spatially connected burned regions associated with high-response cores while excluding unsupported patches. These mechanisms provide an interpretable framework for burned-area segmentation under heterogeneous scene conditions.

Evaluation across 12 fire events using S2 and L89 imagery demonstrated the effectiveness of BNCS under varying scene conditions. The method achieved macro F1 scores of 91.83% and 89.56% for S2 and L89, respectively, outperforming the best conventional automatic thresholding method by 5.55 and 6.70 pp. Threshold-transfer experiments further showed that BNCS outperformed transferred fixed thresholds by 3.06–3.40 pp, indicating reduced dependence on thresholds calibrated from other scenes. Ablation experiments revealed that the underlying adaptive thresholding procedure provided substantial segmentation capability, while core support contributed the clearest conditional accuracy improvement. As the analysis extent increased from three to nine times the reference bounding-rectangle area, BNCS macro F1 decreased by 2.28 and 3.23 pp for S2 and L89, respectively, while remaining higher than that of all evaluated comparison methods. Comparisons with GABAM and GLAD GFL further showed that BNCS achieved more complete coverage of the reference burned areas in 11 matched fire events, although differences in mapping objectives and temporal definitions should be considered. Overall, the results support the potential of BNCS as an interpretable approach to event-specific burned-area delineation across heterogeneous scenes and optical satellite sensors.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18234123/s1. The supplementary material file contains Figures S1–S23 and Tables S1–S3.

Author Contributions: X.H. and X.Q. provided the conceptualization and methodology. X.H., Y.S., S.L., F.J. and L.Y. collected the experimental materials. X.H. and X.Q. performed the data collation and analysis. X.H. wrote the manuscript. X.Q. and S.H. contributed with suggestions. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, grant number "2023YFD2202000", and Natural Science Foundation of Inner Mongolia Autonomous Region, grant number "2026ZD049".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Sentinel-2 MultiSpectral Instrument and Landsat 8/9 Operational Land Imager data used in this study were obtained from Google Earth Engine (https://code.earthengine.google.com/, accessed on 9 September 2026).

Acknowledgments: We acknowledge Google Earth Engine for providing the satellite data access and processing platform, and the European Space Agency (ESA), the National Aeronautics and Space Administration (NASA), and the China Centre for Resources Satellite Data and Application (CRESDA) for providing satellite imagery used in this study.

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

References

1. Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanase, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I.; et al. Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sens. Environ.* **2019**, *225*, 45–64. [CrossRef]
2. Hawbaker, T.J.; Vanderhoof, M.K.; Schmidt, G.L.; Beal, Y.J.; Picotte, J.J.; Takacs, J.D.; Falgout, J.T.; Dwyer, J.L. The Landsat Burned Area algorithm and products for the conterminous United States. *Remote Sens. Environ.* **2020**, *244*, 111801. [CrossRef]
3. Qin, X.L.; Li, X.T.; Liu, S.C.; Liu, Q.; Li, Z.Y. Forest fire early warning and monitoring techniques using satellite remote sensing in China. *Natl. Remote Sens. Bull.* **2020**, *24*, 511–520. [CrossRef]
4. Li, Z.Y.; Chen, E.X.; Qin, X.L.; Guo, Y.; Tian, X.; Liu, Q.W.; Sun, B.; Zhao, L.; Cai, S.S.; Du, L.M.; et al. Research and application of forestry and grassland remote sensing technology in China: Recent progress, challenges and countermeasures. *Natl. Remote Sens. Bull.* **2025**, *29*, 1804–1830. [CrossRef]
5. Chen, Y.; Morton, D.C.; Randerson, J.T. Remote sensing for wildfire monitoring: Insights into burned area, emissions, and fire dynamics. *One Earth* **2024**, *7*, 1022–1028. [CrossRef]
6. Lizundia-Loiola, J.; Franquesa, M.; Khairoun, A.; Chuvieco, E. Global burned area mapping from Sentinel-3 Synergy and VIIRS active fires. *Remote Sens. Environ.* **2022**, *282*, 113298. [CrossRef]
7. Roteta, E.; Bastarrika, A.; Franquesa, M.; Chuvieco, E. Landsat and Sentinel-2 based burned area mapping tools in Google Earth Engine. *Remote Sens.* **2021**, *13*, 816. [CrossRef]
8. Bastarrika, A.; Rodriguez-Montellano, A.; Roteta, E.; Hantson, S.; Franquesa, M.; Torre, L.; Gonzalez-Ibarzabal, J.; Artano, K.; Martinez-Blanco, P.; Mesanza, A.; et al. An automatic procedure for mapping burned areas globally using Sentinel-2 and VIIRS/MODIS active fires in Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 232–245. [CrossRef]
9. Miller, J.D.; Thode, A.E. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sens. Environ.* **2007**, *109*, 66–80. [CrossRef]
10. Parks, S.A.; Dillon, G.K.; Miller, C. A new metric for quantifying burn severity: The Relativized Burn Ratio. *Remote Sens.* **2014**, *6*, 1827–1844. [CrossRef]
11. Llorens, R.; Sobrino, J.A.; Fernández, C.; Fernández-Alonso, J.M.; Vega, J.A. A methodology to estimate forest fires burned areas and burn severity degrees using Sentinel-2 data. Application to the October 2017 fires in the Iberian Peninsula. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102243. [CrossRef]
12. Hu, X.Y.; Jiang, F.; Qin, X.L.; Huang, S.S.; Meng, F.X.; Yu, L.F. Exploration of suitable spectral bands and indices for forest fire severity evaluation using ZY-1 hyperspectral data. *Forests* **2025**, *16*, 640. [CrossRef]

13. Chatzopoulos-Vouzoglanis, K.; Reinke, K.J.; Soto-Berelov, M.; Jones, S.D. Are fire intensity and burn severity associated? Advancing our understanding of FRP and NBR metrics from Himawari-8/9 and Sentinel-2. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *127*, 103673. [CrossRef]
14. Viedma, O.; Chico, F.; Fernández, J.J.; Madrigal, C.; Safford, H.D.; Moreno, J.M. Disentangling the role of prefire vegetation vs. burning conditions on fire severity in a large forest fire in SE Spain. *Remote Sens. Environ.* **2020**, *247*, 111891. [CrossRef]
15. Lee, K.; van Leeuwen, W.J.D.; Gillan, J.K.; Falk, D.A. Examining the impacts of pre-fire forest conditions on burn severity using multiple remote sensing platforms. *Remote Sens.* **2024**, *16*, 1803. [CrossRef]
16. Chen, D.; Fu, C.; Hall, J.V.; Hoy, E.E.; Loboda, T.V. Spatio-temporal patterns of optimal Landsat data for burn severity index calculations: Implications for high northern latitudes wildfire research. *Remote Sens. Environ.* **2021**, *258*, 112393. [CrossRef]
17. Morresi, D.; Marzano, R.; Lingua, E.; Motta, R.; Garbarino, M. Mapping burn severity in the western Italian Alps through phenologically coherent reflectance composites derived from Sentinel-2 imagery. *Remote Sens. Environ.* **2022**, *269*, 112800. [CrossRef]
18. Howe, A.A.; Parks, S.A.; Harvey, B.J.; Saberi, S.J.; Lutz, J.A.; Yocom, L.L. Comparing Sentinel-2 and Landsat 8 for burn severity mapping in western North America. *Remote Sens.* **2022**, *14*, 5249. [CrossRef]
19. Picotte, J.J.; Cansler, C.A.; Kolden, C.A.; Lutz, J.A.; Key, C.; Benson, N.C.; Robertson, K.M. Determination of burn severity models ranging from regional to national scales for the conterminous United States. *Remote Sens. Environ.* **2021**, *263*, 112569. [CrossRef]
20. Tang, M.; Zhu, X.; Zhao, C.; Song, J.; Xiao, C.; Guo, R. REBAI: Development and validation of a novel indicator for burned area detection using Sentinel-2 images. *Ecol. Indic.* **2026**, *182*, 114501. [CrossRef]
21. Sezgin, M.; Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **2004**, *13*, 146–168. [CrossRef]
22. Gupta, L.; Sortrakul, T. A Gaussian-mixture-based image segmentation algorithm. *Pattern Recognit.* **1998**, *31*, 315–325. [CrossRef]
23. Woźniak, E.; Aleksandrowicz, S. Self-adjusting thresholding for burnt area detection based on optical images. *Remote Sens.* **2019**, *11*, 2669. [CrossRef]
24. Sismanis, M.; Chadoulis, R.T.; Manakos, I.; Drosou, A. An unsupervised burned area mapping approach using Sentinel-2 images. *Land* **2023**, *12*, 379. [CrossRef]
25. Milczarek, M.; Aleksandrowicz, S.; Kita, A.; Chadoulis, R.-T.; Manakos, I.; Woźniak, E. Object- versus pixel-based unsupervised fire burn scar mapping under different biogeographical conditions in Europe. *Land* **2023**, *12*, 1087. [CrossRef]
26. Duan, Q.; Liu, R.G.; Chen, J.L.; Wei, X.X.; Liu, Y.; Zou, X. Burned area detection from a single satellite image using an adaptive thresholds algorithm. *Int. J. Digit. Earth* **2024**, *17*, 2376275. [CrossRef]
27. Liu, P.; Liu, Y.X.; Guo, X.X.; Zhao, W.J.; Wu, H.S.; Xu, W.X. Burned area detection and mapping using time series Sentinel-2 multispectral images. *Remote Sens. Environ.* **2023**, *296*, 113753. [CrossRef]
28. Kulinan, A.S.; Cho, Y.; Park, M.; Park, S. Rapid wildfire damage estimation using integrated object-based classification with auto-generated training samples from Sentinel-2 imagery on Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *126*, 103628. [CrossRef]
29. Gibson, R.; Danaher, T.; Hehir, W.; Collins, L. A remote sensing approach to mapping fire severity in south-eastern Australia using Sentinel-2 and Random Forest. *Remote Sens. Environ.* **2020**, *240*, 111702. [CrossRef]
30. Roodsarabi, Z.; Farhadi, H.; Ebadi, H.; Kiani, A. Machine learning-based burned area detection using Sentinel-2 imagery and spectral indices. *Environ. Monit. Assess.* **2025**, *197*, 905. [CrossRef]
31. Martins, V.S.; Roy, D.P.; Huang, H.; Boschetti, L.; Zhang, H.K.; Yan, L. Deep learning high resolution burned area mapping by transfer learning from Landsat-8 to PlanetScope. *Remote Sens. Environ.* **2022**, *280*, 113203. [CrossRef]
32. Sui, T.; Huang, Q.; Wu, M.; Wu, M.; Zhang, Z. BiAU-Net: Wildfire burnt area mapping using bi-temporal Sentinel-2 imagery and U-Net with attention mechanism. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *132*, 104034. [CrossRef]
33. Padilla, M.; Ramo, R.; Gomez-Dans, J.L.; Sierra, S.; Mota, B.; Lacaze, R.; Tansey, K. Near-real time monitoring of burned area at global scale based on deep learning. *Int. J. Remote Sens.* **2025**, *46*, 5996–6038. [CrossRef]
34. Liu, P.; Liu, Y.; Guo, X.; Liu, Y.; Zhao, W.; Xu, W. Faster, better, and more accurate mapping of burned areas using Sentinel-2 multispectral images. *Remote Sens. Environ.* **2026**, *333*, 115137. [CrossRef]
35. Liu, S.; Xue, Y.; Chen, H.; Chen, Y.; Zhan, T. Segmentation of Any Fire Event (SAFE): A rapid and high-precision approach for burned area extraction using Sentinel-2 imagery. *Remote Sens.* **2025**, *17*, 54. [CrossRef]
36. Seo, Y.; Kim, S.H.; Kafatos, M.; Kim, J.; Lee, Y. Deep learning-based burned area mapping of California wildfires using Sentinel-2 and Landsat-8 imagery enhanced with super-resolution techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *153*, 105526. [CrossRef]
37. Campagnolo, M.L.; Libonati, R.; Rodrigues, J.A.; Pereira, J.M.C. A comprehensive characterization of MODIS daily burned area mapping accuracy across fire sizes in tropical savannas. *Remote Sens. Environ.* **2021**, *252*, 112115. [CrossRef]
38. Franquesa, M.; Stehman, S.V.; Chuvieco, E. Assessment and characterization of sources of error impacting the accuracy of global burned area products. *Remote Sens. Environ.* **2022**, *280*, 113214. [CrossRef]

39. Zack, G.W.; Rogers, W.E.; Latt, S.A. Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **1977**, *25*, 741–753. [CrossRef] [PubMed]
40. Rosin, P.L. Unimodal thresholding. *Pattern Recognit.* **2001**, *34*, 2083–2096. [CrossRef]
41. Satopää, V.; Albrecht, J.; Irwin, D.; Raghavan, B. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, MN, USA, 20–24 June 2011; pp. 166–171. [CrossRef]
42. Long, T.; Zhang, Z.; He, G.; Jiao, W.; Tang, C.; Wu, B.; Zhang, X.; Wang, G.; Yin, R. 30 m resolution global annual burned area mapping based on Landsat images and Google Earth Engine. *Remote Sens.* **2019**, *11*, 489. [CrossRef]
43. Tyukavina, A.; Potapov, P.; Hansen, M.C.; Pickens, A.; Stehman, S.V.; Turubanova, S.; Parker, D.; Zalles, V.; Lima, A.; Kommareddy, I.; et al. Global trends of forest loss due to fire, 2001–2019. *Front. Remote Sens.* **2022**, *3*, 825190. [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.