

BAM: A physics-informed self-supervised framework for near-real-time wildfire burned area mapping from multi-source earth observation

Mirza Waleed^a, Muhammad Bilal^{b,c,*}

^a Department of Geography, Hong Kong Baptist University, Hong Kong Special Administrative Region of China, China

^b Architecture and City Design Department, College of Design and Built Environment, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

^c Center for Aviation & Space Exploration, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

ARTICLE INFO

Keywords:

Burned area

Self-supervised learning

Earth observation

Near-real-time monitoring

Google earth engine

ABSTRACT

The operational monitoring of global wildfires is frequently constrained by the delayed availability of reliable reference labels, the latency of standard surface reflectance products, and spectral confusion between charred biomass and naturally dark surfaces. These limitations are especially problematic for near-real-time wildfire monitoring, where burned-area maps must be produced before manual training data or fully processed reference products are available. To address these limitations, this study presents the Burn Area Mapper (BAM), a fully automated, physics-informed machine learning framework implemented in Google Earth Engine (GEE). The framework utilizes a Physics-Informed Self-Supervision paradigm that eliminates the need for manual annotation by leveraging the stoichiometry of combustion to generate scene-specific training labels. Central to this approach is the novel Automated Temporal Burn Index (ATBI), which utilizes multiplicative scaling of NIR and SWIR bands to enforce a strict bimodal distribution, enabling an adaptive Otsu algorithm to automatically isolate high confidence burn signals. These physics-derived labels are refined by a Gradient Tree Boost (GTB) classifier incorporating textural and topographic features. We validated the framework using internal spatially blocked assessment across 15 wildfire events spanning 6 continents, achieving a global mean F1-score of 0.994 without regional retraining. Comparative benchmarking against the MODIS MCD64A1 coarse-resolution product revealed that BAM captures up to 93% more burned area in fragmented landscapes. Furthermore, for one test event, validation against 3 m PlanetScope imagery at 30 m aggregated scale confirms ($R^2 = 0.80$) that BAM's probabilistic output effectively proxies the sub-pixel burn fraction, resolving fine-scale fire fronts and unburned refugia that are invisible to coarse-resolution systems. By combining rapidly available Landsat 8/9 TOA observations, SREM atmospheric correction, automated pseudo-label generation, and GTB refinement within GEE, BAM provides a scalable open-source framework for near-real-time burned-area mapping and post-fire assessment.

1. Introduction

The escalating frequency and intensity of global wildfire events have necessitated a paradigm shift in Earth observation, moving from historical post-fire assessment to active, Near-Real-Time (NRT) monitoring (Bowman et al., 2020). The operational viability of any global wildfire mapping framework is fundamentally constrained by data latency and the ability to function in data-scarce environments (Chen et al., 2024). For disaster response agencies, the utility of satellite-derived burn perimeters is often compromised by the processing delays associated with standard Surface Reflectance (SR) products, which require ancillary

atmospheric data that may not be available during the critical window of active fire suppression (Vermote et al., 2016). Furthermore, these standard workflows often employ conservative cloud masking algorithms that aggressively discard pixels in hazy or smoky conditions, precisely the atmospheric states most prevalent during active combustion (Zhu and Woodcock, 2012). Consequently, a significant portion of optical imagery becomes unusable, creating observational gaps that hinder continuous tracking of fire fronts in tropical and high-latitude regions. To bridge this gap, there is an urgent need for automated systems that can extract valid burn signals from noisy NRT data without relying on external atmospheric models.

* Corresponding author at: Architecture and City Design Department, College of Design and Built Environment, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia.

E-mail address: muhammad.bilal@kfupm.edu.sa (M. Bilal).

<https://doi.org/10.1016/j.jag.2026.105517>

Received 10 April 2026; Received in revised form 5 July 2026; Accepted 4 August 2026

1569-8432/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The pursuit of automated mapping has long been obstructed by a fundamental radiometric challenge: the spectral ambiguity between charred biomass and other naturally occurring dark surface features. In heterogeneous landscapes, identifying burn scars is not merely a matter of detecting change but of distinguishing pyrogenic carbonization from deep-water bodies, topographic shadows, dark organic soils, and ephemeral cloud shadows (Liu et al., 2023). Traditional spectral indices, most notably the Differenced Normalized Burn Ratio (dNBR), have served as the industry standard for decades (Holsinger et al., 2022). However, these indices typically rely on an additive-difference formulation that implicitly assumes a linear trajectory of surface change.

Comprehensive analysis of global wildfire events suggests that fire-induced surface alterations are, in contrast, fundamentally non-linear (Alonso-González and Fernández-García, 2021; Artés et al., 2019; Bhattarai et al., 2025). Combustion triggers two simultaneous spectral shifts: a sharp reduction in NIR plateau reflectance attributable to the collapse of spongy mesophyll structures, and a concurrent increase in SWIR reflectance driven by the removal of moisture-laden vegetation and the exposure of dry soil, ash, and carbonized biomass. Linear indices frequently fail to capture this complex stoichiometry, leading to significant histogram overlap between burned and unburned pixel populations (Balakavi et al., 2025). This transition zone uncertainty is particularly pronounced in complex environments, such as temperate broadleaf forests, where the spectral similarity between canopy shadow and scorch can cause separation indices to collapse, resulting in substantial omission errors or the need for manual thresholding (Artés et al., 2019).

To bypass this non-linear bottleneck, operational frameworks must integrate robust preprocessing strategies that can extract surface signals from noisy data. The use of physics-based inversions, such as the Simplified and Robust Surface Reflectance Estimation Method (SREM), offers a pathway to eliminate dependencies on external atmospheric models (Bilal et al., 2019). By performing inversion using only image-based priors, it becomes possible to recover valid burn perimeters even in transboundary events characterized by extreme atmospheric opacity, such as the heavy smoke accumulation observed in the 2024 India-Nepal border fires (Balakavi et al., 2025).

While Machine Learning (ML) offers the potential to resolve complex non-linear spectral relationships, its application to global wildfire mapping is currently hindered by the scarcity of high-quality, pixel-level training data. A critical limitation of many existing regional algorithms is their tendency to overfit to specific vegetation types or atmospheric conditions, resulting in significant performance degradation, often referred to as domain shift, when applied to unseen biomes (Chen et al., 2024; Costa-Saura et al., 2022; He et al., 2024; Quintero et al., 2025; Shi and Touge, 2022). Consequently, models trained on North American datasets frequently fail to generalize to the distinct fire regimes of the African Savanna or the Mongolian Steppe.

To overcome the dependency on manual annotation, recent research has pivoted toward self-supervised paradigms where physics-informed feature engineering replaces human labeling (Bilal, 2025a, 2025b; Defourny et al., 2019; Liu et al., 2023). This approach is essential not only for automation but for resolving the resolution gap inherent in current global baselines. While the MODIS MCD64A1 product serves as the standard for global carbon emission estimates, its coarse 500-meter spatial resolution inherently limits its capacity to resolve sub-pixel heterogeneity. This limitation leads to substantial omission errors in heterogeneous landscapes, where dendritic fire spread patterns along narrow valley bottoms or agricultural interfaces remain effectively invisible to coarse sensors (Alonso-González and Fernández-García, 2021; Bilal, 2025a). The ability to recover these fragmented burn scars is critical for accurate ecological assessment, particularly in identifying unburned refugia that serve as seed sources for regeneration.

To address these challenges, this study introduces the Burn Area Mapper (BAM), a hierarchical, automated framework that synthesizes physics-based spectral analysis with data-driven ML refinement. BAM

adopts an automated self-supervision paradigm where the physical properties of the fire itself are used to train a scene-specific classifier, eliminating the need for external training data. Central to this approach is the novel Automated Temporal Burn Index (ATBI), which integrates a multiplicative scaling factor derived from the SWIR and NIR bands (Bilal, 2025a). By acting as a noise reducer, the ATBI enforces a strict bimodal distribution in the feature space, enabling an adaptive statistical Otsu algorithm to generate reliable labels without human intervention.

These physics-derived labels are subsequently refined by an ML classifier, which ingests textural (Gray-Level Co-occurrence Matrix (GLCM)), spectral, and topographic features to resolve spatial ambiguities and filter noise. To address these atmospheric constraints and bypass the latency associated with standard products, this study integrates SREM to enable robust atmospheric correction. This ensures operational continuity even in data-scarce, hazy environments.

The specific objectives of this research are to:

- 1) Develop a fully automated framework that leverages physics-informed self-supervision to generate high-precision burn perimeters in near-real-time.
- 2) Validate the zero-shot transferability of the framework across heterogeneous ecological gradients, assessing its stability in diverse biomes ranging from the boreal taiga to tropical deciduous forests.
- 3) Quantify the resolution enhancement provided by the 30-meter Landsat-based assessment compared to the operational MODIS MCD64A1 baseline, particularly in identifying fragmented burns and unburned refugia.

2. Methodology

2.1. BAM framework overview

The Burn Area Mapper (BAM) framework (Fig. 1) operates as a sequential, automated three-stage pipeline that synthesizes physics-based spectral constraints with data-driven machine-learning refinement. First, Landsat 8/9 imagery and ancillary global datasets are ingested, atmospherically corrected using SREM, and filtered with a unified exclusion mask. Second, scene-specific pseudo-labels are generated from the ATBI using adaptive Otsu thresholding. Third, a Gradient Tree Boost (GTB) classifier refines these labels using spectral, textural, and topographic predictors to produce 30 m burn-probability and binary burned-area maps. This modular design enables rapid scene-specific mapping without manual annotation or regional retraining. Fig. 1 provides a conceptual overview of the BAM workflow, while the detailed implementation choices, including feature-stack composition, pseudo-label sampling, parameter settings, and validation design, are described in Sections 2.3–2.7 and Supplementary Tables S2–S4.

2.2. Global validation sites and sensor fusion

To evaluate the efficacy and global transferability of the framework, this study analyzes a stratified dataset of 15 distinct wildfire events spanning 6 continents and biome types. The selection criteria prioritized biogeographical diversity to ensure the framework is tested against a wide spectrum of environmental complexities, including heterogeneous topography, atmospheric interference, and varying vegetation densities. Detailed metadata for the 15 selected wildfire events and the complete multi-source input datasets (Landsat 8/9, JRC Water, ESA WorldCover, and FABDEM) are provided in Supplementary Tables S1 and S2.

The study area encompasses representative samples from Boreal forests in Canada and Russia, Temperate conifer forests in North America, Mediterranean sclerophyll in Chile and Europe, and Subtropical broadleaf systems in Asia. This global distribution is illustrated in Fig. 2 (Panel A).

To ensure the temporal density required for active fire monitoring,

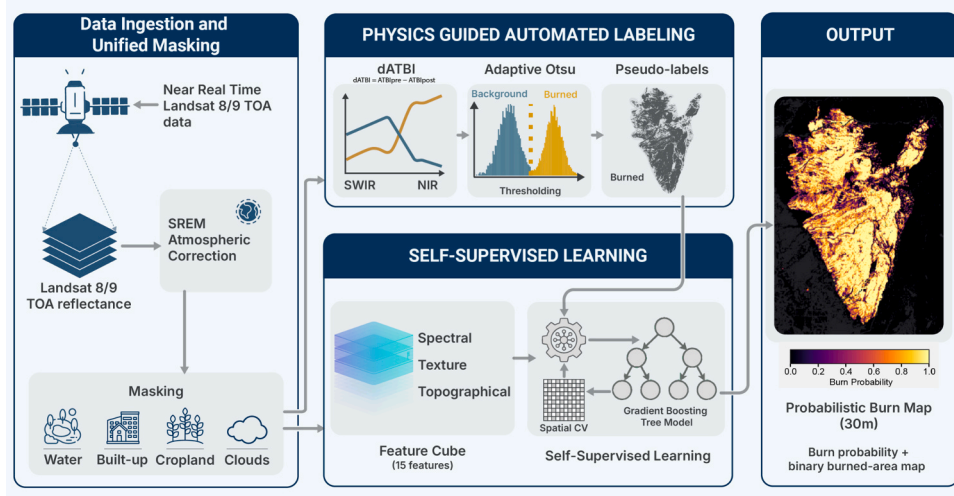


Fig. 1. Conceptual workflow of the BAM framework. Near-real-time Landsat 8/9 TOA observations are converted to surface reflectance using SREM and filtered using unified exclusion masks. The corrected pre- and post-fire observations are transformed into dATBI, which is thresholded using adaptive Otsu to generate conservative pseudo-labels. A scene-specific GTB classifier then refines these pseudo-labels using 15 spectral, textural, and topographic predictors to produce 30 m burn-probability and binary burned-area outputs.

the framework employs a sensor fusion approach that leverages both the Landsat 8 (OLI) and Landsat 9 (OLI-2) constellations. By interleaving acquisitions from both platforms, the system achieves a nominal revisit period of 8 days. Fig. 2 (Panel B) quantifies operational data availability for 2025, demonstrating that the combined constellation provides sufficient observational frequency to track active fire fronts across most global burn regimes, significantly reducing the latency gaps inherent in single-sensor workflows.

3. Data acquisition, preprocessing, and unified masking

3.1. Multi-source data acquisition and sensor fusion

The pipeline ingests Landsat 8/9 Collection 2 Tier 1 top-of-atmosphere (TOA) reflectance data because these observations are available more quickly than standard surface reflectance products. TOA scenes are converted to surface reflectance using the Simplified and Robust Surface Reflectance Estimation Method (SREM), which performs image-based inversion without external atmospheric inputs and is therefore suitable for smoky or hazy NRT conditions (Bilal et al., 2019). Topographic context is represented using FABDEM at 30 m (Hawker et al., 2022) to derive elevation, slope, and aspect, which help account for terrain-driven illumination and fire-spread effects.

3.2. Unified exclusion masking

To reduce commission errors during pseudo-label generation, BAM applies a unified exclusion mask immediately after atmospheric correction. Permanent and seasonal water are removed using JRC Global Surface Water, with shoreline buffering to reduce spectral mixing; a global ocean mask is also applied. Built-up areas were excluded to reduce false positives due to urban shadows and dark impervious surfaces. Croplands were also excluded in the default BAM configuration because the present study focuses on wildland and forest burned-area mapping, while harvested fields, bare agricultural soils, and seasonal crop removal can produce dATBI-like spectral changes unrelated to wildfire. However, this choice can remove true agricultural burns or cropland–wildland interface fires where ESA WorldCover classifies the surface as cropland. Therefore, cropland masking should be interpreted as a conservative default option rather than a universal requirement. Details of the input datasets and their sources are provided in Table S2.

3.3. Physics-informed self-supervised label generation

BAM adopts a self-supervised strategy in which the target scene provides its own initial training signal. Pseudo-labels are generated from ATBI, a non-linear spectral feature designed to amplify the radiometric signature of combustion and improve the separability required for automated thresholding.

3.4. The automated temporal burn index (ATBI)

Standard indices such as the NBR rely on an additive difference formulation, which often yields insufficient separability in complex environments where the spectral signature of burns overlaps with dark soils or shadows (Bilal, 2025b). To address this, we introduce the ATBI, a non-linear spectral feature designed to physically amplify the radiometric signal of carbonized biomass (Bilal, 2025a).

ATBI was calculated separately for the pre-fire and post-fire epochs using Landsat 8/9 near-infrared (NIR; Band 5) and shortwave infrared 2 (SWIR2; Band 7) reflectance after SREM correction. In this study, the term “NBR-like component” refers to the normalized NIR–SWIR2 contrast term, while the physical scaling factor refers to the multiplicative SWIR2/NIR ratio. ATBI is defined as Equation (1):

$$ATBI_t = \left(\frac{\rho_{NIR,t} - \rho_{SWIR2,t}}{\rho_{NIR,t} + \rho_{SWIR2,t}} \right) \times \left(\frac{\rho_{SWIR2,t}}{\rho_{NIR,t}} \right) \quad (1)$$

where $\rho_{NIR,t}$ and $\rho_{SWIR2,t}$ are the SREM-corrected reflectance values at epoch t . The first term is structurally similar to the Normalized Burn Ratio (NBR), but the second term non-linearly amplifies the combustion signal. In burned pixels, NIR reflectance decreases due to vegetation structure collapse, while SWIR2 reflectance increases due to moisture loss, ash exposure, and carbonized material. Therefore, the SWIR2/NIR scaling term increases the separation between burned and unburned surfaces.

The temporal ATBI anomaly was then computed as Equation (2):

$$dATBI = ATBI_{pre} - ATBI_{post} \quad (2)$$

This convention produces positive dATBI values for burned pixels because post-fire ATBI decreases strongly after combustion. The dATBI layer was used only for pseudo-label generation and was excluded from the subsequent GTB feature stack to reduce leakage between pseudo-labeling and model refinement.

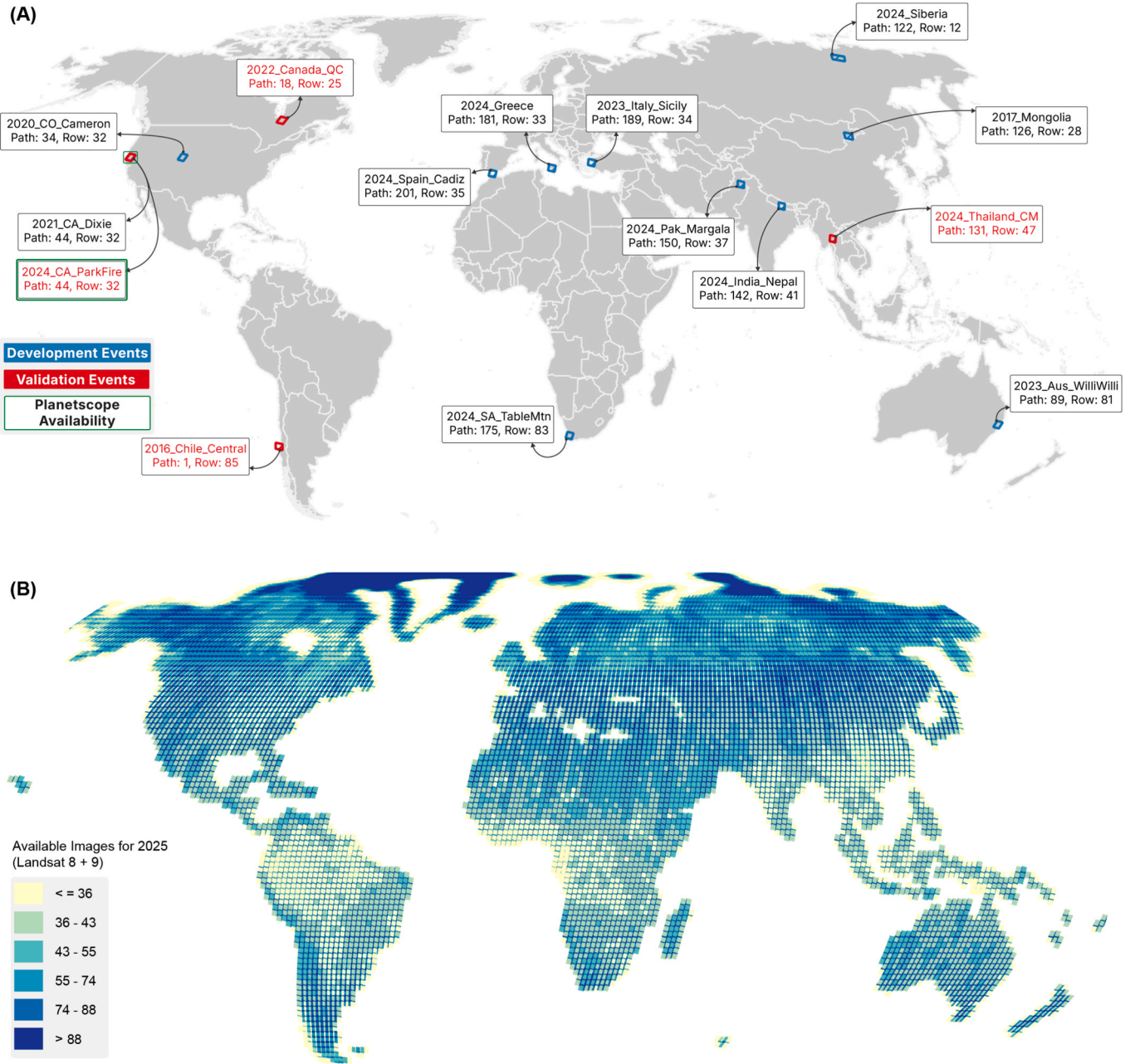


Fig. 2. Study design and Landsat acquisition density. (A) Global distribution of the 15 wildfire events used for development and validation. (B) Annual acquisition density for combined Landsat 8/9 observations, illustrating nominal NRT coverage.

3.5. Adaptive thresholding and pseudo-label extraction

The separation of burned and unburned training samples is performed dynamically using Otsu's method, which treats the *dATBI* histogram as a bimodal distribution and iteratively searches for the threshold t^* that maximizes the inter-class variance. However, standard Otsu thresholding can be sensitive to histogram noise in low-severity fire events. To ensure the generation of high-precision labels, the framework implements a Robust Parameterized Thresholding strategy:

1. **Conservative Bias Injection:** A positive bias ($\delta = 0.02$) is added to the calculated Otsu threshold ($t_{final} = t_{otsu} + \delta$). This constraint enforces a conservative selection strategy, prioritizing User's Accuracy (Precision) to ensure that the burned training samples represent core fire areas with high certainty, rather than ambiguous edge pixels.

2. **Minimum Stability Floor:** To prevent over-segmentation in scenes with negligible fire activity, a minimum threshold floor $t_{min} = 0.03$ is strictly enforced. If the calculated threshold falls below this value, it indicates a unimodal (no fire) distribution, and the system defaults to a background-only state.

The resulting binary mask serves as the automated ATBI guided label set, providing the instructional signal necessary for the subsequent supervised learning phase.

To quantify burned-unburned separability, we used two complementary metrics: the Histogram Separation Index (HSI) and the M-statistic. For each index, burned and unburned values were converted into normalized distributions using common histogram bins or kernel density estimates. The overlap coefficient was calculated as Equation (3):

$$OVL = \sum_{i=1}^n \min(p_{b,i}, p_{u,i}) \quad (3)$$

where $p_{b,i}$ and $p_{u,i}$ are the normalized burned and unburned probabilities in bin i . The Histogram Separation Index was then calculated as Equation (4):

$$HSI = 1 - OVL \quad (4)$$

where values closer to 1 indicate stronger class separation and values closer to 0 indicate greater overlap. We also calculated the M-statistic using Equation (5):

$$M = \frac{|\mu_b - \mu_u|}{\sigma_b + \sigma_u} \quad (5)$$

where μ_b and μ_u are the class means and σ_b and σ_u are the corresponding standard deviations for burned and unburned samples.

3.6. Multi-dimensional feature engineering and self-supervised refinement

Because spectral thresholding alone remains vulnerable to ambiguity in heterogeneous landscapes, BAM refines the pseudo-labels using a supervised model trained directly on the target scene. This scene-specific refinement step learns the spatial and structural context of fire, enabling rejection of false positives such as cloud shadow, dark soil, and other spectrally similar surfaces.

3.7. High-dimensional feature space construction

BAM utilizes a 15-variable feature stack integrating multi-modal data to capture physical, structural, and topographic surface characteristics (details provided in Supplementary file as Table S3). The refinement feature stack combines post-fire reflectance, bi-temporal spectral change variables, GLCM texture metrics derived from post-fire NIR, and terrain variables from FABDEM. These predictors capture spectral, structural, and topographic information that helps distinguish true burn scars from confounding surfaces. To prevent data leakage, dATBI is excluded from the refinement feature space, forcing the model to derive decision boundaries from independent evidence.

The feature order was fixed across all events to ensure reproducibility. Spectral predictors were placed first because they represent the primary physical burn signal. Temporal change predictors were included next to isolate fire-induced disturbance from static background reflectance. Texture predictors were derived from post-fire NIR because burned surfaces often exhibit spatial fragmentation and perimeter roughness that cannot be captured by pixel-level spectra alone. Terrain variables were included last to account for topographic illumination, shadowing, and fire-spread context. The dATBI layer was not included as a GTB predictor because it was already used to generate the pseudo-labels; excluding it reduces leakage between label generation and model refinement.

3.8. Gradient tree boost classification

GTB model was selected because it satisfies three requirements of the BAM framework: non-linear discrimination, robustness to weak pseudo-labels, and operational compatibility with Google Earth Engine (GEE). Tree boosting is suitable for this setting because it sequentially fits weak learners to residual errors, allowing the model to capture non-linear interactions among spectral, temporal, texture, and terrain predictors (Friedman, 2001; Hastie et al., 2001). Compared with a single CART model, boosting reduces instability; compared with kernel-based models such as SVM, it scales more efficiently within the GEE environment; and compared with deep learning, it does not require large manually labeled training datasets or external GPU-based inference (Chen and Guestrin, 2016; Hastie et al., 2001).

Additionally, we benchmarked GTB against the available in-platform

classifiers used in this study, including Random Forest, CART, SVM, Naive Bayes, and k-NN (Supplementary Fig. S1). GTB produced the most consistent performance across the 15 events and was therefore selected for the final BAM implementation. XGBoost and LightGBM were not used because the objective was a fully server-side GEE implementation without exporting features to an external machine-learning environment. Deep learning was not adopted because the framework is designed for scene-specific self-supervised learning with limited pseudo-labels, not for supervised training on a large manually labeled burn database.

For each wildfire event, BAM was trained as a scene-specific classifier using 10,000 stratified pseudo-labeled pixels, consisting of 5,000 burned and 5,000 unburned samples. The 10,000-sample size was selected to balance class representation, spatial diversity, and computational feasibility in Google Earth Engine. Sensitivity testing showed that smaller sample sizes produced less stable decision boundaries in fragmented burns, while increasing the sample size beyond 10,000 produced limited performance gains but increased runtime. These labels were not obtained from external reference products or manual digitization. Instead, they were generated automatically from the conservative dATBI-Otsu pseudo-labeling procedure described in Section 2.4.2. Thus, the GTB model learns from physics-derived pseudo-labels generated within the target scene, which is the basis of the self-supervised design. Across the 15 wildfire events, this produced 150,000 pseudo-labeled training candidates before spatially blocked internal assessment.

The GTB hyperparameters were selected using a fixed global sensitivity analysis rather than event-specific tuning, because the objective of BAM is operational transferability without per-fire calibration. We tested alternative values for the number of trees, learning rate, maximum leaf nodes, training sample size, and thresholding parameters across the event set (Table S3). The final configuration of 100 trees, learning rate of 0.1, maximum of 10 leaf nodes, and least-absolute-deviation loss was selected because it provided stable F1-score and Intersection over Union (IoU) while limiting overfitting to weak pseudo-labels and maintaining feasible runtime in Google Earth Engine. Increasing model complexity beyond this configuration produced negligible accuracy gains but increased computation time. Lastly, BAM produces two outputs: a continuous burn-probability layer and a binary burned-area map derived from that probability surface.

3.9. Validation and benchmarking

Model performance is evaluated by using fire-aware spatially stratified cross-validation to reduce spatial leakage. Each scene is partitioned into 3×3 km blocks, which are then assigned to three strata based on burn density: core fire (>20%), edge/transition (1–20%), and background (<1%). Ten-fold cross-validation samples proportionally from these strata so that each fold contains a representative mixture of interior burns, complex perimeters, and unburned background. Each fold must contain at least 50 total samples, including a minimum of 10 burned and 10 unburned samples. This out-of-fold internal assessment assesses spatial generalization within each event; external validation is addressed separately using PlanetScope imagery.

We distinguish between internal assessment, external validation, and operational benchmarking. Internal assessment uses held-out spatial blocks derived from the pseudo-label set; therefore, it evaluates spatial generalization and pseudo-label consistency within each event but is not treated as fully independent ground truth validation. Independent external validation was conducted using 3 m PlanetScope imagery over the 2024 Park Fire. MODIS MCD64A1 was not used as a ground-truth reference because its 500 m resolution is not directly comparable with BAM's 30 m output. Instead, MCD64A1 was used only as a coarse-resolution operational benchmark to evaluate differences in spatial detail, fragmented burns, and unburned refugia. More details are provided in Supplementary file as Table S5.

3.10. Post-processing and implementation

Raw GTB outputs are regularized using 3×3 morphological opening and closing followed by a connected-component sieve with an 8-pixel minimum mapping unit (0.72 ha). BAM is implemented in Google Earth Engine for scalable NRT processing. Platform access details are provided in the Data and Code Availability section. The operational implementation of BAM in Google Earth Engine is illustrated in Fig. 3, which shows the post-SREM input, ATBI signal, pseudo-label mask, final burned-area output, and web interface used for visualization and export.

4. Results

4.1. Physical foundation: Spectral separability and feature space dynamics

The success of BAM depends on whether dATBI can generate pseudo-labels that are sufficiently separable for automated thresholding. Fig. 4 shows that dATBI expands the burned class more clearly than dNBR in the bi-spectral feature space, especially in Mediterranean and subtropical examples where class overlap is otherwise substantial. In the representative Kernel Density Estimation (KDE) comparison, dATBI produces a clearer bimodal distribution, with the Otsu threshold falling in a pronounced low-density valley between burned and unburned modes. This property is important because pseudo-label quality, rather than raw index magnitude alone, determines the stability of the

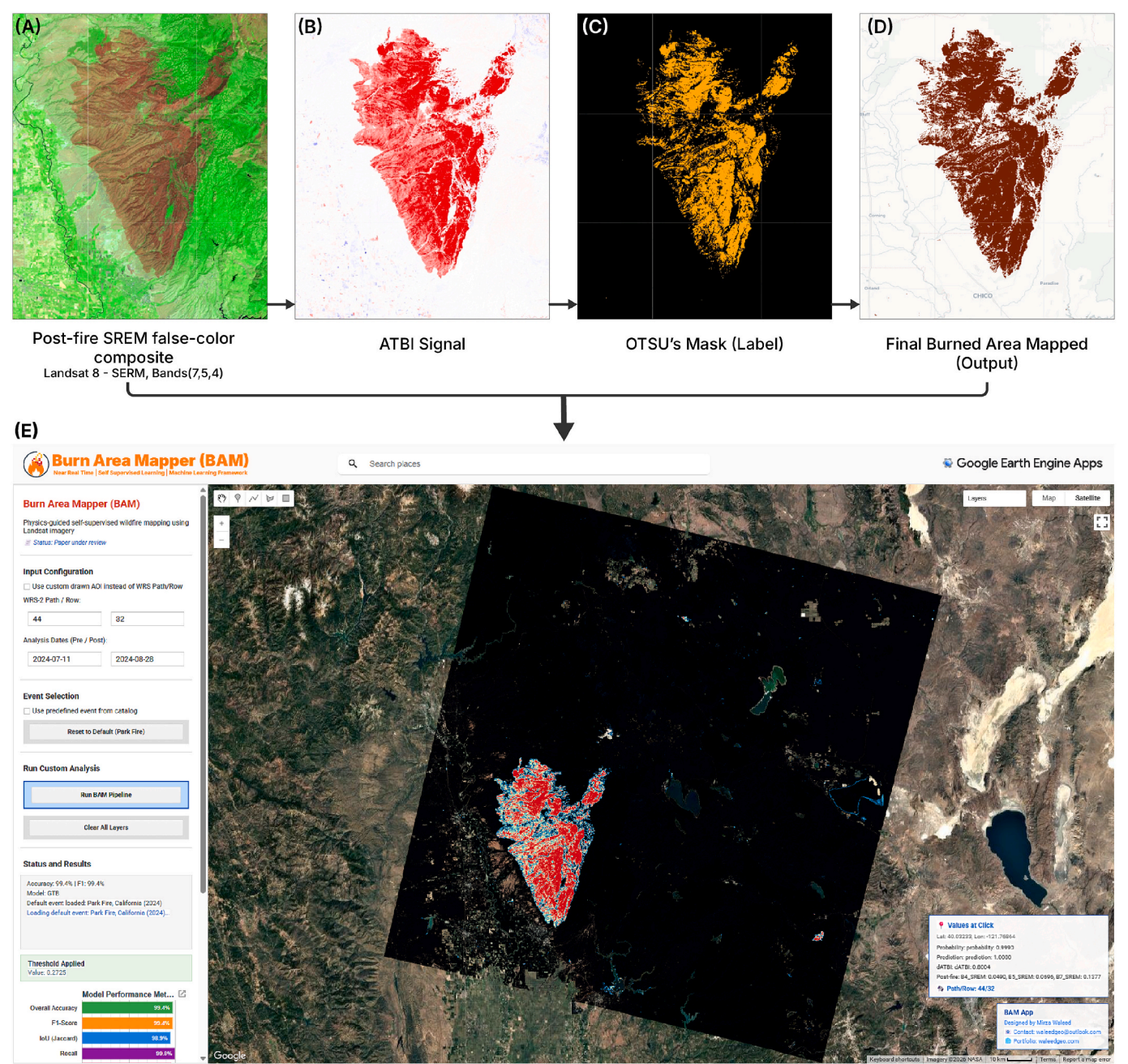


Fig. 3. Operational BAM implementation in Google Earth Engine. The top row (A to D) shows post-SREM input, ATBI signal, pseudo-labels, and final BAM output; the bottom panel (E) shows the BAM web interface used for near-real-time visualization and export.

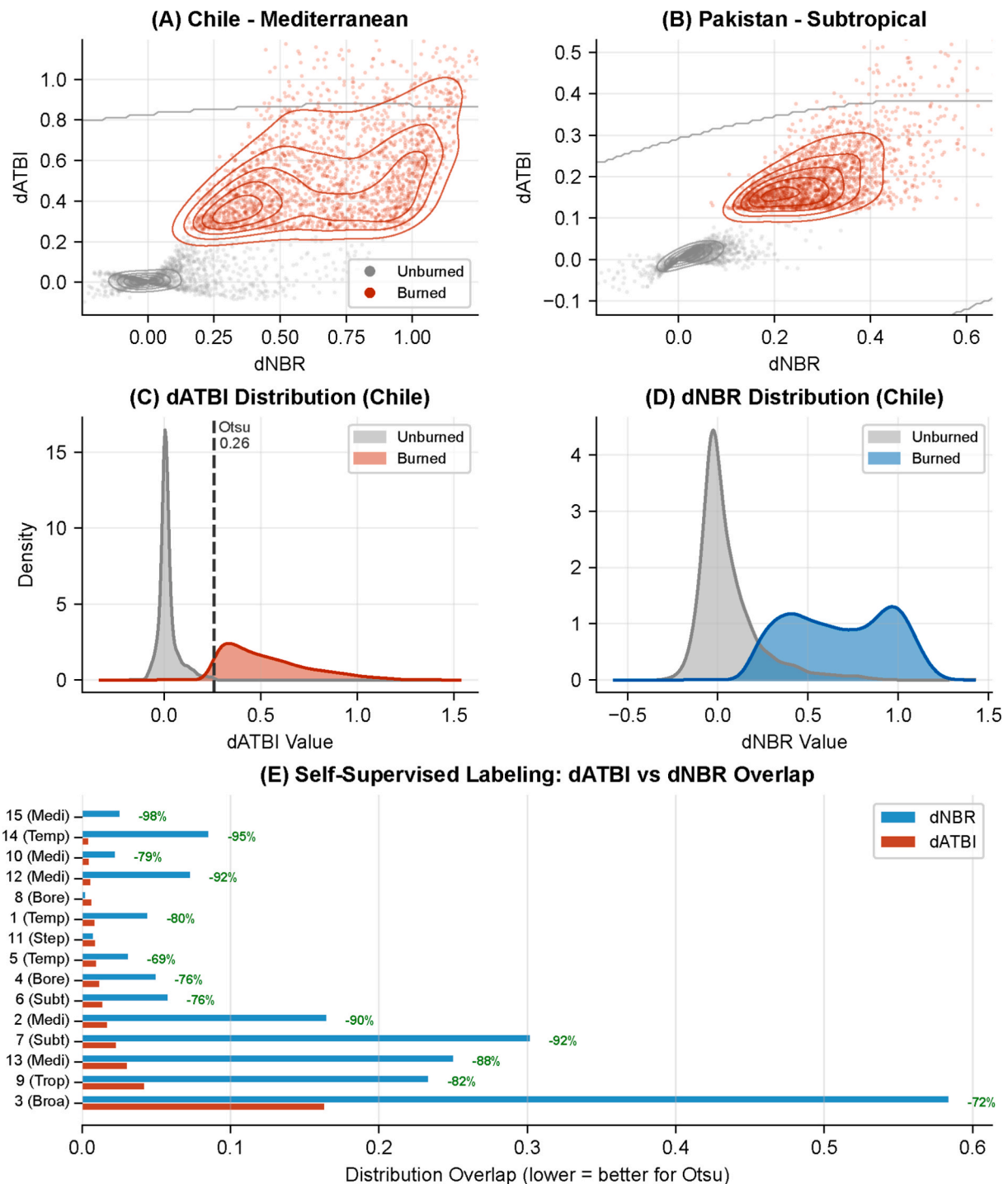


Fig. 4. Comparison of burned-unburned separability for dATBI and dNBR. (A–B) Representative bivariate density plots. (C–D) Kernel Density Estimation (KDE) distributions showing cleaner thresholding for dATBI. (E) Distribution-overlap summary across 15 events, where lower values indicate better separability.

subsequent self-supervised refinement.

To quantify whether ATBI improved separability relative to conventional NBR-based change detection, we compared dATBI with dNBR using the Histogram Separation Index and M-statistic (Table S6). Across the 15-event benchmark, dATBI produced stronger burned-unburned separation than dNBR in 13 events, with a mean separation index of 0.99 compared with 0.85 for dNBR. The improvement was most pronounced in spectrally complex cases such as Willi Willi and India–Nepal, where dNBR exhibited stronger overlap between burned and unburned distributions. The improved separability of dATBI is consistent with previous

burned-area studies showing that fire-induced spectral change is primarily governed by NIR reduction and SWIR enhancement (Farhadi et al., 2022). However, unlike conventional dNBR-based approaches, ATBI explicitly multiplies the normalized NIR–SWIR2 contrast by the SWIR2/NIR ratio, which increases the distance between burned and unburned distributions when combustion produces simultaneous vegetation structure loss and surface drying. This explains why the largest gains occurred in spectrally complex events, where conventional normalized differences are more likely to suffer from class overlap (Bilal et al., 2019).

4.2. Self-supervised denoising and biogeographical generalization

The efficacy of the BAM framework relies on the capacity of the self-supervision framework to transform physics-constrained spectral anomalies into spatially coherent burn probabilities. Fig. 5 shows this refinement process across four distinct pyromes, tracking the progression from raw spectral inputs to calibrated probabilistic outputs. The initial labels, derived via adaptive Otsu thresholding of the dATBI signal, serve as a coarse heuristic that successfully delineates the approximate fire perimeter but remains susceptible to radiometric noise and fragmentation. As evidenced in the 2024 Park Fire and 2016 Central Chile panels, the binary labels exhibit characteristic salt-and-pepper artifacts in transition zones where spectral values oscillate near the global threshold. The subsequent integration of the GTB classifier functions as a spatial filter, effectively decoupling the structural topology of the burn scar from these stochastic artifacts. Therefore, by ingesting textural and topographic covariates alongside the spectral data, the ML model suppresses false positives in the unburned matrix while reinforcing the confidence of detection within the core burn area. This results in the smooth, high-fidelity probability gradients observed in column 4, confirming that the model learns to identify the morphological invariance of wildfire rather than memorizing the noise inherent in the training labels.

The analysis further substantiates the framework's zero-shot transferability across extreme ecological gradients. The 2024 Chiang Mai event represents a significant adversarial test case, characterized by a complex tropical deciduous landscape where rapid vegetative regrowth and steep topography typically attenuate the post-fire spectral signal. Despite the reduced magnitude of the dATBI anomaly relative to the temperate examples, the adaptive nature of the local thresholding mechanism, combined with the classifier's non-linear decision

boundaries, successfully recovers the fragmented burn patterns. Similarly, at the 2022 Quebec event, the framework distinguishes between the spectral signature of deep boreal burn scars and that of optically similar dark water bodies typical of the Canadian Shield. The consistent performance across these disparate biomes indicates that the physics-guided feature space captures the fundamental spectral-structural properties of fire disturbance, enabling robust mapping without the need for region-specific parameter tuning or supervised retraining.

4.3. Probabilistic calibration and model reliability

To validate that the self-supervised learning GTB model does not merely overfit to the binary automated labels but learns a robust probabilistic representation of fire occurrence, we evaluated the model's calibration using reliability diagrams and sharpness histograms (Fig. 6). The global reliability analysis (Panel A) yields a Brier Score of 0.0017, a value significantly lower than the reference baseline for random guessing ($BS = 0.25$), confirming that the calibrated probabilities closely approximate the true empirical burn frequency. The reliability curve exhibits a characteristic S-shaped deviation from the diagonal, indicating that the model is sharp; it tends to predict probabilities close to 0 or 1 rather than hedging in the uncertain middle range. This behavior is operationally desirable for automated mapping, as it reflects high decisiveness in classification decisions while maintaining strong calibration at the extremes where the vast majority of pixels reside.

The sharpness of the classifier is further quantified by the probability distribution histogram (Panel B), which displays a pronounced U-shaped profile. Across the 150,000 stratified samples (from 15 events), 99.3% of predictions fall into high-confidence categories ($\$P < 0.2\$$ or $\$P > 0.8\$$), leaving only 0.7% of pixels in the intermediate transition zone.

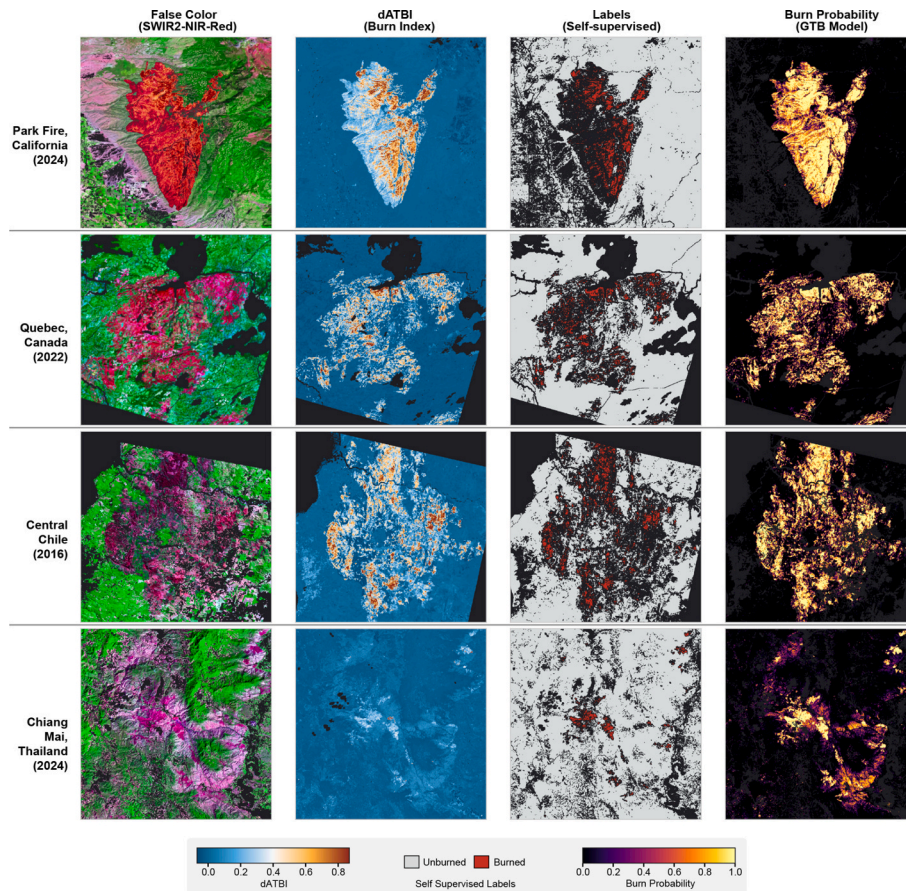


Fig. 5. Examples of BAM refinement across four biomes. Columns show post-fire composite, dATBI, pseudo-labels, and final GTB burn probability, illustrating conversion of coarse spectral labels into spatially coherent probabilities.

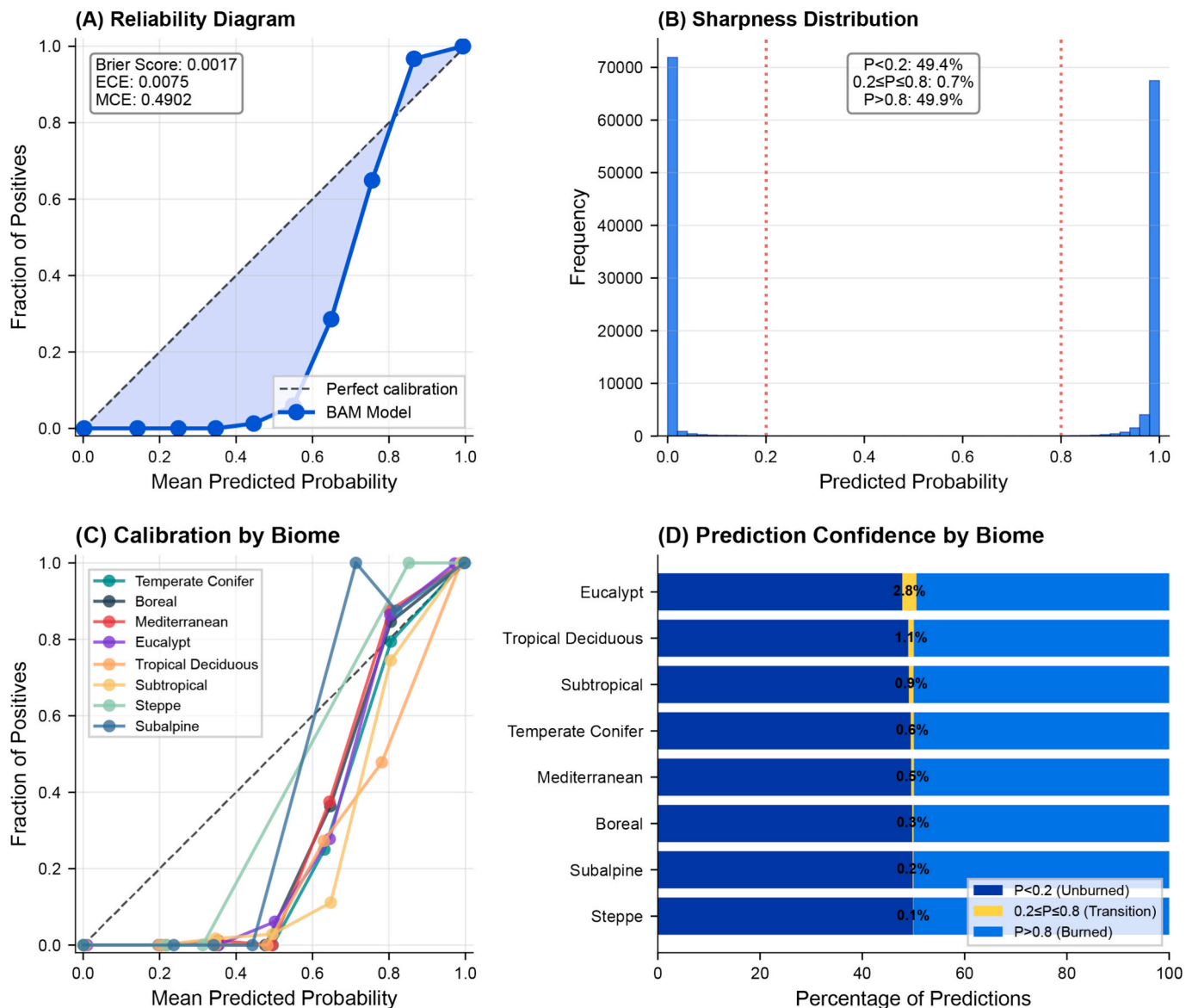


Fig. 6. Probability calibration and model sharpness across 15 global events (sample size = 150,000). (A) Global reliability diagram with Brier Score = 0.0017 and Expected Calibration Error (ECE) = 0.0075. (B) Sharpness distribution showing 99.3% high-confidence predictions ($P < 0.2$ or $P > 0.8$). (C) Per-biome calibration curves demonstrating transferability across eight vegetation types. (D) Confidence breakdown by biome illustrating that transition-zone uncertainty remains below 2.8%.

Spatially, these intermediate probabilities are strictly confined to the perimeter of burn scars, accurately reflecting the sub-pixel heterogeneity of mixed edge pixels rather than global model uncertainty. Critically, this calibration stability is maintained across disparate ecological regimes. The per-biome calibration curves (Panel C) demonstrate that the functional relationship between predicted probability and true burn rate remains consistent across all eight tested vegetation types, from the sparse canopy of the Temperate Steppe to the dense biomass of the Boreal Forest. The analysis of transition zone variance (Panel D) reveals that while complex canopy structures like Eucalypt forests exhibit slightly higher uncertainty (2.8% intermediate pixels) due to mixed spectral signals, the model retains extremely high confidence ($>99\%$) in simpler biomes such as Steppe and Mediterranean Scrub. This cross-biome consistency confirms that the self-supervised mechanism successfully normalizes the output probability space, ensuring that a $P = 0.9$ prediction reliably indicates burn detection regardless of the underlying landscape.

4.4. Global accuracy and biome-agnostic transferability

The operational viability of a global wildfire mapping framework is contingent upon its stability across heterogeneous ecological gradients. A critical limitation of many regional algorithms is their tendency to overfit to specific vegetation types or atmospheric conditions, resulting in significant performance degradation when applied to unseen biomes. To evaluate the zero-shot transferability of the BAM framework, we conducted a rigorous cross-validation assessment across 15 distinct wildfire events spanning varying regimes of biomass, moisture content, and topographic complexity. Fig. 7 presents the comprehensive performance matrix, quantifying the agreement between the self-supervised GTB model predictions and the out-of-fold internal assessment subsets (prepared through spatial stratified sampling, as discussed in Section 2.6. "Fire-Aware Spatially Stratified Validation").

The heatmap visualization in Panel A reveals a high degree of uniformity in model performance, with most event-metric combinations yielding scores exceeding 0.99. The global mean F1-score of 0.9940,

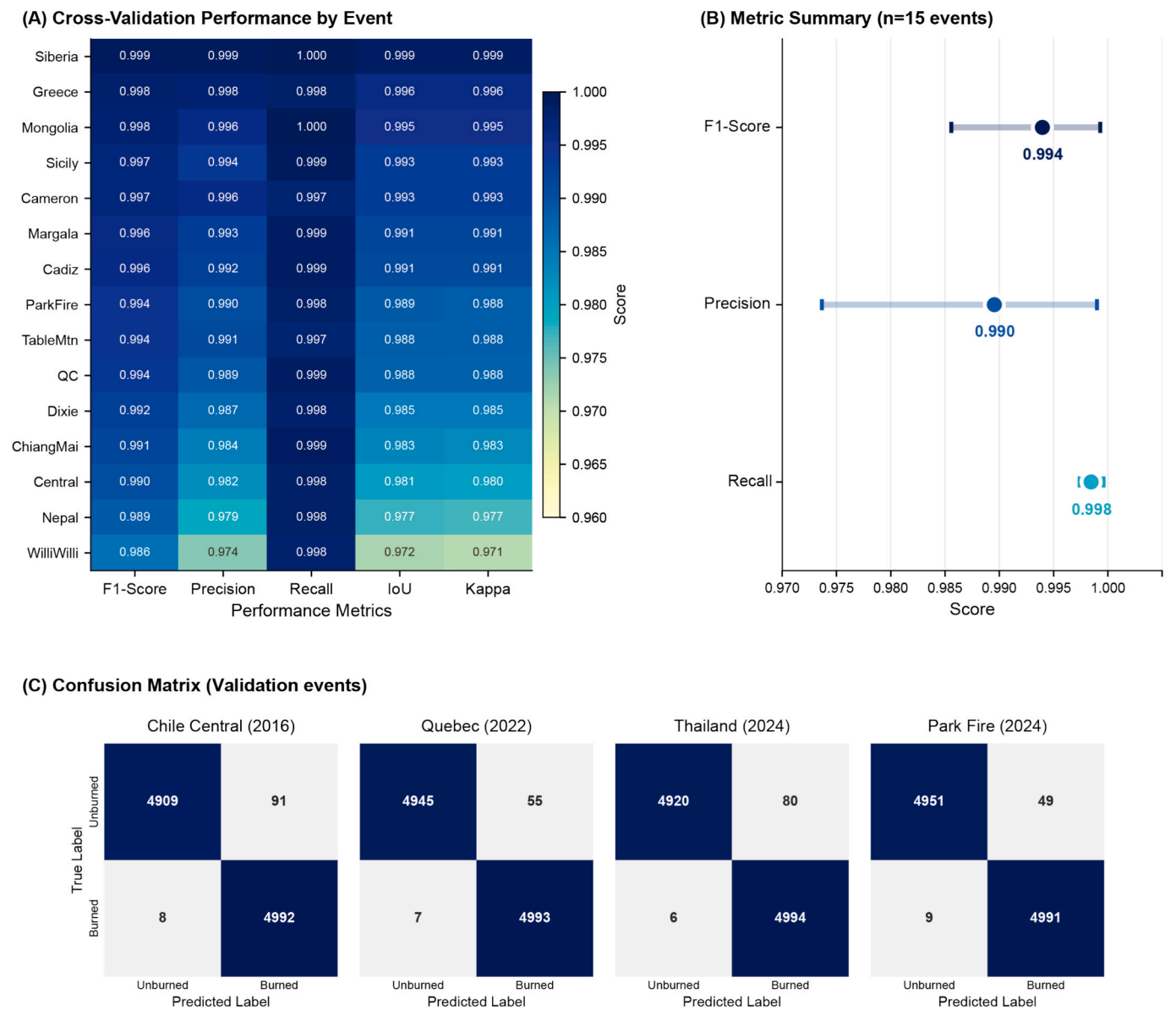


Fig. 7. Internal performance from spatially stratified cross-validation. Event-level metrics (A), distribution summaries (B), and representative confusion matrices (C) show consistently strong performance and low omission error.

with a tight range between 0.9856 and 0.9993, indicates that the framework maintains operational fidelity regardless of the underlying landscape. Notably, the model achieves high internal agreement under spatially blocked assessment in high-latitude boreal environments, such as the 2024 Siberia event, with an F1-score of 0.999. This suggests that the inclusion of texture and topographic features successfully mitigates the common spectral confusion between dark lichen, permafrost, and burn scars that often confound traditional indices in these regions.

The distribution of performance metrics, summarized in Panel B, further clarifies the classifier's behavioral characteristics. Across the 15 events, the internal spatially blocked assessment produced an F1-score of 0.994 ± 0.004 (mean $\pm$ standard deviation), with a range of 0.986–0.999. Precision was 0.990 ± 0.007 (0.974–0.999), and recall was 0.998 ± 0.001 (0.997–1.000). IoU was 0.988 ± 0.007 , and kappa was 0.988 ± 0.008 . These uncertainty summaries show that the high mean performance was not driven by a small number of easy events; however, the values should be interpreted as internal spatially blocked agreement against pseudo-labels rather than fully independent external accuracy.

To further examine the pixel-level classification structure, we analyzed the confusion matrices for four representative validation events shown in Panel C. These matrices reveal a consistent pattern in which True Positives and True Negatives dominate the classification structure, accounting for over 99% of the 10,000 stratified samples for each event. A critical operational insight is the systematic minimization of False Negatives, which range from only 6 pixels in the Tropical Dry Deciduous biome of Chiang Mai to 9 pixels in the Chaparral of the Park Fire. This confirms the model's capacity to capture the complete fire perimeter with negligible omission. While False Positives are slightly more frequent, ranging from 49 to 91 pixels, they remain operationally low and are primarily attributed to sub-pixel mixing at the burn edges. The ability of the model to maintain this high-fidelity separation equally well in the dense boreal forests of Quebec and the sparse sclerophyll shrublands of Central Chile confirms that the learned decision boundaries are driven by the physical invariant of the burn signal rather than biome-specific artifacts.

4.5. Qualitative assessment of signal refinement and noise suppression

To evaluate the refinement of self-supervised learning, we compared raw spectral inputs with the BAM probabilistic outputs for the 2024 Park Fire (Fig. 8). While the False Color Composite (Panel A) reveals charred biomass, the raw reflectance is contaminated by smoke and topographic shadows that spectrally mimic burn signals in rugged terrain. Standard indices like the NBR (Panel B) capture the general fire extent but exhibit significant salt-and-pepper noise in unburned areas due to sub-pixel variability in soil moisture and canopy shadow. Such spectral confusion typically leads to fragmented perimeters and thresholding ambiguity.

In contrast, the BAM probability surface (Panel C) demonstrates high spatial coherence. By integrating textural (GLCM) and topographic covariates, the GTB classifier suppresses stochastic noise and clarifies signals where NBR values are ambiguous ($0.1 < \text{NBR} < 0.3$). This distillation process produces a sharp decision boundary that decouples the burn topology from radiometric interference. The final binary mask (Panel D) enforces physical validity, rejecting false positives in urban or water features while preserving unburned refugia essential for post-fire regeneration. These visual results confirm BAM's superior balance of sensitivity and specificity compared to traditional index-based methods.

4.6. Comparative benchmarking and resolution enhancement

To position the BAM framework within the operational landscape, we benchmarked its performance against the industry-standard MODIS MCD64A1 burned area product. While MCD64A1 serves as a critical baseline for global carbon emission estimates, its 500 m resolution inherently limits its capacity to resolve sub-pixel heterogeneity. Fig. 9 illustrates this resolution gap across four distinct fire events, revealing that discrepancies between the two products are primarily driven by the superior spatial granularity of the 30 m Landsat-based assessment. The comparative analysis relies on a spatial agreement map where consensus pixels appear in orange, while divergences are classified as either resolution gain (blue), where BAM detects burns missed by MODIS, or omission (red), where MODIS aggregates unburned islands into the fire perimeter.

The most significant divergence is observed in fragmented fire regimes where burn scars do not form contiguous monotypic blocks. In the 2024 Park Fire (California) and Chiang Mai (Thailand) events, the framework demonstrates a substantial resolution gain of 93% and 68% respectively. As shown in the agreement maps, the MODIS product frequently fails to detect the fine-scale fingers of the fire front and small, isolated spot fires that occur in complex topography. In the Chiang Mai event, which occurred in a tropical dry deciduous forest characterized by steep terrain, the coarse sensor footprint of MODIS could not resolve the scattered burn patches that define the fire regime in this biome.

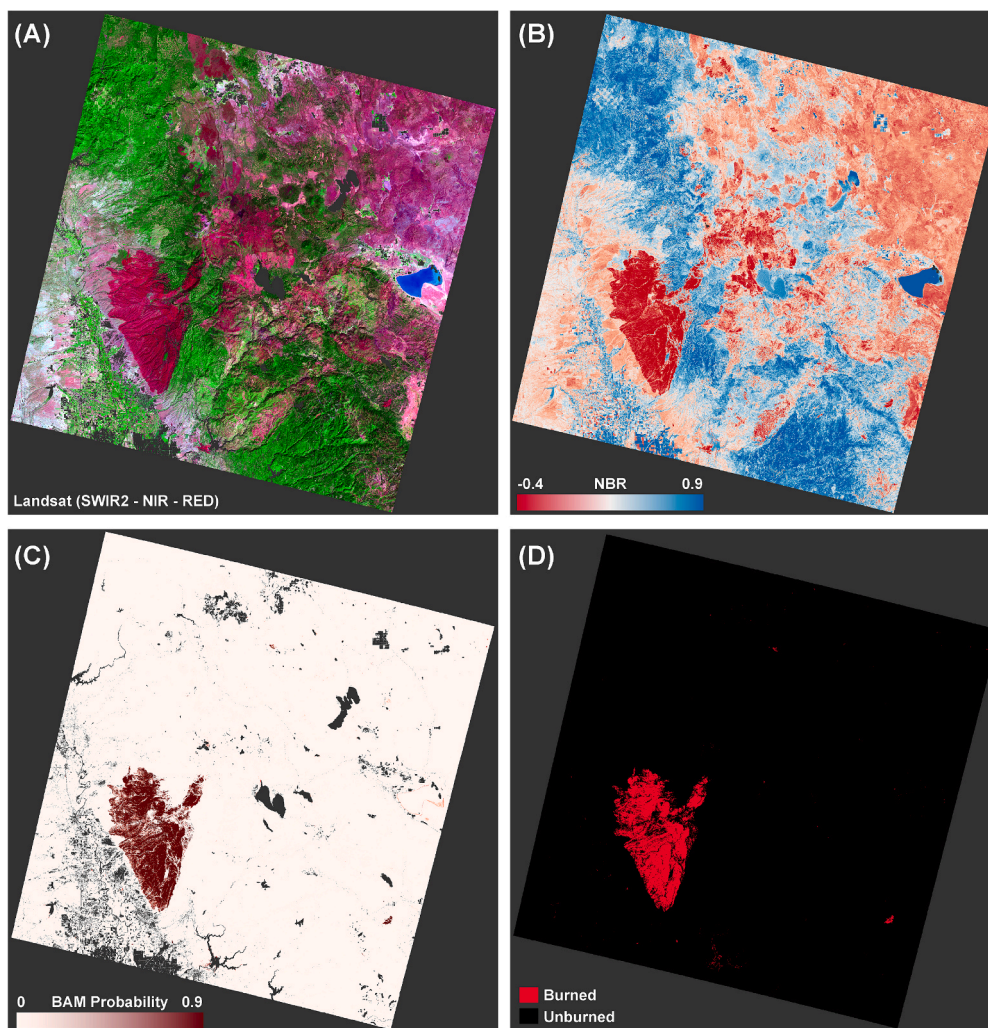


Fig. 8. Spectral-to-probabilistic workflow for the 2024 Park Fire, CA. (A) FCC (SWIR2-NIR-R) spectral input; (B) dNBR index exhibiting background noise and poor signal separation; (C) BAM probability surface ($P > 0.9$ core; $P < 0.1$ background) illustrating ML-based denoising; and (D) final BAM binary mask.

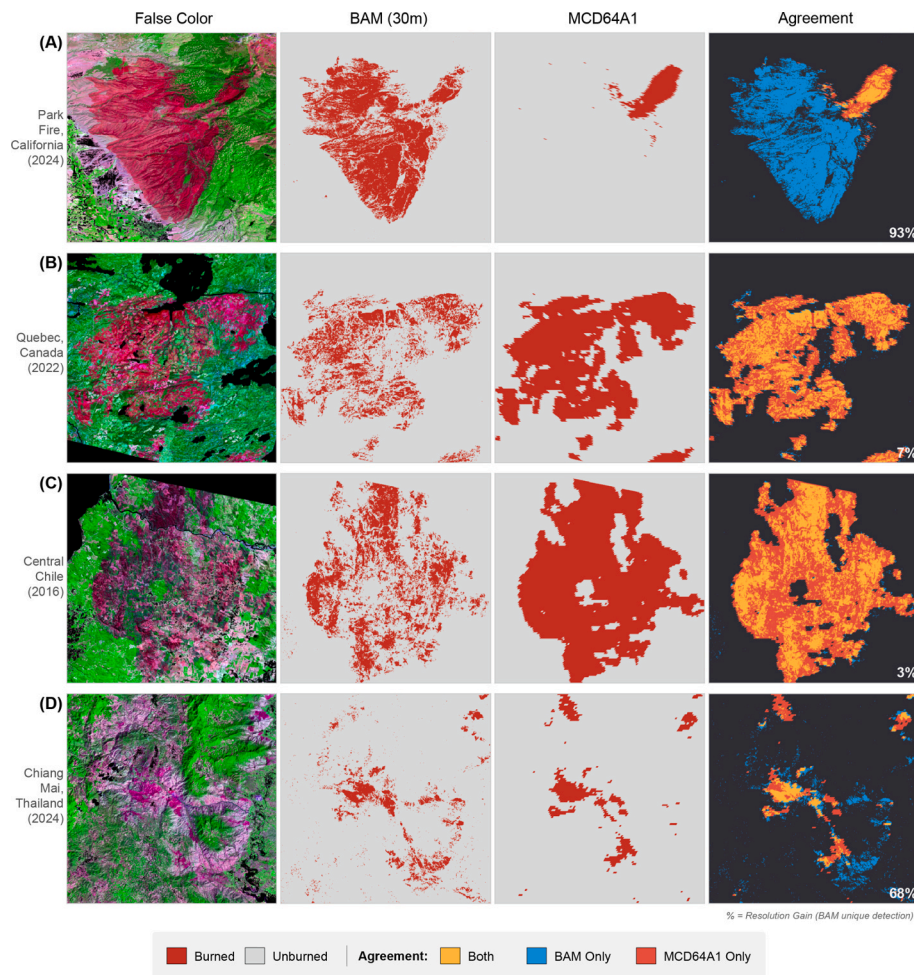


Fig. 9. Comparison of BAM and MODIS MCD64A1 across representative fires. BAM shows the greatest gains in fragmented burns and better delineates unburned islands within contiguous burn scars.

Consequently, BAM recovered a significant area of burned biomass that was effectively invisible to the operational baseline.

Conversely, in the massive contiguous burns of the 2022 Quebec (Boreal) and 2016 Central Chile (Mediterranean) events, the two products show stronger agreement on the core fire extent, with resolution gains of 7% and 3%, respectively. However, a closer inspection of the spatial topology reveals a consistent over-aggregation artifact in the MODIS product. The red regions in the agreement maps indicate areas where the 500 m pixels subsume unburned refugia and water bodies into the burned class. The BAM framework, operating at 30 m, correctly delineates these unburned islands, providing a more ecologically precise estimate of net vegetation loss. This ability to resolve the fragmentation of the burn scar is not merely an aesthetic improvement but a critical functional enhancement for post-fire recovery planning, as it accurately identifies surviving vegetation patches that serve as seed sources for regeneration.

4.7. High-resolution validation and sub-pixel fidelity assessment

To rigorously validate the framework's capacity to resolve fine-scale fire heterogeneity beyond the limits of moderate-resolution baselines, we benchmarked BAM against high-resolution commercial PlanetScope imagery acquired over the 2024 Park Fire event. Given the absence of validated agency perimeters for this recent event, a supervised proxy ground truth was generated using 3 m multispectral PlanetScope data. We developed a custom XGBoost classifier trained on a feature stack comprising native spectral bands (RGB, NIR) and derived indices (NDVI,

NDWI). The model, trained on 10,000 stratified samples derived from manual expert person interpretation, achieved exceptional performance metrics, yielding an overall accuracy of 94.95%, an F1-score of 0.95, and a ROC AUC of 0.99, thereby establishing a reliable high-granularity baseline for validation.

Fig. 10 presents a comparative assessment between this 3 m baseline and the 30 m BAM output. Despite the tenfold difference in spatial resolution, the self-supervised BAM framework demonstrates remarkable topological consistency with the commercial standard. As evidenced in the magnified insets, BAM successfully preserves the structural integrity of the fire perimeter, correctly identifying narrow, dendritic fire fingers and small unburned refugia that are typically homogenized by coarser sensors. While the 30-meter output inherently aggregates the high-frequency “salt-and-pepper” noise observed in the 3 m classification (Panel B), the BAM probability gradients (Panel C) closely mirror the spatial density of the burn, suggesting that the model's probabilistic output effectively functions as a calibrated proxy for sub-pixel burn severity.

To move beyond visual assessment, we quantified the spatial agreement between the BAM predictions and the high-resolution ground truth using standard confusion matrix metrics. The spatial error map (Fig. 11, Panel A) reveals that the vast majority of disagreements are concentrated at the burn scar perimeter, a pattern consistent with the mixed-pixel effect where the 30 m Landsat footprint aggregates sub-pixel spectral variation. Despite this resolution constraint, the framework achieved a high Precision of 0.952, indicating that commission errors (false positives) are negligible and that the model rarely

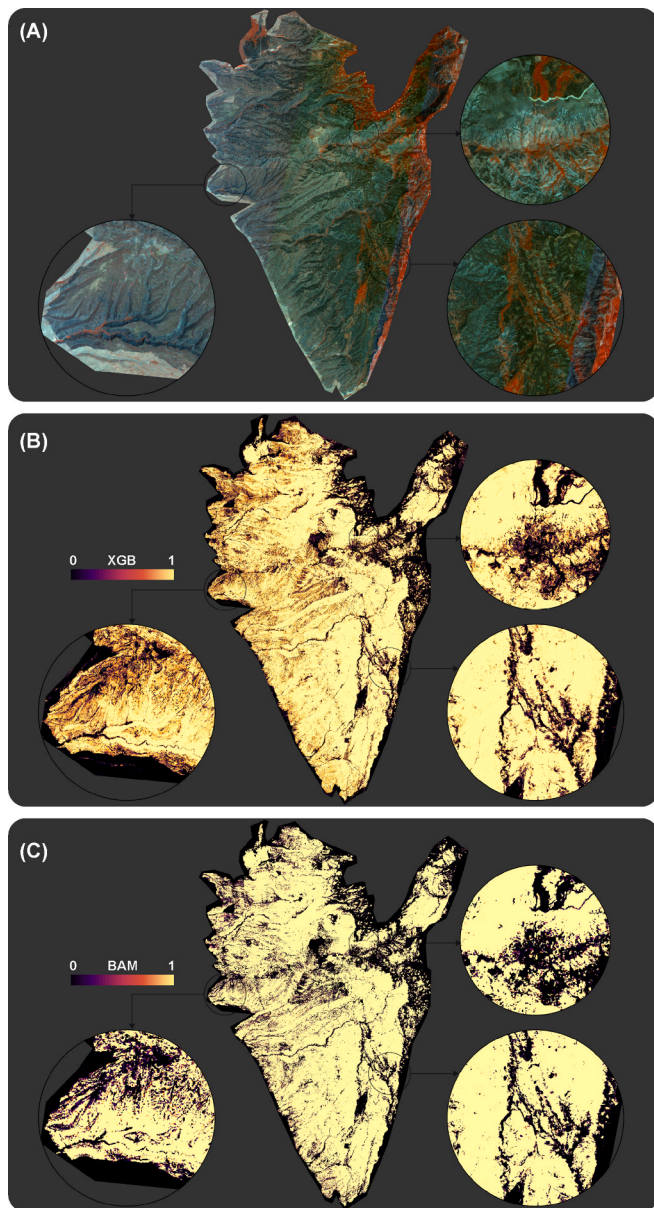


Fig. 10. High-resolution validation over the 2024 Park Fire. (A) PlanetScope 3 m false-color composite (NIR–Green–Blue). (B) PlanetScope-derived 3 m supervised reference map generated with XGBoost (Accuracy = 94.95%, F1 = 0.95). (C) BAM burn-probability output at 30 m. Insets highlight strong agreement in major fire morphology, including narrow burn fronts and internal unburned islands, despite the resolution difference.

hallucinates fire in unburned areas. The Recall of 0.832 highlights a moderate omission rate, primarily attributable to the inability of the 30 m sensor to resolve the microscopic, low-severity surface burns detected by the 3 m PlanetScope baseline. Overall, the model maintained a robust F1-score of 0.888 and an Intersection over Union (IoU) of 0.799, confirming that the core burn extent is captured with high fidelity.

Recognizing that direct pixel-to-pixel comparisons at differing resolutions are often confounded by geometric registration noise, we performed a multi-scale correlation analysis to isolate the true predictive signal. Fig. 11 (Panel B) illustrates the coefficient of determination (R^2) as a function of spatial aggregation. While the pixel-level correlation is moderate ($R^2 = 0.578$), the agreement improves dramatically with minimal aggregation, reaching the standard validation threshold of $R^2 = 0.80$ at a resolution of just 90 m (3 by 3 pixels). The curve asymptotes near $R^2 = 0.92$ at coarser scales, demonstrating that the residual

variance is driven by local spatial noise rather than systematic model bias. This structural agreement is further validated by the density scatter plot at 90 m resolution (Fig. 11, Panel C), which exhibits a strong linear relationship (Pearson's $r = 0.892$). The regression slope of 0.96 is near-unity, confirming that the BAM probability scores are well-calibrated and effectively serve as a linear proxy for sub-pixel burn fraction, rather than merely acting as a binary classifier.

4.8. Physical drivers of classification: Feature importance hierarchy

To elucidate the decision-making logic of the self-supervised model, we analyzed the Gini importance of the input features across the global validation set. Fig. 12 visualizes this importance score through biome-specific polar plots, which confirm that the classification is primarily driven by physical changes in surface reflectance rather than scene-specific artifacts. Across all events, the model exhibits a strong reliance on post-fire near Infrared and shortwave infrared bands, which collectively account for the majority of the information gain. This aligns with the fundamental spectral behavior of fire, where the combustion of vegetation results in a drastic reduction in NIR reflectance due to cell structure collapse and a concurrent increase in SWIR reflectance due to charring and drying. The dominance of these post-fire spectral bands confirms that the GTB model successfully learned the biophysical signature of burn scars as the primary discriminant.

The biome-stratified analysis reveals adaptive shifts in feature prioritization that correspond to distinct ecological characteristics. In high-biomass environments such as Tropical and Temperate forests, the model effectively relies almost exclusively on spectral data, particularly the post-fire SWIR signal, to identify the stark contrast between the canopy and the burn scar. However, in the Grassland biome, a notable divergence occurs where topographic variables, specifically slope, emerge as significant predictors. This suggests that in landscapes with subtle spectral changes—where the signal of burned grass can resemble dry soil or topographic shadows, the model utilizes terrain information to differentiate between naturally occurring shadow regions and true burn scars on hillslopes. Similarly, in the Boreal biome, vegetation indices such as changes in NDVI play a more prominent role, likely aiding discrimination between burn scars and the naturally dark, slow-growing vegetation typical of high latitudes. While textural features such as GLCM contrast and entropy contribute a smaller fraction to the overall importance, their consistent presence across biomes suggests they function as critical local filters for rejecting high-frequency noise that spectral data alone cannot resolve.

5. Discussion

5.1. The physics of spectral discrimination and automated label generation

The main contribution of BAM is not the use of another burn index in isolation, but the use of a physically targeted feature to generate reliable pseudo-labels without manual annotation. The multiplicative SWIR/NIR component in ATBI amplifies the simultaneous SWIR increase and NIR collapse associated with combustion, which improves burned–unburned separation relative to additive indices such as dNBR. This matters operationally because self-supervised refinement is only as good as the weak labels that initiate it. By producing a clearer histogram valley for Otsu thresholding, dATBI increases pseudo-label precision and reduces the propagation of noise into the GTB model. The improvement observed in 13 of 15 events indicates that this mechanism is robust across a wide range of fire environments.

5.2. Biogeographical invariance and zero-shot transferability

The cross-biome stability of BAM appears to stem from three design choices. First, pseudo-labels are generated scene-by-scene rather than

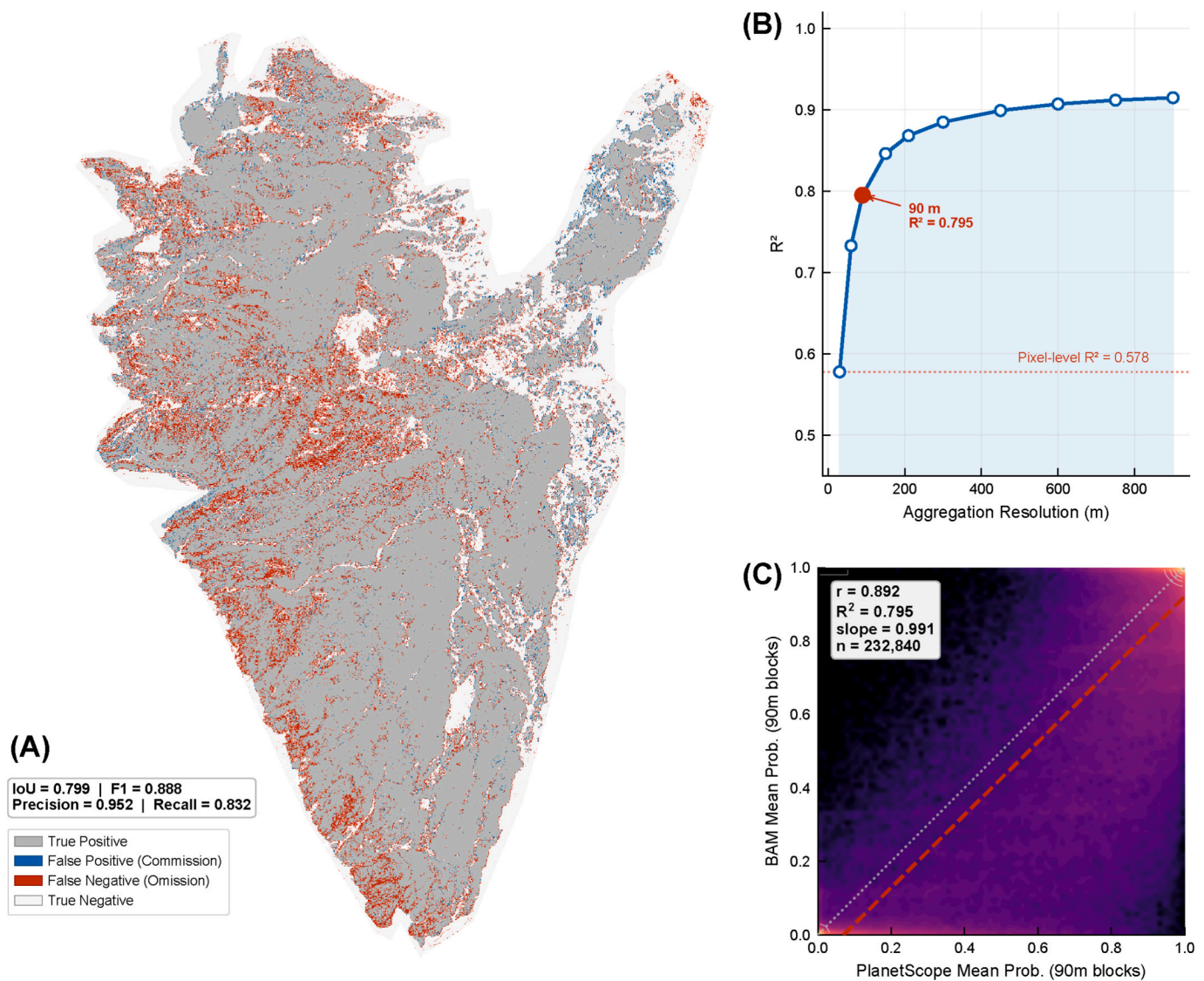


Fig. 11. Statistical validation of BAM 30 m probability against 3 m PlanetScope reference data. (A) Spatial error map showing True Positives (grey), False Positives (blue), and False Negatives (red), with disagreements largely confined to boundary pixels (Precision = 0.952). (B) Multi-scale correlation analysis indicating increasing agreement with spatial aggregation ($R^2 = 0.58$ at 30 m to 0.80 at 90 m), suggesting resolution-driven variability rather than systematic bias. (C) Density scatter plot at 90 m showing strong correlation ($r = 0.892$) and near-unity slope (0.96), consistent with well-calibrated probabilistic output.

imported from a fixed regional training set, thereby reducing domain mismatch. Second, the refinement model integrates complementary spectral, textural, and topographic predictors rather than relying on a single index. Third, the validation strategy enforces spatial separation within each fire, making the reported performance less vulnerable to local autocorrelation. Together, these choices help explain why the framework maintained high internal accuracy across boreal, Mediterranean, tropical, temperate, and grassland environments without regional retraining.

5.3. Operational latency and the role of robust preprocessing

The integration of SREM directly supports the NRT purpose of BAM. Because SREM relies on image-based priors rather than external atmospheric products, the framework can operate on rapidly available Landsat TOA scenes and remain usable under smoky conditions that would otherwise reduce data availability. The India–Nepal case illustrates the importance of this preprocessing step for preserving operational continuity during active events. The NRT capability of BAM is therefore methodological rather than simply computational: the

workflow avoids manual reference-label preparation, avoids dependence on delayed surface reflectance products, and uses scene-specific pseudo-label generation directly from newly available Landsat observations.

The present implementation was restricted to Landsat 8/9 to maintain a sensor-consistent ATBI–SREM–GTB workflow. This design uses harmonized OLI/OLI-2 TOA observations, 30 m optical bands, and the Landsat thermal band included in the feature stack. Sentinel-2 offers clear advantages for burned-area mapping, including higher spatial resolution, red-edge information, and shorter revisit frequency; however, its integration would require cross-sensor harmonization, adjustment of the ATBI/SREM preprocessing chain, treatment of mixed Sentinel-2 band resolutions, and redesign of the feature stack because Sentinel-2 does not provide a Landsat-like thermal infrared band. We therefore treat Sentinel-2 integration as an important future extension rather than part of the present Landsat-based BAM version.

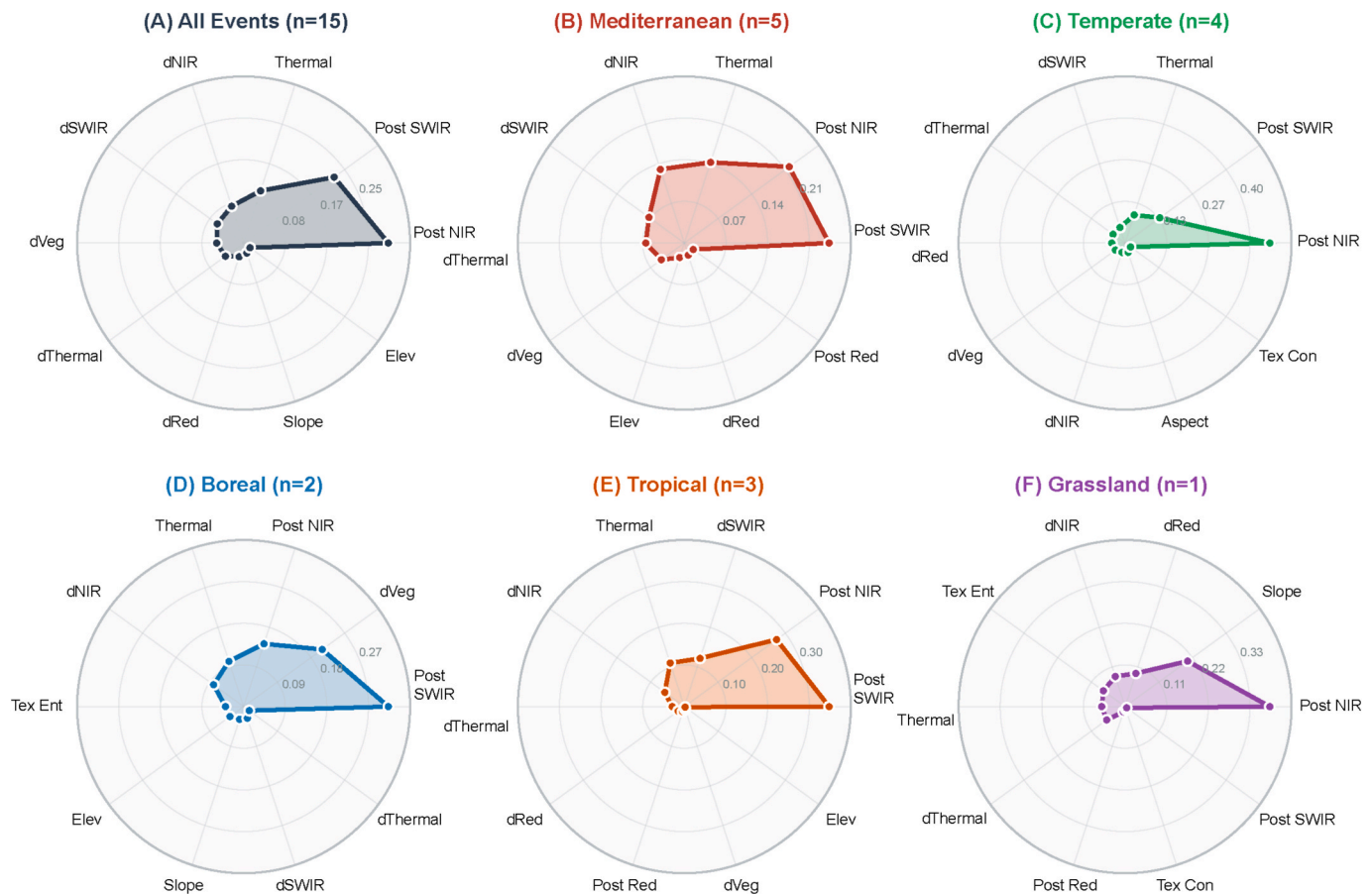


Fig. 12. Biome-stratified feature importance in BAM. Post-fire NIR and SWIR dominate overall, while topographic and vegetation variables gain importance in spectrally subtle environments.

5.4. Resolving sub-pixel heterogeneity: Comparison with MODIS and PlanetScope

The comparison with MODIS and PlanetScope shows that BAM adds value at two complementary scales. Compared with MODIS, the key benefit lies in resolving fragmented burns and internal refugia that coarse global products smooth over or omit. Against PlanetScope, the benefit lies in preserving meaningful fire structure even when the Landsat footprint necessarily aggregates fine-scale variability. The probability layer is particularly useful because it expresses spatial gradation along burn boundaries and correlates with aggregated reference burn fraction, making it more informative than a binary perimeter alone.

Previous burned-area mapping studies have demonstrated the value of spectral indices, Sentinel-2 time series, and machine-learning classifiers for regional fire mapping. Index-based approaches such as NBR, dNBR, Burned Area Index for Sentinel-2 (BAIS2), and Burn Area Detection Index (BADi) provide physically interpretable burn-sensitive features, while supervised machine-learning approaches can improve classification by combining multiple spectral and contextual variables (Bilal, 2025b; Farhadi et al., 2022; Holsinger et al., 2022; Liu et al., 2023; Roodsarabi et al., 2025). BAM differs from these approaches by using a physically constrained index not as the final classifier, but as an automated pseudo-label generator for scene-specific self-supervised learning. This distinction is important because it reduces dependence on manually prepared training data while retaining the interpretability of spectral burn physics. The comparison with MODIS further shows that BAM's main operational advantage is not simply higher total burned area, but improved delineation of fragmented burn scars and internal unburned refugia.

5.5. Limitations and future trajectories

While the BAM framework exhibits robust performance, it remains limited by the physical constraints of optical remote sensing. Persistent cloud cover can still produce temporal gaps, and independent high-resolution validation is currently limited to one event. Future work should therefore prioritize integrating microwave sensing in cloud-affected regions and extend external validation to additional pyromes. The self-supervised label-generation strategy may also support future deep learning models, but this should be framed as a downstream opportunity rather than as a core claim of the present study. Finally, a limitation of the default exclusion mask used for BAM is that cropland fires may be underestimated, particularly where burned croplands are mapped as class 40 in ESA WorldCover or where land-cover classification errors occur. For agricultural fire applications, the cropland mask should be disabled or replaced with a crop-aware configuration, which should then be validated separately.

6. Conclusions

This study introduced the Burn Area Mapper (BAM), a physics-informed self-supervised framework for near-real-time wildfire burned-area mapping from Landsat 8/9 observations. The core contribution of BAM is the use of the Automated Temporal Burn Index (ATBI) as a physically constrained pseudo-label generator, allowing scene-specific model training without manually prepared reference labels or region-specific supervised calibration. By combining ATBI-based pseudo-labeling, SREM atmospheric correction, and Gradient Tree Boost refinement using spectral, temporal, textural, and topographic predictors, BAM produces both binary burned-area maps and

continuous burn-probability outputs at 30 m resolution.

The results show that dATBI improves burned–unburned separability relative to conventional dNBR in most evaluated events, supporting more stable automated thresholding and pseudo-label extraction. Across 15 wildfire events spanning diverse biomes and fire regimes, the internal spatially blocked assessment demonstrated consistently high agreement between BAM predictions and held-out pseudo-labels, indicating that the refinement model generalizes beyond the initial training samples within each scene. Independent evaluation using 3 m PlanetScope data further showed that BAM preserves major burn morphology at 30 m and that its probability output provides meaningful information on sub-pixel burn fraction when aggregated to coarser scales.

Compared with MODIS MCD64A1, BAM provides finer spatial delineation of fragmented burn scars, narrow fire fronts, and internal unburned refugia that are often generalized in coarse-resolution products. These characteristics make BAM particularly useful for rapid post-fire assessment, ecological impact analysis, and local-to-regional disaster response, while complementing existing global burned-area products designed for long-term climate and emissions applications.

BAM remains subject to the limitations of optical remote sensing, including persistent cloud cover, smoke opacity, and data gaps when suitable post-fire observations are unavailable. The default cropland mask also prioritizes conservative wildland-fire mapping and may require adjustment for agricultural-fire applications. Future work should extend independent high-resolution validation across additional pyromes, evaluate cropland-sensitive configurations, and explore Sentinel-2 and SAR integration to improve revisit frequency, spatial detail, and robustness under cloud-affected conditions. Overall, BAM provides a scalable, interpretable, and operationally practical pathway for automated 30 m near-real-time burned-area mapping.

CRedit authorship contribution statement

Mirza Waleed: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Muhammad Bilal:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge USGS and NASA for Landsat data, ESA for WorldCover, and Planet Labs PBC for PlanetScope imagery. We also thank the Deanship of Research at King Fahd University of Petroleum and Minerals for funding support, and Prof. Yizhou Zhuang for constructive suggestions on the study design.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jag.2026.105517>.

Data availability

The code for BAM, including SREM, ATBI, and classification modules, is available at <https://github.com/waleedgeo/BAM>. The operational Google Earth Engine application is available at <https://waleedgis.users.earthengine.app/view/bam>. Landsat 8/9 and FABDEM data are publicly accessible through Google Earth Engine. PlanetScope imagery is commercial and subject to Planet Labs licensing; derived validation

data are available from the corresponding author upon reasonable request.

References

- Alonso-González, E., Fernández-García, V., 2021. MOSEV: a global burn severity database from MODIS (2000–2020). *Earth Syst. Sci. Data* 13, 1925–1938. <https://doi.org/10.5194/essd-13-1925-2021>.
- Artés, T., Oom, D., De Rigo, D., Durrant, T.H., Maianti, P., Libertà, G., San-Miguel-Ayanz, J., 2019. A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Sci. Data* 6, 296. <https://doi.org/10.1038/s41597-019-0312-2>.
- Balakavi, S., Vadrevu, V., Lasko, K., 2025. Mapping burnt areas using very high-resolution imagery and deep learning algorithms - a case study in Bandipur, India. *Plos One* 20, e0327125. <https://doi.org/10.1371/journal.pone.0327125>.
- Bhattarai, H., Val Martin, M., Sitch, S., Yung, D.H., Tai, A.P., 2025. Global wildfire patterns and drivers under climate change. *Biogeosciences* 22, 7591–7610. <https://doi.org/10.5194/bg-22-7591-2025>.
- Bilal, M., 2025a. The automated temporal burn index (ATBI) for accurate and scalable burned area mapping. *Int. J. Appl. Earth Obs. Geoinf.* 144, 104866. <https://doi.org/10.1016/j.jag.2025.104866>.
- Bilal, M., 2025b. Spectral indices alone are not enough: a critical assessment of pre-fire wildfire risk mapping. *Ecol. Inform.* 91, 103435. <https://doi.org/10.1016/j.ecoinf.2025.103435>.
- Bilal, M., Nazeer, M., Nichol, J.E., Bleiweiss, M.P., Qiu, Z., Jäkel, E., Campbell, J.R., Atique, L., Huang, X., Lolli, S., 2019. A simplified and robust surface reflectance estimation method (SREM) for use over diverse land surfaces using multi-sensor data. *Remote Sens.* 11. <https://doi.org/10.3390/rs11111344>.
- Bowman, D.M.J.S., Kolden, C.A., Abatzoglou, J.T., Johnston, F.H., van der Werf, G.R., Flannigan, M., 2020. Vegetation fires in the Anthropocene. *Nat. Rev. Earth Environ.* 1, 500–515. <https://doi.org/10.1038/s43017-020-0085-3>.
- Chen, Tianqi, Carlos, Guestrin, 2016. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM*, pp. 785–794.
- Chen, B., Wu, S., Jin, Y., Song, Y., Wu, C., Venevsky, S., Xu, B., Webster, C., Gong, P., 2024. Wildfire risk for global wildland–urban interface areas. *Nat. Sustain.* 7, 474–484. <https://doi.org/10.1038/s41893-024-01291-0>.
- Costa-Saura, J.M., Bacciu, V., Ribotta, C., Spano, D., Massai, A., Sirca, C., 2022. Predicting and mapping potential fire severity for risk analysis at regional level using google earth engine. *Remote Sens.* 14, 4812. <https://doi.org/10.3390/rs14194812>.
- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udrou, C., Valero, S., Bégue, A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* 221, 551–568. <https://doi.org/10.1016/j.rse.2018.11.007>.
- Farhadi, Hadi, Mehdi Mokhtarzade, Hamid Ebadi, Behnam Asghari, Beirami, 2022. Rapid and automatic burned area detection using sentinel-2 time-series images in google earth engine cloud platform: A case study over the Andika and Behbahan Regions, Iran. *Environ. Monit. Assess* 194 (5), 369. <https://doi.org/10.1007/s10661-022-10045-4>.
- Friedman, Jerome H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *JSTOR*.
- Hastie, Trevor, Jerome Friedman, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York.
- Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J., 2022. A 30 m global map of elevation with forests and buildings removed. *Environ. Res. Lett.* 17, 24016. <https://doi.org/10.1088/1748-9326/ac4d4f>.
- He, K., Shen, X., Merow, C., Nikolopoulos, E., Gallagher, R.V., Yang, F., Anagnostou, E. N., 2024. Improving fire severity prediction in south-eastern Australia using vegetation-specific information. *Nat. Hazards Earth Syst. Sci.* 24, 3337–3355. <https://doi.org/10.5194/nhess-24-3337-2024>.
- Holsinger, L.M., Parks, S.A., Saperstein, L.B., Loehman, R.A., Whitman, E., Barnes, J., Parisien, M.-A., 2022. Improved fire severity mapping in the north American boreal forest using a hybrid composite method. *Remote Sens. Ecol. Conserv.* 8, 222–235. <https://doi.org/10.1002/rse2.238>.
- Liu, P., Liu, Y., Guo, X., Zhao, W., Wu, H., Xu, W., 2023. Burned area detection and mapping using time series Sentinel-2 multispectral images. *Remote Sens. Environ.* 296, 113753. <https://doi.org/10.1016/j.rse.2023.113753>.
- Quintero, N., Viedma, O., Veraverbeke, S., Moreno, J.M., 2025. Optimising fire severity mapping using pixel-based image compositing. *Remote Sens. Environ.* 321, 114687. <https://doi.org/10.1016/j.rse.2025.114687>.
- Roodsarabi, Zohreh, Hadi Farhadi, Hamid Ebadi, Abbas, Kiani, 2025. Machine learning-based burned area detection using sentinel-2 imagery and spectral indices. *Environ. Monit. Assess* 197 (8), 905. <https://doi.org/10.1007/s10661-025-14332-8>.
- Shi, K., Touge, Y., 2022. Characterization of global wildfire burned area spatiotemporal patterns and underlying climatic causes. *Sci. Rep.* 12, 644. <https://doi.org/10.1038/s41598-021-04726-2>.
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>.