

Data-Driven Fire-Zone Segmentation for Improved Short-Term Wildfire Prediction

Nicolas Caron^{1*}, Hassan Noura¹, Christophe Guyeux¹, and Benjamin Aynes²

¹ Université Marie et Louis Pasteur, CNRS, institut FEMTO-ST (UMR 6174)
F-90000 Belfort, France

² SAD Marketing, Lille, France

Abstract. Wildfire prediction models typically discretize study areas into uniform grids, ignoring the heterogeneous spatial distribution of ignitions. We challenge this paradigm by showing that *how* data is discretized matters more than *which* model is used. We propose an unsupervised fire-zone segmentation algorithm combining watershed detection with K-means clustering to define prediction units directly from historical fire patterns. Experiments across six French departments and six forecasting models show that fire-zone segmentation consistently outperforms grid-based approaches, with mean IoU improvements of +3–6% depending on spatial scale. The method is computationally lightweight (<10s per configuration) and fully parallelizable. Our results demonstrate that optimizing spatial discretization yields significant, reproducible performance gains for short-term wildfire forecasting.

Keywords: Wildfire prediction · Spatial segmentation · Watershed algorithm · Unsupervised learning · Fire-zone detection.

1 Introduction

Wildfires affect every continent, arising from climatic, ecological, and anthropogenic factors. Globally, nearly half a million events are reported annually, with over 90% attributed to human activity [9]. The economic and humanitarian burden is substantial: suppression costs, infrastructure damage, and public-health impacts routinely exceed billions of US dollars per fire season [4]. Early warning systems integrating IoT sensor networks with artificial intelligence (AI) have shown considerable potential for improving detection and resource allocation [10].

This work addresses the detection and spatial segmentation of fire-prone zones as a preprocessing step for wildfire prediction. Specifically, we investigate whether partitioning the study area into semantically meaningful fire zones—rather than relying on arbitrary uniform grids—can improve the accuracy of short-term wildfire forecasts generated by AI models. Code and supplementary materials of this article, including a synthetic-data test, are available on github.

* Corresponding author. Email: nicolas.caron@umlp.fr

1.1 State of the art

Artificial intelligence and Internet-of-Things technologies already support wildfire detection and preparedness, yet predictive performance depends critically on how space is discretized [10]. CNN and ULSTM variants have been explored at kilometre-scale resolutions in California and the Mediterranean [15, 16]; however, results at such fine scales are difficult to interpret due to the overwhelming prevalence of fire-free cells, which necessitates undersampling and biases validation metrics. Several studies [13, 20] report declining prediction quality when targeting very fine resolution, calling into question the reliability of pixel-level forecasts. Furthermore, aggregated information, such as total fire counts or cumulative burned area, is lost at this granularity. Coarser spatial resolutions can mitigate these issues, but the question of how to delineate zones then arises. Contemporary work typically relies on uniform grids: Michail et al. [19] coupled GraphCast [17] with a temporal encoder on a 0.25° grid; Chen et al. [7] partitioned Portugal into ten cells to predict burned area; and Koh et al. [14] analyzed U.S. risk on a 0.5° grid.

Recent work has increasingly framed wildfire mapping as a *supervised* image-segmentation task using deep learning architectures such as U-Net, Vision Transformers (ViT), and attention-based models [2, 12, 18]. These methods excel at detecting active fires or delineating burned areas from multispectral satellite imagery when pixel-level ground-truth labels are available. However, they address a *different problem*: detecting or segmenting fires *after* ignition, rather than defining prediction zones *before* forecasting.

A key limitation of the state of the art in daily wildfire prediction is its reliance on regular grid-based segmentation, which disregards the complex, irregular spatial distribution of wildfires. Such grids introduce noise into model training (water bodies, unreliable historical encoding, etc.), potentially degrading predictive performance. In this article, we address this gap by proposing a fire-zone segmentation algorithm that leverages image-processing techniques to define semantically meaningful prediction zones. Our results demonstrate clear improvements over conventional grid-based approaches, highlighting the importance of spatially aware segmentation for wildfire prediction. Table 1 summarizes key differences between our approach and recent related work.

1.2 Contributions

This article addresses a central issue in wildfire modeling: the definition of prediction units (grid cells vs. fire zones) conditions the model’s capacity to learn ignition patterns. We demonstrate that grid-based spatial discretization constitutes a major bottleneck in wildfire prediction, it dilutes the ignition signal and introduces spatial noise. To overcome this limitation, we introduce a fire-area segmentation algorithm that combines watershed detection with clustering to align prediction zones with historical ignition patterns and thereby improve short-term forecasting. Models trained on this segmentation consistently outperform those using a standard grid across multiple spatial scales. While

Table 1: Comparison of spatial discretization approaches for wildfire prediction.

Study	Resolution	Segmentation	Supervised	Task
Michail [19]	0.25°	Grid	N/A	Forecast
Chen [7]	10 cells	Grid	N/A	Forecast
Koh [14]	0.5°	Grid	N/A	Risk
Huot [12]	Pixel	None	Yes	Detection
Ban [2]	Pixel	None	Yes	Detection
Ours	0.2-0.4°	Fire-zone	No	Forecast

this partitioning approach enhances predictive performance, we do not claim optimality, and no systematic comparison with alternative segmentation strategies has yet been conducted. Moreover, no public dataset provides a validated spatial segmentation to support standardized evaluation. This work opens a different research perspective in wildfire prediction, one that does not seek new models but rather optimizes how the dataset itself is constructed. Figure 1 illustrates the pipeline developed in this study. Figure 2 illustrates the proposed algorithm to obtain fire zone segmentation.

2 Proposed Method: Fire-Area Segmentation

Throughout this study, we use the French *department* as the reference territorial unit. Departments are administrative divisions located between regions and municipalities. This level is retained because it matches the organization of French fire services, the scale at which operational risk is commonly assessed, and the structure of the available datasets.

Conventional grid-based discretization methods that partition the study area into uniform cells are suboptimal because fire ignitions follow complex, non-uniform spatial patterns. Accordingly, our approach prioritizes cluster size, ensuring adaptability across departments and scales. Since no public dataset exists specifically for *pre-fire* fire-zone segmentation, we rely on unsupervised segmentation algorithms. The proposed pipeline segments fire-prone areas prior to any predictive modeling. Our method comprises three main stages: (A) generating a continuous 3D signal of fire risk, (B) detecting fire-prone areas, and (C) merging fire areas to match a specified target size.

Figure 2 illustrates the detailed steps of the image-processing method. The objective of this algorithm is to construct optimal prediction zones, i.e., zones in which the number of fires can be predicted, for a specified zone size, based on the locations of known fires. The same algorithm can be applied to risk locations provided by upstream AI models when prediction zones are not known in advance.

The method takes three input parameters:

1. **Scale (s)**: Determines the size of the areas in degrees. Smaller scale values result in more zones, while larger values reduce their number by merging areas.

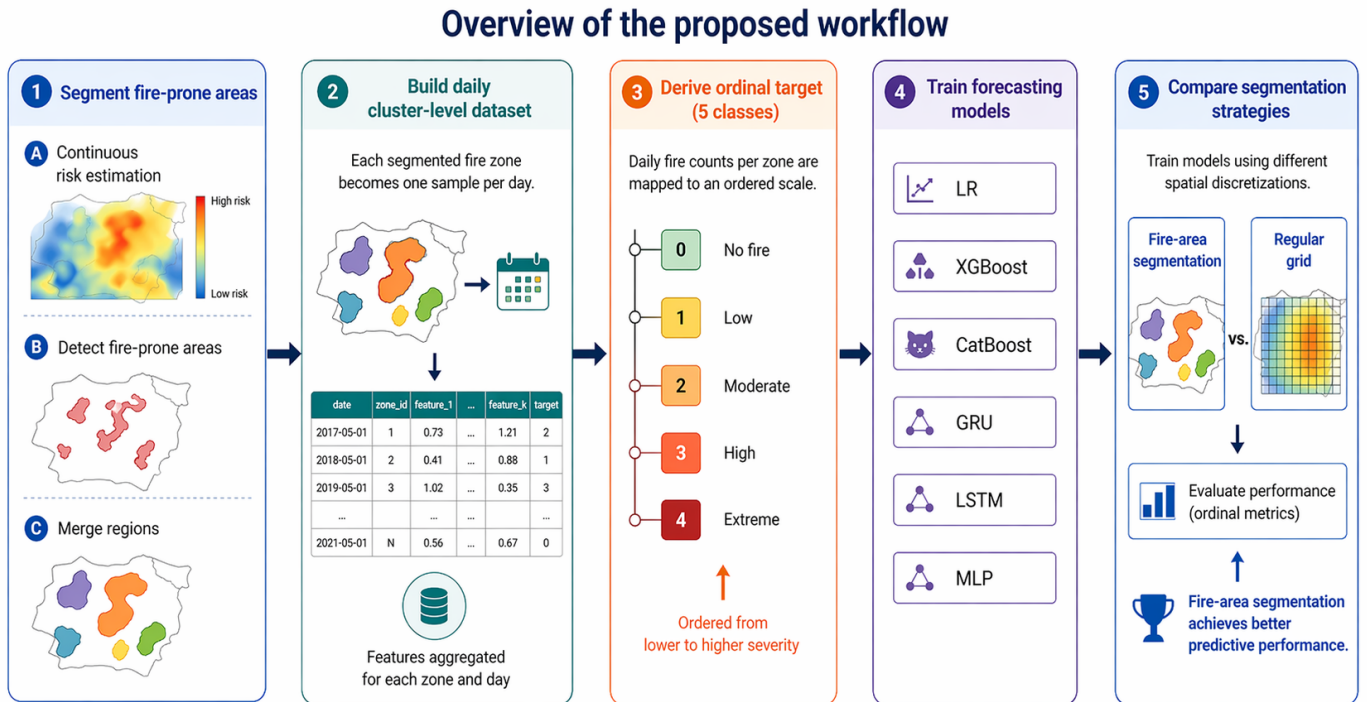


Fig. 1: Process pipeline. First, the algorithm detects and segments fire-affected zones in which risk will be forecast. Next, a daily dataset is built with each fire cluster treated as an individual sample. Finally, each model is trained on both the fire-area segmentation and a conventional regular grid to demonstrate the advantages of our method.

The number of pixels (*size*) corresponds to the number of 2 km resolution pixels within a square of $\text{scale} \times \text{scale}$ degrees.

2. **intensity_levels** (*r*): Controls the number of intensity thresholds for fire regions. Higher values generate larger intensity zones, while lower values focus on localizing the highest-risk areas.
3. **max_dilations** (*a*): Specifies the maximum number of dilation iterations applied to merge undersized zones. Higher values produce more connected regions but may introduce noise.
4. **tol** (*t*): Defines a tolerance on the region size. Since the data are organized by department, it is difficult to obtain regions that exactly match the desired size (pixel-perfect), as fire zones do not necessarily follow particular sizes and may vary across departments. However, some may be sufficiently close. This parameter allows defining a maximum and minimum size such that $\text{size} \pm \text{size} \times \text{tol}$ with *size* being the number of pixels of 2km resolution in a $\text{scale} \pm \text{scale}$ square.

2.1 Continuous risk signal

The raw fire occurrence raster is too sparse and discrete to be directly used for spatial segmentation, as ignition points are isolated and do not form continuous regions. To enable the watershed algorithm to detect coherent fire-prone areas, we first transform this discrete signal into a continuous spatial risk surface. Risk is first estimated for each pixel using a Laplacian distribution-based filter applied to a 3D raster of fire occurrences. It smooths the signal based on the average duration of fire sequences within a 20 km radius (computed seasonally and regionally). Summing along the temporal axis yields cumulative risk.

2.2 Detect fire-prone area

This step reduces noise and helps identify the main fire areas. The cumulative risk is clustered into *n_reducor_class_class* groups using K-means. The *n_reducor_class_class* parameter is used to select the intensity of the fire-prone areas detected by the watershed algorithm.

2.3 Merge fire-prone areas

The purpose of this step is to ensure that each detected fire-prone region matches a target size corresponding to the spatial scale at which predictions will be made. In other words, the algorithm does not only detect fire-prone areas, but reshapes them so that they become valid prediction units. The pseudo code, Python code, and a detailed figure are available in the supplementary materials. Figure 3 shows a simple diagram of the merge algorithm used in this article.

Zones smaller than the minimum size are merged with neighboring zones. The algorithm selects the largest neighboring zone that yields a cluster of the required size. When no valid neighbor is found, areas are expanded (dilated) to reach more

distant neighbors. Undersized zones persisting after *max_dilations* iterations are removed, while oversized zones are eroded. After each merge, the algorithm reruns until every cluster is validated. If none of the zones meet the criteria, the output is an image containing only the background (i.e., no segmented regions). Background pixels (i.e., non-fire pixels) are recursively split until all clusters reach the required size ($\text{size} \pm \text{size} \times \text{tol}$).

The output of the algorithm is a labeled raster in which each pixel belongs to a fire-zone cluster that defines a prediction unit. On average, the algorithm completes a segmentation in 40 seconds, including I/O operations. This benchmark was performed on a Dell Precision 7780 with 32 GB of RAM and a 13th-generation Intel Core i7-13850HX (28 cores). Computation time scales with the size of the study area. Since the algorithm does not explicitly account for the size of the original study region, no modifications are required to apply it at larger scales (e.g., an entire country). Applying the algorithm at finer resolution (e.g., 1×1 km) likewise requires no changes, because the number of pixels within a zone is the relevant quantity and not the pixel resolution itself.

The grid-based segmentation was obtained by recursively splitting each department in two until clusters of size $\text{size} \pm \text{tol} \times \text{size}$ were produced. The splitting was performed using a 2-cluster K-Means algorithm. A split is done only when doing so makes the resulting cluster sizes closer to the target size than the original cluster.

2.4 Algorithmic design choices

We selected watershed segmentation combined with K-means clustering over alternative approaches such as SLIC superpixels [1], Mean-Shift [8], or DBSCAN for the following reasons. First, watershed naturally detects “basins” in the density surface generated from historical ignitions, producing zones that follow the topography of fire risk rather than imposing arbitrary geometric shapes. Second, K-means provides explicit control over the *number* of merged regions via the `intensity_levels` parameter, which is essential to match operational constraints (departmental boundaries, target zone sizes). Third, both algorithms are computationally lightweight and deterministic given a fixed random seed, enabling reproducible experiments across multiple configurations. SLIC, while effective for image oversegmentation, does not adapt well to sparse, irregularly distributed point data; Mean-Shift is sensitive to bandwidth selection and can produce unpredictable cluster counts; DBSCAN struggles when density varies significantly across the study area. Our two-stage approach (watershed + K-means merging) balances spatial fidelity with controllable granularity.

3 Datasets

This section describes the two datasets used to validate our method. Both are organized by French *departments*, administrative divisions situated between the regional and municipal levels.

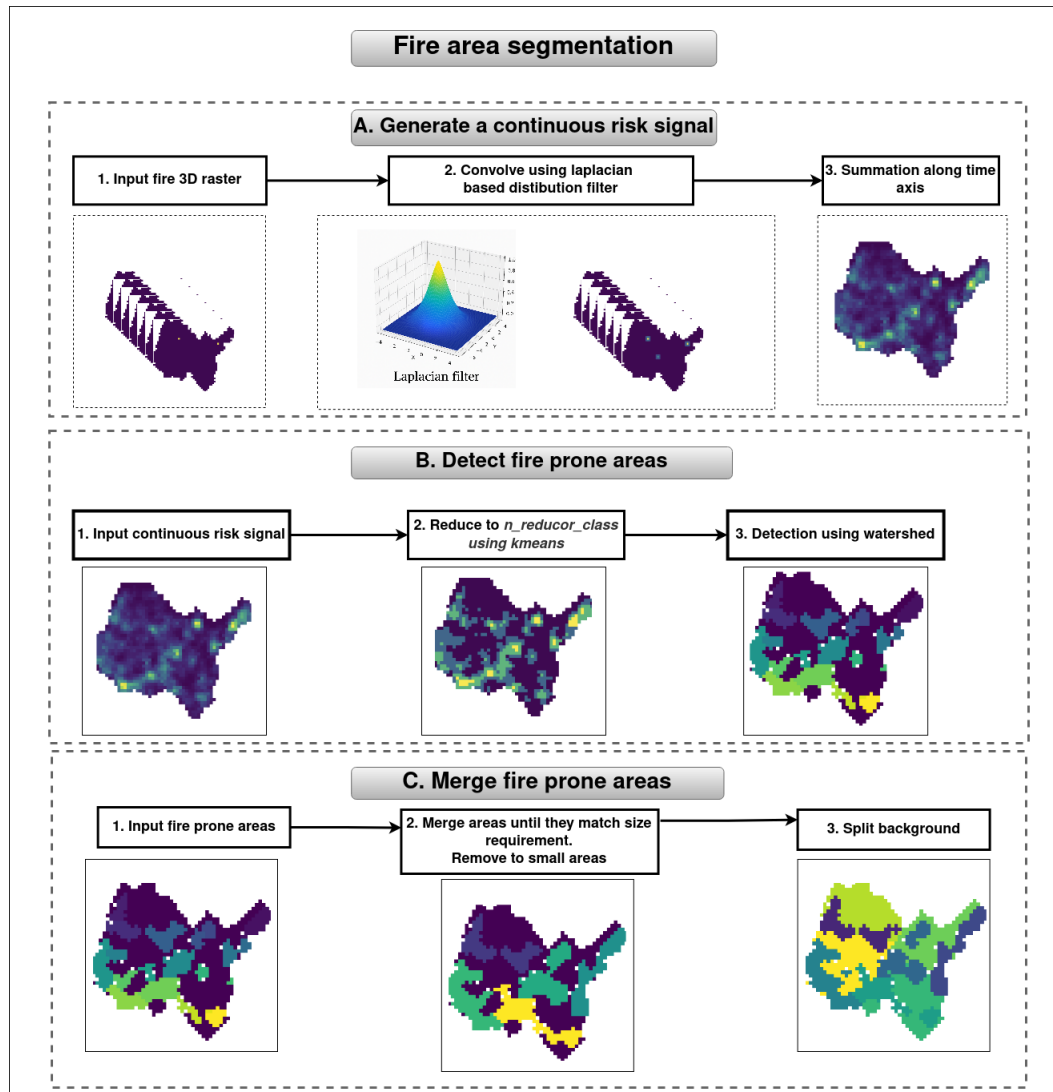


Fig. 2: Detailed steps of the image-processing method implemented in this article. White pixels represent hexagons that were incorrectly rasterized at 2 km resolution and were removed.

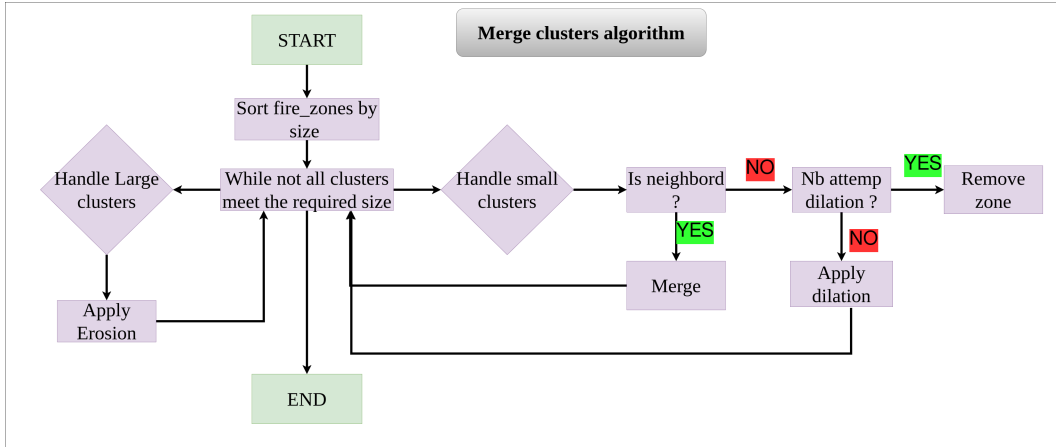


Fig. 3: Simple scheme of the merging cluster algorithm used in this article.

3.1 Fire data

Firefighter interventions. This dataset compiles wildfires that required firefighter intervention across four French departments: Ain (01), Doubs (25), Rhône (69), and Yvelines (78). These regions have recently experienced increasingly severe fire seasons with a rising number of recorded incidents; however, they have received little research attention. Most fires result from human activity, although their precise causes remain unknown. The recorded location corresponds to the position from which the fire was reported. Data span from 2017 to 2024, except for Rhône, which ends in 2022.

BDIFF. The Forest Fire database has centralized data on forest fires in France since 2006. Each fire is recorded based on the municipality in which it occurred. We selected the Bouches-du-Rhône and Hérault departments, which are well known for their vulnerability to fires. Note that forest fires may not fully represent a department’s fire risk, as other types of fires (e.g., wildland or natural-area fires) are not included. To the best of our knowledge, no prior study has made this distinction. Data span from 2017 to 2023.

These datasets are well suited to evaluate spatial discretization because fire locations are imprecise and heterogeneous.

3.2 Features and Target

The features and the target are loaded as was done in the article by Caron et al. [5]. Wildfire forecasting is treated as an ordinal multi-class task. For every department, a five-level occurrence label is created: days with no fires are placed in class 0, while all positive instances are grouped by K-means into four ordered categories—Normal, Medium, High, and Extreme. This scheme emphasizes typical

fire-activity patterns instead of relying on absolute counts. The ordinal target is unrelated to the segmentation process and simply reflects how wildfire occurrence is modeled in this study.

The features computed for training the models are grouped into six categories, as shown in Table 2: Meteorological, Topographic, Socio-Economic, Air Quality, Hydrological, and Historical. Feature processing is the same method as in the original article. We add *Air quality* and *Hydrological* variables for *Firefighter* dataset. The integration follows the same scheme: features are transformed into a 3D raster with a resolution of 2 km, aggregated using average, maximum, and minimum. Features with low variance or high correlation (measured with Pearson, Spearman, and Kendall methods, keeping those with the highest variance) are removed. The 2 km resolution is not a requirement of the segmentation algorithm. It was chosen as a practical compromise between spatial precision and the uncertainty in reported fire locations, ensuring that each pixel reasonably covers the area where a fire likely occurred.

The datasets are split into training (2017–2021), validation (2022), and testing (2023) subsets. All preprocessing was performed on the training set and then applied to the remaining data. The **Firefighters** and **BDIFF** datasets are treated separately. Each row consists of the aggregated features and the target risk for a specific cluster on a specific day. The segmentation was performed using the training set only.

Table 2: Summary of features used in this study. '-' means the same as above.

Variables	Frequency Source		Variables	Frequency Source	
Meteorological			Topographic		
Temperature	12h, 16h	Meteostat	Elevation	Static	IGN
Dew Point	-	-	Forest landcover	-	-
Precipitation	-	-	Landcover	-	Cosia
Wind Direction	-	-	NDVI, NDSI, NDMI, NDBI, NDWI 7 days	-	GEE (landsat 1+2)
Wind Speed	-	-	Swelling-shrinking of clays	-	-
Precipitation in Last 24 hours	-	-			
Snow height	-	-			
Sum of last 7 days rain drop	-	-			
Day since last rain	12h	-			
Nesterov	-	firedanger			
Munger	-	-			
KBDI	-	-			
Angstroem	-	-			
BUI, ISI, FFMC, DMC, FWI,	-	-			
Daily severity rating	-	-			
Precipitation Index last 3, 5, and 9 days	-	Calculated			
Socio-Economic			Historical		
Highway	Static	OSMNX [3]	Past risk	Daily	Calculated
Population	-	Kontur	Past risk burned area	-	-
Calendar	Daily	-	Cluster	Static	-
			Department	-	-
Air Quality (Only for Firefighter dataset)					
O3	Daily	Geodair	NO2	-	-
PM10	-	-	PM25	-	-
Hydrology (Only for Firefighter dataset)					
River height	Daily	vigierues	Groundwater depth	Daily	eaufrance

4 Training

This section presents the training scheme employed to validate the segmentation methods. For each segmentation approach (watershed-based or grid-based), every model is trained to predict daily fire risk within each cluster.

Table 4 lists the models evaluated in this study. Detailed hyperparameter values for each model are provided in the supplementary materials, along with the Python code.

The objective of this study is to demonstrate the benefits of segmenting fire-affected areas prior to short-term prediction. To that end, we selected fast-to-train models so as to keep the hyperparameter search within an acceptable time budget, on the order of a few days, using five AMD EPYC 7513 (Zen 3) CPUs with 256 GB of RAM. This choice is further justified by recent benchmarks showing that gradient-boosted decision trees (XGBoost, CatBoost) remain highly competitive with, and often outperform, deep learning models on tabular data [11]. More computationally demanding architectures such as Transformers or CNNs could be evaluated in a subsequent phase once the segmentation parameters have been finalized and the concept validated.

Because the segmentation algorithm can produce different clusters on each run, optimizing these clusters is essential. A grid-search strategy over *max_dilations* and *intensity_levels* was used to generate a large number of candidate partitions for training each model. For each segmentation method (grid-based or fire-area-based), a new dataset is constructed in which each sample corresponds to one zone on one day. All models are then trained and evaluated under identical conditions. Consequently, the comparison is between two different ways of constructing the dataset rather than between models. To this end, we firstly fixed the parameter *tol* at 0.3, the highest tolerance value for which no overlap is observed in the cluster-size distribution across scales, with a step of 0.1° (see table 3). The search grid was then specified as

$$\begin{aligned} \text{max_dilations} &\in \{1, 2, 3, 4, 5\}, \\ \text{intensity_levels} &\in \{2, 3, 4, 5, 6\}. \end{aligned}$$

Analysis were done on three different scales (in degrees):

$$\text{scales} \in \{0.2\checkmark, 0.3\checkmark, 0.4\checkmark\},$$

By comparing results across a regular parameter grid, we identified the combination that maximizes predictive performance, separately for each spatial scale and forecasting horizon. All experiments used a fixed random seed of 42 to ensure reproducibility.

We observe that the optimal parameters change across all configurations of models, departments, scales, and forecast horizons. Optimal values of *max_dilations* and *intensity_levels* are provided in the supplementary materials.

To address class imbalance, we tested different proportions of zero-class samples (from 0.05 to 1.0) and selected the proportion yielding the best validation score. Across most configurations, a proportion of 0.3–0.5 proved optimal, balancing signal retention with noise reduction.

Table 3: Range of cluster sizes at each scale studied with a tolerance of 0.3

Scale	Size (avg.)	Minimum size	Maximum size
0.2	169 px	84 px	157 px
0.3	256 px	179 px	332 px
0.4	677 px	338 px	629 px

Table 4: List of Models Used (LR = Logistic Regression)

Model	Type
LR (Logistic Regression)	Linear model
XGBoost	Decision tree (Boosting)
CatBoost	Decision tree (Boosting)
GRU	Recurrent neural network
LSTM	Recurrent neural network.
MLP	Multi Layer Perceptron

5 Results

The models’ performance is evaluated using the IoU metric which compute the intersection over union between the predicted and the real signal. It is well-suited for multiclass wildfire prediction as it accounts for class uncertainty and preserves class ordinality—predicting class 1 instead of 4 is penalized less than predicting 0 [6]. The binary F1 score is shown in the supplementary materials and leads to a similar conclusion.

5.1 Firefighter interventions

Table 6 compares the performance between grid and fire area segmentation in the **Firefighter** dataset. Table 5 shows the average IoU gain obtained using fire area segmentation. Across the three departments, the *fire-area* strategy consistently improves IoU over the grid baseline. Doubs shows the strongest uplift ($\Delta = +0.05$ across all scales); Ain benefits only at the mid- and high-thresholds, while Yvelines gains mainly at mid-scale.

5.2 BDIFF

Table 7 compares grid and fire-area segmentation performance in the **BDIFF** dataset. Table 8 shows the average IoU gain obtained using fire-area segmentation. The *fire-area* approach systematically outperforms the grid baseline. Bouches-du-Rhône gains most at the lower scale ($\Delta = +0.06$ at $s = 0.2$) while Hérault peaks at the mid-scale ($\Delta = +0.06$ at $s = 0.3$).

Table 5: Mean IoU gain Δ (fire area – grid) per department and lead time.

Department	$s = 0.2$	$s = 0.3$	$s = 0.4$
Ain	0.00	0.04	0.04
Doubs	0.04	0.04	0.05
Yvelines	0.03	0.04	0.01

Table 6: IoU performance by department and scale on **Firefighter** dataset (0-day forecast).

Model	Ain			Doubs			Yvelines		
	0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
Fire area									
LR	0.14	0.24	0.34	0.12	0.19	0.25	0.17	0.26	0.32
XGBoost	0.15	0.26	0.33	0.12	0.18	0.25	0.18	0.26	0.32
CatBoost	0.16	0.25	0.35	0.12	0.18	0.25	0.17	0.26	0.34
GRU	0.14	0.21	0.33	0.12	0.17	0.25	0.17	0.24	0.33
LSTM	0.14	0.24	0.33	0.12	0.15	0.26	0.17	0.25	0.35
MLP	0.14	0.25	0.30	0.12	0.19	0.24	0.17	0.28	0.33
Grid									
LR	0.13	0.21	0.32	0.10	0.14	0.20	0.13	0.20	0.32
XGBoost	0.14	0.20	0.31	0.11	0.14	0.22	0.15	0.23	0.32
CatBoost	0.14	0.22	0.30	0.10	0.14	0.20	0.15	0.22	0.33
GRU	0.13	0.21	0.31	0.10	0.13	0.20	0.15	0.21	0.32
LSTM	0.14	0.20	0.29	0.09	0.13	0.21	0.13	0.22	0.33
MLP	0.13	0.19	0.23	0.10	0.14	0.16	0.15	0.22	0.31

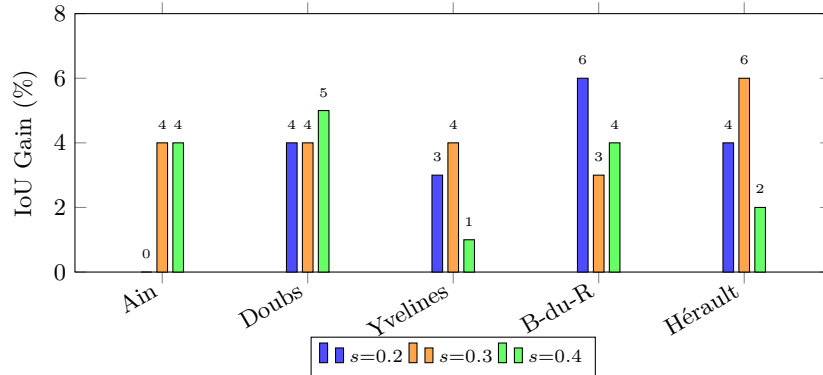
Fig. 4: Mean IoU improvement (fire-area over grid segmentation) across all departments and spatial scales. The medium scale ($s=0.3^\circ$) yields the most consistent gains.

Table 7: IoU performance by department and scale on **BDIFF** dataset (0-day forecast). Red values indicate the best values for a given scale.

Model	Bouches-du-Rhone			Herault		
	0.2	0.3	0.4	0.2	0.3	0.4
Fire area						
LR	0.26	0.30	0.34	0.11	0.19	0.23
XGBoost	0.25	0.29	0.34	0.13	0.17	0.22
CatBoost	0.25	0.30	0.35	0.12	0.19	0.24
GRU	0.23	0.28	0.34	0.11	0.16	0.22
LSTM	0.23	0.30	0.34	0.12	0.16	0.21
MLP	0.23	0.28	0.33	0.11	0.17	0.21
Grid						
LR	0.17	0.29	0.32	0.065	0.12	0.20
XGBoost	0.20	0.25	0.32	0.086	0.11	0.20
CatBoost	0.19	0.28	0.32	0.066	0.12	0.23
GRU	0.18	0.25	0.28	0.065	0.11	0.16
LSTM	0.16	0.26	0.29	0.060	0.11	0.18
MLP	0.19	0.25	0.29	0.10	0.13	0.23

Table 8: Mean IoU gain Δ (fire area – grid) per department and lead time.

Department	$s = 0.2$	$s = 0.3$	$s = 0.4$
Bouches-du-Rhone	0.06	0.03	0.04
Herault	0.04	0.06	0.02

We recommend fire-area segmentation as the primary method, a “no-regrets” improvement over grid-based discretization. The medium scale ($s = 0.3^\circ$) yields the most robust cross-department performance (see Figure 4). At finer scales ($s = 0.2^\circ$), gains remain positive but metrics are more volatile; at coarser scales ($s = 0.4^\circ$), the performance gap narrows. In no scenario does grid-based segmentation surpass fire-area segmentation. Figure 5 compares both approaches for Doubs and Bouches-du-Rhône.

A reproducibility test using logistic regression revealed result variability of approximately 0.01 for both approaches, confirming that observed differences reflect genuine methodological improvements.

5.3 Discussion

The results highlight the key advantage of segmenting fire-prone areas prior to prediction: improved model learning. Our approach is complementary to supervised deep-learning methods (e.g., U-Net, ViT): it can serve as an unsupervised preprocessing step that defines semantically meaningful prediction units *before* any downstream segmentation or classification model is applied. However, several limitations remain and will be addressed in future work.

Parameter coupling across departments. The current parameter optimization assumes that the algorithm’s optimal parameters are identical for every department, which is incorrect, for example, Bouches-du-Rhône requires 2 dilations, whereas Hérault requires 3. Detailed per-department configurations and their cross-application effects are provided in the supplementary materials. Two avenues could mitigate this coupling:

1. **Exhaustive cross-configuration experiment.** Rather than applying a single segmentation to all regions, a separate experiment can be run for *every possible combination* of departmental configurations at the same scale. Although this increases computational cost, the procedure systematically explores the entire space of segmentation choices. Depending on the model and hardware, training could easily require several weeks.
2. **Department-specific models.** Training one model per department would achieve the same isolation more economically; however, the smaller number of fire events per dataset could hurt generalization and degrade overall performance, potentially leading to suboptimal segmentation.

Scalability considerations. The segmentation algorithm has modest computational requirements. On our hardware (AMD EPYC 7513, 256 GB RAM), processing a single department at 2 km resolution ($\approx 100 \times 100$ pixels) takes less than 10 seconds per configuration; the full grid-search over 25 configurations completes in under 5 minutes per department. Memory footprint remains below 2 GB even for the largest departments. For national-scale deployment, the algorithm can be applied department-by-department in a fully parallelizable fashion. At finer resolutions (< 1 km), memory and time costs would increase quadratically with pixel count, suggesting that tiled processing or downsampled density estimation would be required.

Cross-boundary zones. We acknowledge that the current experiments do not handle zones that straddle departmental boundaries, a limitation arising from three considerations: (i) French fire services are organized at the departmental level, so operational risk assessments must be reported by department; (ii) aggregating by department reduces the volume of data that must be stored in memory; and (iii) we lack access to a complete, nationwide record of fire-brigade interventions.

Data-scarce regions. When a region has very limited wildfire experience or when no historical fire records are available, the segmentation process becomes challenging with the current algorithm due to the absence of ignition points. To address this issue, we consider two strategies. First, a spatial model (e.g., CNN, U-Net, SwinUNet) can be trained to learn the segmentation produced by the algorithm on data-rich regions and then used to infer a segmentation for previously unseen or data-scarce regions. Second, an upstream fire-susceptibility map could be constructed and used as input to the segmentation algorithm.

Scope of the segmentation comparison. This study primarily aims to highlight the underperformance of regular grid-based segmentation in wildfire prediction, rather than to claim the optimality of the proposed method; future work should therefore systematically compare it with alternative spatial partitioning strategies, including graph-based segmentation, hierarchical clustering, and adaptive grid approaches. Beyond accuracy gains, segmentation design must also satisfy operational and methodological constraints. In practice, prediction zones should remain compatible with administrative wildfire management structures and have coherent sizes across the study area. In our case, the main limitation is that segmentation parameters must currently be optimized jointly with the predictive model, requiring a full training procedure for each candidate configuration. Developing strategies to decouple segmentation optimization from model training would therefore represent a substantial contribution to the literature.

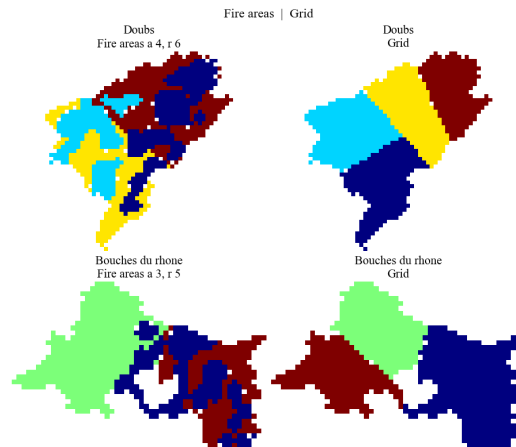


Fig. 5: Comparison of watershed and grid segmentation with $\text{scale}=0.3$ and $\text{tolerance}=0.3$. Colors indicate unique IDs with no other meaning. $a = \text{max_dilations}$, $r = \text{intensity_levels}$.

6 Conclusion

In this study, we demonstrate that conventional grid-based segmentation is poorly suited to wildfire prediction because it introduces spatial noise that weakens the model’s ability to learn ignition patterns. To address this limitation, we propose a fire-zone segmentation method that defines prediction units directly from historical fire distributions.

Evaluated across six French departments, six forecasting models, three spatial scales, and multiple lead times, this data-driven segmentation consistently outperforms the static grid approach with mean IoU improvements of +3–6%. The method requires no labeled segmentation data, runs in under 10 seconds per configuration, and integrates seamlessly into existing prediction pipelines.

The parameter search further indicates that no single configuration is optimal for all departments, highlighting the importance of region-specific tuning. For operational deployment, we recommend the medium scale ($s = 0.3^\circ$) as a robust default. Future work may explore exhaustive cross-configuration testing, transfer learning from data-rich to data-scarce regions, or coupling the segmentation with downstream deep-learning models for end-to-end optimization.

Data Availability

The Firemen dataset used in this study is not publicly available, as it contains operational intervention records shared under restricted access. The BDIFF data are publicly accessible through the official BDIFF portal. Details on the data collection, preprocessing, and construction pipeline used in this work can also be found in Caron et al. [5].

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11), 2274–2282 (2012)
2. Ban, Y., Zhang, P., Nascetti, A., Bevington, A.R., Wulder, M.A.: Near real-time wildfire progression monitoring with sentinel-1 sar time series and deep learning. *Scientific Reports* **10**(1), 1322 (2020)
3. Boeing, G.: Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* **65**, 126–139 (2017). <https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2017.05.004>, <https://www.sciencedirect.com/science/article/pii/S0198971516303970>
4. Bot, K., Borges, J.G.: A systematic review of applications of machine learning techniques for wildfire management decision support. *Inventions* **7**(1) (2022). <https://doi.org/10.3390/inventions7010015>, <https://www.mdpi.com/2411-5134/7/1/15>

5. Caron, N., Noura, H., Guyeux, C., Aynes, B.: Localized forest fire risk prediction: A department-aware approach for operational decision support (Dec 2025). <https://doi.org/10.1109/comcomap68359.2025.11353174>, <http://dx.doi.org/10.1109/ComComAp68359.2025.11353174>
6. Caron, N., Noura, H.N., Nakache, L., Guyeux, C., Aynes, B.: Ai for wild-fire management: From prediction to detection, simulation, and impact analysis—bridging lab metrics and real-world validation. *AI* **6**(10) (2025). <https://doi.org/10.3390/ai6100253>, <https://www.mdpi.com/2673-2688/6/10/253>
7. Chen, J., Yang, Y., Peng, L., Chen, L., Ge, X.: Knowledge graph representation learning-based forest fire prediction. *Remote Sensing* **14**(17) (2022). <https://doi.org/10.3390/rs14174391>, <https://www.mdpi.com/2072-4292/14/17/4391>
8. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5), 603–619 (2002)
9. Costafreda-Aumedes, S., Comas, C., Vega-Garcia, C.: Human-caused fire occurrence modelling in perspective: A review. *International Journal of Wildland Fire* **26**, 983 (01 2017). <https://doi.org/10.1071/WF17026>
10. Giannakidou, S., Radoglou-Grammatikis, P., Lagkas, T., Argyriou, V., Goudos, S., Markakis, E.K., Sarigiannidis, P.: Leveraging the power of internet of things and artificial intelligence in forest fire prevention, detection, and restoration: A comprehensive survey. *Internet of Things* **26**, 101171 (2024). <https://doi.org/https://doi.org/10.1016/j.iot.2024.101171>, <https://www.sciencedirect.com/science/article/pii/S2542660524001124>
11. Grinsztajn, L., Oyallon, E., Varoquaux, G.: Why do tree-based models still outperform deep learning on typical tabular data? (2022)
12. Huot, F., Hu, R.L., Gober, N., Yao, L., Lyons, C., Di Luca, M., et al.: Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–13 (2022)
13. Jain, P., Coogan, S.C., Subramanian, S.G., Crowley, M., Taylor, S.W., Flannigan, M.D.: Part i: Improving wildfire occurrence prediction for conus using deep learning and fire weather variables. *Fire* **3**(4), 65 (2020)
14. Koh, J.: Gradient boosting with extreme-value theory for wildfire prediction. *Extremes (Boston)* **26**(2), 273–299 (Jan 21 2023). <https://doi.org/10.1007/s10687-022-00454-6>, epub ahead of print
15. Kondylatos, S., Prapas, I., Camps-Valls, G., Papoutsis, I.: Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the mediterranean (2023), <https://arxiv.org/abs/2306.05144>
16. Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., Fernandez-Torres, M.A., Carvalhais, N.: Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters* **49**(17), e2022GL099368 (2022). <https://doi.org/https://doi.org/10.1029/2022GL099368>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022GL099368>, e2022GL099368 2022GL099368
17. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., Battaglia, P.: Graphcast: Learning skillful medium-range global weather forecasting (2023), <https://arxiv.org/abs/2212.12794>

18. Luo, K., Lian, I.b.: Building a vision transformer-based damage severity classifier with ground-level imagery of homes affected by california wildfires. *Fire* **7**(4) (2024). <https://doi.org/10.3390/fire7040133>, <https://www.mdpi.com/2571-6255/7/4/133>
19. Michail, D., Davalas, C., Panagiotou, L.I., Prapas, I., Kondylatos, S., Bountos, N.I., Papoutsis, I.: Firecastnet: Earth-as-a-graph for seasonal fire prediction (2025), <https://arxiv.org/abs/2502.01550>
20. Vilar, L., Woolford, D.G., Martell, D.L., Martín, M.P.: A model for predicting human-caused wildfire occurrence in the region of madrid, spain. *International Journal of Wildland Fire* **19**(3), 325–337 (2010). <https://doi.org/10.1071/WF09030>, <https://doi.org/10.1071/WF09030>