

Spatial Clustering of Wildfire Hotspot Intensity in West Kalimantan, Indonesia: A K-Means Approach Using Google Earth Engine Data

Muhammad Azhar^{a,*}

^aHong Kong Shue Yan University, Hong Kong SAR, China

ARTICLE INFO

Keywords:

wildfire hotspot
K-means clustering
Google Earth Engine
West Kalimantan
peatland fire
spatial analysis
remote sensing

ABSTRACT

Recurrent vegetation and peatland fires in Kalimantan, Indonesia, generate transboundary haze that affects public health and regional air quality every dry season. This study presents a reproducible, remote-sensing-based workflow for characterising the spatial pattern of fire activity across the twelve regencies (*kabupaten/kota*) of West Kalimantan Province during September 2026, a month of intensified burning under El Niño-influenced dry conditions. Satellite-derived active-fire hotspot counts were aggregated to administrative boundaries using Google Earth Engine, and regency centroids were extracted from the associated polygon geometries. A K-means clustering model was fitted to a three-dimensional feature space combining geographic location (latitude, longitude) and log-transformed hotspot intensity, with the optimal number of clusters selected through the elbow method and silhouette analysis. The analysis identified two statistically and spatially coherent clusters. A high-intensity cluster comprising five regencies (Ketapang, Pontianak, Melawi, Sintang and Kapuas Hulu) accounted for 687 of 712 recorded hotspots (96.5%), with Ketapang Regency alone contributing 483 hotspots (67.8% of the provincial total). A low-intensity cluster of six regencies in the northern and central parts of the province accounted for the remaining 3.5%. These results delineate a concentrated fire-risk corridor running through the southern peatland- and plantation-dominated regencies of West Kalimantan and support the prioritisation of monitoring and mitigation resources toward a small number of administrative units rather than uniform province-wide deployment. The proposed workflow is lightweight, fully reproducible from open satellite data, and can be re-run monthly to track the evolution of provincial fire-risk geography.

1. Introduction

Indonesia is home to more than one-third of the world's tropical peatlands, and the island of Borneo (Kalimantan) contains the largest share of this carbon-dense ecosystem (Schmidt et al., 2024). Decades of peatland drainage, deforestation and land conversion for smallholder and industrial agriculture have made these landscapes highly susceptible to fire, particularly during the dry season and in years affected by the El Niño–Southern Oscillation (Schmidt et al., 2024; United Nations Office for Disaster Risk Reduction (UNDRR), 2024). Fire is frequently used as a low-cost method of land clearing, and once ignited, peat fires can smoulder below the surface for weeks, releasing large quantities of smoke, particulate matter and greenhouse gases (Bilotta et al., 2024).

The 2026 dry season again brought a marked increase in fire activity across Sumatra and Kalimantan under an intensifying El Niño, with satellite observations showing

fire detections tracking sharply upward in a pattern reminiscent of the severe 2015 and 2023 fire seasons (National Aeronautics and Space Administration (NASA) Earth Observatory, 2026). West Kalimantan, in particular, experienced hazardous air quality: the provincial capital, Pontianak, recorded air quality index readings among the worst in the region during this period, prompting renewed regional concern over transboundary haze (Al Jazeera, 2026). Reliable, up-to-date characterisation of where fire activity concentrates within a province is therefore essential for directing limited monitoring, firefighting and public-health resources.

A growing body of work has applied unsupervised machine-learning methods to satellite-derived hotspot data to reveal such spatial patterns. K-means clustering, in particular, has been used to delineate fire-prone areas from NASA MODIS and VIIRS active-fire products in Malang, East Java (Kirana and Bhawiyuga, 2024), across Sumatra (Kirana et al., 2020), and in comparative studies of centroid-based, hierarchical and density-based clustering algorithms for wildfire and smoke-detection hotspots (Rahman et al., 2025; Density-driven wildfire hotspot mapping). Complementary approaches for Kalimantan specifically have used

*Corresponding author

 Azhar@hksyu.edu (M. Azhar)

ORCID(s):

spatiotemporal scan statistics (Sitanggang et al., 2020), generalised space–time autoregressive models (Pratiwi et al., 2025), Bayesian conditional-autoregressive models (Rohimahastuti et al., 2025; Debataraja et al., 2025), and hybrid random-forest/conditional-autoregressive frameworks that incorporate meteorological covariates such as rainfall, air temperature and consecutive dry days (Azis et al., 2025, 2026). These studies consistently find that fire occurrence in Kalimantan is spatially clustered rather than randomly distributed, and that peatland cover, land-use pressure and dry-season meteorology are the dominant drivers (Schmidt et al., 2024; United Nations Office for Disaster Risk Reduction (UNDRR), 2024; Azis et al., 2026).

Building on this literature, the present study focuses specifically on West Kalimantan Province and asks a narrower, operational question: at the administrative (regency) level, which parts of the province formed a coherent high-intensity fire cluster during the September 2026 fire season, and how much of the provincial fire load did that cluster account for? Rather than modelling the drivers of fire occurrence, the objective here is descriptive and diagnostic—to provide a reproducible, low-cost spatial screening tool that converts a single month of satellite hotspot counts, exported directly from Google Earth Engine (GEE), into an actionable ranking and grouping of regencies for resource-allocation purposes. The workflow combines geographic coordinates and log-transformed hotspot magnitude in a single K-means feature space, an approach conceptually related to that of Kirana et al. (2020) and to global fire-regime clustering studies (Fernández-García et al., 2022), but applied here at fine administrative granularity for a single province and month.

2. Study Area and Data

West Kalimantan (Kalimantan Barat) is a province on the western part of the island of Borneo, Indonesia, comprising twelve second-level administrative units (ten regencies, or *kabupaten*, and two cities, or *kota*): Landak, Pontianak, Bengkayang, Kota Singkawang, Melawi, Sanggau, Sekadau, Sintang, Kapuas Hulu, Ketapang, Kota Pontianak and Sambas. The province spans lowland peat swamp forest along its southern and western coastal margins through to hilly interior terrain along the Kalimantan–Sarawak border, and hosts extensive oil-palm and other agricultural plantations that are frequently established or maintained using fire.

Active-fire hotspot counts for September 2026 were compiled per regency using Google Earth Engine, which aggregates satellite active-fire detections (consistent with the MODIS/VIIIRS active-fire family of products widely used for operational fire monitoring (Giglio et al., 2016; Bastarrika et al., 2024)) within each administrative polygon sourced from a standard Level-2 administrative boundary (ADM2) layer. The exported table contained, for each regency, the hotspot count, the polygon area and perimeter, and a GeoJSON representation of the polygon geometry, from which the regency centroid (latitude and longitude) was subsequently derived.

Table 1

Hotspot count and provincial share by regency, West Kalimantan, September 2026, with assigned cluster membership (Section 4).

Regency	Hotspot count	Share of total (%)	Cluster
Ketapang	483	67.8	1
Pontianak	104	14.6	1
Melawi	41	5.8	1
Sintang	37	5.2	1
Kapuas	22	3.1	1
Hulu			
Sambas	10	1.4	2
Landak	5	0.7	2
Sekadau	4	0.6	2
Sanggau	3	0.4	2
Kota	2	0.3	2
Singkawang			
Bengkayang	1	0.1	2

Twelve regencies were present in the raw export; one (Kota Pontianak, the small urban core of the provincial capital) recorded zero hotspots during the study month and was excluded from the clustering step, leaving eleven regencies with at least one detected hotspot and a combined total of 712 hotspots for September 2026. Table 1 summarises the hotspot count and provincial share for each regency, together with its assigned cluster (see Section 4).

3. Methodology

3.1. Data preprocessing and feature construction

Regency centroids were extracted from the GeoJSON polygon field associated with each administrative unit. Because the distribution of raw hotspot counts is strongly right-skewed—a small number of regencies register very high counts while most register only a handful—a $\log(1 + x)$ transformation was applied to the hotspot count prior to clustering, which is standard practice for count data of this kind and prevents a single extreme value from dominating the distance metric (Kirana et al., 2020; Fernández-García et al., 2022). The final feature set comprised three variables per regency: latitude, longitude and log-transformed hotspot count. All three features were standardised to zero mean and unit variance using a z-score scaler, ensuring that the geographic and intensity dimensions contributed comparably to the Euclidean distance used by the clustering algorithm.

3.2. K-means clustering and model selection

K-means partitions n observations into k clusters by iteratively assigning each point to the nearest of k centroids and updating centroids to the mean of their assigned points, minimising the within-cluster sum of squares (WCSS) (MacQueen, 1967). Because the standardised feature space combines location and fire intensity, resulting clusters can be interpreted as groups of regencies that are simultaneously close in space and similar in fire magnitude—an approach

used previously to delineate hotspot-prone areas from satellite fire data (Kirana et al., 2020; [Density-driven wildfire hotspot mapping](#)).

The number of clusters, k , was selected jointly using the elbow method (the point beyond which additional clusters yield diminishing reductions in WCSS) and the average silhouette coefficient, which measures how well each observation fits its assigned cluster relative to the nearest alternative cluster and ranges from -1 to 1 , with higher values indicating better-separated clusters (Rousseeuw, 1987). Candidate values of k from 2 to 7 were evaluated (an upper bound of 7 was imposed because only eleven regencies had non-zero hotspot counts), with K-means run using 10 random centroid initialisations (k-means++) at each candidate k to reduce sensitivity to initial conditions, following standard practice (Kirana et al., 2020; [Density-driven wildfire hotspot mapping](#)). The value of k with the highest average silhouette score was retained as the final model.

3.3. Software and reproducibility

All processing was implemented in Python 3 using pandas for data handling and scikit-learn for standardisation, K-means clustering and silhouette evaluation, with raw hotspot aggregation performed in Google Earth Engine. The complete workflow—data import, centroid extraction, feature engineering, model selection and cluster interpretation—is scripted end to end and can be re-executed for any subsequent month by substituting an updated GEE export, supporting routine, low-cost operational monitoring.

4. Results and Discussion

4.1. Fire activity distribution

Of the 712 hotspots recorded across West Kalimantan in September 2026, activity was heavily concentrated in a small number of regencies. Ketapang alone accounted for 483 hotspots (67.8% of the provincial total), followed by Pontianak (104 hotspots, 14.6%), Melawi (41, 5.8%), Sintang (37, 5.2%) and Kapuas Hulu (22, 3.1%). The remaining six regencies—Sambas, Landak, Sekadau, Sanggau, Kota Singkawang and Bengkayang—together contributed only 25 hotspots (3.5% of the provincial total), each recording ten or fewer detections (Table 1; Fig. 2b). This highly skewed distribution is consistent with prior findings that fire activity in Kalimantan is not spatially uniform but is instead governed by the distribution of degraded peatland, plantation land use and dry-season meteorology (Schmidt et al., 2024; [United Nations Office for Disaster Risk Reduction \(UNDRR\), 2024](#); Azis et al., 2026).

4.2. Optimal number of clusters

The elbow curve of within-cluster sum of squares decreased steadily with increasing k without a single sharply defined bend (Fig. 1a), while the average silhouette score was maximised at $k = 2$ (silhouette = 0.392) and declined monotonically for larger k (Fig. 1b). Both diagnostics therefore support a two-cluster solution as the most robust partition of the eleven regencies, indicating that, at the

Table 2

Summary statistics for the two K-means clusters of West Kalimantan regencies, September 2026.

Cluster	No. of regencies	Total hotspots	Mean hotspots / regency	Share of total (%)
1 (high intensity)	5	687	137.4	96.49
2 (low intensity)	6	25	4.2	3.51

provincial scale and for this single month, fire activity is best described as a binary contrast between a compact high-intensity group and a diffuse low-intensity group rather than a finer gradation of intermediate categories.

4.3. Cluster characteristics

Applying K-means with $k = 2$ to the standardised feature space produced the partition summarised in Table 2. Cluster 1 comprised five regencies—Ketapang, Pontianak, Melawi, Sintang and Kapuas Hulu—which together recorded 687 hotspots, or 96.5% of the provincial total, at a mean intensity of 137.4 hotspots per regency. Cluster 2 comprised the remaining six regencies—Sambas, Landak, Sekadau, Sanggau, Kota Singkawang and Bengkayang—with a combined 25 hotspots (3.5% of the total) and a mean intensity of only 4.2 hotspots per regency.

Geographically, the high-intensity cluster traces a corridor along the southern and south-eastern portion of the province—from Ketapang on the south-western coast, through Melawi and Sintang in the interior, to Kapuas Hulu on the eastern border—together with Pontianak Regency surrounding the provincial capital (Fig. 2a). The low-intensity cluster occupies the northern and north-western regencies bordering the South China Sea and Sarawak (Malaysia), together with the interior regency of Sanggau and the small urban unit of Kota Singkawang. This spatial pattern is broadly consistent with the known distribution of degraded peatland and plantation agriculture in southern and central Kalimantan, which prior studies identify as the principal drivers of chronic fire risk in the region (Schmidt et al., 2024; [United Nations Office for Disaster Risk Reduction \(UNDRR\), 2024](#); Azis et al., 2026), and with West Kalimantan-specific hotspot-modelling studies that likewise find fire risk concentrated in the province's southern peatland regencies rather than distributed evenly across it (Pratiwi et al., 2025; Debataraja et al., 2025).

4.4. Implications for fire management

From an operational standpoint, the clustering result implies that a large majority of West Kalimantan's September 2026 fire load was concentrated in only five of the province's twelve administrative units. This supports a risk-weighted allocation of monitoring assets—ground patrols, aerial water-bombing readiness and community fire-awareness campaigns—toward the Ketapang–Pontianak–Melawi–Sintang–Kapuas

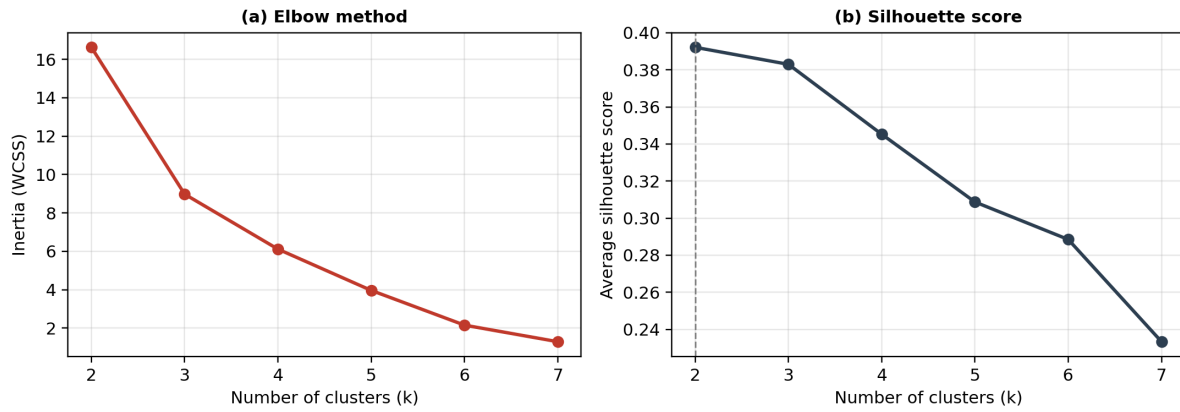


Figure 1: Model-selection diagnostics for K-means clustering. (a) Within-cluster sum of squares (inertia) as a function of the number of clusters, k . (b) Average silhouette score as a function of k , with the selected value ($k = 2$) marked by the dashed line.

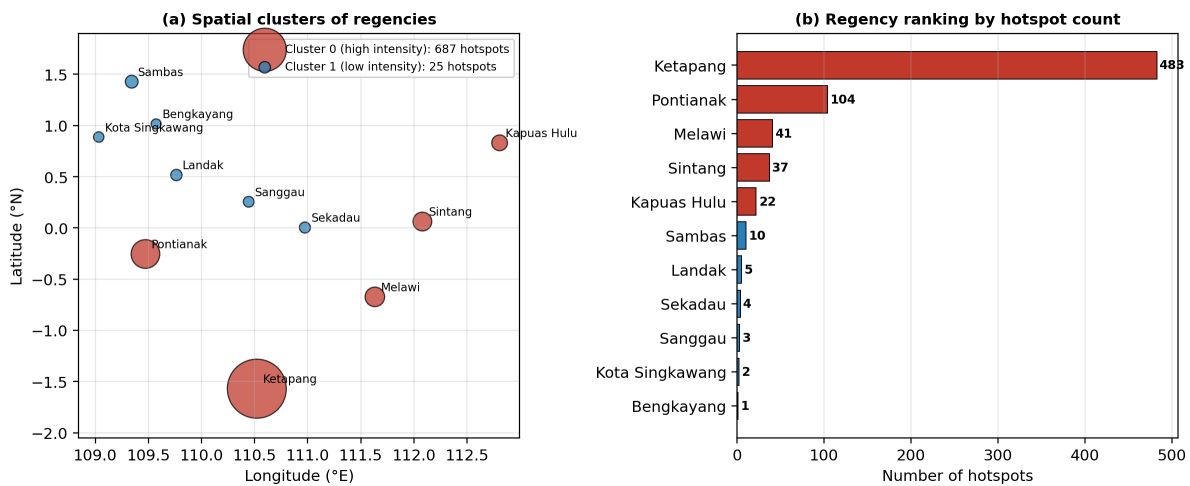


Figure 2: (a) Spatial distribution of the two K-means clusters, with marker size proportional to regency hotspot count and marker colour denoting cluster membership. (b) Ranking of all eleven regencies by hotspot count, coloured by cluster membership.

Hulu corridor ahead of and during the peak dry-season months, while maintaining baseline surveillance elsewhere. Because the same GEE-to-K-means pipeline can be re-run each month at negligible marginal cost, provincial disaster-management agencies could use it as a recurring screening layer to track whether the high-intensity cluster is expanding, contracting or shifting geographically from one fire season to the next, complementing more computationally intensive predictive approaches such as conditional-autoregressive random forests (Azis et al., 2025, 2026) or space-time autoregressive models (Pratiwi et al., 2025) where those are available.

4.5. Limitations

Several limitations should be noted. First, the analysis is based on a single month of aggregated, regency-level hotspot counts; it describes the spatial pattern of fire activity during that month rather than a long-term or seasonally averaged fire-risk climatology, and results for other

months or years may differ, particularly outside El Niño-influenced dry seasons (United Nations Office for Disaster Risk Reduction (UNDRR), 2024). Second, clustering was performed on administrative-unit centroids and aggregate counts rather than on individual point-level hotspot detections, which coarsens the spatial resolution relative to point-level or grid-based approaches used elsewhere in the literature (Sitanggang et al., 2020; Azis et al., 2025). Third, with only eleven non-zero regencies, the candidate range of k was necessarily narrow, and the two-cluster solution, while well supported by the silhouette criterion, should be interpreted as a coarse province-wide contrast rather than a fine-grained typology. Finally, the model incorporates no meteorological, land-cover or peat-depth covariates; incorporating such covariates, as in recent Kalimantan-focused predictive studies (Azis et al., 2025, 2026), would allow the descriptive clustering presented here to be extended toward an explanatory or predictive framework in future work.

5. Conclusion

This study demonstrated a lightweight, reproducible workflow that couples Google Earth Engine hotspot aggregation with K-means clustering to characterise the spatial concentration of wildfire activity across West Kalimantan Province during the September 2026 fire season. Model-selection diagnostics identified two clusters as the optimal partition of the province's fire-active regencies: a compact high-intensity cluster of five regencies (Ketapang, Pontianak, Melawi, Sintang and Kapuas Hulu) responsible for 96.5% of all recorded hotspots, and a diffuse low-intensity cluster of six remaining regencies accounting for the balance. These findings reinforce, at fine administrative granularity, the broader literature's identification of southern and central Kalimantan's peatland- and plantation-dominated landscapes as the primary locus of fire risk on the island, and offer disaster-management agencies a simple, low-cost, and repeatable tool for prioritising monitoring and mitigation resources within the province. Future work should extend the approach to multi-year, multi-month panels; incorporate point-level satellite detections and meteorological or land-cover covariates; and benchmark K-means against density-based and model-based clustering alternatives to assess the robustness of the identified fire-risk corridor.

Declarations

Funding: This research received no external funding.

Conflict of interest: The author declares no conflict of interest.

Data availability: The regency-level hotspot dataset analysed in this study was derived from Google Earth Engine active-fire and administrative-boundary layers and is available from the corresponding author upon reasonable request.

References

- Al Jazeera, 2026. Toxic haze from Indonesian forest fires chokes Malaysian Borneo. Al Jazeera, Doha.
- Azis, I., Djuraidah, A., Aidi, M.N., Indahwati, Sopaheluwakan, A., 2026. Hybrid conditional autoregressive generalized random forest for overdispersed spatial counts: evidence from Kalimantan wildfire hotspots. *Hacetep Journal of Mathematics and Statistics* 55, 734–751. doi:10.15672/hujms.1849711.
- Azis, I., Djuraidah, A., Aidi, M.N., Sopaheluwakan, A., 2025. Spatial prediction of the number of wildfire hotspots using negative binomial conditional autoregressive random forest. *Global Journal of Environmental Science and Management* 11, 1–18.
- Bastarika, A., Rodriguez-Montellano, A., Roteta, E., Hantson, S., Franquesa, M., Torre, L., Gonzalez-Ibarzabal, J., Artano, K., Martinez-Blanco, P., Mesanza, A., Anaya, J.A., Chuvieco, E., 2024. An automatic procedure for mapping burned areas globally using Sentinel-2 and VIIRS/MODIS active fires in Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing* 218, 232–245. doi:10.1016/j.isprsjprs.2024.08.019.
- Bilotta, N.G., Nurhidayah, L., et al., 2024. Updated smoke exposure estimate for Indonesian peatland fires using a network of low-cost PM_{2.5} sensors and a regional air quality model. *GeoHealth* 8, e2024GH001125. doi:10.1029/2024GH001125.
- Debataraja, N.N., Kusnandar, D., Simanjuntak, M., 2025. Spatial assessment of peat-land fires utilizing binary logistics regression in West Kalimantan. *BAREKENG: Journal of Mathematics and Applications* 19. In press.
- Density-driven wildfire hotspot mapping, 2025. Density-driven wildfire hotspot mapping and energy-aware sensor deployment, in: *Proceedings of the International Conference on Wireless Sensor Networks for Environmental Monitoring*, Springer, Cham. pp. 251–268. doi:10.1007/978-3-032-23341-7_24.
- Fernández-García, V., Marcos, E., Fulé, P.Z., Reyes, O., Santana, V.M., Calvo, L., 2022. Characterizing global fire regimes from satellite-derived products. *Forests* 13, 699. doi:10.3390/f13050699.
- Giglio, L., Schroeder, W., Justice, C.O., 2016. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment* 178, 31–41.
- Kirana, A.P., Bhawiyuga, A., 2024. Spatio-temporal pattern analysis of forest fire in Malang based on remote sensing using K-means clustering. *International Journal of Multidisciplinary: Applied Business and Education Research* 5, 1732–1742.
- Kirana, A.P., Sitanggang, I.S., Syaifina, L., Bhawiyuga, A., 2020. Spatial and temporal clustering analysis of hotspot pattern distribution of critical land in Kalimantan, Indonesia. *IOP Conference Series: Earth and Environmental Science* 528, 012042. doi:10.1088/1755-1315/528/1/012042.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley. pp. 281–297.
- National Aeronautics and Space Administration (NASA) Earth Observatory, 2026. Peatland fires darken skies in Indonesia. NASA Science, Washington, DC.
- Pratiwi, H., Imro'ah, N., Huda, N.M., Ayyash, M.Y., 2025. Comparison of weight matrix in hotspot modeling in West Kalimantan using the GSTAR method. *Jurnal Matematika UNAND* 14, 31–45. doi:10.25077/jmua.14.1.31-45.2025.
- Rahman, F., et al., 2025. An improved clustering method based on K-means for hotspot data: clustering active fire hotspots detected by NASA FIRMS VIIRS across Indonesia. *Journal of Data Science and Applications*.
- Rohimahastuti, F., Djuraidah, A., Wijayanto, H., 2025. Spatiotemporal Bayes model for estimating the number of hotspots as an indicator of forest and land fires in Kalimantan Island, Indonesia. *IPB University Repository*, Bogor.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 53–65. doi:10.1016/0377-0427(87)90125-7.
- Schmidt, A., Ellsworth, L.M., Boisen, G.A., Novita, N., Malik, A., Gangga, A., Albar, I., Nurhayati, A.D., Ritonga, R.P., Asyhari, A., Kauffman, J.B., 2024. Fire frequency, intensity, and burn severity in Kalimantan's threatened peatland areas over two decades. *Frontiers in Forests and Global Change* 7, 1221797. doi:10.3389/ffgc.2024.1221797.
- Sitanggang, I.S., Syaifina, L., Bhawiyuga, A., 2020. Spatial and temporal clustering analysis of hotspot pattern distribution of critical land in Kalimantan, Indonesia, in: *2nd International Conference on Environment and Forest Conservation (ICEFC 2019)*, Chicago.
- United Nations Office for Disaster Risk Reduction (UNDRR), 2024. Indonesia wildfires, 2023 – forensic analysis. Technical Report. UNDRR. Geneva.