



1 **A Geographically Validated Administrative Wildfire Inventory for Türkiye (2013–2025)**

2 Bikem Ekberzade¹

3 ¹Istanbul Technical University, Eurasia Institute of Earth Sciences, Istanbul, Turkey

4 Correspondence to: ekberzade19@itu.edu.tr

5 ORCID: 0000-0002-3271-2040

6 **Abstract**

7 Administrative wildfire records provide an essential foundation for understanding fire regimes,
8 evaluating fire management practices, and supporting Earth system research. However, their
9 scientific utility is often limited by typographical errors, inaccurate geographic coordinates, and
10 inconsistencies between recorded administrative locations and GPS positions. This dataset
11 presents a geographically validated administrative wildfire inventory for Türkiye spanning 2013–
12 2025, derived from official records maintained by the General Directorate of Forestry (OGM). The
13 original archive was processed using a reproducible five-stage geographic validation workflow
14 designed to systematically identify, reconcile, and document spatial inconsistencies while
15 preserving the integrity of the underlying fire records. The workflow combines administrative text
16 harmonization, spatial verification against official district boundaries, satellite-assisted
17 coordinate recovery using NASA VIIRS Active Fires / Thermal Anomalies detections, distance-
18 based border assessment, and conservative on-land coordinate imputation for irrecoverable
19 records. Rather than discarding records with erroneous coordinates, the methodology prioritizes
20 recovery of validated fire locations and, where this is not possible, preserves the statistical record
21 while explicitly preventing misleading spatial interpretations. To ensure complete transparency,
22 every wildfire record includes a provenance flag indicating whether its coordinates are original
23 (administratively validated), satellite-corrected, border-accepted, or conservatively imputed.
24 The resulting inventory substantially improves the spatial consistency and analytical reliability of
25 Türkiye’s historical wildfire archive while preserving its original event attributes. This openly
26 available dataset serves as a valuable contribution to wildfire regime analysis, validation of
27 remote sensing products and Earth system model outputs, and operational wildfire research
28 throughout Türkiye and the eastern Mediterranean.

29 **Keywords:**

30 historical wildfire inventory; administrative wildfire records; spatial validation; data provenance;
31 VIIRS; Turkey



32 **1 Introduction**

33 As wildfire activity intensifies across many regions of the world, the demand for reliable historical
34 fire records has become increasingly important for both scientific investigation and operational
35 fire management. Regional wildfire inventories constitute one of the primary observational
36 foundations for understanding fire regimes, and their evolution across spatial and temporal
37 scales (Gincheva et al. 2024). They underpin analyses of fire occurrence, burned area, ignition
38 causes, suppression effectiveness, and fire-weather relationships, while providing essential
39 reference data for evaluating satellite products, calibrating fire models, and supporting Earth
40 system research (Taylor et al., 2013).

41 National forestry agencies often maintain detailed administrative wildfire archives that include
42 information unavailable from satellite observations, such as ignition sources, wildfire duration
43 including suppression times, responding agencies, and burned area estimates. These records
44 represent an indispensable complement to Earth observation products. However, because these
45 archives are compiled through operational reporting systems to primarily support operational fire
46 management rather than scientific workflows, they may contain formatting or typographical
47 inconsistencies, transcription errors, coordinate inaccuracies, and incomplete geographic
48 information that limit their direct scientific application (Short, 2014). Resolving such
49 inconsistencies requires evaluating multiple sources of spatial evidence while preserving the
50 integrity of the original administrative record. This challenge is rarely addressed explicitly in
51 operational wildfire archives, despite its implications for subsequent spatial analyses.

52 While there is no doubt that satellite-derived fire products have transformed wildfire monitoring
53 by providing spatially consistent and near-real-time observations of active fire, they cannot fully
54 replace administrative archives (Schroeder et al. 2014; Giglio et al. 2018). This complementary
55 role is already recognized in European wildfire data systems, including the European Forest Fire
56 Information System (EFFIS), which maintains nationally supplied fire information alongside
57 separate satellite-derived fire products (Camia et al. 2014; San-Miguel-Ayanz et al. 2023).
58 Thermal anomaly products provide independent observations of active fire but may omit small or
59 short-lived events and do not have access to the operational metadata contained within forestry
60 archives (Giglio et al. 2018; Gincheva et al. 2024). Moreover, they include other forms of biomass
61 burning, such as agricultural residue fires, prescribed burning, or industrial heat sources,
62 depending on the product and application (Hall et al. 2016; Giglio et al. 2018; Ekberzade 2026a).
63 Administrative wildfire archives and satellite-derived fire products therefore provide
64 complementary, rather than interchangeable, representations of wildfire activity (Short 2014;



65 Gincheva et al., 2024). Administrative records establish the occurrence (in space and time) and
66 operational characteristics of documented forest fires, while satellite observations help verify the
67 location of the documented event with close likelihood, providing valuable evidence when the
68 national record lacks proper georeferencing. Neither source alone is sufficient to maximize the
69 scientific value of historical wildfire records, but together they provide a more reliable
70 representation of wildfire occurrence than either source in isolation (Gincheva et al. 2024)

71 Forest fires in Türkiye are overwhelmingly human-caused: frequently stemming from negligence,
72 agricultural activities (mainly stubble burning), and infrastructure-related faults, with
73 anthropogenic climate change contributing to the lengthening of the fire season and increasing
74 fire severity (Keeping et al. 2026). The General Directorate of Forestry (OGM) of Türkiye manages
75 fire prevention, suppression, and post-fire rehabilitation across the country's diverse
76 landscapes, maintaining an extensive operational reporting network. Within this network,
77 detailed fire occurrence data have historically been maintained at the national level by
78 jurisdictions, with date, ignition source, location (province, district, and, from 2024 onward,
79 locality¹). Explicit GPS coordinates started to be included in the national archive from 2013
80 onward.

81 Building upon the considerable scientific value of the OGM operational archive, this study
82 presents the first public release of the archive's most recent 12 years (covering 2013-2025) as a
83 systematically quality-controlled geospatial dataset, with documented spatial validation and
84 coordinate provenance, thereby enabling its broader application in wildfire, climate, and Earth
85 system research. The methodology used for validation is designed to preserve statistical integrity
86 and provides explicit provenance tracking for every geographic coordinate, while ensuring
87 adaptability and reproducibility in different regions.

88 **2 Methods**

89 **2.1 Dataset's geographical coverage**

90 Türkiye provides an ideal case study for fire related research due to its diverse ecoregions
91 exhibiting various wildfire dynamics, a centralized wildfire reporting system, and increasing
92 scientific importance in defining the eastern Mediterranean fire regime. The peninsula occupies
93 a climatically diverse region of northeastern Mediterranean, encompassing over 20 million ha of

¹ While the national records from 2024-2025 use the word "köy" (village) as the column name for these locations, following a 2012 legislation, villages were officially converted to neighbourhoods in Türkiye. For purposes of clarity, this manuscript therefore will refer to these entries as "neighbourhood" from here on, while retaining the original "köy" terminology in the source datasets.



94 forest, which represents nearly 30% of the country's total land area (OGM 2020; OGM 2024). The
 95 country's wildfire hotspots are historically concentrated in its southern and western regions (the
 96 Mediterranean and Aegean coasts), where hot, dry summers driven by the Mediterranean climate
 97 create highly flammable conditions between late June and September, and where landlocked
 98 arid and semi-arid landscapes of the southeast also contribute to the annual fire season with
 99 considerable burned areas (Atalay et al. 2024; Ekberzade 2026a). These fire-prone coastal and
 100 montane landscapes are characterized by rapid urban-wildland transitions and steep elevation
 101 gradients (Ekberzade et al. 2024 and 2025). The dominant vegetation driving regional fire
 102 dynamics in the Mediterranean climate-type regions in the country is the extensive maquis
 103 shrublands and forests composed primarily of Turkish red pine (*Pinus brutia*) at lower elevations
 104 (0–1,500 m) and European black pine (*P. nigra*) or Cedar of Lebanon (*Cedrus libani*) at higher
 105 altitudes (Ekberzade et al. 2022 and 2025). However, with the current state of global warming and
 106 persistent long-term drought, frequent wildfire activity is observed in the country's northern
 107 region, where a temperate rainforest with broadleaf species and a historically low fire return
 108 interval exists (Keeping et al. 2026).

109 2.2 Data processing workflow

110 The original OGM dataset consisted of separate Microsoft Excel files kept for each year between
 111 2013-2023, with records for 2024 and 2025 merged in a single file². The common variables among
 112 the separate files were province and district names (with only a number of years including locality
 113 entered as village (köy) or neighborhood (mahalle)), date and time of incident, geographic
 114 coordinates and ignition source. Each file had a different header structure (e.g. il_adı or ili used
 115 interchangeably to denote province across different years) and column formatting (e.g. the
 116 coordinates being recorded as character). Certain records lacked coordinates and there were
 117 discrepancies between the different written versions of place names (either through changes of
 118 a place name over time, or certain place names with two words being written jointly or separately,
 119 such as Kahramanmaraş, Kahraman Maraş, or written in its short form only as Maraş). In cross
 120 border wildfires, if the incident spread to a neighboring city or district, only one place name was
 121 entered in the records (often the name of the district where the first responders were dispatched
 122 from), and if the coordinates belonged to one of the other districts it created a spatial discrepancy
 123 which needed to be validated.

² OGM started including geographic coordinates in their wildfire archives from 2013 on. Earlier years only have province and district names in the records, the validation of which would require a different approach. Therefore, this study takes into consideration only on records between 2013-2025.



124 Therefore, to ensure the highest possible spatial fidelity for the historical wildfire inventory, the
 125 raw OGM archive was processed through a custom, five-stage geographic validation and rescue
 126 pipeline. Because manually entered administrative data were prone to spatial input errors, the
 127 designed algorithm helped to systematically identify errors, verify true locations via satellite data,
 128 and preserve statistical integrity without introducing macro-level spatial distortion. All stages of
 129 the data clean up and aggregation were handled using R programming language, and ArcGIS Pro
 130 was used for spatial verification of the results.

131 **2.2.1 Initial file clean-up: text normalization and pre-join dictionary audit**

132 Administrative district names lacked strict data-entry validation, leading to frequent
 133 typographical and spelling variations. Before spatial processing, all Turkish administrative text
 134 (province and district names) was stripped of special characters, capitalized, and normalized to
 135 standard ASCII format (e.g., "ŞARKIŞLA" to "SARKISLA"). A pre-join dictionary builder was then
 136 applied to systematically audit and change historical OGM versions to official civil names (e.g.,
 137 "K.DENİZ EREGLİ" to "EREGLİ"). Identifying and standardizing these text discrepancies before the
 138 spatial join prevented valid GPS coordinates from being falsely flagged as spatial errors due solely
 139 to human clerical mistakes (sample R code is provided as Supplementary Material).

140 **2.2.2 Strict Spatial Verification**

141 Following text normalization, raw Degrees/Minutes/Seconds (DMS) coordinates were converted
 142 to decimal degrees and mapped against a high-accuracy Turkish Administrative (adm2) shapefile
 143 (OCHA 2025). A strict boolean flag (*geom_mismatch*) was generated if the geographic coordinate
 144 landed in a different province or district than the text indicated. While OGM teams operate within
 145 broad jurisdictions that often span multiple civil districts, the algorithm strictly evaluated district
 146 boundaries, and actively rejected broader "Forestry Hierarchy" text matching to prevent the
 147 masking of significant GPS drift. Central metropolitan municipalities listed generically (e.g.,
 148 "Muğla-Muğla", without the mention of a district or subdistrict) were treated as exceptions and
 149 trusted to the underlying GPS coordinate. All other spatial contradictions were treated as
 150 fundamental spatial errors requiring satellite verification.

151 **2.2.3 Buffered Satellite Rescue (VIIRS integration)**

152 Records that were flagged as spatial mismatches were cross-referenced against the NASA VIIRS
 153 Active Fires / Thermal Anomalies dataset (Giglio, 2024) within a temporal window of ± 3 days. The
 154 algorithm searched for satellite-detected active fires specifically within a strict 10 km buffer of



155 the intended district polygon. If a match was found, the erroneous OGM coordinates were
 156 corrected with the precise VIIRS coordinates.

157 The strict 10 km search buffer was purposefully implemented to prevent the algorithm from
 158 searching the entirety of a province for a missing fire, risking a misinterpretation of agricultural
 159 practices. For example, during August in regions like Adana, massive agricultural stubble fires
 160 (*aniz*) frequently trigger satellite anomalies. Thus, applying a strict 10 km search buffer prevented
 161 the algorithm from artificially teleporting a small missing forest fire 20+ kilometers away to the
 162 site of a completely unrelated agricultural burn. The length of the search buffer was decided upon
 163 consultation with OGM officials.

164 **2.2.4 Distance-Based Tolerance Filtering (Orphan Analysis)**

165 The mismatched records that VIIRS could not detect (often due to the incident being <1 ha or
 166 obscured by cloud cover) were subjected to a geodesic distance analysis. The distance from the
 167 erroneous GPS point to the intended district boundary was calculated (R code in Supplementary
 168 Material). Records within a 10 km radius that successfully intersected a valid terrestrial district
 169 polygon were classified as border disputes and accepted using their original coordinates.
 170 Maritime boundary discrepancies (coordinates landing in a water body) were processed further
 171 to correct coastal drift: clerical errors within the 10 km tolerance radius were geographically
 172 snapped to the nearest shoreline of their intended district using the *st_nearest_points* function.
 173 Both inland and coastal boundary offsets were classified under the *OGM_Border_Accepted*
 174 category. Points that landed beyond the 10 km threshold or landed strictly outside national
 175 terrestrial boundaries were explicitly flagged as major spatial errors and subjected to imputation.

176 Forestry teams operating in mountainous terrain or answering to complex wildfire emergencies
 177 frequently extinguish fires that sit precisely on the administrative border of neighboring districts
 178 and thus can misreport the location of the fires. Upon discussion with the OGM officials, it was
 179 decided that a 10 km tolerance radius sufficiently accommodated for natural GPS drift (as
 180 explained above) and minor administrative border disputes without corrupting the spatial
 181 integrity of the broader dataset.

182 **2.2.5 Spatially Randomized Imputation for Unrecoverable Errors**

183 Records that contained large and unrecoverable GPS errors (e.g., coordinate swaps landing in
 184 different countries) that failed the VIIRS rescue loop were subjected to a guaranteed on-land
 185 imputation. Discarding these unrecoverable records would result in the unacceptable loss of vital
 186 statistical data such as ignition dates, causes, historical fire counts and total burned area for a



187 given district, and would thus skew the final analyses. However, during their reintroduction to the
 188 national inventory, standard mathematical centroids were avoided, as they often place the center
 189 of coastal or irregularly shaped districts (e.g., peninsulas and islands) into the sea. Assigning a
 190 single geometric center point also causes multiple erroneous records from the same district to
 191 artificially stack on a single coordinate, misrepresenting spatial density.

192 To resolve this, the workflow utilized a spatial random sampling function (*st_sample* via the *sf*
 193 package) to artificially distribute these unrecoverable records onto randomly generated points
 194 guaranteed to fall within the confirmed landmass of the intended terrestrial district polygon. In
 195 this manner, the statistical and historical counts for that province and district were preserved
 196 without creating artificial point-clusters (misleading hotspots) or distorting macro-level spatial
 197 density maps.

198 2.2.6 Data Provenance and Transparency

199 To ensure complete analytical transparency, the final dataset includes a *data_source* column
 200 tracking the origin of every coordinate: *OGM_Original* refers to records whose GPS coordinates
 201 fall within their recorded administrative location; *OGM_Border_Accepted* refers to records where
 202 GPS coordinates were retained or geographically snapped due to verified terrestrial edge-case
 203 proximity (<10 km); *VIIRS_Rescued* refers to human GPS errors successfully corrected via NASA
 204 satellite detection; and *Imputed_On_Land_Surface* refers to a terrestrial placeholder coordinate
 205 assigned to unrecoverable GPS errors (orphans), preserving burned area counts and event
 206 metadata. Additionally, a *coastal_adjustment* boolean column (TRUE/FALSE) explicitly flags
 207 records that were snapped from the sea to the coastline, allowing end-users to precisely
 208 distinguish between inland border conflicts and corrected maritime drift.

209 To preserve institutional confidentiality while facilitating international accessibility, internal
 210 administrative metadata and intermediate processing variables were excluded from the final
 211 dataset, and all retained column headers were standardized to English. A comprehensive data
 212 dictionary detailing all variables, data types, and descriptions of the final dataset is provided in
 213 Table A1.

214 3 Results

215 Between 2013 and 2025 there were 36,147 entries for wildfire into the national archive. Following
 216 text normalization and initial spatial audit, 30,212 records (approximately 84%) demonstrated
 217 high spatial fidelity, landing strictly within their stated administrative boundaries. The remaining
 218 5,927 records were flagged as major spatial anomalies and reserved for spatial review (e.g.,



geographic contradictions or offshore coordinates). Using the buffered satellite rescue algorithm, 3,295 of these erroneous records (approximately 56% of total spatial errors) were recovered using NASA VIIRS Active Fires / Thermal Anomalies dataset and appended with new satellite coordinates. 1,705 records were identified as acceptable border disputes (around 29% of total spatial errors), and retained using their original coordinates or geographically snapped to the nearest intended coastline to correct maritime drift. The final 927 unrecoverable records (constituting around 16% of total spatial errors) which couldn't be verified via satellite, often due to small fire size or cloud cover, were systematically and randomly imputed to the terrestrial surface of their intended district. Consequently, the pipeline retained 99.98% of the historical fire records for statistical analysis without introducing macro-level spatial distortion, resulting in a finalized dataset of 36,139 records, with only eight records excluded due to severe administrative text corruption (Table 1).

Table 1. Spatial validation and data rescue results by year. Total coordinate errors are further subdivided into: geographic errors corrected using satellite data; coordinates retained or snapped due to the 10 km terrestrial tolerance radius; and unrecoverable entries which are subjected to imputation. The last column denotes the final dataset size after spatially randomized on-land surface imputation was applied to the orphans. Red values highlight years where a negligible fraction of records (8 total across the 13yr period) was completely excluded from the final dataset due to severe administrative text corruption.

Year	Total raw OGM records	Spatially verified	Total coordinate errors	Recovered via VIIRS	Accepted border dispute + coastal drift	Major Spatial Error (orphans)	Final Dataset (Post-Imputation)
2013	3755	2945	809	429	204 + 11	165	3754
2014	2149	1700	449	201	158 + 2	88	2149
2015	2150	1683	466	250	132 + 3	81	2149
2016	3188	2478	708	392	192 + 7	117	3186
2017	2411	1923	488	251	149 + 6	82	2411
2018	2167	1756	410	259	84 + 9	58	2166
2019	2695	2350	342	203	111 + 2	26	2692
2020	3412	2846	566	361	130	75	3412
2021	2793	2368	425	248	117	60	2793
2022	2160	1869	291	143	93 + 2	53	2160
2023	2262	1980	282	155	86	41	2262
2024	7005	6314	691	403	205 + 2	81	7005
2025							
Total	36147	30212	5927	3295	1705	927	36139

3.1 Spatial Distribution and Localized Validation



By color-coding the final map according to the *data_source* provenance flag, it is possible to assess whether the spatial corrections introduce systematic geographic patterns. The spatial distribution of the finalized dataset is broadly consistent with the established geographic realities of Türkiye's fire regime (Fig. 1). Fire events concentrate along the Aegean and Mediterranean coasts, corresponding with the highly flammable maquis shrublands and *Pinus brutia* forests dominant in these regions (Ekberzade 2022; Atalay et al. 2024). A similar trend is also observed in the dense temperate conifer forests in the northern part of the country, with the region exhibiting an increase in wildfire activity in recent years (Keeping et al. 2026). The rescued and imputed records are distributed within the broader spatial structure of the original inventory rather than forming distinct spatial clusters, suggesting that the validation process did not introduce obvious large-scale spatial artifacts. Notably, the application of spatially randomized imputation successfully prevented the artificial clustering of unrecoverable records that would otherwise occur using standard geometric centroids.

To demonstrate the functionality of the geographic validation pipeline at a finer spatial scale, the inset in Figure 1 highlights a sample of records from Muğla province for the year 2023. The map serves as a visual confirmation of the retention of original valid records, the acceptance of edge-case geographic coordinates along the Milas-Menteşe administrative boundary, and the relocation of erroneous entries using NASA VIIRS satellite detections. Similar micro-scale visual checks were conducted alongside the automated pipeline to manually audit randomly selected records across different provinces, providing an additional check against systemic spatial artifacts.

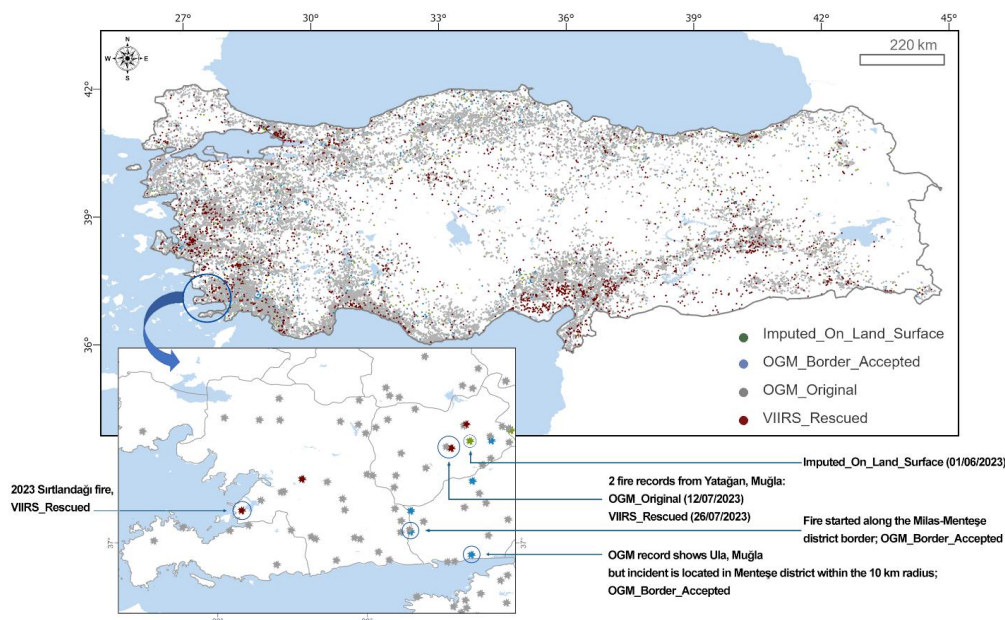


Figure 1. Spatial distribution and geographic validation of the finalized wildfire inventory for Türkiye (2013-2025). The country-wide map shows the spatial distribution of all retained wildfire records, categorized by coordinate provenance: *OGM_Original* (gray), *OGM_Border_Accepted* (blue), *VIIRS_Rescued* (red), and *Imputed_On_Land_Surface* (green). The inset shows a local scale example from Muğla province for the year 2023, highlighting the retention of original OGM



coordinates; acceptance of a record near the Milas–Menteşe administrative boundary; VIIRS-based relocation of an erroneous coordinate for the Sirtlandağı and Yatağan fires; and on-land imputation of an unrecoverable record.

3.2 Quantification of Spatial Corrections

Analyzing the geodesic distances between the originally flawed OGM coordinates and their final corrected locations reveals the substantial geographic displacement associated with the corrected records. For the 3,295 records successfully relocated via satellite thermal anomalies, the algorithm corrected spatial errors by a median distance of 27.7 km and a mean distance of 114 km, with maximum extreme coordinate outliers exceeding 7,500 km (Fig. 2). Had these raw coordinates been utilized without spatial verification, many historical fire events would have been falsely mapped into incorrect ecoregions, different countries, or oceanic bodies, potentially compromising downstream ecological and climatic modelling.

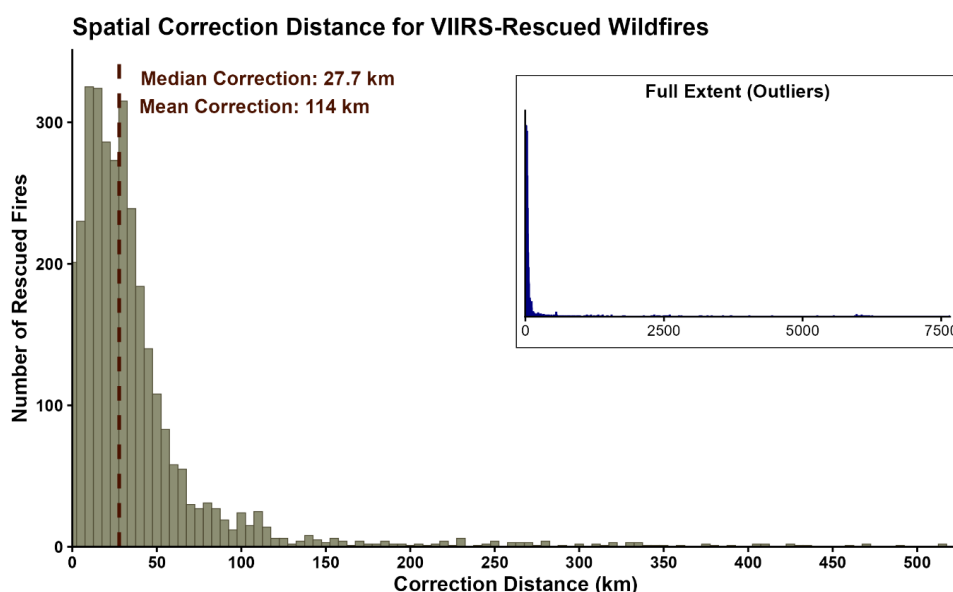


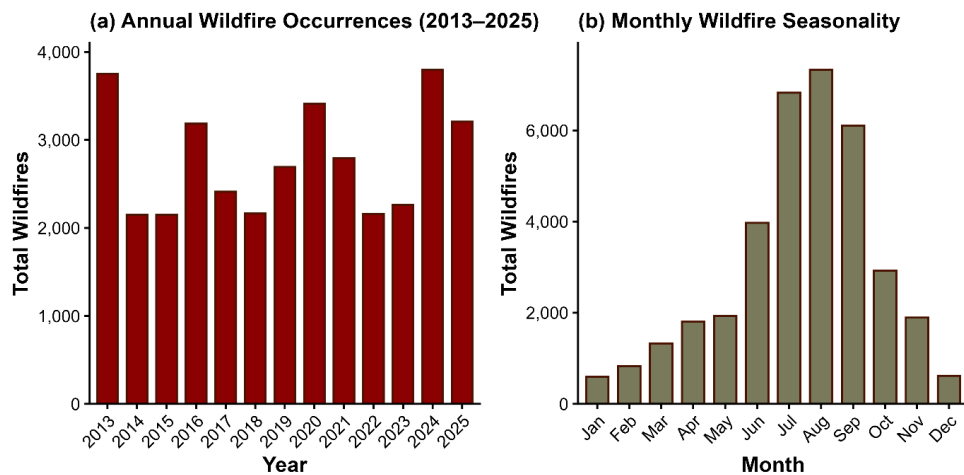
Figure 2. Geodesic spatial correction distance for wildfire records recovered via NASA VIIRS satellite cross-referencing. The main histogram is scaled to 0–500 km to illustrate the distribution of typical geographic drift and localized coordinate corrections, with the median displacement (27.7 km) marked by the dashed line. The inset plot displays the full extent of the data, highlighting severe administrative clerical errors (e.g., coordinate swaps or missing decimals) that result in extreme spatial outliers exceeding 7,000 km, driving the mean correction distance to 114 km.

3.3 Temporal and Seasonal Trends

In addition to high spatial fidelity, the final dataset preserves the temporal integrity of the original archive. The interannual distribution of fire events from 2013 to 2025 captures the historical fluctuations (Fig. 3a). Seasonally, the dataset shows a pronounced summer concentration of fire activity, with ignition frequencies peaking between July and September, in line with the Mediterranean climate-type fire regime (Fig. 3b). Because the 927 "orphan" records were



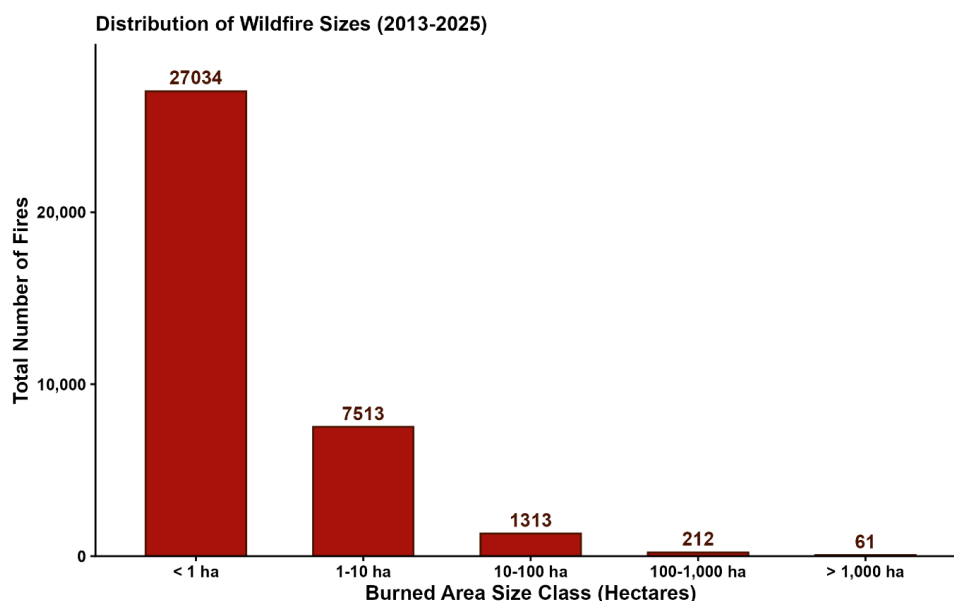
289 subjected to spatially randomized terrestrial imputation rather than deletion, these interannual
 290 and seasonal statistical counts remain intact for aggregate administrative analyses.



291 **Figure 3.** Temporal distribution of historical wildfires in Türkiye (2013–2025) based on the finalized
 292 spatially verified dataset. (a) Annual wildfire occurrences, highlighting interannual variability; (b)
 293 monthly wildfire seasonality, demonstrating the distinct concentration of fire events during the
 294 hot, dry Mediterranean summer period. Records lacking explicit temporal data in the raw archive
 295 were excluded from this visualization.

296 3.4 The Dominance of Small Fires

297 A critical advantage of combining geographically validated administrative records with satellite
 298 detection is the retention of small-scale fire events that may be missed by automated global
 299 burned area (BA) products. An analysis of the BA distribution (Fig. 4) reveals that the
 300 overwhelming majority of fires in the dataset fall into the <1 ha category. Preserving these highly
 301 localized events substantially increases the representation of fire frequency at the small-event
 302 end of the fire-size distribution, providing information that is difficult to obtain from burned area
 303 products alone.



304

305 **Figure 4.** Distribution of historical wildfires categorized by total burned area (ha). The high
 306 frequency of small fires (<1 ha) highlights the dataset's value for assessing regional fire frequency,
 307 as these small and often short-lived events are more likely to be underrepresented in coarse-
 308 resolution burned-area products.

309 4 Discussion

310 As wildfire regimes in the Eastern Mediterranean intensify under a warming climate, accurate
 311 historical baselines are essential for predictive modelling, resource allocation, and ecological
 312 risk assessment (Ascoli et al. 2021; Trucchia et al. 2023). National wildfire archives are
 313 indispensable in this context; they provide an exceptionally high-resolution picture of regional fire
 314 activity and provide valuable reference data for evaluating and validating global satellite products
 315 (Larkin et al. 2020; McClure et al. 2023). However, their susceptibility to human clerical error and
 316 geographic drift during manual data entry often compromises their direct integration into
 317 statistical and scientific research.

318 Therefore, the creation of this spatially verified national wildfire archive represents a critical step
 319 in transforming fragmented administrative logs into a robust, research-grade geospatial resource.
 320 By systematically resolving clerical errors, mainly geographic drift, the validation pipeline
 321 presented here recovered thousands of wildfire records that would have otherwise been
 322 excluded from spatial analyses. However, the transition from operational, manually entered data
 323 to a standardized scientific product inherently introduces specific structural nuances, detailed
 324 below, that end-users must consider.



325 **4.1 Dataset Uncertainties and Limitations**

326 While the data validation pipeline significantly enhances the spatial reliability of the historical
327 wildfire archive, researchers using this dataset must be aware of inherent uncertainties
328 stemming from the integration of manually inputted administrative records with satellite
329 detections of active wildfires.

330 **4.1.1 The Imputed Records and Micro-Spatial Modelling**

331 The most critical limitation pertains to the 927 unrecoverable records designated as
332 *Imputed_On_Land_Surface*. These records were subjected to terrestrial imputation to preserve
333 statistical information (such as total provincial burned area, temporal frequencies, and ignition
334 causes) which would have otherwise been lost to human clerical errors. However, because their
335 final coordinates are spatially randomized placeholders within an administrative boundary, they
336 do not represent true physical ignition points, and therefore may lead to erroneous results in high
337 resolution hotspot analyses or micro-spatial modelling (such as fuel-type ignition probability).
338 End-users conducting high-resolution topographic or pixel-level spatial modelling are therefore
339 strongly advised to filter the dataset using the *data_source* column to exclude imputed records.

340 **4.1.2 Spatial Tolerance and Geographic Drift**

341 Records flagged as *OGM_Border_Accepted* constitute those that successfully intersected a valid
342 terrestrial district polygon within a 10 km search radius. While this threshold was selected in
343 consultation with OGM officials to account for natural GPS drift, cross-border wildfire spread,
344 and complex topography, these coordinates may still possess minor spatial inaccuracies up to
345 the 10 km limit. Additionally, since maritime drift records were geographically snapped to the
346 nearest shoreline to preserve terrestrial validity, they do not represent the precise ignition point.
347 Therefore, end-users requiring absolute coordinate fidelity can isolate or exclude these specific
348 records using the *coastal_adjustment* boolean flag. Similarly, records categorized as
349 *VIIRS_Rescued* are subject to the nominal 375 m spatial resolution of the VIIRS Active Fires /
350 Thermal Anomalies product, meaning that the corrected coordinate represents the location of
351 the detected thermal anomaly rather than the exact ignition point.

352 **4.1.3 Temporal Gaps**

353 The dataset contains a known temporal limitation regarding diurnal data. While *ignition_date* is
354 complete across the entire 13 yr period, the *ignition_time* is absent for the years 2019 and 2020
355 due to recording protocol changes in the raw national archive during those years. These missing
356 values have been explicitly standardized as NA in the final dataset. Researchers focusing on the



357 diurnal timing of ignitions or hourly fire weather indices should exclude 2019 and 2020 from their
 358 hourly calculations, or, where appropriate, use the satellite-derived *viirs_frp* timestamp for
 359 rescued records during those years.

360 By utilizing the *data_source* provenance column, researchers can subset the dataset, balancing
 361 the trade-off between absolute spatial precision and comprehensive statistical coverage
 362 depending on their specific analytical needs.

363 4.2 The "Micro-Fire" Detection Gap

364 While automated global satellite products, such as the widely used MODIS (MCD64A1) and VIIRS
 365 burned area products, provide invaluable data for global fire modelling, their ability to capture
 366 small fires is constrained by spatial resolution, observation timing, canopy and cloud
 367 obscuration, and their short duration (Chuvieco et al. 2022). Consequently, small burned areas
 368 can be underrepresented or missed by satellite-derived burned-area products, particularly at the
 369 sub-hectare scale (Giglio et al. 2018; Gincheva et al. 2024).

370 By grounding this inventory in geographically validated administrative records from OGM, the
 371 finalized dataset helps address this “micro-fire” detection gap. As demonstrated by the
 372 distribution of burned area (Fig. 4), the overwhelming majority of historical wildfires in Türkiye fall
 373 within the sub-hectare category, highlighting a component of fire occurrence that may be
 374 underrepresented in coarse-resolution satellite burned-area products. Preserving these
 375 localized events provides an important resource for researchers investigating fire frequency and
 376 small-scale spatial patterns that cannot be fully resolved from burned-area products alone.
 377 Consequently, this dataset provides a valuable resource for micro-spatial studies, including
 378 Wildland-Urban Interface (WUI) vulnerability assessments and analyses of localized
 379 anthropogenic ignition patterns.

380 4.3 Broader Implications for Mediterranean Fire Science

381 Beyond resolving local administrative inaccuracies, this dataset provides a critical missing link
 382 for broader Mediterranean fire ecology and climate change research. Regional monitoring
 383 systems, such as the European Forest Fire Information System (EFFIS), recognize the
 384 complementary value of administrative fire information supplied at the national level and
 385 satellite-derived observations, maintaining these data streams separately within the European
 386 Fire Information Framework (Camia et al. 2014; San-Miguel-Ayanz et al. 2023). At a different level
 387 of spatial and temporal granularity, the present dataset complements such regional systems by
 388 providing an event-level, geographically validated release of Türkiye’s administrative wildfire



389 records, with explicit coordinate provenance. This provides researchers with a detailed national
390 reference dataset for evaluating satellite-derived products and investigating localized fire
391 dynamics that may be underrepresented in coarser-resolution observations.

392 Furthermore, as persistent long-term drought and global warming push extreme fire weather
393 beyond traditional boundaries (Keeping et al. 2026), having precise historical baselines becomes
394 paramount. This dataset captures not only the dense fire activity in the classic Mediterranean and
395 Aegean sclerophyllous ecosystems, but also tracks the emerging shift of wildfire activity into
396 Türkiye's northern temperate rainforests that have historically exhibited low fire activity and are
397 currently facing novel climate vulnerabilities (Ekberzade et al. 2026; Keeping et al. 2026).
398 Consequently, this archive serves as a foundational tool for modelling future Eastern
399 Mediterranean fire regimes and developing proactive climate adaptation strategies.

400 **5 Conclusion**

401 The integration of national administrative wildfire archives into Earth system modelling and
402 spatial fire research is often limited by human clerical errors, geographic drift, and inconsistent
403 formatting. This study demonstrates that these operational limitations can be systematically
404 overcome through a rigorous, algorithmically driven spatial validation framework. By cross-
405 referencing over 36,000 historical records from the Turkish General Directorate of Forestry (OGM)
406 against civil administrative boundaries and NASA VIIRS active-fire detections, this methodology
407 successfully recovered thousands of severely mislocated fire records while retaining 99.98% of
408 the original archive for statistical analysis. Crucially, the pipeline preserved the small fires (<1 ha)
409 that dominate the national records but may be underrepresented in automated global burned-
410 area products.

411 The resulting 2013-2025 dataset provides a comprehensive and spatially verified wildfire
412 inventory for Türkiye. By bridging administrative fire records with independent satellite
413 observations and explicit coordinate provenance, this open-access archive provides a valuable
414 resource for advanced spatial modelling, wildland-urban interface (WUI) risk assessments, and
415 climate-driven fire regime projections. Ultimately, this study demonstrates that with careful
416 spatial curation, historical administrative logs can be transformed into highly reliable, research-
417 grade geospatial resources capable of advancing global fire science.

418

419

420



421 Appendix A

422 **Table A1.** Data dictionary for the finalized historical wildfire dataset (2013-2025). *Note: In the
423 original OGM archive, records for the years 2019 and 2020 lack reported ignition times; therefore,
424 these entries were populated with NA to maintain a continuous, standardized data structure
425 across the final dataset.

Column Name	Data Type	Description
fire_id	Integer	Unique sequential identifier assigned to each fire record
ignition_date	Date	The verified date of fire ignition (YYYY-MM-DD)
ignition_time*	Character	The reported local time of fire ignition (HH:MM)
province	Character	Standardized, ASCII-normalized province name
district	Character	Standardized, ASCII-normalized administrative district name
ignition_cause	Character	The reported cause of the fire (e.g., natural/lightning, accident, intentional)
burned_area_ha	Numeric	The total area burned, recorded in ha.
original_lon	Numeric	The unedited longitude coordinate derived from the raw archive (decimal degrees, WGS84)
original_lat	Numeric	The unedited latitude coordinate derived from the raw archive (decimal degrees, WGS84)
final_lon	Numeric	The verified, corrected, or imputed longitude coordinate used for final geographic placement
final_lat	Numeric	The verified, corrected, or imputed latitude coordinate used for final geographic placement
data_source	Character	Provenance flag indicating how the final coordinate was derived: OGM_Original, VIIRS_Rescued, OGM_Border_Accepted, Imputed_On_Land_Surface
coastal_adjustment	Character	Boolean flag indicating whether or not there has been coastal adjustment applied to the major spatial errors
viirs_frp	Numeric	Fire Radiative Power (FRP; MW) extracted from NASA VIIRS data, populated only for the rescued records

426

427

428



429 **Data Availability**

430 The spatially verified final historical wildfire dataset for Turkey (2013–2025) is publicly available
 431 via Zenodo at <https://doi.org/10.5281/zenodo.21887740> (Ekberzade, 2026b; embargoed until
 432 publication with reviewer access granted via: <https://tinyurl.com/3em3muye>). The repository
 433 includes the data in tabular (.csv) and spatial (.gpkg) formats, alongside a ReadMe file including
 434 a comprehensive data dictionary. The custom R code and spatial validation algorithms developed
 435 for the data rescue pipeline are provided as supplementary material and within the Zenodo
 436 repository to ensure full methodological reproducibility. The dataset is distributed under a
 437 Creative Commons Attribution 4.0 International (CC BY 4.0) license.

438 **Author contributions**

439 BE conceived and designed the methodology, acquired and validated the data, ran all quality
 440 controls, and wrote the manuscript.

441 **Competing interests**

442 The author declares no competing interests

443 **Acknowledgements**

444 The author would like to sincerely thank the officials at the Department of Combating Forest Fires
 445 within the General Directorate of Forestry of Türkiye (Orman Genel Müdürlüğü, Orman
 446 Yangınlarıyla Mücadele Daire Başkanlığı) for providing the raw OGM data files and for their
 447 continuous support and feedback during the development of this methodology.

448 **Financial support**

449 Research conducted for this study has been made possible by grant number MGA-2026-47431.

450 **References**

- 451 Ascoli, D., Moris, J., Marchetti, M., Sallustio, L.: Land use change towards forests and wooded
 452 land correlates with large and frequent wildfires in Italy. *Annals of Silvicultural Research*, 46(2),
 453 177-188, <https://doi.org/10.12899/asr-2264>, 2021.
- 454 Atalay, H., Dervisoglu, A., Sunar, A.F.: Exploring Forest Fire Dynamics: Fire Danger Mapping in
 455 Antalya Region, Türkiye. *ISPRS International Journal of Geo-Information*, 13(3), 74,
 456 <https://doi.org/10.3390/ijgi13030074>, 2024.
- 457 Camia, A., Durrant, T., San-Miguel-Ayanz, J.: The European fire database: technical specifications
 458 and data submission: executive report, Publications Office.
 459 <https://data.europa.eu/doi/10.2788/2175>, 2014.
- 460 Chuvieco, E., Roteta, E., Sali, M., Stroppiana, D., Boettcher, M., Kirches, G., Storm, T., Khairoun,
 461 A., Pettinari, M. L., Franquesa, M., and Albergel, C.: Building a small fire database for Sub-Saharan



- 462 Africa from Sentinel-2 high-resolution images, *The Science of the Total Environment*, 845,
 463 157139, <https://doi.org/10.1016/j.scitotenv.2022.157139>, 2022.
- 464 Ekberzade, B., Yetemen, O., Sen, O. L., & Dalfes, H. N.: Simulating the potential forest ranges in
 465 an old land: the case for Turkey's forests. *Biodiversity and Conservation*, 31(13-14), 3217-3236,
 466 <https://doi.org/10.1007/s10531-022-02485-8>, 2022.
- 467 Ekberzade, B., Yetemen, O., Ezber, Y., Sen, O. L., & Dalfes, H. N.: Latitude or altitude as the future
 468 refugium? A case for the future of forests in Asia Minor and its surroundings. *Ecology and*
 469 *Evolution*, 14, e11131, <https://doi.org/10.1002/ece3.11131>, 2024.
- 470 Ekberzade, B., Görüm, T., Karabacak, F. et al.: Up in flames: the human factor behind a megafire
 471 in Mediterranean Türkiye. *npj Nat. Hazards* 2, 65, <https://doi.org/10.1038/s44304-025-00120-4>,
 472 2025.
- 473 Ekberzade, B.: Forecasting the Flame: Simulating Future Wildfire Regimes in the Northeast
 474 Mediterranean, *EGUsphere* [preprint], <https://doi.org/10.5194/egusphere-2026-1498>, 2026a.
- 475 Ekberzade, B.: A Geographically Validated Administrative Wildfire Inventory for Turkey (2013–
 476 2025) [Dataset] Zenodo. <https://doi.org/10.5281/zenodo.21887740>, 2026b.
- 477 Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS
 478 burned area mapping algorithm and product, *Remote Sensing of Environment*, 217, 72–
 479 85, <https://doi.org/10.1016/j.rse.2018.08.005>, 2018.
- 480 Giglio, L.: *VIIRS/NPP Active Fires 6-Min L2 Swath 375m V002*. 2024, distributed by NASA EOSDIS
 481 Land Processes Distributed Active Archive Center, [online] available from:
 482 <https://doi.org/10.5067/VIIRS/VNP14IMG.002> (Accessed 6 June 2026), 2024
- 483 Gincheva A, Pausas JG, Edwards A, Provenzale A, Cerdà A, Hanes C, Royé D, Chuvieco E, Mouillot
 484 F, Vissio G, Rodrigo J, Bedía J, Abatzoglou JT, Senciales González JM, Short KC, Baudena M, Llasat
 485 MC, Magnani M, Boer MM, González ME, Torres-Vázquez MÁ, Fiorucci P, Jacklyn P, Libonati R, Trigo
 486 RM, Herrera S, Jerez S, Wang X, Turco M.: A monthly gridded burned area database of national
 487 wildland fire data. *Sci Data*. 11(1):352, <https://doi.org/10.1038/s41597-024-03141-2>, 2024.
- 488 Hall, J. V., Loboda, T. V., Giglio, L., and McCarty, G. W.: A MODIS-based burned area assessment
 489 for Russian croplands: Mapping requirements and challenges, *Remote Sensing of Environment*,
 490 184, 506–521, <https://doi.org/10.1016/j.rse.2016.07.022>, 2016.
- 491 Keeping, T. R., Zachariah, M., Haas, O., Grillakis, M., Barnes, C., Clay, G. D., Ekberzade, B., Jaupaj,
 492 O., Ribeiro, A., Trigo, R., Voulgarakis, A., and Otto, F. E. L.: Attribution of the record breaking 2025



- 493 European fire season to climate change, *Nat. Hazards Earth Syst. Sci.*, 26, 3761–3813,
494 <https://doi.org/10.5194/nhess-26-3761-2026>, 2026.
- 495 Larkin, N. K., Raffuse, S. M., Huang, S., Pavlovic, N., Lahm, P., and Rao, V.: The Comprehensive
496 Fire Information Reconciled Emissions (CFIRE) inventory: Wildland fire emissions developed for
497 the 2011 and 2014 U.S. National Emissions Inventory, *Journal of the Air & Waste Management*
498 *Association*, 70, 1165–1185, <https://doi.org/10.1080/10962247.2020.1802365>, 2020.
- 499 McClure CD, Pavlovic NR, Huang S, Chaveste M, Wang N.: Consistent, high-accuracy mapping of
500 daily and sub-daily wildfire growth with satellite observations. *International Journal of Wildland*
501 *Fire* 32, 694–708, <https://doi.org/10.1071/WF22048>, 2023.
- 502 OCHA: Türkiye - Subnational Administrative Boundaries, [online] Available from:
503 <https://data.humdata.org/dataset/cod-ab-tur> (Accessed 6 June 2026), 2025
- 504 OGM: 2020 Türkiye Orman Varlığı, Tarım ve Orman Bakanlığı Orman Genel Müdürlüğü Orman
505 İdaresi ve Planlama Dairesi Başkanlığı, [online] Available from: [https://](https://www.ogm.gov.tr/tr/ormanlarimiz-sitesi/TurkiyeOrmanVarligi/Yayinlar/2020%20Türkiye%20Orman%20Varligi.pdf)
506 [www.ogm.gov.tr/tr/ormanlarimiz-](https://www.ogm.gov.tr/tr/ormanlarimiz-sitesi/TurkiyeOrmanVarligi/Yayinlar/2020%20Türkiye%20Orman%20Varligi.pdf)
507 [sitesi/TurkiyeOrmanVarligi/Yayinlar/2020%20Türkiye%20Orman%20Varligi.pdf](https://www.ogm.gov.tr/tr/ormanlarimiz-sitesi/TurkiyeOrmanVarligi/Yayinlar/2020%20Türkiye%20Orman%20Varligi.pdf) (Accessed 6 June
508 2026), 2020
- 509 OGM: Ormancılık İstatistikleri, Tarım ve Orman Bakanlığı Orman Genel Müdürlüğü Tarım Reformu
510 Genel Müdürlüğü, [online] Available from: [https:// istatistik.tarimorman.gov.tr/Sayfa/Detay/2045](https://istatistik.tarimorman.gov.tr/Sayfa/Detay/2045)
511 (Accessed 6 June 2026), 2025
- 512 San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Liberta, G., Jacome Felix Oom, D., Branco,
513 A., De Rigo, D., Suarez-Moreno, M., Ferrari, D., Roglia, E., Scionti, N., Broglia, M., Onida, M.,
514 Tistan, A. and Loffler, P.: Forest Fires in Europe, Middle East and North Africa 2023, Publications
515 Office of the European Union, Luxembourg, 2024, <https://doi.org/10.2760/8027062>, JRC139704,
516 2023.
- 517 Schroeder, W., Oliva, P., Giglio, L., and Csiszar, I. A.: The New VIIRS 375 m active fire detection
518 data product: Algorithm description and initial assessment, *Remote Sensing of Environment*,
519 143, 85–96, <https://doi.org/10.1016/j.rse.2013.12.008>, 2014.
- 520 Short, K. C.: A spatial database of wildfires in the United States, 1992–2011, *Earth Syst. Sci. Data*,
521 6, 1–27, <https://doi.org/10.5194/essd-6-1-2014>, 2014.



- 522 Taylor, S.W., Woolford, D.G., Dean, C.B., Martell, D.L.: "Wildfire Prediction to Inform Fire
523 Management: Statistical Science Challenges." *Statist. Sci.* 28 (4) 586 - 615,
524 <https://doi.org/10.1214/13-STS451>, 2013.
- 525 Trucchia A, Meschi G, Fiorucci P, Provenzale A, Tonini M, Pernice U.: Wildfire hazard mapping in
526 the eastern Mediterranean landscape. *International Journal of Wildland Fire* 32, 417–
527 434. <https://doi.org/10.1071/WF22138>, 2023.