

International Association of Wildland Fire

An improved spatio-temporal clustering method for extracting fire footprints based on MCD64A1 in the Daxing'anling Area of north-eastern China

Huiyi Su^A^(D), Xiu Ma^{B,C} and Mingshi Li^{A,B,*}

For full list of author affiliations and declarations see end of paper

*Correspondence to: Mingshi Li College of Forestry, Nanjing Forestry University, Nanjing 210037, China Email: nfulms@njfu.edu.cn

ABSTRACT

Background. Understanding the spatio-temporal dynamics associated with a wildfire event is essential for projecting a clear profile of its potential ecological influences. **Aims.** To develop a reliable framework to extract fire footprints from MODIS-based burn products to facilitate the understanding of fire event evolution. **Methods.** This study integrated the Jenks natural breaks classification method and the density-based spatial clustering of applications with noise (DBSCAN) algorithm to extract the fire footprints in Daxing'anling region of China between 2001 and 2006 from MCD64A1 burned area data. **Key results.** The results showed that the fire footprints extracted by the model gained an overall accuracy of 80% in spatial and temporal domains after an intensive validation by using the historical fire records provided by the local agency. The agreement of burned area between the extracted fire patches and the historical fire records for those matched fire points was characterised by an overall determination coefficient R^2 at 0.91. **Conclusions.** The proposed framework serves as an efficient and convenient wildfire management tool for areas requiring large-scale and long-term wildfire monitoring. **Implications.** The current framework can be used to create a reliable large-scale fire event database by providing an important alternative for the improvement of field investigation.

Keywords: clustering, Daxing'anling, DBSCAN, fire footprint, Jenks natural breaks, MCD64A1, remote sensing, wildfire.

Introduction

As one of the largest potential mechanisms to release carbon from forest ecosystems, wildfires play a pivotal role in determining both the forest ecosystem's function and structure (Gruber et al. 2004). The frequency, extent and intensity of fires are drivers of regional ecosystem's evolutions (Goldammer and Furyaev 1996; Van Mantgem et al. 2011), affecting the carbon and nitrogen cycles as well as the energy balance and climate change (Grogan et al. 2000; Brockway et al. 2002; Girardin and Mudelsee 2008). Traditional field investigations to document the information on fires are timeconsuming, labour-intensive, accessibility-limited and extremely dangerous. Remotesensing technology has the characteristics of large-scale observation and diverse spatial and temporal resolutions, which can effectively describe the surface process in different levels of detail and offset the deficiency of reported statistical data in informatisation and spatialisation of fire management, providing an effective technical means for studying wildfire (Chuvieco and Congalton 1989; Peng et al. 2007; Mitri and Gitas 2013; Mallinis et al. 2018). Moderate Resolution Imaging Spectroradiometer (MODIS) has been widely used for fire detection, fire risk mapping and post-fire vegetation restoration assessment in large areas due to its advantages of high temporal resolution, medium spatial resolution for land surface observation application and free accessibility (Loboda et al. 2007; Van Leeuwen 2008; Hope et al. 2012; Jin et al. 2012; Maier et al. 2013). One of the prevalent satellite-based global burned area products available today is the MODIS

Received: 28 September 2022 Accepted: 29 March 2023 Published: 26 April 2023

Cite this:

Su H et al. (2023) International Journal of Wildland Fire 32(5), 679–693. doi:10.1071/WF22198

© 2023 The Author(s) (or their employer(s)). Published by CSIRO Publishing on behalf of IAWF. This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND)

OPEN ACCESS

MCD64A1 burned area product, which is based on daily surface reflectance imagery from the sun-synchronous, polar-orbiting MODIS Aqua and Terra multispectral satellite constellations with a 2-day observation frequency (Giglio et al. 2009). The MCD64A1 burned area product uses near daily MODIS surface reflectance imagery trained with the MODIS active fire product. It has been extensively validated in several regions of the world and is also regularly subject to algorithm refinements (Boschetti et al. 2008; Roy et al. 2008; Schroeder et al. 2008a; Roy and Boschetti 2009). Most studies require appropriate burn data to assist the refinement. The commonly used burn data are often based on pixel information statistics rather than the fire event itself. This makes it impossible to study specific independent fire events in a large area and misses some important information due to not taking full advantage of the large-scale monitoring of remote-sensing technology.

Fire Footprint refers to the spatial and temporal information of each Fire event. Winoto-Lewin et al. (2020) determined the vegetation regeneration, maturation and fire incidence and severity across forest types, following wildfires in January and February 2019, by setting a random point survey in the fire footprint. French et al. (2016) surveyed the fire footprint, and an adjacent management burn, to investigate the drivers of fire severity in sedge-heathland and to assess the regeneration response of woody vegetation and how these were influenced by antecedent fire histories. Bird and Cali (1998) analysed fire footprints from the accumulation of several years of landscape use visible on aerial photographs. Currently, the acquisition of fire footprint information still relies mainly on field surveys, which are time-consuming, labour-intensive and difficult to implement in large areas and remote areas, and are prone to fire footprint's information loss or omission and inaccuracy.

Density Based Spatial Clustering of Applications with Noise (DBSCAN) is an outstanding clustering method well known for its ability to identify groups of arbitrary shapes and deal with noisy datasets (Ester et al. 1996; Patil and Vaidya 2012), and it has been applied in fire ecology with an emphasis on clustering application models. For example, Artés et al. (2019) used DBSCAN algorithm to cluster hotspot data of Kalimantan Island and South Sumatra Province in Indonesia during 2002-2003. Kristianto et al. (2020) used land surface temperature (LST) data and local agency statistics to cluster the local fire high-risk areas. Sheng et al. (2021) presented a method based on DBSCAN and convolutional neural network to recognise flame and smoke modes connected to fire stages. Vatresia et al. (2020) proposed a spatio-temporal clustering method with DBSCAN to cluster hotspot data over Sulawesi Island from 2016 to 2018. Usman et al. (2015) studied the distribution pattern of hotspot data in Sumatra Island from 2002 to 2013 based on DBSCAN clustering algorithm. However, there are few studies on clustering based on burned pixels with temporal attributes to obtain fire footprints.

The Jenks optimisation method, also called the Jenks natural breaks classification method (here after referred to as Jenks), is a data classification method designed to determine the best arrangement of values into different classes. Previous studies have proved that Jenks has good adaptability and high accuracy in the geographical environment unit division (Chen et al. 2013; Yalcin and Kilic Gul 2017; Zhou et al. 2020). Jaafari et al. (2019) used the Jenks method to classify the probability values of fire risk and generated the distribution maps of five different wildfire occurrence grades. Anchang et al. (2016) proposed an effective urban vegetation-mapping method based on IKONOS image by combining Jenks method with Agglomerative Hierarchical clustering (AHC) algorithm. Based on these backgrounds, the Jenks optimisation method can be used to classify burned pixels to maximise the difference of combustion time attributes among groups, and then divide one image into multiple pieces for clustering optimisation parameters to make full use of burn date information.

Considering the strengths and limitations highlighted in the above techniques, the purpose of this study was to combine DBSCAN and Jenks to develop a framework that can extract fire footprints from remote-sensing-based fire products, to accurately extract the location of fire events case-bycase in a large area and to obtain the exact occurrence time and duration of fire events. Daxing'anling is China's major fire management area. The forest ecosystem in Northeast China holds a large carbon pool and plays an important role in mitigating global climate change (Xing et al. 2015; Zhang et al. 2016). We used MODIS MCD64A1 data from 2001 to 2006 to extract fire footprints in the Daxing'anling, then verified fire footprints according to the historical fire records provided by the local institutions. It is expected that this fire footprint extraction method would not only help to systematically assess the burning mechanisms of fires in the region, but also contribute to fire prevention and suppression policy development, as well as to carbon emission accounting from the historical fire events.

Material and methods

Study area

The Daxing'anling area has a geographical span of $50^{\circ}10'-53^{\circ}33'$ N, $121^{\circ}12'-127^{\circ}0'$ E, which covers 8.3×10^{4} km². It is located in the northwestern Heilongjiang Province and northeastern Inner Mongolia Autonomous Region of Northeast China (Fig. 1). The average elevation of the study area is 573 m, and the elevation of the highest hill is 1528 m above sea level, with gentle slopes < 15° accounting for >80% of the area. The terrain is undulating and is higher in the western and central parts and lower in the eastern, northern and southern parts. Forest cover of the area is about 76.3%. Dominant tree species include *Larix gmelinii, Betula platyphylla* and *Pinus sylvestris var. mongolica*.



Fig. 1. (a) Location of the Daxing'anling partial region in China. (b) Administrative division map of the study area. (c) Elevation map of partial Daxing'anling region.

According to the meteorological administration of China, the annual mean temperature of the study area is -3° C, with the lowest temperature at -48° C and the highest temperature at 36° C. The frost-free period is 90–110 days, and the annual precipitation ranges from 400 to 550 mm, 2/ 3 of which occurs from June to August. The study area has a temperate continental monsoon climate of cool and humid summers, with southeast airflows from the Pacific highpressure system and cold and dry winters under the control of the Mongolia high pressure system (Tian *et al.* 2011). The annual average wind speed is 2 m/s, with a maximum 7–8 m/s that often occurs in spring (Tian *et al.* 2013). The details on the study area were obtained from the Daxing'anling District Administration Office (http://www. dxal.gov.cn/rwxa (accessed on 24 May 2022)).

Data and preprocessing

Remote-sensing-based MCD64A1 and MCD12Q1 products

We used the Collection 6 MCD64A1 product in this study, with a spatial resolution of 500 m. The product is based on a dynamic threshold applied to the surface reflectancedependent burn-sensitive vegetation index (VI) (Eqn 1) and incorporates temporal variability, which is guided by the MODIS active fire hotspot locations (Roy *et al.* 2008; Giglio *et al.* 2009; Humber *et al.* 2019). The data layers include burn date, burn data uncertainty, quality assurance, and first Day and last Day of reliable change detection of the year. The MCD64A1 product was used in this study because it is the reference product for burnt area monitoring that is publicly available. The layer information we used for this analysis was the burn date of each pixel in the day of year (DOY).

$$VI = \frac{\rho_5 - \rho_7}{\rho_5 + \rho_7}$$
(1)

where ρ_5 and ρ_7 are the surface reflectance of MODIS short wavelength infrared band 5 and band 7 respectively.

The 500 m MCD12Q1 characterises five Global Land Cover classification systems. It depicts the Land Cover types by spanning a year's input of observation data from the Terra and Aqua satellites. In this work, Land Cover Type 1 products classified by IGBP (International Geosphere-Biosphere Programme) theme were selected. A total of 17 land cover classes are specified, including 11 natural vegetation classes, three developed and mosaicked land classes, and three non-vegetated land classes. According to the land cover type corresponding to the pixel value of data products (Sulla-Menashe and Friedl 2018), it is assumed that the pixel classification codes larger than 10 correspond to those noncombustible types in this study for mask preparation.

The MODIS burned area product (MCD64A1) and land cover product (MCD12Q1) are available from the United States Geological Survey (USGS) (https://lpdaac.usgs.gov/ products/). The product numbers of MODIS dataset required for the study area are h25v03 and h26v03. In terms of data preprocessing, the MODIS Reprojection Tool (MRT) provided by NASA's EOSDIS (Earth Observing System Data and Information System) was used to perform format conversion, re-projection and image mosaicking from the original hierarchical data format (HDF) with the sinusoidal (SIN) projection, then to extract the required data layers from the transformed data.

Validation data

The validation data used in this study were provided by the Shenyang Institute of Applied Ecology, Chinese Academy of Sciences. The data consisted of all fire events recorded during 2001-2006 in the Daxing'anling region, including the central latitude and longitude coordinates of the fire events, the estimated burned area and the fire occurrence time documented in specific month. The burned area of the fire was visually measured on the ground by local investigators or professionals after the fire, and the latitude and longitude of the fire location were obtained by monitoring towers and patrols in the forest area using GPS instruments to record the position of smoke and flames. We compiled the locational coordinates of each fire into the database in ArcMap package, and transformed the projection information to co-register the MODISderived products. These vectorised fire points were used for subsequent model validations in this study.

Because the spatial resolution of MODIS-based products is 500 m and the corresponding land cover area of each burned pixel is 25 ha, for verification purposes we only used those fire events with burned areas larger than 25 ha in the fire database. According to the coordinates of each fire event provided by the local agency, it is considered that the fire occurrence point recorded by manual work may have a positional error, and the transformation of projection mode would lead to a certain degree of position deviation when it is re-projected. In order to ensure the matching between field fire points and burned patches extracted from MODIS-based products during the verification process, we took each vector point as the centre and established a buffer with a radius of 1 km. If the buffer overlapped with certain clustered burning patch, it would be determined as a coincided fire event. If multiple burned patches coincided with the buffer, the patch with the largest overlapped pixels was defined as the coincided or matched fire event.

Methodology

Based on the spatial location of MCD64A1 burned pixels and the burn date information represented by the pixel value, this study proposed an improved spatial-temporal clustering method to extract fire footprints. This method was mainly divided into two steps, namely time classification and spatial clustering (Fig. 2). Firstly, the classification model was used to classify the annual burning pixels in the study area into three fire periods based on the burn date, then different parameters were selected for three fire periods for spatial clustering to finally obtain the fire footprints.



Fig. 2. Research framework of the study.

Classification method of fire occurrence period

The Jenks method seeks to minimise the average deviation from the class mean in each class while maximising the deviation among different classes, which means that the method reduces the variance within classes and maximises the variance between classes (McMaster 1997). This grouping method divides the data into classes, and for those classes, the boundaries (thresholds) are set at locations where the data values are relatively different.

Step 1: Calculating the sum of squaring deviations from the mean of the complete dataset (SDAM) (Eqn 3) for the array of a certain class in the classification results, and denoting an array as A, whose mean value \bar{X} is (Eqn 2):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2}$$

$$SDAM = \sum_{i=1}^{n} (X_i - \bar{X})^2$$
(3)

where *n* is the number of elements in array A, and X_i is the value of the *i*-th element.

Step 2: Getting all cases of datasets classification after the breakpoint combination, and the sum of squared deviations from the class means (SDCM) calculated for each combination situation to select the smallest of them. Dividing *n* elements into *k* classes, so that *k* subsets can be obtained, for example $[X_1 X_2 \cdots X_i], [X_{i+1} X_{i+2} \cdots X_j], \cdots [X_{j+1} X_{j+2} \cdots X_n]$, calculating the sum of the total deviation squares of each subset SDCM_{*i*}, SDCM_{*j*}, ..., SDCM_{*n*}, and summing up them to get SDCM₁ (Eqn 4):

$$SDCM_1 = SDCM_i + SDCM_i + \dots + SDCM_n$$
 (4)

International Journal of Wildland Fire

In the case of other combinations divided into k classes, the value of SDCM₂, ..., SDCM_{C_n^k} can be calculated successively in the same way, and the smallest value is selected as the final result SDCM_{min}. Therefore, this classification range is the best classification.

Step 3: Calculating the gradient goodness of variance fit (GVF) of each classification (Eqn 5):

$$GVF_i = \frac{SDAM - SDCM_i}{SDAM}$$
(5)

the GVF_i ranges from 0 (worst fit) to 1 (perfect fit). The higher the gradient is, the larger the difference between classes, and the test proves that the SDCM_{\min} in step 2 is adopted. When the gradient value is the largest, it can be concluded that the classification result of Jenks classification method is most ideal.

The algorithm is commonly used in Geographic Information Systems (GIS) applications. In this study, we implemented the Jenks algorithm based on ArcGIS software by using a raster reclassification tool in the ArcMap Spatial Analyst Toolbox. Then, the MCD64A1 burn date data of each year were reclassified according to the breakpoint value obtained by using the grid calculator tool, and the spatial information of each burned pixel was retained to obtain the combustion map of the three periods. The following labels were assigned to the three unlabelled classes by ranking the number of burned pixels: high frequency fire period; medium frequency fire period; and low frequency fire period.

Clustering algorithm and parameter tuning

DBSCAN requires two basic parameters including ε (Eps) and minPoints (MinPts). Eps describes the minimum distance between two points that can be considered as neighbours. MinPts is the minimum number of points to form a group region (Ester 2009). More simply, the step-by-step description of the DBSCAN algorithm is as follows:

Step 1: Find the neighbourhoods of every point with the parameter Eps;

Step 2: Identify the core points that have more neighbours than MinPts;

Step 3: Find components of core points connected with the neighbour graph;

Step 4: Assign non-core points to the nearby cluster with nearest Eps; if there is no nearby cluster meeting the Eps, assign them to noise.

DBSCAN algorithm performance is very sensitive to Eps and MinPts parameters, and improper values of Eps and MinPts will lead to poor or even incorrect clustering effect. If MinPts remains stable, too-large Eps value will lead to most points being clustered in the same cluster, and too-small Eps will lead to the split of a cluster. If Eps is unchanged and the value of MinPts is too large, the midpoint of the same cluster will be marked as an outlier, or two neighbouring clusters with high density may be merged into one cluster; if the MinPts value is too small, it will lead to the discovery of a large number of core points (Sander et al. 1998). To tune the two parameters properly, the K-distance (k-dist) graph was used in the study. The specific principle of the *k*-dist graph is as follows: set a dataset $P = \{p(i); i = 0, 1, ..., n\}$. For any point p(i), calculate the distance between point p(i) and all the points in subset $S = \{p(1), p(2), ..., p(i-1), p(i+1), ..., p\}$ (*n*)} of set *D*. The distances are sorted in the ascending order. Assuming that the sorted distance set $D = \{d(1), d(2), ..., d\}$ (k - 1), d(k), d(k + 1), ..., d(n), then d(k) is called k-dist. The *k*-dist should be calculated for each point *p*(*i*) in the set *P*, and the k-dist graph of the clustering model can be drawn by ordering the k-dist values in the descending manner. For datasets with high density level, the variation depends on the density of the cluster and the random distribution of points, but for points with the same density level, the range of variation will not be large, and there may be large variation between two density levels (Liu et al. 2007). Therefore, the kdist graph is a smooth curve consisting of line segments composed with different d(k). The density level of noise points will be obviously different from that of clustered points, which can be distinguished by the maximum slope change position of the curve (Fig. 3).

In the *k*-dist graph, the distance corresponding to the point with the largest change in slope is determined as Eps value. A guiding principle for the selection of MinPts is that $MinPts \ge dim + 1$, where dim represents the dimension of the data to be clustered (Nisa et al. 2014; Wang et al. 2019). In the MCD64A1 dataset used in this study, the time interval of different burning pixels (DOY difference) is also a dimension for calculating the distance between two points in the clustering model in addition to the horizontal and vertical coordinates. The dimension of data in the study is 3 and the MinPts \geq 4. Combined with the spatial resolution of remote-sensing products, the area of ground objects corresponding to one pixel is 25 ha (500 m \times 500 m) and the area of 4 pixels is 100 ha, which is the area threshold for the fire classification of 'serious' type in The Forest Fire Prevention *Regulations* issued by the Chinese government. Therefore, we set MinPts to 4 for model parameters of all datasets to be clustered, and determined the threshold point one by one against the 4-dist diagram to set the corresponding Eps value.



Number of points sorted on descending by k-distance

Fig. 3. K-dist distribution curve and its interpretation.

This study implemented the DBSCAN clustering algorithm for remote-sensing-detected burned-pixel dataset based on the Geospatial Data Abstraction Library (GDAL) package combined with *sklearn* packages in Python environment.

Model fitting and evaluation

In this study, the silhouette coefficient and noise were used as the evaluation index for the clustering model performance. The *noise%* is the ratio of pixels classified as noise to the pixels clustered as burned pixels. Silhouette coefficient is an index for evaluating the effect of clustering, which can be understood as an index to describe the contour definition of each category after clustering (Rousseeuw 1987). There are two factors involving cohesion and separation. The cohesion reflects the closeness between a sample point and elements in the class, also known as the inner class distance. The separation represents the degree of closeness between a sample point and elements outside the class, also known as the distance between classes. The formula of silhouette coefficient is as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$
(6)

where a(i) represents the cohesion of the sample point, which is the average of the degree of dissimilarity between *i* vector and other points in the same cluster A;, and b(i) is the minimum of the average dissimilarity of the *i* vector to other clusters. The value range of $S(i) \in [-1, 1]$, and the larger the silhouette coefficient, the better the clustering effect. Averaging the silhouette coefficients of all points gives the total silhouette coefficient *S* of the clustering result.

For the obtained fire footprint statistics by DBSCAN, the pixel value with the smallest DOY value in each patch was used as the fire occurrence time and converted to the corresponding month, and the fire burned area was counted by the number of burned pixels. We used several statistical measures to quantify the model performance, including the R^2 , (Eqn 7), the mean absolute error (MAE) (Eqn 8) and the root mean square error (RMSE) (Eqn 9).

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_{i} - y_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}$$
(7)

MAE =
$$\frac{1}{n} \sum_{i=1}^{n} |(\hat{y}_i - y_i)|$$
 (8)

RMSE =
$$\sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_{i} - y_{i})^{2}}{n}}$$
 (9)

where *n* is the number of samples, \hat{y}_i is the burned area predicted by footprint, y_i is the burned area verified by the local survey agency, \bar{y} is the arithmetic mean of all verified burned area.

Results

Jenks classification of DOY

We obtained the frequency distribution of all burned image elements in DOY in the Daxing'anling region from the MODIS burn date dataset during the period 2001-2006 using the Jenks method. Fig. 4 shows the corresponding statistics. Mean represents the average date of occurrence of burning pixels in the current year and St.D reflects the dispersion degree of burning pixels in time. The difference between the two DOY breakpoint values in 2003 and 2006 was smaller compared with the other 4 years, within 28 and 21 days, respectively. The highest standard deviation was observed in 2001. The line graph of the annual variation of the Jenks classification breakpoint values shows that 2003 had the smallest mean DOY value and the smallest two breakpoint values, and the time period covered by the burned pixels was earlier than in other years. The St.D was largest in 2001, indicating that the time distribution of burned pixels in that year was relatively dispersed, and smallest in 2005, indicating that the time distribution of burned pixels in that year was relatively concentrated. The DOY mean in 2003 was smallest, but the DOY mean in the remaining years was greater than 150, indicating that the fires in the study area mainly occurred after June between 2001 and 2006.

According to the above principles of categorical statistics and the obtained DOY breakpoint values, the number of burned pixels in the three periods of each year is summarised in Table 1. Ranking the number of burned pixels in descending order, ** is the high frequency fire periods (HF), * is the medium frequency fire periods (MF) and the rest is the low frequency fire periods (LF). The highest number of pixels was burned in 2003, and the lowest in 2004. After classification, the number of pixels in the HF period ranged from 141 to 21 184, the number of pixels in the MF period ranged from 81 to 17 719 and the number of pixels in the LF period ranged from 38 to 2005. It can also be seen that the HF in almost all years was the third period, except for 2006. That means that most of the burning in 2001–2005 was in the later part of the annual burn date range.



Fig. 4. Annual changes and statistics of burned pixels and breakpoints from 2001 to 2006.

The MCD64A1 burn date dataset in the study area from 2001 to 2006 was classified through the above preparation. The division of the three fire periods in each year is shown in Fig. 5. It can be seen that after the Jenks classification of DOY, pixels in the three burn periods have been clearly distinguished for each year. According to the burning period, the burned pixels of each year were divided into three images for subsequent spatial clustering.

DBSCAN parameter training and fire clustering

With all MinPts setting to 4, k-dist graphs were constructed to identify turning points for Eps parameter selection for all MCD64A1 burn date images in the three periods of each

 Table I.
 Number of MCD64A1 pixels classified according to the breakpoints from 2001 to 2006.

Year	Total number of burned pixels	Period I	Period 2	Period 3
2001	470	125*	82	263**
2002	2836	113	598 *	2125**
2003	40 908	17719*	2005	21 184**
2004	260	38	81*	4 **
2005	1983	87	534*	1362**
2006	1854	1605**	163*	86

* Medium frequency fire periods (MF), ** high frequency fire periods (HF), no asterisk low frequency fire periods (LF).

year after reclassification (Fig. 6). The Eps range of all the clustering models obtained by *k*-dist graph selection was from 0.007 to 0.082, the smallest Eps value at 0.007 was observed in the HF period of 2001 and the highest Eps value at 0.082 was identified for the MF period in 2002. The noise ranged from 0.05 to 13.95%, and most of models had less than 5% noise, except for the LF period in 2001, 2002 and 2006. The silhouette coefficient was almost larger than 0.9, except for the value of 0.821 observed in the MF period of 2003. The number of clusters (denoted as n_clusters in the table) per year was largest in 2003 (39 categories) and smallest in 2004 (11 categories) (Table 2). The number of clusters in the high frequency fire period in 2001 (two categories), and largest in the high frequency fire period in 2003 (14 categories).

Accuracy verification of fire events

We masked out the non-combustible pixels of the above images after Jenks classification and DBSCAN clustering by using the corresponding land cover data of the current year, then counted the pixel number of each fire event to calculate its corresponding area. The fire events provided by the local agency were compared with the fire events obtained by Jenks–DBSCAN models to derive the performance measures. Fig. 7 shows a typical example that reflects the consistency between the burned pixels extracted by the Jenks–DBSCAN model from MODIS products and the realworld fire profiles depicted by higher spatial resolution remote-sensing natural colour composites. Obviously,



Fig. 5. Classifying fire occurrence periods using Jenks algorithm from MCD64AI burn date dataset in the study area.



Fig. 6. 4-dist graph for Eps parameter selection in the three periods of each year from 2001 to 2006.

MODIS fire pixels could match the shapes of post-fire vegetation loss in the true colour images, and the locations of the two fire events also coincided with the clustered fire footprints in space. In terms of multi-temporal higher spatial resolution images, it can also be seen that there is a degree of differentiation of fire areas left by two burning events, with footprint 1 (Fig. 7a) clustering two burning patches into one fire event. Table 3 lists the performance measures of the fitted Jenks-DBSCAN models during the period 2001–2006. Among the 112 fire footprints obtained by our models, 44 of the 55 fire events were in agreement with the actual fire conditions of validation, including not only the space, but also the conversion of the DOY of each fire event to the Gregorian calendar month that could correspond to the validation data. This accounted for 39.28% of the total clustered fire events and 80% of the coincided validation data. But the 2001 modelled results were not consistent with the validation data. The scattergram in Fig. 8 shows that although the overall accuracy derived from the verification data and clustering results of the fire area is good in terms of fitting parameter ($R^2 = 0.91$), there are still many points where the estimated value deviates from the observation value greatly (far from the trendline). From Fig. 9 we can see that most of the burned pixels with complete shape are well clustered into one fire footprint; the burned pixels could be divided into different fire footprints even in

adjacent spaces. In particular, for the two mega-fires in the east (one of which was named '319 prairie-forest fire event'), the local authorities provided nine fire coordinates because of its wide burning range and rapid diffusion, so we combined them into one fire event during accuracy verification, indicated on the map by a black thumbtack symbol.

Discussion

Effect of fire footprints clustering model

By analysing the parameters of DBSCAN model (Fig. 6), we found that the 18 Eps identified in this study are all less than 0.1, which is much lower than other clustering studies (Nisa et al. 2014; Usman et al. 2015) based on hotspot data, but match the range (0.025–0.08) identified in Yanuarsyah et al. (2016). We speculate that variations in the data type and density to be clustered are responsible for the Eps difference. Hotspot is one of the forest and land fire indicators widely used for developing fire early warning system to prevent, suppress and monitor wildfire in areas where there is a high fire risk, such as peat land (Usman et al. 2015). Thus, there may be a high correlation between hot spots and fire points, but in fact not all hot spots indicates fire points (Vatresia et al. 2020). The clustering analysis of fire point data is based on the Geography Tobbler's law (Tobler 2004). However, for burned pixels, due to the occurrence and diffusion principle of wildfire (Thompson and Calkin 2011; Demange et al. 2022), there is a strong spatial connectivity between pixels in a fire footprint - this is why the modelled results in this study have very high silhouette coefficients (>0.9). We also find that bigger noise% corresponds to a relatively small n_cluster, whereas smaller noise% leads to larger n_cluster (Table 2). Therefore, in practical applications, it is necessary to balance the parameters to achieve a better model effect.

Two of the most apparent drawbacks of DBSCAN algorithm are its low execution speed and its sensitivity to parameter selection. To overcome these problems, an initial grouping was applied to the data in this work through the Jenks algorithm to form three clearly distinguished fire occurrence frequency periods for each year (Fig. 5). Clustering the different periods separately can reduce the operational complexity of the program running on the distance between elements one by one, and can also distinguish between those two fire footprints in adjacent spatial locations to avoid mixing them into one fire event; this enhances the useability of the temporal properties of burned pixels in the MODIS products. The classification of three fire frequencies year by year is helpful to study the dynamic changes of fire in long temporal series.

Tables 1 and 2 show that DBSCAN tends to create too many clusters and redundancy when there are fewer fires in a year. The reason for this is that DBSCAN is applicable for

Year Period type n clusters Noise% Metrics Eps 2 1.14 2001 HF 0.007 0.988 5 MF 0.041 3.20 0.954 LF 0.018 6 6.01 0.966 2002 HF 0.047 8 0.24 0.926 MF 0.082 6 1.00 0.963 LF 4 7.08 0.971 0.056 2003 14 0.05 0.911 HF 0.033 13 0.10 0.821 MF 0.051 12 LF 0.048 0.30 0.968 HF 5 0.982 2004 0.025 4.26 3 MF 0.013 3.70 0.985 LF 0.025 3 2.63 0.960 2005 HF 0.019 4 0.15 0.950 7 MF 0.023 0.56 0.980 LF 0.025 5 4.60 0.976 4 2006 HF 0.030 0.50 0.940 7 0.031 1.23 0.978 MF LF 0.015 4 13.95 0.962

www.publish.csiro.au/wf



Fig. 7. Comparison of the model-extracted fire events and multiple real validation situations. (a) Spatial match presentation of partial fire event clustering results and the validation fires in 2004. (b-d) High-resolution true-colour images of the target region obtained from Google Earth in 2003/12, 2004/12 and 2005/12, respectively.

Year	Fire footprints frequency			Burned area		
	DBSCAN n_clustering	Validation data	Matched clustering results	R ²	RMSE (ha)	MAE (ha)
2001	13	Ι	0	_	_	-
2002	18	6	5	0.70	392.65	1757.74
2003	39	27	22	0.99	146.90	13 307.60
2004	П	4	3	0.59	132.79	103.33
2005	16	10	8	0.95	18.56	9733.44
2006	15	7	6	0.99	1355.96	15 047.57
Total	112	55	44	0.91	130.88	10 740.58

Table 3.	Accuracy	verification	of fire	time and	fire area
----------	----------	--------------	---------	----------	-----------

relatively dense data. In particular, the excessive noise values of the LF periods in 2001, 2002 and 2006 indicated a sparse spatial distribution of too few effective pixels to be clustered (Fig. 5). Before clustering, the Jenks was used for periodic classification. As can be clearly seen from the segmentation results in Fig. 5, after the Jenks classification, part of the fires very close to each other spatially had been pre-divided over three periods, which is helpful to reduce misclassification in subsequent spatial clustering. Taking

2003 as an example, the contour shape of fires is intact (Fig. 9). In 2003, there was a very large fire in the study area. Its fire footprint was extensive but there were many small fire patches (Fig. 9; MF 7), making the model silhouette coefficients low in that period compared with other years. In addition, the parameters used in this study can also serve as essential information to expand the global-scale estimation of fire footprint extraction, and to improve the parameterisation of other fire models.

Analysis of the fire footprints

Most of the fires that occurred in the area during the study period were of short duration, and large areas of fire scars were concentrated in the southeast area (Fig. 5). The fire footprints obtained by clustering in this study match, to a certain extent, the fire traces obtained by Tian *et al.* (2013). Fires with larger areas also matched the records obtained by



Fig. 8. Scatterplots of the clustering area and the verified area in matched fire events.

Fang et al. (2015) from the China Forestry Science Center. Although the validation accuracy of the burned area showed excellent performance (R^2 at 0.91), the frequency of the fire footprints obtained by clustering in this study is still much higher than that of the local records; only 44 out of 112 fire footprints match the validation data (Table 3). Additionally, some small-scale burnings were not investigated and were judged as fire events, so it is more difficult to balance the relationship between noise ratio and cluster number when adjusting the parameters of the DBSCAN model when the dataset was not dense enough (Patil and Vaidya 2012), and it was easy to have redundancy in the number of clusters in years when the fires were sparse. There was also the uncertainty in the MCD64A1 burned pixels. Studies consistently found that MODIS fire detection algorithms had trouble detecting smaller/cooler fires and frequently detected false alarms (Schroeder et al. 2008b). The clustering effect of the model on the fire footprint was more effective for large-scale and spatially continuous fire footprints in a year than small and dispersed fire footprints. In addition to the redundancy of clustering category number caused by the DBSCAN model itself, the model's outcomes could still compensate for the omission errors of large-scale wildfire monitoring implemented by the local agency on the ground. Most forest fires in this region are caused by lightning and man-made fires (Zhao et al. 2020). The peak fire seasons occur in spring and autumn, including the periods from early March to June and from late September to mid-October. In drought years, forest fires



Fig. 9. The map of the fire footprints and their validations extracted by Jenks–DBSCAN within the three periods in 2003.

may also occur in summer. After the 'Great Black Dragon Fire' (Xu 1987) in the Daxing'anling in 1987, the establishment of professional fire control institutions and the improvement of fire management level in China were significantly improved (Zong *et al.* 2022), which gradually reduced the number of fires caused by human factors in the region. There are still many spring fire hazards associated with the burning of souvenirs for the traditional Chinese Qingming Festival (Qi *et al.* 2022). This valuable information regarding fire occurrence occasions and locations – in partnership with the fire footprints extracted from the models in the current work – will act as an essential input for the formulation of targeted fire prevention and suppression actions.

In the study period, the 3.19 prairie-forest fire event was a typical mega-fire, which lasted for 1 month and caused great economic losses to the local area. In addition to human mismanagement of the local forests, the corresponding climate conditions and the accumulation of combustibles also contributed to its occurrence. The high-severity canopy fires occured in the presence of tall, flammable P. pumila (Shu et al. 2004) (Fig. 7). Zhang and Hu (2008) indicated that lightning-caused fires accounted for 36.3% of the total fire events in this region, and more than 75% of them occurred in summer. This corresponds to the information we got from the validation data. The fire footprints extracted from this study, which are reliable in spatial correspondence, accurate in the time of fire occurrence and have high R^2 in area statistics, can be used in a variety of applications and are beneficial for fire condition monitoring and biomass burning, as well as carbon emissions and related air quality studies.

Utility and application of the technology

Global fire patch data derived from pixel-level burned area information have recently emerged as an important source of information for climate, vegetation and carbon cycle modelling (Archibald et al. 2013; Loepfe et al. 2014; Hantson et al. 2015). Fire regimes could be more precisely described and modelled with an assemblage of fire traits rather than simple burned area (Laurent et al. 2018). Fire footprint characteristic is closely linked to meteorological and topographical attributes, as well as fuel continuity (Cary et al. 2006). Similarly, fire footprint characteristic databases (morphology, time span) that reflect potential propagation processes could be used to explore functional fire geography as a framework for analysis and modelling. Some researchers have developed algorithms for establishing fire database based on remote-sensing data. Chuvieco et al. (2022) expanded the MODIS Burned Area Product fire database using the high resolution of Sentinel-2 data. Laurent et al. (2018) used burned area from MODIS and MERIS sensors to build a global database of fire patch functional traits. Although the proposed fire footprint extraction framework in the current work was limited by the spatial resolution of single remote-sensing data that fails to take into account small-scale fires, it can efficiently establish a large fire event database with long time series and large regional scales.

The proposed framework is designed to obtain the fire footprint using MCD64A1 data and extract the combustion pixels into patches by the Jenks–DBSCAN method, which can be used for characteristic mining. Compared with the local field surveys, the database management of large-area fires often has manual errors in the measurement of burnt land, fire location and occurrence time (Lentini 2006). The current framework was validated in this study, and can be used to create a reliable large-scale fire event database, providing an important tool for the improvement of field investigations and contributing to fire prevention management and forest resource protection, all of which are quite helpful for fire regime research.

Limitations and prospects

Although limited by the time span of validation data, this study accurately determined the time and space of a single fire event as a test of spatio-temporal clustering method; the impact of large-scale and long-term fires in the study area was not explored in depth. The selection of remote-sensing products is relatively simple, but there are factors (including cloud-snow coverage and burned area uncertainty) that can affect the accuracy of fire footprint result. The burn date uncertainty is often associated with the lack of a clear signal in the MODIS surface reflectance imagery due to cloud contamination or snow coverage. A minimum of 1 day uncertainty is applied for all burned pixels (Giglio et al. 2016). Using remote-sensing data to analyse large-scale fires to extract fire footprints is a more feasible method, but the accuracy of obtained fire area is limited by the spatial resolution of satellite data. There is still substantial uncertainty about the interannual variability of fire emissions from fire combustion, the spatio-temporal variability of their patterns and their long-term trends (Wiedinmyer and Neff 2007; Wiedinmyer et al. 2011). There was an overestimation in the number of footprints obtained by Jenks-DBSCAN method compared withthe number of fires investigated on the ground, and in addition to the possibility of omissions in field investigations, techniques need to be improved to reduce the uncertainty of burned pixels. At the same time, the limitations of the single clustering model in dealing with noise make it possible to further optimise the fire footprint extraction framework.

Due to the limitations of the spatial resolution of MODIS data and the imperfect parameter tuning of the clustering model, only fire footprints with large areas can be extracted in this study. Making full use of high temporal resolution and combining high spatial resolution remote-sensing products to enable the extraction of small fire footprints will be the focus of in-depth research in the future. Future studies should also consider combining other types of remotesensing data to obtain more accurate environmental variables in each fire event to study the diffusion mechanism of fire occurrence and driving forces, as well as the spread of driving force to assess a region's dynamics (such as forest disturbance and forest restoration), thus providing information for fire prevention administrations in disaster mitigation and forest management.

Conclusions

This study developed an algorithm that combines both temporal segmentation and spatial clustering to better capture individual fire events from the MODIS MCD64A1 burned area dataset, and tested it in the Daxing'anling area of China during the period 2001–2006. The algorithm used the time information (DOY) of MCD64A1 and Jenks method to classify three fire frequency periods in each year. After meticulous parameter tuning, the DBSCAN models were used to cluster the burned pixels in each period to obtain fire events one by one. The model-extracted results matched 80% of the fire events provided by the verification data in space and time domains, and the overall R^2 of the fire area reached 0.91, with the highest value of 0.99 in 2003. The proposed Jenks-DBSCAN model contributes to the establishment of a reliable large-scale fire event database, and provides an important tool for data completion, fire prevention management and forest resource protection in the region.

References

- Anchang JY, Ananga EO, Pu R (2016) An efficient unsupervised index based approach for mapping urban vegetation from IKONOS imagery. International Journal of Applied Earth Observation and Geoinformation 50, 211–220. doi:10.1016/j.jag.2016.04.001
- Archibald S, Lehmann CER, Gómez-Dans JL, Bradstock RA (2013) Defining pyromes and global syndromes of fire regimes. *Proceedings* of the National Academy of Sciences 110, 6442–6447. doi:10.1073/ pnas.1211466110
- Artés T, Oom D, de Rigo D, Durrant TH, Maianti P, Libertà G, San-Miguel-Ayanz J (2019) A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Sci Data* 6, 296. doi:10.1038/ s41597-019-0312-2
- Bird MI, Cali JA (1998) A million-year record of fire in sub-Saharan Africa. *Nature* **394**, 767–769. doi:10.1038/29507
- Boschetti L, Roy D, Barbosa P, Boca R, Justice C (2008) A MODIS assessment of the summer 2007 extent burned in Greece. *International Journal of Remote Sensing* **29**, 2433–2436. doi:10.1080/01431160701874561
- Brockway DG, Gatewood RG, Paris RB (2002) Restoring fire as an ecological process in shortgrass prairie ecosystems: initial effects of prescribed burning during the dormant and growing seasons. *Journal of Environmental Management* **65**, 135–152. doi:10.1006/jema. 2002.0540
- Cary GJ, Keane RE, Gardner RH, Lavorel S, Flannigan MD, Davies ID, Li C, Lenihan JM, Rupp TS, Mouillot F (2006) Comparison of the sensitivity of landscape-fire-succession models to variation in terrain, fuel pattern, climate and weather. *Landscape Ecology* **21**, 121–137. doi:10.1007/s10980-005-7302-9
- Chen J, Yang ST, Li HW, Zhang B, Lv JR (2013) Research on geographical environment unit division based on the method of natural breaks (Jenks). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XL-4/W3**, 47–50. doi:10.5194/isprsarchives-XL-4-W3-47-2013

- Chuvieco E, Congalton RG (1989) Application of remote sensing and geographic information systems to forest fire hazard mapping. *Remote Sensing of Environment* **29**, 147–159. doi:10.1016/0034-4257(89)90023-0
- Chuvieco E, Roteta E, Sali M, Stroppiana D, Boettcher M, Kirches G, Storm T, Khairoun A, Pettinari ML, Franquesa M, Albergel C (2022) Building a small fire database for Sub-Saharan Africa from Sentinel-2 high-resolution images. *Science of the Total Environment* **845**, 157139. doi:10.1016/j.scitotenv.2022.157139
- Demange M, Di Fonso A, Di Stefano G, Vittorini P (2022) A graph theoretical approach to the firebreak locating problem. *Theoretical Computer Science* **914**, 47–72. doi:10.1016/j.tcs.2022.02.012
- Ester M (2009) Density-based clustering. In 'Encyclopedia of database systems'. (Eds L Liu, MT ÖZsu) pp. 795–799. (Springer Publishing: Boston, MA, USA)
- Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In 'Proceedings of the second international conference on knowledge discovery and data mining'. pp. 226–231. (AAAI Publishing: Portland, OR, USA)
- Fang L, Yang J, Zu J, Li G, Zhang J (2015) Quantifying influences and relative importance of fire weather, topography, and vegetation on fire size and fire severity in a Chinese boreal forest landscape. *Forest Ecology and Management* **356**, 2–12. doi:10.1016/j.foreco.2015. 01.011
- French BJ, Prior LD, Williamson GJ, Bowman DMJS (2016) Cause and effects of a megafire in sedge-heathland in the Tasmanian temperate wilderness. *Australian Journal of Botany* **64**, 513–525. doi:10.1071/BT16087
- Giglio L, Loboda T, Roy DP, Quayle B, Justice CO (2009) An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sensing of Environment* **113**, 408–420. doi:10.1016/j.rse.2008. 10.006
- Giglio L, Schroeder W, Justice CO (2016) The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment* **178**, 31–41. doi:10.1016/j.rse.2016.02.054
- Girardin MP, Mudelsee M (2008) Past and future changes in Canadian boreal wildfire activity. *Ecological Applications* **18**, 391–406. doi:10.1890/07-0747.1
- Goldammer JG, Furyaev VV (1996) Fire in ecosystems of boreal Eurasia: ecological impacts and links to the global system. In 'Fire in ecosystems of Boreal Eurasia'. (Eds JG Goldammer, VV Furyaev) pp. 1–20. (Springer Publishing: Dordrecht, Netherlands)
- Grogan P, Burns TD, Chapin IIi FS (2000) Fire effects on ecosystem nitrogen cycling in a Californian bishop pine forest. *Oecologia* **122**, 537–544. doi:10.1007/s004420050977
- Gruber N, Friedlingstein P, Field CB, Valentini R, Heimann M, Richey JE, Lankao PR, Schulze E-D, Chen C-TA (2004) The vulnerability of the carbon cycle in the 21st century: an assessment of carbon-climate-human interactions. In 'Scope-Scientific committee on problems of the environment international council of scientific unions. Vol. 62'. (Eds CB Field, MR Raupach) pp. 45–76. (Island Press)
- Hantson S, Pueyo S, Chuvieco E (2015) Global fire size distribution is driven by human impact and climate. *Global Ecology and Biogeography* **24**, 77–86. doi:10.1111/geb.12246
- Hope A, Albers N, Bart R (2012) Characterizing post-fire recovery of fynbos vegetation in the Western Cape Region of South Africa using MODIS data. *International Journal of Remote Sensing* 33, 979–999. doi:10.1080/01431161.2010.543184
- Humber ML, Boschetti L, Giglio L, Justice CO (2019) Spatial and temporal intercomparison of four global burned area products. *International Journal of Digital Earth* 12, 460–484. doi:10.1080/ 17538947.2018.1433727
- Jaafari A, Zenner EK, Panahi M, Shahabi H (2019) Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agricultural and Forest Meteorology* **266–267**, 198–207. doi:10.1016/j.agrformet.2018.12.015
- Jin Y, Randerson JT, Goetz SJ, Beck PSA, Loranty MM, Goulden ML (2012) The influence of burn severity on postfire vegetation recovery and albedo change during early succession in North American boreal forests. *Journal of Geophysical Research: Biogeosciences* 117, G01036. doi:10.1029/2011JG001886

- Kristianto A, Sediyono E, Hartomo KD (2020) Implementation DBSCAN algorithm to clustering satellite surface temperature data in Indonesia. *Register: Jurnal Ilmiah Teknologi Sistem Informasi* 6, 109–118. doi:10.26594/register.v6i2.1913
- Laurent P, Mouillot F, Yue C, Ciais P, Moreno MV, Nogueira JMP (2018) FRY, a global database of fire patch functional traits derived from space-borne burned area products. *Scientific Data* **5**, 180132. doi:10.1038/sdata.2018.132
- Lentini JJ (2006) 'Scientific protocols for fire investigation.' (CRC Publishing: Melbourne, VIC, Australia)
- Liu P, Zhou D, Wu N (2007) VDBSCAN: varied density based spatial clustering of applications with noise, In '2007 International conference on service systems and service management', 9–11 June 2007. pp. 1–4. (IEEE Publishing: Chengdu, China)
- Loboda T, O'Neal KJ, Csiszar I (2007) Regionally adaptable dNBR-based algorithm for burned area mapping from MODIS data. *Remote Sensing* of Environment **109**, 429–442. doi:10.1016/j.rse.2007.01.017
- Loepfe L, Rodrigo A, Lloret F (2014) Two thresholds determine climatic control of forest fire size in Europe and northern Africa. *Regional Environmental Change* 14, 1395–1404. doi:10.1007/s10113-013-0583-7
- Maier SW, Russell-Smith J, Edwards AC, Yates C (2013) Sensitivity of the MODIS fire detection algorithm (MOD14) in the savanna region of the Northern Territory, Australia. *ISPRS Journal of Photogrammetry and Remote Sensing* **76**, 11–16. doi:10.1016/j.isprsjprs.2012.11.005
- Mallinis G, Mitsopoulos I, Chrysafi I (2018) Evaluating and comparing Sentinel 2A and Landsat-8 Operational Land Imager (OLI) spectral indices for estimating fire severity in a Mediterranean pine ecosystem of Greece. *GIScience & Remote Sensing* 55, 1–18. doi:10.1080/ 15481603.2017.1354803
- McMaster R (1997) In memoriam: George F. Jenks (1916-1996). Cartography and Geographic Information Systems 24, 56–59. doi:10.1559/152304097782438764
- Mitri GH, Gitas IZ (2013) Mapping post-fire forest regeneration and vegetation recovery using a combination of very high spatial resolution and hyperspectral satellite imagery. *International Journal of Applied Earth Observation and Geoinformation* **20**, 60–66. doi:10.1016/j.jag.2011.09.001
- Nisa KK, Andrianto HA, Mardhiyyah R (2014) Hotspot clustering using DBSCAN algorithm and shiny web framework. In '2014 international conference on advanced computer science and information system', 18–19 October 2014. pp. 129–132. (IEEE Publishing: Jakarta, Indonesia)
- Patil YS, Vaidya M (2012) A technical survey on cluster analysis in data mining. Journal of Emerging Technology and Advance Engineering 2, 503–513. doi:10.1145/1980022.1980143
- Peng G, Li J, Chen Y, Norizan AP (2007) A forest fire risk assessment using ASTER images in Peninsular Malaysia. *Journal of China* University of Mining and Technology 17, 232–237. doi:10.1016/ S1006-1266(07)60078-9
- Qi X, Wei M, Wang Z, Jiang T, Wang P, Teng M, Yan Z (2022) Effects of the COVID-19 lockdown in Hubei, China: cessation of incense burning reduces regional landscape fire. *Ecology and Society* **27**, 5. doi:10.5751/ES-13386-270305
- Rousseeuw PJ (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
- Roy DP, Boschetti L (2009) Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products. *IEEE Transactions on Geoscience and Remote Sensing* **47**, 1032–1044. doi:10.1109/TGRS. 2008.2009000
- Roy DP, Boschetti L, Justice CO, Ju J (2008) The collection 5 MODIS burned area product — Global evaluation by comparison with the MODIS active fire product. *Remote Sensing of Environment* **112**, 3690–3707. doi:10.1016/j.rse.2008.05.013
- Sander J, Ester M, Kriegel H-P, Xu X (1998) Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery* **2**, 169–194. doi:10.1023/A:1009745219419
- Schroeder W, Csiszar I, Morisette J (2008a) Quantifying the impact of cloud obscuration on remote sensing of active fires in the Brazilian Amazon. *Remote Sensing of Environment* **112**, 456–470. doi:10.1016/ j.rse.2007.05.004

- Schroeder W, Prins E, Giglio L, Csiszar I, Schmidt C, Morisette J, Morton D (2008b) Validation of GOES and MODIS active fire detection products using ASTER and ETM+ data. *Remote Sensing* of Environment 112, 2711–2726. doi:10.1016/j.rse.2008.01.005
- Sheng D, Deng J, Xiang J (2021) Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network. *IEEE Access* 9, 63933–63942. doi:10.1109/ACCESS.2021.3075731
- Shu L, Wang M, Li Z, Xiao R, Tian X (2004) Dwarf Siberian pine forest fire environment in Daxingan Mountains. *Journal of Mountain Science* **22**, 36–39.
- Sulla-Menashe D, Friedl MA (2018) 'User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. Vol. 1'. 18 pp. (USGS: Reston, VA, USA) Available at https://lpdaac.usgs.gov/ documents/101/MCD12_User_Guide_V6.pdf [Accessed 10 April 2022]
- Thompson MP, Calkin DE (2011) Uncertainty and risk in wildland fire management: a review. *Journal of Environmental Management* **92**, 1895–1909. doi:10.1016/j.jenvman.2011.03.015
- Tian X, McRae DJ, Jin J, Shu L, Zhao F, Wang M (2011) Wildfires and the Canadian Forest Fire Weather Index system for the Daxing'anling region of China. *International Journal of Wildland Fire* **20**, 963–973. doi:10.1071/WF09120
- Tian X, Shu L, Wang M, Zhao F, Chen L (2013) The fire danger and fire regime for the Daxing'anling region for 1987-2010. *Procedia Engineering* **62**, 1023–1031. doi:10.1016/j.proeng.2013.08.157
- Tobler W (2004) On the first law of geography: a reply. Annals of the Association of American Geographers **94**, 304–310. doi:10.1111/j. 1467-8306.2004.09402009.x
- Usman M, Sitanggang IS, Syaufina L (2015) Hotspot distribution analyses based on peat characteristics using density-based spatial clustering. *Procedia Environmental Sciences* 24, 132–140. doi:10.1016/j. proenv.2015.03.018
- Van Leeuwen WJD (2008) Monitoring the effects of forest restoration treatments on post-fire vegetation recovery with MODIS multitemporal data. Sensors 8, 2017–2042. doi:10.3390/s8032017
- Van Mantgem PJ, Stephenson NL, Knapp E, Battles J, Keeley JE (2011) Long-term effects of prescribed fire on mixed conifer forest structure in the Sierra Nevada, California. *Forest Ecology and Management* 261, 989–994. doi:10.1016/j.foreco.2010.12.013
- Vatresia A, Rais RR, Miliana RS (2020) Spatio-Temporal Clustering for Fire Hot Spots Occurrence Over Sulawesi Island. In '2020 FORTEIinternational conference on electrical engineering (FORTEI-ICEE)', 23–24 September 2020. pp. 156–161. (IEEE Publishing: Bandung, Indonesia)
- Wang S, Wang G, Zhang J (2019) Data analysis method of terrorist attacks based on AHP-DBSCAN method. *Journal of Physics: Conference Series* 1168, 032029. doi:10.1088/1742-6596/1168/3/ 032029
- Wiedinmyer C, Neff JC (2007) Estimates of CO_2 from fires in the United States: implications for carbon management. *Carbon Balance and Management* **2**, 10. doi:10.1186/1750-0680-2-10
- Wiedinmyer C, Akagi SK, Yokelson RJ, Emmons LK, Al-Saadi JA, Orlando JJ, Soja AJ (2011) The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development* 4, 625–641. doi:10.5194/ gmd-4-625-2011
- Winoto-Lewin S, Sanger JC, Kirkpatrick JB (2020) Propensities of old growth, mature and regrowth wet eucalypt forest, and *Eucalyptus nitens* plantation, to burn during wildfire and suffer fire-induced crown death. *Fire* **3**, 13. doi:10.3390/fire3020013
- Xing W, Bao K, Gallego-Sala AV, Charman DJ, Zhang Z, Gao C, Lu X, Wang G (2015) Climate controls on carbon accumulation in peatlands of Northeast China. *Quaternary Science Reviews* **115**, 78–88. doi:10.1016/j.quascirev.2015.03.005
- Xu Y (1987) Daxinganling recovered after the great fire. In 'China forestry yearbook 1987'. (Eds M Gao, Y Qian) pp. 524–526. (China Forestry Press Publishing: Beijing, China) [In Chinese]
- Yalcin M, Kilic Gul F (2017) A GIS-based multi criteria decision analysis approach for exploring geothermal resources: Akarcay basin (Afyonkarahisar). *Geothermics* 67, 18–28. doi:10.1016/j.geothermics. 2017.01.002
- Yanuarsyah I, Suwarno Y, Hudjimartsu S (2016) The use of hotspot spatial clustering and multitemporal satellite imagery to facilitate

peat land degradation in West Kalimantan, Indonesia (Case Study in Mensiku Miniwatershed of Kapuas River). In 'IOP Conference Series: Earth and Environmental Science. Vol. 47'. p. 012046. (IOP Publishing: Bristol, UK)

- Zhang Y, Hu H (2008) Climatic change and its impact on forest fire in Daxing'anling Mountain. *Journal of Northeast Forestry University* **36**, 29–31. [In Chinese]
- Zhang Z, Craft CB, Xue Z, Tong S, Lu X (2016) Regulating effects of climate, net primary productivity, and nitrogen on carbon sequestration rates in temperate wetlands, Northeast China. *Ecological Indicators* **70**, 114–124. doi:10.1016/j.ecolind.2016.05.041
- Zhao F, Liu Y, Shu L (2020) Change in the fire season pattern from bimodal to unimodal under climate change: the case of Daxing'anling in Northeast China. *Agricultural and Forest Meteorology* **291**, 108075. doi:10.1016/j.agrformet.2020.108075
- Zhou H, Qu S, Yang X, Yuan Q (2020) Regional credit, technological innovation, and economic growth in China: a spatial panel analysis. *Discrete Dynamics in Nature and Society* **2020**, 1738279. doi:10.1155/2020/1738279
- Zong X, Tian X, Yao Q, Brown PM (2022) An analysis of fatalities from forest fires in China, 1951–2018. *International Journal of Wildland Fire* **31**, 507–517. doi:10.1071/WF21137

Data availability. Publicly available datasets were analysed in this study. These data can be found at: https://lpdaac.usgs.gov/products/ and http://www.dxal. gov.cn/rwxa. Fire verification data were obtained from the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences and are available with the permission of the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences.

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

Declaration of funding. This work was jointly funded by the following grants: The National Key R&D Program of China (2019YFA0606600); the Natural Science Foundation of China (31971577); and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Acknowledgements. The authors acknowledge the United States Geological Survey (USGS) and National Aeronautics and Space Administration (NASA) for providing the image data. Special thanks to the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences for sharing their fire data. We also thank engineer Xinyu Wang for his help with code issues in this research.

Author contributions. HS designed the study, data preparation, data processing and analysis, and wrote the paper. XM reviewed the manuscript and improved English language. ML provided the project design and the paper editing. All the authors read and approved the final manuscript.

Author affiliations

^ACollege of Forestry, Nanjing Forestry University, Nanjing 210037, China.

^BCo-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China.

^CEast China Inventory and Planning Institute of the State Administration of Forestry and Grassland, Hangzhou 310019, China.