Research on Identification Methods of Industrial Heat Source Integrating Thermal Anomaly Features

Yu Liu [1], Xinyuan Gao [1], Lijuan Zheng[1,] *, Hailun Dai[2], Lijing Xu [1]

[1] Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of P.R.China Beijing, 100048, China- (867532660, 969520377, 283991924, 2825333121)@qq.com
[2] Beijing Satlmage Information Technology Co. Ltd. Beijing, 100048, China- 99261010@qq.com

Commission III, WG III/2

Keywords: Land Surface Temperature, Industrial Heat Source, DBSCAN, Logistic Regression.

Abstract

To address the limitation of existing industrial heat sources extraction methods based on high-temperature data are difficult to identify low temperature (<500 K) and small-scale factories, this paper proposes an industrial heat sources identification method that integrates thermal anomaly features from long-term remote sensing image. Firstly, for the surface temperature data inverted from satellite remote sensing image data, a global threshold combined with temperature-feature dynamic detection approach is used to extract potential industrial thermal anomaly points. Secondly,based on the annual scale, we integrated results from multiple periods of industrial thermal anomaly points to construct an industrial heat sources identification model combining DBSCAN clustering and logistic regression.Finally, taking Tangshan City, China as the research area, industrial heat sources identification was conducted to verify the accuracy of the model. The results showed that the accuracy of the industrial heat sources identification model was 86%, which significantly improved the ability to identify low-temperature and small-scale factories compared to the extraction results based on VIIRS Active Fire products.

1. Introduction

At present, China's carbon emissions mainly come from energy consumption dominated by fossil fuels. Energy is the main battlefield for achieving carbon peaking and carbon neutrality, while the production activities of heat source enterprises such as coal, oil and gas, and steel are the main sources of fossil energy consumption. Research has shown that industrial production activities are one of the major sources of heat emissions (Sun et al.2024). Accurately grasping the spatial distribution of industrial heat sources provides important data reference for the government to steadily promote carbon peaking and carbon neutrality.

The spatiotemporal distribution of industrial heat sources is an important indicator for assessing levels of energy consumption and air pollution (Ma et al.2024). Industrial heat sources provide the primary means of production for the material and technical bases of all sectors of a nation’ s economy (Deng et al.2018). Industrial heat sources have the characteristics of large quantity, complex types, and wide distribution.Traditional periodic surveys for obtaining industrial heat source data are time-consuming, labor-intensive, and slow to update, resulting in unclear quantities and uncertain locations. Thermal infrared remote sensing has been widely applied in forest fire monitoring, geothermal resource exploration, and urban heat island assessment due to its ability to efficiently and cost-effectively identify the surface temperature anomalies(Liu et al.2018,Li et al.2021). It is now gradually being applied to the identification of industrial heat sources.Industrial heat source extraction methods can be broadly categorized into two types: spatiotemporal clustering methods and index extraction methods (Ma et al.2020). Clustering based detection methods commonly utilize data from the VIIRS Nightfire and VIIRS Activefire products,which use mid infrared or short infrared channel data for fire point detection with low spatial resolution of 750 m and 375 m. Their ability to identify thermal anomalies with temperatures below 500 K is limited, and using these products to identify industrial heat sources may miss some small or colder factories. Index based detection methods utilizing the band characteristics of sensor data,a specific thermal anomaly index is designed to identify thermal anomaly points and extract industrial heat sources, which can detect factories with lower temperatures. However, due to the different band designs of thermal infrared sensors for different satellites, the universality of the method is low.

To address the above-mentioned problems, this study takes Landsat temperature products as the main data source and Tangshan City as the research object, and proposes an industrial heat source identification method integrating thermal anomaly characteristics. It aims to improve the ability to identify small and low-temperature industrial heat sources and increase the diversity of methods for identifying industrial heat sources using thermal infrared remote sensing technology.

2. Study Area and Data Sources

2.1 Study Area

Tangshan City is located in the eastern part of Hebei Province, China, as shown in Figure 1,which is between 117°31′-119°19′ E longitude and 38°55′-40°28′ N latitude, with a total area of 1,3472 square kilometers. As a heavy-industry city, Tangshan has maintained strong economic growth in recent years, with a continuously increasing dependence on heavy industries such as steel. It has a large total consumption of fossil energy, and its

* Corresponding author

economic growth relies heavily on energy consumption. The rapid growth of the heavy-industry economy and the massive consumption of fossil energy have led to a huge total carbon emission in the region, thus putting Tangshan City face with arduous emission reduction tasks. Therefore, taking Tangshan City as the study area to clarify the current distribution of industrial heat sources will provide reference data for better promoting the achievement of carbon peaking and carbon neutrality in the industrial sector.

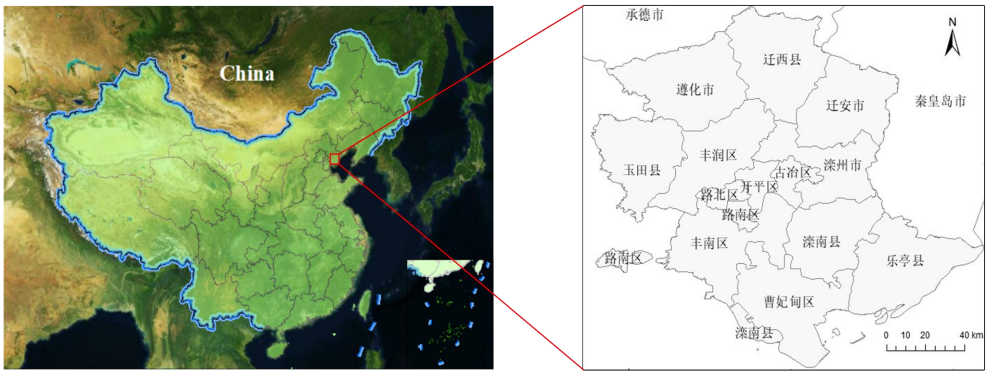


Figure 1. The location of study area

2.2 Data Sources and Preprocessing

A total of 124 cloud-free Landsat Collection 2 Level 2 ST data covering Tangshan City from 2020 to 2023 were selected for temperature conversion, invalid value removal. The obtained land surface temperature data of the study area with a resolution of 100 m was used as the main data source for thermal anomaly identification.

By using the average-masked product from the Version 2 of VIIRS/DNB Nighttime Lights data (Elvidge et al.2021) and the global 10m land cover type product obtained from ESA Sentinel-2 images (Karra et al.2021), the interference of nighttime thermal anomalies caused by wild biomass burning and natural heat sources in the single-period thermal anomaly results was initially removed.

3. Methods

This study mainly uses Landsat-8/9 temperature products to identify industrial heat sources. Firstly, the global threshold combined with temperature feature dynamic detection approach is used to extract and construct a dataset of potential industrial thermal anomaly points from single-period image data. Secondly, based on multi-temporal (annual scale) thermal anomaly information, DBSCAN clustering and Logistic model are applied to remove interference information and identify industrial heat sources. Finally, the verification of extraction results is carried out. The accuracy of the identification results is evaluated by combining existing industrial heat source datasets, industrial heat source extraction results from VIIRS Active Fire products, and high-resolution images from Google Earth.

3.1 Thermal Anomaly Detection Using Single-period Data

3.1.1 Preliminary Identification of Thermal Anomalies: Boxplot is a statistical method that displays the dispersion of a univariate sample dataset, which can be used to detect outliers in the dataset. Unlike the standard deviation method, the actual data for the boxplot method does not need to satisfy a normal distribution, nor does it require preassuming that the data follows a specific distribution.Therefore, in this study, the upper boundary value of temperature data for each period calculated by the boxplot method is used as the global threshold for the preliminary identification of thermal anomalies. Pixels with a temperature higher than the upper boundary value of the box (Out_{upper}) are labeled as potential thermal anomaly pixels, while those not exceeding the upper boundary value are labeled as non-thermal anomaly pixels. This threshold can eliminate a large number of non-thermal anomaly pixels, thereby accelerating the subsequent extraction speed of industrial thermal anomaly points. The formula for determining the global threshold of thermal anomalies using the boxplot method is as follows:

$$Out_{upper} = Q3 + 1.5IQR \quad (1)$$

$$IQR = Q3 - Q1 \quad (2)$$

Among them: Q1 represents the first quartile (25%), and Q3 represents the third quartile (75%).

3.1.2 Dynamic Detection of Potential Thermal Anomalies: It is difficult to accurately extract industrial thermal anomaly points using only the single temperature feature of Landsat Collection 2 Level 2 ST data, a dynamic detection method based on temperature characteristics is proposed. By setting adaptive threshold, this method avoids the omission of results caused by setting a single threshold. Meanwhile, The thermal anomaly features of the background window are extracted, which providing an important basis for industrial heat source identification. The main principle is to calculate the temperature difference between each preliminarily identified thermal anomaly pixel and the average temperature of all pixels in its 21 × 21 background window; pixels with a temperature difference greater than 2K are regarded as potential thermal anomaly pixels.

$$T(i) - \mu(i) > 2K \quad (3)$$

Where $T(i)$ and $\mu(i)$ respectively represent the temperature of the i-th preliminarily identified thermal anomaly pixel and the average temperature of all pixels within its background window.

The temperature characteristic values of the background window where the potential thermal anomaly pixel is located are calculated, and the results are taken as the characteristic attributes. The temperature characteristics of the background

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-377-2026 | © Author(s) 2026. CC BY 4.0 License.

window include:the average temperature of the background window (*WMean*), the temperature difference between thermal anomaly points pixels and the average temperature of the background window (*Tdiff*), the proportion of pixels in the background window with a temperature close to that of thermal anomaly points pixels relative to the total number of pixels in the window (*Proportion*),the number of pixels contained in each temperature grade within the window, which are obtained by classifying the pixels in the background window into 5 grades using the mean-standard deviation grading method based on pixel temperature. The temperature classification criteria are shown in Table 1.

Grades	Indicator
LTnumber	$\mu \geq T$
S1	$\mu+\sigma \geq T > \mu$
S2	$\mu+2\times\sigma \geq T > \mu+\sigma$
S3	$\mu+3\times\sigma \geq T > \mu+2\times\sigma$
HTnumber	$T \geq \mu+3\times\sigma$

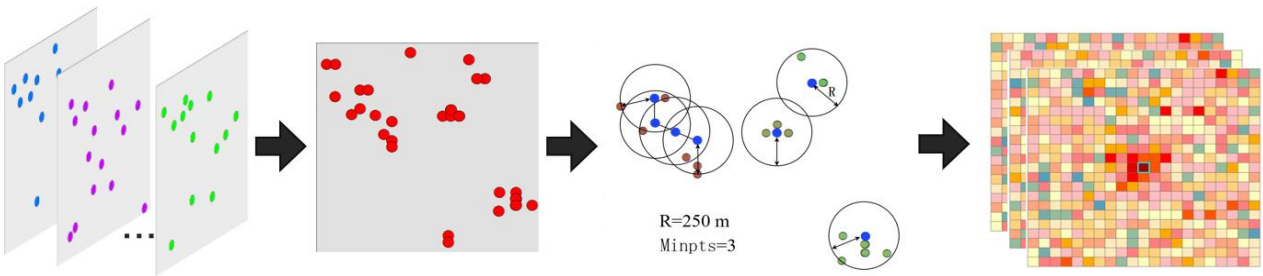
Table 1. Mean-Standard Deviation Gradation

Among the potential thermal potential thermal anomalies extracted through the above steps, there are still a large number of non-industrial thermal anomaly interferences.Therefore, based on two characteristics of heat source enterprises—most of them are concentrated or individually located in built-up areas, and lighting is necessary and stably exists for a long time during their production activities.The potential thermal anomaly point data extracted are clipped using the construction land from the 10m global land cover data and the areas with a threshold greater than 0 in the VIIRS/DNB annual nighttime light data. This is to preliminarily remove thermal anomaly points caused by geothermal resource heating and wild biomass combustion, as shown in Figure 2.

Figure 2. The results after removing non industrial thermal anomalies points

3.2 Industrial Heat Source Identification Using Multi-Temporal Data

On the basis of identifying thermal anomaly points from single-period data, the data of single-period thermal anomaly points throughout the year are merged.According to the spatial aggregation characteristics of industrial heat sources, DBSCAN clustering is used to extract suspected industrial heat source objects.Based on the statistical values of characteristic factors of suspected industrial heat sources, Logistic regression analysis is applied to realize multi-temporal industrial heat source identification. The extraction process is shown in Figure 3.



(a) (b) (c) (d)

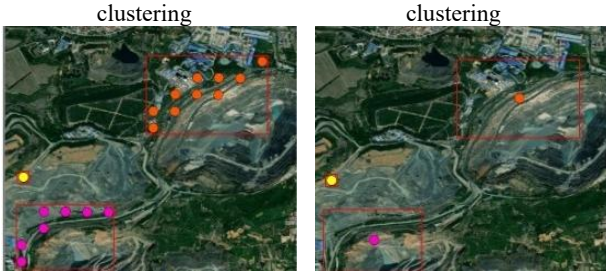
Figure 3. Industrial heat source extraction process

3.2.1 DBSCAN clustering extraction: DBSCAN is the first density based clustering algorithm (Khan et al.,2014).The central idea behind DBSCAN and its extensions and revisions is the notion that points are assigned to the same cluster if they are density reachable from each other (Hahsler et al.,2019). Points within this radius are neighborhood points, and points on the edge of this neighborhood are boundary points.A point is considered a noise point if it is neither a core point nor a boundary point. It describes the compactness of neighborhood sample points through parameters (R, MinPts), as shown in Figure 4(c), where R represents the neighborhood radius of sample points, and MinPts represents the threshold of the number of points within the neighborhood at distance R from a sample point.In this study, the DBSCAN clustering algorithm is used to simplify adjacent points in the single-period thermal anomaly detection results and to cluster the merged data of annual-scale thermal anomaly points, thereby extracting suspected industrial heat source objects.In the results of thermal anomaly detection for single-period data, there are regions where adjacent pixels are all identified as thermal anomaly points. Although most of these regions correspond to industrial thermal anomalies points (Figure 4(a)), some non industrial thermal anomalies points also exist (Figure 4(c)). If such sets of false thermal anomaly points are directly merged into the annual data without processing, they may be misidentified as industrial heat sources in subsequent annual-scale thermal anomaly clustering due to their sufficient quantity. Therefore, before merging, only one thermal anomaly point is retained around the same heat source in single-period data.For thermal anomaly points located in adjacent pixels (considered to belong to the same heat source) in the above results, DBSCAN clustering is used for merging.For each cluster of thermal anomaly points in the results, the point at the geometric center of the minimum bounding rectangle of the cluster replaces all thermal anomaly points in that cluster. Meanwhile, the number of thermal anomaly points contained in the cluster and the mean values of the 8 window temperature characteristics of all thermal anomaly points are taken as the thermal anomaly characteristics of the new point.Considering the characteristics of Landsat Collection 2 Level 2 ST data, the neighborhood radius R is set to 250 m and MinPts to 3 for clustering the annual-scale data, thereby extracting potential industrial heat sources.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-377-2026 | © Author(s) 2026. CC BY 4.0 License.

(a)Industrial thermal anomalies points before clustering

(b)Industrial thermal anomalies points after clustering



(c) Non industrial thermal anomalies points before clustering

(d)Non industrial thermal anomalies points after clustering

Figure 4.The clustering of single-period thermal anomaly points

The following attribute features are calculated for each potential heat source object in the results: the total sum of the "numbers" attribute values of all thermal anomaly points (*Numbers*); the number of periods covered by the single-period thermal anomaly points contained (*IntervalIM*);The mean values of the 8 window temperature characteristics of all thermal anomaly points, which serve as the corresponding attribute features of the potential industrial heat sources.

3.2.2 Logistic Regression for Determining Industrial Heat Source Objects:Regression analysis is a valuable research method because of its versatile application to different study contexts (Stoltzfus et al.,2011) Logistic regression, a common classification algorithm, is often used to solve binary classification problems.When determining whether a suspected heat source object is an industrial object, it can be used to express the relationship between the binary dependent variable—i.e., whether it is an industrial heat source (where t=1 indicates the heat source is industrial and t=0 indicates it is non-industrial)—and a set of independent variables (x1, x2, x3, ⋯), which represent the various characteristic factors of the suspected heat source object. In this study, 1063 heat source objects identified in 2023 were used as sample data. The classification variables of the Logistic model included the window temperature features (*WMean*, *Tdiff, Proportion*, *LTnumber*, *S1*, *S2*, *S4*, *HTnumber*) of potential heat source objects extracted in the previous section, the total number of thermal anomaly points (*Number)*, and the temporal feature (*IntervalIM*). The Logistic regression model is as follows:

$$P = \frac{e^{\beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_m x_m}}{1 + e^{\beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_m x_m}} \quad (4)$$

Where P represents the probability of an industrial heat source object, m denotes the number of covariates, β_1 ... are the coefficients of the independent variables, and x_1 ... are the respective independent variables in the model.

To avoid high correlations among feature factors that could lead to computational errors in the model, this study uses Pearson's correlation coefficient to analyze the relationships between variables. According to Table 2, *Numbers* are highly correlated with *LTnumber* and *S2*, and *LTnumber* are highly correlated with *S2*. Therefore, *Numbers* and *LTnumber* are excluded.

Factors	*WMean*	*Tdiff*	*LTnumber*	*S1*	*S2*	*S3*	*HTnumber*	*Numbers*	*IntervalIM*	*Proportion*
WMean	1	0.19	0	0.22	-0.05	-0.25	-0.03	-0.09	-0.3	0.03
Tdiff		1	-0.15	-0.16	0.03	-0.14	-0.01	0.23	0.15	-0.26
LTnumbers			1	-0.33	-0.67	-0.11	0.01	-0.944	0.03	0.21
S1				1	-0.2	0.05	0	0	-0.21	0.45
S2					1	0.05	0	0.73	0.04	-0.37
S3						1	-0.02	0.1	-0.1	-0.11
HTnumbers							1	-0.01	0.21	0.04
Numbers								1	0.04	-0.37
IntervalIM									1	-0.09
Proportion										1

Table 2. Pearson correlation coefficient

To check whether the remaining influencing factors have multicollinearity, tolerance (T) and variance inflation factor (VIF) are used as indicators to measure whether there is multicollinearity between factors.The calculation formula is as follows:

$$VIF = \frac{1}{1 - A^2} = \frac{1}{T} \quad (5)$$

Where A^2represents the variance between independent variables.

Factors	T	VIF
WMean	0.752	1.33
Tdiff	0.774	1.292

Factors	T	VIF
S1	0.001	1101.3
S2	0.007	146.75
S3	0.058	17.38
HTnumbers	0.956	1.046
IntervalIM	0.777	1.287
Proportion	0.598	1.673

Table 3. Collinearity judgment of influencing factors

According to Table 3, the VIF values of s1, s2, and s3 are all greater than 10, indicating collinearity. Therefore, these three influencing factors should be removed.

4. Results and Discussion

4.1 Results of industrial heat source extraction

Visually interpret and annotate sample data using Google Earth imagery, assigning a value of 0 to non-industrial heat sources and 1 to industrial heat sources.Randomly split the sample data into 50% training data and 50% validation data, repeating the division five times. Apply the Logistic model to fit and predict the sample data from each of the five splits, selecting the model parameters that yield the best prediction results.This model was then applied to classify suspected industrial heat source objects from 2020 to 2023, ultimately identifying 2,702 industrial heat source objects in Tangshan during that period,as shown in Figure 5.

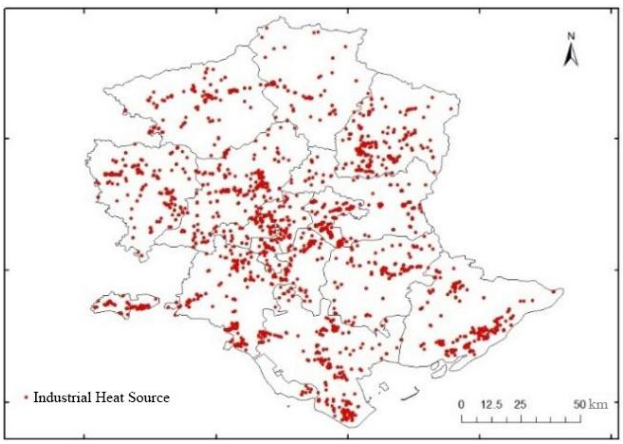
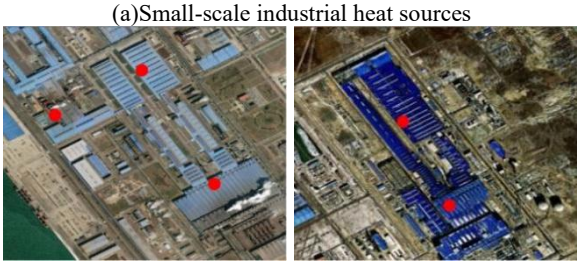


Figure 5.Results of Industrial Heat Source Extraction in Tangshan

4.2 Accuracy Assessment

Using the industrial heat sources in Tangshan City from the 2021 public dataset (Ma et al.,2022) as validation data, the accuracy of the industrial heat sources extracted in this paper is assessed. Through comparison, it was found that there were 82% repetitions in the extraction results. The industrial heat sources in open source datasets are mainly large-scale, high-energy consuming enterprises, while the method proposed in this article has greatly improved the extraction quantity of low-temperature, small-scale industrial heat sources,as shown in Figure 6 (a). Compared with the validation data, the number of newly added industrial heat source points is 231.It should be noted that the number of newly added heat source points is not the number of industrial heat sources, as some industrial heat source may have multiple points identified simultaneously, as shown in Figure 6 (b)

(a)Small-scale industrial heat sources



(b) Multiple points in one industrial heat source

Figure 6. Results of Industrial Heat Source Extraction

To further validate the effectiveness of the method, the industrial heat source extraction results were overlaid with high-resolution images from Google Earth, and the extraction results were verified one by one through visual interpretation. The overall accuracy was 86%, as shown in Table 4.

Year	Extraction number	Interpretation number	Accuracy rate
2020	592	531	94%
2021	354	316	89%
2022	1199	970	81%
2023	1065	886	83%

Table 4. Result of accuracy assessment

5. Conclusions

This article proposes an industrial heat source identification method based on Landsat temperature products that integrates thermal anomaly features, and obtains the following main conclusions:

1)Extracting potential industrial heat anomalies through dynamic thresholding of temperature characteristics, merging multi temporal annual heat anomalies, and combining DBSCAN clustering and logistic models for identifying industrial heat sources is an effective method.This method significantly improves the ability to identify low-temperature and small-scale industrial heat sources

2)Through comparative analysis with open-source datasets, the results show that the method proposed in this study has high accuracy in identifying industrial heat sources. While increasing the number of low-temperature and small-scale industrial heat sources identified, it also achieves good results in identifying large-scale industrial heat sources.

3)In 2020 and 2021, the extraction numbers of industrial heat sources was significantly lower than that in 2022 and 2023, which was consistent with the suspension of production

activities of many factories, especially small factories, during the COVID-19 epidemic in 2020 and 2021.

Acknowledgements

The research was supported by the Project supported by National Key R&D Program of China (Rapid Identification Technology of Security State of Land Space Elements, 2023YFC3804003-01).

References

Deng, H., Xu, B., & Zou, Y. ,2018.The Economic Logic of China’s Industrialization: From Heavy Industry to Comparative Advantage. *J. Econ. Res.* 53: 17 – 31.

Elvidge, C.D., Zhizhin, M., Ghosh, T., Hsu, F. and Taneja, J., 2021. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. *Remote Sens.*, 13(5): 922.

Hahsler, M., Piekenbrock, M. and Doran, D., 2019. dbscan : Fast Density-Based Clustering with *R*. *J. Stat. Softw.*, 91(1).

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P.,2021.Global land use/land cover with Sentinel 2 and deep learning.*IEEE international geoscience and remote sensing symposium IGARSS. IEEE*, 2021: 4704-4707

Khan, K., Rehman, S.U., Aziz, K., Fong, S. and Sarasvady, S., 2014. DBSCAN: Past, present and future. *IEEE*, pp. 232-238.

Li, R., Tao, M., Zhang, M., Chen, L., Wang, L., Wang, Y., He, X., Wei, L., Mei, X. and Wang, J., 2021. Application Potential of Satellite Thermal Anomaly Products in Updating Industrial Emission Inventory of China. *Geophys. Res. Lett.*, 48(8): e2021GL092997.

Liu, Y., Hu, C., Zhan, W., Sun, C., Murch, B. and Ma, L., 2018. Identifying industrial heat sources using time-series of the VIIRS Nightfire product with an object-oriented approach. *Remote Sens. Environ.*, 204: 347-365.

Ma, C., Li, T., Sui, X., Liao, R., Xie, Y., Zhang, P., Wu, M. and Wang, D., 2024.Annual dynamics of global remote industrial heat sources dataset from 2012 to 2021. *Scientific Data*, 11(1): 631.

Ma, C., Sui, X., Zeng, Y., Yang, J., Xie, Y., Li, T. and Zhang, P., 2022. Classification of Industrial Heat Source Objects Based on Active Fire Point Density Segmentation and Spatial Topological Correlation Analysis in the Beijing–Tianjin–Hebei Region. *Sustainability*, 14(18): 11228.

Ma, Y., Ma, C., Liu, P., Yang, J., Wang, Y., Zhu, Y. and Du, X., 2020. Spatial-Temporal Distribution Analysis of Industrial Heat Sources in the US with Geocoded, Tree-Based, Large-Scale Clustering. *Remote Sens.,* 12(18): 3069.

Stoltzfus, J.C., 2011. Logistic Regression: A Brief Primer. *Acad. Emerg. Med.*, 18(10): 1099-1104.

Sun, S., Li, L., Wu, Z., Gautam, A., Li, J., & Zhao, W., 2020. Variation of industrial air pollution emissions based on VIIRS thermal anomaly data. *Atmospheric Research*, 244: 105021.