Denoising Algorithm of LJ1-01 Nightlight Data with Hybrid Chi-square Distribution

Peng Yang[1], Jingyu Liu[2]

[1] Laboratory of Target Microwave Properties, Deqing Academy of Satellite Applications, Huzhou,313200, China

[2] School of Architectural and Surveying & Mapping Engineering, Jiangxi University of Science and Technology, Ganzhou, 341000, China

Abstract: To distinguish the effective light points caused by human activities from the noise points, and improve the extraction accuracy of feature targets, a denoising algorithm of LJ1-01 nightlight with hybrid Chi-square distribution was proposed. It was assumed that LJ1-01 data and noise subsets are superpositions of multiple Chi-square, and based on the principle of least squares, the indirect adjustment equation was constructed by combining effective statistical lighting value, and the proportion of each superposition state was solved. Next, a continuous probability density curve with sharp fitting accuracy was constructed based on the proportion. Finally, the abundance function was constructed. According to the offset of weight centers of different noise subsets in the original image, noise was removed. To verify the universality and robustness of the algorithm, four regions were selected as experimental objects in China according to the reflective characteristics of discrete objects. The feature similarity of the images after noise elimination is higher than 0.82 and the structure similarity is higher than 0.94. Meanwhile, three regions in Beibu Gulf were selected to extract the number of ships to verify the denoising effect. The results demonstrate that the accuracy of ship extraction using hybrid Chi-square denoising algorithm is above 90%.

Keywords: nightlight; LJ1-01; hybrid Chi-square distribution; denoising; extraction of ship; indirect adjustment

*Third Author, E-mail: ljy_jxust@163.com

1. Introduction

The nighttime remote sensing data is a direct reflection of human nighttime activities [1-4]. These captured light spots are subjected to post-processing such as atmospheric and geometric correction to obtain nightlight with geographical coordinates. The conclusion of the literature shows that nighttime lighting values are highly correlated with population [5], GDP and energy consumption in the region [6-12]. Between 1992 and 2013, the Operational Line Scan System (OLS) carried by the US Defense Meteorological Satellite Program (DMSP) enabled the first use of nighttime lighting data[11, 13], which was have been removed non-urban lights when the data was published and this data could be directly used

for related research. It is the start for remote sensing of nightlight. After 2013, the data was provided by the Visible infrared Imaging Radiometer (NPP-VIIRS) carried by the Suomi National Polar-orbiting Partnership satellite [14-17]. Compared with MSP/OLS, it can not only acquire visible and near-infrared images at night, but also increased spatial resolution by 5x and radiation resolution by 250x, and solves the problem of oversaturation of MSP/OLS lighting data. Obtaining nighttime lighting images with high spatial resolution, high temporal resolution and high radiation resolution will have higher scientific value and expand the research area. In June 2018, the scientific experimental microsatellite LJ1-01 jointly developed by Wuhan University of China and CGSTL was launched [18-20]. The LJ1-01 is a new generation of remote sensing satellite with a resolution of 130 meters, a width of 250km*250km and a temporal resolution of 15 days. Compared with NPP-VIIRS and DMSP/OLS, the LJ1-01 has better space and research value for more detailed research on ship extraction, factory extraction and urban edge demarcation [21]. However, there are too much unordered noise in the nightlight data under different weather conditions and different surface reflectance, which would greatly interfere the extraction and statistical research.

The unordered noise in the LJ1-01 affects the accuracy of the analysis, extraction, and evaluation of the data and must be removed. The LJ1-01 utilizes unordered light removal methods including anisotropic baffles and internal structural optimization to properly eliminate solar unordered light. The unordered noise in the imaging system mainly comes from the strong energy of the sun, the diffuse reflection in the cloud layer and the high

surface reflectivity, etc. Those are confused with the normal energy distribution, so that there are most noise in the image obtained by the sensor, which seriously affect the TNL, the extraction of the boundary definition [22] and the availability of data.

At present, when using and analyzing LJ1-01 data, it is found that LJ1-01 contains sporadic spot noise in deserted depopulated zone, cloud diffuse scattered spot noise in cloudy mountainous areas, ocean wave reflected spot noise and refracted spot noise affected by aerosol [23-26]. If these noises were removed, the LJ1-01 can provide high precision service products such as the GDP, the carbon emission index and the urban vacancy rate index, etc. There are many important significance and reference value to the data mining and utilization on denoising of LJ1-01 data. Therefore, this paper studies the probability distribution model of LJ1-01 nightlight, and analyzes the noise probability distribution in desert, mountain, metropolis, and ocean. In addition, different noises are removed, and the ship extraction [27, 28] combining the algorithm is realized in this paper.

2. Materials and Methods

2.1. Research Area and Data Preprocessing

In order to accurately reflect the denoising effect of the algorithm in different regions, four typical areas of Urumqi, Chengdu, Beijing and Beibu Gulf were selected (Figure 1). Around Urumqi is a desertified area with sporadic light spot noise, Chengdu belongs to the basin topography, most of which has scattered light spots on the cloud, Beijing is a typical refracted spot noise area affected by aerosol, Beibu Gulf is affected by wave reflection and cloud scattered light spot noise. As shown in Figure 1.

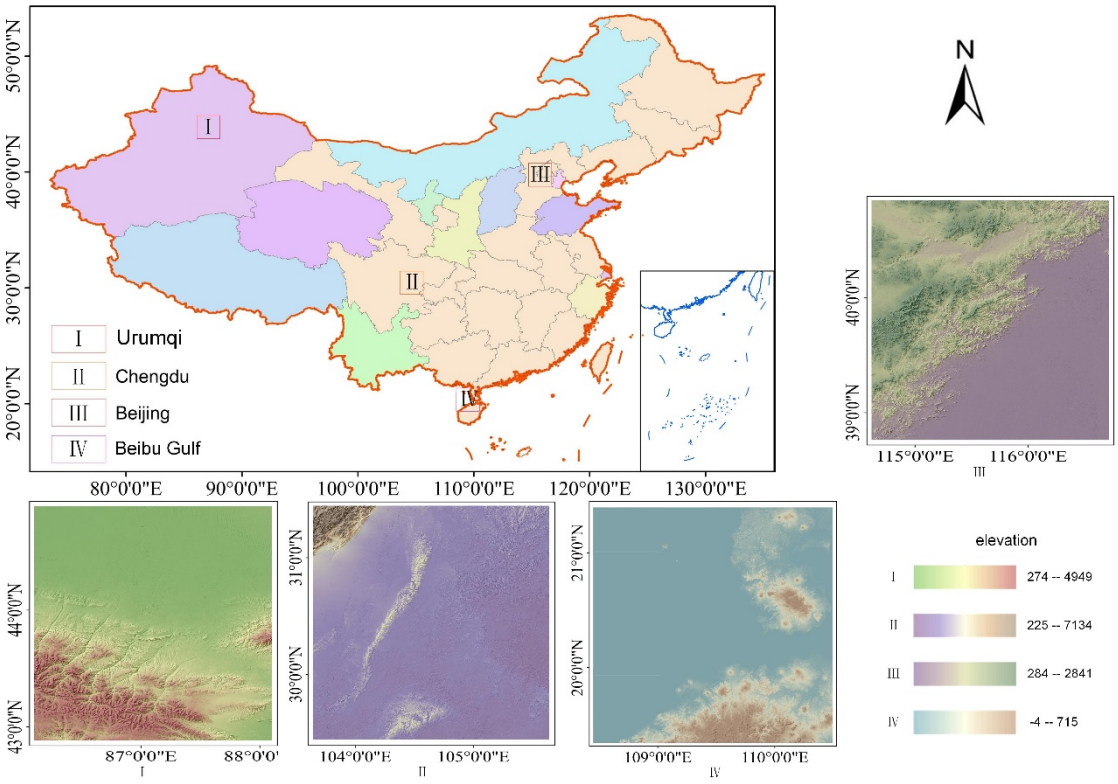

Fig.1 Study area

The LJ1-01 data are provided free of charge by the Hubei High Resolution Earth Observation System Data and Application Network Center (http://59.175.109.173:8888/app/login.html). In this study, the LJ1-01 satellite parameters are shown in Table 1.

Table 1 LJ1-01 parameters

Strip width	**Night-time transit**	**bandpass**	**Value**
~260 km	~22:00(Local time)	Pan 0.46 to 0.98um	Absolute radiation
Spatial resolution	**Calibration**	**Saturation**	**Time series**
3.9 arc-sec (~130 m)	Star calibration	None	2018.6 to date

Although the original LJ1-01 nightlight has been geometrically corrected, the radiation correction is still required before use. The absolute radiation corrected value reflects the light radiation intensity at each point. The radiation conversion formula is as shown in equation (1):

$$L=DN^{3/2}*10^{-10} \quad (1)$$

Where L is the radiance value after absolute radiation correction, the unit is $W/(m^2*sr*\mu m)$, DN is the image gray value of each pixel, and the radiance of LJ1-01 is converted into the center wavelength by this formula.

2.2. Hybrid Chi-square filter model

At present, the LJ1-01 nighttime lighting data is a data type containing effective lighting and noise. Separating the effective light from the noise in this data is the algorithm studied in this paper.

First, since the nightlight satisfies the value of each grid point as a positive number, that is, the light values are distributed in the first quadrant, and both are positive values, and the number of high-brightness grids is only a few. The general probability distribution model does not satisfy this, such as: Normal distribution, T distribution, Cauchy distribution, etc., they still have probability density values in the non-first quadrant. If these probability functions are fitted, the results do not match the actual ones. Moreover, the cumulative probability density of all probability density functions should be 1, that is, the acreage under the density distribution curve is 1.

For any positive integer x, the Chi-square distribution with a degree of freedom $\chi-1$ is a probability distribution with a random variable of X, and Chi-square is distributed within the first quadrant, exhibiting a right skewness, which is a very special Gamma distribution [29-32]. At the same time, the smaller the degree of freedom, the more skewed the distribution, the greater the degree of freedom, the wider the distribution, and the more consistent with the normal distribution. Since the distribution characteristics of Chi-square accord with the distribution structure of LJ1-01 nightlight, this paper chooses the Chi-

square for denoising algorithm research. The probability density formula for the Chi-square is as shown in equation (2):

$$f(\chi,x)=\frac{1}{2^{\chi/2}\Gamma\left(\chi/2\right)}x^{\chi/2-1}e^{-x/2},x>0,\chi\in N^{+} \tag{2}$$

It can be seen from the formula that with the change of the degree of freedom $\chi-1$, there are only 0 or 1 peaks of the Chi-square, and due to the influence of noise, the peak of the probability distribution of the nightlight may be greater than one, therefore, the Chi-square of only one parameter cannot correctly fit the probability distribution of the original light data. At the same time, the existing research is only the synthesis and decomposition of the two peak functions.

In summary, this paper proposes a hypothesis that the probability distribution of the original LJ1-01 data in different regions can be represented by a superposition of a finite number of different parameters of the Chi-square, as shown in equation (3).

$$\begin{cases} g(i,x_i)=\sum_{i=1}^{n}\varepsilon_i f(x_i)=\sum_{i=1}^{n}\varepsilon_i\frac{1}{2^{i/2}\Gamma\left(i/2\right)}x_i^{i/2-1}e^{-x_i/2},x_i>0 \\ g_i(i,x_i)=\varepsilon_i f(x_i),\sum_{i=1}^{n}\varepsilon_i=1,\varepsilon_i>0,i\in(1,2...n) \end{cases} \tag{3}$$

Where $g(i,x_i)$ is the superposition state of each Chi-square distribution, which is the overall probability distribution, $f(x_i)$ is the Chi-square probability distribution of different parameters, ε_i is the weight of the Chi-square of different parameters, $i\in(1,2...n)$ is the parameter of the Chi-square distribution, $g_i(i,x_i)$ is the Chi-square of different parameters probability in the overall distribution. At the same time, it can be verified that the superposition state also meets two initial conditions: 1. Simultaneous integration of both sides of $g(x)$ probability density, the integral result conforms to the cumulative probability density of 1; 2, the starting point probability is 0, and the area where x is less

than 0 has no probability. Since the raw data containing the noise and the effective light is a superposition of the Chi-square of a finite number of different parameters, it is only necessary to solve the weight B of the Chi-square of each single state to fit the high-precision probability density curve. At the same time, it can also solve the problem of probability density coordination when the peak is greater than 1.

Theoretically, it is required to solve n unknowns. The non-least-squares solution requires at least n equations. For a finite number n of weights of the single-state Chi-square distribution, only a limited number of m observations can be observed. The weight assignment between values implements the solution of the unknowns. Since the Chi-square of the states exist independently, if $n > m$, the solution of the weights is the least square approximation solution under the conditional adjustment, if $n \le m$, the solution of the weight is the only solution according to the least squares criterion under indirect adjustment [33-35]. In the case where the observations are selectable, try to use indirect adjustments to get a more accurate solution. The indirect adjustment model is as shown in equation (4):

$$\underset{m\times 1}{V} = \underset{m\times n}{B}\ \underset{n\times 1}{\hat{\varepsilon}} - \underset{m\times 1}{l} \tag{4}$$

$\underset{n\times 1}{\hat{\varepsilon}}$ is n unknowns to be solved, $\underset{m\times n}{B}$ is a coefficient matrix, $\underset{m\times 1}{l}$ is an observation value, $\underset{m\times 1}{V}$ is an error value, $rank(B) = n$,and coefficient b_{ij} in the coefficient matrix is the probability density occupied by x_j in the ith state. The parameters are as shown in equation (5):

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},\quad \hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix},\quad l = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{pmatrix} B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}, \begin{cases} b_{ij} = f(i, x_j) \\ i \in (1, 2...m) \\ j \in (1, 2...n) \end{cases} \tag{5}$$

The basic solution is to make $V^T PV = \min$, P is the weight variance. When the weights of the unknowns are equal, P is the identity matrix, that is, let $V^T V = \min$. It can be obtained by the least squares method: $B^T V = 0$, in summary, $\hat{\varepsilon}=(B^T B)^{-1}*(B^T l)$, the error variance is V after substituting $\hat{\varepsilon}$ into the equation. To satisfy the cumulative probability of the superposition state is 1, the weight should be expressed as: $\varepsilon_i = \hat{\varepsilon}_i / \sum \hat{\varepsilon}$ .At the same time, a representative pure noise is selected as the noise subset A_1 in the same nightlight. The above method is used to separate the weights ε' of the hybrid Chi-square distributions in the noise subset A_1. In order to ensure that the value at each point is greater than 0 when the probability distribution of the original data is subtracted from the probability distribution of pure noise. The probability density of subset A_1 is as shown in equation (6):

$$g'(i, x_i) = \sum_{i=1}^{n} \varepsilon_i \varepsilon_i' f(x_i), x_i > 0 \tag{6}$$

Similarly, the probability density of the k-th noise subset A_k is as shown in equation (7):

$$g^k(i, x_i) = \sum_{i=1}^{n} (\prod_{j=1}^{k} \varepsilon_i^k) f(x_i), x_i > 0 \tag{7}$$

Where ε_i^k is the weight of the k-th noise subset A_k on each Chi-square. At the same time, the sum of the weights must be 1.

$$\sum_{i=1}^{n} \varepsilon_i = 1, \varepsilon_i > 0, \sum_{i=1}^{n} \varepsilon_i' = 1, \varepsilon_i' > 0, i \in (1, 2...n), \sum_{i=1}^{n} \varepsilon_i^k = 1, \varepsilon_i^k > 0, i \in (1, 2...n)$$

The probability density of the light data after denoising is as shown in equation (8):

$$g_0(i, x_i) = g(i, x_i) - \sum_{j=1}^{k} g^j(i, x_i), k \in N^+ \tag{8}$$

Where $g_0(i,x_i)$ is the probability density of the light data after noise is removed; $g(i,x_i)$ is the probability density of the original light data; $g^j(i,x_i)$ is the probability density of the j-th noise subset A_j.

Its effective abundance function $G(i,x_i)$ is expressed as a ratio function of the probability density after denoising and the original probability density: $G(i,x_i)=g_0(i,x_i)/g(i,x_i)$, defining the weight function $C(i)=\varepsilon_i/(2^{i/2}\Gamma(i/2))$, then the abundance function $G(i,x_i)$ can be expressed as shown in equation (9):

$$G(i,x_i)=1-\sum_{i=1}^{k}\frac{\sum C^j(i)x^{i/2-1}e^{-x/2}}{\sum C(i)x^{i/2-1}e^{-x/2}}=1-\sum_{i=1}^{k}\frac{(\sum x^{i/2}C^j(i))/(\sum x^{i/2})}{(\sum x^{i/2}C(i))/(\sum x^{i/2})}=1-\sum_{i=1}^{k}\frac{\overline{C}^j(i)}{\overline{C}(i)} \quad (9)$$

Where $\overline{C}^j(i)$ is the center of gravity of the j-th noise subset A_j weight function, and $\overline{C}(i)$ is the center of gravity of the original light data weight function. It can be seen from the above formula that the effective abundance function is the difference between the gravity data of the original data and the weight function of each noise subset. Therefore, the threshold data can be set according to the content of the abundance function to remove the partial light data of the bias noise, and the final result is the culling noise. After the effective light data, for example, when $G(i,x_i)\geq 90\%$ is valid light data, you can filter out the part of the data with abundance function greater than $\geq 90\%$. As long as the appropriate threshold is set, the original data can be filtered by noise, so that most of the noise is eliminated. The light data after noise elimination can be applied to urban index calculation, GDP forecast and material extraction to obtain better results.

3. Results and Discussion

3.1. Hybrid Chi-square filter

The noise in the lighting data will lead to an increase in the TNL of the area and the practicability of the data, and the noise data containing the noise will seriously affect the extraction accuracy of the object entity. Therefore, the elimination of the lighting noise is necessary. Due to the different causes of noise in different regions, different noise subsets are needed to fit the hybrid distribution. Then the whole image is preliminarily identified by visual discrimination. Finally, the noise is removed according to the deviation of the barycenter of the weight coefficient of the noise subset relative to the original data. Experimental results are shown in Figure 2.A(1), A (2) and A (3) is the verification area, where A (3) is the sample subset and B (1), B (2) and B (3) is the area corresponding to the verification area after noise elimination.

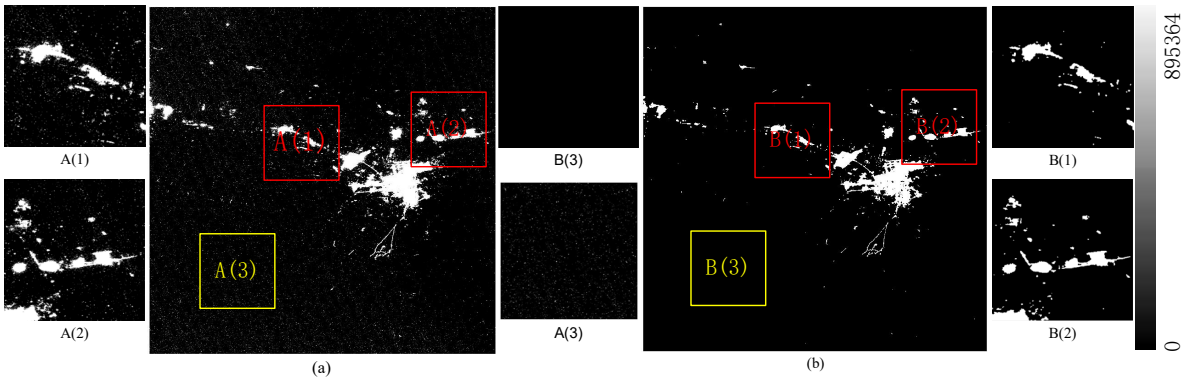

Fig.2 Denoising in Urumqi. (a): Before denoising. (b): After denoising.

In Figure 2, A (3) is located in an uninhabited desert area, and its light value is caused by non-human activities. Compared with A (3) and B (3), the noise points are basically eliminated. It can be seen from A (1), B (1), A (2) and B (2) that the lighting boundary of the urban area is very obvious, the surrounding unordered light removal effect is better, and the characteristics of the urban main light are completely preserved. Therefore, the algorithm has a good rejection effect on the lighting noise caused by non-human activities.

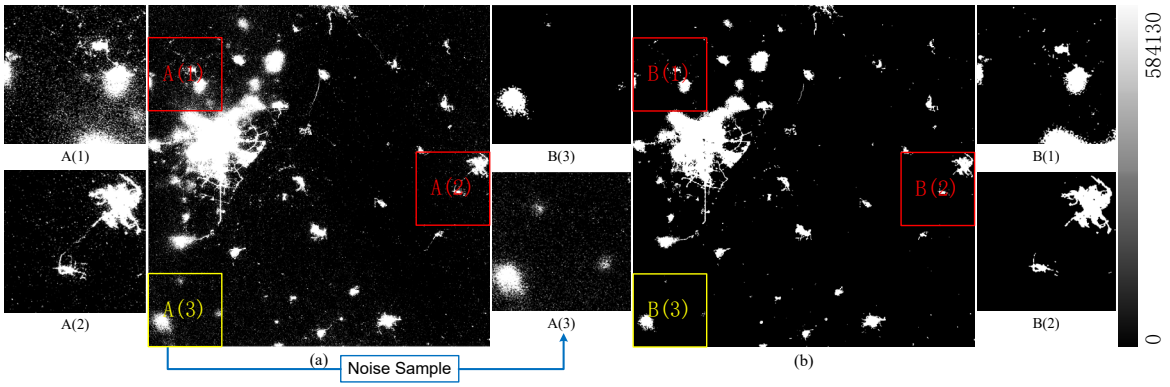

Fig.3 Denoising in Chengdu. (a): Before denoising. (b): After denoising.

In Figure 3, the A (3) region is enveloped by clouds, and its noise is mainly derived from atmospheric scattering and refraction. By comparing the effects before and after the denoising, it can be seen that the unordered light spots are basically eliminated. Meanwhile, the algorithm does not treat all the light values in the noise subset as noise, such as A (3) and B (3), the algorithm follows the center of gravity offset of the weighting coefficient of the noise subset calculates the effective abundance, and then combines the threshold of the effective abundance to determine whether it is a noise point, so the urban area lights in B (3) remain. As shown in A (1) and B (1), if the cloud layer is very thick and the feature information is seriously missing, the algorithm cannot extract the ground light information by eliminating the noise. In other words, the algorithm can eliminate the light scattering noise caused by the thin cloud layer. In A (2) and B (2), the boundary of the urban area is more obvious after noise is removed, and the surrounding unordered light removal effect is better. In summary, the algorithm in this paper has better ability to reject noise in light scattering and refraction caused by thin clouds.

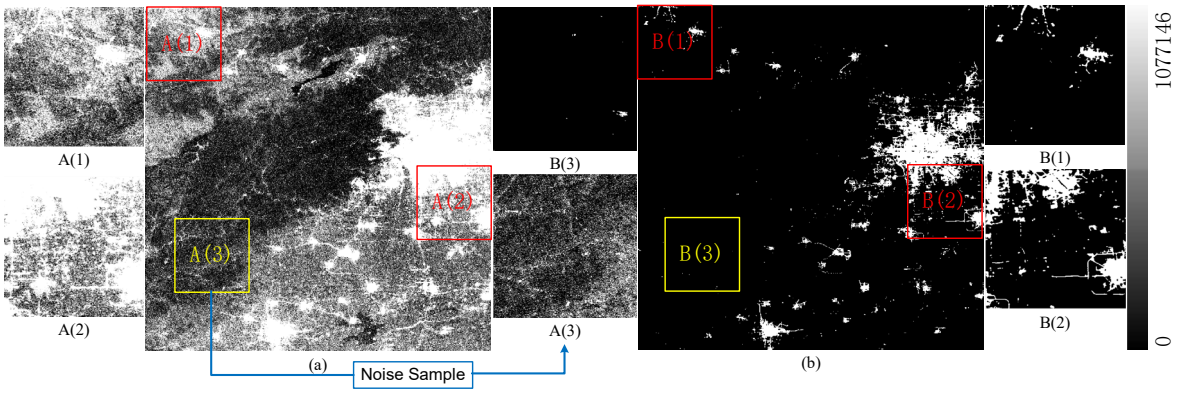

Fig.4 Denoising in Beijing. (a): Before denoising. (b): After denoising.

In Figure 4, A (3) is located in a forest area, and its main source of noise is atmospheric scattering and refraction rather than clouds. From the comparison before and after the noise is removed, the unordered light of this scene in Beijing is obviously more than that of Urumqi and Chengdu. Unlike the local influence of the thin cloud layer, the whole image is affected by atmospheric scattering, refraction and reflection. The algorithm of this paper also has better rejection effect for this kind of noise. As shown in A (3) and B (3), there are more real noises in the noise sample subset, so the rejection effect is better. As shown in A (1) and B (1), due to the influence of scattering, the light data is very messy. After the noise is removed, the city boundary is obvious. In A (2) and B (2), there are more unordered lights in the main city of Beijing, but the algorithm in this paper can still distinguish between Fangshan District and Daxing District. From above, the algorithm also has good rejection ability for light scattering, refraction and reflection caused by large particles in the air.

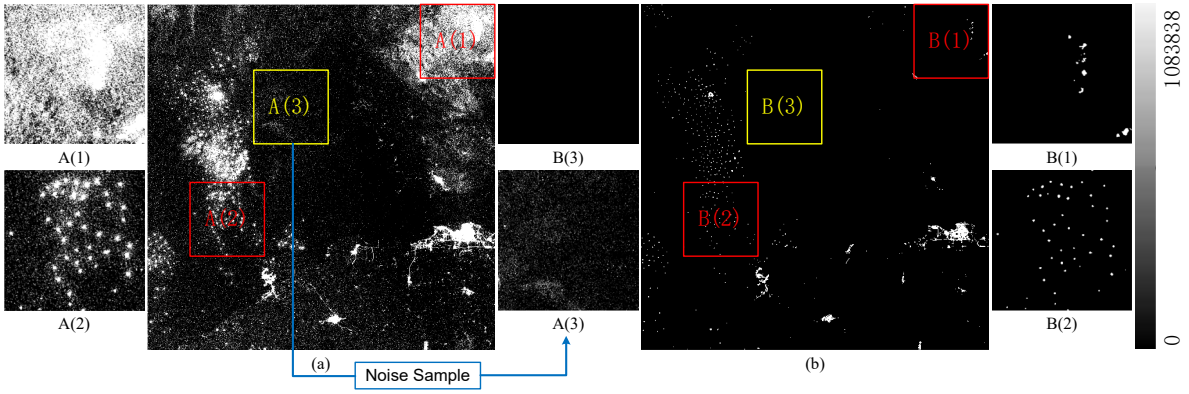

Fig.5 Denoising in Beibu Gulf. (a): Before denoising. (b): After denoising.

In Figure 5, A (3) is located in the immediate vicinity of the Beibu Gulf, and its noise mainly comes from the reflection, refraction, scattering and some clouds in the sea surface. It can be seen from the comparison that the sea surface unordered light removal effect looks good, but the excessive denoising appears on the land due to the influence of thick clouds. As shown in A (3) and B (3), It works really well in noise rejection for unmanned sea surfaces. Shifted to A (1) and B (1), the cloud layer is thicker, and the ground light cannot be correctly classified after the attenuation feature is missing. That is, the light that has been attenuated through the thick cloud layer is also removed as noise. The A (2) and B (2) as shown this is mainly for the fishing lights on the sea surface. After the noise is removed, the fishing boat lights are more prominent, and accurate fishing boat extraction research can be carried out. In conclusion, this algorithm has better ability to remove noise in the light reflection, refraction, scattering and other problems caused by sea waves, but it still cannot solve the problem of light attenuation caused by thick clouds.

Table 2 Weighting factor table for hybrid X distribution

City	Original data						Denoising data					
	ε_1	ε_2	ε_3	ε_4	ε_5	R^2	$\varepsilon_1^{'}$	$\varepsilon_2^{'}$	$\varepsilon_3^{'}$	$\varepsilon_4^{'}$	$\varepsilon_5^{'}$	R^2
Urumqi	0.2309	0.2253	0.2088	0.1831	0.1519	0.8085	0.2148	0.22	0.2119	0.1912	0.1621	0.9098
Chengdu	0.2141	0.2204	0.2125	0.1913	0.1617	0.8895	0.1808	0.208	0.2179	0.2087	0.1847	0.9455
Beijing	0.1954	0.2166	0.2175	0.2	0.1706	0.9136	0.18	0.2103	0.2198	0.2082	0.1817	0.9640
BeibuGulf	0.2232	0.2237	0.2109	0.1866	0.1556	0.8353	0.2161	0.2216	0.2125	0.1901	0.1597	0.8869

As shown in Table 2, the Urumqi region is affected by non-human factors, and the Chi-square distribution with a degree of freedom of 0 is a higher proportion. Under the influence of weather, Chi-square with degree of freedom 1 occupies a higher proportion in Chengdu and Beibu gulf. Affected by atmospheric scattering, and the Chi-square with a degree of freedom of 2 accounts for a higher proportion in Beijing. For the noise subset, the Chi-square with a degree of freedom of 1 is the dominant distribution in the desert terrain. In forest and plain areas, the Chi-square with a degree of freedom of 2 is the dominant distribution, while the Chi-square with a degree of freedom of 0 is the dominant distribution in sea surface. It can be explained that the noise types of different regions have different probability density distributions, and the denoising for different regions should be combined with the distribution of noise in the region. R^2 is accuracy of fitting and they were higher than 0.8.

In order to qualitatively analyze the denoising ability of the proposed algorithm, seven image quality evaluation indicators were selected to evaluate the quality of the image after denoising. PSNR, SSIM and FSIM are indicators with reference evaluation, and Avegrad, Entropy and MSE are indicators without reference evaluation. PSNR is peak signal noise, which is generally 20db-50db, SSIM is structural similarity and the FSIM is feature similarity, both of which are between 0 and 1, the higher the value, the better the image quality after noise elimination. Avegrad is the average gradient, which can sensitively reflect the contrast of small details of the image, and it is mainly used to evaluate the degree of blurring of the image. The Entropy can evaluate the degree of confusion in the data, the larger the value, the more confusing, and the MSE can evaluate the degree of change of the data, the smaller the value, the better the accuracy.

Table 3 Image quality after denoising (Evaluation indicators)

City	Available indicators	Non-referential indicators

	PSNR	SSIM	FSIM	Avegrad	Entropy	MSE	Time consuming
Urumqi	36.3382	0.9976	0.8244	0.001	0.17	6.8794	1.4222
Chengdu	43.5825	0.9962	0.9802	0.0004	0.3	2.5219	2.0204
Beijing	34.6031	0.9417	0.9508	0.0012	0.3	10.4688	2.7540
Beibu Gulf	34.1463	0.983	0.9462	0.001	0.06	4.3634	2.1441

As shown in Table 3, the images with the noise removed by the algorithm have superior image quality. Under different circumstances, the output results maintain a structural similarity greater than 0.94 and a feature similarity greater than 0.82, while having a better peak signal-to-noise ratio, that is, most effective signals in the denoised images are saved. For a single image of 250km*250km, the average time is 2.1s. Experiments show that the algorithm in this paper not only has a high denoising accuracy but also a high operational efficiency, it has practicality in preprocessing of large quantities of data. Moreover, the experiment above only uses the subset of order 1 noise as the noise sample. The noise sub-set space will get better denoising effect if choose higher-order noise subset space.

3.2. Ship extraction

In order to further analyze the application ability of this algorithm to target extraction in light data, the light data mentioned above in Beibu Gulf were selected and the morphological clustering extraction algorithm was combined with the algorithm in this paper for ship extraction. As shown in Figure 6, A (1), A (2) and A (3) are the three sea surfaces in the original data, and B (1), B (2) and B (3) are the corresponding positions after noise elimination. In A (1) and B (1), the positions of the ships are relatively concentrated, and the lights between the similar ships interfere with each other, so the discriminability is poor. For the original data, only relatively discrete ships can be extracted by morphological clustering method, while the image after noise elimination can be extracted completely with complete shape. As shown in A (2), since the high-intensity light passes through the

reflection and refraction of the water surface, the direct extraction will also extract the light reflected from the sea surface as a ship, and the extraction quantity far exceeds the true value. In B (2), the ship can be extracted while the noise is basically eliminated. The discrimination of independent ships will be worse when two ships are very close to each other. After the algorithm removes the noise reflections such as water reflection, scattering and refraction, the independent ship can be distinguished and extracted more effectively.

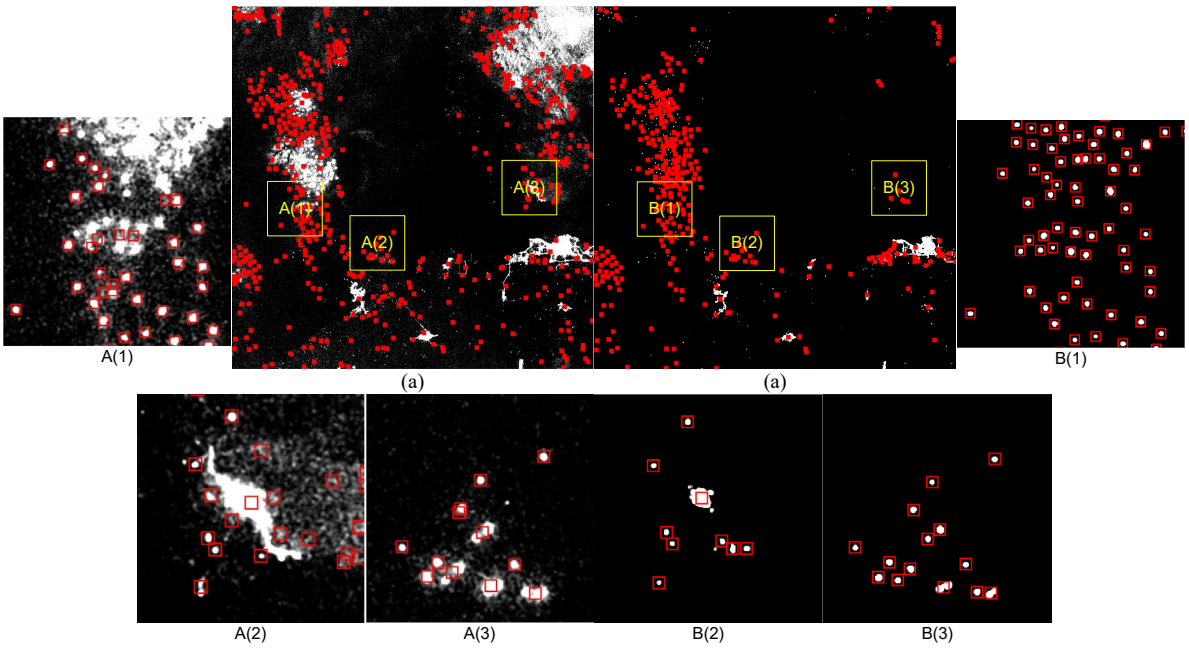

Fig.6 Ship extraction in Beibu Gulf. (a): Before denoising. (b): After denoising.

Experiments show that in the extraction of ships, the effect of direct extraction of raw data is poor, and different noises have a huge impact on the extraction accuracy. The specific extraction accuracy is shown in Table 4.

The actual number of ships is known in the area (1), (2) and (3), and the original map and the number of ships extracted after the noise is removed are obtained. If the extraction algorithm is directly used, the extraction accuracy is only 80%, and the number of extractions in the region (2) is much larger than the actual number. The analysis shows that the identification error is caused by the influence of noise, and the effective accuracy of

region (2) is only 70% after visual discrimination. However, the ship's extraction after hybrid Chi-square distribution filtering algorithm has a minimum extraction accuracy of 90%.

Table 4 Ship extraction accuracy in different areas

Area	True number of ships	Identification number of ships before denoising	Accuracy	Identification number of ships after denoising	Accuracy
(1)	62	35	0.538	62	1.000
(2)	10	17 (7)	1.700 (0.700)	9	0.900
(3)	15	12	0.800	14	0.933

In conclusion, the algorithm in this paper can rapidly remove the noise of LJ1-01 original lighting data, and can effectively eliminate noise under different environments, indicating that the algorithm has better robustness. This algorithm has high extraction accuracy when applied to ship extraction, which indicates that this algorithm has good practicability and effectiveness in practical application.

4. Discussion

4.1. The effect of cloud quantity

The results of experiments in Chengdu and Beibu Gulf show that the light value is greatly attenuated by the thick clouds, which causes the collected images to lose their original features. And the algorithm mainly studies the probability density of statistics, so it cannot correct the influence of cloud cover in physics. However, the temporal resolution of LJ1-01 is 15 days. Assuming that there is no emergency such as power failure, the amount of lights in an area would not change significantly. Therefore, the accuracy of identification and extraction of fixed ground objects can be improved by merging the time series data of more than two scenes. But the integration is only for fixed objects such as cities, not for

movable lights such as ships. The correction process of cloud cover impact is shown in figure 7.

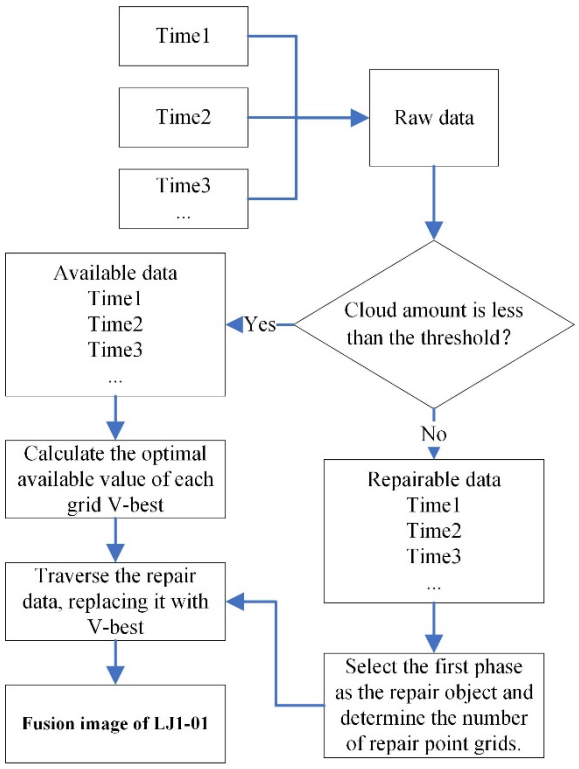

Fig.7 Fuses the method of cloud removal

In Figure 7, the cloud cover of each data is first judged and divided into available data and repairable data. Then, the data and the grid points that needs to be repaired are selected and calculated. Then, the optimal value v-best of the same grid point location is found in the available data, and the final fused image is obtained by replacing one by one.

4.2. The effect of probability density on fitting accuracy

It can be seen from the analysis that the fitting results of the probability density model of hybrid Chi-square distribution selected in this paper have a correlation coefficient of more than 0.8 with the original distribution and the accuracy is high. However, the model is limited. The Chi-square distribution is a special case when alpha takes half of the integer

and beta takes constant 2 in the Gamma distribution, and the Chi-square distribution is shown in Figure.8 (a).

Figure 8 (b), (c) and (d) shows the probability density distribution curves with different values for alpha and beta. With the increase of beta, the overall function value becomes flatter, while the wave peak moves towards the positive half axis of x, so a more accurate probability density curve can be fitted. Therefore, the validation accuracy of noise can be improved by enhancing the fitting accuracy of Gamma distribution.

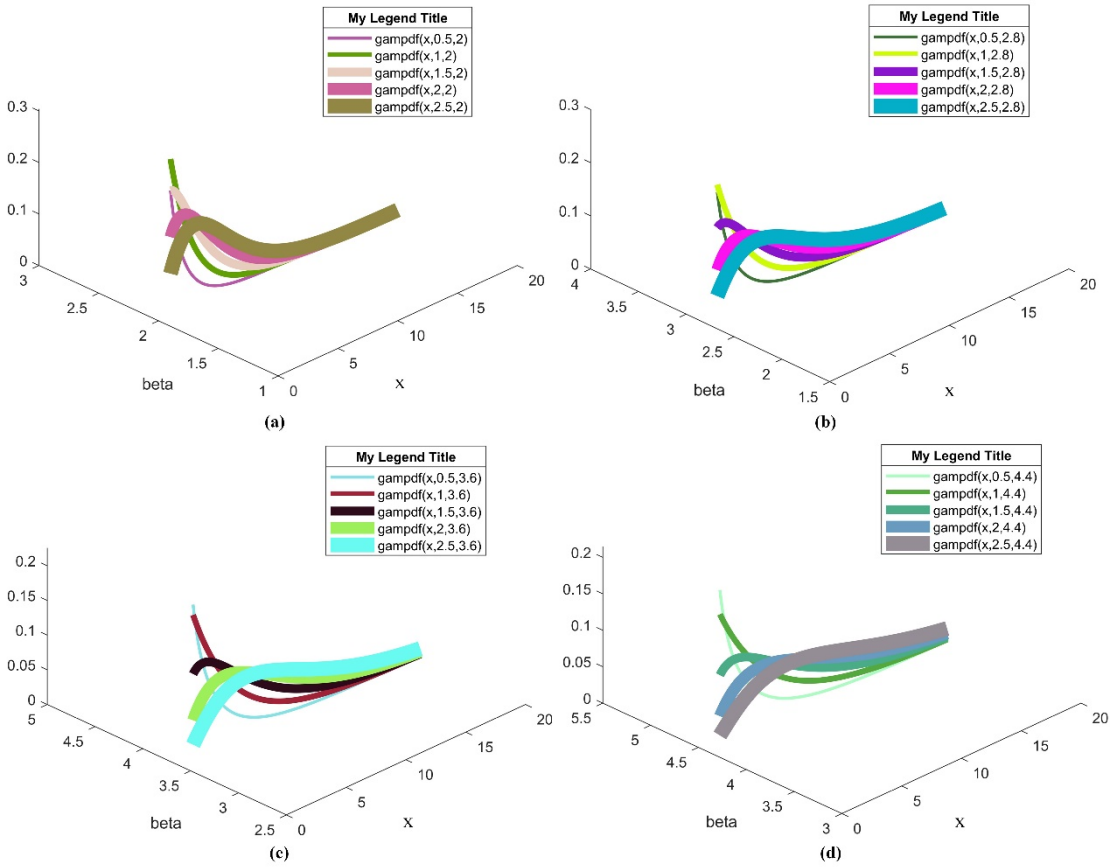

Fig.8 Chi-square distribution in different parameters

Figure 8 (a) shows the Chi-square distribution curve with alpha value of 0.5, 1, 1.5, 2, 2.5 and beta value of 2. Figure 8 (b) shows the Chi-square distribution curve with alpha value of 0.5, 1, 1.5, 2, 2.5 and beta value of 2.8. Figure 8 (c) shows the Chi-square

distribution curve with alpha value of 0.5, 1, 1.5, 2, 2.5 and beta value of 3.6. Figure 8 (d) shows the Chi-square distribution curve of alpha with 0.5, 1, 1.5, 2, 2.5 and beta with 4.

Since Chi-square distribution is a special Gamma distribution, the common Gamma distribution should be selected as the basis of the algorithm in the subsequent research.

4.3. The effect of algorithm on time series

The effect of using this algorithm to eliminate noise has been verified in previous experiments, but the reservation of real lighting data still needs to be verified. Three data of LJ1-01 from Urumqi were selected for the experiment, and (a), (b) and (c) were used as the comparison and verification areas. And the results are shown in Figure.9 below.

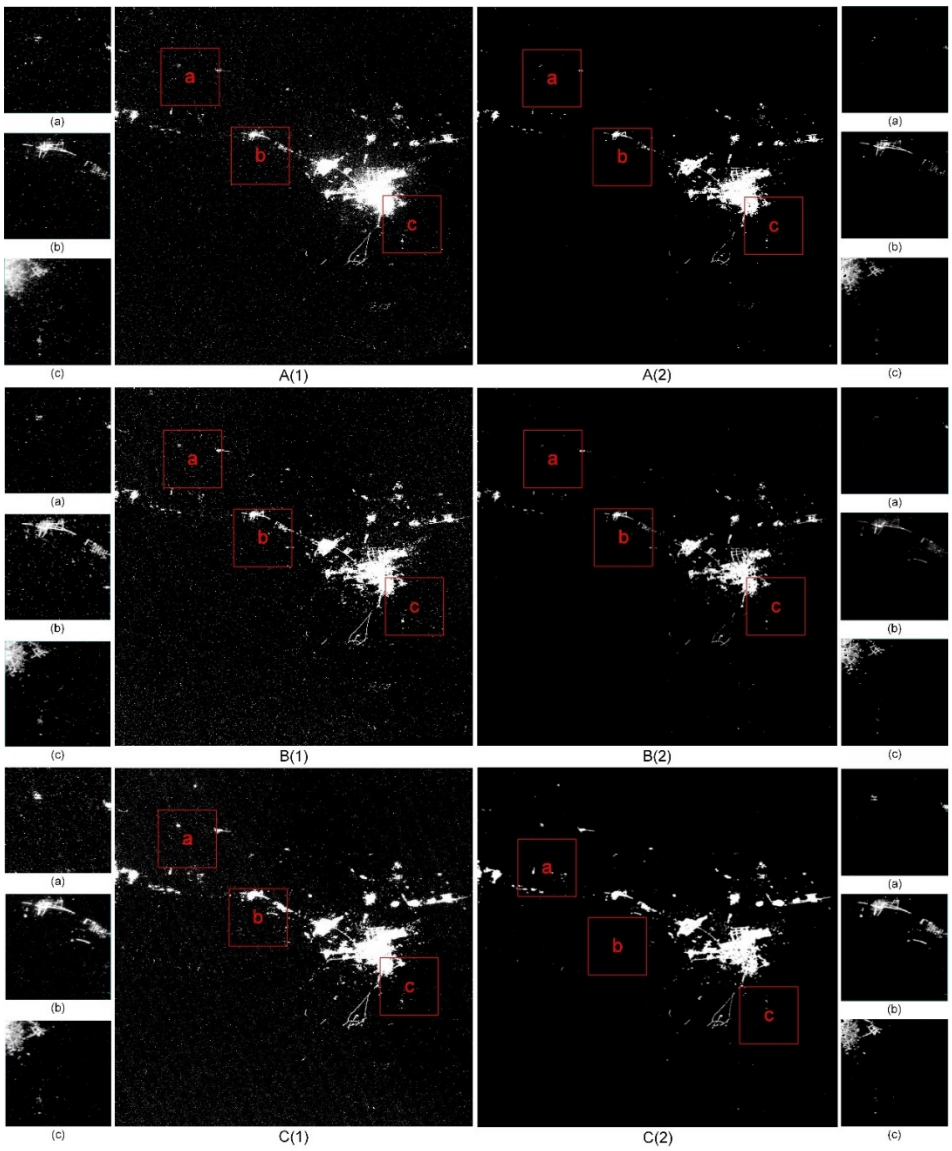

Fig.9 Experimental results of Urumqi on time series. A: Data in 20181013. B: Data in 20181030. C: Data in 20190205. (1): Before denoising. (2): After denoising.

Table 5 The number of light grids in research area

Area	20181013		20181030		20190205	
	A1	A2	B1	B2	C1	C2
a	4708	73	4220	89	10942	82
b	6838	3409	3756	3337	11265	3337
c	5740	3432	4743	3331	4618	3390

As shown in Figure 9, the algorithm presented in this paper has a good noise elimination effect on all the data of the third phase. According to the number of light grids before and after denoising in statistical areas (a), (b) and (c) (Table 5), it is found that

before noise elimination, the number of pixels of data in different periods has a great change, and the effective number of light pixels is relatively stable after noise elimination. It can be seen that this algorithm has good universality and robustness.

5. Conclusion

The filtering of LJ1-01 data noise points is an important part of high-precision extraction. At the same time, the hybrid Chi-square distribution is also an important component of probability density. Based on this, we propose a hybrid Chi-square distribution denoising model. By choosing the original data and noise subset to fit a high precision and continuous probability density curve, and then calculating the abundance function according to the offset of the weight center of the noise subset relative to the original data, and the abundance function determines whether it is contaminated by noise, and finally extracts the area that is not contaminated by noise.

In order to verify the generality and robustness of the algorithm, we consider and take into account the difference of reflective characteristics of different objects. Four typical regions in China are selected as experimental area. Different causes of noise in different areas. For example, t Urumqi is a low-brightness light value caused by a high-reflectivity desert, in Chengdu, there is a reflection and scattering of light values caused by the influence of clouds and mountains; in Beijing, the atmospheric aerosol particles causes the light scattering and reflection; The Beibu Gulf area is affected by the reflection of the waves and the scattering of clouds. According to the noise characteristics of different regions, a suitable noise subset is selected for the algorithm. The results after denoising show that the similarity of features is higher than 0.82 and the similarity of structures is higher than 0.94. That is to say, the feature points are still retained after denoising, and the quality of denoising is better. Considering the application of the algorithm in the actual situation, the neighboring sea area of Beibu Gulf is selected as the research region, and the

advantage of the algorithm is analyzed be compared with the ship extraction algorithm. Due to the algorithm removes the influence of noise in advance, the extraction accuracy of this algorithm is more than 90%, which is obviously higher than the general extraction algorithm.

This is the first time that the hybrid Chi-square distribution method has been applied to LJ1-01 data's denoising. Limited validation results indicate that the accuracy of the hybrid Chi-square distribution method can meet the accuracy requirement of the hydrological, meteorological, feature extraction, urban and rural boundary analysis, etc. The better performance of this method is not to eliminate the noise when dealing with data in different time in the same area. The algorithm is effective except that the accurate light value cannot be recovered proportionally after being attenuated by the atmosphere.

More validation data from different regions in different time with other be collected for further validation in the future. Also, more extensive comparison should be conducted before carried out the high precision fitting using LJ1-01 data。Although the fitting accuracy of this model is high, its essence is just a gamma distribution equation of special point position. In the future research, we will simplify the complex, from special to general, and establish a new algorithm with practical, universal, efficient and high precision.

Acknowledgments: We give thanks to the LJ1-01 night-lighting data provided free of charge by the Hubei High Score Center, the data are available from http://59.175.109.173:8888/app/login.html. The research results of this manuscript benefited from Professor Zhang Guo of Wuhan University and Professor Yu Bailang of East China Normal University in this field. These views, minds, and findings contained in this article are all original to the authors and there is no conflict of interest.

References

1. Lee, D.; Trisopal; Lggy. Application of nightlight satellite imagery for assessing flooding potential area in the Mekong river basin. *Journal of Korea Water Resources Association* **2018,** *51* (7), 565-574.
2. Kim Jung-ah; Cheon Sanghyun. Monitoring NightLight Emission Changes in Seoul Using Satellite NightLight Data to Mitigate Light Pollution. *Journal of Korea Intitute of Spatial Design* **2018,** *13* (6), 165-178.
3. Elshorbagy, A.; Bharath, R.; Lakhanpal, A.; Ceola, S.; Montanari, A.; Lindenschmidt, K.-E., Topography- and nightlight-based national flood risk assessment in Canada. *Hydrology and Earth System Sciences* **2017,** *21* (4), 2219-2232.
4. Zgurovsky, M.; Putrenko, V.; Dzhygyrey, I.; Boldak, A.; Yefremov, K.; Pashynska, N.; Pyshnograiev, I.; Nazarenko, S.; Ieee, *Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine*. 2018; p 8-12.
5. Lloyd, C. T.; Sorichetta, A.; Tatem, A. J., High resolution global gridded data for use in population studies. *Scientific Data* **2017,** *4*.
6. Zhang, G.; Guo, X.; Li, D.; Jiang, B., Evaluating the Potential of LJ1-01 Nighttime Light Data for Modeling Socio-Economic Parameters. *Sensors* **2019,** *19* (6).
7. Li, X.; Zhao, L.; Li, D.; Xu, H., Mapping Urban Extent Using Luojia 1-01 Nighttime Light Imagery. *Sensors* **2018,** *18* (11).
8. Ouyang, Z.; Lin, M.; Chen, J.; Fan, P.; Qian, S. S.; Park, H., Improving estimates of built-up area from night time light across globally distributed cities through hierarchical modeling. *Science of the Total Environment* **2019,** *647*, 1266-1280.
9. Huang, Z.; Du, X.; Castillo, C. S. Z., How does urbanization affect farmland protection? Evidence from China. *Resources Conservation and Recycling* **2019,** *145*, 139-147.
10. Arellano, B.; Roca, J., Defining urban and rural areas: A new approach. In *Remote Sensing Technologies and Applications in Urban Environments Ii*, Erbertseder, T.; Chrysoulakis, N.; Zhang, Y., Eds. 2017; Vol. 10431.
11. Yu, S.; Zhang, Z.; Liu, F., Monitoring Population Evolution in China Using Time-Series DMSP/OLS Nightlight Imagery. *Remote Sensing* **2018,** *10* (2).
12. Liu, X.; Ou, J.; Wang, S.; Li, X.; Yan, Y.; Jiao, L.; Liu, Y., Estimating spatiotemporal variations of city-level energy-related CO2 emissions: An improved disaggregating model based on vegetation adjusted nighttime light data. *Journal of Cleaner Production* **2018,** *177*, 101-114.
13. Li, G.; Cai, Z.; Liu, X.; Liu, J.; Su, S., A comparison of machine learning approaches for identifying high-poverty counties: robust features of DMSP/OLS night-time light imagery. *International Journal of Remote Sensing* **2019,** *40* (15), 5716-5736.
14. Mohan, P.; Strobl, E., The short-term economic impact of tropical Cyclone Pam: an analysis using VIIRS nightlight satellite imagery. *International Journal of Remote Sensing* **2017,** *38* (21), 5992-6006.
15. Zhong Liang, L. X.-s., Yang Peng, Method for SNPP-VIIRS nighttime lights images denoising. *Bulletin of Surveying and Mapping* **2019,** *03*, 21-26.
16. Wang, L.; Fan, H.; Wang, Y., An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data. *International Journal of Remote Sensing* **2019**.
17. Jiang, S.; Li, J.; Duan, P.; Wei, Y., An image layer difference index method to extract light area from NPP/VIIRS nighttime light monthly data. *International Journal of Remote Sensing* **2019,** *40* (12), 4839-4855.
18. Su, Z.; Zhong, X.; Zhang, G.; Li, Y.; He, X.; Wang, Q.; Wei, Z.; He, C.; Li, D., High sensitive night-time light imaging camera design and in-orbit test of luojia1-01 satellite. *Sensors (Switzerland)* **2019,** *19* (4).

19. Zhang, G.; Li, L.; Jiang, Y.; Shen, X.; Li, D., On-Orbit Relative Radiometric Calibration of the Night-Time Sensor of the LuoJia1-01 Satellite. *Sensors* **2018,** *18* (12).

20. Wang Lei, C. R., Li Deren, Yu Baoguo, Quality Assessment of the LEO Navigation Augmentation Signals from Luojia-1A Satellite. *Geomatics and Information Science of Wuhan University* **2018,** *43* (12), 2191-2196.

21. Paranunzio, R.; Ceola, S.; Laio, F.; Montanari, A., Evaluating the Effects of Urbanization Evolution on Air Temperature Trends Using Nightlight Satellite Data. *Atmosphere* **2019,** *10* (3).

22. Liu, X.; Ning, X.; Wang, H.; Wang, C.; Zhang, H.; Meng, J., A Rapid and Automated Urban Boundary Extraction Method Based on Nighttime Light Data in China. *Remote Sensing* **2019,** *11* (9).

23. Hutchins, M. G.; Colby, J. D.; Marland, G.; Marland, E., A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States. *Mitigation and Adaptation Strategies for Global Change* **2017,** *22* (6), 947-972.

24. Kim Jung-ah; Cheon Sanghyun. A Methodology to Produce Light Pollution Map and Diagnose Urban Nightlight Conditions Using International Space Station Nighttime Image Data. *Journal of Korean Institute of Information Technology* **2018,** *16* (12), 13-24.

25. Han, J.; Meng, X.; Liang, H.; Cao, Z.; Dong, L.; Huang, C., An improved nightlight-based method for modeling urban CO2 emissions. *Environmental Modelling & Software* **2018,** *107*, 307-320.

26. Wei, J.; Peng, Y.; Mahmood, R.; Sun, L.; Guo, J., Intercomparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products. *Atmospheric Chemistry and Physics* **2019,** *19* (10), 7183-7207.

27. Salembier, P.; Liesegang, S.; Lopez-Martinez, C., Ship Detection in SAR Images Based on Maxtree Representation and Graph Signal Processing. *Ieee Transactions on Geoscience and Remote Sensing* **2019,** *57* (5), 2709-2724.

28. Lin, Z.; Ji, K.; Leng, X.; Kuang, G., Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images. *Ieee Geoscience and Remote Sensing Letters* **2019,** *16* (5), 751-755.

29. Ding Hai-yong, S. W.-z., Fast N-FINDR algorithm for endmember extraction based on chi-square distribution. *Journal of Remote Sensing* **2013,** *17* (01), 122-137.

30. Fu Chao, H. X.-j., Yu Hong-yan, Lai Yu-yang, Unified multidisciplinary reliability analysis by chi-squared distributions. *Computer Integrated Manufacturing Systems* **2017,** *23* (07), 1439-1446.

31. Wang Xiao-bo, Y. H.-x., Wang Guo-hong, Guan Cheng-bin, AMulti-threshold Classif ication Method for Airborne /Sea Target based on Chi-square Parameter Testing. *Fire Control and Command Control* **2009,** *34* (07), 4-8.

32. Chen Gang, W. M.-j., Asymptotic Expansions of the Probability Density Function and the Distribution Function of Chi-square Distribution. *Journal of Nanjing Normal University(Natural Science Edition)* **2014,** *37* (03), 39-43.

33. Guo Hui, X. L.-j., Analysis and realization of algorithm in traverse network indirect adjustment. *Science of Surveying and Mapping* **2014,** *39* (03), 107-110.

34. De-hu, C., Comparson among Three Algorithms of Parameter Adjustment Model with Inequality Constraints. *Journal of Geodesy and Geodynamics* **2013,** *33* (02), 41-44.

35. Zhao De-shen, G. P., Relationship between adjustment of condition observations and adjustment of indirect observations. *Journal of Liaoning Technical University* **2003,** *03*, 320-322.

Caption List