Operation task extraction method and application of rice operation machinery in South China based on GNSS

Li Xiaobing1,a, Yang Jianfeng1,b*
1China Electronic Product Reliability and Environment Testing Research Institute, Guangzhou, China
*aemail: maybayer@163.com
*Corresponding author email: bwangt@scut.edu.cn, yangjeff2013@163.com

Abstract: According to the characteristics of rice planting and rice operation machinery in South China, the paper discusses how to use the satellite of the operation machinery to position trajectory information, to extract operation task by field based on the clustering algorithm of OPTICS, and to calculate the operation area by broken line buffer algorithm. The paper also discusses about parameter setting of OPTICS. The accuracy of extracting tasks by field can be over 84%. When the area of each task is over 3 mu (0.2 hectares), the calculation accuracy can reach over 95%. On the surface of practical application in Guangdong Province in recent three years, the algorithm can meet the requirements of operation charging and government subsidy giving, and realize accurate task differentiation and area calculation.

1. Introduction
South China is a traditional area of rice planting. In the process of rice planting and harvesting, rice transplanter and harvester have been widely used to complete the operation. At the same time, various sensors and wireless transmission hardware terminal, equipped with GPS / BDS and function of status identification of operation, have been applied to rice operation machinery (Li Hong, et al. 2008, Li Bin, et al. 2012) for remote monitoring and operation counting of agricultural machinery operation. Due to the fee-charging needs of agricultural machinery customers and the government’s demands for subsidies for rice operation machinery such as rice transplanter and harvester, it is necessary to accurately distinguish operation plots (tasks) and calculate the area for billing or accounting.

Real-time spatial data obtained through GPS/BDS equipped on the machinery have been widely used in field of agriculture, including farmland management zoning (Moair Pedroso, et al. 2010, Li Xiang, et al. 2005), soil parameter analysis (Li Yan, et al. 2007) and so on. At present, most researches about application of spatial data in agricultural machinery focus on calculating the operation area based on spatial operation trajectory of agricultural machinery such as obtaining operation area by using distance method (Zhao Bin, et al, 2010), vector buffer algorithm (Liu Hui, et al, 2015) and grid buffer algorithm (Zhu Dengsheng, et al, 2020). In North China, due to the large plots and the fact that these plots often belong to the same farmer, people pay more attention to the area of operation. However, due to the hilly and mountainous landform in South China, the paddy fields in southern China are usually small and irregular. The plots owned by farmers are scattered and discontinuous, and the used rice operation machinery is small. The plots operated by a machine every day may belong to many different farmers. These all leads to a fact that some common operational measurement methods suitable for large plots in North China do not fit the demands of operation segmentation and area
measurement in South China which require the same plot as an independent task. At the same time, current agricultural machinery operation has particularity. For example, the manipulator often forgets to turn off the power supply of the equipment on the machinery, which leads to the agricultural machinery constantly obtaining track points. These track points include effective operation points, and also contain a large number of invalid operation track points, such as staying in the original place, driving on the road, temporarily stopping at the edge of the field plot or going back and forth near the operation field for preparation. Although the additional operation status identification sensors can assist the task differentiation to a certain extent, the judgment of the sensor may be wrong due to the complexity of agricultural machinery operation. At the same time, more sensors increase the cost of hardware system and maintenance.

In this paper, based on the GNSS space trajectory obtained by agricultural machinery operation equipment, the plot (task) division, position information conversion and calculation of the operation area of each plot are automatically carried out in each rice operation machine in continuous time period. The algorithm includes using algorithm of local outlier factor (LOF) to filter non-operation-state points, using the density-based clustering algorithm of OPTICS to divide operation plots and tasks, and using vector buffer algorithm to calculate operation area. The final goal is to improve the operation efficiency of agricultural machinery operators, provide effective help for operation charging and government subsidy accounting, and reduce hardware sensors of agricultural machinery so as to reduce the hardware cost and provide system maintainability.

2. Principles and designs of the algorithm

With the help of GNSS installed on agricultural operation machinery such as GPS or BDS, the real-time longitude and latitude coordinates of agricultural machinery trajectory can be obtained. Generally, according to the operation speed of the machinery, a longitude and latitude coordinate point can be uploaded in 3 to 5 seconds. Usually, WGS-84 coordinate system is used to obtain the coordinates of each trajectory point in the Gauss plane rectangular coordinate system by Gauss Kruger Projection so as to transform the continuous space trajectory points of the agricultural machinery operation into plane continuous discrete points.

2.1. OPTICS (Ordering points to identify the clustering structure)

OPTICS is a kind of density-based clustering algorithm (Mihael Ankerst, et al, 1999), which can be regarded as an improvement of DBSCAN (Density-Based Spatial Clustering of Applications with Noise). DBSCAN is also a very famous kind of density-based clustering algorithm. All the dense regions of the sample points can be found and treated as clusters in algorithm of DBSCAN. Two parameters, neighborhood radius (R) and minimum number of points (minpoints) can actually describe what is called denseness --- when the number of the points in the neighborhood radius R is bigger than minpoints, it reaches denseness. Therefore, we can discover clusters of arbitrary shapes without knowing the number of the clusters. But on the other hand, the clustering results are very sensitive to the values of these two parameters, that is, different values will produce different clustering results. The algorithm of OPTICS overcomes this disadvantage of DBSCAN. It is insensitive to the initial parameter (ε - neighborhood) and the minimum number of points (minpts) and clusters with different densities can be obtained. In addition, the abnormal points can be removed in OPTICS. With this feature, the non-working trajectories of agricultural machinery operation which are regarded as abnormal points such as the walking track on the road can be effectively removed.

2.2. Algorithm of operation area based on broken line buffer

The space track coordinate data collected continuously during the deep ploughing operation forms the actual operation track of the agricultural machinery operation. As shown in Fig. 4, P0, P1,..., P3 are the operation tracks of the rice operation machinery. Because the agricultural machinery operation has a certain width (width of the machinery), the operation process can actually be regarded as the process of forming line segment buffer according to the operation track (Borut Zalik, et al, 2003). The radius R
in the buffer is 1/2 of the operation width. Finally, the process of calculating the cultivated area transforms into the problem of obtaining the buffer boundary according to the trajectory line segment. Then the integral is calculated according to the buffer boundary, and finally the effective area is obtained. Because the boundary of the operation track is obtained, two kinds of abnormal conditions in the process of agricultural machinery operation can be avoided. One is the repetitive operation. In this algorithm, the calculation error caused by repetition can be avoided effectively. The other is the missing operation which is usually reflected as the “hole” in the buffer, and its area will be deducted in the integration process.

![Algorithm principle based on buffer](image)

**3. Algorithm construction**

**3.1. Algorithm**

A) Arrange all the trajectory points on the operation track of continuously operating agricultural machinery in ascending order according to the actual operation time. The Gauss Kruger Projection is carried out in the Gaussian plane rectangular coordinate system to obtain the coordinates of each trajectory point in the coordinate system.

B) Cluster the trajectory points that exclude non-working points, \( X_i (X_i \in D) \), by using the density-based clustering algorithm of OPTICS to obtain the trajectory point set \( Z = Z_1, Z_2, ..., Z_i, ..., Z_n \). \( Z_i \) represents all trajectory points of a field task, and \( Z_i \) is a subset of the all track point set, \( D \). That is, \( Z_i \subseteq D, i = 1, 2, ..., n \), \( n \) is an integer greater than 0;

C) Extract the GPS coordinates of representative trajectory point of each element, \( Z_i \), and obtain the geographic location of the operation task according to “GPS coordinates --- geographic location database DB”;

D) Through vector buffer algorithm, obtain the operation area, \( A_i \), of current operation task, \( Z_i \), to be the final operation area of task \( Z_i \).

**3.2. Parameter setting**

Generally, the speed of rice operation machinery is different according to different types and models. It also means that the density of the coordinate point set is not exactly the same when the obtaining frequency of the latitude and longitude coordinates is the same. Therefore, the parameter selection type of different operation speed is also different.

**3.2.1 OPTIC**

Although the algorithm of OPTICS is not sensitive to the initial parameters, \( \varepsilon \) and \( MinPts \), it is still necessary to select the matching parameters and obtain the appropriate clustering results. Fig. (a) - (e) shows the results of the OPTICS algorithm when the value of \( MinPts \) is fixed, the working speed of the operation machine is 4.5 km/h and the working width is 2 meters. Fig. (a) shows 4 task plots. Fig (b) shows that the clustering result with small \( \varepsilon \) value is 54, which is not consistent with the
situation. Fig. (c) and Fig. (d) show that with appropriate eps values, the number of clusters generated is consistent with the actual situation. The number and of outlier are slightly different, but it can be seen from the table that the final area difference between the two is acceptable in practical application. Fig. (E) shows that when the eps value is high, some outliers generate a new cluster, resulting in the results inconsistent with the actual field task. Fig. (d), (f), (g) show that with a appropriate fixed eps value of 10, the results obtained with smaller (f) or larger (g) MinPts are not consistent with the actual situation.
Fig 2. the results of OPTIC with different eps and MinPts (speed=4.5 km/h; Working width=2m)

Table 1. the results of OPTIC with different eps and MinPts (speed=4.5 km/h; Working width=2m)

| eps | MinPts | Number of clusters | Number of Outliers | Area(667 m²) |
|-----|--------|--------------------|--------------------|--------------|
| 10  | 10     | 7                  | 175                | 43.44        |
| 10  | 15     | 5                  | 202                | 43.44        |
| 10  | 20     | 5                  | 204                | 43.44        |
| 10  | 25     | 5                  | 213                | 43.4         |
| 10  | 30     | 5                  | 246                | 43.39        |

Generally, the density of trajectory points to be clustered is related to the running speed, sampling frequency and operation width of agricultural machinery. In the experiment, select different rice operation machinery, choose the operation speed of 1.5km/h, 3km/h and 6km/h, the sampling frequency of 1s, 2s and 5s, the operation width of 1.8m, 2m and 2.5m, and appropriate eps and MinPts parameter, establish regression model and obtain formula (1) and (2).

\[
\text{eps} = \text{int}(0.571 * s + 0.154 * f + a2 * w + 6.924) \quad (1)
\]

\[
\text{MinPts} = \text{int}(-0.079 * s - 2.897 * f + 29.226) \quad (2)
\]

4. Result and discussion

4.1 Algorithm task splitting effect

Randomly select respective 7 days’ work record of five rice transplanters and five rice combined harvesters in a certain place in South China, totally 305 operation tasks. According to the daily GNSS trajectory, manually account the operation tasks. The results are shown in Table 2. As can be seen from Table 2, the algorithm of OPTIC can achieve good field task splitting. The factors that affect the accuracy of splitting, in addition to the parameter selecting of eps and MinPts corresponding to OPTIC, are usually related to the characteristics of operation plots. For example, because of the close distance, adjacent fields are usually assigned into one operation. In addition, the accuracy of splitting is also often related to the type of the operation. For example, because the rice seedlings are often from a unified supplier in transplanter’s operation and the rice varieties do not change, farmers only need to pay according to the area. This means that agricultural machinery operators often carry out rice seedling transplanting according to adjacent fields rather than farmers to whom the plots belong. As a result, more adjacent fields belonging to different farmers will be assigned into one operation. On the
contrary, fields of rice harvester operations often belong to the same farmer whether they are geographically adjacent or far apart, with a higher accuracy in task splitting.

| Tasks of split | Real paddy fields | Accuracy(%) |
|----------------|-------------------|-------------|
| rice transplanter | 173               | 146         | 84.4         |
| rice combine    | 197               | 192         | 97.5         |

4.2 Accuracy of area calculation

Fig. 3 shows the calculation accuracy results of different operation areas. When the operation area is more than 3 mu (0.2 hectares), the calculation accuracy can reach more than 95%. Generally, the factors affecting the accuracy of the algorithm mainly come from the positioning error of GNSS. When the receiving signal of satellite is good, the civil positioning chip usually has a CEP rate of 3 to 5 meters, which means that the smaller the operation area is, the greater the calculation accuracy is. Due to the mountainous landform in South China, satellite signals and communication signals may become worse, which further decreases the calculation accuracy. Another influencing factor comes from the operation habit and operation method. For example, the transplanter needs to frequently run to the edge of the field to supplement rice seedlings. If the supplementary point of seedlings is far from the operation site, the path that the machine runs for supplement may be calculated as effective operation, which enlarges the area calculation result.

5. Discussion

5.1 Fixed point influencing factors

In practical operation, the manipulator often does not turn off the power supply of the agricultural machinery track acquisition terminal after parking the machinery at the end of the operation, which leads to the continuous acquisition and upload of trajectory data. Due to the characteristics of GPS equipment, the data acquired by GPS will drift in static state. And the data formed by this continuous drift in static state will often form a “cluster” of track points. Due to the different drift sizes of different GPS chips, it is difficult to filter these data in passing speed data, which will lead to errors in the calculation of area and more seriously, lead to abnormal calculation results. Therefore, these data need to be eliminated before calculating the area. It is found that, as shown in Fig. 4 and Fig. 5, when the agricultural machinery is in static state, the longitude or latitude of the satellite positioning chip is always approximate to the parallel line. Therefore, this feature can be used to remove such abnormal data. The method steps are as follows: (1) calculate the difference between the coordinate latitude values of the current and the previous track point. The formula is \( \Delta i = |\text{lati} - \text{lati-1}| \) in which lati
represents the coordinate latitude value of current track point and lati-1 represents the coordinate latitude value of previous track point. If Δi < δ 1, add 1 to DupCnt. DupCnt represents the number of GPS track points with the same latitude value.

Latitude ratio: \( \text{ratio} = \frac{\text{Cnt}}{\text{DupCnt}} \). Cnt represents the number of all GPS track points. If ratio > δ 2, then all trajectory points are considered to be generated by static drift of GPS, otherwise they are normal GPS trajectory points.

Fig. 4. Drift track points of agricultural machinery static chip

Fig. 5. Drift track points of static chip after amplification

5.2. Improvement of calculation efficiency

Fig. 5 shows the calculation time of operations with different numbers of track points. Because of the similar time complexity \( O(m^2) \) between OPTIC and DBSCAN, the time complexity of broken line buffer algorithm is \( o[n \times \log(n)] \). Therefore, with more track points involved in the calculation, the calculation time becomes longer. Especially in busy seasons, there will be a large number of rice machines working at the same time, which brings severe challenges to the computing efficiency of the server. Therefore, it is necessary to filter the original data to improve the calculation efficiency without affecting the area calculation accuracy. Because the boundary is finally obtained by merging the buffers in algorithm of broken line buffer, reducing the trajectory points on the broken line segment can improve the calculation efficiency without affecting the calculation area. The method to filter used in the paper is that if the distance \( d(x_1, x_2) \) between two adjacent time points is less than δ, then throw point x2. It is proved that if δ is set to be less than 0.9, the calculation efficiency can be improved by more than 30%.

Table 3 Comparison of computing efficiency

| Total Points | Before Optimization(ms) | After optimization(ms) | Improvement (%) |
|--------------|-------------------------|------------------------|----------------|
| 7113         | 9420                    | 6469                   | 31.3           |
| 5040         | 7333                    | 4861                   | 33.7           |
| 3020         | 5467                    | 3835                   | 29.9           |
| 1005         | 1350                    | 854                    | 36.7           |

*Run on: CPU: Intel Core(TM) i7-10710U, RAM: 16GB

6. Conclusions

Obtain the operation space track of rice operation machinery through the satellite positioning system of equipment on machinery, and then extract the operation task and measure the operation area by mathematical method. The results of the stable operation of machinery in small plots in Guangdong Province since 2018 are satisfactory.
With OPTIC clustering, divide the task by field. The accuracy of different types of operation can reach over 84%, which greatly reduces the accounting efficiency of traditional manual operations, and provides basis for government subsidy giving and operation charging;

(2) Use the algorithm of broken line buffer to calculate the area of the divided tasks. The calculation accuracy of the operation fields more than 3 mu(0.2 hectares) can reach over 95%, meeting the practical application requirements.

(3) Using the algorithm of filtering to filter the original trajectory points can effectively improve the calculation efficiency by over 30%. Removing the data drift caused by the positioning chip in static status effectively guarantee the robustness and stability of the algorithm.

Acknowledgements
This work is supported by National Key Research and Development Program of China (No. 2017YFD0700303)

References
[1] Li Hong, Yao Guangqiang, Chen Liping. (2008) Farm machinery monitoring and scheduling system based on GPS, GPRS and GIS. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 24(Supp.2): 119－122.(in Chinese with English abstract)
[2] Li Bin, Gong Linong, Yue Dansong. (2012) GPS for vehicle mounted system of agricultural machinery cross region production[J]. Journal of Agricultural Mechanization, 34: 220－222. (in Chinese with English abstract)
[3] Moacir Pedroso, James Taylor, Bruno Tisseyre. (2010) A segmentation algorithm for the delineation of agricultural management zones. Computer and Electronics in Agriculture, 70:199－208.
[4] Li Xiang, Pan Yuchun, Zhao Chunjiang. (2005)Delineation and scale effect of precision agriculture management zones using yield monitor data over four years. Scientia Agricultura Sinica, 38: 1825－1833. (in Chinese with English abstract)
[5] Li Yan, Shi Zhou, Wu Cifang. (2007)Classification of management zones for precision farming in saline soil based on multi data sources to characterize spatial variability of soil properties. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 23:84－89. (in Chinese with English abstract)
[6] Zhao Bin, Kuang Lihong, Zhang Wei. (2010) Development of precision seeder operating distance and area testing instruments. Journal of Agricultural Mechanization Research,6:17-20.(in Chinese with English abstract)
[7] Liu Hui,Meng Zhijun,Wang Pei. (2015)Buffer algorithms for operation area measurement base on global navigation satellite system trajectories of agricultural machinery. Transcations of the Chinese Society of Agricultural Engineering(Transactions of the CSAE),31:180-184.(in Chinese with English abstract)
[8] Han Yu. (2014)Study on the operation area measurement algorithm based on the trajectory of agricultural machinery.
[9] Zhu Dengsheng, Fang Hui, Hu Shaoming, Wang Wenquan, Zhou Yansuo, Wang Hongyan, Liu Fei, He Yong. (2020) Development and Application of an Intelligent Remote Management Platform for Agricultural Machinery. Smart Agriculture, 2: 67-81.(in Chinese with English abstract)
[10] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander. (1999)OPTICS: Ordering Points To Identify the Clustering Structure. In: ACM SIGMOD International Conference on Management of Data, Philadelphia, Pennsylvania, USA, 49-60.
[11] Borut Zalik, Mirko Zadra vec, Gordon J. Clapworthy. (2003) Construction of a non-symmetric geometric buffer from a set of line segments, Computers & Geosciences 29:53–63