AN IMPROVED PIXEL COUNTING METHOD FOR ARBITRARY ZONAL STATISTICS ON GLOBELAND30

C. Xu[1*], J. Chen[2], H. Wu[2], R. Li[2], Y.J. Zhao[3]

1 China University of Mining and Technology, School of Environment Science and Spatial Informatics, Xuzhou, xuchen6091
@126.com
2 National Geomatics Center of China, Beijing, China
3 Chinese University of Mining and Technology (Beijing), Beijing, 100083, China

Commission IV, ICWG IV/III, WG IV/4

ABSTRACT:

Global Land cover is essential for ecological study, climate and environmental changes, resources management, human activities as well as sustainable development. With the finer resolution products arising, e.g. GlobeLand30, quantitative analysis of the earth dynamics based on raster data sets is becoming possible. For arbitrary zonal statistics, pixel counting method is an operational practice widely used for area estimation by counting the number of classified pixels and multiplying by the area represented by each pixel, though uncertainty. For large statistical regions, the error of the projection can not be neglect. Thus an rectified PC method is developed for arbitrary zonal statistics of all raster data sets. The test result shows that the method performance is stable and easy to zonal statistics.

1. INTRODUCTION

Globe land cover data set is one of the most important layer for digital earth as id reflects the status and changes about the biophysical material over the surface of the Earth, such as forests, croplands, water bodies and built-up areas(Pielke et al. 2005; Running et al.,2008). During the past decades, a number of global land cover products at different spatial resolutions have been produced and widely used in scientific research activities such as ecological study, climate and environmental changes, resources management, human activities and sustainable development(Allen, et al, 2018).

Zonal statistic is a common approach to collect data in a given region, many researchers have used it to support urban management and planning, social economic development pattern, land ecology change analysis. The National Agricultural Statistics Services(NASS) of the US Department of Agriculture(USDA) provides timely, accurate, and useful statistical information service for agriculture applications in US through a data access and dissemination application system named "CropScape"(Boryan ,et al,2011; Han,et al., 2012;Han,et al., 2014). It can provide online data navigation, visualization, and analysis capabilities and services. However, "CropScape" is mainly for agriculture applications and provides data only in contiguous US. For global scale, with the finer resolution land cover products arising, e.g. GlobeLand30(Jun Chen et al., 2014), quantitative analysis of the earth dynamics based on raster data sets is becoming possible.

For arbitrary zonal statistics, pixel counting(PC) method is an operational practice widely used for rasterized assessment data of the earth, though uncertainty(Gallego, 2004). PC is used for area estimation by counting the number of classified pixels and multiplying by the area represented by each pixel. When a zonal statistic region is selected on Google Earth or some other GIS software, it is often marked by a polygon object which is

actually a linear string of points in geographical coordinates. Each point is described with latitude and longitude. The zonal statistic process can begin with projecting the polygon to the corresponding images, count the total number of pixels of each land cover category and calculate the final result of the area of each category. If the spatial reference of the target raster data set is not equal area projection, the area of the polygon will not equal to its true area and the statistical result will be biased.To eliminate the error of the projection, a rectified PC(RPC) method is developed for statistics in GlobeLand30 system.

2. METHOD

2.1 Preprocessing

Both the statistical region layer and the index layer are polygon shape layers. If any polygon has topology problem, there will be no results or lead to errors. The first step is to check the topology especially the self-intersection status. A simple method is used to check each polygon must be a closed linear ring with no self-intersection error that points must be arranged in a closed linear ring and the coordinate pairs sequence is in a clockwise or anti-clockwise direction.

2.2 Statistical region division and densely resampling

The GlobeLand30 data set is organized by map sheet in UTM projection strips. Each image covers 5°latitude × 6°longitude in mid-latitude region which dimension and the dimension is about 17000×19000pixels. The statistical region across more than one map sheet must be split into slices corresponding to referred images by overlay analysis between the map sheet index layer and the statistical region layer. The polygon marked the statistical region will be divided into several polygons and each subdivision polygon will be linked to one image by the index. When each subdivision polygon is projected to the image's projection, as finite points in the polygon's linear ring, the polygon's shape will deform. Therefore, the polygon will be

densely resampled to keep its shape in planer coordinate system. Get the mid-point from the start vertice to the end vertice in one of the polygon's edges. Loop the step until the polygon's area change meet the threshold between two steps.
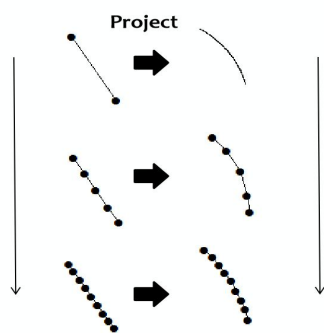

Figure 1. Densely resampling

2.3 RPC

Each subdivision polygon will be projected to the UTM conformal projection to count pixels of land cover categories by masking the image by the polygon. The area of each category will be calculated by multiplying the total number to a single pixel's area. As the projection distortion and the PC statistical error exists, the results deviates from the true area. The error and the relative error are defined as equation(1) and equation (2) respectively:

$$\Delta S = S_{PC} - S_E \qquad (1)$$

$$\Delta S_r = \left| \frac{S_{PC} - S_E}{S_E} \right| \times 100\% \qquad (2)$$

where ΔS=area error

ΔS_r=relative area error

S_{PC}=pixel count area

S_E=the true area calculated in equal area projection

If the relative error exceeds the setting threshold value, the error should be corrected. Here, an empirical method is used to rectify the error.For each category, the correction weight is distributed by its proportion. If n_i denotes the pixel number of the category, N denotes the whole pixel number the subdivision covers, the correction weight of the category is defined in equation(3):

$$w = \frac{n_i}{N} \qquad (3)$$

where w=correction weight

n_i=the pixel number of the category i

N=the whole pixel number the subdivision covers.

Then, the area of each category is calculated as equation(4):

$$S_i = S_{PC_i} - w\Delta S = S_{PC_i} + \frac{n_i}{N}\left(S_E - S_{PC}\right) \qquad (4)$$

where S_i=the area of category i

n_i=the pixel number of the category i

S_{PCi}=the pixel count area of category i.

The final result of the statistical region is calculated by summing each subdivision result.

3. TEST AND RESULT

To test the method above, a C# programme is written in the microsoft visual studio dot net environment using ESRI ArcGIS ArcObjects tool. The operating system is windows 7 64bit.

Software	version
Microsoft Visual Studio	2012
.Net Framework	4.0
ArcGIS ArcObjects	10.2
Operating System	Windows 7

Table 1. Details of the development environment

3.1 Study area

The study area is the boundary of Sweden obtained from globe infrastructural map database and its spatial reference is WGS 1984 geodetic system. As shown in Figure 2, seven maps sheets of GlobeLand30 data are needed to cover the whole Sweden test site. The corresponding index of map sheet is listed on the right side of the same figure.

Figure 2 The geographic location of test site in the accuracy test, demonstrating Sweden consist of seven map sheets. Light purple area outline the area of Sweden.

3.2 Result

Year	PC area	RPC area	Reference area
2000	447499.2897	449206.4455	449206.4456
2010	447505.6716	449206.4455	

Table 2. Results of the test(km)

Category	Year 2000	Year 2010
Cultivated land	39286.866926	39904.378332
Forest	251385.193476	253865.16099
Shrub land	49989.390671	52032.999389
Grass land	884.172646	511.732051
Wet land	27919.376901	29251.684497
Water body	37481.3441	36441.185915
Tundra	34745.098526	29072.595652
Barren land	1062.568218	1280.762209
Artificial Surfaces	5911.398631	5984.585511
Permanent snow and ice	541.035421	861.36097

Table 3. Categories Results of the test(km)

The results are presented in table 2 and table 3. The PC method result area and the RPC method area area calculated in the year of 2000 and 2010 respectively. In table 3, the area of each category is rectified result.The reference area is the true area of the statistical region in equal area projection Behrmann.

4. DISCUSSION

4.1 Accuracy analysis

In GlobeLand30 images, the single pixel area is 30×30, 900 square meters. As the PC method is just count the category number and multiply to single pixle area, so the PC result is times of 900. Considering the table presentation, the results are converted to kilometers unit. However, the difference still can be found between the PC result and the RPC result.

According to equation (2) , the realtive error of PC is 0.38%, the RPC method correct the projection error and keep result in consistency with the reference area. Both of the errors of the projection and the shape are calibrated. From the two year statistical results, the RPC performance is stable. The correction does not lead to large changes in the proportion of the categories.

4.2 Utility analysis

The software developed provides user-friendly interface, users do not need special knowledge about gis. The RPC method is well integrated, users can complete the statistics by selecting the input zone layer and the GlobeLand30 index layer. Compared with many GIS desktop software, the method is labor-saving for that users can avoid complex spatial analysis and work flow design.

The result of each category in the statistical region listed in table (3) can be used to further analysis about the land cover status. For example, with each category's area, the proportion of the category in the local can also be calculated. Comparison of the category proportion between different evaluation unit will be realized.

5. CONCLUSION

The developed RPC method actually correct the error caused by conformal projection. To test the method, experiments with regions in different scales ranging from county scale, province scale, country scale, continental scale to global scale are also considered. Different types of regions were also performed statistic: administration regions, climate regions, ecological regions, grid regions as well as arbitrary regions.

In the future, more test will be carried out and the method will be packaged to a web service for the statistical application of all kinds of data organized in raster format. The method can also be transformed to assist land cover classification and temporal data sets check. For a unique region, different land cover data sets created in different base time varies in the land cover categories and their areas. So the statistical result may help checking the change information between the time.

REFERENCES

Allen, C., G. Metternich and T. Wiedmann, Initial progress in implementing the Sustainable Development Goals (SDGs): a review of evidence from countries. Sustainability Science, 2018. 13(5): p. 1453-1467.

ArcObjects SDK 10 Microsoft .NET Framework-ArcGIS Resource Center. Available from:

http://help.arcgis.com/en/sdk/10.0/arcobjects_net/componenthelp/index.html (accessed 25 September, 2019)

ArcGIS Help-ArcGIS Desktop Statistical analysis. Available from:
http://desktop.arcgis.com/zh-cn/arcmap/latest/analyze/commonly-used-tools/statistical-analysis.htm (accessed 25 September, 2019)

Boryan, C., et al., Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto International, 2011. 26(5): p. 341-358

Chen, Jun, et al. "Global land cover mapping at 30m resolution: A POK-based operational approach." ISPRS Journal of Photogrammetry and Remote Sensing 103 (2015): 7-27.

Gallego, F.J., Remote sensing and land cover area estimation. International Journal of Remote Sensing, 2004. 25(15): p. 3019-3047.

Han, W., et al., CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Computers and Electronics in Agriculture, 2012. 84: p. 111-123.

Han, W., A Geospatial Web Service Approach for Creating On-Demand Cropland Data Layer Thematic Maps. Transactions of the ASABE, 2014. 57(1): p. 239-247.

Jun Chen, Yifang Ban, and Songnian Li. China: Open access to Earth land-cover map. Nature 514.7523 (2014): 434-434.

Pielke, Roger A. Land use and climate change. Science 310.5754 (2005): 1625-1626.

Running, Steven W. Ecosystem disturbance, carbon, and climate. Science321.5889 (2008): 652-653.