QUALITY INSPECTION AND COMMON ISSUES ANALYSIS OF 10 METER RESOLUTION GLOBAL LAND COVER DATA

Wen-jun XIE 1,*, Hai-peng CHEN 1, Lin-bing LV 1, Yu-heng CHEN 1, Miao LI 1

1 National Quality Inspection and Testing Center for Surveying and Mapping Products, Beijing, China - xwj_13@126.com

Commission III, ICWG III/IVb

KEY WORDS: Land Cover, GLC 10, Quality Inspection, Analysis of Quality Issues, Inspection Contents, Quality Improvement.

ABSTRACT:

In order to better carry out the environmental monitoring and resource protection, the 10 meter resolution global land cover data (hereinafter referred to as the GLC 10 data) came into being. The production mode of GLC 10 data is to use vector data, topographic map and other related reference data to get land cover based on digital orthophoto. GLC 10 data is a new type of remote sensing data and its classification system and classification index are also set according to the needs of a new project. Therefore, how to verify and control the quality of this kind of data is an urgent issue to be solved. According to the particularity of GLC 10 data and the new requirements of quality inspection technology, this paper puts forward a set of quality inspection contents and methods of GLC 10 data for large-scale production. And through the way of software automatic inspection combined with human-computer interaction, the inspection requirements are summarized one by one. Then, according to the actual quality inspection work from 2018 to 2020, the common quality issues of GLC 10 data are analyzed and sorted out, which can provide technical reference for the inspection and quality control of GLC 10 data.

1. INTRODUCTION

At present, the development of the world is facing a series of challenges, such as urbanization, agricultural expansion, over exploitation of resources, water source protection and so on. Therefore, the 10 meter resolution global land cover data (hereinafter referred to as the GLC 10 data) came into being, which can obtain more detailed information about the surface, so as to provide help for the maintenance of human health and the realization of the sustainable development goals of the United Nations. It uses digital orthophoto, combined with vector data, topographic map and other reference materials to get land cover data. Its classification system is composed of 10 first-level categories and 21 second-level categories. At present, most of the researches on quality control focus on the land cover classification data with 2 meter resolution and 30 meter resolution. For example, Luo Fujun analyzed the causes of the quality issues of the land cover data with 2 meter resolution and put forward some suggestions for improvement. Li Lei elaborated the inspection contents and methods of 30 meter land cover data, and listed some quality issues found in quality inspection, but there is no in-depth research on the quality control of GLC 10 data. Therefore, according to the particularity of GLC 10 data, combined with the actual technical requirements of quality inspection, this paper puts forward a set of technical methods and key points of quality inspection, analyzes the common quality issues found during the inspection, and gives three quality control measures, which can provide technical reference for the quality control of GLC 10 data.

2. QUALITY CHARACTERISTICS OF GLC 10 DATA

2.1 Production Technology of GLC 10 Data

The production process of GLC 10 data is mainly based on indoor data editing. Its specific process is as follows, using the remote sensing orthophoto with good current situation and suitable ground resolution to compare with the reference data in the mission area. That is, the 2 meter resolution ZY-3 satellite images from 2011 to 2018. The reference data generally including digital surface model data, vector data, topographic map, etc., so as to classify the land cover types. Finally, after passing the primary and secondary inspections, the final version of GLC 10 data that meets the technical requirements can be obtained. The production flow chart is shown in Figure 1.

![Figure 1. The production process of GLC 10 data.](image)

There are two ways of data production, namely, automatic computer classification and manual editing. In general production, different production methods are selected according...
to the length of the production period. When the production period is long and the quality of the image is good, the computer automatic classification combined with manual editing is generally adopted for production, and when the production period is tight and the image quality is poor, for example, there are many cloud shadows or the mountain shadow is serious on the image, in this case, the GLC10 data production is generally performed by manual editing, because the automatic interpretation of the computer is not good when the image quality is not so good, and there will be edge offset or extremely small patches in the classification results, then the amount of subsequent manual editing tasks will increase accordingly.

2.2 GLC 10 Data Content Overview

GLC 10 data is stored in Geodatabase format, and its attribute settings are shown in Table 1. GLC 10 is a set of 10 meter resolution surface land cover classification data, which adopts a classification system of 10 first-level categories and 21 second-level categories. The specific data classification system is shown in Table 2.

| Dataset name | The layer name of the data | Property item | Attribute item meaning |
|--------------|---------------------------|---------------|------------------------|
| GLC 10 (LC) | LCA (Surface data layer of shp format) | CC1 | First-level category code |
|              |                           | CC2 | Second-level category code |

Table 1. Definition of GLC 10 data attribute item name and content.

| First-level category | Second-level category |
|----------------------|-----------------------|
| CC1 code             | CC2 code              | Name        | Name                  |
| 010000               | 010100                | Cultivated land | Paddy field |
| 010200               | 010200                | Woodland      | Dry land |
| 020000               | 020100                | Shrubland     | Arbor |
| 030000               | 030000                | Grassland     | Other woodland |
| 040000               | 040000                | Artificial surface | Building area |
| 050000               | 050200                | Desert and bare surface | Industrial and mining land |
| 050300               | 050300                | Waters        | Airport |
| 050400               | 050400                | Glaciers and snow cover | Port |
| 050500               | 050500                | Wetland       | Reservoir dam |
| 050600               | 050600                | Tundra        | Traffic |
| 060000               | 060000                |                | Other artificial surface |
| 070000               | 070100                |                | Rivers and ditches |
| 070200               | 070200                |                | Lake |
| 070300               | 070300                |                | Reservoir |
| 070400               | 070400                |                | Pond |

Table 2. GLC 10 data classification system.

2.3 Quality Characteristics of GLC 10 Data

Through the quality analysis of GLC 10 data, we can find that there are many factors that affect its quality. In addition to the impact of the classification and editing methods used in production and the complexity of reference data, the poor quality of the original orthophoto and the difficulty in mastering the classification system and indicators have also brought some difficulties to the quality inspection work. The GLC 10 data has the following characteristics due to its production process.

2.3.1 Easy to cause topological errors: When data is edited, polygon vectors will be cut or merged, which is easy to cause topology errors such as extremely small polygon or overlap.

2.3.2 Image quality matters: The image interpretation is limited by the resolution, if the focus of data editing is not clear, it will lead to the situation of broken patches or missing classification of land cover types.

2.3.3 Small patches: For some small patches that do not meet the information extraction index, it is necessary to merge them into the land cover polygon with the same or similar characteristics when editing the data. However, if the classification attribute of a small patch is correct, even if the small patch does not meet the information extraction index, it can be retained without merging.

2.3.4 Foreign objects of the same spectrum exist: In tropical regions, it may be difficult to distinguish between gardens and other vegetation inclusions. During the production process, comprehensive selection will be made according to the main land types in the area, so there will also be foreign objects of the same spectrum, and the inspectors need to comprehensively analyze whether it is reasonable.

2.4 GLC 10 Data Quality Control Process

According to the characteristics of GLC 10 data and the mode of its production, we use four levels of inspection process to control the data quality, which are the first level inspection, the second level inspection, the acceptance inspection and the quality verification. The specific process is shown in Figure 2.
2.4.1 Inspection duration and sampling proportion: Generally, the first level inspection and the second level inspection take a long time, because the inspection work of these two levels is to check all the GLC 10 data, while the follow-up acceptance inspection and quality verification generally take a short time, which is one fourth of the time required for the two level inspection. According to the quality control requirements, the amount of data for the four levels of inspection is also different. First, all the GLC 10 data of the production need to be inspected at the first level and the second level. The number of samples checked in these two inspections are all account for 100% of the total production. The amount of sample data checked in the acceptance inspection accounted for 10% of the total production, and the amount of sample data checked in the quality verification accounted for 25% of the total acceptance inspection.

2.4.2 Responsibility assignment: Moreover, the departments that carry out the four levels of inspection are independent and unrelated to each other. Specifically, the first level and second level inspection are carried out in the production department, and the inspection adopts the way of self inspection and personnel cross inspection. The acceptance inspection is undertaken by an independent third-party quality inspection agency, which is generally completed by the provincial quality inspection station. The final quality verification work is completed by the national quality inspection center. There are also multi-level control measures for the accuracy of the mistakes found by the quality inspection personnel. At the end of each level of inspection, first of all, a reviewer will review all the quality issues to ensure that there is no wrong mention of the quality issues in the inspection records; secondly, the reviewed inspection records will be fed back to the production department, and the technical director and quality director of the department will confirm the quality issues again, and finally the inspection records will be handed over to the operators, and they will modify the errors in the data according to the confirmed inspection records. Therefore, this strict and full coverage quality control system can ensure that GLC 10 data quality can be effectively controlled in all aspects.

3. CONTENTS AND REQUIREMENTS OF GLC 10 DATA QUALITY INSPECTION

3.1 Quality Elements And Inspection Items

The inspection work adopts manual inspection, human-computer interaction, program automatic inspection and other methods to check the contents of the GLC 10 data. The specific inspection quality elements and inspection items are as follows.

3.1.1 The Spatial Reference System: This item focuses on checking whether the coordinates of the map frame, the elevation datum, and the map projection are correct.

3.1.2 Time Accuracy: This item focuses on checking the time accuracy of the original data and final version of the GLC 10 data, to ensure that all the data is up-to-date and meet the requirements of technical design.

3.1.3 The Logical Consistency: This item focuses on checking whether the file name, file storage organization, and data format meet the requirements, whether the data is missing or redundant, and whether the data can be read normally.

3.1.4 Acquisition Accuracy: This item focuses on checking whether the accuracy of the combination of the patch boundary and orthophoto is out of limits, and whether the geometric position of adjacent patches is out of limits.

3.1.5 Classification Accuracy: This item is to check the correctness of land cover classification, whether the patches with secondary class are divided into second-level categories according to the technical design requirements, whether the filling of classification code is standard, whether the patches attributes between two adjacent images are continuous, and whether there are missing or redundant patches.

3.1.6 Characterization Quality: This item is to check whether there are abnormal errors in the geometry of the patches, such as whether there are unreasonable extremely small polygons, or whether there are unreasonable hard folds at the patches boundary.

3.1.7 Quality of The Attachment: This item checks whether the data organization, file naming, format, number of files and file storage order of metadata meet the requirements of technical design, whether the attribute items filled in have errors or omissions, and also checks the completeness and regularity of the related documents.

3.2 Contents And Requirements of Inspection

3.2.1 Contents of Automatic Inspection Using Software: The automatic software inspection combined with manual elimination of false alarms can be used to check the spatial reference system and logical consistency.

Firstly, check the correctness of the spatial reference system, that is, whether the coordinate system is set to WGS-84 coordinate system, whether the unit of the coordinate is degree, and at least 6 digits should be reserved after the decimal point. Then check the logical consistency, that is, whether the file name, file storage organization and data format meet the requirements, whether the data is missing or redundant, and whether the data can be opened and read normally; the data topology tolerance value should be less than 0.5 meter, and the data should ensure that there are no gaps, overlaps and other topology errors.

3.2.2 Contents of Human Computer Interaction Inspection: The human-computer interaction inspection is carried out by using ArcGIS, which checks the time accuracy, acquisition accuracy, classification accuracy, characterization quality and quality of the attachment.

For the inspection of time accuracy, that is to check whether the image data, basic geographic information data, industry thematic data and other data sources are the latest data; check whether the timeliness of the results data meets the technical design requirements. That is to check the image overlap area, whether the image of the corresponding region in the vegetation growing season and the image with good visual quality are preferred. If the latest time requirement cannot be met, we should check whether this special situation has been recorded into the technical summary.

For the inspection of geometric displacement in the acquisition accuracy, that is, to check whether the accuracy of the patch boundary and the orthoimage fit together meets the technical regulations, that is, the maximum position deviation between
the land cover boundary on the image and the boundary of the patches cannot exceed 20 meters.

For the edge joining inspection of vector polygon, that is to check whether the patch after edge joining is smooth and continuous, and avoid hard fold and sharp corner, and the adjacent patches with the same attribute value should be merged into one entity after edge joining.

For the inspection of classification accuracy is to check the consistency of the GLC 10 data with orthoimage, vector data, terrain data and other reference data. For example, vector data can be used to compare and check road and railway patches, and terrain data can be used to judge vegetation and crops. In the case where the second-level category cannot be distinguished due to poor image quality, at least the first-level category of the patch must be ensured to be correct.

For the inspection of characterization quality is to check topology error of patches, such as whether there are overlaps between two polygons, whether there are rigid corners on the edge of the polygons, and whether there are gaps.

For the inspection of quality of the attachment, that is, check the integrity of the metadata, whether it is composed of three parts according to the technical design requirements, and whether the metadata naming is filled in according to the name of the task area, and whether one data file corresponds to only one metadata.

4. COMMON QUALITY ISSUES AND CAUSE ANALYSIS

Based on the actual inspection of GLC 10 data in different mission areas from 2018 to 2020, this paper summarizes some common quality issues, as shown in Figure 3 to Figure 10. It can be found that the quality issues mainly focus on two types of quality elements, that is, the classification accuracy and acquisition accuracy. The main quality issues can be summarized as the following four situations. First, the accuracy of edge line of land cover classification patch is beyond the limit. The second is that some land cover patches which have reached the acquisition index on the image have been missing drawing. The third is that there are some unreasonable patches with foreign objects of the same spectrum. Fourth, there are some land cover patches are classified into wrong categories.

4.1 Acquisition Accuracy Exceeds The Maximum Limit

Due to the influence of human factors or software automatic classification, it is easy to cause the acquisition accuracy exceeds the maximum limit, as shown in Figure 3 (a) (b) (c) (d) (e) (f) below, which are the cases that the patches of dry farm, paddy field, grassland, river and road is not editing according to the image. The yellow line or arrow mark is the mark made by the inspector on the correct range of the patch.

(a) The deviation between the boundary of dry farm patch and the orthoimage is more than 20 meters.

(b) The deviation between the boundary of grassland patch (CC1 code is 040000) and the orthoimage is more than 20 meters.

(c) The boundary of the river (CC2 code is 070100) is not edited according to the image.

(d) There is a local deviation between the road (CC2 code is 050600) sideline and the image.

(e) There is an overall offset between the road (CC2 code is 050600) sideline and the image.

(f) The scope of building area patch (CC2 code is 050100) is too large.
actual shape of the residential area. During data production, some roads in the existing high-precision or same precision vector data will be directly used in GLC 10 data, but in fact, these roads may have changed, which will lead to the mismatch between the road patch and the image, as shown in Figure 3 (d) (e). In view of this situation, we should refer to the actual image to modify the road boundary.

4.2 Land Cover Patches Are Missing Drawn into GLC 10 Data

This type of error is easy to occur in large-scale data production. As shown in Figure 4 (a) (b) (c), the woodland, water and buildings with enough areas index are not drawn into the GLC 10 data. The yellow line mark is the mark made by the inspector for the missing patches.

(a) Some woodland (CC2 code is 020200) are missing drawn into the GLC 10 data.

(b) One pond (CC2 code is 070400) is missing drawn into the GLC 10 data.

(c) The boundary of woodland (CC2 code is 020100) was drawn too large, and houses (CC2 code is 050100) with enough area inside were forgotten to be drawn.

Figure 4. Land cover patches are missing drawn into the GLC 10 data.

Cause analysis: According to the requirements of the technical regulations, trees arranged in rows on both sides of the road, along the river and beside the houses do not need to be classified separately even if they reach the acquisition index. However, Figure 4 (a) does not show this situation. It shows a large area of grassland with some woodland in it which area is more than 25000 square meters, so these woodland should be drawn separately. The error in Figure 4 (b) is that there is an obvious pond with enough area, which is not drawn into the data due to carelessness. Figure 4 (c) shows some buildings in the woodland, which have reached the acquisition index but not drawn into the data. This kind of situation is generally caused by the wrong judgment or carelessness of the image during the data editing.

4.3 Unreasonable Patches with Foreign Objects of The Same Spectrum

The same spectral foreign matter refers to the situation that the land cover with the same texture on the image but the category attributes of the patches in the data are assigned different values. This kind of issue often occurs in the easily mixed categories such as cultivated land, woodland and grassland. As shown in Figure 5, there are some issues of different land classification with the same spectral features. The yellow line mark is the inspector's mark for unreasonable patches with foreign objects of the same spectrum.

(a) Woodland (CC2 code is 020100) and grassland (CC1 code is 040000).

(b) Dry farm (CC2 code is 010200) and grassland (CC1 code is 040000).

(c) Reservoirs (CC2 code is 070300) and lakes (CC2 code is 070200).

(d) Grassland (CC1 code is 040000) and paddy field (CC2 code is 010100).
Grassland (CC1 code is 040000) and industrial and mining land (CC2 code is 050200).

**Figure 5.** Patches with the same texture on the image but are classified into two different land cover categories.

**Cause analysis:** Some patches have the same texture and color features from the image analysis, and their area also reach the acquisition index, but they are classified into different categories in the results. This is because the editing scale of different production operators is not unified, and the definition and classification requirements of land cover types are not well understood by them. Most of the time, this kind of issues are caused by carelessness in the process of data editing.

### 4.4 Inaccurate Classification

Due to the limitation of 10 meter image discrimination and the great difference of ecological environment in different regions, it is easy to cause inaccurate classification of patches. The specific quality issues are shown in Figures 6 to Figure 10. The yellow line or highlighted area labeling is the mark made by the inspector on the incorrectly classified patches.

**Figure 6.** The image does not support classification as grassland.

(a) Cultivated land (CC1 code is 010000) is wrongly classified as grassland (CC1 code is 040000).

(b) Woodland (CC1 code is 020000) is wrongly classified as grassland (CC1 code is 040000).

(c) Industrial and mining land (CC2 code is 050200) is misclassified as grassland (CC1 code is 040000).

**Figure 7.** The lake patch (CC2 code is 070200) is not edited according to the water level on the image.

**Figure 8.** A waterless area is misclassified as a pond (CC2 code is 070400).

**Figure 9.** Shrubs (CC2 code is 020200) are misclassified as arbors (CC2 code is 020100).

(a) Industrial and mining land (CC2 code is 050200) is misclassified as buildings (CC2 code is 050100).

(b) River (CC2 code is 070100) is misclassified as a pond (CC2 code is 070400).

**Figure 10.** Classification error of second-level categories.

**Cause analysis:**

As for Figure 6 (a), according to the image, it can be judged that it should be cultivated land, but the data is wrongly classified as grassland. The area of this wrong patch is large, which has a great impact on the evaluation of the quality of the results. The reason for the error of the attribute of the large patch is due to the wrong operation when editing the data. Generally, when
modifying the data, some small patches around will be merged into a large patch nearby, which may lead to the wrong assignment of the attribute of the small patch to the combined large patch. In addition, inaccurate understanding of technical indicators also caused classification errors. For example, the highlighted patch in Figure 6(b) is woodland, which is unreasonably classified as grassland. Figure 6(c) shows that the patch with obvious traces of manual excavation are misclassified as grassland.

The water in the pond shown in Figure 7 does not reach the maximum water level due to seasonal changes, but the water boundary is wrongly drawn to the maximum water level. According to the requirements of technical regulations, the correct drawing way should be editing the water boundary according to the actual water level of the image, so the bare surface should be classified separately. The quality issue shown in Figure 8 is similar to that in Figure 7. It shows a dried up pond, so it is more reasonable to classify it as grassland.

The highlighted patch in Figure 9 shows the shrubs are misclassified as arbors. According to the image texture, it can be seen that it is a relatively low and sparse vegetation, which obviously does not meet the definition of arbors.

There are also classification errors caused by the production personnel's insufficient ability to interpret the land cover categories. As shown in Figure 10 (a), the open-air industrial facilities and factories should be classified as industrial and mining land, but they are wrongly classified as building areas where people live. Figure 10 (b) shows that the river is misclassified as ponds.

5. QUALITY CONTROL MEASURES

In view of the common quality issues mentioned in the fourth section, we can take the following quality control measures to control the final GLC 10 data quality.

5.1 Carry Out Trainings

In view of the production personnel's inconsistent mastery of technical indicators, we can carry out a series of technical training in the early stage of production. By explaining the technical indicators one by one, and taking the way of examination, to check the production personnel's mastery of the training content, the personnel who fail to pass the examination are not allowed to participate in the data production and quality inspection work, so as to eliminate the influence of artificial factors on the data quality as far as possible.

5.2 Optimize Production Process

In view of the data error caused by manual misoperation, we can consider appropriately increasing the proportion of software automatic classification in the production process. At present, the production process of GLC 10 data is still carried out with 80% to 90% manual editing and 10% to 20% computer automatic classification. The main reason for this arrangement is that the existing automatic classification software has higher requirements for image quality, and if the image quality is poor, the results of automatic classification will appear a series of problems such as boundary mismatch or small patches, and a lot of manual post editing is needed, which is not suitable for large-scale data production. But if the software classification accuracy can be improved, it can effectively improve the production efficiency, so we can study to improve the software automatic classification accuracy, increase the degree of automation in the production process, so as to reduce the impact of manual misoperation.

5.3 The Principle of Screening And Using Original Data Should Be More Clear

As the main data source of GLC 10 production, the spatial and temporal accuracy of images largely determine the quality of data production. There are also other relevant reference materials, which have complex sources and different time. It is necessary to carefully sort out and analyze the useful information, which need to be screened by a special technical director in the early stage of data production and clarify the use principles of all materials.

6. CONCLUSION

According to the technical requirements of quality inspection, this paper deeply analyzes the quality characteristics of GLC 10 data, and puts forward a set of targeted quality control measures. At the same time, this paper summarizes and lists some common quality issues, analyzes the causes and gives three quality control measures, which can provide some technical reference for the quality control of GLC 10 data.

REFERENCES

Maria, A.B., Monia, E.M., Eman, H., Jun, C., Ran, L., 2015: The First Comprehensive Accuracy Assessment of GlobeLand30 at a National Level: Methodology and Results. Remote Sens., 2015(7), 4191-4212.

James, D.W., Stephen, V.S., Leila, G., Jon, D., Joyce, A.F., Timothy, G.W., 2013: Accuracy assessment of NLCD 2006 land cover and impervious surface. Remote Sensing of Environment, 130 (2013), 294-304.

Chen, J., Liu, W.Z., Wu, H., et al. 2019: Basic Issues and Research Agenda of Geospatial Knowledge Service. Geomatics and Information Science of Wuhan University44(1), 38-47.

Gong, P., Zhang, W., Yu, L., Li, C.C., Wang, J., Liang, L., Li, X.C., Ji, L.Y., Bai, Y.Q. 2016: New research paradigm for global land cover mapping. Journal of Remote Sensing, 20(5), 1002-1016.
Xu, G., Wang, L.L., Wang, X., 2018: Discussions of Geographical Conditions Monitoring Data Quality Inspection. Geomatics & Spatial Information Technology, 41(10), 165-167.

Chen, Z.P., Zhao, Y., Zhang L., 2017: Research on the Data Quality Inspection of the Land Cover Classification for Geographic National Conditions Investigation. Geomatics & Spatial Information Technology, 40(8), 150-152.

Hu, J.J., Chen, J., Peng, S., 2018: Hot-spot Analysis of GlobeLand30 Data Studying. Remote Sensing Information, 33(4), 1-7.

Dang, Y., Zhang, J.X., Deng, K.Z., et al. 2017: Study on the Evaluation of Land Cover Classification using Remote Sensing Images Based on AlexNet. Journal of Geo-information Science, 19(11), 1530-1537.

Ma, J.Z., Sun, Q., Xiao, Q., et al. 2016: Accuracy Assessment and Comparative Analysis of GlobeLand30 Dataset in Henan Province. Journal of Geo-information Science, 18(11), 1563-1572.

Li, R., Kuang, W.H., Chen, J., et al. 2016: Spatio Temporal Pattern Analysis of Artificial Surface Use Efficiency Based on GLobeLand30. Science China: Earth Sciences, 46(11), 1436-1445.

Luo, F.J., Li, C., Yan, Y.H., 2016: Reflections on the Check of Land Cover Classification Data. Geomatics & Spatial Information Technology, 39(4), 156-161.

Li, L., Wang, C., Yin, T., 2020: Quality Inspection Method and Techniques for Global Land Cover Data. Geomatics Technology and Equipment. 22(4), 40-46.