

RESEARCH ON QUALITY IMPROVEMENT OF GLOBELAND30 UPDATE DATA

Wen-jun XIE^{1,*}, Li ZHANG¹, Hai-peng CHEN¹

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: GlobeLand30 Update Data, Land Cover, Quality Inspection, Quality Improvement

ABSTRACT:

GlobeLand30 update data is one of the important products of the Construction Maintenance and Update of Geographic Information Resources Project (hereinafter referred to as the Update Project). It provides important basic information for national geographical conditions monitoring, eco-environmental assessment and global geographic information integrated services. Different from the land cover classification data in National Geoinformation Surveying and Monitoring Project, GlobeLand30 update data is a completely new type of scientific research outcomes with new setup of resolution, its data format and data structure are newly set according to the requirements of the Update Project. Therefore, current inspecting methods are of limitations and incompatibility and cannot be firmly reasonable for inspecting the accuracy of GlobeLand30 update data. Combining with the practice of quality control for the Update Project, this paper proposes a set of methods and processes for the quality inspection of the GlobeLand30 update data. It also summarizes the key points of its inspection and analyses the common errors found in the actual inspection practice of the Update Project from 2017 to 2018. It can provide a certain technical reference for the quality control and quality improvement of GlobeLand30 update data in the Update Project.

1. INTRODUCTION

The Construction Maintenance and Update of Geographic Information Resources Project (hereinafter referred to as the Update Project) is of great significance for building a global geographic information construction technology system and promoting the construction of the Belt and Road as a comprehensive service for global geographic information. The GlobeLand30 update data is one of the six important achievements of the Update Project, and also a new scientific research product. The production of GlobeLand30 update data is mainly based on Landsat-8 satellite remote sensing images, and the 2010 version of GlobeLand30 is used as the data update basis. It is based on the optimum selection and processing of Landsat-8 satellite imagery for full global coverage, service-oriented integration of all available reference data and auxiliary information, object-based precise land cover characterization, and knowledge-based data quality controlling, so as to complete the production of ten categories of global surface coverage information including arable land, forest, wetland, etc. The final achievement is the GlobeLand30 update data of 2015 version. In order to ensure that the quality of the achievements meets the design requirements of the Update Project, in the process of project implementation, we have established a quality control system, that is, two-level inspection, first-level acceptance, and post-inspection review in accordance with the principles of the whole process, hierarchical, and quality management principles. Strict quality inspection of data is also carried out during all stages of the project, such as acceptance, review and database construction.

At present, some scholars have carried out relevant research on the inspection of GlobeLand30. For example, Ji-yu LIU (2015) built a knowledge-based data quality inspection rule to check the quality of farmland data of GlobeLand30. Xu CHEN (2016) used the inspection rules based on knowledge to check the artificial cover data of GlobeLand30. Fu-jun LUO (2016)

summarized the quality issues of the land cover classification data in the National Geoinformation Surveying and Monitoring Project and put forward the principles of dealing with the related issues. Generally speaking, these existing methods are mostly for the extraction or inspection of a single class in the land cover classification data, which has certain limitations and does not form a systematic quality inspection process or method. Moreover, compared with the land cover classification data in National Geoinformation Surveying and Monitoring Project, the production process and quality of the GlobeLand30 update data are more special. In addition to the difficulty brought by the regional differences on the global scale, the contents and methods of the quality inspection of the GlobeLand30 update achievements also have special requirements. Therefore, in view of the particularity of the GlobeLand30 update data, and combined with the actual quality inspection technical requirements of the Update Project, this paper puts forward a set of targeted technical process, methods and main points of quality inspection, and also sorts out and analyzes the common quality issues found in the actual inspection work, and puts forward the targeted solutions.

2. QUALITY ANALYSIS OF GLOBELAND30 UPDATE DATA

Different from the production mode of land cover classification data in National Geoinformation Surveying and Monitoring Project, that is, most of the work is mainly based on indoor data editing, supplemented by field verification, the production of GlobeLand30 update data is almost impossible to carry out field verification. Therefore, its data production mode is mainly based on images and all kinds of thematic materials for reference, and is generated by updating and editing the data of the previous period. The specific production process is shown in Figure 1.

* Corresponding author

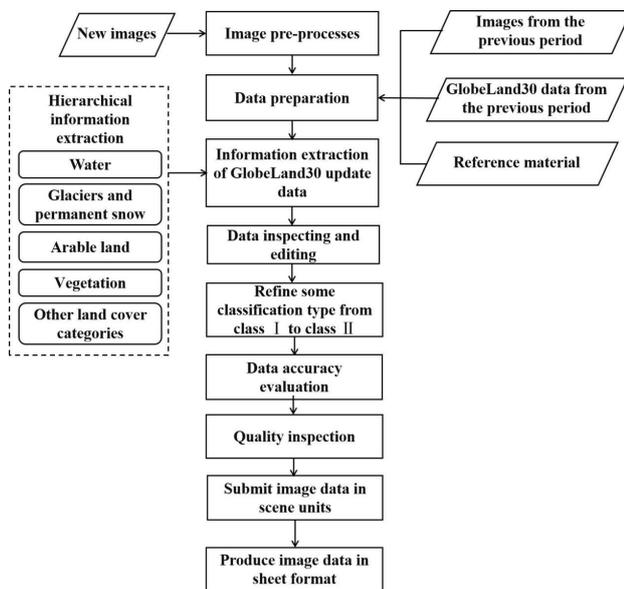


Figure 1. Production process of GlobeLand30 update data

From the production process in Figure 1, it can be seen that the data processing process of GlobeLand30 update data is complex, and the classification system designed for this project is more professional, with 10 categories of the first-class (class I) and three of the categories from the first-class are subdivided to secondary category (class II). Through the quality analysis of the GlobeLand30 update data, we found that there are many factors that affect the quality of the achievements. For example, first, its data production uses classification extraction and update methods, which is not only complex in technology, but also complex in process. Second, it is difficult to master the classification indicators and requirements of the Update Project. Third, the quality of the original image is poor, and sometime the phase consistency between the updated image and the original image is also poor. Fourth, the achievements are divided into several batches for production and submission, which is easy to cause a large number of documents and attachments. Each of the above factors will bring certain challenges to our quality inspection.

3. CONTENTS, METHODS AND PROCESSES OF INSPECTION

Considering the various factors that affect the quality of GlobeLand30 update data, this paper proposes a set of targeted quality inspection technical processes and methods, and comprehensively sorts out the contents and main points of the quality inspection of achievements. According to the project inspection standards, the main inspection contents of the GlobeLand30 update data should include: spatial reference system, time accuracy, position accuracy, logical consistency, classification accuracy, quality of raster data, and quality of the attachments. The main inspection methods can be software automatic inspection combined with human-computer interaction and manual inspection.

3.1 Software Automatic Inspection

The quality elements such as spatial reference system, logical consistency and raster data can be checked automatically by software and manually eliminate the errors of computer misjudgment.

3.1.1 The Correctness of Spatial Reference System: This item needs to check whether the geodetic datum is set to WGS-84 coordinate system, whether the projection method is UTM projection, whether the banding method is 6 degree zoning, and whether the coordinate unit is meter.

3.1.2 The Logical Consistency: This item needs to check whether the data file name, data file storage organization, data format meet the design requirements of the project, whether the data is missing or redundant, and whether the data can be read normally.

3.1.3 Quality of raster data: This item needs to check whether the achievements files use lossless GeoTIFF compression format and 8-bit 256 color index mode, whether the assignment of class I is an integer value, and whether the assignment of class II is a value corresponding to the integer value of class I, whether the attribute of the no-data area is represented by the number 0, and whether the sea area is represented by the number 255.

3.2 Human-computer Interaction and Manual Inspection

Generally, ArcGIS is used in human-computer interaction and manual inspection to complete the inspection of quality elements such as time accuracy, image data source, position accuracy, classification accuracy, and quality of the attachments.

3.2.1 Time accuracy:

3.2.2 Quality of image data source: As for the plane accuracy, it is necessary to check whether the mean square error of absolute positioning of ground feature points is better than 100 meters, and this index can be relaxed by 0.5 times in particularly difficult areas such as high mountains. As for the registration accuracy, it is necessary to check whether the mean square error of positioning of ground feature points is less than 60 meters, and this index can be relaxed by 0.5 times in particularly difficult areas such as high mountains. As for image quality, it is necessary to check whether the multispectral image has complete bands with no clouds or only a few clouds, and the images are clear, no noise, and no data is missing. When poor quality images are used in some areas due to the difficulty of image acquisition, it is necessary to check whether the current quality issues of the image and its usage are recorded in the metadata in detail.

3.2.3 Position Accuracy: This item needs to check whether the non overlapping area of the achievements meets the design requirements of the project. Without the influence of image factors, for the linear land cover data, the error of edge joint should be less than or equal to one pixel. For the patchy land cover data, the edge should transition naturally, without hard edge, and should keep the coordination between adjacent land cover types.

3.2.4 Classification Accuracy: The overall accuracy verification mainly uses methods such as checking the production materials and cross checking the reference materials. The random sampling method based on landscape index is used to obtain sample points, and then the confusion matrix is calculated according to the sample points to get the overall accuracy evaluation results. The inspection of classification accuracy also needs to extract misclassified area, which refers to the comparison and inspection between the achievements and the images used in production. It is necessary to collect the reference data of multi-source geographic information in the mission area, as well as the field photos, high-resolution images, humanities, phenology and other comprehensive data of the relevant land cover types, so as to form the multi-source knowledge of the specific land cover types in the mission area, and use the method of multi-source knowledge comprehensive check to sketch and calculate the area of the misclassified area.

3.2.5 Quality of Attachments: This item needs to check whether the contents filled in metadata meet the requirements of the project, and whether other documents submitted are complete, reasonable and correct. If there are special circumstances in the images or final achievements, it is also necessary to check whether these situations have been recorded in the technical summary.

3.3 Quality Inspection Process

According to the contents and key points of inspection in Section 3.2, and combined with the actual situation of GlobeLand30 update data in the Update Project, this paper summarizes a set of quality inspection process, which is mainly divided into four stages, sampling, inspection, review and assessment, as shown in Figure 2.

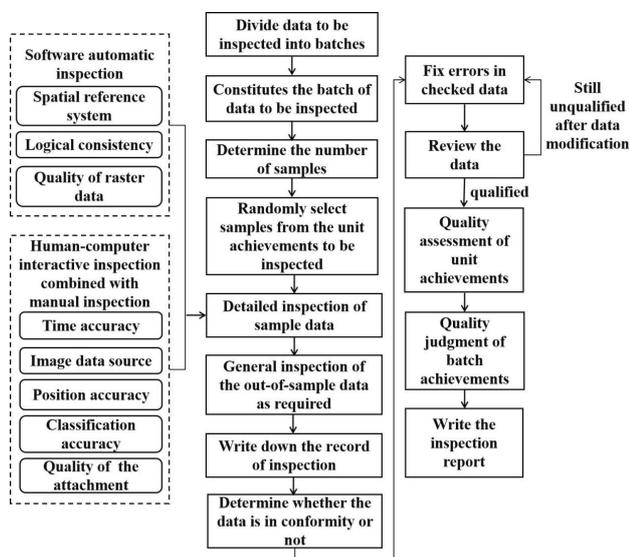


Figure 2. Inspection process of GlobeLand30 update data

3.3.1 Sampling: When the actual inspection work is carried out, first of all, the achievements are divided into batches to form a batch of data to be inspected. The batch of achievements shall be the unit achievements of the same level and specification produced under the guidance of the same technical design document. At the same time when sampling, comprehensive consideration should be given to the topography and landform categories, human factors, landscape categories and other conditions in the inspected mission areas.

3.3.2 Inspection: The specific inspection can be carried out in the way of sampling for detailed inspection combined with general inspection for out-of-sample data. First of all, check whether each quality element of the unit achievements in the sample meets the design requirements. The importance and tendency errors of the sample data can be summarized through the way of software automatic inspection combined with human-computer interaction and manual inspection, then record these errors in the inspection record tables, and judge whether the unit achievement is qualified. Then, according to the needs of some important inspection items or elements, we may carry out a general inspection for out-of-sample data and fill in the inspection record form. For batch of achievements that are in conformity after detailed sampling inspection and general inspection, the data manufacturer is required to modify all the errors in the data according to the inspection record.

3.3.3 Review: The modified achievements need to be rechecked and the unqualified achievements shall continue to be returned for modification until they are qualified.

3.3.4 Assessment: Finally, the quality judgment of the batch achievements that are qualified in the review should be assessed and the corresponding inspection report should be prepared.

4. THE UPDATE PROJECT INSPECTION IN 2017-2018

We have successfully used the inspection method proposed in this paper to check the GlobeLand30 update data of the Update Project from 2017 to 2018, which can effectively find the quality issues existing in the achievements.

4.1 Summary of Common Quality Issues

Through the practical inspection of the Update Project from 2017 to 2018, this paper summarizes some common quality issues as shown in Figure 3 to figure 9. These errors found in the achievements are mainly concentrated on the quality element of classification accuracy, which is specifically manifested in the quality issues such as land cover data are been wrong updated or missing updated, the same spectrum foreign matters and the land cover categories is discontinuous between two adjacent images.

4.1.1 Some Land Cover Data Are Classified Incorrectly:

Because of the limitations of the 30 meter resolution image used in the artificial recognition of land cover categories, and the complexity of the geomorphology in different regions of the world, which is easy to cause incorrect classification of land cover. For example, the following pictures are all cases of incorrect updating of land cover data. Figure 3 shows an open pit is misclassified as wetlands and lakes. Figure 4 shows some cloud shadows are misclassified as lakes. Figure 5 shows lakes are misclassified as rivers.

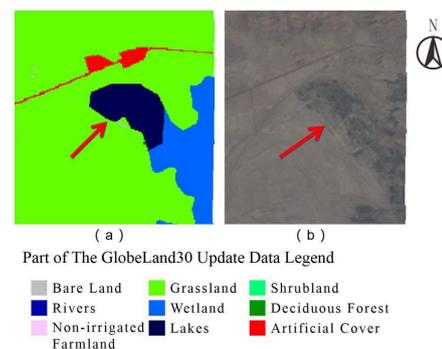


Figure 3. Open pit is misclassified as wetlands and lakes (a) GlobeLand30 update data (b) The image used for data updating

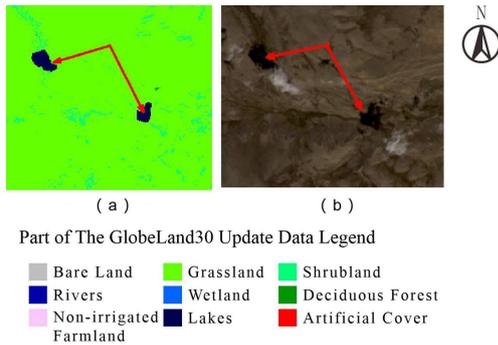


Figure 4. Cloud shadows are misclassified as lakes (a) GlobeLand30 update data (b) The image used for data updating

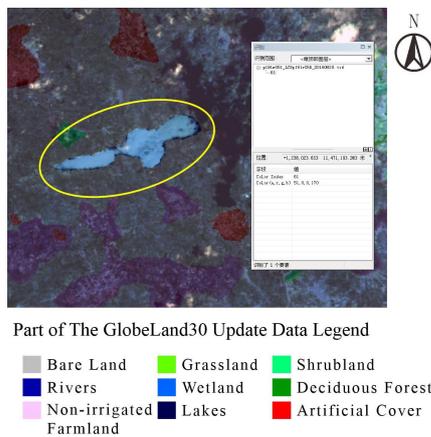


Figure 5. Lakes are misclassified as rivers (The image used for data updating is superimposed together with GlobeLand30 update data)

4.1.2 Some Land Cover Data Are Missing Updated: It is easy to cause this kind of issues when editing in a large range of manual work. As shown in Figure 6, some artificial covers are missing updated. Figure 7 shows some rivers are missing updated.

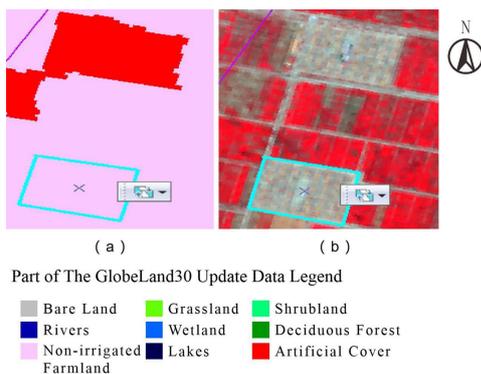


Figure 6. Artificial covers are missing updated (a) GlobeLand30 update data (b) The image used for data updating

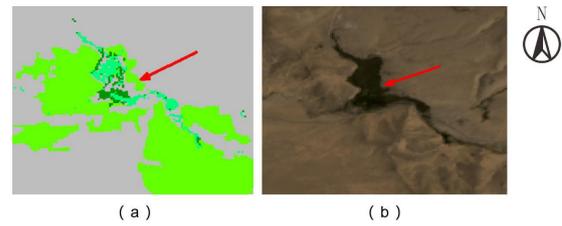


Figure 7. Rivers are missing updated (a) GlobeLand30 update data (b) The image used for data updating

4.1.3 The Same Spectrum Foreign Matters: The same spectrum foreign matters refer to the situation where the regions are covered with same textures on the image, but its category attributes are assigned different values. Such issues often occur in categories that are difficult to distinguish on the image, such as farmland, woodland, and grassland. Figure 8 shows regions with the same texture on the image are classified into two different classes.

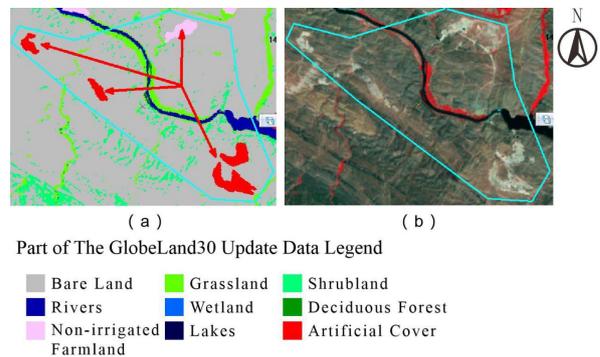


Figure 8. Regions with the same texture on the image are classified into two different classes (a) GlobeLand30 update data (b) The image used for data updating

4.1.4 The Land Cover Categories Is Discontinuous between Two Adjacent Images: Generally, in the process of actual data production, the production range will be divided into several smaller areas for data update, and then the land cover categories between adjacent images will be merged to maintain the coordination and consistency of categories. As shown in Figure 9, the land classification is discontinuous between two adjacent images.

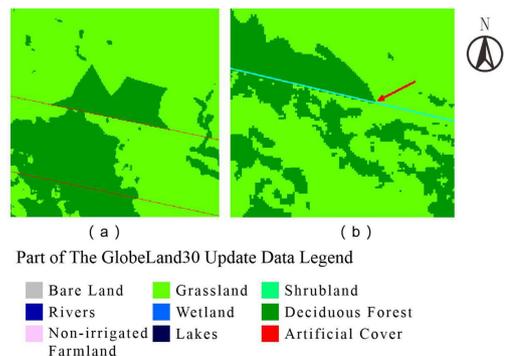


Figure 9. The land classification is discontinuous between two adjacent images (a) Two adjacent images with GlobeLand30 update data (b) Two adjacent images with GlobeLand30 update data

4.2 Analysis on The Causes of Quality Issues

The first reason is that there is a deviation in personnel's understanding of the production technology standards of this project. The Update Project is a new type of scientific research project. Its related technical requirements are different from the production standards of traditional basic surveying and mapping projects. Therefore, in the process of project development, the production personnel do not have the same understanding of relevant technical regulations, which will cause improper processing of GlobeLand30 update data, and ultimately affect the quality of the achievements.

The second reason is that the production software used in this project needs to be improved. Due to the tight time, heavy task and heavy workload of the Update Project, software automatic classification is often used to extract some updated land categories, such as newly added waters, roads and other categories. If the classification software can not identify a certain land category well, the extraction results will affect the subsequent process.

The third reason is that the inspection focus of quality inspectors is not comprehensive and accurate. At present, the inspection of this project mainly relies on manual inspection, which will lead to omissions caused by human factors. For example, inspectors may tend to pay too much attention to the mixture of forest and grassland during inspection, and fail to achieve sufficient and comprehensive quality inspection, will result in some quality issues in the final achievements.

4.3 Solutions

Based on the analysis of the causes of the quality issues and the current production mode of the Update Project, three suggestions are put forward for reference.

First, during the data production process, the quality issues caused by the software should be continuously and timely summarized, and the software should be updated and improved as soon as possible to improve production efficiency.

Second, before the start of large-scale production every year, all production personnel shall be called together to conduct unified training on production technical specifications, so as to ensure that each production personnel can master the technical requirements of the project, and need to formulate unified operation instructions and quality standards, so as to unify the standards of personnel in dealing with same issues.

The third is to pay attention to the key points of the Update Project quality inspection, that is to say, the focus should be on whether the changed areas in the two periods of data have been updated and whether the updated land categories are correct. For example, the change of man-made surface in urban areas is relatively large every year, so more attention can be paid to the change of artificial cover, waters and farmland during quality inspection. We should also strengthen the quality inspection of achievements, reduce the occurrence of systematic and universal errors, and make full use of quality inspection software to effectively control the quality of achievements.

5. CONCLUSION

This paper puts forward a set of targeted technical processes and methods for the quality inspection of GlobeLand30 update data

of the Update Project, summarizes the inspection contents and key points, which has been successfully used in the actual inspection of the Update Project from 2017 to 2018, and has proved its effectiveness and operability, and can effectively find the quality issues of the achievements. This paper also sorts out and analyzes some common issues, and puts forward the corresponding solutions, which can provide some technical reference for the quality control of the GlobeLand30 update data of the Update Project.

ACKNOWLEDGEMENTS

This work was supported by the 2019 Young Talent Subsidy of Technology Innovation Talent Training Program of the Ministry of Natural Resources.

REFERENCES

- Chen, J., Chen, J., Liao, A.P., et al, 2014: Concepts and Key Techniques for 30 m Global Land Cover Mapping. *Acta Geodaetica et Cartographica Sinica*, 43(6), 551-557.
- Chen, X., Zhu, L., Peng, S., Hou, L.L., 2016: Knowledge Construction of the Data Quality Checking Rules of GlobeLand30 Artificial Surfaces Type. *Journal of Beijing University of Civil Engineering and Architecture*, 32(4), 39-46.
- Chen, T., Li, G.Y., Tao, S., Zhou, X., Bi, K., 2018: A Method for Extracting Land Cover Change Information Oriented to Geographic National Conditions Monitoring. *Journal of Geomatics*, 43(3), 103-107.
- Chen, H.P., Zhang, L., Zhao, Y.S., Xie, W.J., Gao, W.C., 2018: Methods and Practices of Quality Inspection of Land Cover Classification Product. *Science of Surveying and Mapping*, 43(9), 58-62.
- 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.
- Li, Y., 2016: Basic Problems and Analysis of Land Cover Data Acquisition in Geographical Conditions. *Geomatics & Spatial Information Technology*, 39(7), 152-154.
- Liu, J.Y., Peng, S., Chen, J., Liao, A.P., Zhang, Y.S., 2015: Knowledge Based Quality Checking Method and Engineering Practice of GlobeLand30 Cropland Data. *Bulletin of Surveying and Mapping*, 2015(4), 42-48.
- 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.
- 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.
- Xia, L., Min, X.F., Chen, W.P., 2018: On the Large-scale Information Extraction for Land-cover Classifications in Americas. *Standardization of Surveying and Mapping*, 34(2), 19-23.