Generate Integrated Land Cover Product for Regional Climate Model by Fusing Different Land Cover Products

Hao Gao[1(✉)], Gensuo Jia[2], and Yu Fu[3]

[1] National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China
gaohao@cma.gov.cn

[2] Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
jiong@tea.ac.cn

[3] Climate Change Research Center, Chinese Academy of Sciences, Beijing 100029, China
fuyu@mail.iap.ac.cn

Abstract. Land cover and its change play an important role as a critical variables in global change studies and environment sciences. With many satellite-derived land cover products have become available at global and regional scales, it is possible to improve the quality of the product by integrating existing ones. In this paper, we propose a methodology to generate integrated land cover product for regional climate model over China. The integration rules were established based on the accuracies information from the ground truth. The accuracy of the integrated product was greatly improved, with overall accuracy of 68.7%, and class-specific accuracy ranging from 25.7% to 91.2%. Additionally, the spatial patterns of the land cover over China were well captured, and good agreement with the Landsat-based classification was achieved. Further, our results implicates the quality of the land cover products for integration are significant critical in our approach, the accuracy of the integrated land cover product is dependant on accuracy of the original products used for integration, especially of the local products. This integrated product could potentially improve the performance of regional climate models by providing better estimates of key land surface parameters over the region.

Keywords: Land cover · Satellite · Integration · Fusion · China

1 Introduction

Land use/cover change is considered as a critical variable in global change studies and environment science, which represents the influence of human activity and environment change and drives the biosphere-atmosphere interaction [1]. Land cover classification

Funded by: National Key Research and Development Program of China (Grant No. 2016YFA0600303), and the National Natural Science Foundation of China (Grant No. 41571425).

S. Wang and A. Zhou (Eds.): CollaborateCom 2016, LNICST 201, pp. 665–675, 2017.
DOI: 10.1007/978-3-319-59288-6_67

provides vital baseline information on the biophysical features of the Earth's surface, which plays an important role in regional to global scale Earth models, because land cover information is required to provide the initial conditions for parameterizing land surface processes [2]. Changes in land cover may also lead to significant consequences in environmental quality, climate change, and ecosystem health [3].

Currently, a number of global and regional scales land cover products have been developed from satellite data for various scientific and management needs [4–8]. They were widely used in a range of scientific studies including ecosystem, environment, climate change, and sustainable development. Although some progresses have been made for improving spatial parameterization, the quality of land cover products are still considered as a major bootle-neck that compromise the performance of regional models [9, 10]. Previous evaluation and inter-comparison of the land cover products showed obvious discrepancies and uncertainties existed among different land cover products, because of the differences in input data, spatial coverage, classifier algorithm, spatial resolution, classification schemes, and acquisition period of satellite data [11–13]. A more reliable and consistent land cover product are particularly important and essential for various applications.

Several international organizations have fostered land cover harmonization and strategies for interoperability and synergy among existing land cover products [14]. Recently, there were some attempts for end users to develop reliable land cover data by integrating different products based on fuzzy agreement scoring [15], Dempster-Shafer evidential reasoning [8], the majority rules [16], and synergistic approach [17, 18]. Despite of insufficient reference data available for thoroughly validating, the integrated products were considered more accurate than the originals. Though there are still some uncertainties with these methods, mainly because the existed products were produced using different input data, classifier algorithm, spatial resolution and classification schemes, developing integrated products from multi-satellite data is a reasonable alternative approach for land cover quality improvement in short-term.

The objective of this paper is to assess the overall and class-specific accuracy of the existing satellite-derived land cover products based on the validation points over China, and to develop a new framework for generating an integrated land cover product with IGBP classification scheme for regional climate model, according to specific fusion rules established in terms of accuracies of the various satellite-derived land cover products. A key focus is on assessing the accuracy improvement of the new integrated land cover products.

2 Data and Method

2.1 Land Cover Products

In this study, the 2001 MODIS Collection 51 global land cover product (MC51 hereafter) at 500-m spatial resolution with IGBP classification scheme, Global Land Cover 2000 (GLC2000) at 1-km spatial resolution with the Land Cover Classification System (LCCS), and some regional scale land cover products, such as the National Land Cover Datasets (NLCD) at 1-km spatial resolution, the Environment and Ecological Science

Data Center for Western China (WESTDC) land cover at 1-km spatial resolution with IGBP classification scheme were selected for integration.

2.2 Validation Points

A specific effort was made to establish validation points from multiple sources, including field investigation, coordinated enhanced observation project in arid and semi-arid China, Asia Fluxnet sites, Chinese Ecosystem Research Network (CERN) sites, Taiwan long term Ecological Research Network (TERN), Terrestrial Ecosystem Monitoring Sites (TEMS), and Degree Confluence Project (DCP). All the land cover categories of the validation points were interpreted according to the IGBP classification scheme. Finally, 1254 validation points were selected throughout China for subsequent validation efforts.

2.3 Evaluating Accuracy of the Land Cover Products

Most of the land cover products are validated, and their overall accuracy and class-specific accuracy are reported [5, 19]. Because different validation database and methods were used, the reported accuracy measures are not comparable and may not be considered as truly robust quantitative estimates. Here, all the land cover products for integration were evaluated using confusion matrix based on the validation points, the overall accuracy and class-specific accuracies were calculated and served as the foundation for integration. The integrated products were also evaluated against the validation points. Moreover, three subsets were selected for examining the spatial pattern of integrated product against the classification map interpreted from the Lansat TM, and for quantitative analyzing the each land cover class.

2.4 Land Cover Product Integration

Our integration approach was based on an idea that the more agreement of same land cover class among the products, the higher probability or likelihood that the specific land cover class exists at that pixel [15–17]. The process of integration was divided into two phases. First, all land cover products were assessed with the validation points, and each land cover categories were ranked based on the accuracy of the land cover categories before integration (Table 1). Then, each land cover categories was integrated based on rank value. The integrated process for each land cover categories based on rank value was illustrated in Fig. 1a. Land cover product #1 (LC#1) had the highest accuracy of the specific land cover categories evaluated by ground truth, and followed by LC#2, LC#3 and LC#4. The black pixels indicate that specific land cover category exist in each product. If all 4 products agree on the specific land cover categories exist in the pixel, a value of 1 is assigned for this pixel. A value of 2 was assigned to the pixel where the LC#1, LC#2 and LC#3 agree the specific land cover categories but LC#4 does not. There are 15 possible rank value with 4 land cover products, which was shown in Table 1. Finally, 17-class integrated layer, which were generated in the second phase, were integrated for the final land cover integration based on the value of rank assigned to each of pixels (Fig. 1b). A specific land cover category was assigned to a pixel if that category has the highest priority.

Table 1. The datasets ranking for integrating four land cover products.

Rank value	Land Cover#1	Land Cover#2	Land Cover#3	Land Cover#4
1	+	+	+	+
2	+	+	+	−
3	+	+	−	+
4	+	−	+	+
5	−	+	+	+
6	+	+	−	−
7	+	−	+	−
8	+	−	−	+
9	−	+	+	−
10	−	+	−	+
11	−	−	+	+
12	+	−	−	−
13	−	+	−	−
14	−	−	+	−
15	−	−	−	+

Land Cover#1 has the highest accuracy of the specific land cover category, followed by Land Cover#2, #2 and #4. The symbol "+" indicates the presence of the specific land cover category at the pixel and "−" indicates absence.

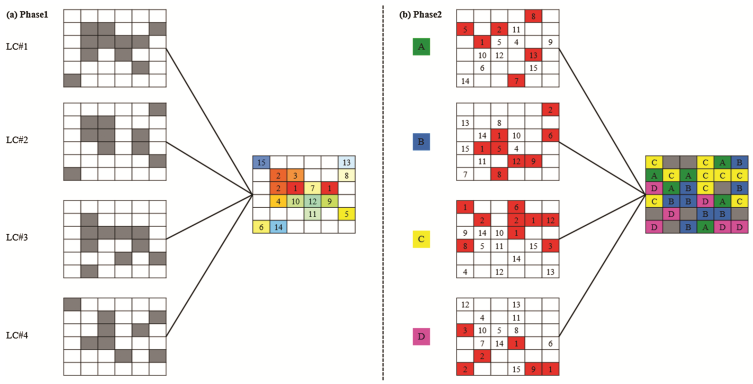

Fig. 1. Multi-satellite product integration based on rank value. (a) Integration of each land cover categories; The black pixels indicate that specific land cover category exist in each product, LC#1 has the highest accuracy assessed with the validation points followed by LC#2, LC#3 and LC#4. The numbers in pixels are the rank value which means the priority for integration, the smaller of the number, the higher priority for integration. (b) Final land cover product integration. A, B, C and D are the integrated land cover categories with the rank value generated from the Phase 1, and the numbers in the pixels indicate the priority of the specific land cover category.

Table 2. The overall and class-specific accuracy of the land cover products for integration

Land cover class	Accuracy (%)			
	MC51 2000	WESTDC	GLC2000	NLCD2000
0. Water	64.0	80.0	54.0	48.0
1. Evergreen needleleaf forest	33.3	35.9	38.5	/
2. Evergreen broadleaf forest	35.0	20.0	45.0	/
3. Deciduous needleleaf forest	22.7	18.2	31.8	/
4. Deciduous broadleaf forest	25.9	22.2	59.3	/
5. Mixed forest	46.8	25.5	21.3	/
6. Closed shrublands	19.6	30.4	16.1	30.0
7. Open shrublands	35.3	26.5	14.7	/
8. Woody savannas	41.2	44.1	14.7	/
9. Savannas	19.2	38.5	0.0	/
10. Grasslands	67.6	80.3	57.6	72.3
11. Permanent wetlands	16.1	32.3	19.4	19.4
12. Croplands	60.3	77.6	53.0	70.6
13. Urban and built-up	65.6	78.5	15.1	39.8
14. Cropland/natural vegetation mosaic	25.4	19.7	0.0	/
15. Snow and ice	44.4	55.6	50.0	61.1
16. Barren or sparsely vegetated	78.8	75.4	68.6	63.6
Overall	53.7	61.7	41.9	60.9

3 Results

3.1 Accuracy of the Land Cover Products

The overall and class-specific accuracy of land cover products used for integration were validated by validation points (Table 2). The overall accuracy of the MC51 2001, GLC2000, WESTDC, and NLCD 2000 are 53.7%, 41.9%, 61.7%, and 60.9%, respectively in China, which are remarkably lower than global area-weighted accuracy. The class-specific accuracies varied significantly. As expected, the accuracies of water, grasslands, croplands, snow and ice, and barren or sparsely vegetated were generally greater than the other categories. By contrast, the accuracies of remaining land cover categories were much lower, and varied significantly among products. Generally, the accuracies of forests in GLC2000 were much greater than in others. Other high accuracy cases include closed shrublands in WESTDC and NLCD 2000, open shrublands in MC51 2001 and WESTDC, permanent wetlands and urban and built-up in WESTDC, and snow and ice in NLCD 2000. Clearly, each land cover product has its advantages and limitations with specific land cover categories, which provided great opportunity and possibility for us to improve the quality of land cover data or integrate a new dataset that contained the advantages of each dataset based on the identified uncertain areas.

3.2 Comparison Between the MC51 and Integrated Land Cover Product

An integrated land cover product were created based on multi-satellite land cover products (Fig. 2). Remarkable differences were found between the MC51 products and integrated product. The major difference was found on the Tibetan Plateau, Inner Mongolia, Loess Plateau, and the edge area of barren area of northwest China, where grasslands replaced the open shrublands and barren or sparsely vegetated. Integrated product demonstrated similar land cover classification as previous results [8, 20], and similar to the 1:1000000 scale China vegetation map. Meanwhile, more oases in the desert are identified in integrated product. It seems to be reliably mapped, grasslands were clearly underestimated by MODIS, due to the presence of mixed classes such as Natural/cropland vegetation, and also the different definition of the grassland.

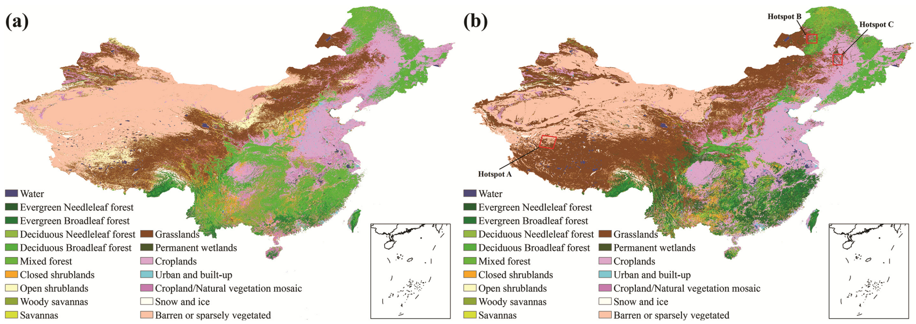

Fig. 2. Overview of the MODIS land cover products and integrated product. (a) MC51 2001; (b) Integrated land cover product for 2001.

Main plant functional types, and spatial pattern of forests in China are captured in the integrated product. Evergreen needleleaf forest are located in the south China and the edge area of Tibetan Plateau; Evergreen Broadleaf forest are mainly dominant in south China; Deciduous needleleaf forest are mainly placed in the Daxing'an Mountains of northwest China; Deciduous broadleaf forest are located in major mountain ranges in northwest and north China. By contrast with MC51 product, we found that much mixed forest are replaced by different forest types. The main patterns of forest in China are better captured by our integrated product, which are much similar to the China vegetation map. Such achievements are mainly contributed by GLC2000, due to its high accuracy of the forest types, which were produced by the regional and local experts. MODIS clearly overestimated the areas of mixed forest in China. Indeed, it is very difficult to distinguish classes with similar spectro-temporal-texture signals, especially for mapping continuum into discreted forest categories.

In addition, more inland water and permanent wetlands are presented in Sanjiang Plain, the largest wetland in China; and more inland water and permanent wetlands are identified in the flood plains of the Yellow River and Yangtze River. Meanwhile, more permanent wetlands are captured on the Tibetan Plateau, which are consistent with local studies [20, 21]. Further, more pixels were categorized into cropland on the Loess Plateau, and edge of the Mongolia plateau. All these characteristics are better captured in our product, indicated by greater agreements with local products and validation points. Indeed, the classes with much higher accuracy in local land cover products are well represented in the integrated product. Local land cover product played an important role in the integration process, and their contribution are significant. The quality of the participant land cover products are critical in our integration method, because their classification confidence could transferred to the integrated product, and affected the quality of the final product.

3.3 Accuracy of the Integrated Land Cover Product

The integrated land cover product for 2001 were also validated with validation points (Table 3). The overall accuracy of the integrated product were 68.7%, a 15.0% increase from original MC51 product. The class-specific accuracies differed significantly, ranging from 25.7% (cropland/natural vegetation mosaic) to 91.2% (grasslands). Generally, the accuracy of most categories was significantly increased in integrated product. However, the accuracy of mixed forest, open shrubland, woody savannas, and barren or sparsely vegetated were decreased. The grassland has the greatest increase of accuracy (23.6%), followed by permanent wetlands (22.6%), deciduous broadleaf forest (22.2%), water (18%), croplands (17.9%), savannas (15.4%), deciduous needleleaf forest (13.7%), closed shrublands (12.5%), snow and ice (11.7%), and evergreen broadleaf forest (10.0%). In addition, the percentage area of forests, grassland, built-up in integrated products generally have high agreements with the statistical yearbook, the forest inventory data (1999–2003), and the 1:1000000 scale China vegetation map. It is very interesting to see that major improvements of accuracy were achieved in classes with high accuracy in the local land cover product, such as water, closed shrubland, grassland, permanent wetlands, croplands, urban and built-up, snow and ice. Obviously, the local land cover product with high accuracy play an important role in the integration procedure. On the other hand, some land cover classes still have low accuracy, such as open scrublands, woody savannas, savannas, cropland/natural vegetation mosaic, and mixed forests. These classes may be easily confused due to the spectral similarity. For woody savannas, and savannas, only one classification scheme was used for MODIS product, and their accuracies are much lower than others. Discrepancy of the accuracy among the individual categories largely resulted from the significant difference of participant products, highly supports from local land cover products are significantly important for our integration method.

Table 3. The overall and class-specific accuracy of the integrated product validated by validation points

Land cover class	Accuracy (%)
0. Water	82.0
1. Evergreen needleleaf forest	34.8
2. Evergreen broadleaf forest	45.0
3. Deciduous needleleaf forest	36.4
4. Deciduous broadleaf forest	48.1
5. Mixed forest	31.5
6. Closed shrublands	32.1
7. Open shrublands	30.5
8. Woody savannas	26.5
9. Savannas	34.6
10. Grasslands	91.2
11. Permanent wetlands	38.7
12. Croplands	78.2
13. Urban and built-up	67.7
14. Cropland/natural vegetation mosaic	25.7
15. Snow and ice	55.6
16. Barren or sparsely vegetated	78.3
Overall	68.7

3.4 Detail Comparison of the Products in Subsets

Three local subsets were selected for comparing with 1:1000000 scale China vegetation map, and classification products based on the Landsat-7 satellite data to further evaluate the integrated product (Fig. 3). Generally, the land cover pattern of the integrated products in each subset appears similar to the China vegetation map and Landsat-7 classification. The integrated product was validated against the Landsat-7 classification. The accuracy of the integrated product was 53.4%, 52.8%, and 55.3% for region A, B and C, respectively, which have improved by about 8.2% to 24.8%. For region A in the Tibetan Plateau, the distribution pattern of the land cover classes in integrated product presented good agreement with Landsat-based classification and national vegetation map. The MC51 products apparently underestimated grassland. For region B in the boundary area of the Daxin'anling Mountains, the MC51 products significantly overestimated croplands, and the transition zone were also not clear. The extent of the cropland in integrated product is accordant with the reality, and forest types were also classified at high accuracy. For region C in the Sanjiang plain, MC51 product underestimated water, permanent wetlands, and overestimated the grassland. But in integrated product, they agreed with reference data well. Overall, our integrated product achieved greater overall accuracy than MC51 product. But, some land cover classes were more or less overestimated in some regions. It is interesting where the grassland or wetlands was overestimated in local product, they were also overestimated in integrated product. Apparently, the accuracy of the participant products also can affect the quality of the integrated product.

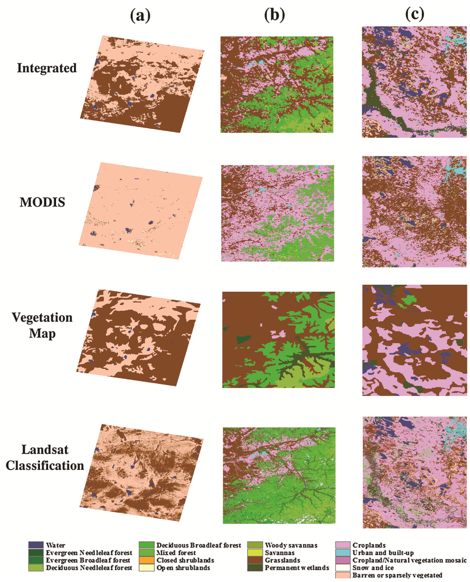

Fig. 3. Comparison of the three local subsets among four land cover map. (a) Subset A in the Tibetan Plateau of northwest China; (b) Subset B in the Daxin'anling Mountain of northeast China; (c) Subset C in the Sanjiang plain of northeast China.

4 Discussion and Conclusion

A method based on the accuracy of multi-satellite land cover products was proposed and applied to integrate a land cover product for 2001 over China. The accuracy of the MODIS, GLC2000, WESTDC, and NLCD2000 land cover products were validated with validation points to establish integration rules. Generally, the distribution patterns of the land cover in integrated product were captured reliably. The overall accuracy of the integrated product was 68.7%, a major improvement compared to the original MC51 product. Most of the class-specific accuracy in integrated product increased. Additionally, the distribution patterns of the land cover in our integrated product showed good agreement with Landsat-based classification and 1:1000000 scale national vegetation map. The quality of the land cover products for integration are critical in our integration method, the accuracy of the integrated product is dependant on the accuracy of participant land cover products for integration. Local land cover products played an important role in the integration process, their contribution are significant. Major improvement of classification accuracy was found in classes that are supported by the local products.

The integration rules in our study are established in terms of the accuracy validated by the ground truth, which are more objective and credible, but some limitations still need to be considered. First, in the preprocessing process, the LCCS classification schemes used in the GLC2000 product was translated to IGBP classification scheme according to recommended relationship. Meanwhile, the different spatial resolution with the products were resampled to the same spatial resolution, which may have introduced additional uncertainty. Second, the accuracy of participant land cover products need to be further validated as additional reference data become available. Our results implicate high accuracy of the participant products could bring great improvement of the quality of the integrated product. Finally, the integrated product also need to be further rigorous validated against ground truth. With high quality land cover products become available in the future, we also expect to improve methods to produce more accurate and up-to-date land cover products for climate modeling.

Acknowledgments. The study was supported by the National Key Research and Development Program of China (Grant No. 2016YFA0600303). We thank the Land Processes Distributed Active Archive Center (LP DAAC), the Joint Research Centre (JRC) of Global Vegetation Monitoring Units, the Data Sharing Infrastructure of Earth System Science, and the Environmental & Ecological Science Data Center for West China, National Natural Science Foundation of China for their helpful response to our inquiry on satellite product.

References

1. Turner, B.L., Lambin, E.F., Reenberg, A.: The emergence of land change science for global environmental change and sustainability. Proc. Natl. Acad. Sci. **104**, 20666–20671 (2007)
2. Bonan, G.B., Levis, S., Kergoat, L., Oleson, K.W.: Landscapes as patches of plant functional types: an integrating concept for climate and ecosystem models. Global Biogeochem. Cycles **16**, 1021 (2002)
3. Feddema, J.J., Oleson, K.W., Bonan, G.B., et al.: The importance of land-cover change in simulating future climates. Science **310**, 1674–1678 (2005)
4. Bartholomé, E., Belward, A.S.: GLC2000: a new approach to global land cover mapping from earth observation data. Inter. J. Remote Sens. **26**, 1959–1977 (2005)
5. Friedl, M.A., Sulla-Menashe, D., Tan, B., et al.: Modis collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. **114**, 168–182 (2010)
6. GLOBCOVER 2009: Products Description and Validation Report. http://ionia1.esrin.esa.int/docs/GLOBCOVER2009_Validation_Report_2.2.pdf
7. Liu, J., Liu, M., Deng, X., Zhuang, D., Zhang, Z., Luo, D.: The land use and land cover change database and its relative studies in China. J. Geog. Sci. **12**, 275–282 (2002)
8. Ran, Y.H., Li, X., Lu, L., Li, Z.Y.: Large-scale land cover mapping with the integration of multi-source information based on the Dempster-Shafer theory. Int. J. Geogr. Inf. Sci. **26**, 169–191 (2012)
9. Ge, J., Qi, J., Lofgren, B.M., et al.: Impacts of land use/cover classification accuracy on regional climate simulations. J. Geophys. Res. **112**, D05107 (2007)
10. Sertel, E., Robock, A., Ormeci, C.: Impacts of land cover data quality on regional climate simulations. Int. J. Climatol. **30**, 1942–1953 (2010)
11. McCallum, I., Obersteiner, M., Nilsson, S., Shvidenko, A.: A spatial comparison of four satellite derived 1 km global land cover datasets. Int. J. Appl. Earth Obs. Geoinf. **8**, 246–255 (2006)
12. Herold, M., Mayaux, P., Woodcock, C.E., et al.: Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ. **112**, 2538–2556 (2008)

13. Kaptué, T.A.T., Roujean, J.L., De Jong, S.M.: Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. Int. J. Appl. Earth Obs. Geoinf. **13**, 207–219 (2011)
14. Herold, M., Woodcock, C.E., Gregorio, A.D., et al.: A joint initiative for harmonization and validation of land cover datasets. IEEE Trans. Geosci. Remote Sens. **44**, 1719–1727 (2006)
15. Jung, M., Henkel, K., Herold, M., Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens. Environ. **101**, 534–553 (2006)
16. Kinoshita, T., Iwao, K., Yamagata, Y.: Creation of a global land cover and a probability map through a new map integration method. Int. J. Appl. Earth Obs. Geoinf. **28**, 70–77 (2014)
17. Fritz, S., You, L., Bun, A., et al.: Cropland for sub-Saharan Africa: a synergistic approach using five land cover data sets. Geophys. Res. Lett. **38**, L04404 (2011)
18. Pérez-Hoyos, A., García-Haro, F.J., San-Miguel-Ayanz, J.: A methodology to generate a synergetic land-cover map by fusion of different land-cover products. Int. J. Appl. Earth Obs. Geoinf. **19**, 72–87 (2012)
19. Mayaux, P.: Validation of the global land cover 2000 map. IEEE Trans. Geosci. Remote Sens. **44**, 1728–1739 (2006)
20. Zhang, Z., Wang, X., Zhao, X., et al.: A 2010 update of National Land Use/Cover Database of China at 1:100000 scale using medium spatial resolution satellite images. Remote Sens. Environ. **149**, 142–154 (2014)
21. Zhang, S.: An introduction of wetland science database in China. Sci. Geogr. Sin. **22**, 188–189 (2002)