National Scale Land Cover Classification Using the Semiautomatic High-Quality Reference Sample Generation (HRSG) Method and an Adaptive Supervised Classification Scheme

Amin Naboureh, *Member, IEEE*, Ainong Li, *Member, IEEE*, Jinhu Bian, *Member, IEEE*, and Guangbin Lei

***Abstract*—The advent of new high-performance cloud computing platforms [e.g., Google Earth Engine (GEE)] and freely available satellite data provides a great opportunity for land cover (LC) mapping over large-scale areas. However, the shortage of reliable and sufficient reference samples still hinders large-scale LC classification. Here, selecting Turkey as the case study, we presented a semiautomatic high-quality reference sample generation (HRSG) method using the publicly available scientific LC products and the linear spectral unmixing analysis to generate high-quality ground samples for the years 1995 and 2020 within the GEE platform. Furthermore, we developed an adaptive random forest classification scheme based on Köppen–Geiger climate zone classification system. Our rationale was related to the fact that large-scale study areas often contain multiple climate zones where the spectral signature of the same LC class may vary within different climate zones that can lead to a poor LC classification accuracy. To have a robust assessment, the generated LC maps were evaluated against independent test datasets. In regard to the proposed sample generation method, it was observed that HRSG can generate high-quality samples independent of the characteristics of scientific LC products. The high overall accuracy of 92% for 2020 and 90% for 1995 and satisfactory results for producer's accuracy (ranging between 83.4% and 99.3%) and user's accuracy (ranging between 86.1% and 99.7%) of nine LC classes demonstrated the effectiveness of the proposed framework. The presented methodologies can be incorporated into future studies related to large-scale LC mapping and LC change monitoring studies.**

***Index Terms*—Global land cover (LC) products, Google Earth Engine (GEE), LC classification, linear spectral unmixing (LSU), reference samples.**

Manuscript received 8 November 2022; revised 9 January 2023; accepted 30 January 2023. Date of publication 2 February 2023; date of current version 15 February 2023. This work was supported in part by the National Science Foundation of China under Grant 42090015 and Grant 42171382, in part by the National Key Research and Development Program of China under Grant 2020YFA0608702, in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19030303, and in part by the Youth Innovation Promotion Association CAS under Grant 2019365. *(Corresponding author: Ainong Li.)*

The authors are with the Research Center for Digital Mountain and Remote Sensing Application, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China (e-mail: amin.nabore@mails.ucas.ac.cn; ainongli@imde.ac.cn; bianjinhu@imde.ac.cn; leiguangbin@imde.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3241620

I. Introduction

LAND cover (LC) map is among the key component of different research, such as climate change, sustainable development, natural hazards risk assessments, and ecological protection [1], [2]. Remote sensing (RS) data have long been used to observe, identify, and monitor LC types at various spatial and temporal scales [3]. In this manner, the launch of Landsat-1 in 1972 marked the beginning of a new era in LC mapping, as the first systematic observations of the earth's surface were made available [4]. Over the last few decades, significant progress has been made in the development of up-to-date and accurate LC mapping. In particular, the availability of high-resolution data (i.e., Landsat and Sentinel) and the emergence of cloud computing platforms [i.e., Google Earth Engine (GEE)] have sped up the generation of LC maps over large-scale studies [5], [6].

Access to high-quality training and validation sample is among the fundamental parts of the LC classification procedure [7]. Traditionally, these sets of samples (training and validation) were often small and acquired via field surveys and/or expert visual interpretation of very high-resolution imagery. However, in the case of large-scale studies (e.g., national and continental scales) and inaccessible/remote regions, it is very challenging to organize field surveys involving thousands of field plots and to access very high-spatial resolution satellite imagery [8]. Moreover, in recent years, machine learning (ML) algorithms have been broadly implemented in LC mapping, where the collection of a large set of training samples is required in order to be effective [9]. Therefore, there is a pressing need for developing robust methodologies to deal with the shortage of reliable and sufficient reference samples for large-scale LC mapping.

To deal with this obstacle, a range of publicly available datasets and techniques have been introduced by the RS community. For instance, the first all-season sample (FAST) dataset [10], providing nearly 90 000 samples, and the Geo-wiki dataset [11], including about 150 000 samples, were two valuable datasets that have been developed in recent years. Such datasets undoubtedly fill a crucial gap, but they cannot be directly used in the classification task of another year(s) as they are limited to a specific year [12]. For instance, the FAST dataset was collected based on the high-resolution Google Earth imagery circa 2015,

and therefore may not be applicable for LC mapping in the years before or after 2015.

In recent years, several methodologies have also been developed to address the shortage of high-quality reference samples in LC mapping. For instance, Radoux et al. [13] proposed an automated approach to produce reference samples using the MODIS LC products. Additionally, Zhang and Roy [14] collected only high-quality pixels (stable pixels between 2009 and 2011) from 500-m MODIS to produce a 30-m resolution LC map at a continental scale. However, the quality of the generated reference sample data using these methodologies is criticized due to the inherent classification errors (i.e., misclassified and misplaced pixels) from the coarse resolution LC products [11]. Recently, Li et al. [15] developed an automatic phenology learning (APL) method for producing high-quality reference samples using publicly available scientific datasets. However, the APL method highly relies on the quality of selected datasets, whereas only global and national LC products with moderate accuracy and spatial resolutions are available for most parts of the world. Therefore, the APL method may not achieve reasonable results at a global scale level. Additionally, numerous studies have been devoted to generate samples for a specific LC class. These methodologies are, however, restricted to a specific LC class. For example, Lin et al. [16] and Ma et al. [17] introduced robust techniques to generate reliable reference samples for built-up class based on GlobeLand30 products. Generally speaking, although the proposed methods and datasets are valuable assets in monitoring LC mapping, they cannot fully meet the users' needs for large-scale LC classification.

Generating up-to-date, reliable, and cost-effective LC maps for large-scale areas require the use of highly automated advanced algorithms [18]. To this end, the emergence of GEE, which not only offers easy visualizations and analysis of big RS data but also hosts a variety of classification algorithms [5], has effectively aided in LC mapping over large-scale areas [19], [20]. Besides all the advantages, several limitations of GEE in large-scale LC mapping have been reported. For example, Ghorbanian et al. [20] and Amani et al. [21] stated that large-scale LC mapping within GEE may involve considerable practical difficulties when a large number of features is required due to the user memory limit of GEE. In another study, Shafizadeh-Moghadam et al. [22] claimed that large-scale LC classification using GEE is still a time-consuming task and may fail to execute when the number of input features is high. Given that increasing the number of input features (e.g., vegetation indices and topographical data) is one of the most common ways to boost LC classification accuracy [2], [23], employing a classification scheme that can handle large-scale LC classification without decreasing the number of input features is critical.

Large areas (e.g., national and continental scales) naturally comprise multiple climate zones where the spectral signature within the same LC class often varies at different climate zones [22]. It has been reported that the intraclass variability of spectral signatures in different climates, which is caused by variations of vegetation covers and artificial materials, may result to poor LC classification accuracy [24]. To address this issue, employing an adaptive classification scheme that executes a large-scale area into smaller sections based on the climate zones can be considered as an effective solution [25]. This procedure can shorten processing time and substantially decrease the impact of climate zone variation in large-scale LC classification.

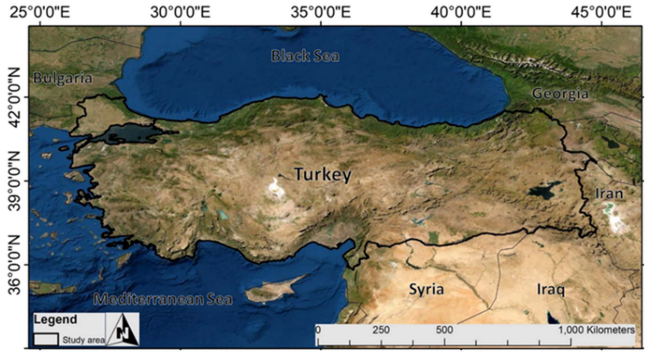


Fig. 1. Location of the study area.

As the accessibility to LC data is limited over large areas, the available global LC products (i.e., GlobeLand30) are widely used in different studies. For example, Chen et al. [26] claimed that the GlobeLand30 products have been utilized by scholars and users over 120 countries across five continents. However, the accuracy of global LC products varies for different parts of the world. For example, although the satisfactory overall accuracy (OA) of GlobeLand-30 was reported as 77.90% for Iran [27], 80% for Italy [28], 80.1% for Nepal [29], and 82.4% for China [30], Sun et al. [31] reported a poor OA of 46% for Central Asia region (Kazakhstan, Turkmenistan, Tajikistan, Kyrgyzstan, and Uzbekistan). Given that the accuracy of recently released the GlobeLand-30 product for the year 2020 has yet not been investigated over a large-scale area, the present study seeks to evaluate the accuracy of GlobeLand-30 version 2020 at a national scale.

Here, we present a semiautomatic high-quality reference sample generation (HRSG) method and an adaptive classification scheme to generate LC maps for Turkey, as an experiment site, in 2020 and 1995 at 30-m resolution. We have selected the year 2020 because there are several available global LC products (e.g., GlobeLand-30) for the year 2020 that can be used for comparison purposes. On the other hand, we have selected the year 1995 to assess whether or not the proposed method (HRSG) can identify high-quality samples using outdated LC products. Our ultimate goals are 1) introducing a robust workflow for generating high-quality reference sample data applicable for large-scale studies in different parts of the world and 2) introducing an efficient classification scheme for large-scale LC mapping using GEE.

II. Study Area

The focus of the present work is Turkey, as a transcontinental country in both Asia and Europe, which is home to nearly 84 million inhabitants with an area of about 783 562 km^2. This country is extended between 26°–45° E and 36°–42° N (see Fig. 1). The reason for taking Turkey as the case study is twofold:

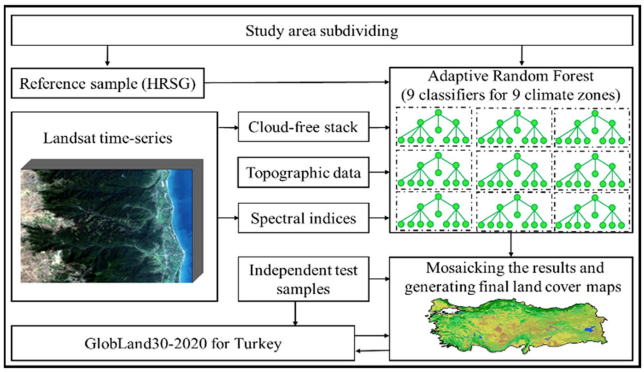


Fig. 2. General steps of the present work.

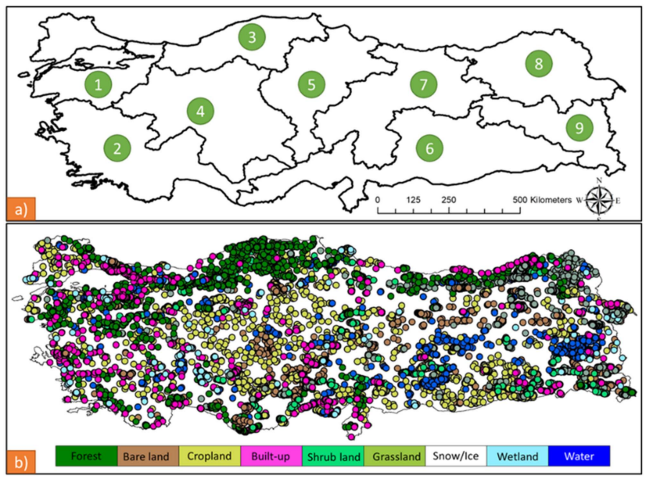


Fig. 3. (a) Location of introduced subareas based on the Köppen–Geiger climate classification. (b) Spatial distribution of tests samples.

to the best of our knowledge, Turkey lacks a country-wide LC map at 30-m resolution, and, the country has a wide range of climatic conditions, elevation levels, and spatial distributions of LC classes that allow us to comprehensively evaluate the robustness of the proposed methods in this study. Turkey consists of a wide range of LC classes, dominated by forest in the northern part and cropland, grassland, and bare land in the center and southern parts. The mean annual minimum/maximum air temperature varies from –30 °C in February in the eastern part to 45 °C in July in the south-eastern part. Mean annual precipitation ranges from 2220 mm in the north-eastern part (Black Sea costs) to 258 mm in the central and south-eastern parts. It is worth mentioning that nine major LC classes namely forest, grassland, shrub land, cropland, built-up, water bodies, wetland, permanent snow and ice, and bare land were considered to generate national scale LC maps for Turkey in 2020 and 1995.

III. Materials and Methods

As shown in Fig. 2, the general steps of the present work can be divided into five parts: 1) study area subdividing, 2) generating input features, 3) collecting reference samples using the HRSG method, 4) applying the adaptive classification scheme, and 5) accuracy assessment of the generated LC maps and GlobeLand30 version 2020 product for Turkey.

A. Study Area Subdividing

The large extent of the study area and an immense size of input features caused "computational time out" and "user memory limited" errors while processing the entire Turkey at once. Our initial analysis also illustrated that climate variation could lead to poor classification accuracy for some spectrally similar classes, such as grassland, cropland, and shrub lands. Therefore, to address the aforementioned issues, the study area was divided into nine subareas based on the major climate zones of the study area as suggested by Shafizadeh-Moghadam et al. [22] and Ebrahimy et al. [25]. To this end, the Köppen–Geiger climate classification system was used to divide the study area into smaller subareas [see Fig. 3(a)]. The Köppen–Geiger [32] integrates the average monthly and yearly precipitation and temperatures as well as seasonal precipitation to split the study region into relatively homogeneous areas.

B. Input Features

In this study, 3413 and 3400 atmospherically-corrected Tier 1 Landsat-8 and Landsat-5 scenes with $<75\%$ cloud cover were processed to generate LC maps for Turkey in 2020 and 1995. Enhanced by Calderon-Loor et al. [33], to minimize cloud interference and missing data, at each time-step, a stack of Landsat images for the target year $\pm$ one year was generated (e.g., Landsat-5 images from 1994 to 1996 were used to generate the year 1995 mosaic). It should be noted that, to ensure the reliability of generated samples, normalized difference vegetation index (NDVI) values of each candidate sample were derived from Landsat images (see Section III-C4) in the growing seasons (from February 15th to September 15th) because the growing season is the most suitable time for spectral change detection [12].

From the available spectral bands of Landsat images, six spectral bands, including near-infrared, green, blue, red, short-wave infrared 1, and short-wave infrared 2, were utilized. Then, the FMASK algorithm [34], which is available in the GEE platform, was applied to remove clouds and cloud shadows. Furthermore, a mean function was applied to all Landsat images to merge all datasets into a single mosaic image for the classification task. Additionally, based on a random forest (RF) feature selection method [35], the following variables were appended to the cloud-free stack: soil-adjusted total vegetation index (SATVI), NDVI, modified normalized difference water index (MNDWI), normalized difference built-up index (NDBI), modified soil-adjusted vegetation index (MSAVI), enhanced vegetation index (EVI), slope, elevation, and aspect. The calculation formulas of these indices can refer to Table I.

C. Reference Sample Generation

As discussed earlier, the reliability and sufficiency of reference sample data are the foundation of an accurate supervised LC classification task [23], [42]. However, it is very difficult to collect the *in-situ* sample for LC mapping over large areas using the traditional reference collecting methods. Here,

TABLE I
VEGETATION INDEX FORMULAE

Abbrev.	Formula	Reference
NDVI	$\frac{NIR - Red}{NIR + Red}$	[36]
NDBI	$\frac{SWIR - NIR}{SWIR + NIR}$	[37]
SATVI	$1.5 \times \frac{SWIR1 - Red}{SWIR + Red + 0.5} + \frac{SWIR2}{2}$	[38]
MSAVI	$\frac{2 \times NIR + 1 - \sqrt{b(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}}{2}$	[39]
MNDWI	$\frac{Green - SWIR1}{Green + SWIR1}$	[40]
EVI	$2.5 \times \frac{NIR - Red}{NIR + 6 \times Red - 7.5 \times Blue + 1}$	[41]

we presented the HRSG method to deal with the shortage of reliable and sufficient reference samples for LC mapping over large-scale areas. The HRSG method assumes that none of the scientific LC products are error-free [43] and limited training samples/endmembers (representative spectra of each LC class) samples can represent intraclass diversity [44]. Therefore, to identify high-quality samples for each LC class, the HRSG extracts estimated pixels of the target class from the available LC products/maps, first. Then, based on a linear spectral unmixing (LSU) analysis and limited endmembers, it identifies high-quality pixels for the target class. Fig. 4 shows the main steps of HRSG as follows: 1) extracting all pixels of the target class from the available LC product(s) and clipping cloud-free stacks based on the obtained pixels; 2) introducing limited endmembers and applying the LSU analysis; 3) producing candidate samples; and 4) identifying high-quality samples.

1) Extracting All Pixels of the Target Class From the Available LC Product(s): In this study, MCD12Q1, GlobeLand30, JRC Global Surface Water, and the Tsinghua FROM-GLC Year of Change to Impervious Surface datasets were allocated to identify high-quality samples for different LC classes based on their characteristics (see Table II). It should be noted that the classification systems of the MCD12Q1 and GlobeLand30 were reclassified based on the classification system of the present work (see Table III). For example, to generate samples for forest class, six different forest classes of MCD12Q1 product merged into one class and labeled as forest class according to the target classification system of this study.

As shown in Table II, since the JRC provides surface water bodies since 1985 at global scale (30-m resolution) [45], the JRC dataset solely was used for generating samples for water class. Following the same logic, the Tsinghua FROM-GLC Year of Change to Impervious Surface [46] dataset was used for generating samples for the built-up class. For the remaining classes, the MCD12Q1 and GlobeLand30 products were used to generate samples for other classes. More specifically, in the case of the year 2020, the MCD12Q1 version 2020 and GlobeLand30 version 2020 were overlaid, and their common pixels were extracted, whereas for the year 1995, the MCD12Q1 version 2001 and GlobeLand30 version 2000 products were utilized. After extracting pixels for the target class, the final stacks of the target years (1995 and 2020) (see Section III-B), including six spectral bands, six vegetation indices, and three topographical data, were clipped based on the extracted pixels.

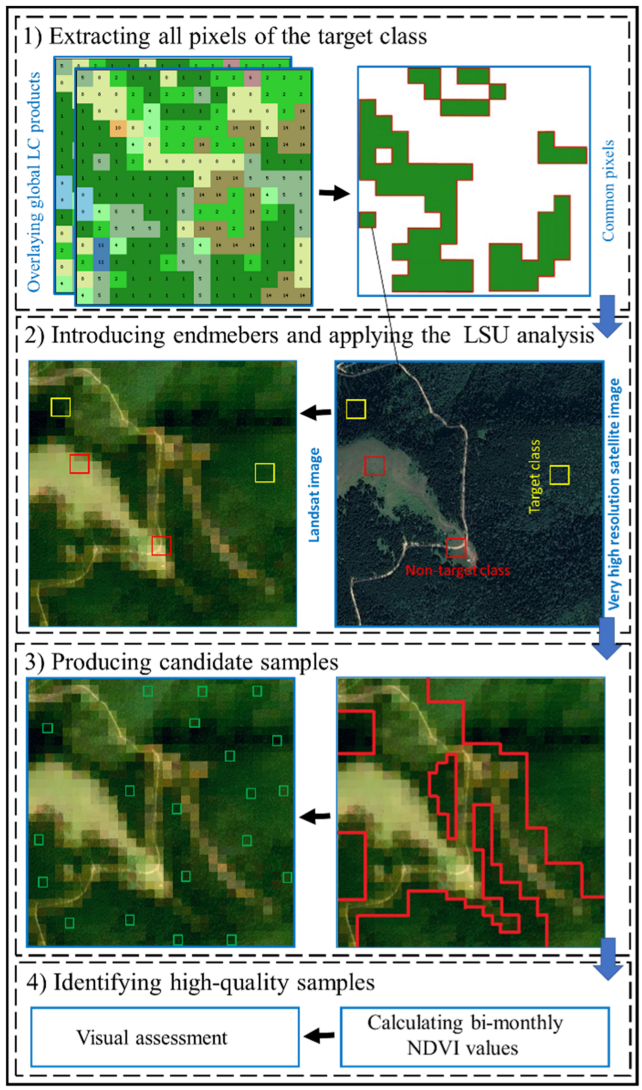


Fig. 4. Simplified framework describing the proposed sample generation method.

2) Introducing Limited Endmembers and Applying the LSU Analysis: After extracting pixels for the target class, limited training samples (pixels/objects) were introduced as the endmembers to contribute to the LSU analysis. Given that intraclass variation of RS images is one of the main factors of uncertainty in LSU analysis [47], using the precise interpretation, multiple endmembers were selected using the random sampling technique. For manual interpretation, as suggested by Yang and Huang [48], we referred to freely available very high spatial resolution Google Earth images, Landsat images, and their NDVI time series. In some more detail, first, the samples were collected based on Google Earth images and Landsat images. Then, their NDVI times series were calculated based on Landsat images. If the NDVI time-series of a given endmember was stable, it was determined as the true label. For each class, the number of endmembers was limited to 50 endmembers (pixels and/or

TABLE II
INFORMATION OF USED LC PRODUCTS IN THIS STUDY

Global LC product	Resolution	Temporal	Class
Tsinghua FROM-GLC Year of Change to Impervious Surface	30 m	1985–2018	Built-up
JRC Yearly Water Classification	30 m	1985–2020	Water
MCD12Q1.006 MODIS Land Cover Type Yearly Global	500 m	2001–2019	Forest Cropland Shrub land Grassland Barren Snow/Ice Wetland
GLOBELAND-30	30 m	2000, 2010, 2020	

TABLE III
REDESIGNING THE CLASSIFICATION SYSTEM OF THE GLOBAL LC PRODUCTS BASED ON THE TARGET CLASSIFICATION SYSTEM OF THE PRESENT STUDY

Target legend	GLOBELAND30	MCD12Q1.006
Forest	Forest (ID=20)	Evergreen needleleaf forests (ID=1) Evergreen broadleaf forests (ID=2) Deciduous needleleaf forests (ID=3) Deciduous broadleaf forests (ID=4) Mixed forests (ID=5) Woody savannas (ID=8)
Shrub land	Shrub land (ID=40)	Closed shrub lands (ID=6) Open shrub lands (ID=7) Savannas (ID=9)
Cropland	Cultivated Land (ID=10)	Croplands (ID=12) Cropland/natural vegetation (ID=14)
Grasslands	Grasslands (ID=30)	Grasslands (ID=10)
Wetlands	Wetlands (ID=50)	Permanent wetlands (ID=11)
Bare land	Tundra (ID=70) Bare land (ID=90)	Barren (ID=16)
Water bodies	Water bodies (ID=60)	Water bodies (ID=17)
Built-up	Artificial Surfaces (ID=80)	Urban and built-up lands (ID=13)
Snow and ice	Snow and ice (ID=100)	Snow/ice (ID=15)

objects) covering at least 1% of the pixels obtained in the previous stage, following Manohar Kumar et al. [49]. Based on the collected endmembers, the LSU method was utilized to generate an initial batch of high-quality reference samples. The LSU model assumes that the reflectance of a given pixel equals the linear weighted sum of the pure spectra of the components present in that pixel weighted by their proportional coverage [50]. The LSU model is mathematically described as follows:

$$P_i = \sum_{i=1}^{nc} (\mathrm{r}_{ci}.\mathrm{f}_c) + e_i, \quad i = 1, \ldots, nb \tag{1}$$

where P_i represents the value (reflectance) of a pixel in band i, r_{ci} represents the value of endmember c in band i, f_c is the fraction of endmember C, e_i illustrates the residual error in band i, nc is the number of endmembers, and nb is the number of spectral bands.

The LSU model can be represented in matrix form as follows:

$$P_{(nb\times 1)} = R_{(nb\times nc)}.\, F_{(nc\times 1)} + E_{(nb\times 1)}. \tag{2}$$

3) Producing Candidate High-Quality Samples: In the present work, the high-quality pixels of the target class were determined through a series of trials and errors. After acquiring pixels of the target class, a stratified random sample generation technique was used to produce the first batch of the candidate samples for the target class from the obtained area. To be included in the initial batch of candidate high-quality samples of the target class, a pixel must meet two requirements. First, its values for one of the target endmembers (considering the intraclass variation of the target class) have to be higher than 0.6. Second, its values for all nontarget endmembers should be less than 0.1.

4) Identifying High-Quality Samples: Since the first batch of candidate samples may include potential errors, two different techniques were used to ensure the reliability of generated

samples. First, time-series of NDVI values (bimonthly) in the growing season (see Section III-B) were calculated for each candidate sample because it is reported that NDVI time-series are sensitive to changes in vegetation covers [51]. To this end, based on the literature, several rules were introduced to confirm the candidate samples. For example, the maximum NDVI value of 0.15 was implemented to filter out any possible vegetation samples from the bare land class [52]. Additionally, cropland samples with NDVI values lower than 0.3 in the growing season were also eliminated because the highest NDVI values of cropland classes often appear in the growing season because of the management role [48]. Second, following Huang et al. [12], samples were finalized after visual interpretation assessments by senior image interpreters who had experience in land change monitoring and image interpretation practices. The equation was used to evaluate the accuracy of the samples generated for each LC class

$$Accuracy = \frac{N^r}{N^r + N^w} \tag{3}$$

where N^r represents the total number of correctly identified samples for a given LC class and N^w represents the number of incorrectly identified samples for that class.

D. Adaptive RF Classification

RF, as a computationally efficient ensemble classifier [53], effectively distinguishes among spectrally similar LC classes [54]. Generally speaking, when a sample is entered into the RF classifier, each decision tree model determines the category of the sample, and the algorithm ultimately selects the most frequently reported category (majority voting) as the class category. It has been reported that RF often outperforms single decision tree models while maintaining some of the advantages of decision tree models (e.g., ability to interpret relationships between predictors and outcomes) [54]. The low sensitivity of RF to the normality of training data and its high performance in different landscapes have led to widely utilizing RF for LC classification over large areas [20], [55], [56] where various kinds of landscapes can be found and collecting normal distribution training data is a time and budget-consuming task.

Considering two main assumptions, an adaptive RF classification method was adopted in this study. First, spectral signatures of LC types are different at various climate zones. Second, producing an LC map for large-scale studies is time-consuming and challenging [22] when the number of input features is high. Accordingly, specific RF classifiers were developed for all nine subareas (see Section III-A), and the results for subareas were mosaicked to generate the final LC map at each time step. It should be noted that the number of decision trees and the number of variables per split were adjusted at 250 and the square root of the number of variables, respectively, after testing various tuning settings for the RF classifiers.

E. Performance Evaluation

To date, visual assessment and comparison of generated samples against the trusted and well-known LC products have been used to evaluate the quality of sample generation methods. To this end, a visual assessment was conducted to assess the accuracy of the acquired samples for the years 1995 and 2020. In this study, we did not use the available global LC products for the performance evaluation of the HRSG because it has been reported that all the available LC products include a certain amount of errors [12], [23]. For example, Fig. 5 shows that the MCD12Q1.006 product misses a large extent of snow cover in 2020 for the northeastern part of Turkey.

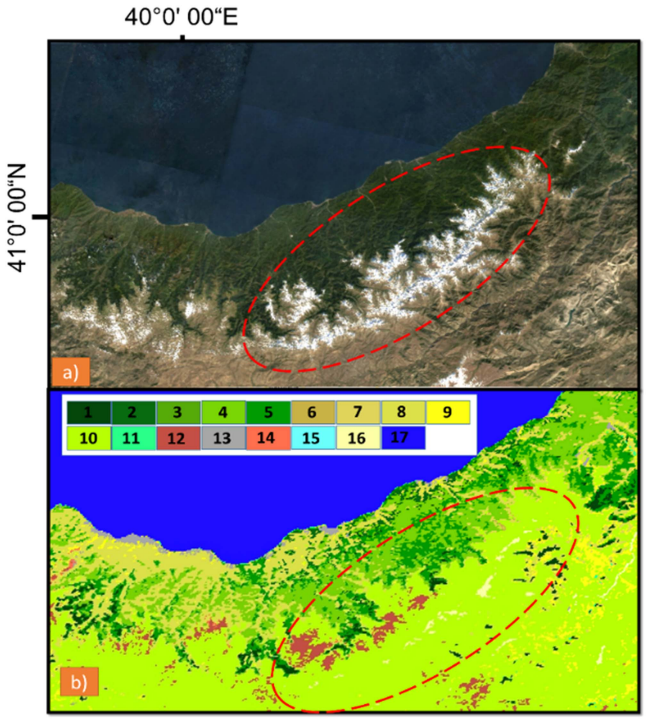


Fig. 5. Example of poor classification accuracy of MCD12Q1-2020 product in the experiment site. (a) Cloud-free Landsat stack for 2020. (b) MCD12Q1 product for 2020 (detailed information about the LC classes is available in Table II).

Three well-known accuracy assessment metrics, namely OA, user's accuracy (UA), and producer's accuracy (PA), were calculated based on independent test datasets (over 6000 samples at each time step) to evaluate the generated LC maps of Turkey (2020 and 1995) and GlobeLand30 (2020 version) [see Fig. 3(b)]. The validity of each sample (true label) is determined based on the visual interpretation of the very high-resolution images in Google Earth, Landsat images and their NDVI time series.

IV. Results

A. Performance Evaluation of HRSG

In this study, the HRSG method was presented to deal with the shortage of high-quality reference samples in LC classification over large areas. Turkey was selected as the study area to assess the performance of the HRSG method as this country comprises various climate zones and elevation ranges. Using LSU and multiple endmembers, a total of nearly 22 000 samples were

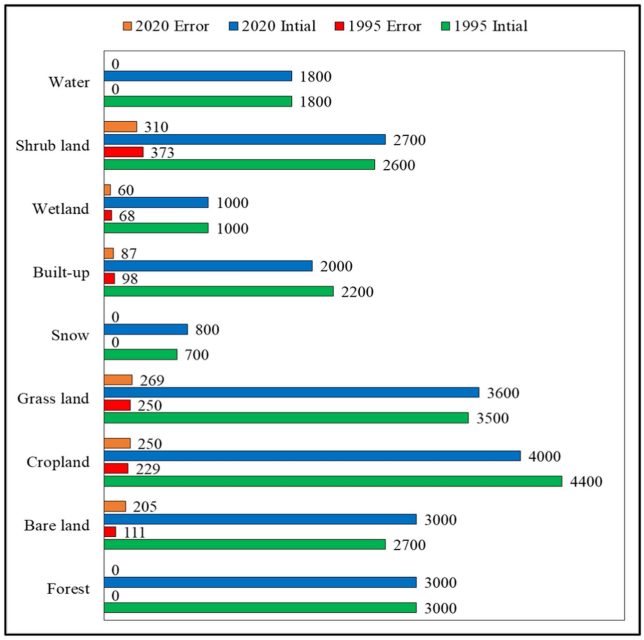


Fig. 6. Results for visual assessments of the generated samples by the HRSG method.

generated for the years 2020 and 1995. Later, a visual assessment was conducted to generate the final dataset. The visual assessment of generated samples for the years 2020 and 1995 illustrated that less than 5.5% of the generated samples [1129 samples in 1995 ($\sim$ 5.1%) and 1181 samples in 2020 ($\sim$ 5.4%)] by the HRSGM method were noisy (see Fig. 6). The highest errors were witnessed in shrub land (11.5% error in 2020 and 14.3% error in 1995) and grassland (7.5% error in 2020 and 7.1% error in 1995) classes. In contrast, forest, snow, and water classes were noise-free. These results show that HRSG is a viable and effective technique for generating abundant and high-quality reference samples.

B. LC Classification

Within GEE, an adaptive RF classification scheme was adopted to generate LC maps for the entire of Turkey in 2020 and 1995 (see Section III-D). As discussed earlier, precision and vision assessments were used to evaluate the performance of generated maps. Our analysis showed that the produced final LC maps for the years 2020 and 1995 are noise-free and provide satisfactory representations of all nine LC classes, according to visual interpretation with freely available very high-resolution photographs (see Fig. 6). The estimated areas for each LC class can be seen in Fig. 7. The dominant LC class was cropland covering 28.3% and 33.7% of Turkey in 1995 and 2020, respectively. In contrast, the least area of the country was covered by snow class (1995 = 0.1% and 2020 = 0.1%). Additionally, the generated maps showed OA of about 92% and 90% for the years 2020 and 1995, respectively. As shown in Table IV, the PA of different LC classes ranged from 86.8% to 99.3% for 2020 and ranged from 83.4% to 98.2% in 1995, whereas the UA of the nine LC classes varied between 87.5% and 99.7% in 2020, and ranged from 86.1% to 97.7% in 1995. The water class showed the highest PA ($>$ 98.2%) and UA ($>$ 97.7%) values among all the LC classes. In contrast, the lowest UA ($\sim$ 86.1%) and PA ($\sim$ 83.4%) values belonged to the grassland class.

C. GlobeLand-30 for Turkey

In this study, the GlobeLand-30 (2020 version) product, as one of the widely used and well-known global LC products, was used to compare with the generated LC map of Turkey in 2020. Table V presents accuracy assessment results for the GlobeLand-30 product version 2020 based on the independent reference test dataset (see Section III-E). As can be seen in Table V, the GlobeLand-30 product showed OA of nearly 60% where UA and PA accuracy considerably varied among different LC classes. Water class showed the highest PA and UA values with 93.7% and 97.3%, respectively. In contrast, the bare land class had the lowest UA ($\sim$ 3.4%) and the shrub land class showed the lowest PA ($\sim$20%). Regarding the mapped area proportions calculated for each LC class (see Fig. 8), it was witnessed that the grassland class covers most part of Turkey ($\sim$40%), and the wetland and snow classes cover the least of Turkey (nearly 1% for each class).

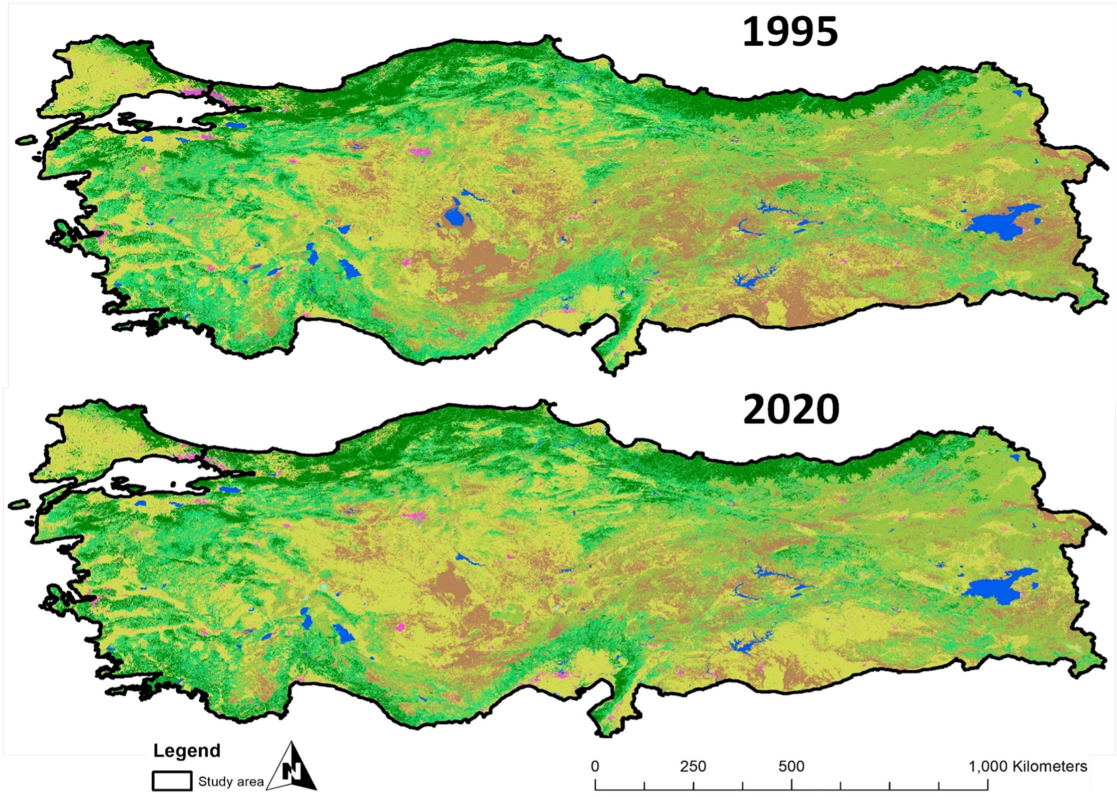
V. Discussion

A. Performance of the HRSG

This study proposed a semiautomatic method for generating reference samples, namely HRSG, based on LSU and multiple endmembers, for LC classification over large areas. Based on the conducted experiments, the maximum number of per class endmembers was limited to 50 (pixel/object) because the additional numbers of endmembers no further significantly increased/decreased the mean spectral signatures for each LC class that was used in target/nontarget class separation process. Adopting limited endmembers has two main advantages: it can decrease the computational efficiency because less spectra need to be processed, and increase the accuracy and less spectral confusion may occur [44], [57].

The bimonthly values of NDVI time-series were used in the sample generation procedure because the NDVI value is sensitive to changes in vegetation covers [51]. Our findings showed that a distinct temporal pattern for generating NDVI time-series could help to increase the accuracy of generated samples. For example, samples for the bare land class should not have obvious differences in their NDVI values during a year. The cropland class had the highest variation in NDVI values because of the management role. In contrast, samples for grassland/shrub land showed a peak greenness in March and April. This finding is in line with [15] where the importance of NDVI time series in generating reliable reference samples for LC classification has been stressed.

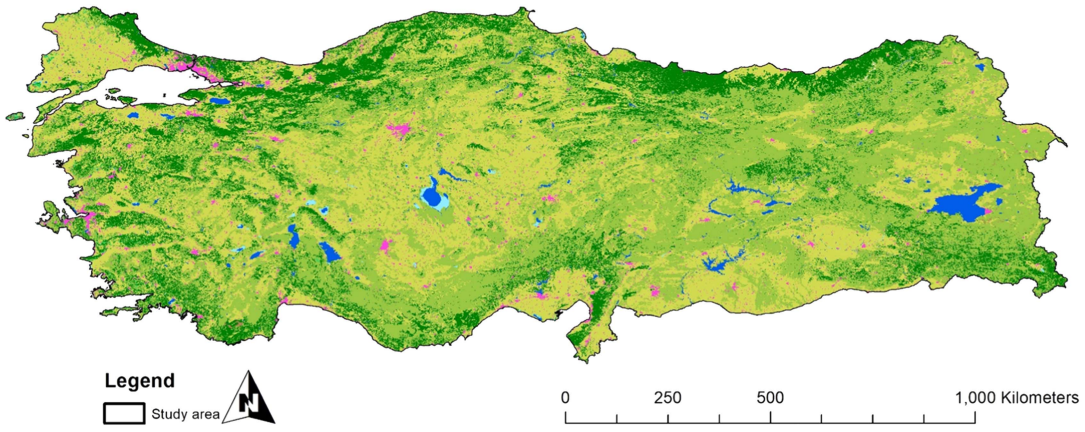
Using the HRSG method, approximately 22 000 reference samples of nine LC classes for 2020 and 1995 were initially generated where about 5.5% of them (each year) were removed during the visual assessment procedure (see Fig. 6). The obtained reasonable accuracy for the generated reference samples



	Forest	Bare land	Cropland	Built-up	Shrub land	Grassland	Snow/Ice	Wetland	Water
1995	91334	130573	222133	5374	126387	194724	1554	1004	10705
2020	90536	90178	263632	7890	113729	204697	1071	1007	11048

Area (Square kilometer)

Fig. 7. Generated LC map for Turkey in 2020 and their estimated area for each LC class.



Forest	Bare land	Cropland	Built-up	Shrub land	Grassland	Snow/Ice	Wetland	Water
139048	1008	219813	16591	1466	320970	2015	88	12787

Area (Square kilometer)

Fig. 8. GlobeLand-30 LC product for Turkey in 2020.

TABLE IV
(A) ACCURACY ASSESSMENT RESULTS FOR THE GENERATED MAP OF ENTIRE TURKEY (1995) BASED ON THE INDEPENDENT TEST DATASET (THE UNCERTAINTY OF THE ACCURACY IS PROVIDED AT 95% CONFIDENCE LEVEL). (B) ACCURACY ASSESSMENT RESULTS FOR THE GENERATED MAP OF ENTIRE TURKEY (2020) BASED ON THE INDEPENDENT TEST DATASET (THE UNCERTAINTY OF THE ACCURACY IS PROVIDED AT 95% CONFIDENCE LEVEL)

(a)	Forest	Bare land	Cropland	Built-up	Shrub land	Grassland	Snow/Ice	Wetland	Water	Total	UA
Forest	**827**	9	9	0	26	34	0	3	0	908	91.1 ± 1.9
Bare land	0	**1030**	41	42	0	48	0	0	0	1161	88.7 ± 1.8
Cropland	13	21	**956**	8	13	40	0	1	0	1052	90.8 ± 1.7
Built-up	0	35	20	**698**	0	3	0	0	0	756	92.3 ± 1.9
Shrub land	31	0	11	0	**558**	27	0	5	0	632	88.3 ± 2.5
Grassland	0	22	51	6	45	**804**	6	0	0	934	86.1 ± 2.2
Snow/Ice	0	3	0	0	2	4	**95**	0	3	107	88.7 ± 6
Wetland	2	0	6	0	3	3	0	**135**	5	154	87.6 ± 5.2
Water	0	3	0	0	0	1	0	6	**437**	447	97.7 ± 1.4
Total	1369	1126	1099	754	647	964	96	147	445	**6151**	
PA	94.7± 1.5	91.7± 1.6	87.4 ± 2	92.6± 1.9	86.2± 2.6	83.4 ±2.3	94.0 ±4.7	90 ±4.8	98.2 ±1.2	OA= 90.1 ±0.7	

(b)	Forest	Bare land	Cropland	Built-up	Shrub land	Grassland	Snow/Ice	Wetland	Water	Total	UA
Forest	**845**	5	5	0	21	28	0	4	0	908	93 ± 1.6
Bare land	0	**1094**	28	13	0	24	1	0	1	1161	94.2 ± 1.3
Cropland	1	30	**963**	6	9	42	0	1	0	1052	91.5 ± 1.7
Built-up	0	25	25	**701**	0	5	0	0	0	756	92.7 ± 1.8
Shrub land	32	0	12	0	**570**	15	0	3	0	632	90.2 ± 2.3
Grassland	3	24	47	5	33	**817**	5	0	0	934	87.5 ± 2.1
Snow/Ice	0	4	0	0	1	6	**96**	0	0	107	89.7 ± 5.7
Wetland	1	1	10	0	1	3	0	**136**	2	154	88.3 ± 0.5
Water	0	0	0	0	0	1	0	0	**449**	447	99.7 ± 1.2
Total	882	1183	1096	725	635	941	102	140	447	**6151**	
PA	95.7 ± 1.3	92.5 ±1.5	88.3±1.9	96.7 ±1.3	89.7 ± 2.3	86.8 ±2.1	94.1±4.6	94.3±3.8	99.3 ±0.7	OA= 92.1 ±0.6	

TABLE V
ACCURACY ASSESSMENT RESULTS FOR THE GLOBELAND-30 PRODUCT FOR TURKEY (2020) BASED ON THE INDEPENDENT TEST DATASET (THE UNCERTAINTY OF THE ACCURACY IS PROVIDED AT 95% CONFIDENCE LEVEL)

	Forest	Bare land	Cropland	Built-up	Shrub land	Grassland	Snow/Ice	Wetland	Water	Total	UA
Forest	**813**	0	14	0	0	81	0	0	0	908	89.5 ± 1.9
Bare land	18	**40**	85	85	1	830	1	88	13	1161	3.4 ± 1
Cropland	34	1	**994**	13	0	10	0	0	0	1052	94.4 ± 1.4
Built-up	10	0	51	**677**	0	16	0	2	0	756	89.5 ± 2.2
Shrub land	295	0	37	0	**1**	299	0	0	0	632	0.15 ± 0.3
Grassland	176	5	113	17	1	**619**	1	2	0	934	66.2 ± 3
Snow/Ice	17	3	0	0	0	50	**37**	0	0	107	34.5 ± 9
Wetland	6	0	28	0	2	12	0	**90**	16	154	58.4 ± 7.8
Water	0	0	5	0	0	5	0	2	**435**	447	97.3 ± 1.5
Total	1369	49	1327	792	5	1922	39	184	464	**6151**	
PA	59.3 ±2.6	81.6 ±10	74.9 ±2	85.4 ±2.4	20 ± 35	32.2 ±2.1	94.8 ±2	48.9 ±7	93.7 ±2.2	OA= 60.2 ±1.2	

for the year 1995 based on the MCD12Q1 version 2001 and GlobeLand-30 version 2000 revealed the feasibility of generating high-quality reference samples using outdated LC products. The reason can be related to this issue is that the HRSG method applies the LSU method to bypass the inherent classification errors from the outdated LC products. As can be seen in Fig. 6, the shrub land and grassland classes had the highest rates of error. Based on the visual assessment, the majority of errors in shrub land samples belong to grassland and vice versa. Besides, the accuracy of available LC products for shrub land class is relatively low, making sample generation a quite challenging task because this class required more endmembers extraction to reach the satisfactory output. Therefore, similar to Li et al. [15], it is suggested to integrate these two classes (shrub land and grassland) into a single class. On the other hand, the zero error for water, forest, and snow classes was reasonable due to the distinguishable spectral response of these classes from other classes (nontarget/target) as well as the high ability of the NDVI index in extracting different these classes from the other LC classes [58], [59].

Generally speaking, the HRSG method's effectiveness was related to a combination of LSU analysis, NDVI variations, and visual interoperation, all of which have been shown to play a significant role in sample generation [15], [60]. Additionally, because the presented methodology is based on the GEE and freely available global LC products, the high expense and complexity of traditional reference sample collection for LC classification are sustainably solved by the proposed method.

Overall, the advantages of HRSG can be concluded as follows.

1) It generates high-quality samples independent of the characteristics of scientific LC products.
2) It is applicable in different parts of the world.
3) The high cost and difficulty of traditional reference sample collection can be effectively obviated by HRSG.

B. Performance of the Proposed Classification Scheme

In this study, two 30-m resolution LC map of Turkey for 2020 and 1995 were generated using an adaptive classification schedule within the GEE platform. Unlike the majority of previous LC studies over large-scale areas that adopted a single classifier for LC mapping using the GEE platform [33], [60], this study merged the results of nine specific RF classifications to generate an LC map for Turkey at each time step (see Section III-D). The high accuracy assessment outcomes (see Tables IV and V) illustrate the effectiveness of the proposed classification scheme.

As can be seen in Tables IV and V, both minority and majority classes show acceptable UA and PA values, whereas it has been reported that standard ML classifiers (i.e., RF) often fail to achieve reasonable accuracies for minority classes [61], [62]. The reason can be that the proposed classification scheme divides the study area into smaller subareas and substantially avoids the occurrence of the data imbalance issue. For example, in the case of snow class, this class might have been influenced by the data imbalanced issue if the whole Turkey was classified using a single classifier because it only covers less than 1% of the whole country. However, based on the conducted subdividing strategy, the snow class placed in subclass number 8 (see Fig. 2) and covered nearly 7% of this subzone, which lead to high UA (94% in 1995 and 94.1% in 2020) and PA (88.7% in 1995 and 89.7% in 2020) values for this class.

Compared with the GlobeLand-30 product, it was observed that the proposed method achieving OA = 92.1% in 2020 outperformed the GlobeLand-30 product (OA = 60.2%). Except the PA value for cropland class and UA value for snow class, the obtained UA and PA values for all LC classes of the present work were higher than GlobeLand-30 product (see Fig. 9).

In the case of visual assessment, the generated LC maps in this study highly reflected the true LC type in the satellite images (see Fig. 10). As can be seen in Fig. 10, we compared the obtained LC map for Turkey in 2020 in this study with other LC products, including GlobeLand-30, Dynamic World [63], and GLC-FC30 [64]. More specifically, the overall layout of cropland, forest, and built-up classes of this study is roughly equivalent to other LC products. In contrast, in the remaining classes, the outcomes are different among the products that can

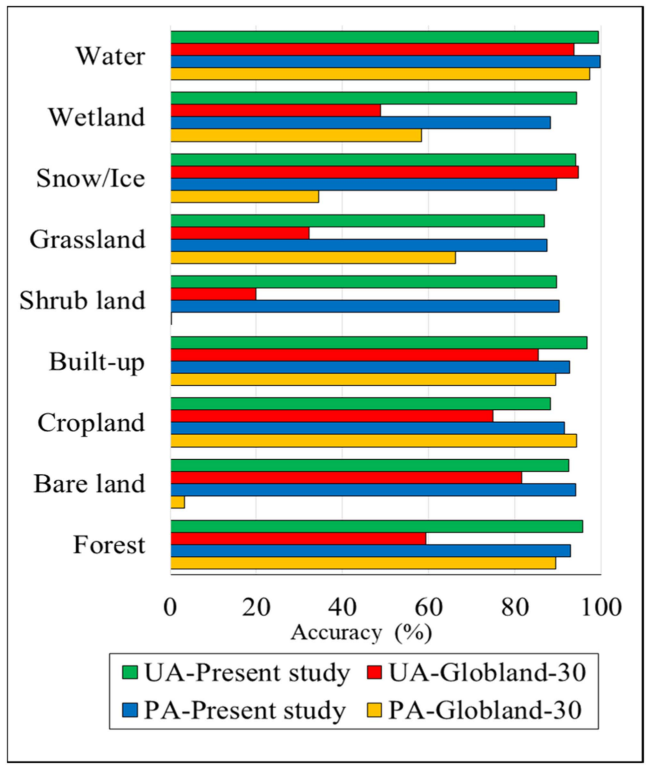


Fig. 9. Comparison in UA and PA values between the GlobeLand30 product and the generated LC map of present work for Turkey.

be related to the inconsistency in the definition of LC types for different LC classes. Compared with GlobeLand-30 version 2020, the main difference lies in water and shrub land classes. For example, Fig. 10 shows some misclassification in water class for the GlobeLand-30 product. In the case of the GLC-FCS30 product, we identify an overestimation of shrub land class in some cases. For example, as shown in Fig. 10(a), the GLC-FC30 classified cropland areas as shrub land class. Dynamic World could identify more built-up and cropland classes than our study that can be linked to the higher spatial resolution of the Dynamic World product (10-m resolution). In Fig. 10(a), it can be seen that bare land areas (dried lake's bed) were wrongly labeled as the snow/ice class by the Dynamic World product. Overall, given the high accuracy assessment results (vision and precision) and satisfactory visual assessment results, it can be concluded that our workflow is a robust scheme for large-scale LC mapping.

C. Limitations and Future Work

The HRSG method mainly relies on the quality of the selected endmembers. Even though the limited endmembers were used for generating reference samples, it required the manual interpretation by trained image interpreters to achieve reasonable accuracy. To deal with these issues, further studies can focus on using the available spectral information. For instance, for various cities at different parts of the world, multiple urban spectral libraries have been collected based on various sensors and spatial resolutions [65], [66]. Additionally, future studies

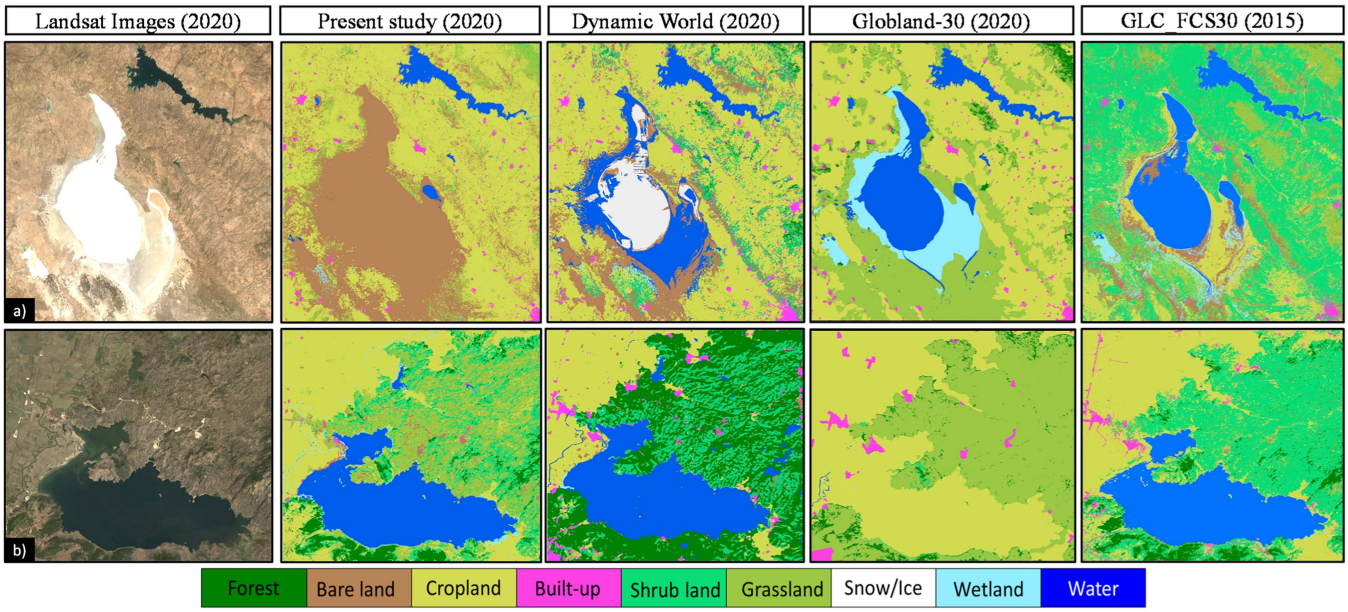


Fig. 10. Comparison of our results against GlobeLand-30, Dynamic World, and GLC-FCS30 products for the year 2020. The corresponding satellite images for the two zoomed areas are Landsat-8 annual median compositions in 2020. Center location: (a) = 38° 51′ 20″ N – 33° 30′ 36″ E, (b) = 37° 31′ 58″ N – 27° 27′ 27″ E.

can focus on automatic spectral unmixing techniques that are free of the need of endmember extraction [67]. On the other hand, HRSG requires time-series of Landsat images (see Section III-C). However, there was a shortage of high-quality observation for some candidate samples, particularly for 1995 because of the shortage of Landsat 5 TM scenes in some part of Turkey (mainly in mountainous areas). Finally, in this study, only NDVI values in the growing season were used to eliminate possible errors from the initial batch of candidate high-quality samples. Future studies can consider the role of other ancillary data in this manner.

VI. Conclusion

This article presents a novel method for generating high-quality reference samples over large areas, namely HRSG. This method uses LSU and global LC products (i.e., MCD12Q1 and GlobeLand30) to generate reference sample for LC mapping. Additionally, choosing Turkey as a case study, due to the large extent and climatic variation of the study area, an adaptive classification scheme was adopted to generate LC maps for Turkey in 1995 and 2020. Based on a visual assessment, we found that HRSG can effectively address the shortage of reliable and sufficient reference samples for LC classification over large areas. The success of this method is attributed to the quality of endmembers, NDVI values, and visual assessment. By evaluating the generated LC maps of Turkey in 1995 and 2020 with two independent datasets obtained, it was observed that the generated LC maps achieved high OA, UA, and PA values. Given the obtained high accuracies, the proposed methods can implement in other parts of the world for generating reliable LC maps.

Acknowledgment

Data Availability: The generated LC maps are available at: https://code.earthengine.google.com/?scriptPath=users%2Faminnabore%2Faminucas%3ALand-CoverTurkey

References

[1] J. C. Gómez, M. A. Wulder, and J. C. White, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 55–72, 2016, doi: 10.1016/j.isprsjprs.2016.03.008.

[2] R. Khatami, G. Mountrakis, and S. V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sens. Environ.*, vol. 177, pp. 89–100, 2016.

[3] Y. Ban, P. Gong, and C. Giri, "Global land cover mapping using Earth observation satellite data: Recent progresses and challenges," *ISPRS J. Photogrammetry Remote Sens.*, vol. 103, no. 1, pp. 1–6. 2015.

[4] S. I. Gordon, "Utilizing Landsat imagery to monitor land-use change: A case study in Ohio," *Remote Sens. Environ.*, vol. 9, no. 3, pp. 189–196, May 1980.

[5] N. Gorelick et al., "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017, doi: 10.1016/j.rse.2017.06.031.

[6] Q. Zhao et al., "Progress and trends in the application of Google Earth and Google Earth Engine," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3778.

[7] G. M. Foody and M. K. Arora, "An evaluation of some factors affecting the accuracy of classification by an artificial neural network," *Int. J. Remote Sens.*, vol. 18, pp. 799–810, 1997, doi: 10.1080/014311697218764.

[8] C. P. Giri, *Remote Sensing of Land Use and Land Cover: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2012.

[9] Q. Zhou et al., "Training data selection for annual land cover classification for the land change monitoring, assessment, and projection (LCMAP) initiative," *Remote Sens.*, vol. 12, 2020, Art. no. 699, doi: 10.3390/rs12040699.

[10] C. Li et al., "The first all-season sample set for mapping global land cover with Landsat-8 data," *Sci. Bull.*, vol. 62, pp. 508–515, 2017.

[11] S. Fritz et al., "A global dataset of crowdsourced land cover and land use reference data," *Sci. Data*, vol. 4, 2017, Art. no. 170075.

[12] H. Huang et al., "The migration of training samples towards dynamic global land cover mapping," *ISPRS J. Photogrammetry Remote Sens.*, vol. 161, pp. 27–36, 2020.
[13] J. Radoux et al., "Automated training sample extraction for global land cover mapping," *Remote Sens.*, vol. 6, pp. 3965–3987, 2014.
[14] H. K. Zhang and D. P. Roy, "Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification," *Remote Sens. Environ.*, vol. 197, pp. 15–34, 2017.
[15] C. Li, G. Xian, Q. Zhou, and B. W. Pengra, "A novel automatic phenology learning (APL) method of training sample selection using multiple datasets for time-series land cover mapping," *Remote Sens. Environ.*, vol. 266, 2021, Art. no. 112670.
[16] C. Lin, P. Du, A. Samat, E. Li, X. Wang, and J. Xia, "Automatic updating of land cover maps in rapidly urbanizing regions by relational knowledge transferring from GlobeLand30," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1397.
[17] X. Ma, X. Tong, S. Liu, X. Luo, H. Xie, and C. Li, "Optimized sample selection in SVM classification by combining with DMSP-OLS, Landsat NDVI and GlobeLand30 products for extracting urban built-up areas," *Remote Sens.*, vol. 9, no. 3, 2017, Art. no. 236.
[18] T. Hermosilla, M. A. Wulder, J. C. White, and N. C. Coops, "Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112780.
[19] J. Bian, A. Li, G. Lei, Z. Zhang, and X. Nan, "Global high-resolution mountain green cover index mapping based on Landsat images and Google Earth Engine," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 63–76, 2020.
[20] A. Ghorbanian, M. Kakooei, M. Amani, S. Mahdavi, A. Mohammadzadeh, and M. Hasanlou, "Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 276–288, 2020.
[21] M. Amani et al., "Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5326–5350, Sep. 2020, doi: 10.1109/JSTARS.2020.3021052.
[22] H. Shafizadeh-Moghadam, M. Khazaei, S. K. Alavipanah, and Q. Weng, "Google Earth Engine for large-scale land use and land cover mapping: An object-based classification approach using spectral, textural and topographical factors," *GISci. Remote Sens.*, vol. 58, no. 6, pp. 914–928, 2021.
[23] A. Naboureh, J. Bian, G. Lei, and A. Li, "A review of land use/land cover change mapping in the China-Central Asia-West Asia economic corridor countries," *Big Earth Data*, vol. 5, no. 2, pp. 237–257, 2021.
[24] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 154, pp. 151–162, 2019.
[25] H. Ebrahimy, H. Aghighi, M. Azadbakht, M. Amani, S. Mahdavi, and A. A. Matkan, "Downscaling MODIS land surface temperature product using an adaptive random forest regression method and Google Earth Engine for a 19-years spatiotemporal trend analysis over Iran," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2103–2112, Jan. 2021, doi: 10.1109/JSTARS.2021.3051422.
[26] F. Chen et al., "A landscape shape index-based sampling approach for land cover accuracy assessment," *Sci. China Earth Sci.*, vol. 46, pp. 1413–1425, 2016.
[27] J. J. Arsanjani, A. Tayyebi, and E. Vaz, "GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries," *Habitat Int.*, vol. 55, pp. 25–31, 2016.
[28] M. A. Brovelli et al., "The first comprehensive accuracy assessment of GlobeLand30 at a national level: Methodology and results," *Remote Sens.*, vol. 7, pp. 4191–4212, 2015.
[29] X. Cao et al., "Land cover mapping and spatial pattern analysis with remote sensing in Nepal," *J. GeoInf. Sci.*, vol. 18, pp. 1384–1398, 2016.
[30] Y. Yang, P. Xiao, X. Feng, and H. Li, "Accuracy assessment of seven global land cover datasets over China," *ISPRS J. Photogrammetry Remote Sens.*, vol. 125, pp. 156–173, 2017.
[31] B. Sun, X. Chen, and Q. Zhou, "Uncertainty assessment of GlobeLand30 land cover data set over Central Asia," *ISPRS Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 1313–1317, 2016.
[32] H. E. Beck et al., "Present and future Köppen-Geiger climate classification maps at 1-km resolution," *Sci. Data*, vol. 5, 2018, Art. no. 180214.
[33] M. Calderon-Loor, M. Hadjikakou, and B. A. Bryan, "High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112148.
[34] D. Frantz et al., "Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects," *Remote Sens. Environ.*, vol. 215, pp. 471–481, 2018.
[35] F. Zhang and X. Yang, "Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection," *Remote Sens. Environ.*, vol. 251, 2020, Art. no. 112105.
[36] J. W. Rouse Jr. et al., "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation," NASA, Washington, DC, USA, Tech. Rep. E75-10354, 1974.
[37] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.*, vol. 24, pp. 583–594, 2003.
[38] R. C. Marsett et al., "Remote sensing for grassland management in the arid southwest," *Rangeland Ecol. Manage.*, vol. 59, pp. 530–540, 2006.
[39] J. Qi et al., "A modified soil adjusted vegetation index," *Remote Sens. Environ.*, vol. 48, pp. 119–126, 1994.
[40] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, pp. 3025–3033, 2006.
[41] T. N. Carlson and D. A. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote Sens. Environ.*, vol. 62, pp. 241–252, 1997.
[42] G. Lei, A. Li, J. Bian, and Z. Zhang, "The roles of criteria, data and classification methods in designing land cover classification systems: Evidence from existing land cover data sets," *Int. J. Remote Sens.*, vol. 41, pp. 5062–5082, 2020.
[43] K. Winkler, R. Fuchs, M. Rounsevell, and M. Herold, "Global land use changes are four times greater than previously estimated," *Nature Commun.*, vol. 12, pp. 1–10, 2021.
[44] J. Degerickx et al., "A novel spectral library pruning technique for spectral unmixing of urban land cover," *Remote Sens.*, vol. 9, 2017, Art. no. 565.
[45] J. F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, pp. 418–422, 2016.
[46] P. Gong et al., "Annual maps of global artificial impervious area (GAIA) between 1985 and 2018," *Remote Sens. Environ.*, vol. 236, 2020, Art. no. 111510.
[47] C. Zhang et al., "Assessing the impact of endmember variability on linear spectral mixture analysis (LSMA): A theoretical and simulation analysis," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111471.
[48] J. Yang and X. Huang, "The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019," *Earth Syst. Sci. Data*, vol. 13, no. 8, pp. 3907–3925, 2021.
[49] C. Manohar Kumar, S. S. Jha, R. R. Nidamanuri, and V. K. Dadhwal, "Benchmark studies on pixel-level spectral unmixing of multi-resolution hyperspectral imagery," *Int. J. Remote Sens.*, vol. 43, pp. 1451–1484, 2022.
[50] J. J. Settle and N. A. Drake, "Linear mixing and the estimation of ground cover proportions," *Int. J. Remote Sens.*, vol. 14, pp. 1159–1177, 1993.
[51] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.
[52] J. R. Eastman et al., "Global trends in seasonality of normalized difference vegetation index (NDVI), 1982–2011," *Remote Sens.*, vol. 5, pp. 4799–4818, 2013.
[53] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[54] M. Beligiu and L. Dragut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.
[55] M. Mahdianpari et al., "A large-scale change monitoring of wetlands using time series Landsat imagery on Google Earth Engine: A case study in Newfoundland," *GISci. Remote Sens.*, vol. 57, pp. 1102–1124, 2020.
[56] H. Yang, X. Zhang, and A. J. Zehnder, "Water scarcity, pricing mechanism and institutional reform in northern China irrigated agriculture," *Agricultural Water Manage.*, vol. 61, pp. 143–161, 2003.
[57] C. Deng, "Automated construction of multiple regional libraries for neighborhoodwise local multiple endmember unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4232–4246, Sep. 2016.
[58] F. Maselli, "Monitoring forest conditions in a protected Mediterranean coastal area by the analysis of multiyear NDVI data," *Remote Sens. Environ.*, vol. 89, pp. 423–433, 2014.

[59] S. Szabo, Z. Gacsi, and B. Balazs, "Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories," *Landscape Environ.*, vol. 10, pp. 194–202, 2016.
[60] A. Naboureh et al., "Assessing the effects of irrigated agricultural expansions on Lake Urmia using multi-decadal Landsat imagery and a sample migration technique within Google Earth Engine," *Int. J. Appl. Earth Observ. GeoInf.*, vol. 105, 2021, Art. no. 102607.
[61] M. Azadbakht, C. S. Fraser, and K. Khoshelham, "Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform LiDAR data," *Int. J. Appl. Earth Observ. GeoInf.*, vol. 73, pp. 277–291, 2018.
[62] A. Naboureh, A. Li, J. Bian, G. Lei, and M. Amani, "A hybrid data balancing method for classification of imbalanced training data within Google Earth Engine: Case studies from mountainous regions," *Remote Sens.*, vol. 12, no. 20, 2020, Art. no. 3301.
[63] C. F. Brown et al., "Dynamic World, near real-time global 10 m land use land cover mapping," *Sci. Data*, vol. 9, no. 1, pp. 1–17, 2022.
[64] X. Zhang et al., "GWL_FCS30: Global 30 m wetland map with fine classification system using multi-sourced and time-series remote sensing imagery in 2020," *Earth Syst. Sci. Data Discuss.*, vol. 15, no. 1, pp. 1–31, 2022.
[65] U. Heiden, K. Segl, S. Roessner, and H. Kaufmann, "Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data," *Remote Sens. Environ.*, vol. 111, pp. 537–552, 2007.
[66] F. Priem and F. Canters, "Synergistic use of LiDAR and APEX hyperspectral data for high-resolution urban land cover mapping," *Remote Sens.*, vol. 8, 2016, Art. no. 787.
[67] Q. Wang, X. Ding, X. Tong, and P. M. Atkinson, "Spatio-temporal spectral unmixing of time-series images," *Remote Sens. Environ.*, vol. 259, 2021, Art. no. 112407.

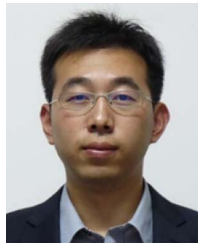
Amin Naboureh (Member, IEEE) received the Ph.D. degree in physical geography from the University of Chinese Academy of Sciences, Beijing, China, in 2022.

He is a Postdoctoral Researcher with the Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China. His research interests include, but are not limited to, land cover/use change monitoring, big remote sensing data processing, and utilizing advanced machine learning for natural hazards assessment and environmental management. A list of his research works can be found at https://www.researchgate.net/profile/Amin-Naboureh.

Dr. Naboureh was the recipient of the Scholarship Award of the CAS-TWAS President's fellowship program in 2018. He was also the recipient of the excellent international student and graduate titles from University of Chinese Academy of Sciences in 2021 and 2022, respectively.

Ainong Li (Member, IEEE) received the Ph.D. degree in physical geography from the University of Chinese Academy of Sciences, Beijing, China, in 2007.

He is currently a Professor with the University of Chinese Academy of Sciences. He has authored/coauthored more than 200 peer-reviewed journal papers and four monographs so far. Currently, he is undertaking four national research projects. His research interests include remote sensing images preprocessing and spatial-temporal-spectral fusion, ecological environment assessment, biodiversity conservation, disaster risk assessment, and emergency response investigation. A list of his research works can be found at http://sourcedb.imde.cas.cn/yw/rc/fas/201708/t20170817_4849614.html.



Jinhu Bian (Member, IEEE) received the M.S. and Ph.D. degrees in physical geography from the Institute of Mountain Hazards and Environment (IMHE), Chinese Academy of Sciences, Chengdu, China, in 2010 and 2017, respectively.

From 2016 to 2017, he was a Visiting Scholar with the University of Maryland, College Park, MD, USA. He is currently a Research Associate Professor with IMHE, Chinese Academy of Sciences. His research interests include spatiotemporal fusion methods and their integrated applications in mountainous areas. A list of his research works can be found at https://www.researchgate.net/profile/Jinhu_Bian.

Guangbin Lei received the B.S. degree in geography information system from the Sichuan Normal University, Chengdu, China, in 2008, and the Ph.D. degree in physical geography (mountain quantitative remote sensing) from the Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China, in 2016.

He is currently a Research Assistant Professor with the Institute of Mountain Hazards and Environment, Chinese Academy of Sciences. His current research focuses on land use/land cover change mapping. A list of his research works can be found at https://www.researchgate.net/profile/Guangbin_Lei.