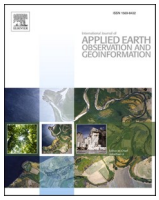
International Journal of Applied Earth Observations and Geoinformation 109 (2022) 102787

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Toward accurate mapping of 30-m time-series global impervious surface area (GISA)

Xin Huang [a,b], Yihong Song [a], Jie Yang [a,*], Wenrui Wang [a], Huiqun Ren [a], Mengjie Dong [a], Yujin Feng [a], Haidan Yin [a], Jiayi Li [a]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, PR China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

ARTICLE INFO

Keywords:
Impervious surface
Landsat
Urban expansion
Global
Google Earth Engine

ABSTRACT

As an important indicator of urbanization, accurate and long-term global artificial impervious surface area (ISA) monitoring is vital to biodiversity, water quality assessment, urban heat island, etc. However, the existing several 30-m global ISA datasets exhibit large inconsistencies, due to their differences in training samples and mapping methods. In this context, we proposed a global ISA mapping method by considering the inconsistency of the existing products, based on which we further generated a new 30-m global ISA dataset (GISA 2.0). Specifically, we divided the mapping area into A-Grids and M-Grids in terms of their consistency degrees. An automatic mapping method was proposed for classifying the A-Grids, by extracting training samples from the consistent regions of existing datasets. In the case of M-Grids, where the existing ISA datasets showed large inconsistency, we proposed to add manually interpreted samples, to strengthen the classification in these areas. We randomly selected over 120,000 test samples from 207 global grids. The results showed that GISA 2.0 achieved a F1-score of 0.935, better than GISA 1.0 (0.893), GAIA (0.721) and GAUD (0.809). A further assessment based on 118,822 ZY-3 test samples indicated that the overall accuracy and F1-score of GISA 2.0 outperformed the existing ones. GISA 2.0 will be freely available at irsip.whu.edu.cn/resources/resources_en_v2.php.

1. Introduction

Impervious surfaces area (ISA) generally refers to artificial structures such as parking lots, roads, roofs, open spaces, etc. As an important indicator of urban sprawl and ecological quality, ISA monitoring can not only provide information on human activities but can be also used for ecological management, such as water quality assessment and storm-water taxation (Luo and Lau, 2019; Song et al., 2020). In addition, it plays an important role in land use and cover change studies, and is valuable for urban infrastructure construction and sustainable development. Therefore, it is vital to document accurate ISA extent and dynamics to achieve a range of urban studies for a better understanding of anthropogenic implications associated with urbanization (Friedl et al., 2010; J. Gong et al., 2020).

During the 70s and 80s in the last century, ISA mapping was mainly conducted via aerial observation platform. With the development of satellite remote sensing, satellite imagery has been widely used to estimate ISA at different scales, such as Moderate Resolution Imaging Spectroradiometer (MODIS), Defence Meteorological Satellite Program-Operational Line Scan System (DMSP-OLS) Night Light (NTL) data. The commonly used methods for ISA estimation from MODIS imagery included machine learning and spectral mixture analysis. NTL data, on the other hand, can effectively reflect socio-economic development, and hence was also widely used for estimating impervious surface. For instance, thresholding algorithms were employed to extract ISA from NTL data (Lu et al., 2014), but it was difficult to determine the optimal thresholds and deal with mixed pixels. Meanwhile, due to the saturation phenomenon, lack of onboard radiometric calibration, and blooming effect, ISA mapping via DMSP-OLS data may result in overestimation of ISA extent (Zhuo et al., 2015). To improve the accuracy of ISA mapping, existing literature has investigated the incorporation of NTL and MODIS data, through the construction of the urban or impervious index, e.g., NUACI (Shao and Liu, 2014), MNDISI (Liu et al., 2013), or by considering auxiliary census data (Lee et al., 2018). However, due to their relatively coarse spatial resolution, small and scattered villages in sub-urban or rural regions cannot be well characterized via MODIS or DMSP-

* Corresponding author.
E-mail address: yang9tn@163.com (J. Yang).

https://doi.org/10.1016/j.jag.2022.102787
Received 5 February 2022; Received in revised form 12 April 2022; Accepted 13 April 2022
Available online 27 April 2022

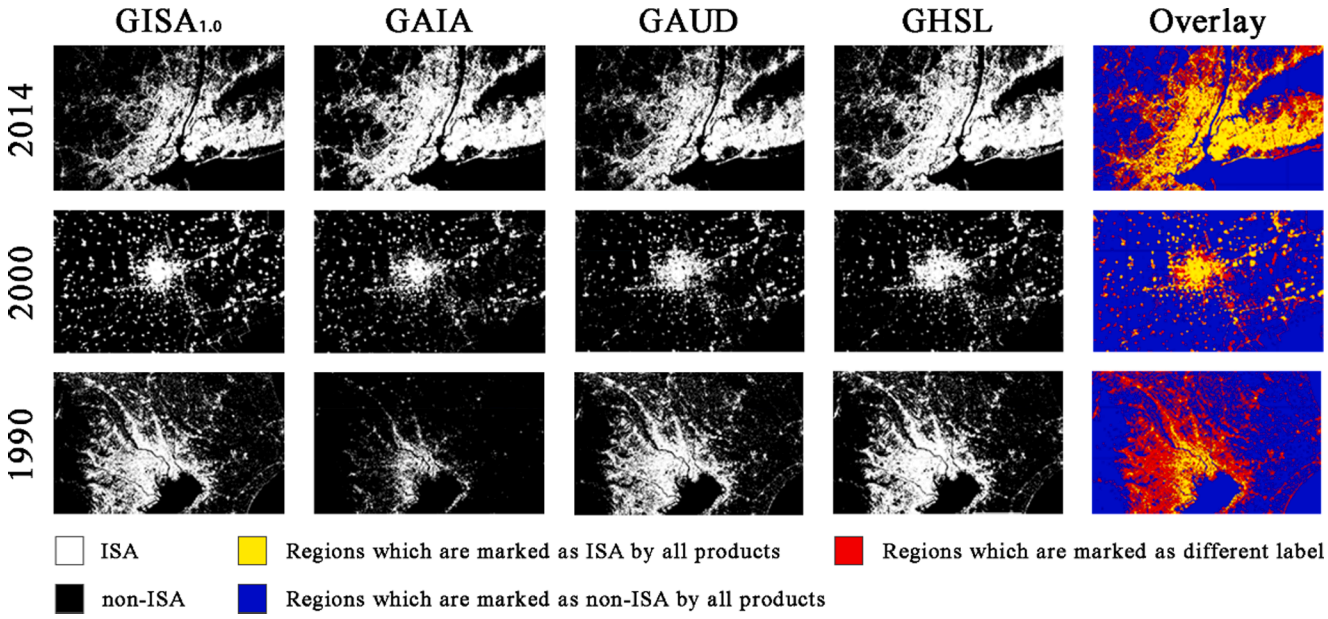


Fig. 1. Comparison of the existing global ISA datasets.

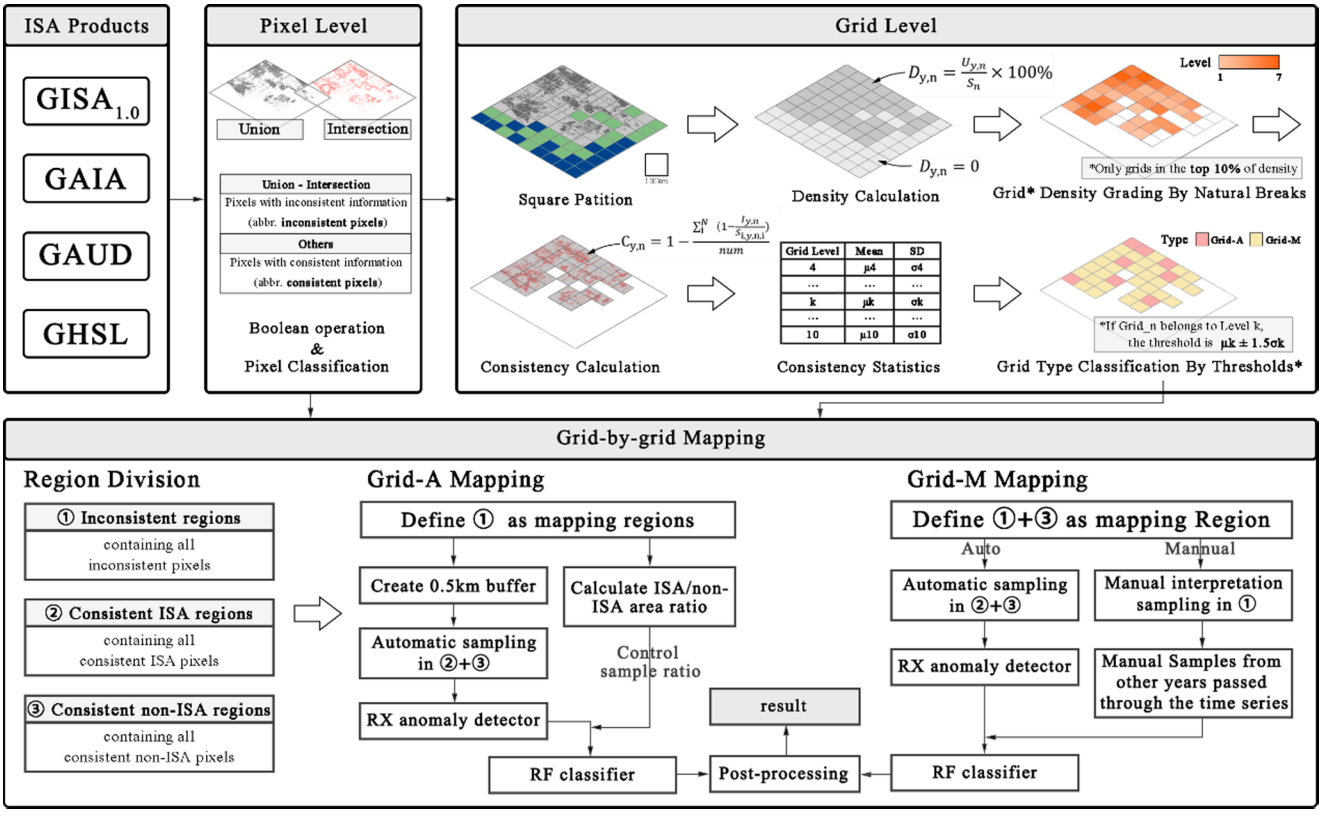


Fig. 2. The flowchart for the production of GISA 2.0.

OLS data.

Landsat data are more capable of detecting ISA at various scales (local, regional, or global), due to its wide spatial coverage, high spatial resolution, and fine revisit frequency. Traditional methods, such as spectral indices, spectral mixture analysis, machine learning, and regression analysis, have been adopted to perform pixel-wise or sub-pixel ISA mapping (Kaspersen et al., 2015; Sun et al., 2011; Wu and Murray, 2003). Likewise, Landsat images were often combined with multi-sources data to improve the ISA mapping performance, such as synthetic aperture radar (SAR) data that is sensitive to urban geometrical structures (Zhang et al., 2014). At the sub-pixel scale, Xu et al., (2018) proposed an improved LSMA (Linear Spectral Mixture Analysis) method by combining the Normalized Difference Building Index (NDBI) and the Normalized Difference Vegetation Index (NDVI), resulting in better mapping accuracy.

High-resolution (HR) images, captured by QuickBird, IKONOS,

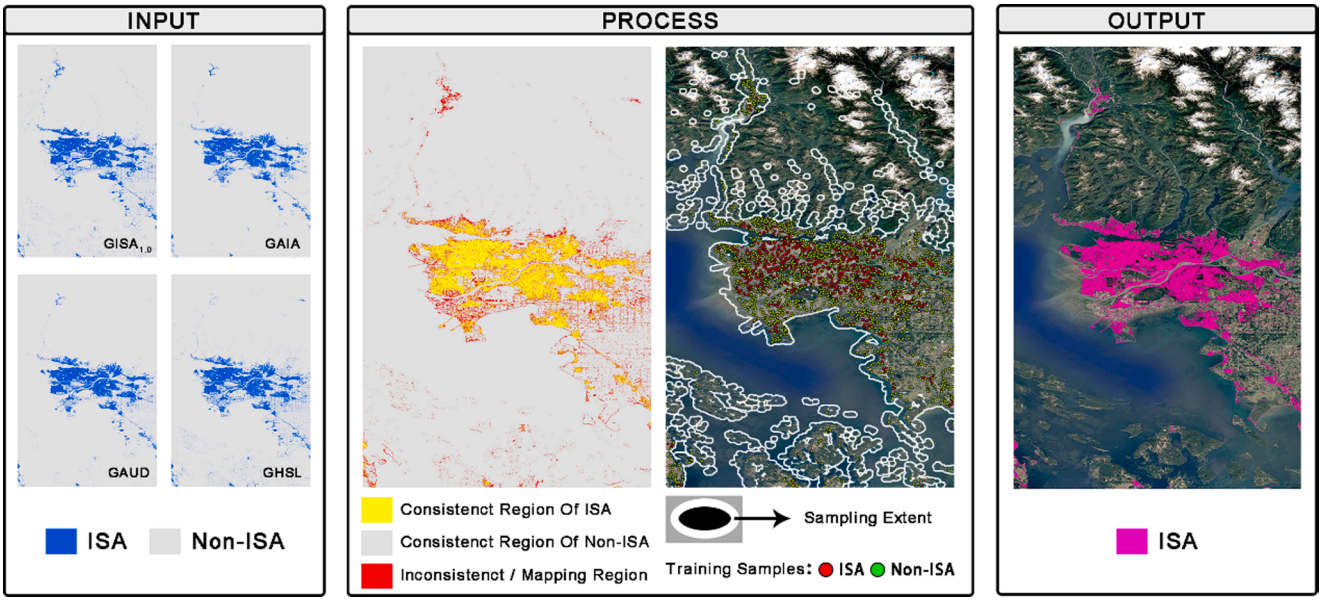


Fig. 3. Classification of A-Grids (taking Vancouver, Canada as an example).

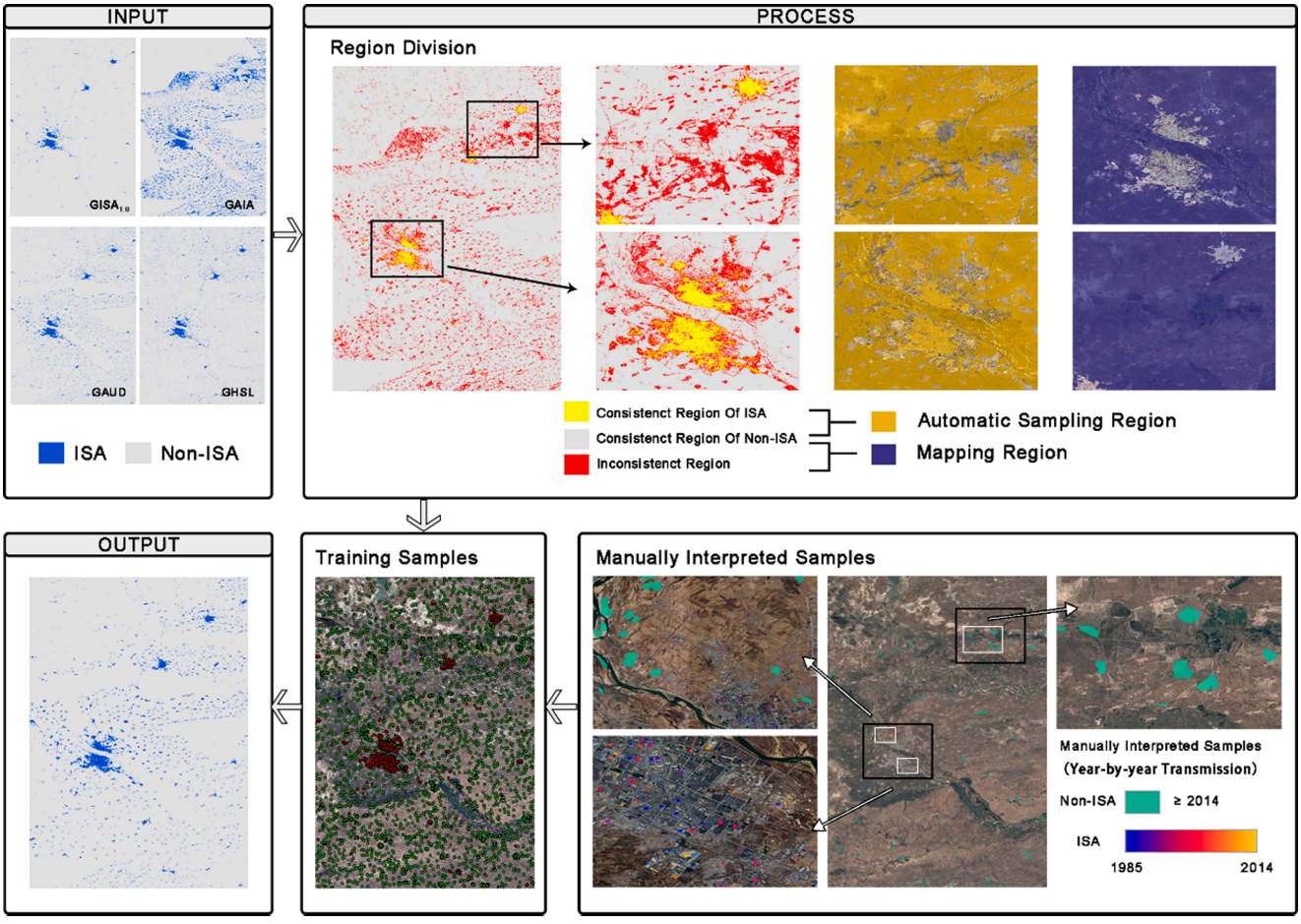


Fig. 4. Process of GISA 2.0 mapping for M-Grids (taking Songyuan, China, as an example).

WorldView, ZY-3, etc. are potential for more precise extraction of ISA (Hamedianfar and Shafri, 2016). However, owing to the limitations in spatial coverage, revisit cycles, data cost, and processing burden, HR imagery was mainly used for ISA mapping at local scales. Notably, with the development of computing and storage capabilities, 10-m global ISA mapping has become achievable, such as the global human settlement layers (GUP) generated by the German Aerospace Center with a spatial resolution of 12 m (Esch et al., 2013), and the 10-m Finer Resolution

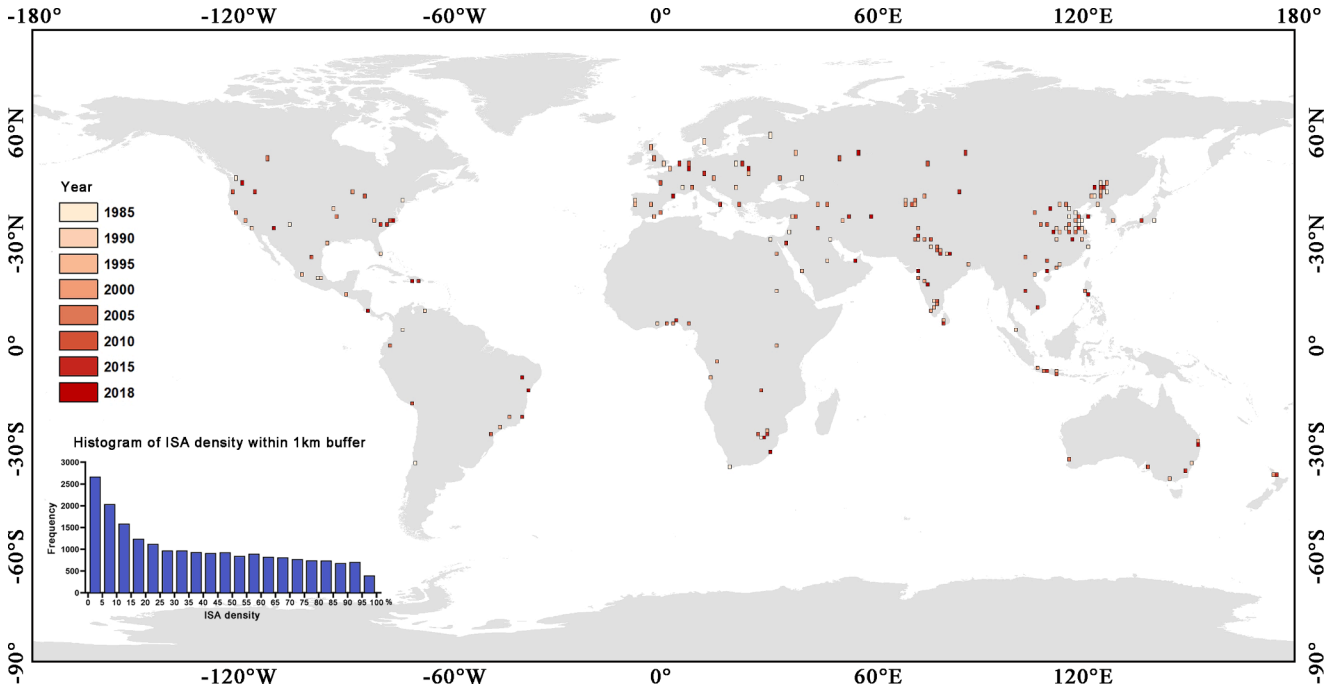


Fig. 5. Global distribution of test samples and ISA density for ISA test samples.

Observation and Monitoring of Global Land Cover (FROM_GLC10) obtained from Sentinel-2 data (Gong et al., 2019). However, it should be pointed out that, although high-resolution ISA datasets provided us with more spatial details, they generally lacked time-series updating, and suffered from image mosaic, cloud contamination, as well as other data quality issues.

In general, medium-resolution Landsat imagery remains the most appropriate data source for large-scale ISA dynamic monitoring. Meanwhile, the appearance and development of the Google Earth Engine (GEE) platform (Gorelick et al., 2017) greatly promoted large-scale ISA mapping, by courtesy of its strong ability of cloud computing and data storage. Till now, several Landsat-derived global ISA datasets have been produced, including: (1) Global Human Settlement Layer (GHSL) (Pesaresi et al., 2016), which was constructed by the Joint Research Centre of European Commission using symbolic machine learning methods with training data extracted from coarse resolution dataset; (2) Global Urban Dynamics Map (GAUD, 1985–2015) (Liu et al., 2020), where urban dynamics were extracted using temporal segmentation from annual Normalized Urban Area Convergence Index (NUACI); (3) Global Artificial Impervious Area (GAIA, 1985–2018) (P. Gong et al., 2020), which was extracted using an "Exclusion-Inclusion" algorithm; (4) Global ISA dataset (GISA 1.0, 1972–2019) (Huang et al., 2021), where training samples were extracted from existing products and results were obtained by adaptive RF classifiers. In particular, GISA 1.0 documented the global ISA map before 1985 and thus possessed a relatively longer time span (1972 to 2019) than other global datasets.

Although these existing datasets showed consistency to some degree, they exhibited a large amount of differences. For instance, these datasets were more consistent in the city cores, but showed relatively large disagreements over suburban or rural regions (see Fig. 1). This can be attributed to their differences in mapping methods and training samples. Therefore, the main objective of this paper was to generate a more accurate and consistent global ISA dataset (written as GISA 2.0) by focusing on the inconsistent regions between these existing global ISA datasets. To this end, in this study, we proposed an integrated classification approach for global ISA mapping based on the spatial–temporal consistency of the existing products. Sepecifically, the novelty of this study includes:

i) A new method that aims to integrate the existing global ISA datasets is proposed for improving the accuracy of global ISA mapping. The proposed method divides the mapping grids into M-grids and A-grids, in terms of the disagreement degree between the existing ISA datasets. Different mapping strategies are proposed for M- and A-grids, respectively.

ii) For the A-grids, an automatic mapping method is proposed, by extracting and refining the training samples from the consistent regions of the existing datasets.

iii) For the M-grids, a different mapping strategy is proposed, by integrating both automatically and manually generated samples. In the meantime, a time-series sample transfer method is designed, to make full use of the manually-interpreted samples.

The rest of this paper was organized as follows. We introduced the method framework in Section 2. Results were reported by visual comparison and accuracy assessment in Section 3, followed by discussions in Section 4. Finally, Section 5 concluded the paper. The GISA 2.0 is freely available at: irsip.whu.edu.cn/resources/resources_en_v2.php.

2. Methodology

The flowchart for the production of GISA 2.0 is shown in Fig. 2. We first extracted consistent regions from existing annual or multi-temporal global datasets and divided the global terrestrial surface into grids based on the degree of consistency. The mapping result was generated for each grid and the classification strategies were different in terms of the consistency degrees. Finally, GISA 2.0 was assessed by independent test samples and compared with existing datasets.

2.1. Integrated classification approach for global ISA mapping

2.1.1. Consistency of existing datasets

The core idea of this study is to generate a new global ISA dataset (GISA 2.0) by focusing on the disagreement (or inconsistent) regions of existing datasets, since the consistent regions often have high classification certainty and accuracy, but the inconsistent regions often show classification uncertainty and errors. To this end, we obtained the inconsistent grids in terms of the existing global ISA datasets and

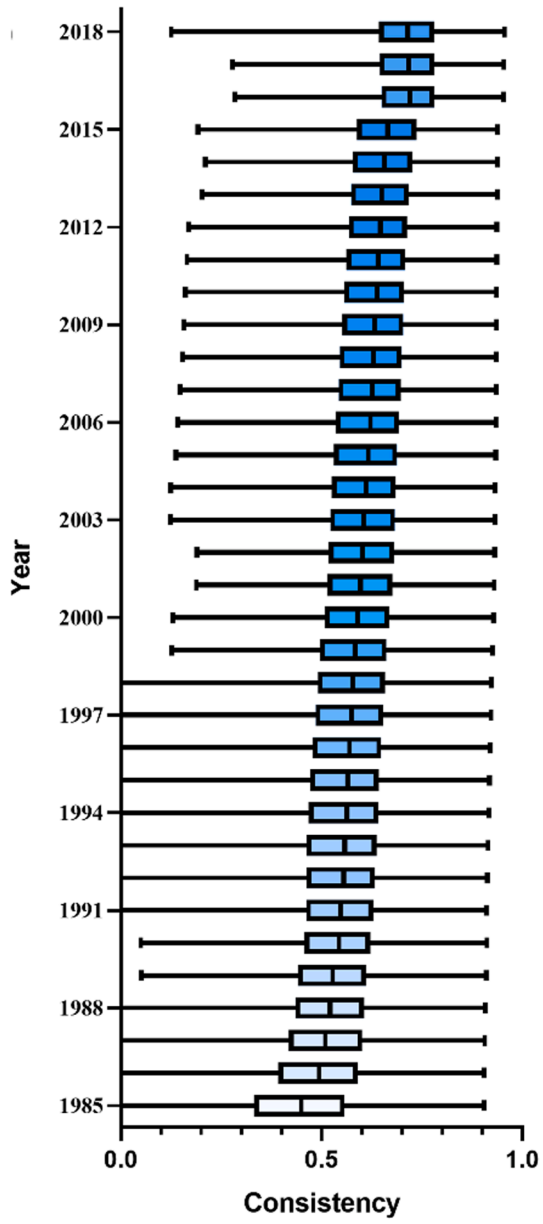


Fig. 6. Consistency between existing global ISA datasets during 1985–2018.

proposed two different mapping methods for consistent and inconsistent grids, respectively. Details are introduced as the following steps:

Step 1: Given the 185 km swath width of a Landsat image, we divided the global terrestrial surface with square grids of 100 km (about half of the swath) side length. The 100 km square grid was much smaller than those employed by existing literature (e.g. 3.5° grid in GAIA), which can better characterize local geographic landscapes. Specifically, for each year, we first calculated the ISA density within each grid as:

$$D_{y,n} = \frac{U_{y,n}}{S_n} \times 100\% \tag{1}$$

where $U_{y,n}$ denoted the total area of ISA union of all the datasets for the grid n in year y, and S_n represented the area of the grid. The small and subtle partition (i.e., 100 km for each grid) might result in a large number of grids with low ISA density, which can significantly increase the computational cost. Therefore, in this research, we focused mainly on the top 10% of grids with the highest ISA density (1353 grids in total), which would be re-mapped and updated. In addition, the selected grids were further divided into 7 grades according to the ISA density using the Jenks natural breaks classification method (Hu and Huang, 2019).

Step 2: The degree of consistency between existing datasets was then calculated for each grid. Specifically, we aggregated the pixel area that was labeled as ISA in all the datasets in year y ($I_{y,n}$), and also obtained the ISA area in each dataset ($S_{i,y,n}$). The consistency for multiple datasets in that year (i.e., $C_{y,n}$) was defined as:

$$C_{y,n} = 1 - \frac{\sum_i^N P_{i,y,n}}{N} \tag{2}$$

$$\text{with } P_{i,y,n} = \frac{I_{y,n}}{S_{i,y,n}} \tag{3}$$

where N referred to the number of existing datasets available in the year y. A higher value of $C_{y,n}$ (e.g., closer to 1) signified better consistency of the existing global ISA datasets within the grid.

Step 3: Finally, we classified the grids within each grade into two categories, namely A-Grids and M-Grids according to the following criteria:

$$\text{A} - \text{Grids}: \quad C_{y,n} < \mu + 1.5\sigma \tag{4}$$

$$\text{M} - \text{Grids}: \quad C_{y,n} \geq \mu + 1.5\sigma \tag{5}$$

where μ and σ denoted the mean and standard deviation of the consistency degree for the considered grids, respectively. Grids with consistency value larger than the threshold were determined as M-Grids, which exhibited higher disagreement among the existing datasets and were therefore difficult to classify. For the M-Grids, we strengthened their classification by introducing a large number of manually-interpreted samples. On the other hand, the grids with relatively larger consistency were defined as A-Grids, where we automatically extracted training samples from the consistent regions, in order to adaptively improve the current results.

2.1.2. Automatic ISA mapping for A-grids

The accuracy and reliability of the existing datasets in the A-Grids were relatively better compared to the M-Grids, and thus ISA mapping in the A-Grids was carried out in a fully automated manner, so as to reduce the workload of sample collection.

We first determined the potential mapping regions by extracting all pixels with inconsistent labels in the grids (e.g., red color in Fig. 3). Specifically, the ISA and non-ISA training samples were randomly selected within a 500-m buffer of the consistent regions among the existing global datasets (white rings in Fig. 3). Finally, the possible outliers within training samples were filtered out via the temporal-spatial-spectral method adopted in the GISA 1.0 (Huang et al., 2021). In A-Grids, the training samples were extracted from the consistent regions of multiple existing datasets, and in this way, the workload of manual sampling can be effectively reduced. This is necessary for the global-scale mapping, and also guarantees that our new products (written as GISA 2.0) are synthesized and updated based on the existing ISA datasets.

Classification results were obtained by the random forest classifiers, and the input features consisted of the spectral channels, spectral index, time-series and phenological information, elevation, and slope calculated from the digital elevation model (DEM). Please notice that the classifiers as well as the features used in GISA 2.0 were the same as those in GISA 1.0 (Huang et al., 2021).

2.1.3. ISA mapping for M-Grids

Different from A-Grids, in the M-Grids, there existed a large number of omission and commission errors in the existing datasets, and their consistent regions were very small. Therefore, the mapping reliability of the existing datasets in the M-Grids was low. Considering this, the mapping strategy we proposed for M-Grids is different from A-Grids (Fig. 4). Firstly, we considered the inconsistent pixels (same as A-Grids)

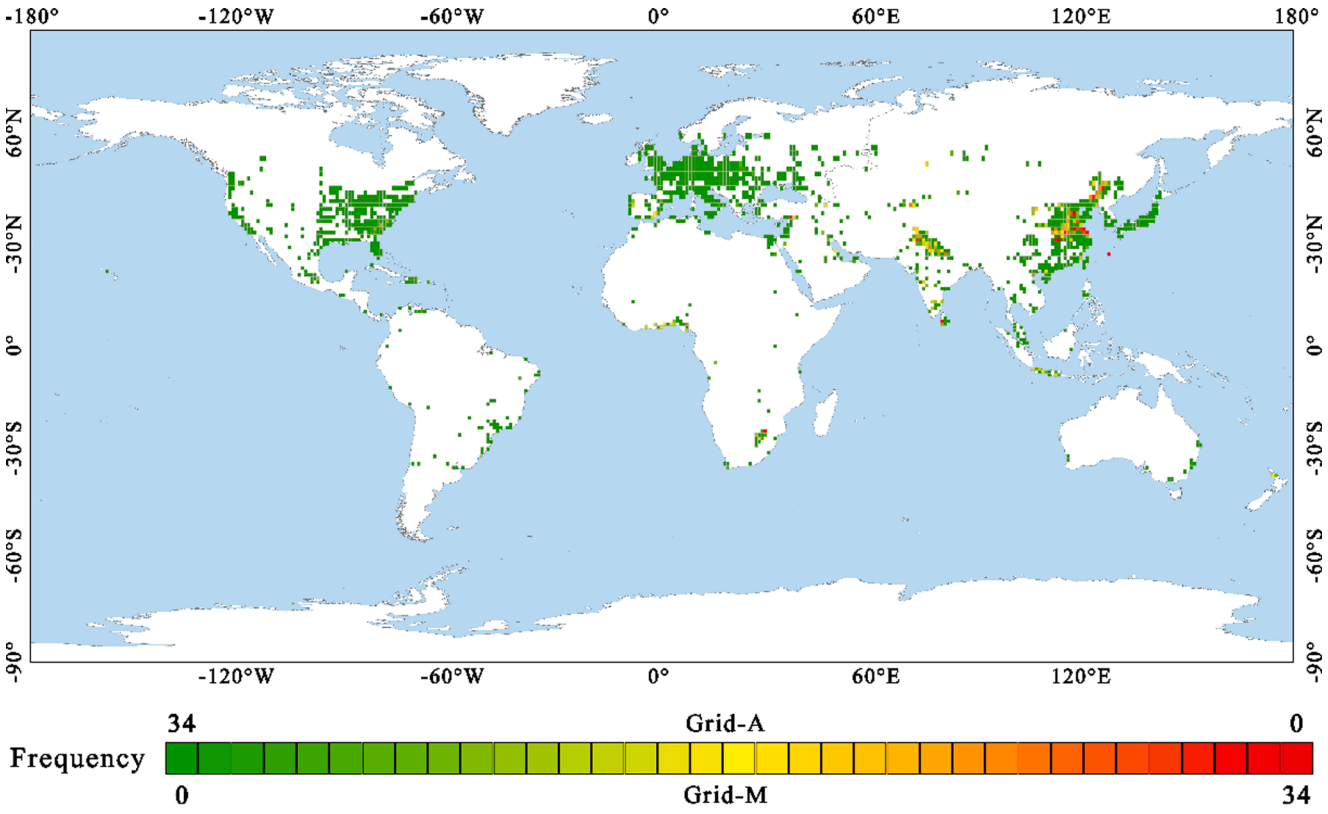


Fig. 7. Global distribution and occurrence frequency of M-Grids and A-Grids during 1985–2018.

Table 1
Accuracy assessment of GISA 2.0 compared to existing datasets based on the manually-interpreted test samples.

A-Grids & M-Grids							
GISA 2.0	ISA	Non-ISA	UA (%)	**GISA 1.0**	ISA	Non-ISA	UA (%)
ISA	16,384	577	96.60	ISA	15,555	1143	93.15
Non-ISA	1715	89,914	98.13	Non-ISA	2544	89,348	97.23
PA (%)	90.52	99.36		PA (%)	85.94	98.73	
OA (%)	**97.89**	**F1-score of ISA**	**0.935**	**OA (%)**	**96.60**	**F1-score of ISA**	**0.894**
GAIA	ISA	Non-ISA	UA (%)	**GAUD**	ISA	Non-ISA	UA (%)
ISA	11,253	1882	85.67	ISA	13,220	1353	90.72
Non-ISA	6846	88,609	92.82	Non-ISA	4879	89,138	94.81
PA (%)	62.17	97.92		PA (%)	73.04	98.50	
OA (%)	**91.96**	**F1-score of ISA**	**0.721**	**OA (%)**	**94.26**	**F1-score of ISA**	**0.809**
A-Grids							
GISA 2.0	ISA	Non-ISA	UA (%)	**GISA 1.0**	ISA	Non-ISA	UA (%)
ISA	8475	313	96.44	ISA	8171	520	94.02
Non-ISA	625	45,187	98.64	Non-ISA	929	44,980	97.98
PA (%)	93.13	99.31		PA (%)	89.79	98.86	
OA (%)	**98.28**	**F1-score of ISA**	**0.948**	**OA (%)**	**97.35**	**F1-score of ISA**	**0.919**
GAIA	ISA	Non-ISA	**UA (%)**	**GAUD**	ISA	Non-ISA	**UA (%)**
ISA	6458	689	**90.36**	ISA	7760	964	**88.95**
Non-ISA	2642	44,811	94.43	Non-ISA	1340	44,536	97.08
PA (%)	**70.97**	98.49		**PA (%)**	**85.27**	97.88	
OA (%)	**93.90**	**F1-score of ISA**	**0.795**	**OA (%)**	**95.78**	**F1-score of ISA**	**0.870**
M-Grids							
GISA 2.0	ISA	Non-ISA	UA (%)	**GISA 1.0**	ISA	Non-ISA	UA (%)
ISA	7909	264	96.77	ISA	7384	623	92.22
Non-ISA	1090	44,727	97.62	Non-ISA	1615	44,368	96.49
PA (%)	87.89	99.41		PA (%)	82.05	98.62	
OA (%)	**97.49**	**F1-score of ISA**	**0.921**	**OA (%)**	**95.85**	**F1-score of ISA**	**0.868**
GAIA	ISA	Non-ISA	UA (%)	**GAUD**	ISA	Non-ISA	UA (%)
ISA	4795	1193	80.08	ISA	5460	389	93.35
Non-ISA	4204	43,798	91.24	Non-ISA	3539	44,602	92.65
PA (%)	53.28	97.35		PA (%)	60.67	99.14	
OA (%)	**90.00**	**F1-score of ISA**	**0.640**	**OA (%)**	**92.72**	**F1-score of ISA**	**0.735**

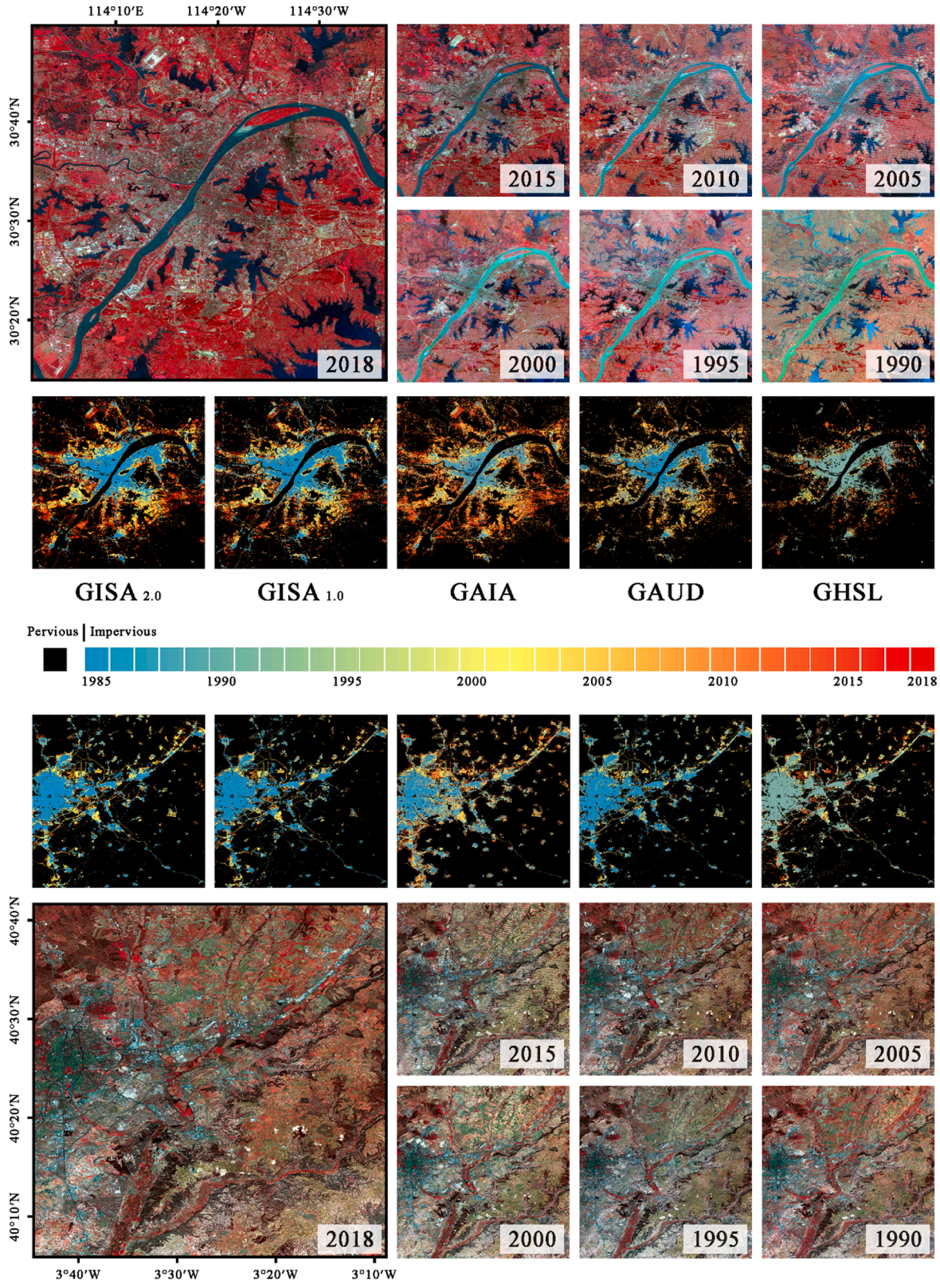


Fig. 8. Comparison between different global ISA datasets (taking Wuhan and Madrid as examples).

and Non-ISA region in M-Grids as mapping regions (blue color in Fig. 4). In order to strengthen the mapping accuracy of M-Grids, we employed both automatically-generated (see Section 2.1.2) and manually-interpreted training samples. The manually added samples were mainly located at: i) regions with ISA omission, and ii) inconsistent regions among existing ISA datasets, in order to focus on the errors existed in the current ISA datasets. In terms of the sample size, a total of 50 sample units (polygons) were manually added to each M-grid, with an average 10 pixels for each unit (see Section 4.3 for details). In particular, the manually-interpreted samples should be located in the

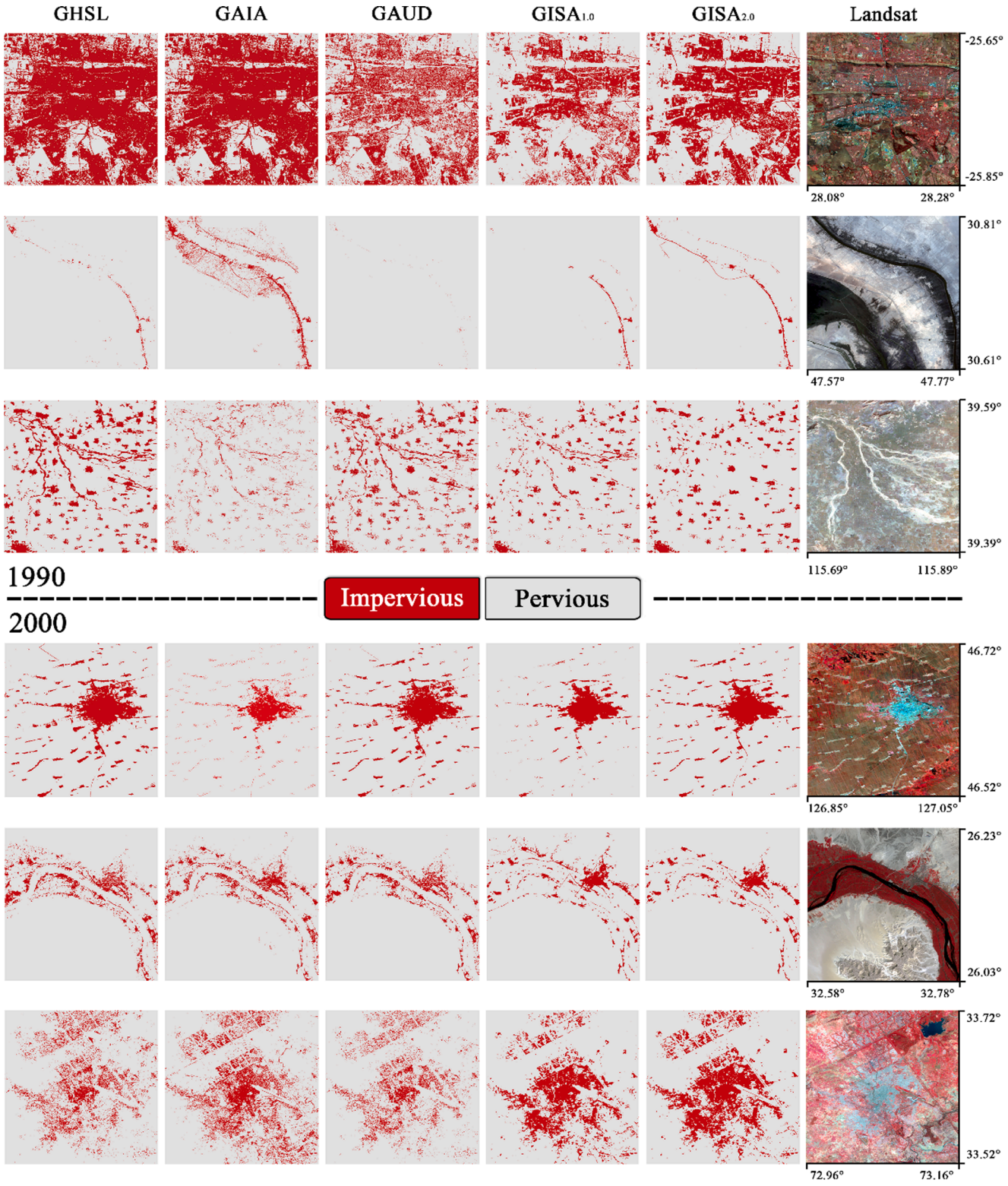


Fig. 9. Comparison of GISA 2.0 with GHSL, GAUD, GAIA and GISA 1.0 in the M-Grids. The Landsat 5 images are shown with the false-color combination (band 4, 3, and 2).

inconsistent regions to correct the classification errors and reduce the uncertainties.

Given the high reliability of manually-interpreted samples, we transferred them to previous or subsequent time periods according to the irreversibility assumption, i.e., the transition from ISA to Non-ISA rarely happens (P. Gong et al., 2020; Huang et al., 2021; Li et al., 2015). Specifically, taking the year *y* for instance, the ISA samples generated by visual interpretation in M-Grids were used for ISA mapping in the year *y* and its subsequent years, while visually-interpreted Non-ISA samples were used in the year *y* and its previous years. Note that the transferred samples should be outside the consistent regions among the existing ISA datasets. In this way, we tried to make full use of visually-interpreted

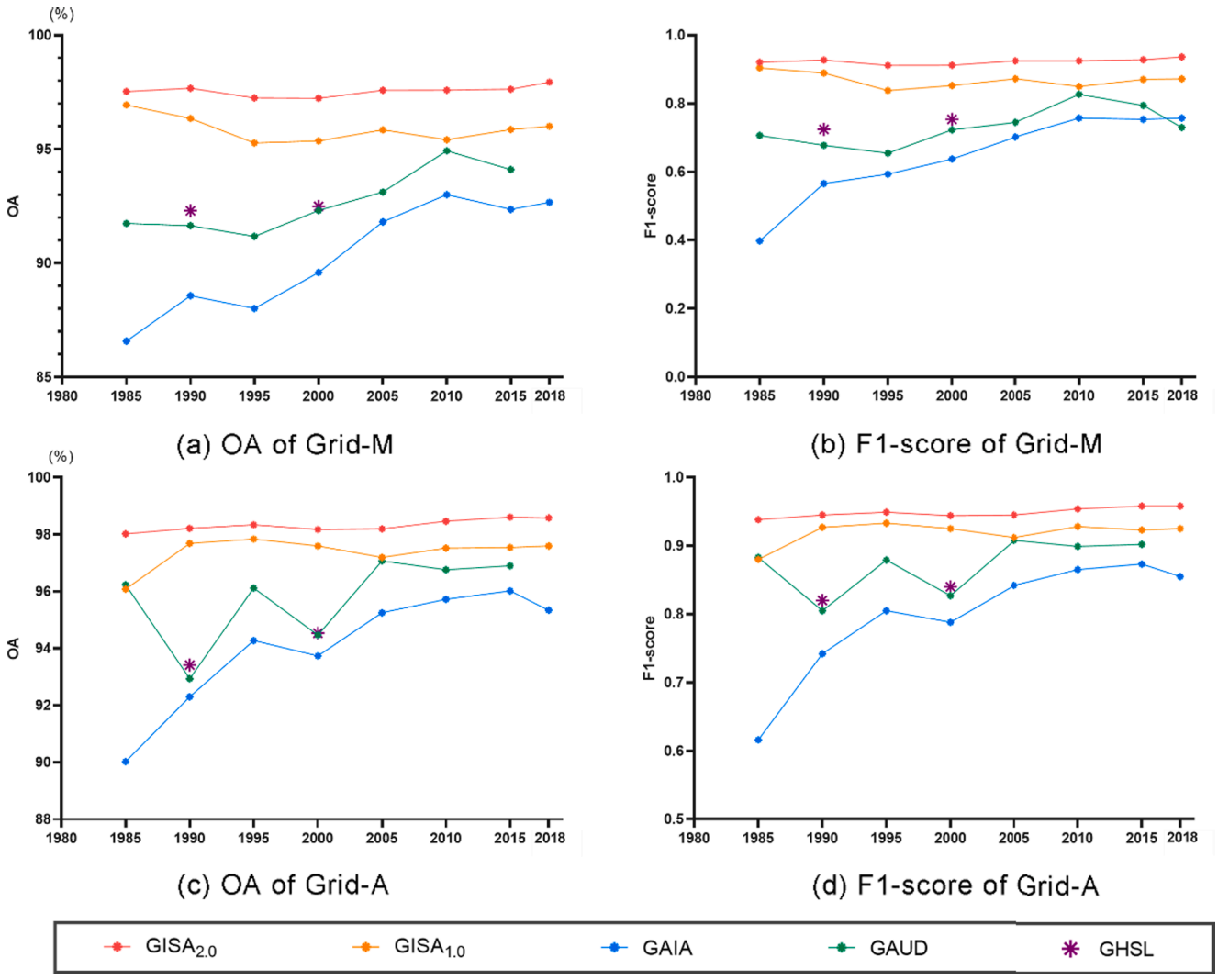


Fig. 10. OA and F1-score of ISA between GISA 2.0 and existing datasets in A- and M-Grids.

Table 2
Accuracy assessment of GISA 2.0 compared to existing datasets based on the ZY-3 test samples.

ZY-3 test samples							
GISA 2.0	ISA	Non-ISA	UA (%)	**GISA 1.0**	ISA	Non-ISA	UA (%)
ISA	34,308	6421	84.23	ISA	32,541	5630	85.25
Non-ISA	5169	72,924	93.38	Non-ISA	6936	73,715	91.41
PA (%)	86.91	91.91		PA (%)	82.43	92.91	
OA (%)	**90.24**	**F1-score of ISA**	**0.855**	**OA (%)**	**89.42**	**F1-score of ISA**	**0.838**
GAIA	ISA	Non-ISA	UA (%)	**GAUD**	ISA	Non-ISA	UA (%)
ISA	32,816	7710	80.98	ISA	32,765	5755	85.06
Non-ISA	6661	71,635	91.49	Non-ISA	6712	73,590	91.64
PA (%)	83.13	90.29		PA (%)	83.01	92.75	
OA (%)	**87.91**	**F1-score of ISA**	**0.821**	**OA (%)**	**89.51**	**F1-score of ISA**	**0.809**

samples through the time-series transfer. The classification results in the M-Grids were then acquired via the random forest classifier and multi-source features, as described in Section 2.1.2.

2.2. Accuracy assessment

The accuracy of GISA 2.0 was assessed by 124,190 samples that are independent from training samples. As demonstrated by Stehman and Foody (2019), we used cluster sampling to determine the location of these test samples, involving 207 randomly selected grids. The sample collection was carried out at five-year intervals (i.e. 1985, 1990, 1995, 2000, 2005, 2010, 2015, and 2018). In each period, the grids (A-Grids and M-Grids) used for accuracy assessment were randomly selected. In particular, the selection of grids on each continent was stratified by its population density. For example, A-Grids for Africa in 2015 were first sorted by the population density, and a grid with its population density above the median and another grid with its population density below the median were randomly selected. In such a way, samples from different urban sizes and densities were selected for validation, but not only from big urban areas. Subsequently, within each grid, 100 ISA and 500 Non-ISA samples were visually interpreted, by referring to the Google Earth and Landsat images (Huang et al., 2021). The spatial distribution of the test samples and ISA density around the ISA test samples (1 km buffer) was shown in Fig. 5. It can be seen that the test samples involved not only high-density ISA samples in large cities, but a large number of low-

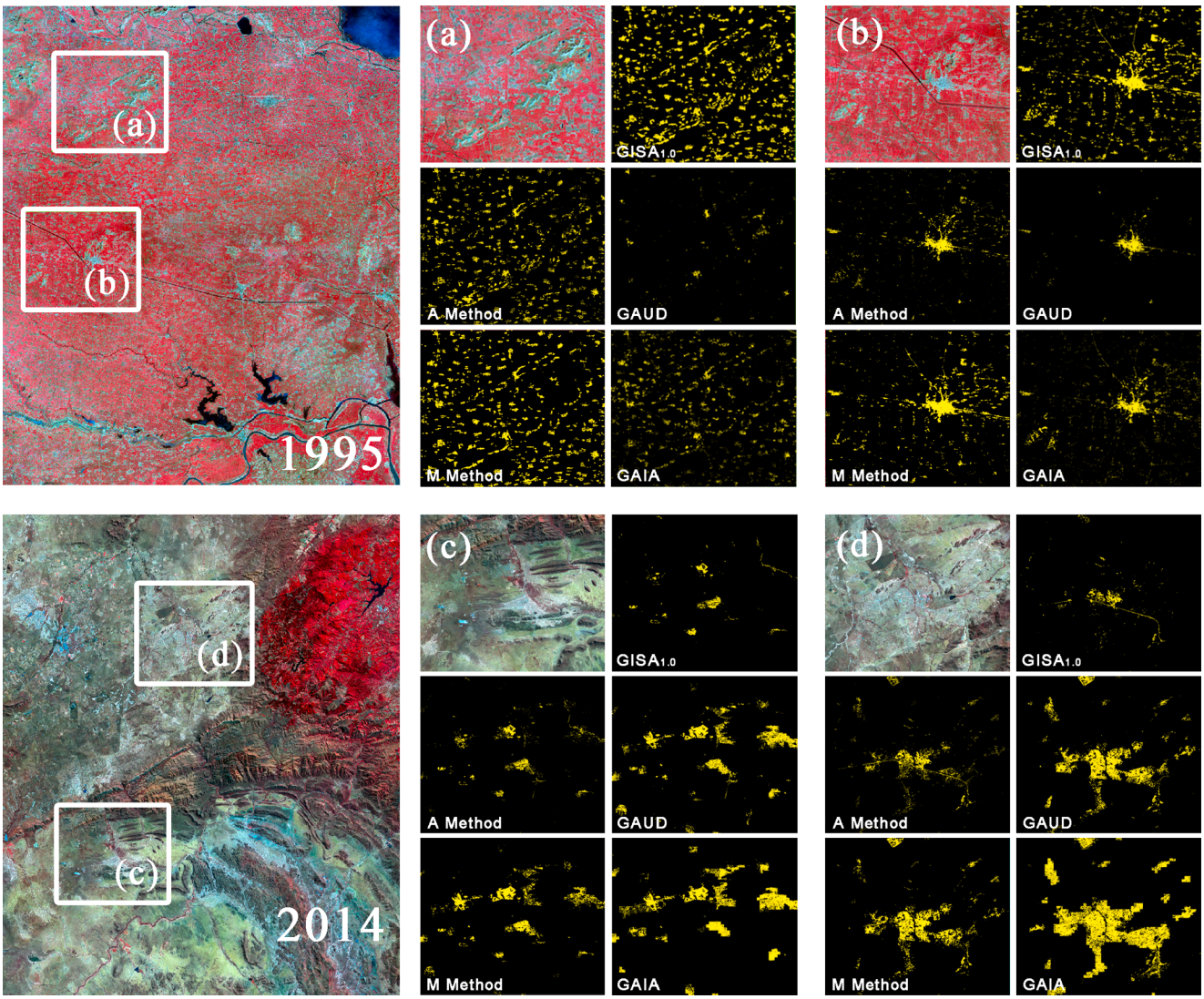


Fig. 11. Comparison of M-Method and A-Method in two M-Grids in northern China and southern Africa, respectively, as examples, in order to demonstrate the effects of our proposed mapping strategies.

density samples in suburban and rural regions. Moreover, the distance between samples should be larger than 1 km to guarantee their spatial independence. In such way, 124,190 global test samples were generated. The accuracy of GISA 2.0 was assessed by overall accuracy (OA), User's Accuracy (UA), Producer's Accuracy (PA), and F1-score. The F1-score indicates the balance between UA and PA.

Moreover, we also employed over 110,000 test samples extracted from the ZiYuan-3 (ZY-3) global built-up dataset for the accuracy assessment of GISA 2.0. Liu et al., (2019) proposed a multi-angle built-up index to extract built-up areas from high resolution (2 m) ZY-3 data in 45 global cities. This built-up dataset was used as a test sample set in this research, due to its high overall accuracy and high spatial resolution (2 m). We defined an ISA test sample if more than 80% of its area (30 m × 30 m) was identified as built-up, while a Non-ISA sample was chosen when no built-up pixel was found within it. Moreover, the distance between test samples should be greater than 300 m to ensure their spatial independence, and each sample was visually inspected in terms of the high-resolution ZY-3 images. In this way, we obtained 39,477 ISA and 79,345 Non-ISA test samples.

3. Results

3.1. Consistency analysis of existing global datasets

In this study, we analyzed the spatial–temporal consistency between existing global ISA datasets. As seen from Fig. 6, generally, the consistency between the existing datasets showed an increasing trend during 1985–2018. This phenomenon can be attributed to the lower image quality and data availability in the early stages, and along with the launch of Landsat-7 and Landsat-8, the consistency between existing ISA datasets was raised gradually.

Fig. 7 illustrates the global distribution of M-Grids and A-Grids, and the colors of the grids represent their occurrence frequency during 1985 to 2018. In general, it can be seen that the inconsistent regions are distributed widely over the world, including both rapidly urbanized areas, e.g., Asia, and developed areas, e.g., North America, and Europe. In terms of grid types, M-Grids are more frequent in Asia and Africa, such as North China Plain and Gulf of Guinea (Fig. 7), whereas A-Grids are more common in North America and Europe. The M-Grids exhibit large inconsistency among existing datasets and should be focused on in our research.

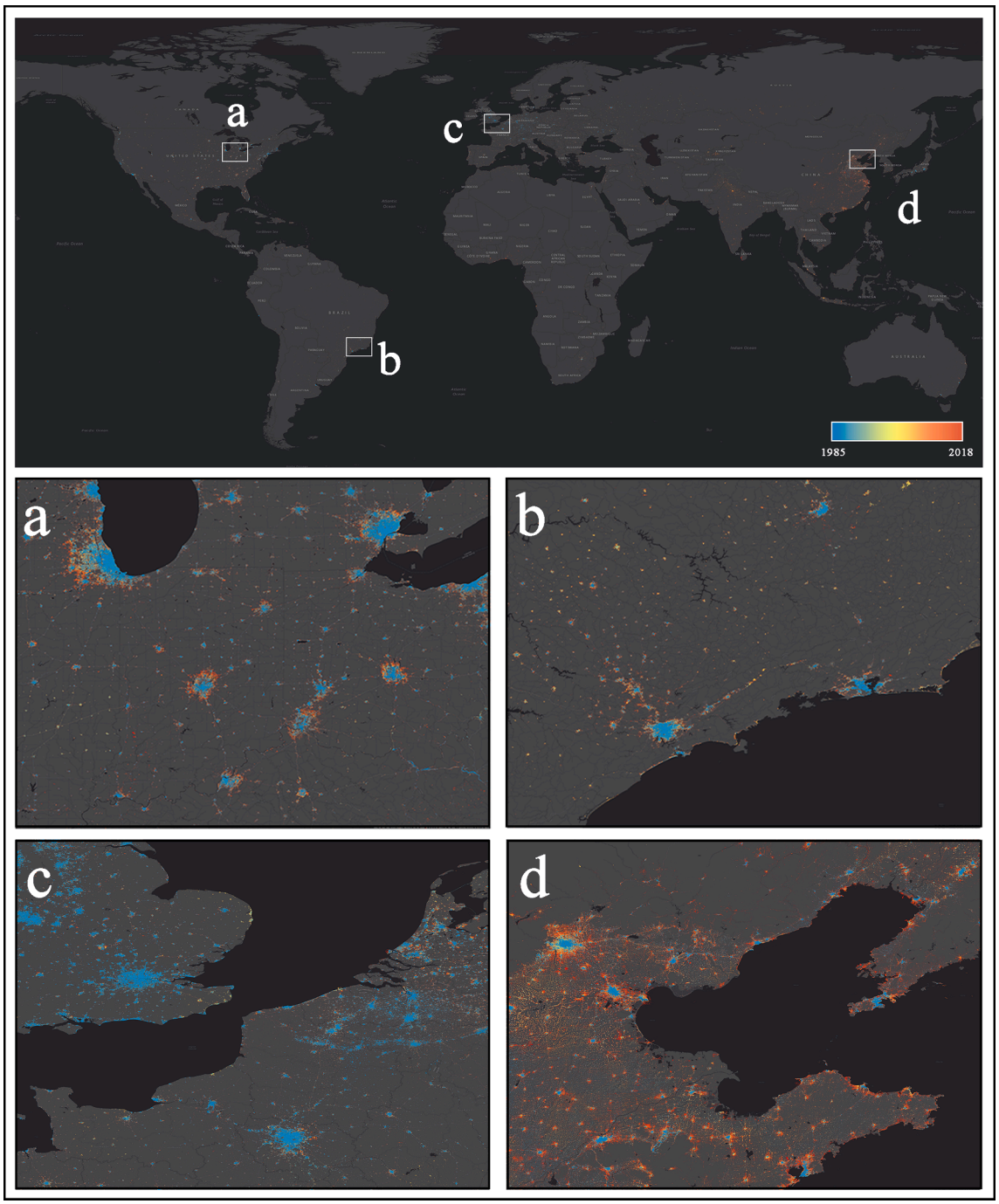


Fig. 12. Global ISA expansion observed from GISA 2.0.

Table 3
Accuracy comparisons of different mapping method.

Mapping method	OA	F1-score of ISA
GISA 1.0	95.726%	0.8560
GISA 2.0 A-Method	96.672%	0.8861
GISA 2.0 M-Method	97.472%	0.9204

3.2. Accuracy assessment of GISA 2.0

3.2.1. Accuracy assessment using visually-interpreted test samples

The accuracy assessment is shown in Table 1. GISA 2.0 achieved the highest OA of 97.89%, with an increase of 1.29%, 5.93%, and 3.63%, with respect to GISA 1.0, GAIA, and GAUD, respectively. The ISA F1-score of GISA 2.0 is 0.935, which exceeds GISA 1.0, GAIA, and GAUD by 0.041, 0.214, and 0.124, respectively. It should be noted that the UAs of ISA among these global ISA datasets are generally similar, whereas GISA 2.0 notably outperforms GISA 1.0, GAIA and GAUD, by 4.58%, 28.35%, and 17.48%, respectively, in terms of ISA PA. This suggests that GISA 2.0 suffers less from ISA omissions. From Fig. 8, taking a closer look at ISA dynamics, GISA 2.0 shows a generally similar expansion pattern with other datasets.

In the case of M-Grids, the manually-interpreted samples were introduced to improve the mapping accuracy of inconsistent regions. As can be seen in Fig. 9, GISA 2.0 was able to obtain more accurate results than the existing ones. For instance, GISA 2.0 succeeded to identify the

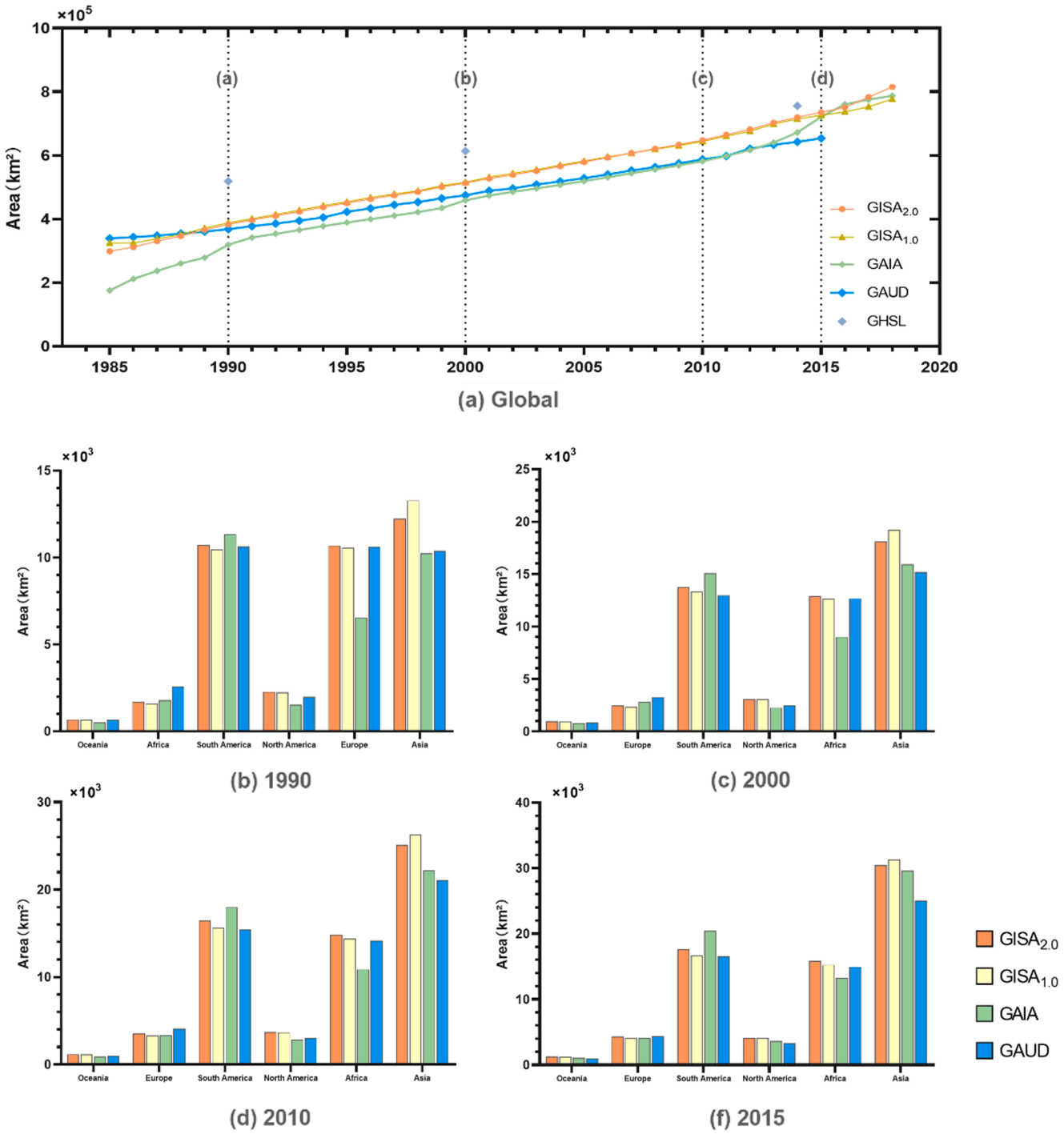


Fig. 13. Comparison of GISA 2.0 with existing global datasets at (a) global and (b)-(f) continental scale.

dry river channels as Non-ISA while the other ones showed false alarms. The accuracy assessment revealed that the overall accuracy (OA) of GISA 2.0 was 97.49% in the M-Grids, and it achieved the highest ISA F-score of 0.921, with an increase of 0.053, 0.281, and 0.186, with respect to GISA 1.0, GAIA and GAUD, respectively (Table 1). In particular, the OA of GISA 2.0 was higher than 97.5% in most of years, and its F1-socres all exceeded 0.9. Compared with other datasets, the accuracy of GISA2.0 was better and more stable throughout the years (Fig. 10(a)-(b)), indicating that our proposed method as well as the generated dataset achieved satisfactory results in inconsistent regions (i.e. M-Grids). This also demonstrated that the classification performance can be effectively improved by enhancing the quality and quantity of training samples from the inconsistent regions.

As for the A-Grids, by considering the efficiency of global mapping, the training samples were collected automatically from the consistent regions between the existing datasets. In terms of the test samples, GISA 2.0 exhibited the highest and more stable OA in A-Grids compared with other datasets (Fig. 10), which illustrated the effectiveness of our proposed automatic mapping approach for the A-Grids. Notably, the accuracy of ISA datasets in the early periods was relatively lower owing to the availability and quality of Landsat 5 images (P. Gong et al., 2020; Loveland and Dwyer, 2012; Pekel et al., 2016; Yang and Huang, 2021). However, it can be seen that GISA 2.0 had the highest values in OA and F1-score for 1985 (Fig. 10(c)-(d)), by strengthening the training samples

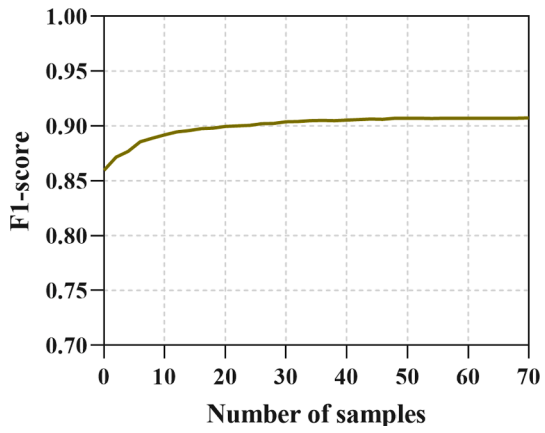


Fig. 14. The F1-Score as a function of number of samples.

from the inconsistent regions. Besides, the robustness of our method can be also reflected from the results after 2000, and the accuracy gain derived from GISA 2.0 was stable.

3.2.2. Accuracy assessment using ZY-3 test samples

In the case of the ZY-3 test samples, the overall accuracy of GISA 2.0 was 90.24%, still outperforming the existing ones (Table 2). In addition, GISA 2.0 achieved the highest ISA F1-score (0.855) against GISA 1.0 (0.838), GAUD (0.809), and GAIA (0.821). The ZY-3 samples mainly covered urban and suburban regions (Liu et al., 2019), and the better accuracy of GISA 2.0 suggested that our method and dataset effectively reduced the ISA mapping uncertainty in these areas.

4. Discussions

4.1. Effect of adding manually-interpreted samples

In order to improve the mapping accuracy of M-Grids (inconsistent regions), we added a large number of manually interpreted training samples and proposed a time-series sample transfer strategy to fully exploit these samples. In the case of the A-Grids, we automatically extracted training samples from the consistent regions among the existing global ISA datasets in view of their relatively higher reliability. In most cases, manual interpretation can generate more accurate samples but is time-consuming and labor-intensive. Therefore, the manually-interpreted samples were added for the grids with larger inconsistency, and automatically generated samples were used for the grids with less inconsistency. Theoretically, higher accuracy can be obtained when the manually-interpreted samples were used for all grids, which, however, requires enormous workload, and hence becomes impossible for global mapping. Therefore, while producing GISA 2.0, manually interpreted samples were added to the most inconsistent grids

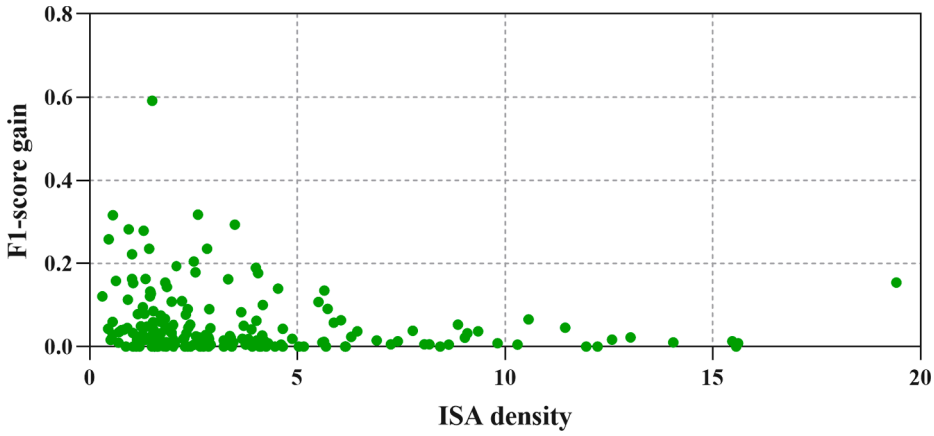


Fig. 15. Relationship between accuracy improvements and ISA density.

Table 4
Accuracy comparison between GISA 2.0 and other existing global ISA datasets.

GLC_FCS2020	ISA	Non-ISA	UA (%)	**FROM_GLC2015**	ISA	Non-ISA	UA (%)
ISA	2276	254	89.96	ISA	2007	190	91.35
Non-ISA	324	12,745	97.52	Non-ISA	593	12,810	95.58
PA (%)	87.53	98.04		PA (%)	77.19	98.54	
OA (%)	96.25	Kappa	0.865	OA (%)	94.98	Kappa	0.807
F1-score (ISA)		0.887			F1-score (ISA)	0.838	
F1-score (Non-ISA)		0.978			F1-score (Non-ISA)	0.970	
FROM_GLC2017	ISA	Non-ISA	UA (%)	**GLC_FCS2015**	ISA	Non-ISA	UA (%)
ISA	1949	245	88.83	ISA	2206	253	89.71
Non-ISA	651	12,754	95.14	Non-ISA	394	12,747	97.00
PA (%)	74.96	98.12		PA (%)	84.85	98.05	
OA (%)	94.26	Kappa	0.779	OA (%)	95.85	Kappa	0.847
F1-score (ISA)		0.813			F1-score (ISA)	0.872	
F1-score (Non-ISA)		0.966			F1-score (Non-ISA)	0.975	
GISA2.0 2018	ISA	Non-ISA	UA (%)	**GISA2.0 2015**	ISA	Non-ISA	UA (%)
ISA	2440	112	95.61	ISA	2421	115	95.47
Non-ISA	160	12,887	98.77	Non-ISA	179	12,885	98.63
PA (%)	93.85	99.13		PA (%)	93.12	99.12	
OA (%)	98.26	Kappa	0.936	OA (%)	95.85	Kappa	0.931
F1-score (ISA)		0.947			F1-score (ISA)	0.943	
F1-score (Non-ISA)		0.990			F1-score (Non-ISA)	0.989	

(M-grids) to achieve accuracy increment more efficiently. In this section, to better demonstrate the effect of the proposed mapping strategy for M-Grids, we used the automatic mapping method adopted for A-Girds to classify the M-Grids. Two M-Grids were randomly chosen for carrying out this comparison experiment. The first case is located at the North China Plain, where it can be seen that the A-Method reduced the ISA omissions of GAIA and GAUD in the scattered rural settlements, and the M-Method further improved the accuracy of detecting ISA from mountains and villages (Fig. 11 (a)-(b)). The other example is selected from the semi-arid region of southern Africa, where the A-Method suppressed the underestimation of GISA 1.0 but wrongly identified bare river beds as ISA (Fig. 11(c)-(d)). However, the proposed M-method can further improve the results by introducing the manually interpreted samples.

To further verify this conclusion, we compared M-method (manually-interpreted samples) and A-method (automatically-chosen samples) over 103 test grids. The results indicated that A-method could improve the mapping accuracy of GISA 1.0, while M-method outperformed A-method with an increment of F1-score by 0.0343 (Table 3). In future study, we will improve the overall accuracy of GISA in an iterative manner, by incorporation of more manully samples and advanced classification methods (e.g., deep learning (Li et al., 2022)), to obtain more accurate global impervious surface.

4.2. Global ISA expansion from GISA 2.0

Based on the 124,190 samples in 207 global grids, the accuracy assessment of GISA 2.0 showed that we further improved the F1-score of the global ISA to 0.936, outperforming GISA 1.0 (0.893), GAIA (0.720), and GAUD (0.809).

Based on GISA 2.0, the global ISA reached 815,452 km^2 in 2018, which was nearly three times that in 1985 (298,777 km^2). Although the existing datasets exhibited a general ISA increasing trend (Fig. 13(a)), they were different in the ISA spatial distribution. At the continental scale, GAIA showed relatively more disagreement than others, such as the underestimation in Europe and the slight overestimation in South America. GISA 1.0 and 2.0 detected relatively more ISA in Asia (Fig. 13 (b)-(f)). This can be attributed to the large number of rural training samples used in the GISA dataset, which is more effective for the extraction of small villages (Fig. 12 (d)). Asia experienced the most ISA expansion over the study period, with an increment of 271,633 km^2 from 1985 to 2018, followed by Europe (77,392 km^2). Oceania had the least growth (8,410 km^2). It should be noted that the detailed global ISA expansion analysis is beyond the scope of this paper, and here we only made a preliminary comparison between GISA 2.0 and other existing global datasets. More sophisticated data analysis is planned in future research.

4.3. Pros and cons of the proposed method as well as GISA 2.0

From a methodological point of view, we proposed a new global ISA mapping framework that aims to improve the global ISA mapping in terms of the consistency of existing ISA datasets. The merits of our method include:

i) In the case of mapping area, ISA was extracted from the global terrestrial surface instead of pre-defined urban masks, which, therefore, reduces the ISA omissions.

ii) In this study, we divided global terrestrial surface into grids with a side length of 100 km, which is smaller than that employed in the existing datasets (e.g., 3.5° grids in GAIA). Thus, the local properties of ISA can be characterized better, which is beneficial for improving the mapping performance.

We quantitatively discussed the effect of the number of the manually interpreted training samples. The results showed that generally the manually interpreted samples effectively increased mapping accuracy. The accuracy curve grows slowly as number of samples (polygons) increase, and reaches saturation with around 50 samples (Fig. 14). Therefore, in this paper, we added 50 samples in each M-grid, considering both mapping performance and workload of visual interpretation. Moreover, We also investigated the relationship between the accuracy improvement and ISA density within the test grids. The results reveal that the accuracy gain is higher in the low-density regions, while the gain is relatively lower in the high-density regions (Fig. 15). This conveys the information that the key for improving the accuracy of global ISA mainly lies in the low-density regions.

Current global ISA datasets exhibit high omission errors in some areas, which may be attributed largely to the insufficient spatial resolution (30 m) of Landsat data. When comparing GISA 2.0 with other global ISA datasets (Table 4), it is found that GISA 2.0 have an omission error of 6.88% in 2015, which is lower than GLC_FCS2015 (Global Land Cover Fine Classification System) (Zhang et al., 2020) and FROM_GLC2015 (Finer Resolution Observation and Monitoring of Global Land Cover) (Gong et al., 2019), by 8.27% and 15.93%, respectively. The omission error of GISA 2.0 is 6.15% in 2018, which is much lower than FROM_GLC2017 (25.04%) and GLC_FCS2020 (12.47%) (Zhang et al., 2021). Therefore, it can be stated that GISA 2.0 has less ISA omissions compared to other ISA datasets. Based on the global ISA mapping method proposed in this paper, we are capable of improving the overall accuracy of GISA and reducing the ISA omissions gradually. Meanwhile, in future, we plan to detect the global ISA with high-resolution remote sensing data (e.g., 10-m Sentinel or 3-m Planet) to further reduce the ISA omissions.

5. Conclusion

In this study, we proposed a global ISA mapping method by considering the consistency of existing products, based on which we further generated a new 30-m global ISA dataset (namely, GISA 2.0). To this end, we divided the global terrestrial surface into square grids of 100 km side length to take the local ISA characteristics into account. In order to quantitatively describe the disagreement between existing global datasets, we divided the mapping area into A-Grids and M-Grids in view of their consistency degree. An automatic mapping method was proposed for classifying the A-Grids, by extracting training samples from the consistent regions of existing datasets. On the other hand, in the case of M-Grids, where the existing ISA datasets showed large inconsistency, we proposed to add manually interpreted samples, to strengthen the classification and reduce the uncertainties in these areas. The accuracy assessment and inter-comparison verified the superiority of our methods and results.

It is worth noting that, in order to improve the efficiency, we focused on the regions or grids that were more inconsistent among existing global datasets. Although the selected grids only accounted for 10% of the global ISA grids, GISA 2.0 was able to significantly improve the mapping accuracy compared to the existing datasets. However, the remaining grids may still contain a number of errors, e.g. the regions that were misclassified or omitted by multiple datasets at the same time, which will be further investigated in our future work. Moreover, for the purpose of accurate global ISA mapping, we will continue to improve the quality of GISA according to the feedbacks from users around the world. GISA 2.0 can be freely downloaded from irsip.whu.edu.cn/resources/resources_en_v2.php.

CRediT authorship contribution statement

Xin Huang: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Yihong Song:** Methodology, Software, Investigation, Validation, Writing – original draft. **Jie Yang:** Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Wenrui Wang:** Investigation, Validation, Data curation. **Huiqun Ren:** Investigation, Validation, Data curation. **Mengjie Dong:** Investigation, Validation, Data curation. **Yujin Feng:** Investigation, Validation, Data curation. **Haidan Yin:** Investigation, Validation,

Data curation. **Jiayi Li:** Investigation, Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research was supported by the National Natural Science Foundation of China under Grant 41971295, and the Foundation for Innovative Research Groups of the Natural Science Foundation of Hubei Province under Grant 2020CFA003.

References

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S., 2013. Urban Footprint processor—fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. IEEE Geosci. Remote Sens. Lett. 10, 1617–1621. https://doi.org/10.1109/LGRS.2013.2272953.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114 (1), 168–182.

Gong, J., Liu, C., Huang, X., 2020a. Advances in urban information extraction from high-resolution remote sensing imagery. Sci. China Earth Sci. 63, 463–475. https://doi.org/10.1007/s11430-019-9547-x.

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., Zhou, Y., 2020b. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. Remote Sens. Environ. 236, 111510 https://doi.org/10.1016/j.rse.2019.111510.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci. Bull. 64, 370–373. https://doi.org/10.1016/j.scib.2019.03.002.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Hamedianfar, A., Shafri, H.Z.M., 2016. Integrated approach using data mining-based decision tree and object-based image analysis for high-resolution urban mapping of WorldView-2 satellite sensor data. J. Appl. Remote Sens. 10, 1–21. https://doi.org/10.1117/1.JRS.10.025001.

Hu, T., Huang, X., 2019. A novel locally adaptive method for modeling the spatiotemporal dynamics of global electric power consumption based on DMSP-OLS nighttime stable light data. Appl. Energy 240, 778–792. https://doi.org/10.1016/j.apenergy.2019.02.062.

Huang, X., Li, J., Yang, J., Zhang, Z., Li, D., Liu, X., Xin, H., Jiayi, L., Jie, Y., Zhen, Z., Dongrui, L., Xiaoping, L., 2021. 30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019. Sci. CHINA Earth Sci. https://doi.org/10.1007/s11430-020-9797-9.

Kaspersen, P., Fensholt, R., Drews, M., 2015. Using landsat vegetation indices to estimate impervious surface fractions for european cities. Remote Sens 7 (6), 8224–8249.

Lee, C., Kim, K., Lee, H., 2018. GIS based optimal impervious surface map generation using various spatial data for urban nonpoint source management. J. Environ. Manage. 206, 587–601. https://doi.org/10.1016/j.jenvman.2017.10.076.

Li, J., Zhang, B., Huang, X., 2022. A hierarchical category structure based convolutional recurrent neural network (HCS-ConvRNN) for Land-Cover classification using dense MODIS Time-Series data. Int. J. Appl. Earth Obs. Geoinf. 108, 102744 https://doi.org/10.1016/j.jag.2022.102744.

Li, X., Gong, P., Liang, L., 2015. A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data. Remote Sens. Environ. 166, 78–90. https://doi.org/10.1016/j.rse.2015.06.007.

Liu, C., Huang, X., Zhu, Z., Chen, H., Tang, X., Gong, J., 2019. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: analysis of 45 global cities. Remote Sens. Environ. 226, 51–73. https://doi.org/10.1016/j.rse.2019.03.033.

Liu, C., Shao, Z., Chen, M., Luo, H., 2013. MNDISI: a multi-source composition index for impervious surface area estimation at the individual city scale. Remote Sens. Lett. 4, 803–812. https://doi.org/10.1080/2150704X.2013.798710.

Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A.D., Chen, A., Gong, P., Chen, J., Hu, G., Chen, Y., Wang, S., Wu, Q., Huang, K., Estes, L., Zeng, Z., 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. Nat. Sustain. 3, 564–570. https://doi.org/10.1038/s41893-020-0521-x.

Loveland, T.R., Dwyer, J.L., 2012. Landsat: Building a strong future. Remote Sens. Environ. 122, 22–29. https://doi.org/10.1016/j.rse.2011.09.022.

Lu, D., Li, G., Kuang, W., Moran, E., 2014. Methods to extract impervious surface areas from satellite images. Int. J. Digit. Earth 7, 93–112. https://doi.org/10.1080/17538947.2013.866173.

Luo, M., Lau, N.-C., 2019. Urban expansion and drying climate in an urban agglomeration of east China. Geophys. Res. Lett. 46, 6868–6877. https://doi.org/10.1029/2019GL082736.

Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422. https://doi.org/10.1038/nature20584.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A.J., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., Syrris, V., 2016. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014; JRC Technical Reports EUR 27741 EN. Publ. Off. Eur. Union 62.

Shao, Z., Liu, C., 2014. The integrated use of DMSP-OLS nighttime light and MODIS data for monitoring large-scale impervious surface dynamics: a case study in the Yangtze River Delta. Remote Sens 6 (10), 9359–9378.

Song, Y., Chen, B., Kwan, M.-P., 2020. How does urban expansion impact people's exposure to green environments? A comparative study of 290 Chinese cities. J. Clean. Prod. 246, 119018 https://doi.org/10.1016/j.jclepro.2019.119018.

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199 https://doi.org/10.1016/j.rse.2019.05.018.

Sun, Z., Guo, H., Li, X., Lu, L., Du, X., 2011. Estimating urban impervious surfaces from Landsat-5 TM imagery using multilayer perceptron neural network and support vector machine. J. Appl. Remote Sens. 5, 1–18. https://doi.org/10.1117/1.3539767.

Wu, C., Murray, A.T., 2003. Estimating impervious surface distribution by spectral mixture analysis. Remote Sens. Environ. 84, 493–505. https://doi.org/10.1016/S0034-4257(02)00136-0.

Xu, J., Zhao, Y., Zhong, K., Zhang, F., Liu, X., Sun, C., 2018. Measuring spatio-temporal dynamics of impervious surface in Guangzhou, China, from 1988 to 2015, using time-series Landsat imagery. Sci. Total Environ. 627, 264–281. https://doi.org/10.1016/j.scitotenv.2018.01.155.

Yang, J., Huang, X., 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci. Data 13, 3907–3925. https://doi.org/10.5194/essd-13-3907-2021.

Zhang, X., Liu, L., Wu, C., Chen, X., Gao, Y., Xie, S., Zhang, B., 2020. Development of a global 30m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform. Earth Syst. Sci. Data 12, 1625–1648. https://doi.org/10.5194/essd-12-1625-2020.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. Earth Syst. Sci. Data 13, 2753–2776. https://doi.org/10.5194/essd-13-2753-2021.

Zhang, Y., Zhang, H., Lin, H., 2014. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. Remote Sens. Environ. 141, 155–167. https://doi.org/10.1016/j.rse.2013.10.028.

Zhuo, L., Zheng, J., Zhang, X., Li, J., Liu, L., 2015. An improved method of night-time light saturation reduction based on EVI. Int. J. Remote Sens. 36, 4114–4130. https://doi.org/10.1080/01431161.2015.1073861.