Fractal Theory Based Stratified Sampling for Quality Assessment of Remote-Sensing-Derived Geospatial Data

Yao Lu [ID], Huan Xie [ID], *Senior Member, IEEE*, Jixian Zhang, Yanmin Jin, Yongjiu Feng [ID], Yali Gong, Wenli Han, He Zhang, and Xiaohua Tong [ID], *Senior Member, IEEE*

Abstract—The stratified sampling is widely used in quality assessment of remote-sensing-derived geospatial data (RSGD). Because of the different stratification indicators (also called the stratification variables) used in stratified sampling, it will lead to different evaluation results. By using fractal theory, this article proposes a stratified sampling method based on fractal (SSF) for quality assessment of RSGD. As a stratification variable, fractal dimension is related to and independent of the study variable in the quality assessment of RSGD. This method can quantitatively and accurately stratify the population, which leads to minimizing the intra-stratum variance, acquiring higher estimation accuracy, and estimation efficiency. The proposed SSF method in this article is transformed into three formulated problems: the quantitative calculation of fractal, the optimal solution of the stratum boundary value, and the configuration of sample sizes. The experiment shows a quantitative performance comparison of SSF, stratified sampling based on class, and sample random sampling using the South Sudan Global Core Vector Database 2020. Design effect and root-mean-square error provide a quantitative assessment of the performance in this study. The experimental results verify the feasibility and applicability of the SSF proposed in this article. It also shows higher estimation accuracy and more economical cost.

Index Terms—Fractal dimension, quality assessment, stratification indicators, stratified sampling method, stratum boundary value.

Yao Lu is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China, also with the National Quality Inspection and Testing Center for Surveying and Mapping Products, Beijing 100830, China, and also with the Technology Innovation Center for Remote Sensing Intelligent Verification, Ministry of Natural Resources, Beijing 100830, China (e-mail: 156291029@qq.com).

Huan Xie, Yanmin Jin, Yongjiu Feng, Yali Gong, and Xiaohua Tong are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: huanxie@tongji.edu.cn; jinyanmin@tongji.edu.cn; yjfeng@tongji.edu.cn; 1710643@tongji.edu.cn; xhtong@tongji.edu.cn).

Jixian Zhang, Wenli Han, and He Zhang are with the National Quality Inspection and Testing Center For Surveying and Mapping Products, Beijing 100830, China, and also with the Technology Innovation Center for Remote Sensing Intelligent Verification, Ministry of Natural Resources, Beijing 100830, China (e-mail: zhangjx@casm.ac.cn; 88212479@qq.com; 3885987@qq.com).

Digital Object Identifier 10.1109/JSTARS.2023.3287347

I. INTRODUCTION

WITH the rapid development of remote sensing technology in the world, the amount of data generated by Terra, Landsat, Spot, NOAA, HJ-1B, GF, and other remote sensing satellites has increased exponentially, and the types of products derived from these data have also increased more and more. The remote-sensing-derived geospatial data (RSGD) are produced based on remote sensing technology and other technologies, including digital maps, digital elevation models, landcover products, land surface temperature products, etc. [1], for example, the VCF data developed by the US Geological Survey (USGS), the University of Maryland and other institutions [2], the MODIS-MCD45 burned area product [3], the USGS National Land Cover Database [4], OpenStreetMap (OSM), etc. [5]. The quality of RSGD, which is widely used in various industries, is getting more and more attention. The quality of RSGD is a key factor that affects the end use. For example, in rice planting acreage estimation [6], deforestation acreage statistics [7], global land cover classification statistics [8], and even in autonomous driving, the quality of RSGD will directly affect the accuracy, reliability, inference, and decision-making at the time of using it [9]. Therefore, the quality assessment of RSGD can provide users with evidence to confirm the quality level and determine whether the quality can meet the expected requirements [10]. In the era of Big Data, the time, money, manpower, and equipment spent on all inspection items of each one for quality evaluation could not be borne by both producers and users. In quality evaluation, sampling inspection is an economical and efficient method, which selects a part of the whole for quality inspection, and then infers the overall data or products.

In the research of sampling methods, the commonly used methods are simple random sampling (SRS), systematic sampling, cluster sampling, and stratified sampling [11]. Stratified sampling is a method that divides the population into multiple nonoverlapping subpopulations (each subpopulation is called a stratum) according to stratification indicators (also known as stratification variables), and sampling independently from each stratum [12], [13]. Because it can obtain higher estimated accuracy than SRS, stratified sampling is adopted in large amounts in the work of land cover statistics and accuracy assessment [14], [15], [16].

When the population is stratified by different stratification indicators, different estimators will be obtained when the sample size is the same. The principle of stratified sampling should make the intrastratum variance as small as possible. Xie et al. [1] designed a multilevel stratified spatial sampling approach for the quality assessment of RSGD; Giree et al. [17] used the reference data as the stratification indicator to estimate the gross forest cover loss; Olofsson et al. [18] used the Köppen climate classification and population density as the stratification indicator to validate the global land coverage classification data. These studies follow the principle of large interstratum variance and small intrastratum variance. When these methods comply with the principle, the research works show that the higher the correlation between the stratification variables and the target variables (study variable) is, the higher the final estimation accuracy is [19], [20]. The stratification indicator adopted by Xie et al. [1] depends on the product to be evaluated, but the quality of the product to be evaluated is unknown, which leads to an unknown impact on the estimation accuracy. The stratification indicator adopted by Giree et al. relies on reference data, which would fail when the reference data cannot be obtained. The stratification indicators adopted by Olofsson et al. [18] are climate types, whose benefit improvement of final verification is limited because the correlation between climate types and estimators is unknown. Therefore, exploring a strict stratification indicator is an important research content of stratified sampling. In the quality assessment of RSGD, ideally, stratifying directly according to the quality assessment value of each element is the optimal stratified sampling design [20], [22]. However, in the sampling stage, the quality evaluation value of each element is unknown. If that is known, sampling is not needed to estimate the population quality [23]. Therefore, it is an ideal situation that the target variables are set as stratification variables [24]. The stratified sampling based on class (SSC) with the classification of elements as the stratification indicator is the most widely used sampling method at present. However, the problem with this method is that the correlation between element classifications and quality is not obvious, which has an impact on the reliability of the final estimator. In this case, it is necessary to study and explore a stratification indicator suitable for RSGD quality evaluation. This stratification indicator is not limited and dependent on the quantity to be evaluated, which has a strong correlation with the indicator.

Fractal is a theory across natural science and social science, which was founded in the 1970s. It was first proposed by Mandelbrot [25] in a paper published in the Science in 1967. It can describe some irregular, unstable, or complex objects in nature. Fractal is a common phenomenon whose theory is a branch of nonlinear scientific theory. It is mainly used to find the laws hidden behind complex and chaotic phenomena. Different from strict mathematical fractal, natural objects usually do not show accurate self-similarity [26]. Based on the above characteristics, fractal geometry provides great potential in the representation, measurement, and analysis of complex remote sensing images and geographical elements [27], [28], [29]. For example, Jiang and Brandt discussed the application of fractal in geospatial topology and scale analysis [30]. Bretar et al. [31] tried to use fractal dimension to characterize the surface roughness of volcanic terrain remote sensing images. Cheng et al. [32] studied and calculated the landscape self-similarity index of Taiwan Island based on fractal and obtained the conclusion that the landscape self-similarity index of Taiwan Island changes with altitude. Cosh and Brutsaert [33] calculated the fractal dimension of vegetation density in different land classifications through wavelet analysis, spectral analysis, and other methods. Read and Lam found that fractal dimension is useful to characterize the spatial complexity of Landsat data [34].

In prior studies on fractal-based sampling methods, Ferrandino [35] conducted a study on measuring spatial aggregation in binary epidemics, utilizing a nested fractal design where the sample point locations were determined by a fractal pattern based on squares. The study compared and analyzed the advantages of this nested fractal design in contrast to the grid sampling and the cluster sampling [35]. Marsh and Ewers designed a scalable sampling design utilizing fractal geometry. They employed a sampling frame consisting of a fractal pattern based on equilateral triangles, with the objective of ecological surveys quantifying β-diversity [36]; Jiang et al. [37] proposed a novel sampling scheme based on Cantor sets, a classic fractal. This scheme utilizes a well known and precise fractal structure to accelerate parallel simulation of multithreaded applications [37]. Previous studies on fractal-based sampling methods used the principle of self-similarity to determine sampling locations or sampling frames by means of triangular, square, or hexagonal grids with changeable side lengths. Nonetheless, these methods can only be applied to point elements and are not suitable for assessing the quality of linear or polygonal elements in RSGD.

In the production of RSGD, the higher the fractal dimensions are, the more disordered and complex the corresponding elements are [30]. The more disordered and complex the corresponding elements are in RSGD, the more difficult it is to produce such elements that meet the quality requirements (the degree of difficulty is related to this). The quality assessment of RSGD is based on whether each element meets the quality requirements. It gives the feasibility of using fractal dimension as stratification indicator [38].

This article aims to integrate fractal theory and propose a stratified sampling method based on fractal dimension. This method can use fractal dimension to quantitatively describe the structure of the data to be evaluated. The purpose of this experiment is

1) to verify the feasibility of fractal dimension as a stratification indicator by comparing stratified sampling method based on fractal (SSF) with SRS;
2) to study whether the proposed method can provide more effective quality assessment through the comparison with different sampling methods and different stratification indicators;
3) to explore whether it can improve the sampling efficiency through the comparison with the stratification indicator.

The rest of this paper is organized as follows: in Section II, the SSF method for RSGD quality assessment is studied, and the theoretical proof and conceptual model of the method are given. After that, the experimental area and experimental data

are introduced; in Section III, the proposed SSF method is quantitatively compared with SRS and SSC; in Section IV, the experimental results of the previous section are analyzed and discussed to verify the proposed SSF in this article; in the last section, the performance of SSF method is discussed and summarized.

II. METHODOLOGY

A. Theoretical Basis of the SSF Method

For the quality assessment of RSGD using traditional stratified sampling, the set of each element (also called individual) in the RSGD is defined as a population M of size N, and $q_i(i = 1, 2, 3 \ldots, N)$ represents the quality assessment value of each element. Next, the population is divided into h layers according to the value of q_i, and samples of size n_h are selected from each layer to form a sample population m of size n. The unbiased estimator for the population quality assessment value is constructed by the quality assessment value of the samples, and the unbiased estimator for the population variance of the quality assessment value is given by

$$\hat{V}\left(\bar{q}_{st}\right) = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2\left(q\right)}{n_h} \qquad (1)$$

where N is the size of population, N_h is the size of the stratum h ($h = 1$ to L), n_h is the sample size of the stratum h, W_h is the weight of the stratum h, $s_h^2(q)$ is the sample variance of the quality assessment value in the stratum h [39]. The optimal stratified sampling method should minimize the variance estimator $\hat{V}(\bar{q}_{st})$ for the study variable q_i [40].

It can be observed that when the stratum weight W_h is fixed, $s_h^2(q)$ shows a positive correlation with $\hat{V}(\bar{q}_{st})$. Therefore, reducing $s_h^2(q)$ as much as possible can minimize $\hat{V}(\bar{q}_{st})$. From the definition of variance, it can be known that the sample variance of the quality evaluation value in the stratum h is

$$s_h^2\left(q\right) = \left(\frac{\sum_{i=1}^{n_h}\left(q_i - \bar{q}_h\right)^2}{n_h - 1}\right) \qquad (2)$$

where the study variable q_i is used as the stratification indicator (or stratification variable). However, in practical application, the q_i to be studied or investigated is an unknown quantity.

To address this issue, this article proposes the SSF method, which uses the fractal dimension as a stratification indicator or auxiliary variable, as it is easily obtained. The fractal dimension value of each element in the population M is denoted as $f_i(i = 1, 2, 3 \ldots, n)$, then the sample variance of the fractal dimension value in the stratum h is

$$s_h^2\left(f_i\right) = \left(\frac{\sum_{i=1}^{n_h}\left(f_i - \bar{f}_h\right)^2}{n_h - 1}\right). \qquad (3)$$

If there is a positive or negative correlation between f_i and q_i, minimizing $s_h^2(f_i)$ or maximizing $s_h^2(f_i)$ can make $s_h^2(q_i)$ as small as possible. Considering that f_i and q_i correlation

coefficients is

$$p_{fq} = \frac{Cov\left(f, q\right)}{\sqrt{D\left(f\right)}\sqrt{D\left(q\right)}} = \frac{Cov\left(f, q\right)}{s_h\left(f\right) s_h\left(q\right)} \qquad (4)$$

we can show that

$$s_h\left(q\right) = \frac{Cov\left(f, q\right)}{s_h\left(f\right) p_{fq}}. \qquad (5)$$

It can be seen from (5) that the $s_h(q)$ has a negative correlation with the correlation coefficient p_{fq}. If there is a correlation between the auxiliary variable f_i and the study variable q_i, $\hat{V}(\bar{q}_{st})$ can be reduced and the sampling accuracy can be improved. Moreover, the higher the correlation between f_i and q_i is, the higher the improvement of sampling accuracy is. Therefore, according to (1)–(5), the proof of the feasibility of the SSF method proposed in this article can be transformed into a proof of the correlation between the fractal dimension value f_i and the quality assessment value q_i of the same element.

Fractal dimension is a quantitative description of fractals. Different from the traditional Euclidean Space, it can quantitatively describe geospatial objects in a continuous noninteger dimension. The size of its value is closely related to "complexity," "roughness," etc., [41], [42], [43], [44]. It is difficult to directly prove the correlation between the fractal dimension and the quality assessment, so this article introduces the Shannon Entropy to prove it.

The Shannon Entropy is a rigorous measure of uncertainty in discrete random variables [45], [46], [47], [48]. When the quality of RSGD is evaluated by the SSF method, the quality assessment value of each element can be defined as a discrete random variable. Thus, the Shannon Entropy U_i of each element quality assessment is the quantitative characterization of the quality uncertainty. According to the physical meaning of Shannon Entropy, the larger its value U_i is, the higher the uncertainty is and the lower the quality assessment q_i is. The negative correlation between the Shannon Entropy U_i and the quality assessment value q_i is evident. If it can be proved that the fractal dimension value f_i of the same element has a correlation or an equivalence relation with the Shannon Entropy U_i, there is a correlation between f_i and q_i, that means the proof for the feasibility of the SSF method proposed in this article is completed.

It is known that each element in the population M is expressed as M_i ($i = 1, 2, 3 \ldots, N$), and the corresponding fractal dimension can be expressed as

$$f_i = \lim_{\varepsilon \to 0} \frac{\ln K_i}{-\ln \varepsilon} \qquad (6)$$

where ε represents the measure in a region of size ε, and K_i represents the number of times that the ith element needs to be measured under the ε measure. The Shannon Entropy [49] corresponding to M_i can be expressed as
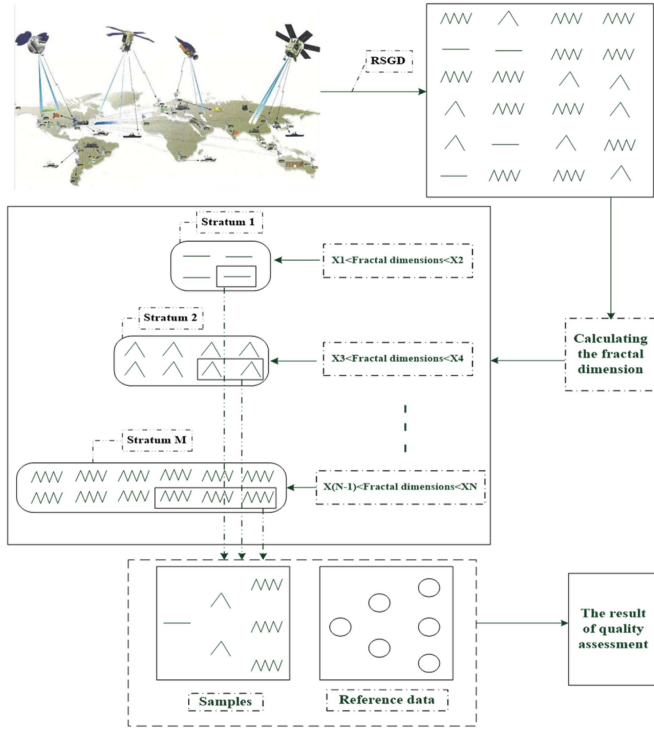
$$U_i = -\sum_{i=1}^{K} P_i \ln P_i. \qquad (7)$$

Fig. 1. SSF conceptual flow chart.

Considering the uniform distribution of disjoint ε under the same measure ε, the Shannon Entropy can be written as

$$U_i = -\sum_{i=1}^{K} \frac{1}{K_i} \ln \frac{1}{K_i} = \ln K_i. \tag{8}$$

According to (6) and (8), we can establish a simple relationship between f_i and U_i that can be written as

$$f_i = \lim_{\varepsilon \to 0} \frac{U_i}{-\ln \varepsilon}. \tag{9}$$

Thus, the fractal dimension f_i grows as the Shannon Entropy U_i from (9), which means both are positive correlation.

Through (6)–(9), the correlation between fractal dimension value f_i and the Shannon Entropy value U_i of the same element in RSGD is proved, which means the feasibility of the SSF method proposed in this article is proved.

B. Implementation of the SSF Method

The conceptual flow chart of the SSF sampling method for RSGD quality evaluation is in Fig. 1. The implementation of SSF method is mainly divided into four steps.

1) First, defining the population and samples, that is, each geospatial object of the data to be evaluated corresponding to each element is defined as population M of size N, and the elements of size n ($n \leq N$) selected from M are called the sample population m.
2) Calculating the fractal dimension of each element in the population and record them as $f_i (i = 1, 2, 3 \ldots, N)$.

3) According to the fractal dimension value f_i of each element, it is divided into nonoverlapping parts called strata with a size of h. Each element corresponding to each f_i is also divided into corresponding strata.
4) The sample size n_h of each stratum is calculated according to a certain sample allocation rule. And then the samples are distributed in each stratum according to the simple random principle.

C. Calculation of the Fractal Dimension

There have been many research works on calculating the fractal dimension, including Variogram method [50], [51], Divider method [52], [53], Box-counting method [54], [55], the Power Spectral method, etc., [56], [57], [58]. The Box-counting method uses square grids with different side lengths ε to cover the studied object. When the ε of the square grid changes, the number $N(\varepsilon)$ of grids covering the studied object changes accordingly and can be written as

$$K(\varepsilon) \sim \varepsilon^{-f}. \tag{10}$$

In a double logarithm graph, a series of points are linearly fitted. The absolute value of the slope of this line is the fractal dimension f. The f can be calculated as

$$\lg K(\varepsilon) = -f \lg \varepsilon + C \tag{11}$$

where ε is the grid side length, $K(\varepsilon)$ is the grid number, and C is the undetermined constant. RSGD is divided into vector data and raster data. Raster data itself is a matrix composed of squares. Vector data can be processed by constructing a square grid. Therefore, in SSF method, the Box-counting method is used to calculate fractal dimension. Figs. 2 and 3 show the calculation results of the traffic and water linear products of South Sudan Global Core Vector Data (GCVD) 2020, which is a kind of RSGD, by using the Box-counting method. The proportion of elements with high fractal dimension in map X is significantly higher than that in map Y.

D. Determining the Number and Boundary of the Strata

To evaluate the quality of RSGD with SSF, the population needs to be stratified. In the previous subsubsection, the most appropriate stratification indicator-fractal dimension has been studied and determined. The following problem is: how many strata should be divided and the boundary between strata should be determined. William G. Cochran deduced the relationship between the population variance of unbiased estimator and the stratum number in 1972. Besides, it is proved that when L is large and 6, the reduction rate of V will be greatly slowed down. Fuller [59] and Zseby [22] also showed that the larger L is, the higher the gain of stratified sampling is,. When L is greater than 4, the gain of stratified sampling tends to be stable.

After the number of strata is determined, it is necessary to solve the problem of strata boundary determination, that is, the value of stratification indicator x, which is the basis for determining strata boundary.

How to determine the stratum boundary? Dividing the cumulative square root of the distribution density function $f(x)$ of
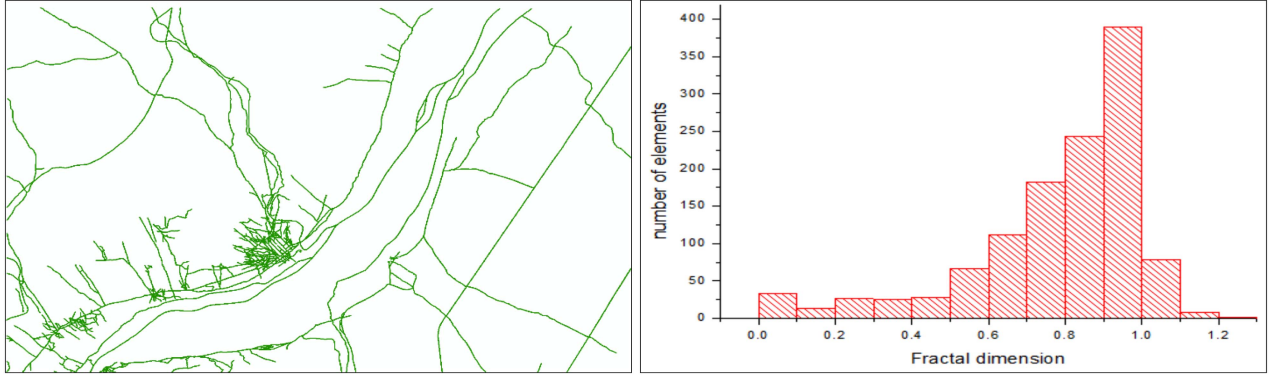
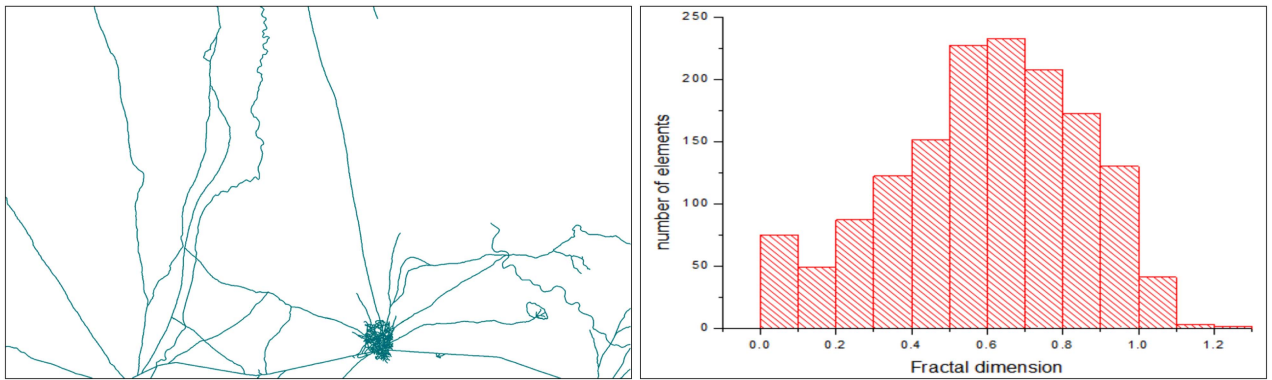Fig. 2.   Map X and its corresponding fractal dimension histogram.

Fig. 3.   Map Y and its corresponding fractal dimension histogram.

the stratification variable x equally can make the variance of x in each stratum as small as possible. This is in line with the principle of small intrastratum variance and large interstratum variance. When the population is reasonably stratified, the sample size allocation can be carried out based on the number of subpopulations in each stratum.

E. Sample Size Allocation

After the boundary of each stratum is determined, the samples need to be reasonably distributed to each stratum in a certain way. In stratified sampling, the different methods of sample allocation in each stratum will have a certain impact on the accuracy of estimators. The main methods are equal allocation, proportional allocation, and Neyman allocation [24]. Equal allocation makes the sample size n_h equal in each stratum, without considering the population characteristics. Proportional allocation is to make the sample size proportional to the subpopulation size N_h of each stratum. The sample size of each stratum n_h can be written as

$$n_h = n \times w_h = n \times W_h = n \times \frac{N_h}{N}. \tag{12}$$

Neyman allocation assumes that the sampling cost is equal. When the sample size n is fixed, the sample size of each stratum

n_h can be written as

$$n_h = n \times w_h = n \times \frac{W_h S_h}{\sum W_h S_h}. \tag{13}$$

F. Effectiveness Evaluation Method of the SSF Method

To assess the advantages and disadvantages of the SSF method, it is necessary to evaluate the gain it generates. Different sampling methods have different effectiveness of estimators. The advantages and disadvantages of different methods are evaluated by design effect (DEFF) and root-mean-square error (RMSE).

DEFF is an index to measure the effect of sample method proposed by Kish [60], which is called design effect. DEFF can be calculate by

$$\text{Deff} = \frac{S}{S_r} \tag{14}$$

where S is the estimator variance of the sample method to be evaluated, and S_r is the estimator variance of SRS under the same sample size.

The idea of using RMSE for evaluation is when the sample size n is the same, the observed value of the sample quality assessment obtained by method A is $\{A_1, A_2, \ldots \ldots, A_n\}$, the observed value of the sample quality assessment obtained by
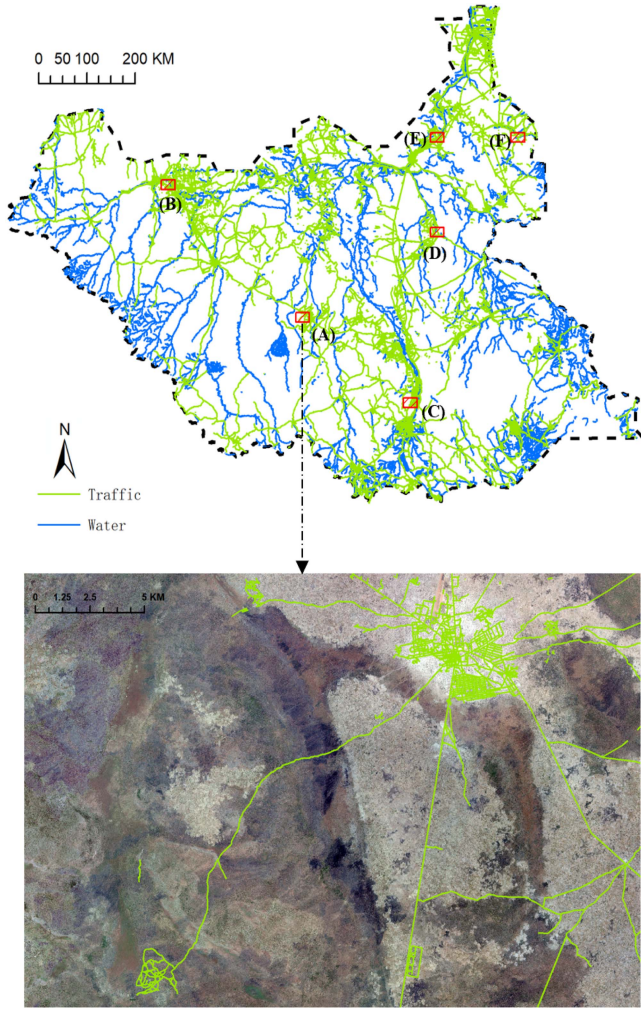
Fig. 4.   Study area and experimental data.

Map	Population	Sample size									
		2%	4%	6%	8%	10%	12%	14%	16%	18%	20%
A	1778	35	71	107	142	177	213	249	284	320	356
B	2514	50	101	151	201	251	302	352	402	453	503
C	2811	56	112	169	225	281	337	394	450	506	562
D	1493	30	60	90	119	149	179	209	238	269	298
E	1197	24	48	72	96	120	144	167	191	215	239
F	1299	26	52	78	104	130	156	182	208	234	260

method B is $\{B_1, B_2, \ldots\ldots, B_n\}$, and the true value of the overall quality assessment is $\{T_1, T_2, \ldots\ldots, T_n\}$

$$\text{RMSE}(A) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - T_i)^2} \qquad (15)$$

$$\text{RMSE}(B) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(B_i - T_i)^2}. \qquad (16)$$

If $\text{RMSE}(A) < \text{RMSE}(B)$, the estimator obtained by sample method A is closer to the true value, with less dispersion degree and higher accuracy.

Map	population	wrong elements	correct elements	correct rate (%)
A	1778	29	1749	98.37%
B	2514	19	2495	99.24%
C	2811	8	2803	99.72%
D	1493	11	1482	99.26%
E	1197	6	1191	99.50%
F	1299	7	1292	99.46%

III. EMPIRICAL CASE STUDIES

A. Study Area

South Sudan selected for the experimental study area is in northeast Africa, between longitude 24°9'7"–35°52'12" and latitude 3°28'52"–12°13'8." It belongs to zone 35 and 36 of UTM projection 6°, and the central meridian is longitude 27° and 33°, respectively. It borders Ethiopia in the East, Kenya, Uganda, and the Democratic Republic of the Congo in the south, the Central African Republic in the West, and Sudan in the north. South Sudan has a tropical grassland climate, and more than 95% of the whole territory is humid and semi-humid areas, with high humidity and a lot of rainfall. The rainy season is from May to October every year, and the temperature is 20–40 °C; the dry season is from November to April, and the temperature is 30–50 °C. Besides, forest resources of South Sudan are very rich.

The study data is the traffic and water linear products of South Sudan GCVD 2020, which is the vector data expressing geographic information elements in the form of lines. The products select main elements according to demand, importance, and operability. Six maps with the same size are randomly selected from the study data as the experimental data, which is consistent with the assessment unit of the quality evaluation organization. The size of each map obtained by measuring tools is 27500 m × 18400 m. The study area and experimental data of this experiment are shown in Fig. 4.

B. Experiment Design

In this experiment, six maps A, B, C, D, E, and F were compared by three methods: SSF, SSC, and SRS. The experimental maps were sampled respectively, and the highest sampling ratio of each map was designed as W% of the population. Each method was used to sample according to the sample sizes of W%. The quality was evaluated according to the reference data, and the corresponding correct rate of different sampling methods under the same sampling ratio was calculated. We repeated the experiment 50 times as the observation values, and the reference data were the reference values. According to (15) and (16), the RMSE of the corresponding correct rate under the three sampling methods can be calculated, and finally, the benefits of the three sampling methods can be evaluated. It should be noted that
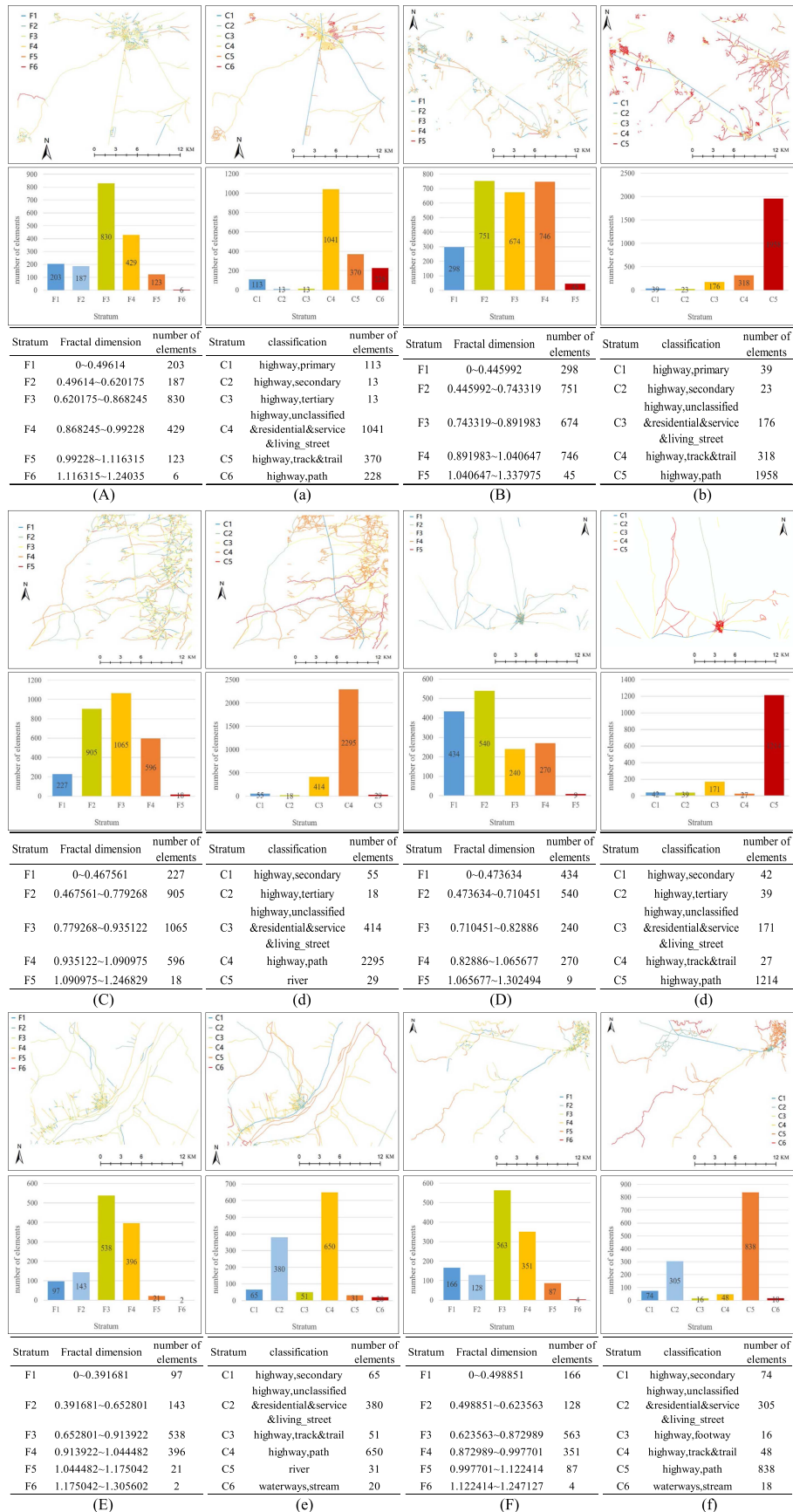
Fig. 5. Stratification results of the experimental data. (A), (B), (C), (D), (E), and (F) are the results of Map A, B, C, D, E and F, respectively, by the SSF method. (a), (b), (c), (d), (e) and (f) are the results of Map A, B, C, D, E and F, respectively, by the SSC method.

TABLE III
DEFF AND RMSE RESULTS OF THE SSF, SSC, AND SRS METHOD

	RMSE								
	Map A			Map B			Map C		
Sampling ratio	SSF	SSC	SRS	SSF	SSC	SRS	SSF	SSC	SRS
2%	0.01830	0.02188	0.02474	0.01088	0.01194	0.01293	0.00682	0.00854	0.00798
4%	0.01520	0.01595	0.01709	0.00719	0.00774	0.00950	0.00371	0.00478	0.00475
6%	0.01083	0.01185	0.01151	0.00581	0.00699	0.00759	0.00338	0.00380	0.00431
8%	0.00942	0.01025	0.00974	0.00557	0.00605	0.00631	0.00305	0.00363	0.00356
10%	0.00829	0.00931	0.00876	0.00534	0.00511	0.00613	0.00304	0.00264	0.00318
12%	0.00582	0.00926	0.00756	0.00432	0.00496	0.00534	0.00246	0.00271	0.00264
14%	0.00700	0.00805	0.00764	0.00363	0.00462	0.00435	0.00243	0.00244	0.00252
16%	0.00646	0.00763	0.00736	0.00376	0.00407	0.00461	0.00210	0.00223	0.00231
18%	0.00589	0.00609	0.00623	0.00332	0.00418	0.00379	0.00190	0.00207	0.00226
20%	0.00549	0.00610	0.00568	0.00310	0.00314	0.00326	0.00169	0.00176	0.00194
	Map D			Map E			Map F		
2%	0.01150	0.01427	0.01612	0.01143	0.01675	0.02051	0.01311	0.01412	0.01676
4%	0.00877	0.00817	0.01128	0.00888	0.00932	0.01342	0.00882	0.01012	0.01358
6%	0.00745	0.01036	0.01061	0.00677	0.00952	0.00877	0.00675	0.00704	0.00838
8%	0.00759	0.00799	0.00782	0.00569	0.00761	0.00763	0.00638	0.00741	0.00707
10%	0.00567	0.00576	0.00703	0.00460	0.00728	0.00690	0.00564	0.00565	0.00615
12%	0.00523	0.00596	0.00527	0.00522	0.00499	0.00650	0.00463	0.00519	0.00568
14%	0.00550	0.00563	0.00610	0.00427	0.00427	0.00458	0.00493	0.00532	0.00557
16%	0.00475	0.00548	0.00522	0.00542	0.00503	0.00595	0.00449	0.00461	0.00464
18%	0.00456	0.00532	0.00497	0.00432	0.00480	0.00505	0.00355	0.00370	0.00390
20%	0.00427	0.00468	0.00481	0.00405	0.00390	0.00414	0.00378	0.00403	0.00386

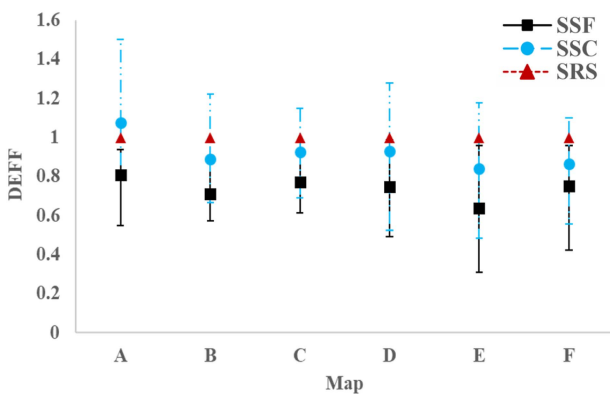|	DEFF								
	Map A			Map B			Map C		
Sampling ratio									
2%	0.54739	0.78276	1	0.70834	0.85247	1	0.73079	1.14693	1
4%	0.79057	0.87010	1	0.57327	0.66360	1	0.61267	1.01473	1
6%	0.88513	1.06133	1	0.58696	0.85012	1	0.61547	0.77803	1
8%	0.93454	1.10757	1	0.77766	0.91819	1	0.73196	1.04027	1
10%	0.89703	1.13145	1	0.75888	0.69583	1	0.91268	0.69027	1
12%	0.59209	1.50029	1	0.65517	0.86103	1	0.86878	1.05283	1
14%	0.83809	1.10846	1	0.69657	1.12891	1	0.92834	0.93928	1
16%	0.77031	1.07381	1	0.66501	0.78114	1	0.82820	0.93465	1
18%	0.89257	0.95427	1	0.77048	1.22024	1	0.70567	0.83623	1
20%	0.93683	1.15517	1	0.90396	0.92802	1	0.76004	0.82352	1
	Map D			Map E			Map F		
2%	0.50899	0.78385	1	0.31037	0.66637	1	0.61191	0.70951	1
4%	0.60416	0.52430	1	0.43816	0.48169	1	0.42217	0.55613	1
6%	0.49195	0.95204	1	0.59664	1.17848	1	0.64823	0.70476	1
8%	0.94336	1.04397	1	0.55741	0.99610	1	0.81469	1.09850	1
10%	0.65023	0.66978	1	0.44541	1.11490	1	0.84101	0.84339	1
12%	0.98532	1.27969	1	0.64508	0.58966	1	0.66510	0.83425	1
14%	0.81555	0.85300	1	0.86626	0.86782	1	0.78380	0.91371	1
16%	0.82981	1.10182	1	0.83245	0.71487	1	0.93319	0.98610	1
18%	0.84219	1.14520	1	0.73217	0.90240	1	0.82924	0.89897	1
20%	0.78758	0.94484	1	0.95916	0.88703	1	0.95585	1.08886	1

Fig. 6. DEFF value ranges and the average DEFF value of the SSF method, the SSC method, and the SRS method in map A, B, C, D, E, and F.

1) to make the comparison between SSF and SSC meaningful in this experiment, the number of strata of the two methods in each map should be kept equal;

2) after sampling, the quality inspection items of the samples are checked, and the variable $Y_i(i = 1, 2, \ldots ., n)$ is used to represent the result of inspection item. This variable can be taken as 0 and 1. "$Y = 1$" indicates that the inspection item is correct, "$Y = 0$" indicates that the inspection item is defective. In order to facilitate the experimental calculation, the quality assessment value of the study variable is simplified to the correct rate. The correct rate p_r of the samples can be calculated by

$$p_r = \frac{r}{n} = \frac{1}{n} \sum_{i=1}^{n} Y_i \qquad (17)$$

where r is the number of correct samples. According to the statistical theory, the sample correct rate p_r is an unbiased estimator P_r of the population

3) to comprehensively evaluate the performance of this experiment under different sample sizes, the W% values were designed as 2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18%, and 20%, respectively. The upper limit of sample size is set to 20% of the population is that, in RSGD quality evaluation, it is very high cost to reach 20%. Usually, the sample size will not exceed 20%; on the other hand, repeating the experiment 50 times can reduce the sampling error, make the experimental results closer to the reference data, and avoid insufficient reliability of the conclusion due to too few observed values [61].

C. Experimental Data

The population and samples: The population sizes of the six maps are 1778, 2514, 2811, 1493, 1197, and 1299, respectively, as shown in Table I. To ensure a fair comparison, it is necessary for the sample size to be consistent across the three sampling methods. The specific sample size of each map is shown in Table I.

The reference data: Referring to Google data and OSM data, we evaluated experimental data quality by means of expert judging all elements of each map. Each element was judged by three experts independently, and two or more consistent results were deemed to be correct, otherwise, it was wrong. According to Section III-B and (17), we calculated the population reference correct rates (or "true" correct rates) of each map (see Table II).

D. Stratification Result Based on Fractal

With the help of ArcGIS software, the experimental data were processed, and the fractal dimension of each element in the six maps was calculated according to (6). The classifications of Map A, B, C, D, E, and F was 6, 5, 5, 5, 6, and 6. Therefore, when the SSF method was adopted, the number of strata of the six maps also took the same value. By the cumulative square root of frequency method, we calculated the cumulative square root values of the six maps, which were 244.4316, 283.7608, 238.3648, 303.4573, 203.7184, and 210.4429, respectively. They were stratified according to the number of strata above. The stratification results of the two methods are shown in Fig. 5.

E. Sampling and Assessment

After obtaining the stratification results, according to (12), the proportional allocation was used to calculate the sample size of each stratum under the ten categories of sample size designed in Section III–B (2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18%, and 20%) respectively. In the SSF, the fractal dimension was used as the stratification indicator. The population of each map was divided according to Table I, and then samples were taken at each stratum through the random principle.

This experiment was a comparative analysis of the gain of different methods, which needed to adopt the same quality evaluation principle. Through the quality evaluation model and reference data, the quality inspection items of each sample in maps A, B, C, D, E, and F were evaluated [62]. Equation (17) can calculate the correct rate of each sampling ratio. We repeated the above process 50 times to obtain a group of observed values of the correct rate and then compared the reference data to calculate the DEFF and RMSE of the correct rate (see Table III).

IV. DISCUSSION

Fig. 6 shows the average DEFF values of the SSF method, the SSC method, and the SRS method in each experimental map. The x-axis represents the map name, the y-axis represents the DEFF value. The black solid line, blue solid-dotted line, and the red dotted line represent the ranges of DEFF values under the SSF method, the SSC method, and the SRS method, respectively.

1) It can be seen from Fig. 6 that the SSF method's DEFF result is less than 1, indicating the effectiveness of the SSF method.
2) All the SSF method's average DEFF values are less than the SSC method's and the SRS method's, indicating the advantages of the SSF method.
3) The DEFF range of the SSF method is smaller than that of the SSC method in each map, indicating that the SSF method is less influenced by the sample size.
4) The average DEFF fluctuation of the SSF method is less than that of the SSC method, indicating that the SSF method is less affected by the spatial location.

Fig. 7 shows the experimental RMSE results of the SSF method, the SSC method, and the SRS method in the six experimental maps. The x-axis represents the sampling ratio, the y-axis represents the RMSE, connecting the RMSE values of the same method with different sample ratios. The black solid line represents the SSF method, the blue solid-dotted line represents the SSC method, and the red dotted line represents the SRS method.

1) By comparing the black solid line and the red dotted line, it is evident that the RMSE of the SSF method is consistently lower than that of the SRS method across all sampling ratios. It indicates that the SSF method proposed in this article is significantly better than the SRS method.
2) Through the comparison between the black solid line and the blue solid-dotted line, it is observed that the RMSE of the SSC method is slightly lower than that of the SSF method, except for the 10% sample rate in Fig. 7(a), the 10% sample rate in Fig. 7(c), the 4% sample rate in Fig. 7(d), and the 16% sample rate in Fig. 7(e). Most other experimental results show that the RMSE of the SSF method is lower than that of the SSC method. It shows that the SSF method proposed in this article is better than the commonly used SSC method.
3) It can be seen from the three lines that with the increase of sampling ratio, the RMSE of the three methods decreases and gradually converges to 0, among which the SSF method converges the fastest.
4) When the RMSE value is same, the sampling ratio corresponding to the SSF method is the smallest, indicating that the SSF method is the most sensitive to the sample size.
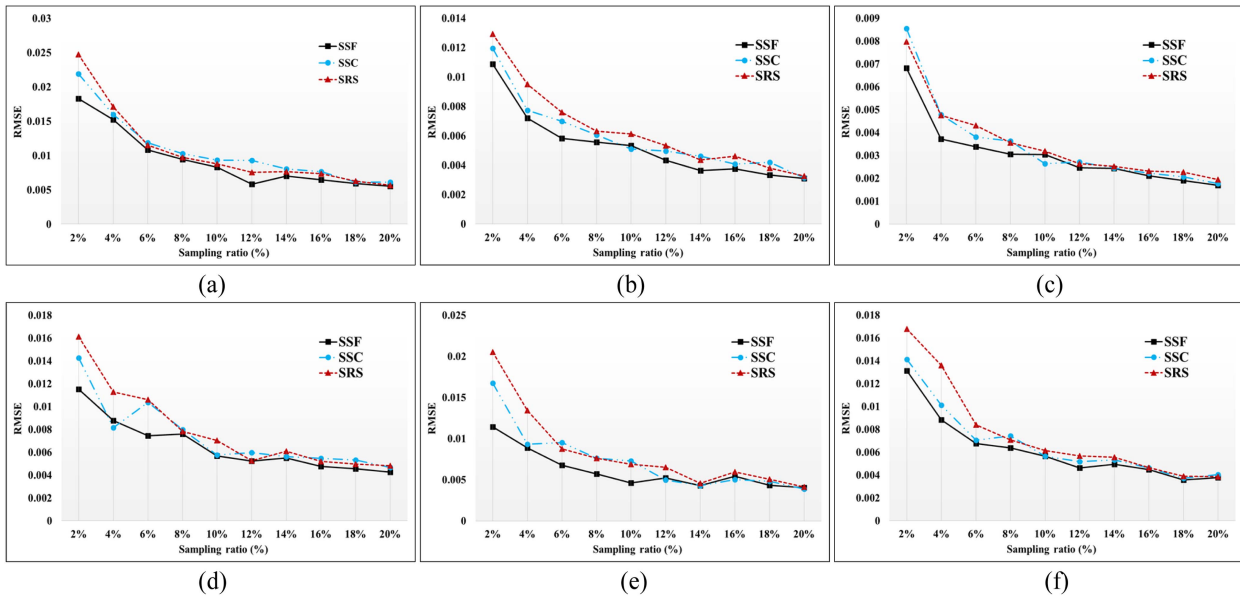
Fig. 7.   (a) Represents the RMSE results using three different sampling methods for quality evaluation when the sampling ratio in map A is 2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18%, and 20% respectively. (b)–(f) RMSE results of map B, C, D, E, and F, respectively.

V. CONCLUSION

Whether in the quality assessment, accuracy verification, or statistical investigation of RSGD, the stratified sampling method is economical and efficient [63]. The SSC method, which is stratified based on classifications of landcover elements, is the most widely used sampling method at present. However, the problem of this method is that the correlation between classifications of landcover elements and quality is not clear, and the improvement of sampling efficiency is limited.

This article demonstrates the feasibility of the proposed SSF method and presents the conceptual flow of the method. In the SSF method, the box-counting method is used to quantitatively characterize the fractal structure of elements in RSGD. The fractal dimension calculated by the box-counting method is used as the stratification indicator, the stratification boundary value is optimized by the cumulative square root of frequency method, and then the sample allocation is optimized by the proportional allocation method.

To validate the effectiveness of the SSF method, this article conducted an experiment using the South Sudan GCVD 2020 dataset across various spatial locations. Through the analysis of the experimental results, the following conclusions were drawn.

1) Based on DEFF, the SSF method designed in this article was compared with the SRS method. The experimental data analysis results showed that the DEFF values were less than 1, which verified the effectiveness of fractal dimension as a stratification indicator.

2) Based on RMSE, the SSF method proposed in this article was compared with the SSC method and the SRS method. The results showed that when the sample size was same, the SSF method could provide more accurate quality evaluation results. It proves that the SSF method is better than widely used the SSC method and the SRS method.

3) The SSF method proposed in this article is more sensitive to the sample size, especially small sample size, the performance of the SSF method is more significant, which proves that this method is more economical and efficient than the SSC method and the SRS method.

It should be noted that the proposed SSF method was tested exclusively on a single remote sensing dataset for the experiment conducted in this article. The multiple remote sensing datasets that vary in terms of bands and resolutions pose a challenge to the applicability of our method. Further research will concentrate on evaluating the representativeness of samples extracted using the proposed SSF method across multiple remote sensing datasets.

Furthermore, the proposed SSF method is currently only suitable for vector data. Further research will focus on whether the fractal dimension calculation in our proposed method is suitable for raster data, whether the element-based representation of samples works well with raster data, and whether the samples extracted by the proposed method are representative of the quality information of raster data.

Finally, the experimental study of the proposed SSF method in South Sudan is helpful to the global RSGD quality assessment.

REFERENCES

[1] H. Xie, X. Tong, W. Meng, D. Liang, Z. Wang, and W. Shi, "A multilevel stratified spatial sampling approach for the quality assessment of remote-sensing-derived products," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4699–4713, Oct. 2015.

[2] B. Pengra et al., "A global reference database from very high resolution commercial satellite data and methodology for application to Landsat derived 30 m continuous field tree cover data," *Remote Sens. Environ.*, vol. 165, pp. 234–248, 2015.

[3] L. Boschetti, S. V. Stehman, and D. P. Roy, "A stratified random sampling design in space and time for regional to global scale burned area product validation," *Remote Sens. Environ.*, vol. 186, pp. 465–478, Dec. 2016.

[4] J. Wickham et al., "Thematic accuracy assessment of the 2011 national land cover database (NLCD)," *Remote Sens. Environ.*, vol. 191, pp. 328–341, 2017.

[5] J.-F. Girres and G. Touya, "Quality assessment of the french Open-StreetMap dataset," *Trans. GIS*, vol. 14, no. 4, pp. 435–459, 2010.

[6] P. Sun et al., "A quantitative performance comparison of paddy rice acreage estimation using stratified sampling strategies with different stratification indicators," *Int. J. Digit. Earth*, vol. 11, no. 10, pp. 1001–1019, 2017.

[7] M. Broich et al., "A comparison of sampling designs for estimating deforestation from Landsat imagery: A case study of the Brazilian legal Amazon," *Remote Sens. Environ.*, vol. 113, no. 11, pp. 2448–2454, Nov. 2009.

[8] S. V. Stehman et al., "A global land-cover validation data set, II: Augmenting a stratified sampling design to estimate accuracy by region and land-cover class," *Int. J. Remote Sens.*, vol. 33, no. 22, pp. 6975–6993, 2012.

[9] Y. Wan et al., "A general framework for spatial data inspection and assessment," *Earth Sci. Inform.*, vol. 8, no. 4, pp. 919–935, 2015.

[10] B. P. M. Duarte and P. M. Saraiva, "An optimization-based approach for designing attribute acceptance sampling plans," *Int. J. Qual. Rel. Manage.*, vol. 25, no. 8, pp. 824–841, 2008.

[11] J.-F. Wang et al., "A review of spatial sampling," *Spatial Statist.*, vol. 2, pp. 1–14, 2012.

[12] E. Johnsen, R. M. Cyert, and H. J. Davidson, *Statistical Sampling for Accounting Information*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1962,.

[13] W. G. Cochran, *Sampling Techniques*, 3th ed. Hoboken NJ, USA: Wiley, 1977.

[14] S. V. Stehman, T. L. Sohl, and T. R. Loveland, "Statistical sampling to characterize recent United States land-cover change," *Remote Sens. Environ.*, vol. 86, no. 4, pp. 517–529, 2003.

[15] M. Padilla et al., "Comparing the accuracies of remote sensing global burned area products using stratified random sampling and estimation," *Remote Sens. Environ.*, vol. 160, pp. 114–121, 2015.

[16] S. V. Stehman et al., "Adapting a global stratified random sample for regional estimation of forest cover change derived from satellite imagery," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 650–658, 2011.

[17] N. Giree et al., "A sample-based forest monitoring strategy using landsat, AVHRR and MODIS data to estimate gross forest cover loss in Malaysia between 1990 and 2005," *Remote Sens.*, vol. 5, no. 4, pp. 1842–1855, 2013.

[18] P. Olofsson et al., "A global land-cover validation data set, part I: Fundamental design principles," *Int. J. Remote Sens.*, vol. 33, no. 18, pp. 5768–5788, 2012.

[19] E. Fabrizi and C. Trivisano, "Efficient stratification based on nonparametric regression methods," *J. Official Statist.*, vol. 23, no. 1, pp. 35–50, 2007.

[20] P. A. Golder and K. A. Yeomans, "The use of cluster analysis for stratification," *Appl. Statist.*, vol. 22, no. 2, pp. 213–219, 1973.

[21] K. G. Reddy, "A procedure for computing optimal stratum boundaries and sample sizes for multivariate surveys," *J. Softw.*, vol. 11, no. 8, pp. 816–832, 2016.

[22] T. Zseby, "Stratification strategies for sampling-based non-intrusive measurement of one-way delay," in *Proc. Passive Act. Meas. Workshop*, 2003, pp. 171–179.

[23] D. Hedlin, "A procedure for stratification by an extended Ekman rule," *J. Official Statist.*, vol. 16, no. 1, pp. 15–29, 2000.

[24] S. V. Stehman and G. M. Foody, "Key issues in rigorous accuracy assessment of land cover products," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111199.

[25] B. Mandelbrot, "How long is the coast of Britain? Statistical self-similarity and fractional dimension," *Science*, vol. 156, no. 3775, pp. 636–638, May 1967.

[26] R. F. Voss, "Fractals in nature: From characterization to simulation," in *The Science of Fractal Images*. Berlin, Germany: Springer, 1988, pp. 21–70.

[27] W. Sun et al., "Fractal analysis of remotely sensed images: A review of methods and applications," *Int. J. Remote Sens.*, vol. 27, no. 22, pp. 4963–4990, 2007.

[28] N. Lam, "Description and measurement of Landsat TM images using fractals," *Photogrammetric Eng. Remote Sens.*, vol. 56, no. 2, pp. 187–195, 1990.

[29] J. J. T. P. R. Hogg, *Quantitative Remote Sensing of Land Surfaces*. Hoboken NJ, USA: Wiley, 2004.

[30] B. Jiang and S. A. Brandt, "A fractal perspective on scale in geography," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 6, p. 95, 2016.

[31] F. Bretar et al., "An advanced photogrammetric method to measure surface roughness: Application to volcanic terrains in the Piton de la Fournaise, Reunion Island," *Remote Sens. Environ.*, vol. 135, pp. 1–11, 2013.

[32] Y. C. Cheng, P. J. Lee, and T. Y. Lee, "Self-similarity dimensions of the Taiwan Island landscape," *Comput. Geosci.*, vol. 25, no. 9, pp. 1043–1050, 1999.

[33] M. H. Cosh and W. Brutsaert, "Microscale structural aspects of vegetation density variability," *J. Hydrol.*, vol. 276, no. 1–4, pp. 128–136, 2003.

[34] J. M. Read and N. S. N. Lam, "Spatial methods for characterising land cover and detecting land-cover changes for the tropics," *Int. J. Remote Sens.*, vol. 23, no. 12, pp. 2457–2474, 2010.

[35] F. J. Ferrandino, "Measuring spatial aggregation in binary epidemics: Correlative analysis and the advantage of fractal-based sampling," *Phytopathology*, vol. 94, no. 11, pp. 1215–1227, 2004.

[36] C. J. Marsh and R. M. Ewers, "A fractal-based sampling design for ecological surveys quantifying β-diversity," *Methods Ecol. Evol.*, vol. 4, no. 1, pp. 63–72, 2013.

[37] C. Jiang et al., "PCantorSim: Accelerating parallel architecture simulation through fractal-based sampling," *ACM Trans. Archit. Code Optim.*, vol. 10, no. 4, pp. 1–24, 2013.

[38] S. O. Samuelsen, H. Anestad, and A. Skrondal, "Stratified case-cohort analysis of general cohort sampling designs," *Scand. J. Statist.*, vol. 34, no. 1, pp. 103–119, Mar. 2007.

[39] C. Wu and M. E. Thompson, "Stratified sampling and cluster sampling," in *Sampling Theory and Practice*. Berlin, Germany: Springer, 2020, pp. 33–56.

[40] M. Padilla et al., "Stratification and sample allocation for reference burned area data," *Remote Sens. Environ.*, vol. 203, pp. 240–255, Dec. 2017.

[41] Y. Xu, D. Liu, Y. Quan, and P. Le Callet, "Fractal analysis for reduced reference image quality assessment," *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2098–2109, Jul. 2015.

[42] G. Shen, "Fractal dimension and fractal growth of urbanized areas," *Int. J. Geographical Inf. Sci.*, vol. 16, no. 5, pp. 419–437, 2002.

[43] S. W. Myint, "Fractal approaches in texture analysis and classification of remotely sensed data: Comparisons with spatial autocorrelation techniques and simple descriptive statistics," *Int. J. Remote Sens.*, vol. 24, no. 9, pp. 1925–1947, 2010.

[44] N. S. N. Lam et al., "Effects of landscape fragmentation on land loss," *Remote Sens. Environ.*, vol. 209, pp. 253–262, 2018.

[45] L. Engelmann et al., "Towards the suitability of information entropy as an LES quality indicator," *Flow, Turbulence Combustion*, vol. 108, no. 2, pp. 353–385, 2021.

[46] M. Cholewa and B. Placzek, "Application of positional entropy to fast Shannon entropy estimation for samples of digital signals," *Entropy (Basel)*, vol. 22, no. 10, Oct. 2020, Art. no. 1173.

[47] V. Majerník, E. Majerníková, and S. Shpyrko, "Uncertainty relations expressed by Shannon-like entropies," *Open Phys.*, vol. 1, no. 3, pp. 393–420, 2003.

[48] Y. Sayyari, "New refinements of Shannon's entropy upper bounds," *J. Inf. Optim. Sci.*, vol. 42, no. 8, pp. 1869–1883, 2021.

[49] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 4, pp. 623–656, 1948.

[50] M.-A. Aitouche, M. Djeddi, and K. Baddari, "Fractal variogram-based time–space of aftershock sequences analysis—Case study: The May 21, 2003 Boumerdes–Algeria earthquake, M w = 6.8," *Arabian J. Geosci.*, vol. 6, no. 7, pp. 2183–2192, 2012.

[51] M. Lozada-Zumaeta et al., "Distribution of petrophysical properties for sandy-clayey reservoirs by fractal interpolation," *Nonlinear Processes Geophys.*, vol. 19, no. 2, pp. 239–250, 2012.

[52] B. Klinkenberg, "A review of methods used to determine the fractal dimension of linear features," *Math. Geol.*, vol. 26, no. 1, pp. 23–46, 1994.

[53] J. R. Carr and W. B. Benzer, "On the practice of estimating fractal dimension," *Math. Geol.*, vol. 23, no. 7, pp. 945–958, 1991.

[54] I. Rodrguez-Iturbe and A. Rinaldo, *Fractal River Basins: Chance and Self-Organization*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[55] G. R. Angeles et al., "Fractal analysis of tidal channels in the Bahía Blanca Estuary (Argentina)," *Geomorphology*, vol. 57, no. 3-4, pp. 263–274, 2004.

[56] J. A. Goff, "Comment on 'Fractal mapping of digitized images: Application to the topography of Arizona and comparison with synthetic images' by J. Huang and D. L. Turcotte," *J. Geophys. Res., Solid Earth*, vol. 95, no. B4, p. 5159, 1990.

[57] W. L. Power and T. E. Tullis, "Euclidean and fractal models for the description of rock surface roughness," *J. Geophysical Res.*, vol. 96, no. B1, pp. 415–424, 1991.

[58] J. Huang and D. L. Turcotte, "Fractal mapping of digitized images: Application to the topography of Arizona and comparisons with synthetic images," *J. Geophysical Res.*, vol. 94, no. B6, pp. 7491–7495, 1989.
[59] W. A. Fuller, "Sampling with random stratum boundaries," *J. Roy. Stat. Soc., Ser. B (Methodological)*, vol. 32, no. 2, pp. 209–226, 1970.
[60] L. Kish, *Survey Sampling*. Hoboken, NJ, USA: Wiley, 1995.
[61] K. L. Chung and K. Zhong, *A Course in Probability Theory*. New York, NY, USA: Academic, 2001.
[62] *Geographic Information - Data Quality*, Standard ISO 19157:2013, International Organization for Standardization, Geneva, Switzerland, 2013.
[63] W. G. Cochran, *Sampling Techniques*. Hoboken, NJ, USA: Wiley, 2007.

Yao Lu received the M.S. degree in signal and information processing from the Aerospace Information Research Institute, Chinese Academy of Science, Beijing, China, in 2014. He is currently working the Ph.D. degree in surveying and mapping with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China.

His research interests include sampling methods and GIS spatial data quality control.

Huan Xie (Senior Member, IEEE) received the B.S. degree in surveying engineering and the M.S. and Ph.D. degrees in cartography and geoinformation from the Tongji University, Shanghai, China, in 2003, 2006, and 2009, respectively.

From 2007 to 2008, she was with the Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Hannover, Germany, funded by the China Scholarship Council, as a Visiting Scholar. Her research interests include satellite laser altimetry and hyperspectral remote sensing.

Jixian Zhang received the B.E., M.S., and Ph.D. degrees in in photogrammetry and remote sensing from the Department of Photogrammetry and Remote Sensing, Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1988, 1991, and 1994, respectively.

He is currently a Chairperson with the National Quality Inspection and Testing Center for Surveying and Mapping Products, Beijing, China. He is also an Adjunct Professor with Wuhan University, Wuhan, China and the China University of Mining and Technology, Beijing, China. He has authored/edited ten books, obtained 11 patents of invention, and authored/coauthored more than 300 research papers. His research interests include photogrammetry, remote sensing, and resources and environment monitoring by remote sensing.

Yanmin Jin received the B.S. degree in geographic information system from the Taiyuan University of Technology, Taiyuan, China, in 2007, and the Ph.D. degree in cartography and geographic information engineering from the Tongji University, Shanghai, China, in 2015, respectively.

Since 2022, she has been an Associate Professor with Tongji University. Her research interests include GIS spatial data processing and data quality control.

Yongjiu Feng received the Ph.D. degree in cartography and geographic Information engineering from the Tongji University, Shanghai, China, in 2009.

He was an Honorary Associate Professor with The University of Queensland, Brisbane, QLD, Australia. From 2015 to 2016, he was a Visiting Academic with the University of Queensland. He is currently a Professor with the College of Surveying and GeoInformatics, Tongji University. His research interests include spatial analysis and remote sensing image processing.

Yali Gong received the M.S. degree in surveying and mapping from the Tongji University, Shanghai, China, in 2017. She is currently working the Ph.D. degree in surveying and mapping with the College of Surveying and Geoinformatics, Tongji University, Shanghai, China.

Her research interests include accuracy assessment of land cover, land change data, and sampling methods.

Wenli Han received the B.S. degree in photogrammetry and remote sensing from the Department of Photogrammetry and Remote Sensing, Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1998.

Her research interests include GIS spatial data processing and data quality control.

He Zhang received the M.S. degree in cartography and geographical information system from the Wuhan University, Wuhan, China, in 2005, She is currently working toward the Ph.D. degree in energy and environmental protection with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China.

Her research interests include quality intelligent inspection of natural resources investigation and monitoring and data quality control.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree in traffic engineering from the Tongji University, Shanghai, China, in 1999.

From 2001 to 2003, he was a Post-Doctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, from 2008 to 2009. His research interests include photogrammetry and remote sensing, trust in spatial data, and image processing for high-resolution satellite images.