Journal of Global Change Data & Discovery. 2024, 8(1): 14-20
www.geodoi.ac.cn

Global Change Research Data Publishing & Repository

Remote Sensing Rao's Q Index Yearly Forest Dataset of China (2000–2017)

Jiang, X.[1,2] Cai, H. Y.[1] Yang, X. H[1,2*]

1. State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The remote sensing Rao's Q index can characterize the functional diversity of macro-forest plants, making it a crucial metric for evaluating ecological quality and essential to efficiently carry out tasks related to regional biodiversity protection. According to the traditional definition of Rao's Q index, the forest plant features are represented by pixel values in the spectral difference of the normalized difference vegetation index (NDVI), and the neighborhood pixel values are used to generate the distance matrix. The yearly remote sensing Rao's Q index dataset of forest in China from 2000 to 2017 was calculated on the R language platform. The temporal resolution of the data was annual, the spatial resolution was 5 km, and the projection was based on Albers Conical Equal Area with the coordinate system of WGS-84. The dataset is archived in the .tif format and consists of 72 data files with data size of 58.2 MB (compressed into one file with 5.17 MB).

Keywords: remote sensing; Rao's Q index; 2000–2017

DOI: https://doi.org/10.3974/geodp.2024.01.02

CSTR: https://cstr.escience.org.cn/CSTR:20146.14.2024.01.02

Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Repository* at: https://doi.org/10.3974/geodb.2024.03.08.V1 or https://cstr.escience.org.cn/CSTR:20146.11.2024.03.08.V1.

1 Introduction

Biodiversity is the core and foundation of ecosystem services. Globally, biodiversity is undergoing significant changes[1], necessitating the immediate clarification of its temporal and spatial variations to formulate and implement effective biodiversity conservation and management decisions. To guide governments at all levels to strengthen biodiversity

Received: 05-01-2024; **Accepted:** 15-03-2024; **Published:** 25-03-2024

Foundations: Ministry of Science and Technology of the P. R. China (2023FY101000, 2017YFC0503803)

***Corresponding Author:** Yang, X. H. AAC-8887-2021, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences; University of Chinese Academy of Sciences, yangxh@igsnrr.ac.cn

Data Citation: [1] Jiang, X., Cai, H. Y., Yang, X. H. Remote sensing Rao's Q index yearly forest dataset of China (2000–2017) [J]. *Journal of Global Change Data & Discovery*, 2024, 8(1): 14–20. https://doi.org/10.3974/geodb.2024.01.02. https://cstr.escience.org.cn/CSTR:20146.14.2024.01.02.

[2] Jiang, X., Cai, H. Y., Yang, X. H. Yearly remote sensing Rao's Q index dataset of forest in China (2000–2017) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. https://doi.org/10.3974/geodb.2024.03.08.V1. https://cstr.escience.org.cn/CSTR:20146.11.2024.03.08.V1.

protection and curb the trend of biodiversity loss and ecosystem degradation, the Ministry of Ecology and Environment recently included biodiversity indicators for the first time into the comprehensive evaluation indicator framework of ecological quality. Against this background, remote sensing has emerged as a crucial tool for biodiversity research due to its ability to provide multiscale, large scale, as well as precise temporal and spatial heterogeneity data[2].

The spectral variation hypothesis has emerged as a significant theory in remote sensing biodiversity research[3–5]. Building upon this theory, Rocchini proposed the remote sensing Rao's Q index in 2017. Specifically, remote sensing pixel values represent functional traits while also considering the impact range of ecological processes to effectively identify differences within a single ecosystem[6, 7]. Its strength lies in monitoring spatiotemporal changes, compensating for traditional biodiversity research shortcomings in large scale quantification over a short period. This index has therefore been extensively employed to investigate the diversity of forest plants in various climatic zones[8, 9] and can effectively characterize the spatiotemporal heterogeneity of Hainan's mangroves[10]. However, comprehensive forest plant diversity datasets covering extensive regional scales and diverse climatic zones are absent in China. Moreover, research focusing on macro-scale forest plant diversity remains scarce. In this context, the remote sensing Rao's Q index is introduced in this study to comprehensively analyze the spatial distribution of forest plant diversity in China and its temporal dynamics[11], with the overall aim to offer insights for guiding overall biodiversity conservation planning and ecological quality assessment.

2 Metadata of the Dataset

Table 1 provides the details of yearly remote sensing Rao's Q index dataset of forest in China (2000–2017)[12]. It includes the dataset full name, short name, authors, year of the dataset, temporal resolution, spatial resolution, data format, data size, data files, data publisher, and data sharing policy, etc.

3 Methods

3.1 Algorithm Principle

Rao's Q index, also known as Rao's quadratic entropy index, is a widely used functional diversity measure in ecological studies, particularly for characterizing variations in functional traits[14]. In 2017, Rocchini introduced the remote sensing Rao's Q index algorithm by integrating remote sensing data with the conventional Rao's Q index[6]. Specifically, the pixel value represents the functional trait, and the proportion of a certain pixel value represents the relative abundance of the trait. Greater uniformity is indicated by closer relative abundances within the window. The distance matrix is created using the disparity in adjacent pixel values; greater discrepancy in pixel values corresponds to greater distance. A higher Rao's Q index signifies more pronounced distinctions in functional traits and greater diversity. The index is calculated as follows:

$$d_{ij} = \frac{1}{n}\sum_{k=1}^{n}\left(X_{ik} - X_{jk}\right)^2 \tag{1}$$

$$Q = \sum_{i=1}^{S-1}\sum_{j=i+1}^{S} d_{ij}\,p_i\,p_j \tag{2}$$

Table 1 Metadata summary of the Yearly remote sensing Rao's Q index dataset of forest in China (2000–2017)

Items	Description		
Dataset full name	Yearly remote sensing Rao's Q index dataset of forest in China (2000–2017)		
Dataset short name	ChinaForest_Rao		
Authors	Jiang, X. AAE-1541-2021, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences; University of Chinese Academy of Sciences, jiangx.20b@igsnrr.ac.cn Cai, H.Y. Y-8555-2019, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, caihy@igsnrr.ac.cn Yang, X. H. AAC-8887-2021, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences; University of Chinese Academy of Sciences, yangxh@igsnrr.ac.cn		
Geographical region	China	Year	2000–2017
Temporal resolution	yearly	Spatial resolution	5 km
Data format	.tif	Data size	5.17 MB (in compression)
Data files	The dataset contains 1 folder: "ChinaForest_Rao" contains the yearly remote sensing Rao's Q index dataset of forest in China (2000–2017) in .tif format. This dataset comprises 72 data files, with file names containing time and phase information. For instance, "ChinaForest_Rao_2000.tif" represents the remote sensing Rao's Q index of China forest in 2000, with a spatial resolution of 5 km. A higher value indicates greater forest diversity		
Foundations	Ministry of Science and Technology of the People's Republic of China (2023FY101000，2017YFC0503803)		
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn		
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China		
Data sharing policy	(1) *Data* are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use *Data* subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute *Data* subject to written permission from the GCdataPR Editorial Office and the issuance of a *Data* redistribution license; and (4) If *Data* are used to compile new datasets, the 'ten per cent principal' should be followed such that *Data* records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset[13]		
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS		

In these equations, Q stands for Rao's Q index, d_{ij} represents the difference in traits i and j, denoted as the dissimilarity between adjacent pixels i and j, n is the total number of pixels in the sliding window, S is the number of pixel values in the sliding window, p_i, p_j are the relative abundances of traits i and j, expressed as the proportion of pixel values of i and j in the sliding window. In this study, Rao's Q index was calculated using the RStudio platform.

3.2 Methodology

The technical route is shown in Figure 1. The dataset preparation process involves the following steps: (1) Preprocessing of NDVI (normalized difference vegetation index) data. Annual NDVI data were obtained from the Resources and Environmental Science and Data Center of the Chinese Academy of Sciences. To facilitate the application of Rao's Q algorithm, the NDVI range was converted from 0–1 to 0–10,000, and the spatial resolution was resampled to 5 km × 5 km. (2) The spatial range of forest was determined. The land cover data acquired from the Space Information Innovation Institute of the Chinese Academy of Sciences were employed for forest cover extraction, and the forest NDVI was obtained by clipping the NDVI. (3) Rao's Q index calculation. The forest Rao's Q index was computed on the RStudio platform using the Rao's Q index algorithm. (4) Data verification. Spatial and numerical comparison of the consistency between Rao's Q index and gymnosperm species richness was conducted. (5) Forest Rao's Q index dataset generation.

4 Data Results and Data Validation

4.1 Dataset Composition

The naming method, data description, data format, number of files and data size of the constituent files of yearly remote sensing Rao's Q index dataset of forest in China (2000–2017) are shown in Table 1.

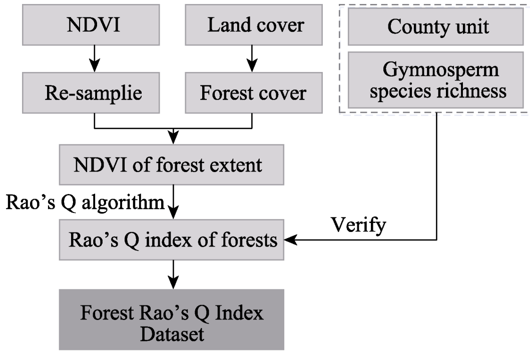

Figure 1 Flowchart of dataset generation process

4.2 Spatial Distribution of Rao's Q Index in Chinese Forests

Figure 2 shows the spatial distribution of Rao's Q index of China's forests for various years, with values from 0.25–2,296. High values were found for the Tianshan Mountains, Luliang Mountains, south of the Himalayas, the Hengduan Mountains, and the Wuyi Mountains-Nanling Mountains, whereas lower values are observed for the Lesser Khingan, Greater Khingan, and Changbai Mountain regions. Consequently, the high-value regions exhibit great variabilities in forest functional traits and boast rich forest plant diversity. Conversely, the corresponding forest plant diversity in the northeastern regions is limited.

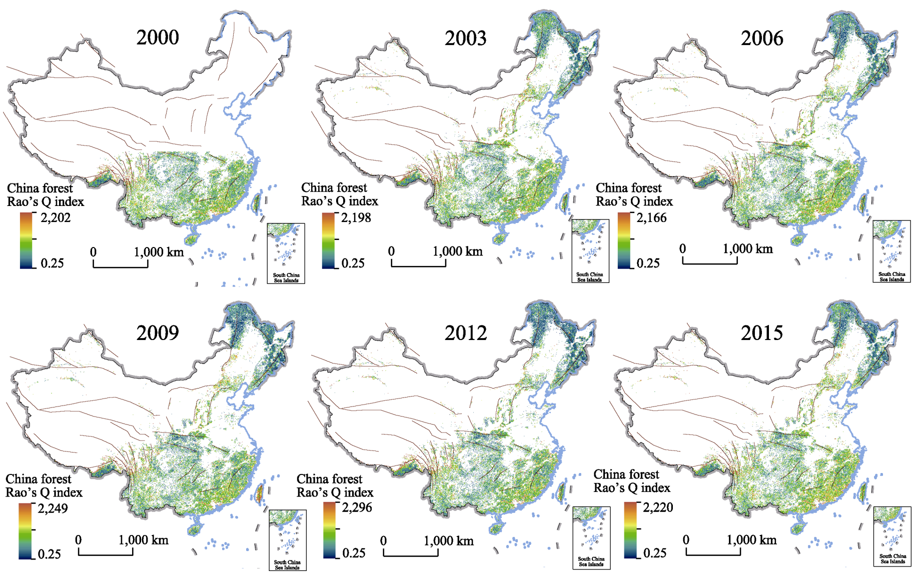

Figure 2 Maps of spatial distribution of Rao's Q index of forests in China

4.3 Data Validation

Large-scale field surveys of biodiversity that take into account temporal and spatial changes face enormous challenges. Biodiversity data across the country are extremely limited, and only spatial distribution data on gymnosperm species richness based on counties as statistical units were available in 2009[15]. Pinaceae, Cupressaceae and Taxaceae are not only the main families of evergreen coniferous forests, but also typical representatives of gymnosperms. The spectral properties of gymnosperms and evergreen coniferous woods are remarkably uniform. Therefore, Rao’s Q index of evergreen coniferous forests was compared with the species richness of gymnosperms from the aspects of spatial distribution and statistics. In terms of space, to be consistent with the species richness statistical unit, Rao’s Q index was first calculated and spatialized at the county level, and subsequently, the two were compared as shown in Figure 3. Approximately 5,400 valid samples were obtained by randomly generating sample sites throughout the entire national forest and extracting the matching Rao’s Q index and species richness. The data are discrete, with a species richness range of 0–35. Rao’s Q index is based on continuous data, ranging from 0.25 to 2,269. There is a range associated with each species′ richness value and the Rao’s Q index value. Table 2 shows the comparison of some species richness values with the maximum, minimum, and mean values of Rao’s Q index. Only the mean values are shown as the median, mode, and mean were close. Rao’s Q algorithm for remote sensing can only accept integers as input data. In this study, NDVI values between 1 and 10,000 are used as input data. The disparity in pixel values reflects highly distinct trait features. Furthermore, inside the 3 × 3 sliding window, there are variations from low to high uniformity. Rao’s Q index for distant sensing, thus, has a larger range of values and more pronounced functional characteristic differences. Additionally, it has a spatial resolution of 5 km, and species richness is a county-level statistical unit. As a result, the minimum and maximum values of Rao’s Q index for n pixels within the same county unit differ significantly.

The spatial distribution of species richness is essentially compatible with Rao’s Q index of evergreen coniferous forests. Whilst the low-value areas are mostly found in the northeast, the high-value areas of both are concentrated in the Wuyi Mountains-Nanling and the southeast Himalayas-Hengduan Mountains. In statistical terms, the highest value of Rao’s Q index fluctuates randomly as species richness rises, and the minimum and mean values gradually increase. The mean of Rao’s Q index increases dramatically, especially when the species richness exceeds 15, leading experts to assume that the mean is more representative. Regression analysis modeling using species richness as the independent variable and the mean Rao’s Q index as the dependent variable showed that the two were highly fitted (R^2 = 0.66, P < 0.001). In summary, Rao’s Q index can adequately characterize forest plant diversity.

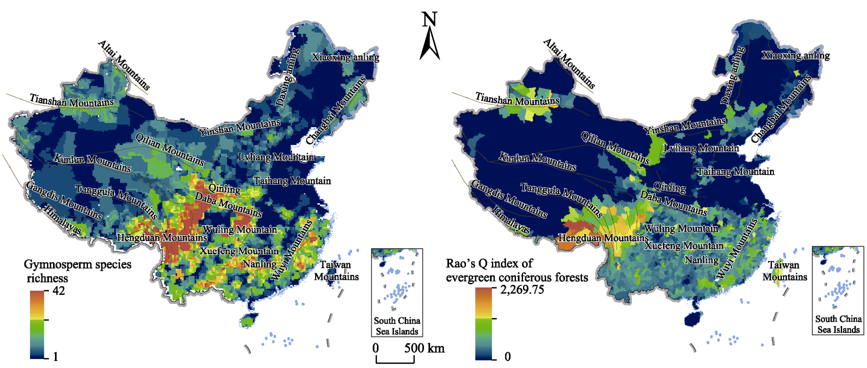

Figure 3 Spatial comparison maps of gymnosperm species richness with Rao’s Q index in evergreen coniferous forest

Table 2 Comparison of gymnosperm species richness and Rao's Q index values of evergreen coniferous forests

gymnosperm species richness	Rao's Q index			gymnosperm species richness	Rao's Q index		
	Max	Min	Mean		Max	Min	Mean
1	1,797.13	0.25	290.91	19	1,948.00	56.40	525.82
3	1,901.88	17.00	326.31	21	1,390.89	17.25	484.65
5	1,405.78	6.75	286.56	23	1,259.11	18.22	420.01
7	1,640.00	35.11	296.00	25	1,211.12	95.00	460.75
9	1,948.20	21.11	318.20	27	1,135.6	114.25	510.12
11	2,095.50	78.75	315.14	29	1,481.56	25.56	482.08
13	1,477.25	59.00	281.84	31	1,481.56	166.14	456.28
15	1,225.11	23.25	295.73	⩾32	894.06	83.55	397.43
17	1,923.56	26.50	399.24				

5 Discussion and Conclusion

In this study, the remote sensing Rao's Q index is introduced, marking the first application of this index to characterize the diversity of forest plants in China. The Rao's Q index dataset of forest remote sensing in China covers the period from 2000 to 2017, with a spatial resolution of 5 km and a temporal resolution of yearly. It provides a fresh monitoring indicator for regional ecological quality assessment, supporting the formulation of relevant policies for forest plant diversity protection and the enhancement of regional ecological quality. Remote sensing technology has greatly promoted the study of biodiversity at the macro scale, and the remote sensing Rao's Q index algorithm is based on the definition of Rao's Q index in ecology. To create a distance matrix that characterizes the spatial range of ecological processes, it makes use of neighborhood pixels and a sliding window[16]. This dataset only covers Chinese forests and uses the NDVI as the data source. However, Rao's Q index values simulated by other remote sensing vegetation indices have been applied to tropical, temperate, and subtropical regions, and this index is not limited to forest ecosystems[8, 9]. In the early stage of the study, the leaf area index, the enhanced vegetation index, and the ratio vegetation index were used as data sources to simulate the LAI-Rao's Q index, EVI-Rao'Q index, and RVI-Rao's Q index of Chinese forests. By comparison, the NDVI-Rao's Q index had a higher consistency with species richness, followed by the LAI-Rao's Q index, whereas the relationship between the EVI-Rao's Q index and the RVI-Rao's Q index and species richness was not obvious. Notably, the remote sensing Rao's Q index is not equivalent to species richness, and here, it only represents the forest plant diversity at the macro scale. At present, the remote sensing Rao's Q index is still in the exploratory stage, and more detailed biodiversity data are required to verify it and to further explore the ecological information it represents.

Author Contributions

Jiang, X. is responsible for the paper writing, the processing of Rao's Q index data as well as data analysis; Cai, H. Y. is responsible for overall design for dataset production, and paper writing; Yang, X. H. carried out for the review of data processing methods, data quality control and manuscript improvement.

Conflicts of Interest

The authors report no conflicts of interest.

References

[1] Pimm, S. L., Russell, G. J., Gittleman, J. L., *et al*. The future of biodiversity [J]. *Science*, 1995, 269(5222): 347–350.

[2] Guo, Q. H., Hu, T. Y., Jiang, Y. Q., *et al*. Advances in remote sensing application for biodiversity research [J]. *Biodiversity Science*, 2018, 26(8): 789–806.

[3] Féret, J. B., Asner, G. P. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy [J]. *Ecological Applications*, 2014, 24(6): 1289–1296.

[4] Wang, R., Gamon, J. A. Remote sensing of terrestrial plant biodiversity [J]. *Remote Sensing of Environment*, 2019, 231: 111218.

[5] Palmer, M. W., Earls, P. G., Hoagland, B. W., *et al.* Quantitative tools for perfecting species lists [J]. *Environmetrics*, 2002, 13(2): 121–137.

[6] Rocchini, D., Marcantonio, M., Ricotta, C. Measuring Rao's Q diversity index from remote sensing: an open source solution [J]. *Ecological Indicators*, 2017, 72: 234–238.

[7] Rocchini, D., Marcantonio, M., Re, D. D., *et al*. Time-lapsing biodiversity: an open source method for measuring diversity changes by remote sensing [J]. *Remote Sensing of Environment*, 2019, 231: 111192.

[8] Khare, S., Latifi, H., Rossi, S. A 15-year spatio-temporal analysis of plant β-diversity using Landsat time series derived Rao's Q index [J]. *Ecological Indicators,* 2021, 121: 107105.

[9] Khare, S., Latifi, H., Rossi, S. Forest beta-diversity analysis by remote sensing: how scale and sensors affect the Rao's Q index [J]. *Ecological Indicator*s, 2019, 106: 105520.

[10] Wang, D. Z., Qiu, P. H., Wan, B., *et al*. Mapping α- and β-diversity of mangrove forests with multispectral and hyperspectral images [J]. *Remote Sensing of Environment*, 2022, 275: 113021.

[11] Jiang, X., Cai, H. Y., Yang, X. H., *et al*. Temporal and spatial evolution of Rao's Q index of forest in China [J]. *Acta Ecologica Sinica*, 2023, 43(8): 3045–3056.

[12] Jiang, X., Cai, H. Y., Yang, X. H. Yearly remote sensing Rao's Q index dataset of forest in China (2000–2017) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. https://doi.org/10.3974/geodb.2024.03.08.V1. https://cstr.escience.org.cn/CSTR:20146.11.2024.03.08.V1.

[13] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. https://doi.org/10.3974/dp.policy.2014.05 (Update 2017).

[14] Petchey, O. L., Gaston, K. J. Functional diversity (FD), species richness and community composition [J]. *Ecology Letters*, 2002, 5(3): 402–411.

[15] Li, G., Shen, Z. H., Ying, J. S., *et al*. The spatial pattern of species richness and diversity centers of gymnosperm in China [J]. *Biodiversity Science*, 2009, 17(3): 272–279.

[16] Hernández-Stefanoni, J. L., Gallardo-Cruz, J. A., Meave, J. A., *et al*. Modeling α- and β-diversity in a tropical forest from remotely sensed and spatial data [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2012, 19: 359–368.