

# The Spatial Distribution Dataset on Ecological Agriculture Patterns of China (2018–2020)

Wang, S.<sup>1</sup> Zhu, Y. Q.<sup>1,2\*</sup> Qian, L.<sup>3</sup> Song, J.<sup>1,2</sup> Yuan, W.<sup>1</sup>

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

2. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China;

3. School of Computer Science, South China Normal University, Guangzhou 510631, China

**Abstract:** Ecological agriculture patterns are agricultural development cases that can be replicated by appropriately utilizing the local natural environment. Surveying the distributions of ecological agriculture patterns can reveal spatial differences, aggregation, and diversity of agricultural development, which is greatly significant to agricultural development planning, agricultural ecological progress, and agricultural sustainable development research. To address this issue, the authors first collect news reports on the topic of Chinese ecological agriculture patterns from official websites such as Yangshi net, Renmin net, and Xinhua net between 2018 and 2020. And then, the authors extract and classify ecological agriculture patterns using natural language processing techniques. Finally, the dotted spatial distribution datasets of Chinese ecological agriculture patterns are produced by parsing spatial and temporal information about the ecological agriculture patterns. The dataset includes the records covering ecological agriculture type, location of each record, report date, keywords, original descriptions, and sources. The dataset is archived in .xlsx and .shp data formats with 33,440 records, and consists of 9 data files with data size of 168 MB (compressed to 21.4 MB).

**Keywords:** ecological agriculture patterns; spatial distribution; news report; 2018–2020

**DOI:** <https://doi.org/10.3974/geodp.2021.02.10>

**CSTR:** <https://cstr.escience.org.cn/CSTR:20146.14.2021.02.10>

**Dataset Availability Statement:**

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: <https://doi.org/10.3974/geodb.2021.06.02.V1> or <https://cstr.escience.org.cn/CSTR:20146.14.2021.06.02.V1>.

## 1 Introduction

Chinese ecological agriculture patterns are agricultural development cases with their own local natural resources and social-economic conditions under the Sustainable Development

---

**Received:** 22-04-2021; **Accepted:** 04-06-2021; **Published:** 25-06-2021

**Foundations:** Chinese Academy of Sciences (XDA23100100); National Natural Science Foundation of China (42050101, 41771430, 41631177, 42101467)

**\*Corresponding Author:** Zhu, Y. Q. L-6116-2016, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, zhuyq@lries.ac.cn

**Data Citation:** [1] Wang, S., Zhu, Y. Q., Qian, L., *et al.* The spatial distribution dataset on ecological agriculture patterns of China (2018–2020) [J]. *Journal of Global Change Data & Discovery*, 2021, 5(2): 181–188. <https://doi.org/10.3974/geodp.2021.02.10>. <https://cstr.escience.org.cn/CSTR:20146.14.2021.02.10>.  
[2] Wang, S., Zhu, Y. Q., Qian, L., *et al.* Spatial distribution dataset on ecological agriculture patterns of China (2018–2020) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. <https://doi.org/10.3974/geodb.2021.06.02.V1>. <https://cstr.escience.org.cn/CSTR:20146.14.2021.06.02.V1>.

Goals of the United Nations<sup>[1,2]</sup>. These ecological agriculture patterns provide outstanding demonstration effects on local agricultural development paths, providing significant meanings to regional sustainable development on local agricultural planting, production, and management<sup>[3]</sup>.

Different areas in China have different ecological, technique, and marketing situations because of its vast territory. These situations bring large different distributions of ecological agriculture patterns, which present an extreme challenge to refer similar patterns to local agriculture development. These differences not only affect the layout revealing of ecological agriculture pattern at the macro perspective, but also reveal the specific local natural environment and social-economic mechanism of the local pattern at the micro perspective. Thus, surveying the distributions of ecological agriculture patterns is significant to reveal the spatial structure and mechanism of each Chinese ecological agriculture pattern, which is related to the planning of ecological agriculture patterns at the national scale.

During 2001–2003, Science and Technology Division of the Ministry of Agriculture of P. R. China (now Ministry of Agriculture and Rural Affairs) conducted a national survey about ecological agriculture patterns. It collected 370 ecological agriculture patterns or techniques with the bottom-up method and published a top-10 Chinese ecological agriculture pattern list (northern quaternity ecological agriculture pattern, southern “animal-biogas-fruits” ecological agriculture pattern, grass restoration and sustainable utilization pattern, farming-forest-livestock pattern, ecological farming, ecological breeding, small watershed hybrid management and utilization, protected agriculture pattern, and agricultural ecological park pattern) evaluated by experts<sup>[4,5]</sup>. Although this method achieved 10 typical Chinese ecological agriculture development patterns, it cannot offer the explicit geographical location and distribution of each pattern<sup>[6]</sup>. This situation limited the accurate recommendation of local agriculture development patterns.

To address this issue, this research developed accurate dotted distribution dataset of Chinese ecological agriculture patterns. Since the outstanding local ecological agriculture patterns may be reported by the news, this research uses news texts as the raw data sources. By using natural language processing, location parsing, and other relevant techniques, this research reveals the type, geographical location, report date, and other information and finally produced the spatial distribution dataset on ecological agriculture patterns of China (2018–2020).

## 2 Metadata of the Dataset

The metadata of the Spatial distribution dataset on ecological agriculture patterns of China (2018–2020)<sup>[7]</sup> is summarized in Table 1. It includes the dataset’s 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 Technical Route

The development technical route of the spatial distribution dataset on ecological agriculture patterns of China shows in Figure 1. It mainly includes two core parts: corpus acquirement and information extraction.

### 3.1.1 Corpus Acquisition

Corpus acquisition consists of two steps: ecological agriculture pattern ontology construction and ecological agriculture pattern corpus crawling. Ecological agriculture pattern ontology is manually constructed with literature, reports, and books. The classification system of ecological agriculture patterns is shown in Table 2.

Ecological agriculture pattern corpus crawling is the process of obtaining news texts of ecological agriculture patterns based on the preset dictionary in the classification system. The news' portals include four sources: government portal, China Media Group, People's Daily Online, and Xinhua News Agency. The government portal selects the news portal of the Ministry of Agriculture and Rural Affairs of the People's Republic of China<sup>1</sup>. China Media Group selects the news portal of the China Media Group<sup>2</sup>. People's Daily Online selects the Search portal of People's Daily Online<sup>3</sup>. Xinhua News Agency selects the Search portal of Xinhua News Agency<sup>4</sup>.

**Table 1** Metadata summary of the Spatial distribution dataset on ecological agriculture patterns of China (2018–2020)

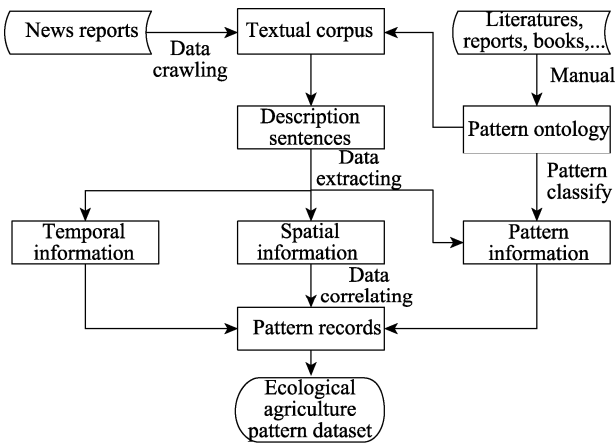
Items	Description		
Dataset full name	Spatial distribution dataset on ecological agriculture patterns of China (2018–2020)		
Dataset short name	CEApatterns_2018-2020		
Authors	Wang, S., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, wangshu@igsnrr.ac.cn Zhu, Y. Q. L-6116-2016, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, zhuyq@lries.ac.cn Qian, L., South China Normal University, 2018022623@m.scnu.edu.cn Song, J., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, songj@igsnrr.ac.cn Yuan, W., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, yuanwen@igsnrr.ac.cn		
Geographical region	China	Year	2018–2020
Temporal resolution	1 day	Spatial resolution	100 m
Data format	.xlsx, .shp		
Data size	168 MB (compressed to 21.4 MB)		
Data files	33,440 records		
Foundations	Chinese Academy of Sciences (XDA23100100); National Natural Science Foundation of China (42050101, 41631177)		
Data publisher	Global Change Research Data Publishing & Repository, <a href="http://www.geodoi.ac.cn">http://www.geodoi.ac.cn</a>		
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China		
Data sharing policy	<b>Data</b> from the Global Change Research Data Publishing & Repository includes metadata, datasets (in the <i>Digital Journal of Global Change Data Repository</i> ), and publications (in the <i>Journal of Global Change Data &amp; Discovery</i> ). <b>Data</b> sharing policy includes: (1) <b>Data</b> are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use <b>Data</b> subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute <b>Data</b> subject to written permission from the GCdataPR Editorial Office and the issuance of a <b>Data</b> redistribution license; and (4) If <b>Data</b> are used to compile new datasets, the 'ten per cent principal' should be followed such that <b>Data</b> records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset <sup>[8]</sup>		
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS		

<sup>1</sup> News portal of the Ministry of Agriculture and Rural Affairs of the People's Republic of China. <http://www.moa.gov.cn/xw/>.

<sup>2</sup> News portal of the China Media Group. <https://news.cctv.com/>.

<sup>3</sup> Search portal of People's Daily Online. <http://search.people.cn/>.

<sup>4</sup> Search portal of Xinhua News Agency. <http://so.xinhuanet.com/>.



**Figure 1** The development technical route of the spatial distribution dataset on ecological agriculture patterns of China

**Table 2** The classification system of ecological agriculture patterns

1 <sup>st</sup> class	2 <sup>nd</sup> class
Ecological farming	Forest-crop intercropping, Forest-medicine intercropping, Forest-vegetable intercropping, Forest-seedling intercropping, Forest-mushrooms intercropping, Forest-grass intercropping, Forest-flowers intercropping, Forest-fruit intercropping, Mushrooms-grass intercropping, Seasonal inter-planting, Space inter-planting, Nutrient inter-planting Fertigation, Drip irrigation, Alley cropping, Rainfall harvesting planting, Drought resistance, Water-efficient agriculture, Fertilizer-efficient agriculture, Precise fertilization, Protected agriculture, Remediation farming, Original ecological cultivation, Technology-assisted reduced fertilization Rice-fish, Rice-livestock, Forest-grass-livestock, Orchard-livestock, Free-range livestock farming, Planting-breeding-processing, Rice-fish-livestock, Animal- biogas- fruits, Multiple crop-livestock, Crop straw recycling farming, Farming- dispersed breeding, Planting-breeding intercropping, Crop-livestock-biogas, Mushrooms/ grass remediation farming, Integrated crop-livestock, Crop-livestock recycling
Ecological breeding	Fermentation bed farming, Livestock manure recycling, Fecal resource-returning field, Fecal resources transformation Waterfowl-aquatic products, Two-stage breeding, Chicken-pig, Dispersed breeding, Polyculture, Protected breeding, Season inter-breeding, Cross-regional culture, Breeding-processing, Farrow-to-finish breeding, Recycling breeding, Remediation breeding
Innovative agriculture	Microbial agriculture, Agriculture + Internet of Things, Photovoltaic agriculture, Industrial farming/breeding, High-quality agriculture, Industrial chain agriculture, High-quantity agriculture, High-tech agriculture, Foodbank Agriculture-internet, Agriculture crowdfunding, Contract farming, Shared agriculture, Agricultural ecological park, Picking tourism, Sci-tech agricultural park

3.1.2 Information Extraction

The process of information extraction includes temporal information extraction, spatial information extraction, pattern extraction, and pattern record aggregation. All these information extracts from news texts and the corresponding algorithms are demonstrated in the next section.

3.2 Algorithm Principle

The dataset development involves the following core algorithms: temporal information extraction algorithm, spatial information extraction algorithm, pattern extraction algorithm, and pattern record aggregation algorithm.

### (1) Temporal information extraction algorithm

Temporal information extraction obtains the report dates of ecological agriculture patterns from news texts. Fortunately, the report date of ecological agriculture pattern has standard representation forms. Thus, report dates can be parsed from the HTML files by using XPATH syntaxes during the dataset development. The XPATH syntaxes of the Ministry of Agriculture and Rural Affairs (MARA) news portal, the China Media Group (CMG) news portal, the People's Daily Online (PDO) search portal, and the Xinhua News Agency (XNA) search portal are (1), (2), (3), and (4), respectively.

$$\text{XPATH}_{\text{MARA}} = // \text{span}[@ \text{class} = "fbsj"] \quad (1)$$

$$\text{XPATH}_{\text{CMG}} = // \text{div}[@ \text{class} = "src - tim"] / \text{span}[@ \text{class} = "tim"] \quad (2)$$

$$\text{XPATH}_{\text{PDO}} = // \text{span}[@ \text{class} = "tip - pubtime"] \quad (3)$$

$$\text{XPATH}_{\text{XAN}} = // \text{div}[@ \text{class} = "easynews"] / \text{div}[@ \text{class} = "newstime"] / \text{span} \quad (4)$$

### (2) Spatial information extraction algorithm

Spatial information extraction obtains the spatial location information of ecological agriculture patterns from news texts. During the development process of the dataset, this research uses NLPiR<sup>5</sup> toolbox to recognize location name (toponym), for example, the “Zhuanglang town” in the sentence of “Zhuanglang town develops a sustainable...”. Then, the recognized toponyms can be parsed into coordinate information using Baidu geo-coding service<sup>6</sup>. Note that the spatial parsing accuracy is 100 m with the Baidu coordinate system (DB09).

### (3) Pattern extraction algorithm

Pattern extraction is to acquire the description texts of ecological agriculture patterns from the news texts. This dataset uses rule-based method to extract patterns by using regular expressions that can be classified into two types: trigger word class and non-trigger word class. Trigger word class uses characteristic features to extract patterns, for example, the regular expression “use {0,1}“(.(.)+”)(.)+pattern”. Non-trigger word class uses the standard sentence structures to extract patterns, for example the regular expression “(“([u4e00-\u9fa5]+)(—([u4e00-\u9fa5]+))+)””. The specific regular expression list is open-sourced in Github<sup>7</sup>.

### (4) Pattern record aggregation algorithm

Pattern record aggregation algorithm is to aggregate structured temporal information, spatial information, and pattern information. The basic principle of aggregation algorithm is associating temporal, spatial, and pattern information within sentences, because the semantic of these information is coherent in the situation of the inner sentence, inner paragraph, and nearby content. And the default information can also be filled in order by sentence, paragraph, and document.

<sup>5</sup> NLPiR. <http://ictclas.nlpir.org/>.

<sup>6</sup> Baidu geocoding service. <http://api.map.baidu.com/geocoding/v3/?address=庄浪县&output=json&ak=ak&callback=showLocation>.

<sup>7</sup> Open-sourced in Github. <https://github.com/shuwang8951/EcoCivMdl>.

4 Data Results and Validation

4.1 Data Composition

The dataset consists of 33,440 ecological agriculture pattern records. Each record in .xlsx includes 22 fields (Table 3).

Table 3 Items of the records

No.	Item	No.	Item
1	ID (serial number)	12	MODE_TYPE_LEVEL_1_ZH (level one class of ecological agriculture pattern in Chinese)
2	DATASOURCE_ZH (data source in Chinese)	13	MODE_TYPE_LEVEL_1_EN (level one class of ecological agriculture pattern in English)
3	DATASOURCE_EN (data source in English)	14	MODE_TYPE_LEVEL_2_ZH (level two class of ecological agriculture pattern in Chinese)
4	URL (URL link)	15	MODE_TYPE_LEVEL_2_EN (level two class of ecological agriculture pattern in Chinese)
5	TITLE_ZH (document title in Chinese)	16	EXTRACT_MODE_ZH (extracted original pattern descriptions in Chinese)
6	TITLE_EN (document title in English)	17	EXTRACT_MODE_EN (extracted original pattern descriptions in English)
7	REPORT_DATE (report date)	18	KEYWORDS_ZH (keywords in Chinese)
8	LOCATION_ZH (location description in Chinese)	19	KEYWORDS_EN (keywords in English)
9	LOCATION_EN (location description in English)	20	CONTENT (document content),
10	LNG (longitude)	21	SHORT_SENTENCE (the short sub-sentence of extracted pattern)
11	LAT (latitude)	22	LONG_SENTENCE (the sentence of extracted pattern)

The dataset .shp file uses a vector data model to store all the .xlsx records as points.

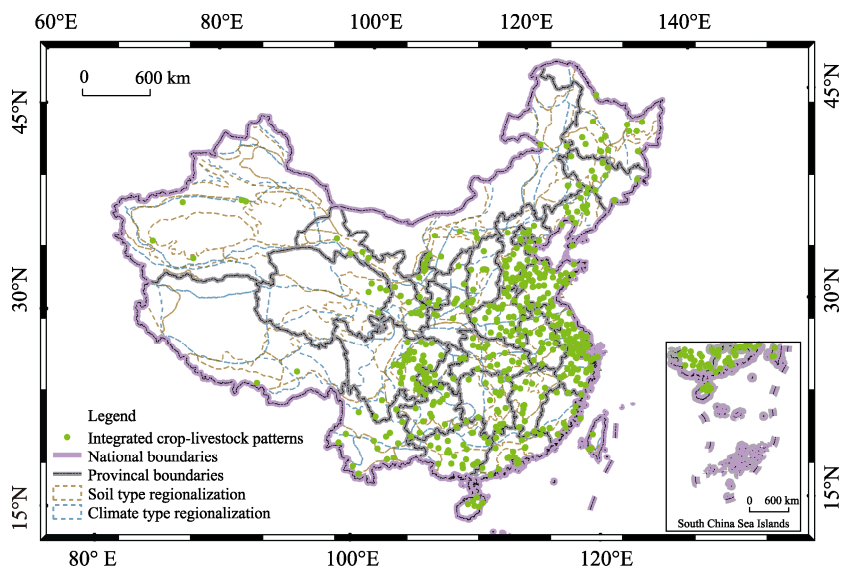
4.2 Data Products

The dataset contains 72 ecological agriculture pattern types. The top-10 types include integrated crop-livestock pattern, animal-biogas-fruits pattern, rice-fish pattern, agricultural ecological park pattern, agriculture+internet pattern, multiple crop-livestock pattern, livestock manure recycling pattern, fertigation pattern, crop-straw utilization pattern, and forest-grass-livestock pattern.

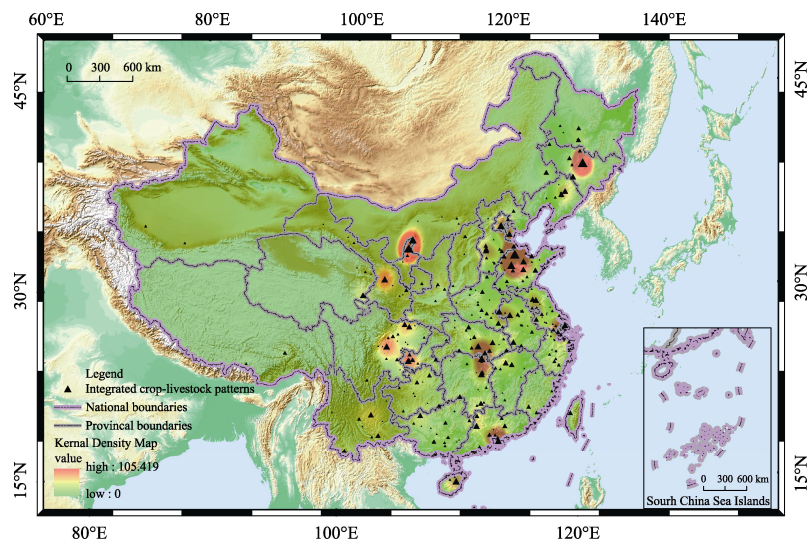
The dotted spatial distribution of the integrated crop-livestock patterns in China is shown in Figure 2. Each point in Figure 2 represents a local application that occurs. To represent the trend of the spatial distribution of integrated crop-livestock patterns, the kernel density map is demonstrated in Figure 3. The black triangles in Figure 3 represent the areas where integrated crop-livestock ecological agriculture patterns occurred more intensively. The bigger triangle means more applications of integrated crop-livestock ecological agriculture patterns.

4.3 Data Validation

For the data validation, we randomly select 150 records from the dataset and manually annotates temporal, spatial, and pattern information from the original news texts. By comparing the annotated and extracted results, all of accuracies of the dataset are listed in Table 4.



**Figure 2** The spatial distribution map of integrated crop-livestock ecological agriculture pattern in China



**Figure 3** The kernel density map of integrated crop-livestock ecological agriculture pattern in China

**Table 4** The different kinds of extraction process accuracies of ecological agriculture pattern

Extraction type	Selected record number	Error record number	Accuracy
Temporal information	150	0	100%
Spatial information	150	7	95.3%
Pattern information	150	8	94.7%

To verify the coverage of ecological agriculture patterns in the dataset, this paper compares typical agricultural ecological park pattern records with current national official pilot region lists about the agricultural ecological park pattern. The lists consist of two parts: 47 national agricultural entrepreneurial and innovation parks (bases) published by the Ministry of Agriculture and Rural Affairs<sup>[9]</sup> and 54 typical outstanding agricultural ecological parks in

literature<sup>[10]</sup>. The coverage rates of the records about agricultural ecological park patterns are listed in Table 5. The average coverage rates of county-level and city-level are 87.13% and 92.08%, respectively.

**Table 5** The coverage rates of the records about agricultural ecological park pattern

Comparative regions	Coverage rate in county-level	Coverage rate in city-level
National agricultural entrepreneurial and innovation parks (bases)	87.03%	92.59%
Agricultural ecological parks (leisure agriculture/tourism agriculture)	87.23%	91.49%
Average	<b>87.13%</b>	<b>92.08%</b>

5 Discussion and Conclusion

To reveal the spatial distributions of Chinese ecological agriculture patterns, the authors collect the news texts about Chinese ecological agriculture patterns in 2018–2020, classify the pattern records among these news texts, parse the relevant temporal report dates and locations, and finally produce the dotted spatial distribution dataset of Chinese ecological agriculture pattern with multiple natural language techniques.

Author Contributions

Zhu, Y. Q. finished the overall design. Wang, S., Song, J., and Yuan, W. designed the algorithms of the dataset. Wang, S. and Qian, L. contributed to the data processing and analysis. Wang, S. did the data validation. Wang, S. wrote the data paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Xu, C. Comparative study of Chinese ecological agriculture and sustainable agriculture [J]. *International Journal of Sustainable Development & World Ecology*, 2004, 11(1): 54–62.

[2] Yin, C., Cheng, L., Yang, X., *et al.* Path decision of agriculture sustainable development based on eco-civilization [J]. *Journal of China Agricultural Resources and Regional Planning*, 2015, 36(1): 15–21.

[3] Liu, Z., Jia, W. Ecological Civilization Concepts and Modes [M]. Beijing: Chemical Industry Press, 2015: 82–87.

[4] Department of Science, Ministry of Agriculture and Rural Affairs of the People’s Republic of China. The top 10 modes and technologies of Chinese ecological agriculture [J]. *Journal of Agricultural Resources and Environment*, 2003(1): 16.

[5] Li, M., Zhang, Y., Xu, M., *et al.* China eco-wisdom: a review of sustainability of agricultural heritage systems on aquatic-ecological conservation [J]. *Sustainability*, 2020, 12(1): 60.

[6] Wang, X. M. Study on the problems of Chinese organic agriculture development history and present situation [C]. International Conference on Advanced Educational Technology and Information Engineering (AETIE). Beijing, 2015: 984–989.

[7] Wang, S., Zhu, Y., Qian, L., *et al.* Spatial distribution dataset on ecological agriculture patterns of China (2018–2020) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. <https://doi.org/10.3974/geodb.2021.06.02.V1>. <https://cstr.science.org.cn/CSTR:20146.14.2021.06.02.V1>.

[8] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (Updated 2017).

[9] Wang, F., Wang K., Chen, T. National agritourism parks in China: distribution, types and spatial optimization [J]. *Research of Agricultural Modernization*. 2016, 37(6): 1035–1044.

[10] Bao, W. Research on the development and industrialization of leisure agricultural resources in China [D]. Qingdao: Ocean University of China, 2013: 175–177.