A Novel Rapid Investigation Method for Ecological Agriculture Patterns Based on Web Text

Shu Wang[1], Yunqiang Zhu[2], Lang Qian[3], Jia Song[4], Wen Yuan[5], Kai Sun[5], Weirong Li[5], and Quanying Cheng[5]

[1]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
[2]Chinese Academy of Sciences (CAS)
[3]Computer School, South China Normal University
[4]Institute of Geograpical Science and Natural Resources Research, Chinese Academy of Sciences
[5]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences

November 21, 2022

Abstract

The investigation of Ecological Agriculture (EA) patterns can reveal the differences, aggregation, and diversity of agriculture development, providing specific paths in agriculture development and environment protection in order to achieve the Sustainable Development Goals. Although field surveys, literature analysis, and administrative statistical methods can be employed to comprehensively investigate EA records and determine EA distributions, they still rely on manual operations that are generally unable to support the rapid and large-scale identification of EA patterns required by current agricultural sustainable researches. To address this issue, this paper proposes a novel and rapid approach for Ecological Agriculture Pattern Investigation Based on Web-text (WEAPI), with the ability to automatically acquire EA pattern records including pattern type, occurrence time, precise location, and other relevant information. The proposed method is employed in a national scale case study to investigate trends in Chinese Ecological Agriculture (CEA). Results reveal the ability of WEAPI to detect new trends in CEA via the latest news, as well as the corresponding distributions. The WEAPI method can also exhibit the unknown patterns of the current Chinese agriculture development. Further validation experiments demonstrate the proposed method to achieve over 95% precision in the pattern parse processes and an 87% coverage rate at the town level of the official CEA pattern list. Moreover, WEAPI can also provide dynamic analyses on the evolution of the EA patterns. Despite limitations under sparse records in partial classes, the results reveal WEAPI to be a promising and powerful tool for agricultural research and agricultural development planning.

A Novel Rapid Investigation Method for Ecological Agriculture Patterns Based on Web Text

Shu. Wang[1], Yunqiang. Zhu[1,2], Lang. Qian[3], Jia. Song[1,2], Wen. Yuan[1,2], Kai. Sun[1], Weirong. Li[1,4], and Quanying. Cheng[1,4]

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

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

[3]Computer School, South China Normal University, Guangzhou, China.

[4]University of Chinese Academy of Sciences, Beijing, China.

Corresponding author: Yunqiang. Zhu (zhuyq@lreis.ac.cn)

Key Points:

- Ecological Agriculture Pattern can be rapidly investigated in a large scale
- A novel Web-text based Ecological Agriculture Pattern Investigation (WEAPI) method is proposed
- The classification and distributions of the Chinese Ecological Agriculture (CEA) pattern are revealed

Abstract

The investigation of Ecological Agriculture (EA) patterns can reveal the differences, aggregation, and diversity of agriculture development, providing specific paths in agriculture development and environment protection in order to achieve the Sustainable Development Goals. Although field surveys, literature analysis, and administrative statistical methods can be employed to comprehensively investigate EA records and determine EA distributions, they still rely on manual operations that are generally unable to support the rapid and large-scale identification of EA patterns required by current agricultural sustainable researches. To address this issue, this paper proposes a novel and rapid approach for Ecological Agriculture Pattern Investigation Based on Web-text (WEAPI), with the ability to automatically acquire EA pattern records including pattern type, occurrence time, precise location, and other relevant information. The proposed method is employed in a national scale case study to investigate trends in Chinese Ecological Agriculture (CEA). Results reveal the ability of WEAPI to detect new trends in CEA via the latest news, as well as the corresponding distributions. The WEAPI method can also exhibit the unknown patterns of the current Chinese agriculture development. Further validation experiments demonstrate the proposed method to achieve over 95% precision in the pattern parse processes and an 87% coverage rate at the town level of the official CEA pattern list. Moreover, WEAPI can also provide dynamic analyses on the evolution of the EA patterns. Despite limitations under sparse records in partial classes, the results reveal WEAPI to be a promising and powerful tool for agricultural research and agricultural development planning.

1 Introduction

Ecological Agriculture (EA) is a modern and efficient agriculture tendency guided by the combined principles of ecology and economics, providing specific paths to implement the Sustainable Development Goals (Ali et al., 2019; Lockeretz, 1989; Priyadarshini & Abhilash, 2020). The distribution of EA patterns can reveal the differences, aggregation, and diversity of regional agriculture development (González-Chang et al., 2020). For example, the agriculture development pattern of the integrated crop-livestock system in China is concentrated in the Jilin, Shandong, and Jiangsu provinces. These distributions not only support development plans and strategy adjustments on a macroscale, but can also provide reliable development templates for similar regions on a microscale (Priyadarshini & Abhilash, 2020; The Ministry of Agriculture of the People's Republic of China, 2003; Yang et al., 2021). Revealing these EA distributions requires the integration of detailed EA patterns with spatial, temporal, and typological information.

Currently, the investigation of EA patterns can be grouped into the following three methods: field surveys, literature analysis and administrative statistics. Field surveys can identify EA patterns via interviews with local farmers, which can then be summarized to form local agriculture patterns and historical trends (Haynes, 2010; P. Liu, Moreno, Song, Hoover, & Harder, 2016; Riley & Harvey, 2007). Literature analysis adopts typical case studies to identify and categorize EA patterns (Li et al., 2019; Y. Liu, Duan, & Yu, 2013; Velten, Leventon, Jager, & Newig, 2015). Administrative statistics can quantify EA patterns through the submission and examination ways level by level from the bottom administrative units using efficient executive orders (H. Wang, Qin, Huang, & Zhang, 2007; Ye, Wang, & Li, 2002).

Although these three methods have advantages including sufficient content, solid data, and an effective execution, they are also limited by several factors. For example, the field survey method relies on the depth and coverage of the interview to achieve sufficient content, which requires time and multiple investigators. Thus, this method is associated with a very low investigation efficiency. In addition, the literature analysis method depends on the data resources from existing research. For example, previous studies have revealed EA patterns, with successful applications of photovoltaic agriculture, rice-fish systems, and circular agriculture in the Fujian, Guizhou, and Shandong provinces (Atinkut et al., 2020; L. Wang, Wang, & Chen, 2019; Yi, 2019; Zheng et al., 2017). Such research is not only limited in quantity, but also focuses on a single or few EA patterns, resulting in a bias of the survey results derived from the literature analysis method. The effective execution of the administrative statistical method comes from administrative orders and government financial support, thus requiring time and money. For example, the MOA (Ministry of Agriculture of the People's Republic of China) China received 500-800 EA pattern cases from nationwide counties and summarized 10 typical Chinese Ecological Agriculture (CEA) patterns via gradual screening and expert examination in 2003. Almost 5 years were required for this project to arrange national eco-agriculture pilot regions and investigate their local EA patterns (The Ministry of Agriculture of the People's Republic of China, 2003). Thus, this method is not sustainable nor economical for the continuous investigation of EA patterns. Furthermore, the level by level submission is subjective to local EA pattern evaluations, which consequently effect the accuracy of the investigations.

Thus, the aforementioned methods are not suitable for the timely investigation of EA patterns to determine current trends and examples of EA development on the macro- or micro-level. Based on this, this is an urgent requirement for a method that can perform the rapid investigation of EA patterns in order to timely represent current ecological agriculture characteristics. Essentially, the existing three methods are not able to support the rapid collection of EA patterns as they all require manual interactions. Manually generated data sources limit the application of rapid EA pattern investigations at the larger scale.

To address this issue, this paper proposes a novel investigation method for ecological agriculture patterns based on web text, which contains extensive agriculture information across locations and research areas (Burton & Riley, 2018; Pan, Yang, Zhou, & Kong, 2020). Web text not only covers a multitude of agriculture patterns, but also possesses the advantages of large-scale applications, timely updating, and easy access. For example, the news article in the official MOA web portal "Leisure Farming Projects Built to Promote Ecological Agriculture in Weimin, Fujian" records the latest EA pattern information with detailed temporal, spatial, and typological descriptions. Massive amounts of these web texts are updated every day in different portals. Therefore, web text has great potential in rapidly investigating EA patterns at a larger scale. By adopting this powerful data source, the main contributions of this paper are as follows:

- To propose a novel rapid Web-text based Ecological Agriculture Pattern Investigation (WEAPI) method.
- To determine the classification and distribution of Chinese Ecological Agriculture (CEA) patterns via the proposed WEAPI method
- To reveal the current top 10 CEA patterns of the 4th level class and their corresponding distributions

The remainder of this paper is organized as follows: Section 2 describes the basic concept and methodology of the proposed WEAPI. Section 3 implements the WEAPI to extract the CEA classifications and distributions, while Section 4 presents the validation experiments and results. Section 5 discusses the underlying features of the WEAPI method and Section 6 concludes the paper.

2 Methodology

2.1 Basic concept

An EA pattern implies a local agriculture development pattern that exhibits a good balance between economic and ecological benefits and can be popularized and widely adopted for reference in similar locations. A formal definition of EA patterns is given in Formula (1):

$$\mathrm{EA} = \left\{a_n \in S_e \cap E_c \cap E_g \middle| n \in \mathbb{R}^+\right\} \quad (1)$$

where a_n denotes local agriculture development patterns and S_e, E_c, and E_g indicate the spatial extension, ecological conservation, and economic growth, respectively. Spatial extension indicates that an EA should have a certain amount of implementation cases distributed in different areas. Ecological conservation restricts EAs from destroying the local ecological environment, while economic growth refers to the promotion of local economic development by EAs. Thus, S_e, E_c, and E_g should satisfy their evaluation thresholds $S_e > \theta_{se}$, $E_c > \theta_{ec}$, and $E_g > \theta_{eg}$. More specifically, a local agriculture development pattern can be regarded as an EA pattern if it simultaneously exhibits spatial extension, ecological conservation, and economic growth.

The investigation of EA pattern definitions requires an EA evaluation process. Therefore, a rapid EA pattern investigation method based on web text should include the following three steps (**Figure 1.**): (a) the information acquisition of massive local agriculture patterns; (b) the information extraction of structured agriculture patterns; and (c) the EA evaluation of the spatial extension, ecological conservation, and economic growth of agriculture patterns. Based on this, information retrieval techniques, Natural Language Processing (NLP) technologies, and an EA evaluation method are required to support the whole framework.

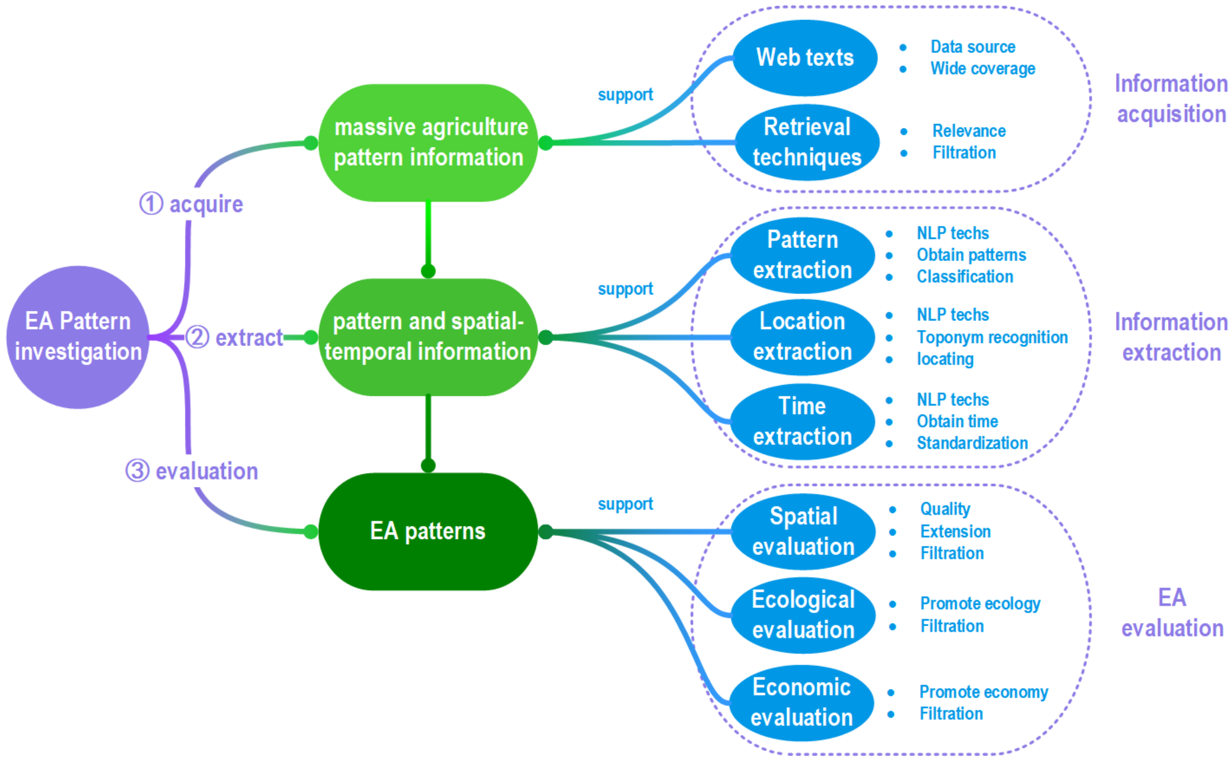


Figure 1. The logic flow of a rapid EA pattern investigation method based on web text.

2.2 WEAPI method

The proposed novel rapid Web-Text Based Ecological Agriculture Pattern Investigation (WEAPI) method is designed based on three core sections: information acquisition, information extraction, and EA determination. **Figure 2.** illustrates the workflow of the WEAPI method and the implementation details are provided in Github (https://github.com/shuwang8951/EA_pattern_analyses).

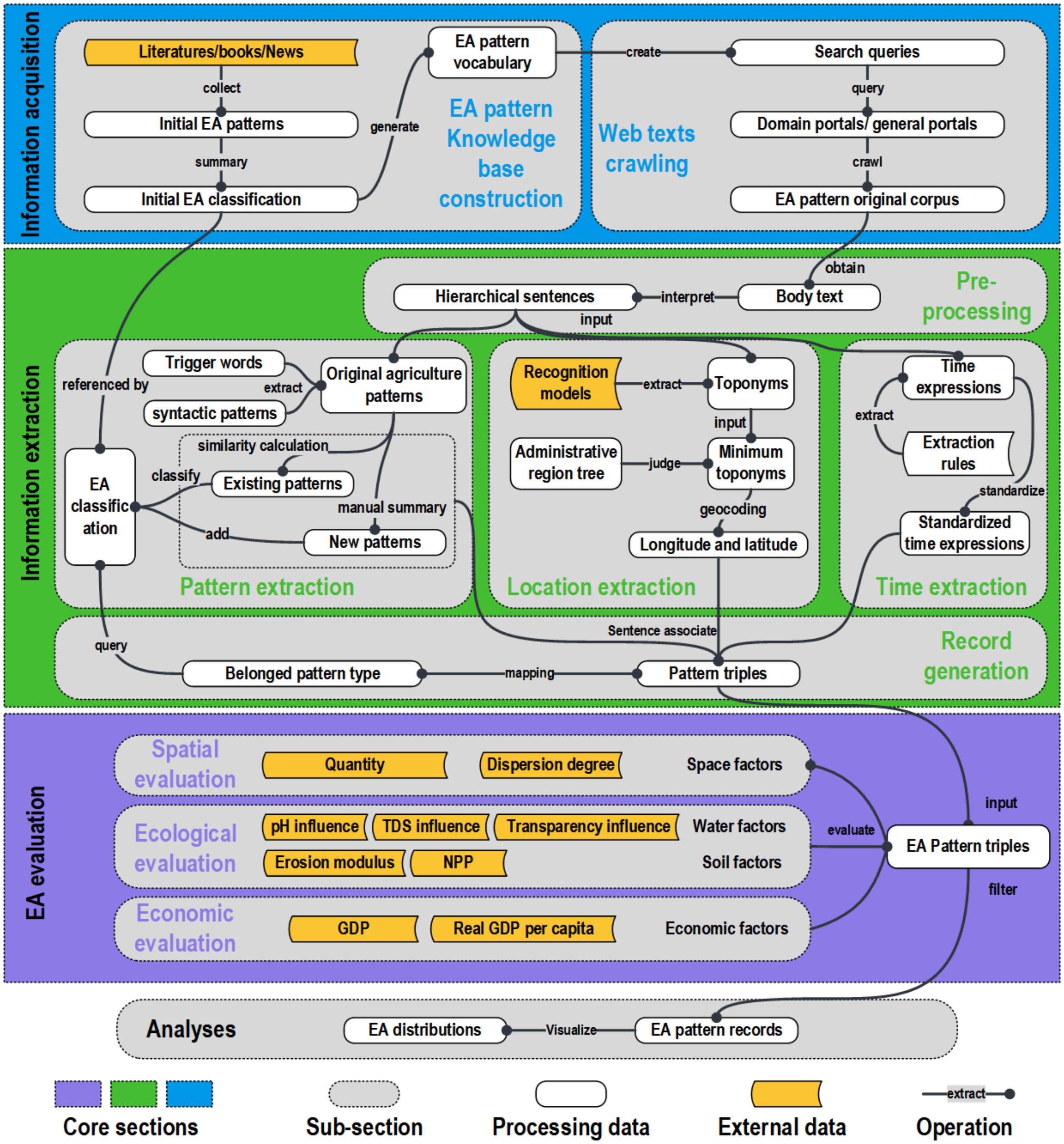


Figure 2. Workflow of the WEAPI method.

The information acquisition section focuses on acquiring massive EA pattern information from web text. This requires sufficient terminology (e.g., "rice-fish co-exist system" and "stereo planting"), as direct queries with the phrase "ecological agriculture" cannot obtain numerous agriculture pattern descriptions. Thus, EA pattern vocabulary is initially generated to support the queries and crawling. The original EA pattern corpus is then crawled.

The information extraction section is designed to generate agriculture pattern triples (pattern description, occurrence location, and occurrence time) by interpreting the content from the original corpus retrieved by the web crawler. Thus, this section contains five sub-sections:

the pre-processing sub-section is implemented to parse the original corpus into sentences; three core sub-sections extract the patterns, locations, and time information; and the last sub-section requires forms the patterns, locations and time information into pattern triples.

The EA evaluation section determines whether the extracted triple belongs to the EA patterns. The filtered EA pattern records (including pattern triples and relevant information) are analyzed to reveal the distributions of the EA patterns and other underlying information.

2.2.1 Information acquisition

The information acquisition process includes two sub-sections: EA pattern knowledge base construction and web text crawling.

EA pattern knowledge base construction obtains vocabulary related to EA patterns from existing data sources including literature, books, and newspapers. We first collect EA pattern vocabulary from the relevant data sources in order to obtain common EA pattern descriptions. Once the initial EA patterns have been determined, the initial EA classification is manually summarized (**Table 1.**) and the EA pattern vocabulary are generated by collecting all the phrases in the classification.

Table 1. *Initial EA classification.*

1st level class	2nd level class	3rd level class	4th level class (examples)
EA pattern	Ecological farming	Reduced fertilization planting patterns	Fertigation, water-efficient agriculture,…
		Circular food chain patterns	Crop straw utilization,…
		3-D intercropping patterns	Forest-crop intercropping,…
	Ecological breeding	Fecal resources utilization patterns	Fecal returning field,…
		Stereoscopic cultivation patterns	Waterfowl-aquatic products,…
	Hybrid farming	Planting-breeding mix patterns	Rice-fish, animal-biogas-fruits,…
	Innovative agriculture	White agriculture patterns	Photovoltaic agriculture,…
		Internet agriculture patterns	Crowdfunding agriculture
		Ecotourism patterns	Ecological park,…

Note. The total number of initial EA classification in the 4th level class is 53.

Web text crawling is key to the acquisition process of the EA pattern original corpus and can be divided into three steps: i) seach queries are created with the EA pattern vocabulary based on Formula (2):

$$\text{search query} = \{EAPVs\} + agriculture \tag{2}$$

where $\{EAPVs\}$ denotes the EA pattern vocabulary that support the queries and $agriculture$ denotes the filter word for general texts. Note that $\{EAPVs\}$ is a list including the designed level and all child level vocabulary; ii) the domain portals and general news portals are then queried using the created queries as EA patterns have a high probability of being reported by these websites. Thus, descriptions of the EA pattern applications can be collected from these websites. The domain portals denote the official portals of the agriculture departments, for example, the national official portal (e.g., http://www.moa.gov.cn/), provincial official portals (e.g., http://coa.jiangsu.gov.cn/), and city-level official portals (e.g., http://nyncj.nanjing.gov.cn/). General news portals indicate the portals of the popular websites (e.g., https://news.baidu.com/); iii) finally, these descriptions are crawled from the portals using the created queries.

2.2.2 Information extraction

The information extraction process is categorized into the following subsections: pre-processing; pattern extraction; location extraction; time extraction; and record generation.

a. Pre-processing

Pre-processing is required to place the EA pattern original corpus into hierarchical sentences, whereby each document in the corpus is interpreted into a tree structure with paragraph nodes, sentence nodes, and sub-sentence nodes. This structure not only simplifies the complex extraction issues into a simple task with short sub-sentences, but also attains clear sentence logic to associate the extracted patterns, locations, and temporal information. The pre-processing step initially obtains the body text of each document from the crawled corpus that can be applied with the existing open-source acquisition tools of published articles (e.g., goose3). The corpus contents are then split into hierarchical sentences for the extraction of patterns, locations, and temporal information.

b. Pattern extraction

Pattern extraction is implemented to extract agriculture patterns from each short sub-sentences via two core steps: the original agriculture pattern extraction and EA pattern classification.

The original agriculture pattern extraction parses all patterns from each short sub-sentence using trigger words and syntactic patterns. Trigger words are indicative vocabulary that point out the descriptions of the agriculture patterns (e.g., "agriculture pattern", "eco-agriculture pattern"). Syntactic patterns denote the grammar rules of the agriculture pattern descriptions (e.g., "using … pattern", "… pattern has been applied in …"). These two sets, which are summarized manually, are employed to extract the original agriculture patterns from short sub-sentences (e.g., "rice+fish", "Shouguang pattern", "Internet+"). Note that these original agriculture patterns are not accurate to the initial EA classification, resulting in several issues: unstandardized descriptions (e.g., "rice+fish" vs "rice-fish"); special semantic descriptions (e.g., "Shouguang pattern" vs "Industrial planting pattern" where "shouguang" is a typical applied place name of this pattern); and un-existing descriptions (e.g., "Internet+" does not exist in the initial EA classification).

In order to address these issues, EA pattern classification adjusts the initial EA classification generated in the information acquisition section. The Levenshtein distance, $lev_{a,b}(i,j)$, is implanted to solve the unstandardized agriculture patterns for the evaluation of the similarity between the extracted and existing patterns.

$$lev_{a,b}(i,j) = \begin{cases} max(i,j) & if\ min(i,j) = 0 \\ min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1(a_i \neq b_j) \end{cases} & otherwise \end{cases} \quad (3)$$

where $lev_{a,b}(i,j)$ refers to the minimum operation account of two words (a,b) at the character level; and $i\ and\ j$ are the character lengths of a and b, respectively. The operations include insertion, deletion, and substitution. For example, the Levenshtein distance between "rice+fish" and "rice-fish" is $lev_{a,b}(i,j) = 1$. Manual interpretation is adopted to solve the remaining original agriculture patterns as the other algorithms (e.g., text clustering and text classification) exhibit larger error rates (exceeding 20%) that will continuously influence the final results. Based on these methods, each agriculture pattern can be extracted from short sub-sentences with the explicit EA classification.

c. Location extraction

As texts do not directly record accurate coordinate information (latitude and longitude) of the agriculture patterns, an indirect method based on toponym recognition, minimum toponym determination, and geocoding is designed.

Toponym recognition obtains all the toponyms from the sub-sentence via an existing recognition model, NLPIR (Zhang, Miao, Liu, Wesson, & Shang, 2020) and outputs either single toponyms (e.g., "Hetong county") or sequence toponyms (e.g., "'Jiangsu' 'Nanjing' 'Gaochun'"). Single toponyms can clear mark the location of the pattern, yet sequence toponyms need to determine the minimum unit among the recognized toponyms. The minimum toponym determination step acquires the minimum unit of the sequence toponym descriptions based on the administrative region tree containing the affiliation relationships between regions (National Bureau of Statistics, 2020b). The sequence toponyms are then compared with the corresponding relationships to select the most accurate toponym. Geocoding is subsequently implemented to achieve accurate coordinate information (latitude and longitude) using the open source geocoding service (e.g., Baidu geocoding service (Baidu, 2020)) and the toponym as the input.

d. Time extraction

Time extraction is grouped into: time expression extraction and time expression standardization. As temporal information has standardized expressions and mature extraction technologies (e.g., rule-based methods) can be used (Leeuwenberg & Moens, 2019), time expressions can be extracted from the sub-sentences via the summarized extraction rules. The extraction rules can be integrated into the NLPIR system and can also be supplemented during the implementation process. The extract time expressions require standardization into a uniform format, e.g., "2020-07-16".

e. Record generation

This step generates uniform pattern records (e.g., "pattern (including types)-time-location"), the triples of the extracted patterns and corresponding classifications, the locations, and the temporal information that require association. Thus, record generation consists of two steps: element association and classification mapping.

Element association is a complex issue that considers the semantics of different elements. Since the descriptions of the elements are organized based on nature language, the sentences have inner semantic levels (sentence level, paragraph level, and document level). Hence, semantic structure can be used to associate different elements. Each pattern can determine its own location and temporal information from its extracted content following the underlying semantic logic. **Figure 3.** presents the element association process. The classification map then identifies the agriculture pattern classification of the extracted patterns.

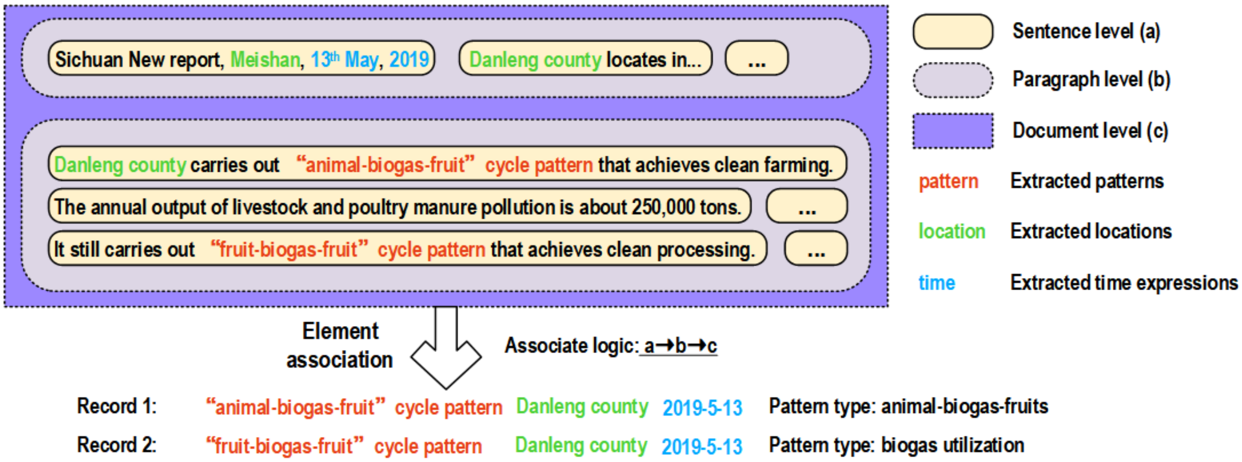


Figure 3. Element association based on inner semantic logic.

2.2.3 EA evaluation

Once the agriculture pattern triples are obtained, a filtration process is required to determine whether the obtained agriculture pattern belongs to the EA pattern. The three conditions in Formula (1) (spatial extension $S_e > \theta_{se}$, ecological conservation $E_c > \theta_{ec}$, and economic growth $E_g > \theta_{eg}$) must be considered. Thus, the evaluation process executes spatial, ecological, and economic evaluations.

The spatial evaluation determines which EA patterns should be popularized and widely used for reference in similar places. In particular, EA patterns include quantity and discreteness as indicators in the spatial scale and thus spatial extension threshold θ_{se} contains two components. In particular, quantity threshold Q is adopted to evaluate the popularity of the EA patterns, while distance threshold D between the different occurrence locations of the EA patterns indicates the EA pattern discreteness. Formula (4) describes spatial extension condition S_e:

$$S_e = \begin{cases} S_e > \theta_{se} & n_{EAPs} > \mathrm{Q} \ \& \ d_{EAPs} > \mathrm{D} \\ S_e \le \theta_{se} & otherwise \end{cases} \tag{4}$$

where n_{EAPs} indicates the numbers of EA patterns; d_{EAPs} indicates the Euclidean distance between EA patterns; Q controls the investigation granularity of the EA patterns and is set to 10 in the WEAPI method; and D determines the spatial resolution of the investigation and is set to 20 km, the typical distance between towns.

The ecological evaluation process indicates that the EA patterns should maintain an ecological environment suitable for sustainable development. The application of agriculture patterns exerts a strong influence on the local water and soil environments and thus the ecological evaluation assesses the extracted patterns based on these two factors. The local water environment includes surface water and ground water. Thus, transparency, pH, and total dissolved solids (TDS) are selected to evaluate the surface water, ground water chemical properties, and physical properties of ground water. The soil environment includes soil erosion and surface vegetation, and thus the erosion modulus, and net primary productivity (NPP) are selected to evaluate the effect of agriculture patterns. **Table 2.** lists the Formulas of the evaluation factors.

Table 2. *Formulas of the ecological evaluation factors*

Type	Factor	Formula	Definition
Water	Transparency (θ_{trs})	$\theta_{trs} = value_{pst} - value_{pre} \geq 0$ (cm)	Improved surface water
Water	pH (θ_{pH})	$\theta_{pH} = \lvert value_{pst} - 7\rvert - \lvert value_{pre} - 7\rvert \leq 0$	Improved groundwater quality
Water	TDS (θ_{TDS})	$\theta_{TDS} = value_{pst} - value_{pre} \leq 0$ (mg/L)	Improved groundwater mineralization
Soil	erosion modulus (θ_{em})	$\theta_{em} = value_{pst} - value_{pre} \leq 0$ (t/(km^2*a))	Improved soil erosion
Soil	NPP (θ_{NPP})	$\theta_{NPP} = value_{pst} - value_{pre} \geq 0$ (g/(m^2*a))	Improved Soil productivity

Note. $value_{pre}$ and $value_{pst}$ denote pre and post factor values.

$$E_c = \begin{cases} E_c > \theta_{ec} & \theta_{trs} \geq 0 \,\&\, \theta_{pH} \leq 0 \,\&\, \theta_{TDS} \leq 0 \,\&\, \theta_{em} \leq 0 \,\&\, \theta_{NPP} \geq 0 \\ E_c \leq \theta_{ec} & otherwise \end{cases} \quad (5)$$

The economic evaluation denotes the promotion of economic development for EA patterns. We select local Gross Domestic Product (GDP) and local real GDP per capita (PERGDP) to evaluate the economic improvements, which calculate the increment between the post and previous values, denoted as $I = value_{pst} - value_{pre}$. The economic growth condition $E_g > \theta_{eg}$ is evaluated as follows:

$$E_g = \begin{cases} E_g > \theta_{eg} & I_{GDP} > 0 \;\&\; I_{PERGDP} > 0 \\ E_g \leq \theta_{eg} & otherwise \end{cases} \quad (6)$$

We thus adopt the definition Formula (1) and conditional calculation Formulas (4), (5), and (6) to investigate the extracted and true EA patterns.

3 Case study

3.1. Research area

The People's Republic of China, a national scale research area, has been selected to investigate the classes and distributions of its Ecological Agriculture (CEA) patterns based on the proposed WEAPI. This region covers approximately 9.6 million km2 of land territory that can be divided into different climate, soil, and landscape zones and thus CEA patterns vary with region (**Figure 4.**).

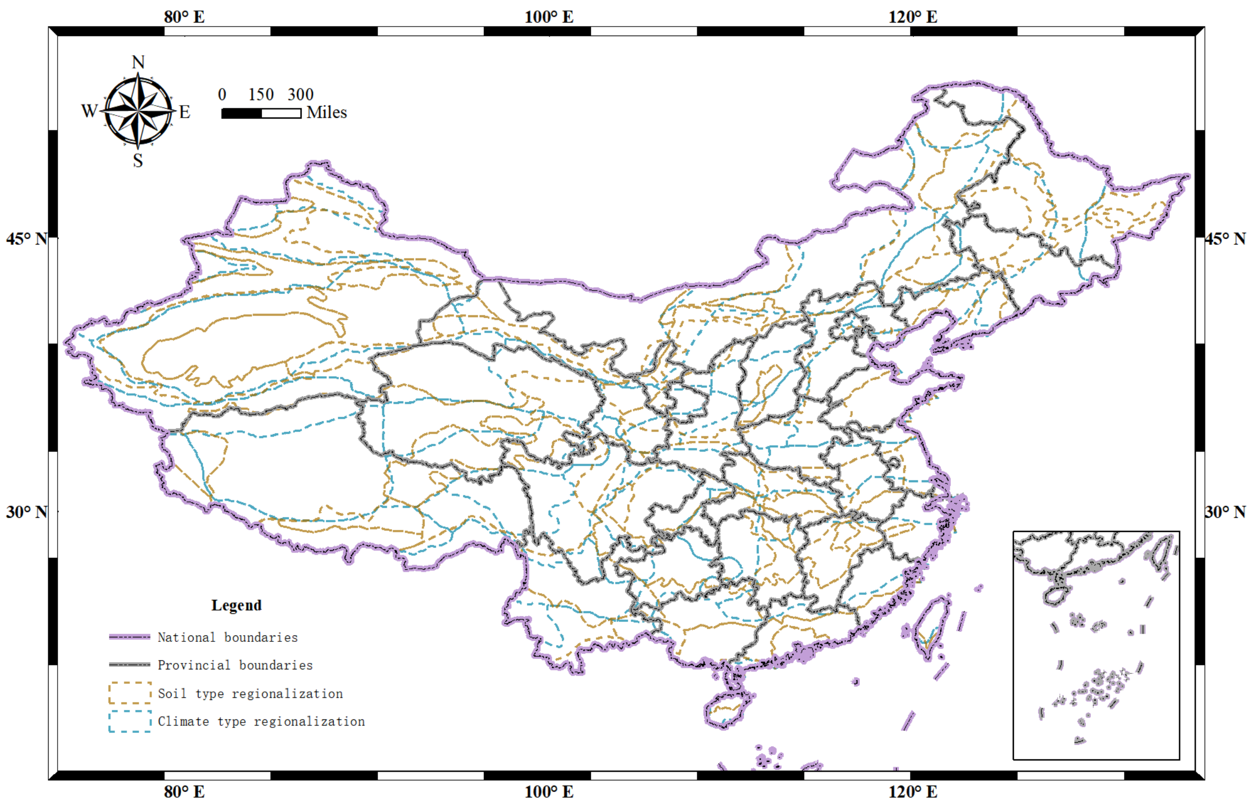


Figure 4. Map of China, the research area, with complex soil and climate regionalization.

3.2. Data sources

Two data sources were employed to investigate the acquisition and evaluation of CEA patterns. Web-based agricultural news is the principle data source of the pattern acquisition process and includes agricultural domain portals and general news portals. The pattern evaluation employs research achievements and statistical data to provide the relevant data. **Table 3.** reports the specific sources.

Table 3. *Data sources used to investigation the CEA patterns.*

	Data sources	Source	Examples
Pattern acquisition	Agricultural domain portals	National official portal Provincial official portals City-level official portals	http://www.moa.gov.cn/ http://coa.jiangsu.gov.cn/ http://nyncj.nanjing.gov.cn/
	General portals	Well-known news portals	https://news.baidu.com/
Pattern evaluation	Research achievements	Research papers	Transparency distribution (D. Liu et al., 2020) NPP (Chen, 2019)
		Research datasets	pH & TDS distribution (Liu Xuyan et al., 2020) Erosion modulus distribution (Resource and Environment Science and Data Center, 2020)
	Statistical data	China statistical yearbook and analysis system	GDP and real GDP per capita (National Bureau of Statistics, 2020a)

Note. Time span of web portals is 2018.1.1-2020.8.31.

3.3. Chinese Ecological Agriculture classification

The WEAPI method identified 75 types of CEA patterns to be discovered, categorized into 4 levels (**Table 4.**). The CEA patterns include four 2nd level classes: ecological farming; ecological breeding; hybrid farming; and innovative agriculture. In particular, ecological farming contains three 3rd level classes (reduced fertilization planting, circular food chain, and 3-D intercropping pattern); ecological breeding consists of fecal resources utilization and stereoscopic cultivation patterns; hybrid farming refers to planting-breeding mix patterns; and innovative agriculture is divided into white agriculture, internet agriculture, and ecotourism patterns. Each of the 3rd level classes include several 4th level classes, whereby 28 new CEA patterns are discovered, compromising 38% of the total CEA patterns (28/75). This indicates that the CEA patterns are continuously changing.

Table 4. *Classification of the 4th level CEA patterns.*

2nd level class	3rd level class	4th level class	
Ecological farming	Reduced fertilization planting pattern	◆ Fertigation	◆ Fertilizer-efficient agriculture
		◆ Drip irrigation	◆ Precise fertilization
		◆ Alley cropping	◇ Protected agriculture
		◆ Rainfall harvesting planting	◇ Remediation farming
		◆ Drought resistance	◇ Original ecological cultivation
		◆ Water-efficient agriculture	◇ Technology-assisted reduced fertilization
	Circular food chain pattern	◆ Crop straw utilization	◆ Biogas utilization
	3-D intercropping pattern	◆ Forest-crop intercropping	◆ Forest-grass intercropping
		◆ Forest-medicine intercropping	◆ Forest-flowers intercropping
		◆ Forest-vegetable intercropping	◆ Forest-bush intercropping
		◆ Forest-seedling intercropping	◆ Mushrooms-grass intercropping
		◆ Forest-mushrooms intercropping	◆ Season inter-planting
Ecological breeding	Fecal resources utilization pattern	◆ Livestock manure recycling	◆ Fermentation bed farming
		◆ Fecal resource-returning field	◆ Fecal resources transformation
	Stereoscopic cultivation pattern	◆ 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
Hybrid farming	Planting-breeding mix pattern	◆ Rice-fish	◆ Crop straw recycling farming
		◆ Rice-livestock	◇ Farming-dispersed breeding
		◆ Forestry-grass-livestock	◇ Rice-frog
		◆ Orchard-livestock	◇ Planting-breeding intercropping
		◆ Free-range livestock farming	◇ Lotus-fish
		◆ Planting-breeding-processing	◇ Crop-livestock-biogas
		◆ Rice-fish-livestock	◇ Mushrooms/grass remediation farming
		◆ Animal-biogas-fruits	◇ Integrated crop-livestock
		◆ Multiple crop-livestock	◇ Crop-livestock recycling
Innovative agriculture	White agriculture pattern	◆ Microbial agriculture	◇ High-tech agriculture
		◆ Agriculture + Internet of Things	◇ High-quality agriculture
		◆ Photovoltaic agriculture	◇ High-quantity agriculture
		◆ Industrial farming/breeding	◇ Farming-processing
		◇ Industrial chain agriculture	◇ Food bank
	Internet agriculture pattern	◇ Agriculture-internet	◇ Contract farming
		◇ Agriculture crowdfunding	◇ Shared agriculture
	Ecotourism pattern	◆ Agricultural ecological park	◇ Picking tourism
			◇ Sci-tech agricultural park

Note. "◆" indicates existing agriculture patterns (manually summarized via pre-2010 data sources) and "◇" indicates new agriculture patterns.

In order to analyze the CEA transformation structure, the new CEA patterns are quantified based on the 2nd and 3rd level classes (**Table 5.**). The statistical analysis reveals the innovative agriculture (70.6%) and hybrid farming (44.4%) as the key CEA transformation components. More specifically, the top 3 3rd level classes are identified as internet agriculture patterns (100%), ecotourism patterns (66.7%), and white agriculture (60%). That indicates internet, tourism, and industry as the main trends of the CEA patterns transformation.

Table 5. *Transformation proportion of the CEA patterns at the 2nd and 3rd level classes.*

2nd level class	Transformation proportion	3rd level class	Transformation proportion
Ecological farming	16.7% (4/24)	Reduced fertilization planting pattern	33.3% (4/12)
		Circular food chain pattern	0% (0/2)
		3-D intercropping pattern	0% (0/10)
Ecological breeding	25% (4/16)	Fecal resource utilization pattern	0% (0/4)
		Stereoscopic cultivation pattern	33.3% (4/12)
Hybrid farming	44.4% (8/18)	Planting-breeding mix pattern	44.4% (8/18)
Innovative agriculture	**70.6% (12/17)**	White agriculture pattern	60.0% (6/10)
		Internet agriculture pattern	**100% (4/4)**
		Ecotourism pattern	66.7% (2/3)

3.4. Chinese Ecological Agriculture distribution

In order to demonstrate the CEA pattern distributions, 25,855 valid CEA pattern records determined via the WEAPI method were classified into four 2nd level classes (ecological farming, ecological breeding, hybrid farming, and innovative agriculture). The maps in **Figure 5.** reveal the explicit differences in trends of the CEA patterns for four of the 2nd level classes. The red regions indicate the CEA pattern clusters, while the black triangles correspond to the CEA pattern records, with their sizes representing the corresponding quantities.

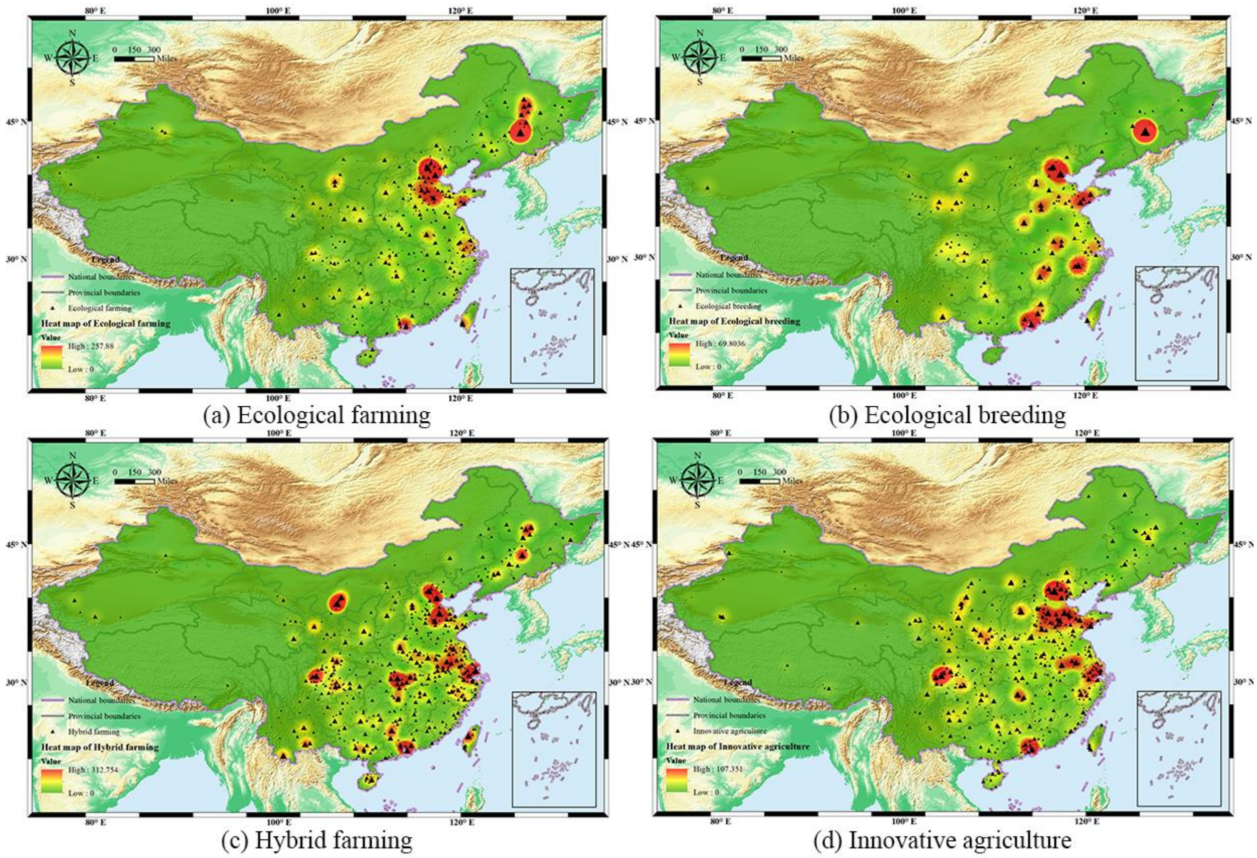


Figure 5. CEA pattern distributions in the four 2nd level classes. (a) ecological farming patterns; (b) ecological breeding patterns; (c) hybrid farming patterns and; (d) innovative agriculture.

Figure 5. a demonstrates two key red regions (Jing-jin-ji around Beijing and Tianjin, and northeast China), indicating that numerous ecological farming patterns are clustered in these areas. More specifically, these two regions are innovation hubs of ecological farming in China. Additional sub-regions colored orange and yellow can be observed to occur in the provincial and regional planting centers (e.g., north Ningxia, central Shanxi, south Jiangsu, southwest Taiwan, etc.). **Figure 5. b** reveals numerous types of ecological breeding to exist in China. In northeast China, the ecological breeding patterns cluster in the northeast plain; in eastern China, key regions are distributed in the Bohai gulf; ecological breeding patterns in southeast China are located along the rivers and lakes; and in central China, the patterns are clustered in the grassland and plain areas. These patterns demonstrate the different Chinese ecological breeding types and their geographic centers. **Figure 5. c** shows five clusters (northeast plain, Huabei plain around the Bohai gulf, Changjiang river basin, Zhujiang river basin, Ningxia plain, and Sichuan basin). These clusters exhibit typical planting and breeding patterns and unique hybrid methods. From **Figure 5. d**, five innovative agriculture centers can be observed, located in the Jing-jin-ji area, the Huabei plain, the Yangtze river delta, the Zhujiang river delta, and the Sichuan basin.

The WEAPI method also reveals the distribution of each popular CEA pattern. **Table 6.** reports the 4th level class CEA patterns based on the CEA classification, representing the top 10 CEA patterns in China during 2018-2020. The top 10 CEA patterns account for 67.56% of the total number of records. **Figure 6.** presents the distribution of each top 10 CEA pattern in the 4th level class determined via the WEAPI method. The distribution of each 4th level CEA pattern class in China is distinct. This indicates that the generation of each CEA pattern requires particular conditions based on the natural resources and social economic environment. The government can employ the WEAPI-generated data and distributions to design an effective plan for future ecological agriculture development, with a clear indication of the CEA patterns and locations that require further investigation.

Table 6. *Proportion of CEA patterns.*

4th level class CEA	Rank	Record number	Proportion (%)	Cumulative proportion (%)
Integrated crop-livestock	1	3572	13.82	13.82
Animal-biogas-fruits	2	2873	11.11	24.93
Rice-fish	3	2723	10.53	35.46
Agricultural ecological park	4	2081	8.05	43.51
Agriculture-internet	5	1216	4.70	48.21
Multiple crop-livestock	6	1174	4.54	52.75
Livestock manure recycling	7	1094	4.23	56.98
Fertigation	8	1037	4.01	60.99
Crop straw utilization	9	887	3.43	64.42
Forestry-grass-livestock	10	811	3.14	67.56
……	……	……	……	……
Rainfall Harvesting Planting	75	11	0.04	100

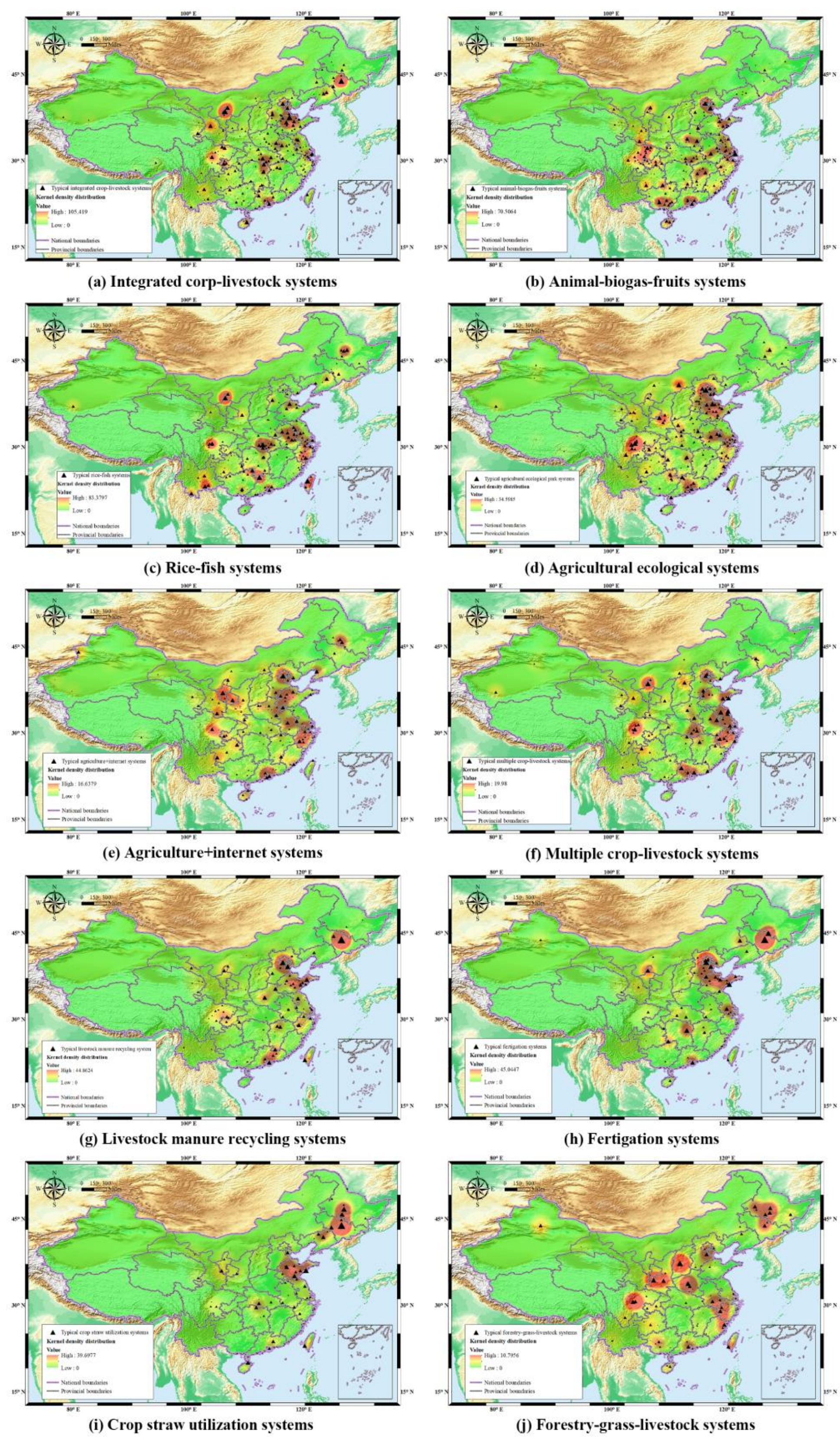


Figure 6. Distributions of the top 10 CEA patterns.

4 Validation experiments

In order to validate the WEAPI method, we tested its precision and coverage.

4.1. Precision validation

Precision validation evaluates the accuracy of the extraction processes. Under the case study of China, we obtained 25,855 valid CEA pattern records, with time, location, and the CEA pattern identified as the core elements. The accuracy of these elements requires verification.

The extensive size of the dataset containing the CEA pattern records does not allow for the manual checking of the records. Thus, a randomly selected data set containing 260 CEA pattern records is used to evaluate the accuracy of the results. The extracted time, locations, and pattern information are manually annotated for each CEA pattern record, and the precision rate is equal to the proportion of the correct number to the total record number. **Table 7.** reports the statistical results. All WEAPI extraction precision rates exceed 95%.

Table 7. *Extracted accuracy of the WEAPI method.*

Type	Record number (RN)	Error number (EN)	Precision rate
Temporal information	260	0	100%
Location	260	12	95.4%
Extracted pattern	260	10	96.2%

4.2. Coverage validation

Coverage validation evaluates the coverage of the CEA pattern records, namely, it verifies the ability of the CEA pattern records to represent the actual local CEA pattern. In order to calculate the CEA pattern record coverage rate, the experiment selects a 4th level ranked CEA pattern (agricultural ecological park), which has an official outstanding list to compare with (Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2017). Additional Chinese national agricultural ecological parks are identified from the relevant literature (Bao, 2013; F. Wang, Wang, & Chen, 2016). **Table 8.** reports the coverage rate of the WEAPI results.

Table 8. *Coverage rate of the WEAPI results.*

Compared dataset	Coverage rate
Official national rural innovation park list	87.03%
Agricultural ecological park list in literature	87.23%
Average	**87.13%**

Note. Coverage rate denotes the proportion of the WEAPI record locations to the locations of the compared dataset and is determined at the town level, indicating that the location errors do not surpass the town area. Different coverage rates are discussed in Section 5.1.

The results of the validation experiments reveal the potential of the WEAPI method as a promising novel rapid technique for the investigation of EA patterns. Despite the minimal errors caused by the extensive amount of automatic crawling and parsing, the whole parsing precision and coverage exceed 95% and 87%, respectively.

5 Discussion

The WEAPI method is demonstrated as powerful tool that can rapidly investigate the ecological agriculture pattern records with fine accuracy and coverage, and reveal the distributions of EA patterns across different levels. However, the proposed method requires improvement in three areas: the ambiguity toponym descriptions of EA patterns in the news; the influence of time labels; and the limited sparse EA pattern records.

5.1. Ambiguity toponym descriptions of EA patterns in the News

The toponym descriptions of EA patterns, which form the basis of locating the EA patterns, are associated with a level of ambiguity. For example, the description of "Rudong town (Yangxi county, Guangdong province) adopts integrated crop-livestock pattern" is always replaced by descriptions of "Yangxi/Guangdong adopts integrated crop-livestock pattern" in different news platform levels. This results in the positional deviations between the real EA pattern occurrences and shifted county/city/province centers. Thus, the extracted EA pattern locations determined by the WEAPI method experience a shift (**Figure 7.**).

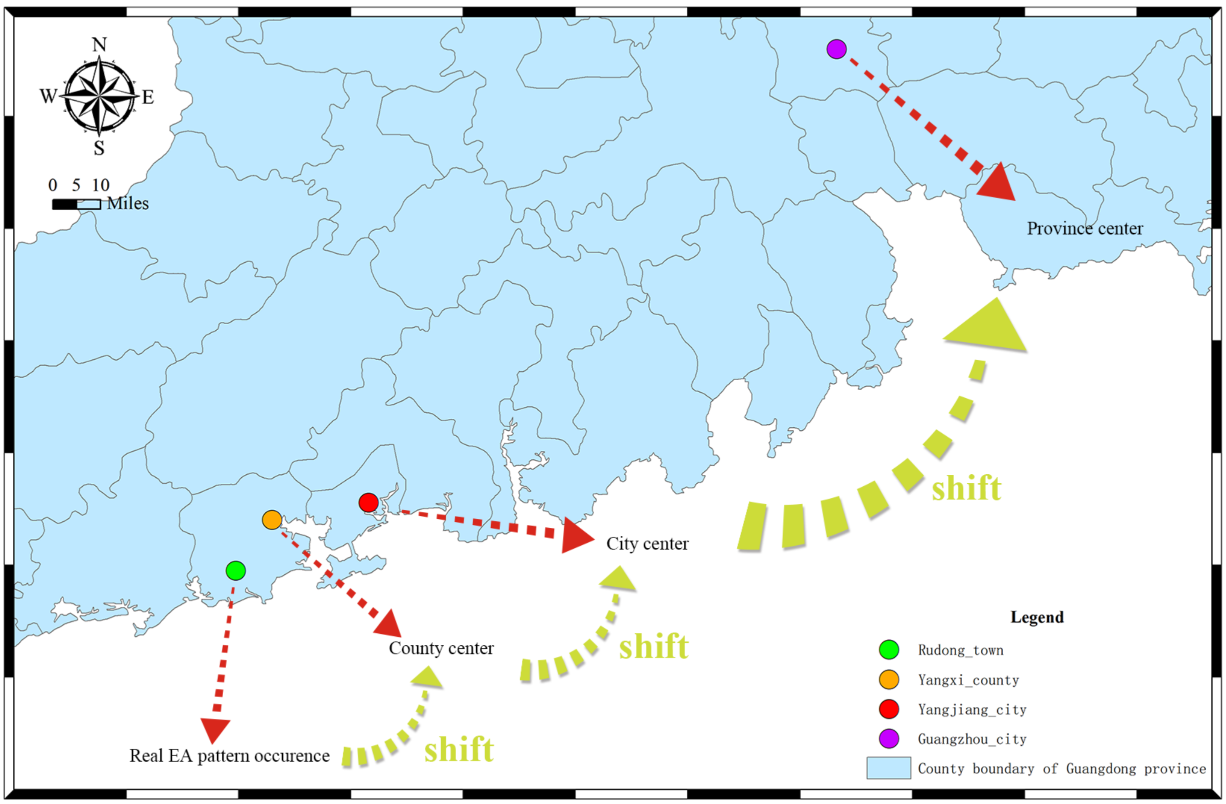


Figure 7. Shift phenomenon of EA pattern locations.

Such a shift results in variations in WEAPI coverage rates across evaluation levels (e.g., town level, city level, and province level, **Table 9.**). The WEAPI element association algorithm (**Figure 3.**) is not able to fully fix these shifts in web texts as ambiguous toponym descriptions are common. Hence, the shifts can be measured using the coverage rates at different levels.

Table 9. *Coverage rates of the WEAPI results at different evaluation levels.*

Compared dataset	Coverage rate (town level)	Coverage rate (city level)	Coverage rate (province level)
Official national rural innovation park list	87.03%	92.59%	100.00%
Agricultural ecological park list in literature	87.23%	91.49%	100.00%
Average	**87.13%**	**92.08%**	**100.00%**

5.2 Influence of the time label in EA pattern records

The time label records the report time of the EA pattern and is crucial for the WEAPI-determined EA pattern records. The results of the case study demonstrate that the EA pattern distributions can be directly represented to reveal the current spatial pattern. However, the distribution varies with time. **Figure 8.** demonstrates the temporal evolution of the "agriculture-internet" pattern, indicating mushroom growth in the quantity and coverage.

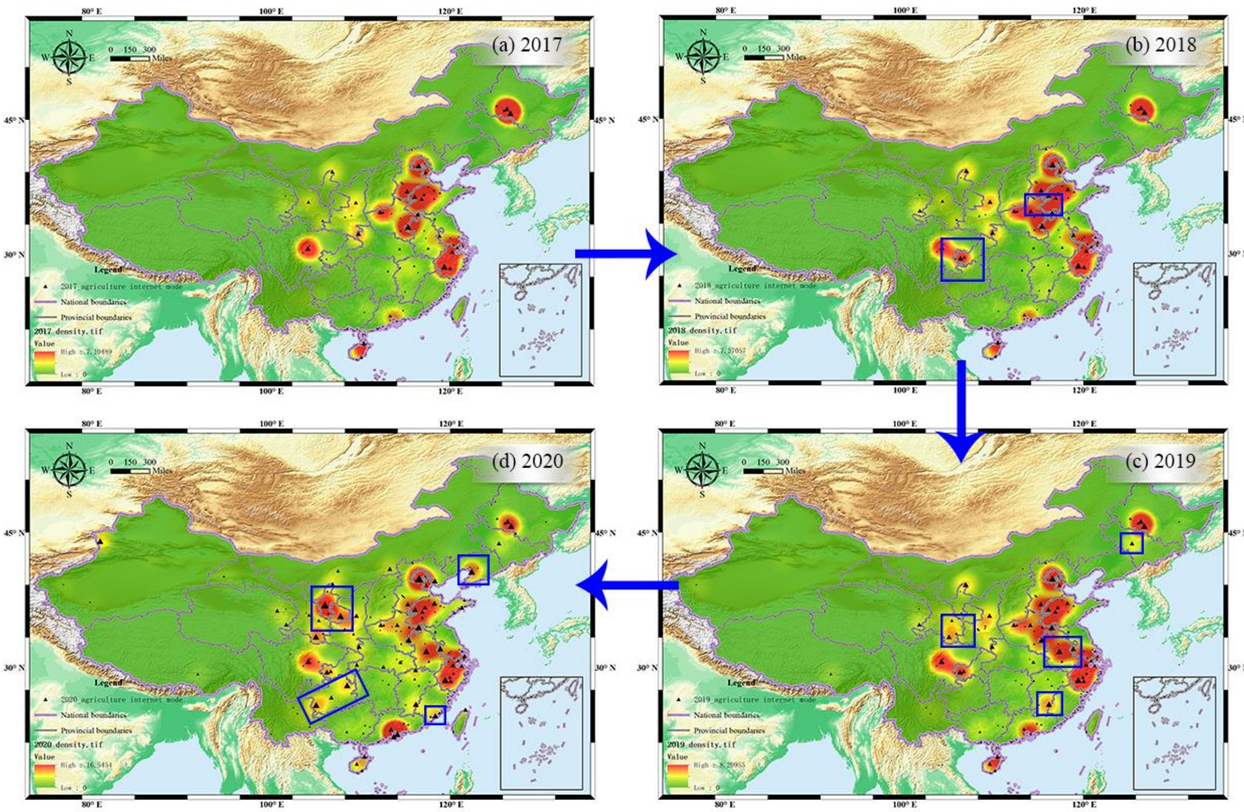


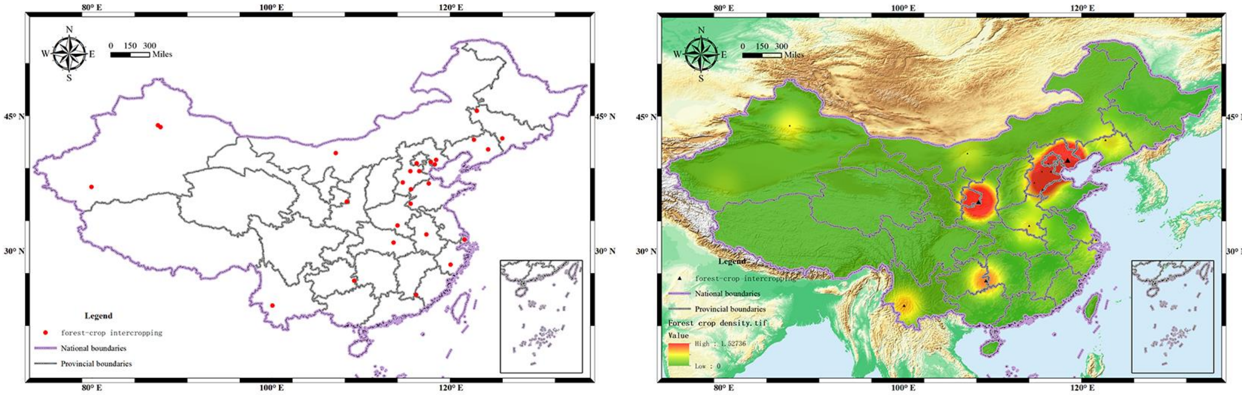
Figure 8. Annual evolution of the agriculture-internet pattern during 2017-2020: (a)-(d) represent 2017-2020.

Figure 8. demonstrates the entire evolution of the agriculture-internet pattern in China during 2017-2020. Pioneer clusters (**Figure 8. a**) and newborn clusters (blue rectangles in **Figure 8.** b**, c, and d**) reveal trends in the agriculture-internet pattern clearly. This indicates that the WEAPI method can not only investigate EA pattern records, but also supports the analyses of the evolution and trends in the EA patterns.

Note that the WEAPI EA pattern records rely on web-based news. However, the news of EA patterns is time sensitive and thus partial reports will be replaced by newer information. The evolution analyses of EA patterns hence require a dynamic monitoring system based on the WEAPI method. This challenge task is reserved for future work.

5.3. Limitations of sparse EA pattern records

The sparse EA pattern record phenomenon denotes the limited amount of records in some of the 4th class EA patterns, which are unable to effectively represent the corresponding distribution. For example, the 4th level class CEA pattern "forest-crop intercropping" consists of 69 records, which is much less than the 2nd class CEA pattern "ecological farming" with 6493 records and the top 4th level class CEA pattern "integrated crop-livestock" with 3572 records. **Figure 9.** presents the spatial and density distributions of the "forest-crop intercropping" pattern. The CEA pattern clusters are revealed in **Figure 9. b**, yet the results are not stable. WEAPI users must be informed of this limitation.



(a) locations (b) distribution

Figure 9. Locations and distribution of the "forest-crop intercropping" pattern.

This limitation is attributed to the corresponding data sources. News reports that record the specific CEA patterns are limited. This issue can be solved by fusing multiple data sources, such as the official reports, social media, etc.. This is a continuous research theme of the next generation WEAPI method.

6 Conclusions

The current research proposes a novel rapid Web-Text Based Ecological Agriculture Pattern Investigation (WEAPI) method. We apply WEAPI to the national region of China to perform the classification of the Chinese Ecological Agriculture (CEA) patterns and reveal the national scaled distributions of the 2nd level CEA patterns. The WEAPI method is also able to investigate popular CEA patterns, for example, the current top 10 CEA patterns in the 4th level class, and the evolution of specific CEA patterns. The results reveal the potential of WEAPI as a powerful tool for the investigation of EA patterns. Two key conclusions were determined from our work:

- Web text can be applied to investigate EA patterns with web crawling and NLP techniques, achieving over 95% precision in pattern parse processes and 87% coverage at the town level for the official CEA pattern list.
- WEAPI can rapidly investigate EA patterns and determine their evolution. This is of great significance to agricultural research and agricultural development planning. Furthermore, a dynamic EA pattern investigation system can be implemented via the WEAPI method.

Currently, the functionality of WEAPI is specific to China, however, overcoming the cross-language problem will allow is application for the determination of global EA pattern distributions.

Acknowledgement

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences [grant number XDA23100101] and National Natural Science Foundation of China [grant number 42050101].

Data availability statement

Supporting codes and data behind the figures can be obtained with a doi: (DOI:10.5281/zenodo.4871390).

References

Ali, M. P., Bari, M. N., Haque, S. S., Kabir, M. M. M., Afrin, S., Nowrin, F., . . . Landis, D. A. (2019). Establishing next-generation pest control services in rice fields: eco-agriculture. *Sci Rep, 9*(1), 10180. doi:10.1038/s41598-019-46688-6

Atinkut, H. B., Yan, T., Zhang, F., Qin, S., Gai, H., & Liu, Q. (2020). Cognition of agriculture waste and payments for a circular agriculture model in Central China. *Sci Rep, 10*(1), 10826. doi:10.1038/s41598-020-67358-y

Baidu. (2020). Baidu Geocoding Service. http://api.map.baidu.com/geocoding/v3/?address=

Bao, W. (2013). *Research on Resource Development and Industrialization of Agri-tourism in China.* (Doctor), Ocean University of China, Qingdao. (175-178)

Burton, R. J. F., & Riley, M. (2018). Traditional Ecological Knowledge from the internet? The case of hay meadows in Europe. *Land Use Policy, 70*, 334-346. doi:https://doi.org/10.1016/j.landusepol.2017.10.014

Chen, P. (2019). Monthly NPP dataset covering China's terrestrial ecosystems at north of 18°N (1985–2015). *Journal of Global Change Data & Discovery, 3*(1), 34-41. doi:10.3974/geodp.2019.01.05

González-Chang, M., Wratten, S. D., Shields, M. W., Costanza, R., Dainese, M., Gurr, G. M., . . . Zhou, W. (2020). Understanding the pathways from biodiversity to agro-ecological outcomes: A new, interactive approach. *Agriculture, Ecosystems & Environment, 301*, 107053. doi:https://doi.org/10.1016/j.agee.2020.107053

Haynes, K. (2010). Other lives in accounting: Critical reflections on oral history methodology in action. *Critical Perspectives on Accounting, 21*(3), 221-231. doi:https://doi.org/10.1016/j.cpa.2009.11.002

Leeuwenberg, A., & Moens, M.-F. (2019). A Survey on Temporal Reasoning for Temporal Information Extraction from Text. *Journal of Artificial Intelligence Research, 66*, 341-380. doi:https://doi.org/10.1613/jair.1.11727

Li, M., Zhang, Y., Xu, M., He, L., Liu, L., & Tang, Q. (2019). China Eco-Wisdom: A Review of Sustainability of Agricultural Heritage Systems on Aquatic-Ecological Conservation. *Sustainability, 12*(1).

Liu, D., Duan, H., Loiselle, S., Hu, C., Zhang, G., Li, J., . . . Han, W. (2020). Observations of water transparency in China's lakes from space. *International Journal of Applied Earth*

Observation and Geoinformation, 92, 102187. doi:https://doi.org/10.1016/j.jag.2020.102187
Liu, P., Moreno, J. M., Song, P., Hoover, E., & Harder, M. K. (2016). The Use of Oral Histories to Identify Criteria for Future Scenarios of Sustainable Farming in the South Yangtze River, China. *Sustainability, 8*(9), 859.
Liu Xuyan, Tang Xinzhai, Zhu Zhilin, Yuan Guofu, Zhang Xinyu, Wang Jiao, . . . Munanhidden, Z. (2020). A dataset of water pH and total dissolved solid contents of Chinese Ecosystem Research Network during 2004–2016. *China Scientific Data, 5*(3). doi:10.11922/csdata.2019.0067.zh.
Liu, Y., Duan, M., & Yu, Z. (2013). Agricultural landscapes and biodiversity in China. *Agriculture, Ecosystems & Environment, 166*, 46-54. doi:https://doi.org/10.1016/j.agee.2011.05.009
Lockeretz, W. (1989). Problems in evaluating the economics of ecological agriculture. *Agriculture, Ecosystems & Environment, 27*(1), 67-75. doi:https://doi.org/10.1016/0167-8809(89)90073-X
Ministry of Agriculture and Rural Affairs of the People's Republic of China. (2017). Notice of the Ministry of Agriculture on the Announement of the Catalogue of National Rural Innovation Parks. Retrieved from http://www.moa.gov.cn/nybgb/2017/dqq/201712/t20171230_6133922.htm
National Bureau of Statistics. (2020a). China statistical yearbook. Retrieved from https://data.cnki.net/YearData/Analysis
National Bureau of Statistics. (2020b). National Administrative Area Divisions and Codes. Retrieved from http://www.stats.gov.cn/tjsj/tjbz/tjyqhdmhcxhfdm/2020/index.html
Pan, D., Yang, J., Zhou, G., & Kong, F. (2020). The influence of COVID-19 on agricultural economy and emergency mitigation measures in China: A text mining analysis. *PLOS ONE, 15*(10), e0241167. doi:10.1371/journal.pone.0241167
Priyadarshini, P., & Abhilash, P. C. (2020). Policy recommendations for enabling transition towards sustainable agriculture in India. *Land Use Policy, 96*, 104718. doi:https://doi.org/10.1016/j.landusepol.2020.104718
Resource and Environment Science and Data Center. (2020). Erosion modulus distribution in China. Retrieved from www.resdc.cn/data.aspx?DATAID=259
Riley, M., & Harvey, D. (2007). Talking geography: on oral history and the practice of geography. *Social & Cultural Geography, 8*(3), 345-351. doi:10.1080/14649360701488765
The Ministry of Agriculture of the People's Republic of China. (2003). The top 10 Chinese Ecological Agricultural modes or technologies. *Journal of Agricultural Resources and Environment*(1), 16.
Velten, S., Leventon, J., Jager, N., & Newig, J. (2015). What Is Sustainable Agriculture? A Systematic Review. *Sustainability, 7*(6), 7833-7865.
Wang, F., Wang, K., & Chen, T. (2016). National agritourism parks in China: Distribution, types and spatial optimization. *Research of Agricultural Modernization, 37*(06), 1035-1044. doi:doi:10.13872/j.1000-0275.2016.0095
Wang, H., Qin, L., Huang, L., & Zhang, L. (2007). Ecological Agriculture in China: Principles and Applications. In D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 94, pp. 181-208): Academic Press.

Wang, L., Wang, Y., & Chen, J. (2019). Assessment of the Ecological Niche of Photovoltaic Agriculture in China. *Sustainability, 11*(8), 2268.

Yang, M., Dong, S., Dong, Q., Xu, Y., Zhi, Y., Liu, W., & Zhao, X. (2021). Trade-offs in ecological, productivity and livelihood dimensions inform sustainable grassland management: Case study from the Qinghai-Tibetan Plateau. *Agriculture, Ecosystems & Environment, 313*, 107377. doi:https://doi.org/10.1016/j.agee.2021.107377

Ye, X. J., Wang, Z. Q., & Li, Q. S. (2002). The ecological agriculture movement in modern China. *Agriculture, Ecosystems & Environment, 92*(2), 261-281. doi:https://doi.org/10.1016/S0167-8809(01)00294-8

Yi, S. (2019). Contingent Valuation of Sustainable Integrated Agriculture–Aquaculture Products: The Case of Rice–Fish Farming Systems in South Korea. *Agronomy, 9*(10), 601.

Zhang, H., Miao, J., Liu, Z., Wesson, I. L., & Shang, J. (2020). *NLPIR-Parser: Making Chinese and English Semantic Analysis Easier and Complete*. Paper presented at the 15th International Conference on the Statistical Analysis of Textual Data, France.

Zheng, H., Huang, H., Chen, C., Fu, Z., Xu, H., Tan, S., . . . Tang, J. (2017). Traditional symbiotic farming technology in China promotes the sustainability of a flooded rice production system. *Sustainability Science, 12*(1), 155-161. doi:10.1007/s11625-016-0399-8