Data in Brief 53 (2024) 110269

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Survey data on livelihoods and inputs and outputs of crop production in Quzhou county on the North China Plain

Zhan Xu[a,b,d,*], Fan Li[c,e], Jiali Cheng[a,b,d], Zhengyuan Liang[a,b,d], Jeroen C.J. Groot[b], Wopke van der Werf[a], Chaochun Zhang[d]

[a] *Centre for Crop Systems Analysis Group, Wageningen University & Research, P.O. Box 430, 6700 AK Wageningen, the Netherlands*
[b] *Farming System Ecology Group, Wageningen University & Research, P.O. Box 430, 6700 AK Wageningen, the Netherlands*
[c] *Development Economics Group, Wageningen University & Research, Hollandseweg 1, 6706 KN Wageningen, the Netherlands*
[d] *State Key Laboratory of Nutrient Use and Management, College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China*
[e] *College of Economics and Management, Huazhong Agricultural University, Wuhan 430070, China*

A R T I C L E I N F O

Article history:
Received 23 November 2023
Revised 20 February 2024
Accepted 23 February 2024
Available online 28 February 2024

Dataset link: Survey data on livelihoods and inputs and outputs of crop production activities in Quzhou county on the North China Plain (Original data)

Keywords:
Village typology
Cropping system
Sustainability
Input-output

A B S T R A C T

Farmers' decisions on crop choice, management practices, and livelihood strategies are essential to agricultural sustainability. This data article describes three datasets on crop production in Quzhou, a county in the central part of North China Plain. The three datasets cover different scales. The village dataset assembles basic data on all 342 villages of Quzhou county, including information on population, land area, crop grown, labour, irrigation and markets. Data was sourced from the yearbook data of 2017 and a village cadres survey in 2018. The village dataset was used to create a village typology from which 35 villages belonging to seven village types (five for each type) were selected for stratified random sampling to collect information on farm characteristics and cropping practices. We surveyed these 35 villages, interviewing fifteen farmer households per village (525 in to-

* Corresponding author.
E-mail address: zhan.xu@wur.nl (Z. Xu).

https://doi.org/10.1016/j.dib.2024.110269

tal) in 2020. The interviewees represented two farm management models: smallholder farms and business farms. The resulting household dataset provides farm-level data, including demographic data of farming decision-makers and the number of household members, land use and machinery resources, crop production management practices, and government subsidies. The crop-level dataset was derived from the household survey and included input-output inventories for each crop grown during one year on each field greater than 1/30th ha (1/2 mu) on the 525 surveyed farms within a year. This dataset comprises information on cropping practices in 1352 fields. The three datasets provide a basis for analyses on cropping practices and sustainability attributes of farms and crops in a typical agricultural county on the North China Plain.

Specifications Table

Subject	Agriculture science
Specific subject area	Crop diversity, Crop production, Management practice, Sustainability
Data format	Raw data
Type of data	CSV files ($n = 5$): i) Village data, ii) Farm household data, iii) Crop data, iv) Codebook_village data, v) Codebook_household and crop data Tables and Figures
How the data were acquired	The data are in three datasets: (1) village dataset, (2) farm dataset, (3) crop dataset. The village-level data (dataset 1) was composed of two sources: yearbook [1] data and village cadres survey, including village Chinese Communist Party (CCP) secretary, village monitor and village accountant. Data for datasets 2 and 3 were collected by conducting personal interviews using a structured questionnaire. We interviewed 525 households in Quzhou county on the North China Plain. These 525 households originated from 35 villages, 5 villages from each of 7 village clusters identified by analysing dataset 1. The dataset consists of 505 valid observations after checking for missing values and potential errors.
Description of data collection	In the first step, we grouped all 342 villages in Quzhou in eight clusters based on similar crop production and socio-economic characteristics according to yearbook data [1] and village cadres survey data. A random sample of five villages was selected for each of seven village types, excluding one semi-urban village type without crop production. We surveyed fifteen farms from each of the 35 villages, including smallholder farms and business farms. We obtained the list of registered business farms from the local government and all the business farms present in the surveyed villages were included in the interview. There were up to two business farms per village, and all business farms in a village were included. Twenty-one of the 35 villages did not have business farms. A random selection was made from the remaining smallholder farms in a village to obtain a sample of fifteen farms per village. A structured questionnaire was administered to all households. In this survey, we collected information on the household and agricultural practices on all plots greater than 0.5 mu (1/30th ha) on a farm. Perennial crops and greenhouse crops were excluded from the survey.
Data source location	Institution: College of Resources and Environmental Sciences, China Agricultural University City/Town/Region: Quzhou County of Handan City, Hebei Province Country: China Latitude and longitude: 36.58 and 114.83

(*continued on next page*)

Data accessibility	Repository name: Mendeley Data Data identification number: Doi:10.17632/jp3v9859cx.1 Direct URL to data: https://data.mendeley.com/datasets/jp3v9859cx/1

1. Value of the Data

- This dataset contains the first publicly available data on cropping practices for a very wide range of crop systems (62 one-year cropping systems were identified) in a representative county on the North China Plain. The North China Plain is an area that is of prime importance for China's wheat and maize grain production, but it also has important production of cotton and a vast range of minor crops.
- The dataset with village characteristics and the associated typology results can be used to select representative villages based on socio-economic data, crop patterns and crop production-related resources when further surveys are done in Quzhou.
- The farm-level data are valuable for analyzing sustainability and input/poutput relationships of crop production in Quzhou as a representative village on the North China Plain.
- Farm level data are furthermore valuable for analysing performance of farm management strategies at the farm household level in response to policy interventions, farm size, socioeconomic conditions, and household demographics
- The crop inventory data was collected for all crops grown within a year on each farm field $>$ 0.5 mu (1/30th ha) operated by the household. So the farm level and crop level data can be combined to be used as a benchmark, and it can be used to analyze the production efficiency of farms.
- The crop level data can be used to design new rotations to better meet stakeholder demands and sustainability objectives.
- The datasets can be a reference source for making policies to promote crop production practices that contribute to improved agricultural sustainability in high-input farming systems.

2. Objective

The objective of the household survey is to provide a comprehensive picture of crop production practices and rural livelihoods in Quzhou county on the North China Plain. The county, located in the center of the North China Plain (Fig. 1), is representative of the North China Plain in terms of its land-use, production practices and sustainability challenges. It was selected in 2019 as an agriculture green development (AGD) focus area [2]. The AGD program aims to explore the sustainability challenges, to promote a series of promising solutions and to share practical experiences across regions with similar conditions throughout China [3]. Specifically, we focused on which crops were grown, and how farmers cultivated these crops. Although the North China Plain is one of the food bowls in China, most previous studies focused on the main cropping system, double cropping of wheat and maize, in which wheat is sown in October and harvested in June the following year while maize is sown immediately after wheat harvest in June and harvested in October. Other crop production activities have not previously been described in detail.

3. Data Description

This data paper reports on three datasets in Quzhou county, covering village, farm household, and crop levels. Data from the 2017 village yearbook, village cadre survey data, and village typology results were included in the village dataset. The farm household dataset provides

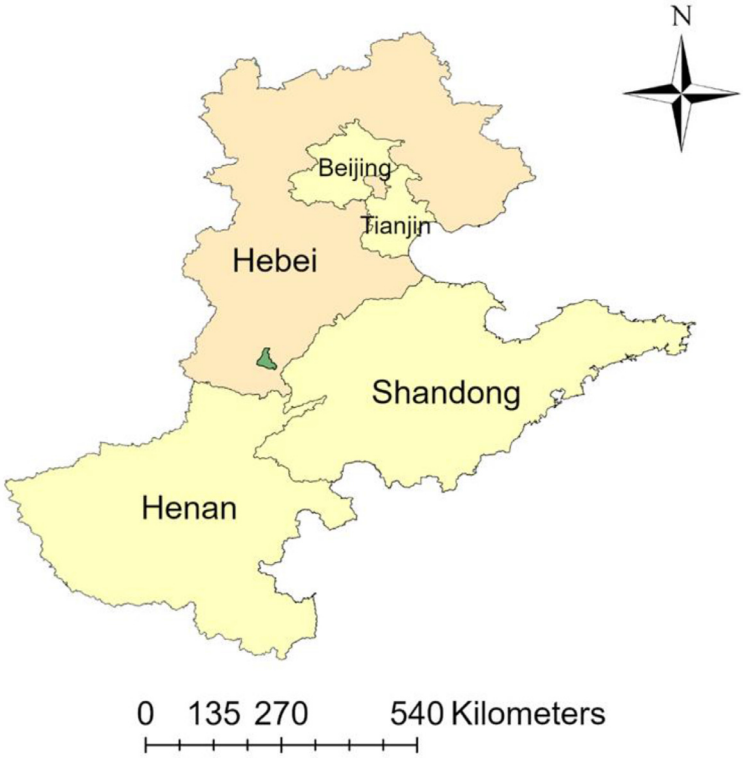

Fig. 1. Map of the North China Plain. The green polygon in south Hebei is Quzhou county.

Table 1
Definition of business farms and three subtypes of business farms included in the dataset.

Farm management models	Characteristics
Business farm	Business farms prioritize economic benefits from marketing their farm produce. Business farms have a relatively large farm size (more than 3 hectares per farm [3]), and they are usually run by a professional manager. Business farms need to register with the local government to obtain a business license.
Family farm	Family farms are run by a family. They hire seasonal labour as needed and the farm size is mostly larger than 50 mu (3.3 hectares) [4].
Cooperative farm	Cooperative farms typically consist of at least five family farms [5] investing together in machinery and wages for hired labour.
Agricultural company	Agricultural companies have the largest farm size. They employ a professional manager for marketing and sales [6]. Some agricultural companies have their own processing facilities.

information on data collected from 505 household surveys across 35 villages on various issues, including the households' socioeconomics, crop production decision makers' characteristics, and crop production information for each major plot (>0.5 mu) on the farm. There are two types of farm management in this dataset: the smallholder model and the business model. The business farms are divided into three subtypes: family farms, cooperative farms, and agricultural companies (Table 1). Based on the data from the household survey, we generated a crop dataset which provides input-output data of all crops grown on the surveyed farms within a year. Five

Table 2
Definition of five cropping system types covered in the dataset.

Cropping system type	Characteristics	Area percentage (%)
Single cropping	A single crop is grown in one plot in one year. E.g. cotton.	43.9 %
Relay intercropping	Relay intercropping entails growing two crop species in the same parcel, but the growing period of the two crops overlaps only partially. E.g. cotton/mung bean relay intercropping.	0.4 %
Double cropping	Double cropping means harvesting two full crops in a calendar year in the same parcel and there is no overlap in growing period. E.g. wheat-maize double cropping.	53.8 %
Relay double cropping	Relay double cropping includes three crops in a parcel in a year, including one crop with a long growing period and two crops with a short growing period. The latter two crops are planted as a double cropping and they are combined with the long growth period crops in a strip intercrop. E.g. cotton-watermelon-cauliflower where cotton is the long season crop and watermelon and cauliflower are grown in the same strip after each other.	0.6 %
Triple cropping	Triple cropping covers three crops grown after each other in one year. E.g. a sequence of three cabbage crops in the same field in one year.	1.3 %

cropping system types (Table 2) and 28 crop species are included in our dataset. Two codebooks are available as supplemental files. The first, *Codebook_village data*, includes units, descriptions, survey questions and the sources for the variables in the village dataset. The second, *Codebook_household and crop data*, includes units, descriptions, and survey questions for the variables in the farm household dataset and crop dataset. The following sections present each dataset in more detail.

3.1. Village dataset

The village dataset comprises 342 records, one for each village in Quzhou, and 16 variables. Fifteen variables were used to build a village typology and the result of the village typology is the 16th variable. The fifteen key variables represented three aspects: village structure (e.g. population, number of households), area of a few main crops (wheat, maize, cotton, vegetables and oil crops) within a village, and crop production-related resources (e.g. land area, irrigatable land, distance to city). Seven variables were derived from the village cadre survey, and the remaining eight were extracted from the yearbook [1] (Table 3). The village cadre survey was conducted in January 2018. These village cadres included the village Chinese Communist Party (CCP) secretary, a village monitor, and an accountant. These cadres collect and manage village statistics and are therefore a primary source of information. The village cadre survey focused on villages' resources endowment, agricultural production structure (crop production and livestock production), agricultural pollutants, and rural waste management. Only the variables related to crop production were included in the village dataset.

Fig. 2 shows the percentage of arable land per village that was cultivated with winter wheat, maize, cotton, vegetables, and oil crops, which are the main crops cultivated in Quzhou. There was large heterogeneity in crop proportions across the county. For instance, vegetable and oil crops were widely cultivated in only a few villages. The areas of maize, cotton, vegetables, and oil crops were sourced from yearbook data, and the wheat area was sourced from the village cadre survey. It was recorded in both the yearbook and the village cadre data that crop production information applied to the year 2017. Thirty villages did not have wheat area data since twelve villages were excluded from the cadre survey because the number of households was too

Table 3
Village dataset variables.

Variable	Unit	Description	Source
Structure			
Household number	Number	Total number of households	Yearbook
Smallholder farmland	%	Percentage of arable land operated by smallholder farms to total arable land	Village cadre survey
Family farms	Number	Total number of family farms	Village cadre survey
Cooperative farms	Number	Total number of cooperative farms	Village cadre survey
Crop pattern			
Area with winter wheat	%	Percent arable land with wheat[1]	Village cadre survey
Area with maize	%	Percent arable land with maize[1]	Yearbook
Area with vegetables	%	Percent arable land with vegetables	Yearbook
Area with cotton	%	Percent arable land with cotton	Yearbook
Area with oil crops	%	Percent arable land with oil crops	Yearbook
Resources			
Arable land	ha	Area of arable land	Yearbook
Irrigatable land	%	Percentage of irrigated land (includes well irrigated and surface water irrigated) to total arable land	Yearbook
Off-farm labour ratio	%	The ratio of off-farm labour to total labour	Village cadre survey
Average income	CNY capita^{-1} year^{-1}	Average net income per capita per year	Village cadre survey
Distance to the city center	km	The distance from the village council to the city council of Quzhou via roads	Village cadre survey
Distance to road net	km	The distance from the village council to the road net. The road net, which includes multiple highways, serves as a transportation network connecting urban and rural areas.	Village cadre survey

[1] Wheat and maize are usually cultivated in a double cropping sequence. Hence, the land can be cultivated twice in a year if the land is cultivated with wheat-maize double cropping. Due to the existence of double and even triple cropping, the crop percentages can add up to more than 100 %.

small (below 100), or because of a high degree of urbanization. Another 18 questionnaires were excluded due to low quality.

3.2. Farm household dataset

The farm household dataset reports farm-level data and was collected in Quzhou County in July 2020. The household data contained 505 records, including data from 486 smallholder farms, 14 family farms, and five cooperative farms. A structured questionnaire was administered to all farm decision-makers in person to collect crop production activity data during a one-year cropping cycle (June 2019 to June 2020). There were four parts to the questionnaire. The answers to the questions in part 1 of the questionnaire resulted in 60 columns in the resulting datafile. Demographic information is divided into two parts. The first part is information on all household members, including the number of household members, their age and sex, occupation, and off-farm income. The second sub-part is information about the person who makes the farming decisions, and we asked about their farming experience, and the number of schooling years. The answers to the questions in part 2 resulted in 17 columns and describes resources, such as land allocation and self-owned machinery. A farm's resources include allocated lands,[1] rental lands,[2]

[1] In China, the rural land is collectively owned by villagers and all of the cropland is allocated for use to the rural population on a pro-rata basis. This cropland is allocated land.

[2] Some villagers choose not to manage part or all of their allocated cropland, e.g. because of their age, and rent it out to earn income. This cropland is rental land.

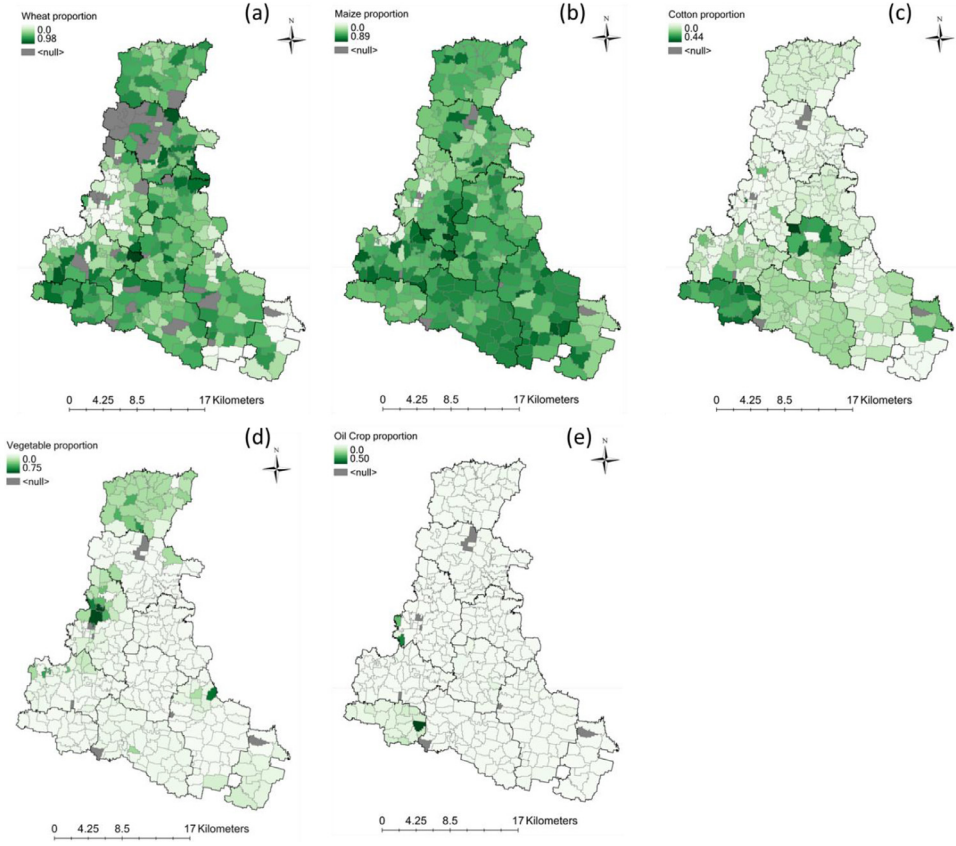

Fig. 2. Percentage of the arable land in Quzhou, cultivated with different crops: (a) wheat, (b) maize, (c) cotton, (d) vegetables, and (e) oil crops. In Quzhou, the typical oil crops are peanuts, soybean, and sunflower grown for oil. Note that wheat and maize are predominantly cultivated in a winter wheat-maize double cropping system.

and rented lands[3] for every cropping system. A detailed input-output inventory for crop production is provided in Part 3, which consists of 864 columns. Additionally, Part 4 includes data on subsidies and land leasing related to crop production. The questions of this part has 9 columns.

Results of the survey indicate that the farmers used five types of one-year cropping system: single cropping, relay-intercropping, double cropping, relay-double cropping, and triple cropping (Table 2). Double cropping and single cropping occupied approximately 54 % and 44 % area of our sample, respectively, and the other three cropping systems occupied a small area percentage. In our data set, the dominant double cropping was wheat-maize double cropping (51 % of the area). The dominant crop species in single cropping was cotton (23 % of the total area).

Table 4 describes the critical socioeconomic characteristics of households and decision-makers. The number of household members of business farms has thirteen missing values since these farms were not run by a family and were operated by several families.

Fig. 3 shows the distribution of farm size as reported by survey respondents. It shows that the mean farm size of sampled farms was 1.23 ha per farm and the median farm size was 0.67 ha per farm. The smallholder farms' mean farm size was 0.71 ha per farm, and the median was

[3] Some villagers want to enlarge their cropland to earn more money. They rent cropland from other villagers. This cropland is rented land.

Table 4
Socioeconomics characteristics of the respondents.

Variable	Category	Frequency (% of valid)	Valid	Missing
Household characteristics				
Household size	1–5	366 (74.4 %)	492	13
	6–10	124 (25.2 %)	(97.4 %)	(2.6 %)
	Above 10	2 (0.4 %)		
Off-farm work	Participating	272 (53.9 %)	505	0
	Not participating	233 (46.1 %)	(100.0 %)	(0.0 %)
Characteristics of farming decision maker				
Gender	Male	374 (74.0 %)	505	0
	Female	131 (26.0 %)	(100.0 %)	(0.0 %)
Age	Youth (18–35)	6 (1.2 %)	505	0
	Adult (36–60)	248 (49.1 %)	(100.0 %)	(0.0 %)
	Elderly (>60)	251 (49.7 %)		
Level of education	No schooling	137 (27.1 %)	505	0
	Primary	146 (28.9 %)	(100.0 %)	(0.0 %)
	Middle	172 (34.1 %)		
	Senior	46 (9.1 %)		
	Bachelor	4 (0.8 %)		
Years of farming experience	1–10	25 (4.9 %)	505	0
	10–20	49 (9.7 %)	(100.0 %)	(0.0 %)
	20–30	114 (22.6 %)		
	>30	317 (62.8 %)		

Table 5
Number of farms with one, two, three or more than three cropping systems. For each case, the table lists in how many cases specific cropping systems (rows) are present on the farm. For instance, of 212 farms have two cropping systems, 174 had cotton, and 191 had winter wheat-maize double cropping.

Cropping systems	Number of cropping systems per farm			
	1 (n = 228)	2 (n = 212)	3 (n = 45)	>3 (n = 20)
Cotton	28	174	37	11
Stevia	4	17	11	5
Chili	3	5	9	6
Maize	4	10	2	3
Vegetable	0	2	6	2
Other single cropping	2	8	8	6
Cotton/mung bean	0	6	5	3
Cotton/ watermelon	0	0	1	1
Winter wheat-maize	186	191	37	12
Vegetable-Vegetable	0	0	3	12
Maize-vegetable	0	3	3	5
Winter wheat-vegetable	0	2	1	1
Other double cropping	0	6	5	8
Relay double cropping	0	0	0	3
Vegetable-vegetable-vegetable	1	1	13	7
Vegetable-maize-vegetable	0	0	3	0

0.63 ha per farm. The business farms' mean farm size was 14.5 ha per farm, and the median was 13.3 ha per farm. Eighty-six out of 505 farms (17 %) rented out their croplands to others and about 22 % of farms rented in croplands from other farms.

Table 5 presents the number of cropping systems cultivated per farm and lists the cropping systems. Forty-five percent of farms grew only one cropping system, usually double cropping of wheat and maize. A total of 212 of the 505 farms cultivated two cropping systems, with wheat-maize double cropping and single cotton being the most commonly used. In 12 % of farms,

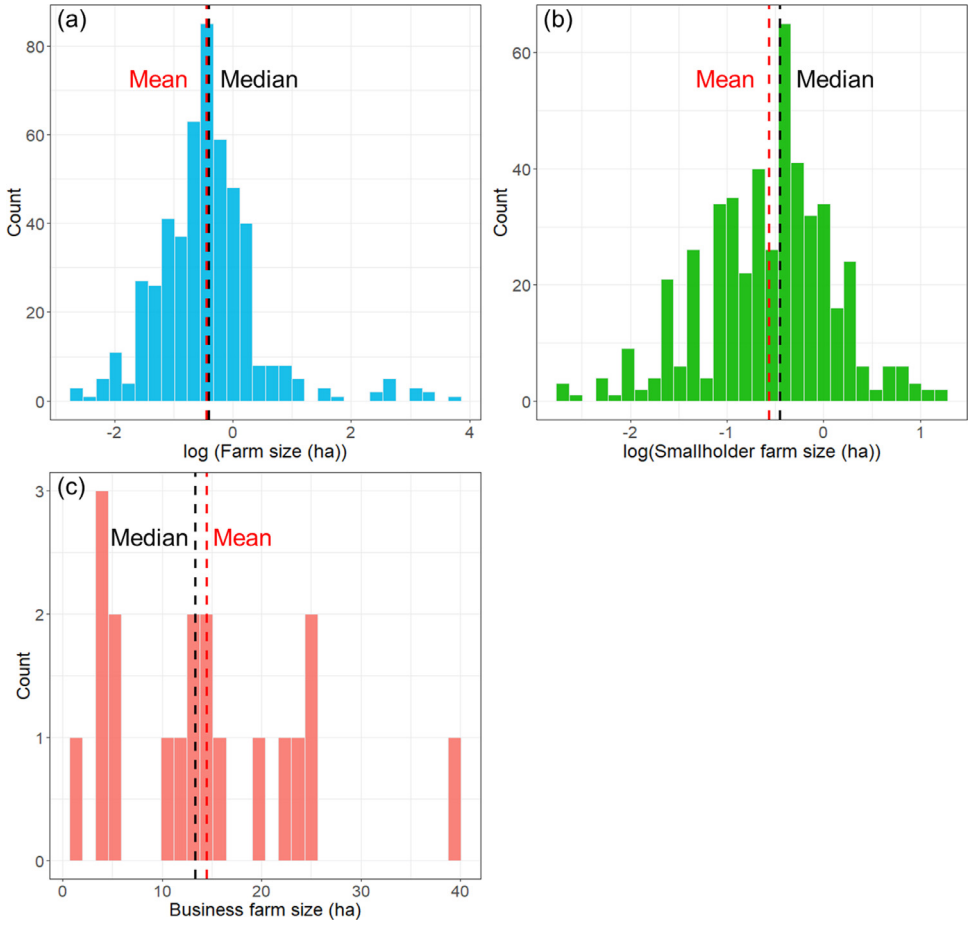

Fig. 3. Farm size distribution as reported by respondents for all farms (a), smallholder farms (b) and business farms (c) The black vertical dashed lines indicate the median farm size and the red vertical dashed lines indicate the mean farm size.

there were more than two cropping systems. Vegetable-based cropping systems were more often found on farms that had at least three cropping systems than on farms that had only one or two cropping systems.

3.3. Crop dataset

The crop dataset was derived from the household survey and applied to the period from June 2019 to June 2020. The dataset provided recalled field-level data with 112 columns and 1352 records. The dataset presents detailed crop input-output inventory and crop management data. The input data includes the quantity and costs of seeds, film mulch, fertilizers, pesticide, irrigation and fuel consumption of each machine, costs of hired labour and purchased machinery service, and the labour consumption of every management practice, including soil preparation, sowing, fertilizer application, pesticide spraying, weeding, harvesting, and other management practices. The output data contains the average crop yield, reported by farmers, as an average

Fig. 4. The number of cultivated crops by respondents.

over all fields[4] of a crop on the farm, and the farm-gate price for every crop. The management includes the timing of sowing and harvesting and application frequency of fertilizers, irrigation, pesticides and weeding, as well as fertilizer type and quantity. With regards to pesticide spraying, we distinguished between herbicides, insecticdes and fungicides; for each category, we collected the information on treatment frequency. For intercropping, the farmers could provide the occupied percentage of every crop component in strips.

Fig. 4 provides information on the cultivated crop species and their records in the dataset. Twenty-eight crop species were cultivated on the interviewed farms, but only nine crops had ten or more records in the crop dataset. Maize, winter wheat, and cotton were the most prevalent crop species grown in the area.

4. Experimental Design, Materials and Methods

Before doing the household survey, we created a village typology based on already available yearbook data on the area of a few main crops in each village of Quzhou county and additional information from a village cadre survey done in 2018. Developing this typology allowed us to target the effort of household survey across the main agricultural village types in Quzhou, to ensure the representativeness of the sample, and avoid that important village types were underrepresented or time would be spent on survey in villages with no or marginal agricultural activity. Creating a typology thus allows us to deal with heterogeneity and to develop a more tailored approach for the household survey [7]. For the purpose of selecting representative villages and

[4] A field is a parcel of open land with one planted crop or planted with an intercrop. In Quzhou, farmers usually cultivate the same crop in multiple fields, and those fields are scattered across the farm.

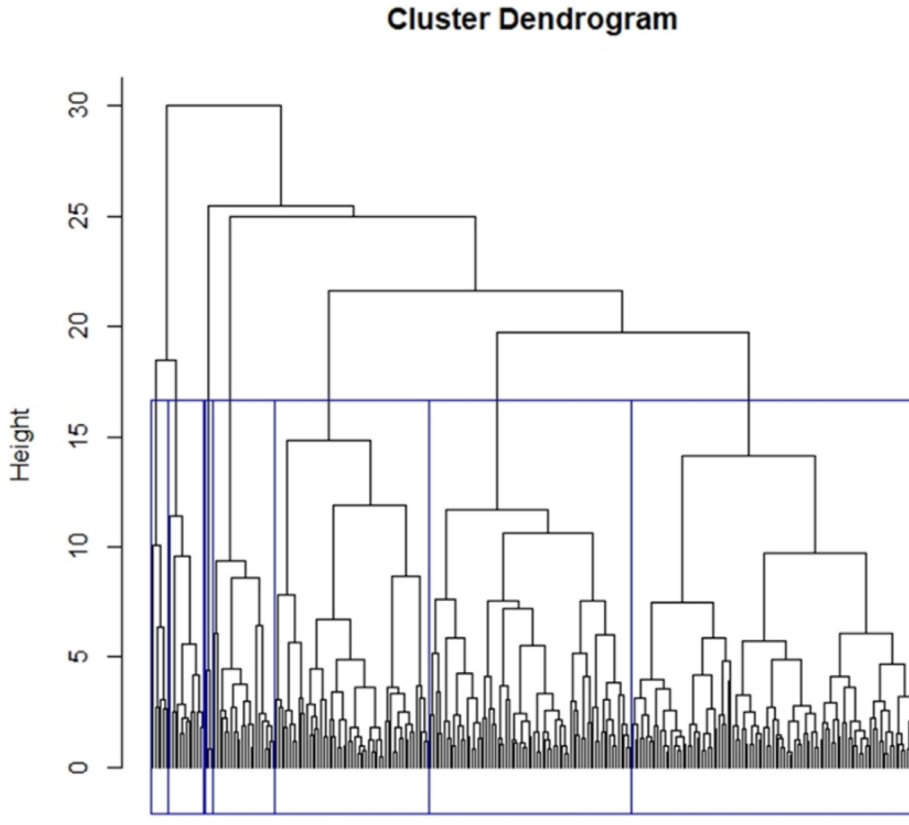

Fig. 5. The dendrogram from hierarchical clustering on 308 villages of Quzhou.

gaining a broader perspective of crop activities in Quzhou, all 342 villages in the county were grouped based on their structure, crop patterns, and resources related to crop production. Eight village types were distinguished (Fig. 5).

4.1. Selection of typology variables

Prior to the farm household survey, a village typology was made based on the village dataset. Considering the classification objective and the available data (amount and accuracy), fifteen variables were calculated from structural variables, crop patterns, and resources (Table 3). The resource variables represented land (arable land), water (irrigatable water), labour (off-farm labour ratio), finance (average income) and market (distance to the city center and distance to the road net). Of 342 villages, 308 were retained for typology analysis because 12 villages were not surveyed in the village cadre survey, 18 villages had low-quality questionnaires; these villages were assigned the type “NA”. Four villages having inconsistent data between yearbook data and surveyed data were assigned the type “Outlier”.

4.2. Multivariate analysis

The typology was generated sequentially using multivariate analysis and cluster analysis: multivariate analysis was used to identify discriminating variables, and cluster analysis was used

to group villages into homogeneous types. All analyses were done with the ade4 package for R version 4.1.0 [8]. Principal component analysis (PCA) was used to reduce the database into non-correlated components based on all fifteen quantitative variables. The number of principal components (PC's) was determined by considering two criteria. The cumulative variance explained by the selected principal components should be >60 %, and all principal components' eigenvalues were greater than one [8]. As a result, we found seven principal components that explained 66 % of the variance among villages. Ward's minimum variance method was used to perform a hierarchical cluster analysis using the PCA output in the form of a reduced dataset based on the retained PC's. A maximum average silhouette was used to determine the number of clusters [9] and seven types of villages were determined. Village type 1 comprised a higher number of villages than the other types. Fig. 5 shows that two subclusters were connected to the biggest cluster in the dendrogram. A second typology analysis was conducted to differentiate the largest clusters. The same classification variables and methods were used to split type 1 into two subtypes. Fig. 6 shows the relationships between village characteristics and village types after PCA and clustering analysis. The distribution of village types across Quzhou is illustrated in Fig. 7.

4.3. Description of village types

Type 1. Small households and small arable land with cereal-based cropping system (119 villages)

In comparison to the other types, this village type had few households (the mean was 192) and less arable land per village (84 ha). The type was subdivided into two subtypes. Farmers in village type 1–1 (sample size of 88 village) allocated significantly less land to vegetables than farmers in village type 1–2 (sample size of 31 villages) (Fig. 8).

Type 2. Large arable land with multiple farm management models (8 villages)

It was the smaller of the two clusters, with a small proportion of off-farm labour (37 %) and a large area of arable land per village (250 ha). This type had the greatest number of business farms, and the proportion of the village area occupied by smallholder farms was the smallest (77 %) among the villages (Fig. 8).

Type 3. Medium number of households and medium arable land with cereal-based cropping system (75 villages)

This was the second largest cluster, with a medium number of households per village (378) and an average area of arable land per village (173 ha). The cultivated crops were primarily maize and wheat, and about 60 % of labourers had other employment (Fig. 8).

Type 4. Close to market and low off-farm labour with substantial vegetable production (60 villages)

Among the villages, type 4 had the smallest distance to the city center and road net. With 96 ha of average arable land per village, this group allocated less than 48 % to wheat cultivation and about 20 % to vegetables grown as arable crop. There was a relatively little off-farm labour in this village type (Fig. 8).

Type 5. High urbanization and nearly no crop production (7 villages)

The type 5 settlements were the closest to the city center, and nearly no land was used for agriculture. Approximately seventy percent of the labour force worked off-farm (Fig. 8). Villages in this cluster were not included in farm survey.

Type 6. Large households and large arable land with far from the city center (17 villages)

In type 6, there were more households per village (786 on average) and arable land area of the village was greater (328 ha). This type was the farthest from the city center (28 km on average), and maize and wheat were the main crops (Fig. 8).

Type 7. Medium arable land with diversified cropping systems (22 villages)

This village type had the highest areas allocated to cotton and oil crops (47 % for cotton and 8 % for oil crops) (Fig. 8).

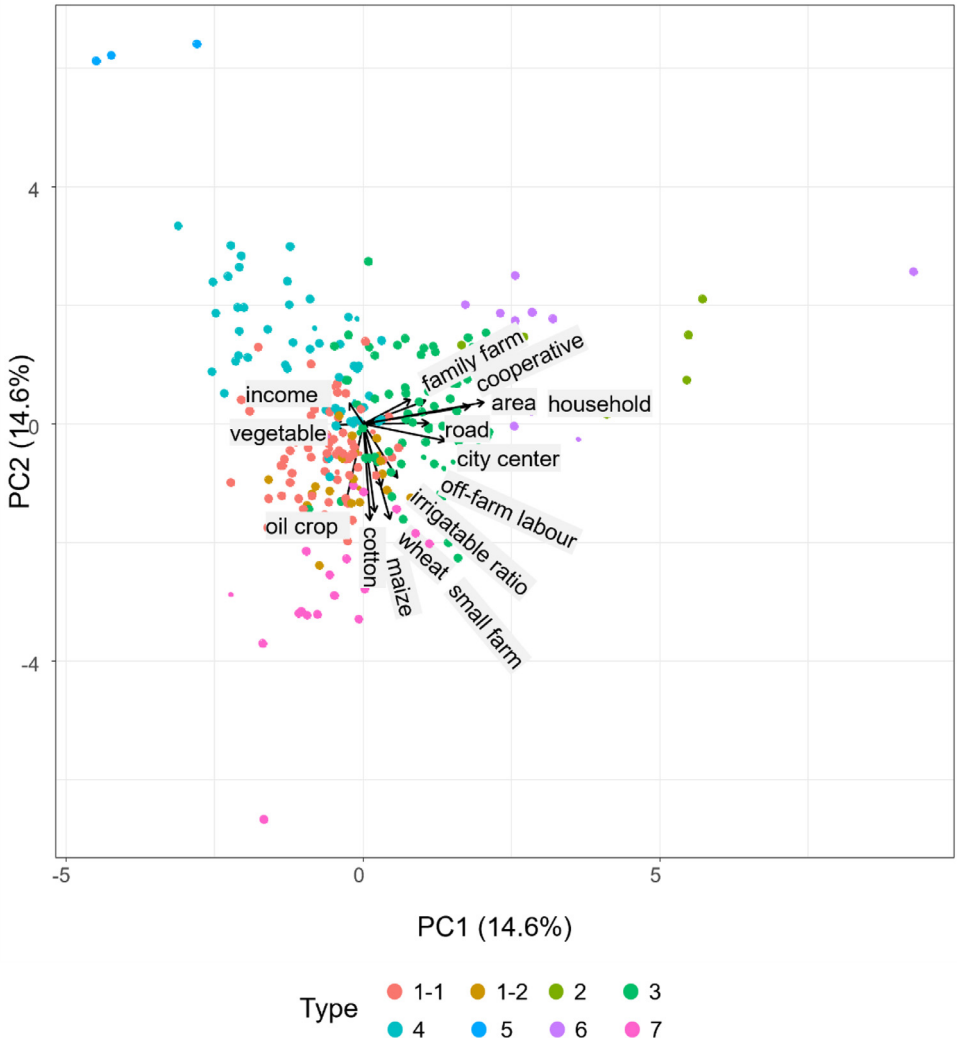

Fig. 6. Principal component analysis of the associations between village characteristics and villages. The filled circles represent villages and different colors represent village types. The arrows represent village characteristic variables, e.g. city center (distance to the city center), road (distance to road net), household (number of households per village), off-farm labour (off-farm labour ratio), income (average income), area (arable land), irrigatable ratio (irrigatable land), wheat (area with wheat), maize (area with maize), cotton (area with cotton), oil crop (area with oil crops), vegetable (area with vegetable), small farm (smallholder farm), cooperative (cooperative farm), and family farm. For a description of variables characterizing the villages, see Table 3. Village types: 1–1 Small villages with small farms, mainly cereal-based (n = 88); 1–2 Small villages with small farms, like type 1 but with more vegetables (n = 31); 2. Large villages with more business farms (n = 8); 3. Medium-sized villages, mainly cereal-based (n = 75); 4. Villages close to the city and with high proportion vegetables (n = 60); 5. Urbanized villages with almost no crop production (excluded from farm survey) (n = 7); 6. Large villages with large farms, far from city, mostly cereal-based (n = 17); 7. Medium-sized villages with diversified cropping systems, including cotton and oil crops) (n = 22). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

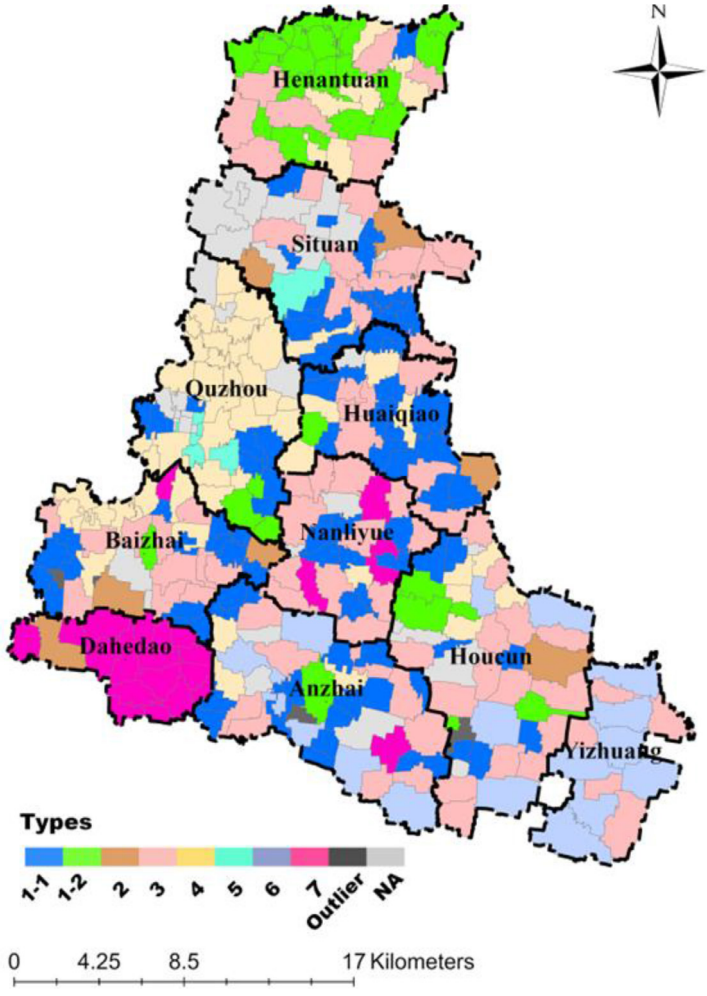

Fig. 7. Village types distribution in Quzhou county. Quzhou county includes ten townships (delineated by bold black borders) and 342 villages (delineated by light grey borders). Township names are listed on the map. Coloring of villages indicates the village type: 1–1 Small villages with small farms, mainly cereal-based ($n = 88$); 1–2 Small villages with small farms, like type 1 but with more vegetables ($n = 31$); 2. Large villages with more business farms ($n = 8$); 3. Medium-sized villages, mainly cereal-based ($n = 75$); 4. Villages close to the city and with high proportion vegetables ($n = 60$); 5. Urbanized villages with almost no crop production (excluded from farm survey) ($n = 7$); 6. Large villages with large farms, far from city, mostly cereal-based ($n = 17$); 7. Medium-sized villages with diversified cropping systems, including cotton and oil crops) ($n = 22$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Use of the typology

The typology was used to select representative villages for a farm household survey that was conducted in Quzhou county from July to August 2020 using stratified random sampling. Five villages were selected at random from each of seven clusters (the highly urbanized cluster 5 was excluded) of villages having crop production. A total of fifteen farms were identified from each selected village. The list of business farms was provided by the local government. All family

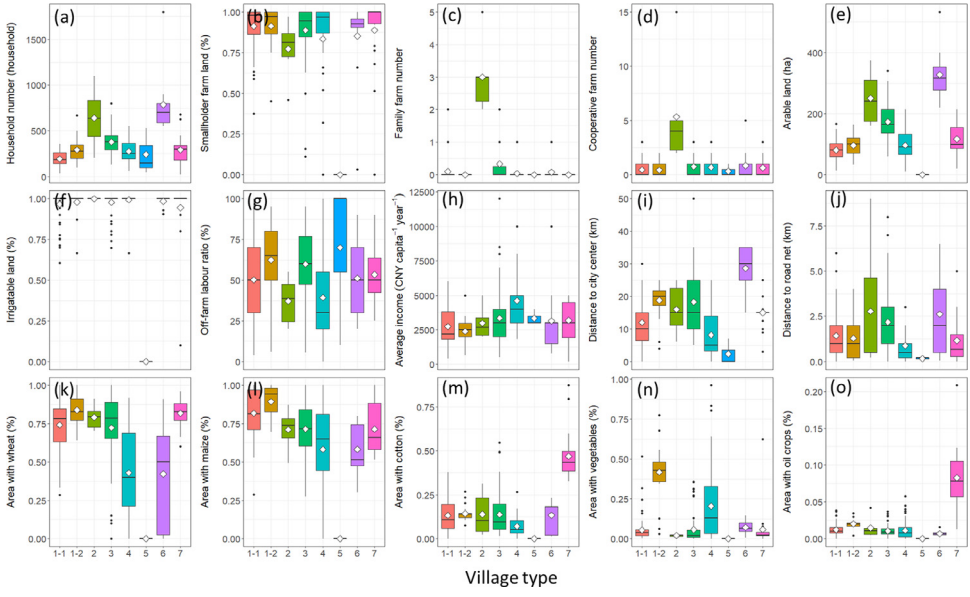

Fig. 8. Village features for the seven village types and two sub-types identified by typology. The white diamond dots represent the mean and the black dots represent outlier values. Oil crops are: peanut, soybean and oil sunflowers. Village types: 1–1 Small villages with small farms, mainly cereal-based (n = 88); 1–2 Small villages with small farms, like type 1 but with more vegetables (n = 31); 2. Large villages with more business farms (n = 8); 3. Medium-sized villages, mainly cereal-based (n = 75); 4. Villages close to the city and with high proportion vegetables (n = 60); 5. Urbanized villages with almost no crop production (excluded from farm survey) (n = 7); 6. Large villages with large farms, far from city, mostly cereal-based (n = 17); 7. Medium-sized villages with diversified cropping systems, including cotton and oil crops) (n = 22).

farms and cooperative farms were always included if they were present in a village. The remaining smallholder farms were randomly selected to obtain a sample of 15 farms per village.

Limitations

The village cadres survey and household survey were conducted in a single county on the North China Plain. Therefore it may not accurately reflect the situation in the whole North China Plain.

In the survey, farmers found it hard to recall some detailed management data, such as cultivar of crops, ingredients and content of pesticides, and the precise date of fertilizer application, pesticide spraying, and irrigation. In the future, some farmers in this survey could be further tracked to collect more detailed data on management practices.

Ethics Statements

After consultation of the ethics experts of Wageningen University and China Agricultural University, we obtained the approval from the local competent authority represented by the Quzhou Experiment Station. All farmers participated in the survey voluntarily. All participants agreed to the survey and the use of the data for non-profit research purposes. All data were collected anonymously.

Data Availability

Survey data on livelihoods and inputs and outputs of crop production activities in Quzhou county on the North China Plain (Original data) (Mendeley Data).

CRediT Author Statement

Zhan Xu: Conceptualization, Investigation, Data curation, Writing – review & editing; **Fan Li:** Data curation, Resources; **Jiali Cheng:** Investigation, Data curation; **Zhengyuan Liang:** Investigation, Data curation; **Jeroen C.J. Groot:** Methodology, Supervision, Writing – review & editing; **Wopke van der Werf:** Supervision, Writing – review & editing; **Chaochun Zhang:** Supervision.

Acknowledgments

This work received financial support from the Cooperative Project of Chinese Academy of Engineering (2020-FJ-XZ-8) and China Scholarship Council (CSC) (grant number is No. 201913043) and Hainan University. In particular, we would like to thank Nico Heerink for the valuable suggestions. We thanked Dongsheng Li, Shuai Han, Yingying Wang and Jiahui Ji for the farm survey data collection.

Declaration of Competing Interest

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

References

[1] National Bureau of Statistics of ChinaChina Statistics Yearbook, Chinese Statistics Press, Beijing, 2017 (in Chinese).
[2] Ministry of Agriculture and Rural Affairs of the People's Republic of ChinaList of National Agricultural Green Development Focus Areas for the Second Round (in Chinese), 2019.
[3] J. Shen, Q. Zhu, X. Jiao, H. Ying, H. Wang, X. Wen, W. Xu, T. Li, W. Cong, X. Liu, Y. Hou, Z. Cui, O. Oenema, W.J. Davies, F. Zhang, Agriculture green development: a model for China and the world, Front. Agric. Sci. Eng. 7 (2020) 5–13.
[4] J. Huang, J. Ding, Institutional innovation and policy support to facilitate small-scale farming transformation in China, Agric. Econ. 47 (2016) 227–237.
[5] M. Shen, J. Shen, Evaluating the cooperative and family farm programs in China: a rural governance perspective, Land Use Policy 79 (2018) 240–250.
[6] Y. Yu, Y. Hu, B. Gu, S. Reis, L. Yang, Reforming smallholder farms to mitigate agricultural pollution, Environ. Sci. Pollut. Res. Int. 29 (2021) 13869–13880.
[7] S. Alvarez, C.J. Timler, M. Michalscheck, W. Paas, K. Descheemaeker, P. Tittonell, J.A. Andersson, J.C. Groot, Capturing farm diversity with hypothesis-based typologies: an innovative methodological framework for farming system typology development, PLoS ONE 13 (2018) e0194757.
[8] R. Core Team, R: A Language and Environment For Statistical Computing, R Foundation for Statistical Computing, 2021.
[9] K.S. Kuivanen, S. Alvarez, M. Michalscheck, S. Adjei-Nsiah, S. Descheemaeker, S. Mellon-Bedi, J.C.J. Groot, Characterizing the diversity of smallholder farming systems and their constraints and opportunities for innovation: a case study from the Northern Region, Ghana, NJAS 78 (1) (2016) 153–166.