Food Research International 164 (2023) 112364

Contents lists available at ScienceDirect

Food Research International

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

Geographic distributions and the regionalization of soybean seed compositions across China

Wenwen Song [a,1], Shi Sun [a,1], Tingting Wu [a], Ruping Yang [b], Shiyan Tian [a], Cailong Xu [a], Bingjun Jiang [a], Shan Yuan [a], Wensheng Hou [a], Cunxiang Wu [a,*], Tianfu Han [a,*]

[a] MARA Key Laboratory of Soybean Biology (Beijing), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[b] Institute of Dryland Agriculture, Gansu Academy of Agricultural Sciences, Lanzhou 730070, China

ARTICLE INFO

Keywords:
Soybean
Nutritional quality
Geographic distribution
Regionalization

ABSTRACT

As one of major food crops, soybean is grown over a broad ecological region in China with considerable variations in environmental conditions, and the seed compositions of soybeans are diverse among different regions. To clarify the spatial patterns of soybean seed compositions, crude oil, protein, and 11 categories of functional components were quantified in 1792 soybean samples collected from a vast range of soybean planting regions across China spanning from 2010 to 2017. The Kriging interpolation maps presented a clear north-to-south (high latitude to low latitude) increasing trend in contents of crude protein and dietary fiber and decreasing trend in contents of crude oil, phospholipids, saponins, and carotenoids. Soybeans with high-level of total oligosaccharide were concentrated in the central region. Based on the geographical distribution of soybean nutritional components, weather conditions, and cultivation systems, the soybean production areas in China were divided into three regions and 10 subregions. This study highlights the geographic distribution of soybean nutritional compositions and provides scientific evidence for guiding the construction of high-quality edible soybean production bases in China.

1. Introduction

Soybean is a major source of high-quality vegetable protein and oil. In addition, soybean contains significant amounts of functional components, including isoflavones, oligosaccharides, saponins, phospholipids, vitamins, etc. (Li, Zhao, Yun, Liu, & Qi, 2016; Medic, Atkinson, & Hurburgh, 2014; Sharma, Kaur, Goyal, & Gill, 2014). These components are reported to have health benefits because they can reduce the risk of some chronic human diseases, such as cancers, osteoporosis, and cardiovascular disease, as well as strengthen the brain and improve gastrointestinal function (Cao, Green-Johnson, Buckley, & Lin, 2019; Hu, Wong, Wu, & Lai, 2020; Namazi, Saneei, Larijani, & Esmaillzadeh, 2018; Wang et al., 2021). In east Asia, domestically-produced soybeans are primarily used to produce various foods, such as tofu, soy milk, soybean sprouts, soy paste, natto, etc. (Cao et al., 2019; Chen et al., 2012). It is confirmed that the multi-nutrient and health-beneficial characteristics have increased the popularity of soyfoods in diets worldwide (Chen et al., 2012). The nutritional quality of soybeans is becoming critical for global soybean products, marketing, and trade (Martey & Goldsmith, 2020). Assessment of soybean nutritional quality will help to meet the demands of the growing soyfood market and maximize their medical and healthcare value worldwide.

China is currently the fourth largest soybean producer after Brazil, United States, and Argentina. In China, soybean planting region covers a wide ecological area with major variations in environmental conditions such as temperature, rainfall, and radiation (Song et al., 2016). It has been widely documented that soybean seed compositions varies significantly for different geographic regions in China and other countries (Song et al., 2016, 2018), and these variations were proven to be attributed to both genetic and environmental factors (Assefa et al., 2019; Bosaz et al., 2019; Carrera, Martínez, Dardanelli, & Balzarini, 2011; Goyal, Sharma, & Mohammed, 2014). Although numerous studies of spatial variation in soybean quality across China exist, a majority of the studies were performed at regional or provincial levels (Ning, Zhang,

Abbreviations: NSR, Northern Spring Planting Region; NESR, Northeast Spring Planting Subregion; NWSR, Northwest Spring Planting Subregion; HHHR, Huang-Huai-Hai River Valley Summer Planting Region; SMCR, South Multiple Cropping Region; BBI, Bowman-Brick trypsin inhibitor.

* Corresponding authors.
E-mail addresses: wucunxiang@caas.cn (C. Wu), hantianfu@caas.cn (T. Han).

[1] Wenwen Song and Shi Sun contributed equally to this work.

https://doi.org/10.1016/j.foodres.2022.112364
Received 17 August 2022; Received in revised form 10 December 2022; Accepted 24 December 2022
Available online 28 December 2022

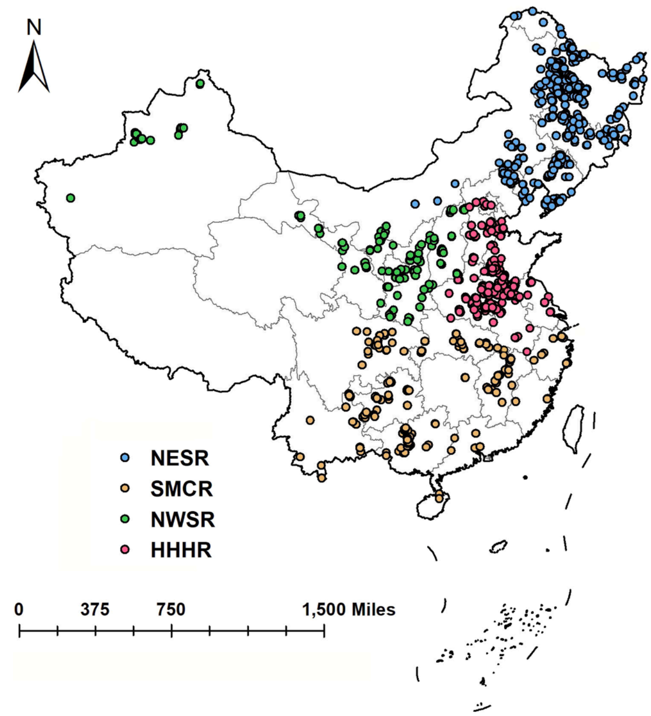

Fig. 1. Distribution of soybean sampling sites in different regions of China.

Hu, Li, & Li, 2007). In the US, geostatistics method was used to analyze the spatial pattern of soybean protein and oil variation at the national scale, in which the Kriging interpolation maps were made to show the variation at spatial levels smaller than regions and states (Rotundo, Miller-Garvin, & Naeve, 2016). In recent years, the Kriging interpolation method has also been applied to study the relationship between some quality traits of soybean germplasm and their origins in China (Abdelghany et al., 2020; Azam et al., 2020). However, previous studies focused on one or a few components, and the spatial distribution of soybean seed multi-nutrient compositions has not been well documented.

The scientific subdivision of soybean production regions according to the spatial variation of soybean quality is an important practice to develop high-quality soybean production for different uses. Based on the cultivation regionalization and ecological zoning of soybean varieties in China, Chang, Zhou, and Qiu (2003) proposed a quality zoning scheme of three zones and 10 subzones by analyzing the characteristics of protein and fat content of soybean in each ecological zone in China and its relationship with environmental conditions and cultivation measures (Chang et al., 2003). This work provided an important basis for soybean quality improvement and industrialization layout in China. Due to the constant replacement of soybean varieties and the adjustment of soybean production layout in the last decade, it has become imperative to readjust the soybean regionalization.

In this study, multi-nutrient compositions of 1792 soybean samples spanning 2010 to 2017 from three soybean planting regions across China were analyzed. The aim of our study was to evaluate spatial variation patterns of seed multi-nutrient compositions in soybean seeds across China using a geostatistical approach. On this basis, a new soybean quality regionalization scheme will be proposed for guiding the production layout of high-quality soybeans in China. It would be helpful for the soybean producers and processors, whose goals are to produce high quality and consistent products.

2. Materials and methods

2.1. Soybean materials

A total of 1792 soybean samples were collected for quality trait determination from 2010 to 2017. For collecting the samples, the soybean production area in China was divided into two-degree longitude-latitude cell grids, and the samples were randomly collected from the large-scale soybean production fields in the corresponding cell grid and an adjacent cell grid. In each field (site), five spots (nearly 1 m^2), seeds of approximately 20 mature soybean plants, were chosen for seed collection (Song et al., 2018). These samples represented three major soybean planting regions of China: Northern Spring Planting Region (NSR), Huang-Huai-Hai River Valley Summer Planting Region (HHHR), and South Multiple Cropping Region (SMCR). The sample number of each region was proportional to the regional production to obtain a nationwide representative quality assessment. The sampling sites covered 28 of the 34 provinces (municipalities or autonomous regions) of China, with a geographical range from 19.71°N to 53.53°N and from 73.74°E to 134.76°E. Distribution of soybean sampling sites within the 28 provinces where samples were collected from 2010 to 2017 is shown in Fig. 1. The number of samples, the maturity groups, and the geographical condition of each site are presented in Table 1. Among the 1792 samples, 269, 187, 146, 206, 240, 236, 281, and 227 samples were collected in 2010, 2011, 2012, 2013, 2014, 2015, 2016, and 2017, respectively. Part of the data was used to analyze the impact of environment on quality (Song et al., 2016, 2018; Zhu et al., 2018). The soybean seeds of each sample were dried, ground, and passed through a 60-mesh sieve in a FOSS 1093 Cyclotec 126 sample mill (FOSS Tecator,

Table 1
Geographic characterization and sample information of the three regions.

Year	NSR		HHHR	SMCR	Total	Compositions determined
	NESR MG 000-VI 38.72–53.53°N 116.88–134.76°E	NWSR MG 00-V 31.75–49.04°N 73.74–114.40°E	MG II-V 32.61–42.58°N 110.60–122.65°E	MG I-VIII 19.71–33.98°N 97.69–122.16°E		
2010	121	50	70	28	269	Crude protein, crude oil, fatty acids, phospholipid, isoflavones, oligosaccharides, dietary fiber, BBI, lectin
2011	86	29	55	17	187	Crude protein, crude oil, water-soluble protein, isoflavones, oligosaccharides
2012	58	20	44	24	146	Crude protein, crude oil, water-soluble protein, isoflavones, oligosaccharides
2013	81	35	49	41	206	Crude protein, crude oil, water-soluble protein, isoflavones, oligosaccharides
2014	101	35	55	49	240	Lunasin
2015	126	21	52	37	236	Lunasin
2016	128	23	64	66	281	Crude protein, crude oil, tocopherols, carotenes, sapnins
2017	101	23	49	54	227	Crude protein, crude oil, tocopherols, carotenes and sapnins
Total	802	236	438	316	1792	

NSR, Northern Spring Planting Region; NESR, Northeast Spring Planting Subregion; NWSR, Northwest Spring Planting Subregion; HHHR, Huang-Huai-Hai River Valley Summer Planting Region; SMCR, South Multiple Cropping Region; MG, Maturity group.

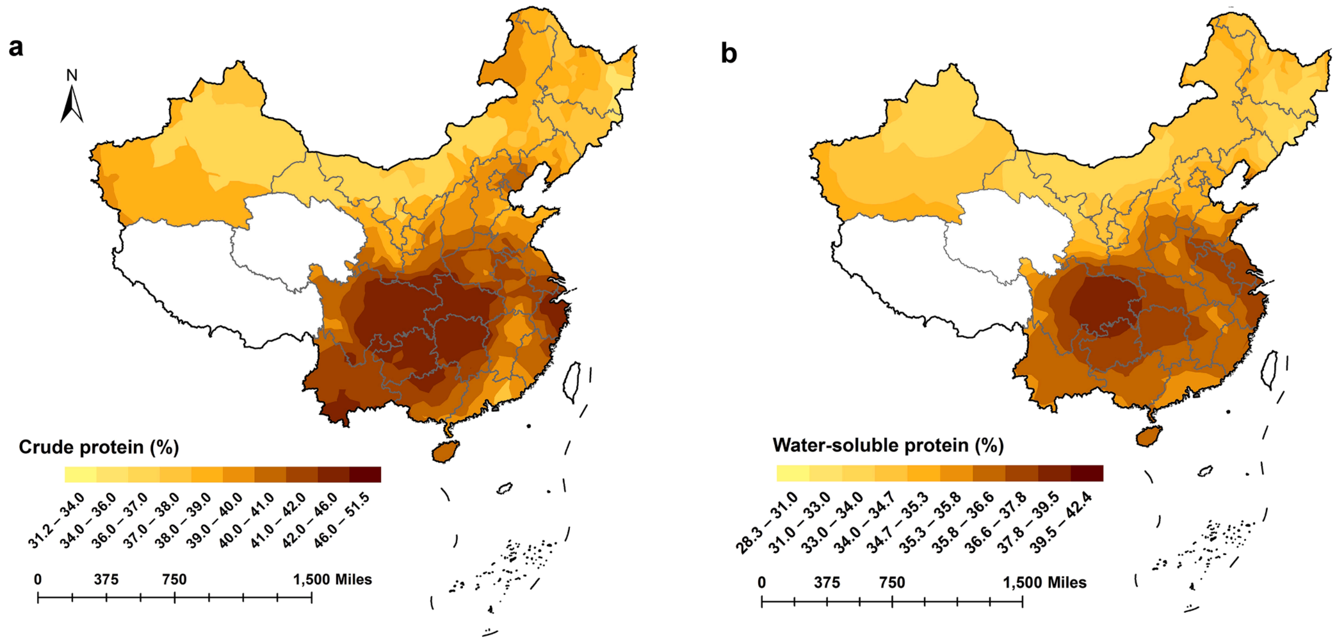

Fig. 2. Geographical distribution of soybean protein contents in different regions of China.

Hoganas, Sweden) for nutritional composition determination.

2.2. Determination of nutritional compositions

The thirty-three components including the majority of nutrients, bioactive constituents, and the important anti-nutritional factors were analyzed in this study. The compositions determined for the samples of different years were shown in Table 1. The methodology for analyzing the thirty-three components of soybean seed constituents consisting of crude protein, water-soluble protein, crude oil, five fatty acids, five isoflavone components, three oligosaccharide components, five saponin components, three phospholipid components, three tocopherol components, two carotenoid components, dietary fiber, Bowman-Brick trypsin inhibitor (BBI), lectin, and lunasin were described below.

The semimicro-Kjeldahl method was used to determine the crude protein and water-soluble protein contents of the 1792 soybean samples. The total oil content was analyzed using the Soxhlet extraction method. Five fatty acids of palmitic, stearic, oleic, linoleic, and linolenic were determined by gas chromatography (GC) equipped with a flame ionization detector described by Qin et al. (2014). The isoflavone components of daidzin, glycitin, genistin, malonyldaidzin, and malonylgenistin were determined in 763 samples according to the study of Sakthivelu et al. (2008). The saponin components of saponin I, saponin II, saponin αg, saponin βg, and saponin βα were determined in 526 samples following Yang, Dong, and Ren (2011). Three oligosaccharide components of sucrose, raffinose, and stachyose were analyzed in 763 samples following Wu et al. (2016). The phospholipid components of phosphatidylinositol, phosphatidylcholine, and phosphatidylethanolamine were extracted and quantified in 269 samples by using the method of Lee, Jin, Park, Chung, Jin, and Lee (2010) with some modifications. The tocopherol components of α-Toc, β + γ-Toc, and δ-Toc were analyzed in 526 samples using the method described by Seguin, Tremblay, Pageau, and Liu (2010) with minor modifications. The carotenoid components of lutein and β-carotene were extracted and quantified in 526 samples according to Wu et al. (2016). The method described by Wu et al. was used for determination of dietary fiber in 269 samples (Hua & Gu, 2004). BBI and lectin were extracted and determined in 269 samples according to the method of Anta, Marina, and García (2010). Lunasin content of the samples was quantified in 476 samples according to Ren, Zhu, Shi, and Li (2017).

2.3. Statistical and spatial analysis

Descriptive statistics were used to summarize the data for each component. A majority of the components are normally distributed. Some of the traits exhibited an approximately normal distribution with a single central peak of frequency. All the data were analyzed using SPSS 19.0 statistical software (IBM Corp., Armonk, NY, USA). Geographical distribution maps of soybean seed multi-nutrient compositions across China soybean production areas were constructed using ArcGIS 10.0 (Allen, 2010). To generate the maps, multi-year data of each component were compiled and exported from excel into a text format capable of being utilized by ArcGIS. The shape-file was then added to a project in ArcMap that contained the shapefile of China as a background reference. The geostatistical analysis method of ordinary Kriging interpolation was used to estimate values in unknown areas by considering both the distance and the degree of variation between known data points. The kriging interpolation formula can be expressed as follow:

$$z(x) = \sum_{i=1}^{n} w_i z_i$$

The z (x) represent the unknown sample point, z_i means the value of the ith known sample point near the unknown sample point, and w_i is the weight of the ith known sample point applied to the unknown sample point. The original map base was downloaded from the website of Ministry of Natrual Resources (https://bzdt.ch.mnr.gov.cn), which is for public use, and the approval number is GS(2019)1822.

2.4. Regionalization

The three-region division scheme was proposed based on the well-known three soybean cultivating regions of NSR, HHHR and SMCR, and combined the soybean quality characteristic of each region. The sub-region division was based on the existing cultivation regionalization (Gai & Wang, 2001; Pu & Pan, 1982) with some modifications according to the soybean quality characteristics.

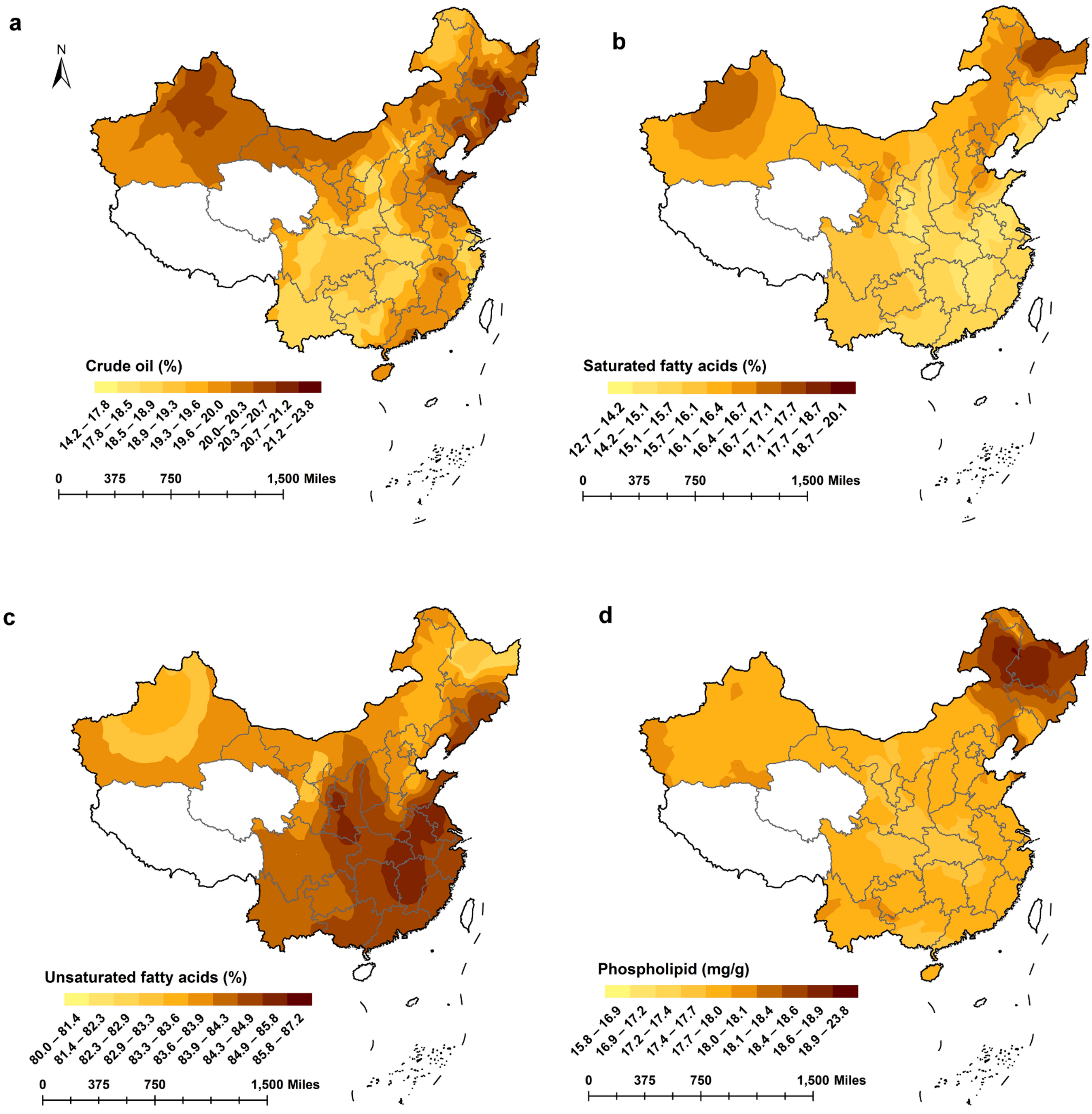

Fig. 3. Geographical distribution of soybean oil and fatty acid compositions in different regions of China.

3. Results

3.1. Geographical distribution of crude protein and soluble-water protein contents in soybean

The geographical distribution maps showed a clear north-to-south increasing trend with the decreasing latitude in both crude protein and soluble-water protein contents across China (Fig. 2a and b). The soybeans with high-level crude protein content were mainly concentrated in the Yangtze River basin and the southwestern mountainous area, including central and eastern Sichuan, Chongqing, Hubei, northern and eastern Guizhou, northwestern Guangxi, central and eastern Zhejiang, and northwestern Yunnan.

3.2. Geographical distribution of crude oil, fat fatty acids, and phospholipids contents in soybean

Kriging interpolation maps for crude fat concentration (Fig. 3a) showed that the soybeans with rich crude fat content were primarily distributed both in the northeastern and northwestern regions of China, including the Sanjiang Plain and the southern region of Heilongjiang, central and eastern Jilin, the northeast and the Bohai Bay area of Liaoning, and northern Xinjiang. Among the fatty acids, the saturated fatty acids showed a decreasing trend from north to south of China, whereas the unsaturated fatty acids showed increasing trends from north to south of China. Saturated fatty acid concentrations were observed in soybeans grown in northernmost parts of China, including the northeastern part of Heilongjiang and the northwestern Xinjiang

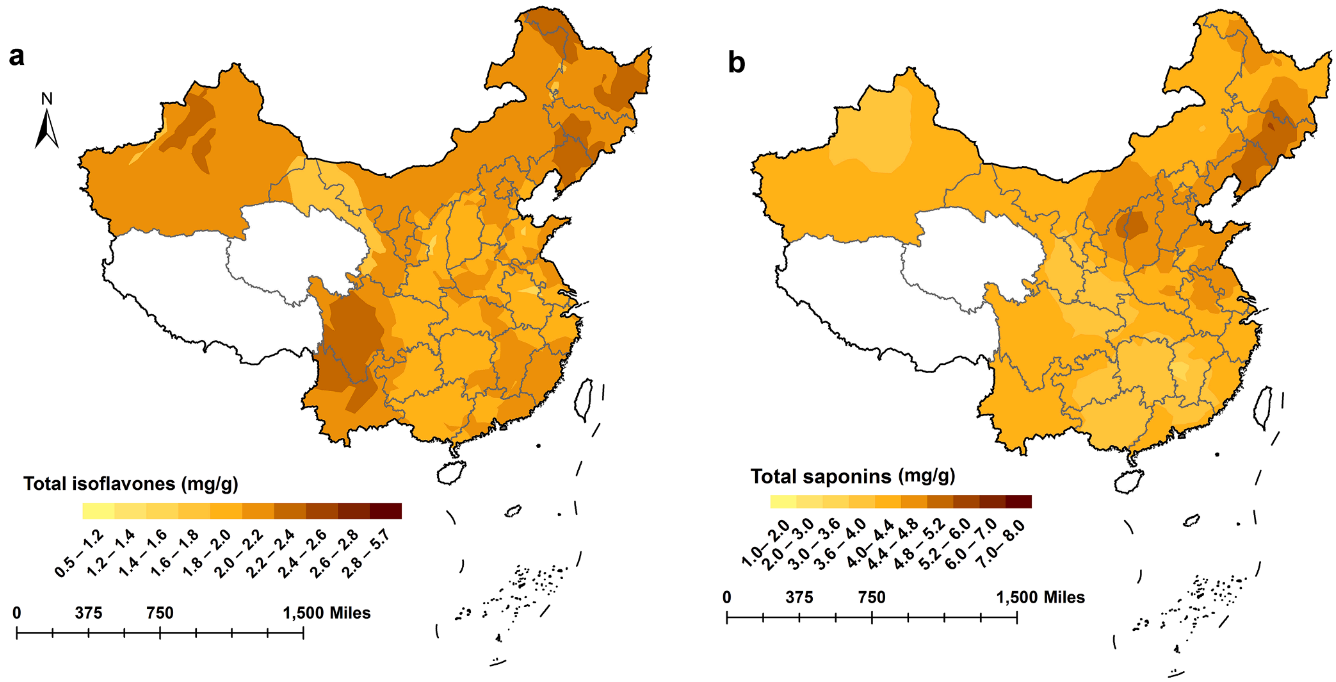

Fig. 4. Geographical distribution of soybean isoflavone and saponin contents in different regions of China.

(Fig. 3b). The soybeans contained high-level unsaturated fatty acids in the central and southern parts of China, including Jilin, eastern Liaoning, western Shanxi, southeastern Gansu, central and southern Shaanxi, and south China (Fig. 3c).

The phospholipid contents in soybeans showed a trend of declining from north to south across China (Fig. 3d). The geographical distribution map showed that soybeans with high-level phospholipid contents were mainly distributed in northeastern regions of China, especially in eastern Hulunbeier of Inner Mongolia and western Heilongjiang. Soybeans with low-level of phospholipid contents were distributed in southwestern Shaanxi, Ningxia, northeastern Sichuan, Chongqing, Hubei, northern Hunan, and northwestern Jiangxi.

3.3. Geographical distribution of soybean isoflavones and saponins contents

The isoflavone contents in soybeans showed a declining trend from north to south and from west to east across China (Fig. 4a). Soybeans with high-level of isoflavone content were distributed mainly in northeastern, northwestern, and southwestern mountainous areas in China, including the Great Xing'anling Mountain area and southeastern regions of Heilongjiang, southern Jilin, central and eastern Liaoning, Yili prefecture of Xinjiang, southwestern Sichuan, and northern Yunnan. Soybeans with low-level of isoflavone were grown mainly in the Hexi Corridor area of Gansu province.

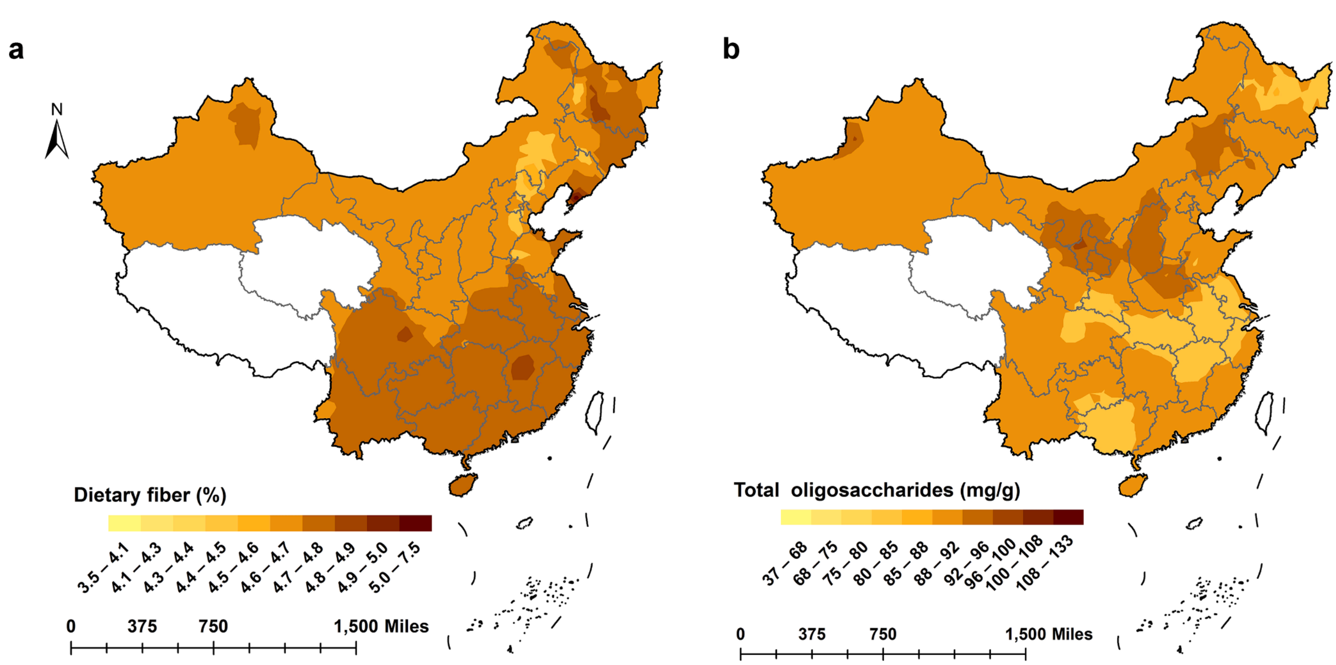

Fig. 5. Geographical distribution of soybean dietary fiber and oligosaccharide contents in different regions of China.

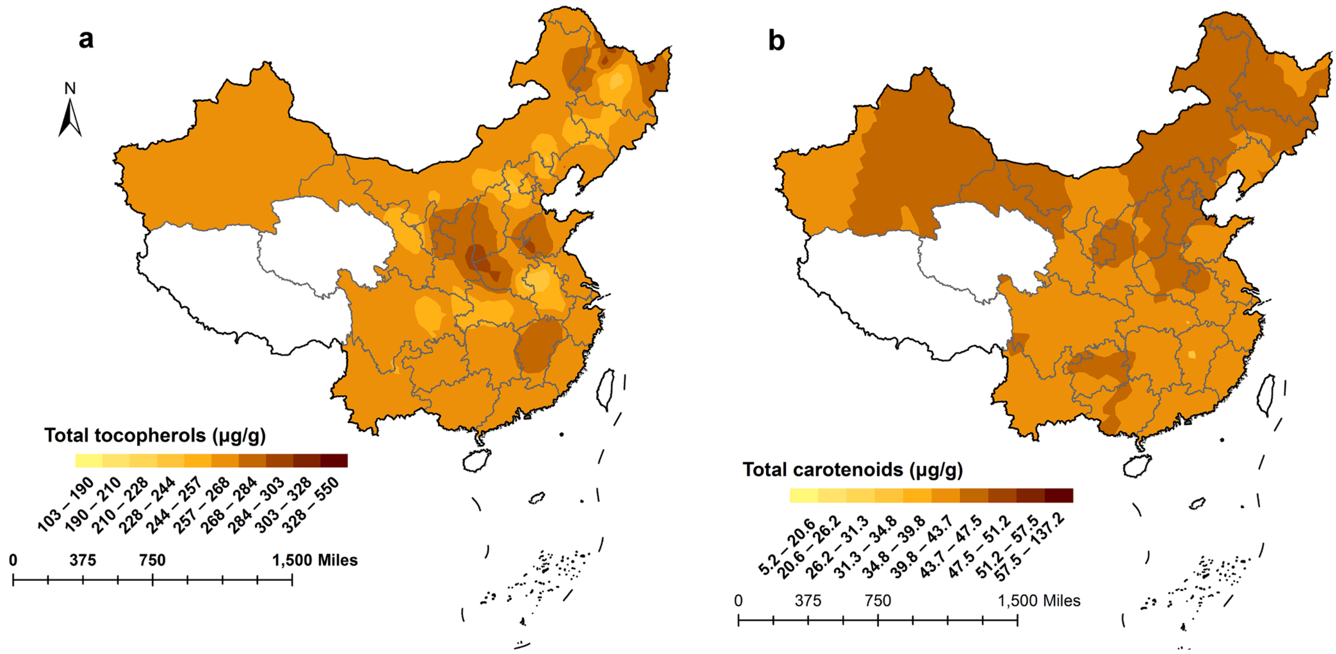

Fig. 6. Geographical distribution of soybean tocopherol and carotenoid contents in different regions of China.

The saponin contents in soybean planted in the south part of northeastern China and Loess Plateau region was higher than in other areas (Fig. 4b). The soybeans with rich saponin constituents were mainly grown in southern Heilongjiang, central and eastern Jilin, central and southern Liaoning, western Shanxi, and northeastern Shaanxi. The saponin contents in soybeans grown south of the Yangtze River and northwestern Xinjiang were generally lower.

3.4. Geographical distribution of soybean dietary fiber and oligosaccharides contents

Kriging interpolation maps showed that northern regions except for the eastern portion of the northeast had lower dietary fiber than southern regions (Fig. 5a). Soybeans with high dietary fiber contents were mainly distributed in the south part of the Yangtze River basin (central Jiangxi and Nanchong of Sichuan) and the eastern part of the northeast (central and southern Heilongjiang, northeastern Jilin, Liaodong Peninsula). The dietary fiber contents in soybeans grown in the southeastern part of Inner Mongolia and the northeastern part of Hebei were generally lower.

Soybeans with high-level of oligosaccharide contents were mainly distributed in Shanxi and Ningxia, northern Henan, central and eastern Gansu, central and western Shaanxi, Hinggan League, Chifeng, Tongliao of Inner Mongolia, and Yining of Xinjiang (Fig. 5b). Soybeans with low-level of oligosaccharide were planted mainly in Sanjiang Plain of Heilongjiang, the Yangtze River Basin, and Guangxi Autonomous Region.

3.5. Geographical distribution of soybean tocopherol and carotenoids contents

Kriging maps did not show clear regional trends in tocopherol contents across China (Fig. 6a), and the distribution of soybeans with rich tocopherol contents was dispersed in some areas, such as northern and eastern Heilongjiang, southeastern Hulunbeier of Inner Mongolia, western Shandong, southwestern Shanxi, northern and central Shaanxi, eastern Ningxia, eastern Gansu, and central and northern Jiangxi.

Generally, the northern regions had higher carotenoid contents than southern regions (Fig. 6b). Soybeans with high carotenoid contents were primarily distributed in central Inner Mongolia, Hebei, central and eastern Henan, northern Shanxi, central Shaanxi, Hexi Corridor and Qingyang region of Gansu, eastern and northern Xinjiang, and parts of Guangxi. Soybeans with low-level carotenoid contents were grown in south part of the Yangtze River Basin.

3.6. Geographical distribution of soybean BBI, lectin, and lunasin contents

Soybeans with high content of BBI type trypsin inhibitor were mainly distributed in Henan, central and western Anhui, and northeastern Hubei. The high-lectin soybeans were mainly concentrated in the northern part of Heilongjiang province (Fig. 7b). Soybeans with high lunasin content were mainly distributed in the central and eastern part of Northeast China (Fig. 7c), mainly including the central-northern part of Heilongjiang province, most parts of Jilin province, and the southeastern part of Inner Mongolia Autonomous Region.

3.7. Regionalization of soybean nutrition quality in China

The regionalization scheme of soybean nutrition quality in China was completed on the basis of current cultivation regionalization of soybeans in China, which has been mapped by Pu and Pan (1982) and modified by Gai and Wang (2001). We completed the quality regionalization by combining existing 3 cultivating regions and 10 subregions and the geographical distribution characteristics of soybean nutritional components in each regions or sub-regions. The soybean production areas in China were divided into 3 regions and 10 subregions (Fig. 8): 1) the Northern Spring-sowing High Oil and Rich Functional Component Soybean Region (Region I); 2) Huang-Huai-Hai Summer-sowing High Protein Soybean Region (Region II); 3) Southern Multiple-Cropping High Protein and Rich Dietary Fiber Soybean Region (Region III). In Region I, four subregions were further divided into 1) The Northern Northeast Spring-sowing High Functional Component Soybean Subregion (Subregion I_1); 2) Middle and South Northeast Spring-sowing High Oil and Functional Component Soybean Subregion (Subregion I_2); 3) North China Plateau Spring-sowing High Oligosaccharide and Vitamin Soybean Subregion (Subregion I_3); 4) Northwest Spring-sowing High Oil Soybean Subregion (Subregion I_4). In Region II, there were two subregions: 1) Northern Huang-Huai-Hai Summer-sowing High Vitamin

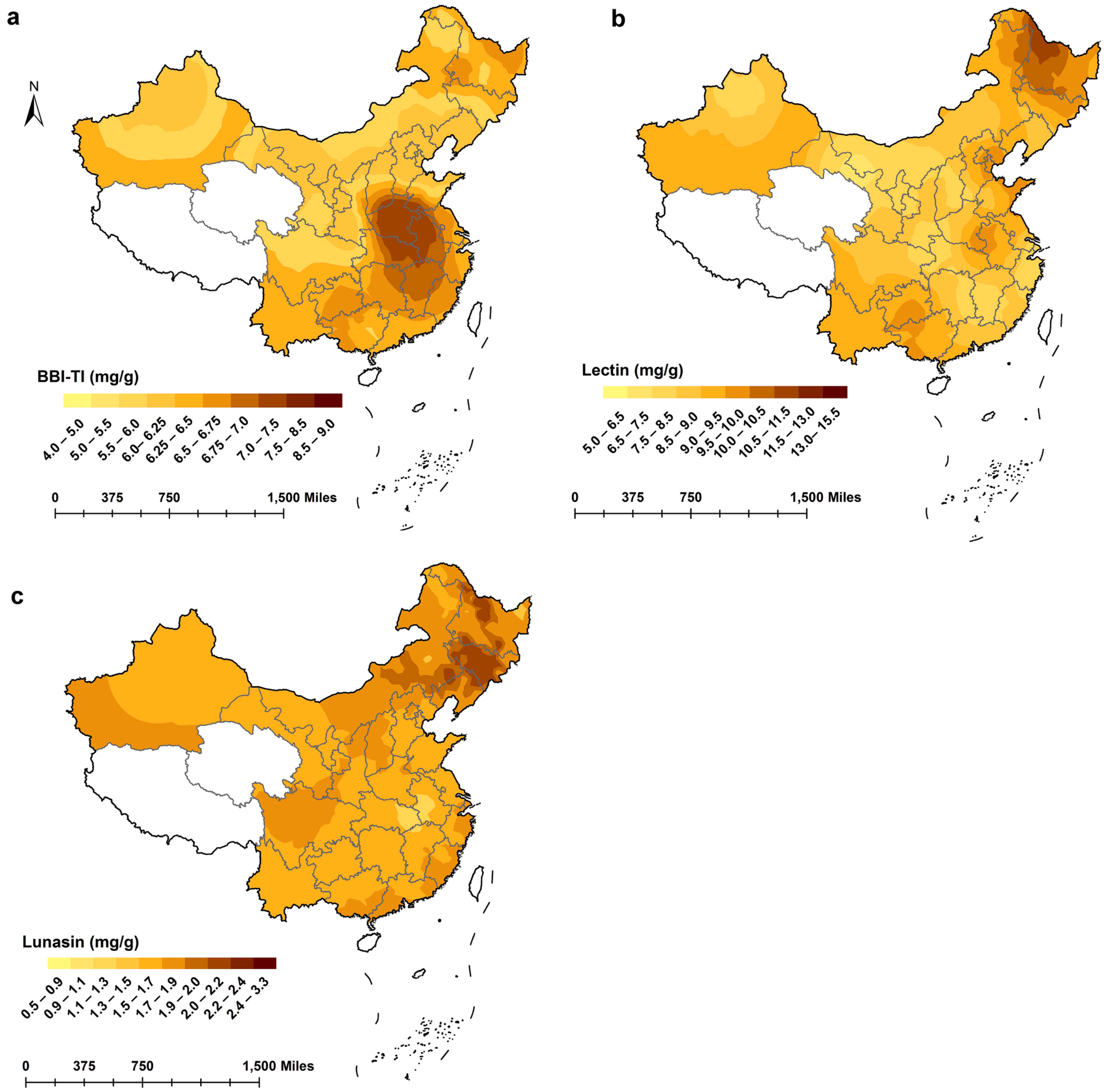

Fig. 7. Geographical distribution of soybean BBI-TI, lectin and Lunasin contents in different regions of China.

Soybean Subregion (Subregion II_5); 2) South Huang-Huai-Hai Summer-sowing High Protein Soybean Subregion (Subregion II_6). Region III consisted of four subregions: 1) Middle and Lower Reaches of Yangtze River Spring- and Summer-sowing High Protein and Dietary Fiber Soybean Subregion (Subregion III_7); 2) Southwest Spring- and Summer-sowing High Protein, Isoflavone, and Dietary Fiber Soybean Subregion (Subregion III_8); 3) Central South Spring and Autumn-sowing High Protein and Dietary Fiber Soybean Subregion (Subregion III_9); 4) South China Multiple Cropping High Dietary Fiber Soybean Subregion (Subregion III_{10}).

4. Discussion

China is the global center of origin for the cultivated soybean with the abundance of soybean germplasms (Han et al., 2016). Soybeans are grown over broad ecological regions with considerable variations in environmental conditions in China (Qin et al., 2014; Song et al., 2016, 2018). China is consequently the country with the most complicated soybean cultivation system. Soybean is planted in spring in NRS, in summer and rotated with wheat in HHHR. In the SMCR, soybeans can be sown in multiple seasons, rotated or intercropped with multiple crops. Scientific and reasonable regionalization has important guiding significance for soybean production in China. The currently used regionalization was made according to the cultivation system (Gai & Wang, 2001; Pu & Pan, 1982). The first quality regionalization scheme by Chang et al. (2003) was not been widely accepted for it was made based on the survey or experiential data. This is the first regionalization based on detection data, and it is of great significance for high-quality soybean production.

This regional pattern of proteins and oil in this study was consistent

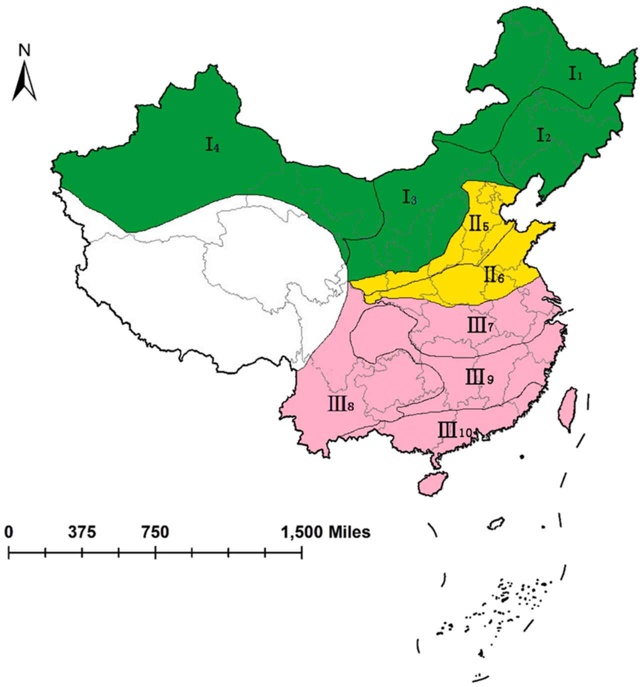

Fig. 8. Regionalization of soybean nutritional quality in China. The soybean production area in China was divided into 3 major regions as below: 1) the Northern Spring-sowing High Oil and Rich Functional Component Soybean Region (Region I, highlight in green); 2) Huang-Huai-Hai Summer-sowing High Protein Soybean Region (Region II, highlight in yellow); 3) Southern Multiple-Cropping High Protein and Rich Dietary Fiber Soybean Region (Region III, highlight in pink). In Region I, four subregions were further divided into 1) The Northern Northeast Spring-sowing High Functional Component Soybean Subregion (Subregion I_1); 2) Middle and South Northeast Spring-sowing High Oil and Functional Component Soybean Subregion (Subregion I_2); 3)North China Plateau Spring-sowing High Oligosaccharide and Vitamin Soybean Subregion (Subregion I_3); 4)Northwest Spring-sowing High Oil Soybean Subregion (Subregion I_4). In Region II, there were two subregions: 1) Northern Huang-Huai-Hai Summer-sowing High Vitamin Soybean Subregion (Subregion II_5); 2) South Huang-Huai-Hai Summer-sowing High Protein Soybean Subregion (Subregion II_6). Region III consisted of four subregions: 1) Middle and Lower Reaches of Yangtze River Spring- and Summer-sowing High Protein and Dietary Fiber Soybean Subregion (Subregion III_7); 2) Southwest Spring- and Summer-sowing High Protein, Isoflavone and Dietary Fiber Soybean Subregion (Subregion III_8); 3) Central South Spring and Autumn-sowing High Protein and Dietary Fiber Soybean Subregion (Subregion III_9); 4) South China Multiple Cropping High Dietary Fiber Soybean Subregion (Subregion III_{10}). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with previous studies (Capelin et al., 2022; Rotundo et al., 2016), but inconsistent with Rotundo et al. (2016). In the study of Rotundo et al. (2016), the oil content did not show clear regional trends. The result of fatty acid agreed with Cherry, Bishop, Hasegawa, and Lefflert (1985). It was different to the result of Abdelghany et al. (2019, 2020) for the different research methods. Abdelghany et al. (2020) obtained the distribution map according to the origins and the fatty acids of the varieties, which were planted in the same site. The isoflavones distribution in this study was also proven similar to the previous studies (Carrão-Panizzi, Berhow, Mandarino, & Oliveira, 2009; Song et al., 2018). Previous research addressed that the fiber content of soybeans was not significantly affected by the geographic region (Grieshop & Fahey, 2001; Karr-Lilienthal, Grieshop, Merchen, Mahan, & Fahey, 2004), however, it was found that the fiber in southern regions was higher than that in the northern region in our study, implying a decreasing trend with the latitude.

Given the cultivation system affected by ecological conditions was not greatly changed by the continuous replacement of varieties and the transfer of production areas, we followed the three-region division scheme of the previous cultivation regionalization (Gai & Wang, 2001; Pu & Pan, 1982) and combined the soybean quality characteristic of each region to form the well-accepted three region scheme. The sub-region division was based on the existing cultivation regionalization (Gai & Wang, 2001; Pu & Pan, 1982) with some modifications according to the soybean quality characteristics. For example, the NESR was divided into two sub-regions according to the characteristics of quality distribution and planting patterns (Gai & Wang, 2001; Pu & Pan, 1982). The advantages of soybean protein and oil contents in the northern region of the NESR were not obvious, but a majority of functional components in soybean seeds here were higher than the southern region. Moreover, this region is an important production base of commercial soybeans in China, in which soybean are commonly cultivated on large scales with high-level mechanization. In the southern region of the NESR, the oil content of soybean is generally high. Soybeans rotated with corn are generally planted at small scale here, and large-scale planting areas are uncommon. The north and south divisions could enable the two areas to form different development characteristics, respectively. The division of the HHHR in this study was also different from previous studies. In previous studies, the region was divided into 3 sub-regions (Chang et al., 2003; Pu & Pan, 1982). Whereas in our study, it was divided into 2 sub-regions by the boundary of the Weihe Plain and the Yellow River, considering the factors of soybean quality distribution patterns, soybean photo-thermal response characteristics, and the topographical features of the region. Soybeans are richer in vitamins in the northern region of HHHR, while the soybeans in the southern region have higher protein content. The division of the SMCR in this study was similar to previous study (Chang et al., 2003; Pu & Pan, 1982). High-level protein content is generally considered as the advantage of soybeans in this region (Qin et al., 2014; Song et al., 2016, 2018). However, it was found that the soybean dietary fiber content, which is associated to the quality of fresh soybeans and various forms of soybean processed foods, is also prominent in this region.

Kriging is an estimation procedure that uses known values and semivariogram to assign optimal weights to determine unknown values. Estimation errors might occur when the quantity of sampling points is small, or the sampling points distribute unevenly (Kleijnen, 2017). Due to the soybean planting area is quite limited in three quarters of the northwest and one third of south China, there were fewer sample sites in these regions. Although these sample sites are representative for these regions, the estimation errors might occur for the small quantity of the sampling points, leading to the differences within regions masked. In the future, this problem can be solved by increasing the number of sampling points, or by masking the accurate radar map of soybean planting area.

Considering the workload and high cost, not all the soybean seed components were determined in the total 1792 samples. Some components such as fatty acids, phospholipids, BBI and lectin were detected in 269 samples. Since the 269 samples were collected respectively from the three soybean planting regions, at least 28 samples for each region, they can still represent the distribution status of soybean in China. Further research will be conducted to improve the accuracy of regionalization in the future when conditions allow us to do it.

5. Conclusions

The spatial variation of primary nutritional components in soybean is illustrated in this study. Most soybean components showed obvious geographical distributing trend, which may be caused by environmental factors. The regionalization scheme of soybean seed quality was proposed according to the spatial patterns of the components and the cultivation systems. The results of this study have important significance for the adjustment of the distribution scheme of soybean production areas in China.

Funding

This study was supported by the National Key R&D Program of China (2018YFD1000900), Earmarked fund for CARS (CARS-04) and the Chinese Academy of Agricultural Sciences (CAAS) Innovation Project.

CRediT authorship contribution statement

Wenwen Song: Conceptualization, Software, Data curation, Writing – original draft. **Shi Sun:** Methodology, Formal analysis, Investigation, Writing – review & editing. **Tingting Wu:** Data curation, Visualization, Writing – review & editing. **Ruping Yang:** Data curation, Investigation, Validation. **Shiyan tian:** Data curation, Investigation, Validation. **Cailong Xu:** Software, Visualization. **Bingjun Jiang:** Writing – review & editing. **Shan Yuan:** Investigation. **Wensheng Hou:** Writing – review & editing. **Cunxiang Wu:** Conceptualization, Supervision. **Tianfu Han:** Supervision, Conceptualization, Investigation, Writing - review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

We thank the CARS soybean experimental stations for collecting the soybean samples. We also thank Guixing Ren, Peiyou Qin, Xiushi Yang, Sancai Liu, Fang Liu, Peng Xue and Yan Li for technical support for composition analysis.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2022.112364.

References

Abdelghany, A. M., Zhang, S., Azam, M., Shaibu, A. S., Feng, Y., Li, Y., … Sun, J. (2020). Profiling of seed fatty acid composition in 1025 Chinese soybean accessions from diverse ecoregions. *The Crop Journal, 8*(4), 635–644.

Abdelghany, A. M., Zhang, S., Azam, M., Shaibu, A. S., Feng, Y., Qi, J., … Sun, J. (2019). Natural variation in fatty acid composition of diverse world soybean germplasms grown in China. *Agronomy, 10*, 24.

Allen, D. W. (2010). GIS tutorial 2: Spatial analysis for ArcGIS 10, ESRI, Redlands. *CA* (pp. 61–124).

Anta, L., Marina, M. L., & García, M. C. (2010). Simultaneous and rapid determination of the anticarcinogenic proteins Bowman-Birk inhibitor and lectin in soybean crops by perfusion RP-HPLC. *Journal of Chromatography A, 1217*(45), 7138–7143.

Assefa, Y., Purcell, L. C., Salmeron, M., Naeve, S., Casteel, S. N., Kovacs, P., … Ciampitti, I. A. (2019). Assessing variation in US soybean seed composition (protein and oil). *Frontiers in Plant Science, 10*, 298.

Azam, M., Zhang, S., Abdelghany, A. M., Shaibu, A. S., Feng, Y., Li, Y., … Sun, J. (2020). Seed isoflavone profiling of 1168 soybean accessions from major growing ecoregions in China. *Food Research International, 130*, Article 108957.

Bosaz, L. B., Gerde, J. A., Borrás, L., Cipriotti, P. A., Ascheri, L., Campos, M., … Rotundo, J. L. (2019). Management and environmental factors explaining soybean seed protein variability in central Argentina. *Field Crops Research, 240*, 34–43.

Cao, Z. H., Green-Johnson, J. M., Buckley, N. D., & Lin, Q. Y. (2019). Bioactivity of soy-based fermented foods: A review. *Biotechnology Advances, 37*(1), 223–238.

Capelin, M. A., Madella, L. A., Panho, M. C., Meira, D., Fernandes, R. A., Colonelli, L. L., … & Benin, G. (2022). Impact of altitude on grain yield, oil, and protein content of soybean. *Australian Journal of Crop Science., 16*(2), 273–279.

Carrão-Panizzi, M. C., Berhow, M., Mandarino, J. M. G., & Oliveira, M. C. N. (2009). Environmental and genetic variation of isoflavone content of soybean seeds grown in Brazil. *Pesquisa Agropecuaria Brasileira, 44*(11), 1444–1451.

Carrera, C., Martínez, M. J., Dardanelli, J., & Balzarini, M. (2011). Environmental variation and correlation of seed components in nontransgenic soybeans: Protein, oil, unsaturated fatty acids, tocopherols, and isoflavones. *Crop Science, 51*(2), 800–809.

Chang, R., Zhou, X., & Qiu, L. (2003). Quality regionalization of soybean in China. In R. Z. Chang (Ed.), *Quality regionalization of soybean in China. Beijing* (pp. 49–54). China Agriculture Press.

Chen, K., Erh, M. H., Su, N., Liu, W., Chou, C., & Cheng, K. (2012). Soyfoods and soybean products: From traditional use to modern applications. *Applied microbiology and biotechnology, 96*, 9–22.

Cherry, J. H., Bishop, L., Hasegawa, P. M., & Lefflert, H. R. (1985). Differences in the fatty acid composition of soybean seed produced in northern and southern areas of the U.S.A. *Phytochemistry, 24*(2), 237–241.

Gai, J., & Wang, Y. (2001). Study on the varietal eco-regions of soybeans in China. *Scientia Agricultura Sinica, 34*(2), 139–145.

Goyal, R., Sharma, S., & Mohammed, J. (2014). Influence of growing environment on antinutrients/bioactive compounds in soybean, *Glycine max* (L.) Merill genotypes. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences, 86*, 343–349.

Grieshop, C. M., & Fahey, G. C. (2001). Comparison of quality characteristics of soybeans from Brazil, China, and the United States. *Journal of Agricultural and Food Chemistry, 49*(5), 2669–2673.

Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D. A., Yang, Z., … Li, W. B. (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist, 209*(2), 871–884.

Hua, Y. F., Gu, Y. X., & Wang, H. J. (2004). Preparation and properties of functional soybean dietary fiber. *China Oils and Fats, 29*(10), 43–46.

Hu, C., Wong, W., Wu, R., & Lai, W. (2020). Biochemistry and use of soybean isoflavones in functional food development. *Critical Reviews in Food Science and Nutrition, 60*(12), 2098–2112.

Karr-Lilienthal, L. K., Grieshop, C. M., Merchen, N. R., Mahan, D. C., & Fahey, G. C. (2004). Chemical composition and protein quality comparisons of soybeans and soybean meals from five leading soybean-producing countries. *Journal of Agricultural and Food Chemistry, 52*(20), 6193–6199.

Kleijnen, J. P. C. (2017). Regression and Kriging metamodels with their experimental designs in simulation: A review. *European Journal of Operational Research, 256*(1), 1–16.

Lee, S. J., Jin, Y. C., Park, S., Chung, J. I., Jin, J. S., Lee, S. J., … Shin, S. C. (2010). Determination of phospholipids in soybean (*Glycine max* (L.) Merr) cultivars by liquid chromatography-tandem mass spectrometry. *Journal of Food Composition and Analysis, 23*(4), 314–318.

Li, Y., Zhao, L., Yun, T., Liu, S., & Qi, W. (2016). Analysis of water-soluble bioactive substances in soybean products. *Journal of Chinese Institute of Food Science and Technology, 16*(2), 258–265.

Martey, E., & Goldsmith, P. (2020). Heterogeneous demand for soybean quality. *African Journal of Agricultural and Resource Economics, 15*(1), 27–50.

Medic, J., Atkinson, C., & Hurburgh, C. R. (2014). Current knowledge in soybean composition. *Journal of the American Oil Chemists' Society, 91*(3), 363–384.

Namazi, N., Saneei, P., Larijani, B., & Esmaillzadeh, A. (2018). Soy product consumption and the risk of all-cause, cardiovascular and cancer mortality: A systematic review and meta-analysis of cohort studies. *Food and Function, 9*(5), 2576–2588.

Ning, H., Zhang, D., Hu, G., Li, W., & Li, W. (2007). Regionization of protein and oil content in soybean (*G. max* Merill) in the Northeast of China. *Soybean Science, 26*(4), 511–516.

Pu, M., & Pan, T. (1982). A study on the regionalization of soybean producing area in China. *Soybean Science, 1*(2), 105–121.

Qin, P., Song, W., Yang, X., Sun, S., Zhou, X. R., Yang, R. P., … Ren, G. (2014). Regional distribution of protein and oil compositions of soybean cultivars in China. *Crop Science, 54*(3), 1139–1146.

Ren, G., Zhu, Y., Shi, Z., & Li, J. (2017). Detection of lunasin in quinoa (*Chenopodium quinoa* Willd.) and the *in vitro* evaluation of its antioxidant and anti-inflammatory activities. *Journal of the Science of Food and Agriculture, 97*(12), 4110–4116.

Rotundo, J. L., Miller-Garvin, J. E., & Naeve, S. L. (2016). Regional and temporal variation in soybean seed protein and oil across the United States. *Crop Science, 56*(2), 797–808.

Sakthivelu, G., Akitha Devi, M. K., Giridhar, P., Rajasekaran, T., Ravishankar, G. A., Nikolova, M. T., … Kosturkova, G. P. (2008). Isoflavone composition, phenol content, and antioxidant activity of soybean seeds from India and Bulgaria. *Journal of Agricultural and Food Chemistry, 56*(6), 2090–2095.

Seguin, P., Tremblay, G., Pageau, D., & Liu, W. (2010). Soybean tocopherol concentrations are affected by crop management. *Journal of Agricultural and Food Chemistry, 58*(9), 5495–5501.

Sharma, S., Kaur, M., Goyal, R., & Gill, B. S. (2014). Physical characteristics and nutritional composition of some new soybean (*Glycine max* (L.) Merrill) genotypes. *Journal of Food Science and Technology, 51*(3), 551–557.

Song, W., Yang, R., Wu, T., Wu, C., Sun, S., Zhang, S., … Han, T. (2016). Analyzing the effects of climate factors on soybean protein, oil contents, and composition by extensive and high-density sampling in China. *Journal of Agricultural and Food Chemistry, 64*(20), 4121–4130.

Song, W., Yang, R., Yang, X., Sun, S., Mentreddy, S. R., Jiang, B., … Han, T. (2018). Spatial differences in soybean bioactive components across China and their influence by weather factors. *The Crop Journal, 6*(6), 659–668.

Wang, X., Yu, C., Lv, J., Li, L., Hu, Y., Liu, K., … Dong, J. (2021). Consumption of soy products and cardiovascular mortality in people with and without cardiovascular disease: a prospective cohort study of 0.5 million individuals. *European Journal of Nutrition, 60*(8), 4429–4438.

Wu, T., Yao, Y., Sun, S., Wang, C., Jia, H., Man, W., … Han, T. (2016). Temporospatial characterization of nutritional and bioactive components of soybean cultivars in China. *Journal of the American Oil Chemists' Society, 93*(12), 1637–1654.

Yang, X., Dong, C., & Ren, G. (2011). Effect of soyasaponins-rich extract from soybean on acute alcohol-induced hepatotoxicity in mice. *Journal of Agricultural and Food Chemistry, 59*(4), 1138–1144.

Zhu, Y., Song, W., Everaert, N., Shi, Z., Han, T., & Ren, G. (2018). Revealing the regional distribution of soybean lunasin content in China and the effects of climate factors by sampling extensively. *Journal of the Science of Food and Agriculture, 99*(6), 2802–2807.