ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
18 March 2021

REVISED
22 June 2021

ACCEPTED FOR PUBLICATION
24 June 2021

PUBLISHED
12 July 2021

LETTER

Relationship of population migration, crop production pattern, and socioeconomic development: evidence from the early 21st century

Yajie Zhang[1,2,*], **Jie Zhao**[3], **Yuxing Yan**[3], **Yu Shi**[3] **and Qiang Yu**[1,4,5,*]

[1] State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Shaanxi 712100, People's Republic of China
[2] Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, People's Republic of China
[3] College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi 712100, People's Republic of China
[4] School of Life Sciences, University of Technology Sydney, Sydney 2007, NSW, Australia
[5] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: zhangyajie1990@yeah.net and yuq@nwafu.edu.cn

Keywords: migration of the center of gravity, population density, crop caloric yield, Fréchet distance, stepwise regression analysis

Supplementary material for this article is available online

Abstract

Global crop production and population distributions have undergone great changes under climate change and socioeconomic development, and have drawn considerable public attention. How to explain the similarity of the migration patterns of crop yield and population density for different countries/regions is still uncertain and worth studying. Here, we estimated the similarity between migrations of main crop caloric yield (i.e. maize, rice, wheat, and soybean) and population density using Fréchet distance, and investigated the regression relationship between Fréchet distance and related climatic and socioeconomic variables for countries/regions with different economic development stages. The results indicated that different countries/regions showed different Fréchet distances during 2000–2015, with a maximum value of 24.44 for Russia and a minimum value of 0.11 for Georgia. For countries/regions with different economic development stages, the built regression models can explain 39%–93% of the variability in the Fréchet distance. Log(land area), log(GDP), and log(land area under cereal production) were always included in regression models and had higher importance in explaining the variability of Fréchet distance. For the model for all countries/regions, both the log(land area) and log(GDP per capita) may positively link to the Fréchet distance. Possible reasons for these results are that countries/regions with high GDP (or GDP per capita) may ease the conflict of land resources between humans and crops to achieve agricultural industrialization, which causes the far connection of the migrations for crop caloric yield and population density. The complicated interactions of crop production, population dynamic, and socioeconomic development should be given greater attention in the future.

1. Introduction

Agriculture provides food supplies to maintain social stability and economic development across the world (FAO 2020). In global agricultural production, maize (*Zea mays* L.), rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), and soybean (*Glycine max* Merr.) are cultivated and produced in most of the croplands (e.g. Zhang *et al* 2021). The distribution of these crops has undergone a great change over time (Sloat *et al* 2020). Migration of production of these main crops is closely related to human living conditions and may correlate with population migration and redistribution, particularly for rural areas and in the least developed countries (De Sherbinin *et al* 2011, Xia *et al* 2019). Population distribution depends upon both natural dynamics and migration decisions (Liu and Yamauchi 2014). This relationship reflects in two opposite aspects: on one hand, population immigration may provide labor to develop agricultural cultivation and production and decrease the areas of farmland abandonment by farmers; on the other hand, population emigration may mitigate the shortage of local cultivated land (Chen 2007, Hazell and Wood 2008, Zhang *et al* 2020a). Agricultural investment and intensification can also fill the gap left by labor outmigration to maintain the cultivation efficiency of food crops (Yang *et al* 2016).

Both migrations of crop production and population are complex phenomena, where many factors acting together, including socioeconomic and political factors in addition to environmental factors (Castelli 2018, Sloat *et al* 2020, Zhang *et al* 2021). Under different levels of economic development, industrialization, and urbanization, and different conditions of topography, climate, and water resources, different countries/regions show different migration patterns of crop yield and population density. Taking China as an example, the historical cultivation and production of major food crops have been identified to northwardly migrate; but the population keeps pouring into the coastal provinces of eastern China (Chen 2007, Ning *et al* 2019). In Najafi *et al* (2018), gross domestic product (GDP) per capita and some climatic factors can effectively explain the variability in historical crop yields for most of the countries from 1961 to 2013, indicating the impact of technology advancement on crop yield fluctuations. For developed nations, the agricultural production is not necessary to close to the population distribution due to the advanced logistics of agricultural products; but it is just the opposite for some underdeveloped nations (Anderson 2010). Now there are still some questions that remained: what do the migration trajectories of main crop yield and population density look like for different countries/regions? Which factors may explain the similarity of these migration patterns for countries/regions with different economic development stages?

Measuring the similarity between trajectories is an important problem that comes up in many research areas such as 3D modeling, handwriting recognition, and applications of geographic information systems. To date, there are dozens of distance measures for trajectory data in the literature, such as Euclidean distance, dynamic time wrapping distance, and Fréchet distance (Su *et al* 2020). The Fréchet distance was originally defined by Maurice Fréchet in 1906 and presented by efficient algorithms in the early 1990s (Fréchet 1906, Alt and Godau 1995). Compared with other trajectory distance measures, the Fréchet distance is a kind of spatiotemporal distance measure that sensitives to the orientation of trajectories (Seyler *et al* 2015, Su *et al* 2020).

In this study, we assessed the migrations of main crop caloric yield (i.e. maize, rice, wheat, and soybean) and population density for the years 2000, 2005, 2010, and 2015 based on corresponding global gridded datasets. The crop caloric yield is the sum of the caloric value of specific crops per unit area. The Fréchet distance was then used to measure the similarity between these two migration trajectories and explained by some related climatic and socioeconomic indicators at the national level. This study aimed to understand the relationship of spatiotemporal changes of crop caloric yield and population density with economic development (and climate change) for countries/regions, which may further help researchers and policymakers to understand rules of socioeconomic development under the double pressure of population growth and food security.

2. Data and methods

The data used in this study were from five main sources (see tables 1 and 2). Based on these data, (a) the total annual caloric yield (unit: calories ha^{-1}) was calculated for each year and each grid cell by multiplying the annual crop yields (unit: t ha^{-1}) from the GDHYv1.2 + v1.3 by the harvested area (unit: ha) from the M3-Crops Data and the crop-specific calorie conversion factors from the 2011 FAO Food Balance Sheets (3580 802.6, 2800 000.0, 3284 000.0, and 3596 499.1 calories t^{-1} for maize, rice, wheat, and soybean, respectively) to get the calorie production (unit: calories) of each crop (Cassidy *et al* 2013); summing these calorie productions across all these four crops; and dividing this sum by the sum of all harvested land (unit: ha) for all these crops from the M3-Crops Data (Iizumi *et al* 2014, Iizumi and Sakai 2020). The total annual caloric yields for the years 2000, 2005, 2010, and 2015 were calculated using the average values of 1999–2001, 2004–2006, 2009–2011, and 2014–2016, respectively. Using these average values may help match the data of population density for followed calculation of the Fréchet distance and to eliminating part of the impact of climate variation on crop production from year to year (Fréchet 1906, Alt and Godau 1995).

(b) Migration of the center of gravity according to spatial coordinates was used to describe the variation of the spatial and temporal patterns of the crop caloric yields and population densities for each country/region during 2000–2015. Assuming that the distribution of crop caloric yields (or population densities) of a country/region consisted of n grid cells, its center of gravity coordinate for a given year was calculated as follows (Zhang *et al* 2017):

$$x = \frac{\sum_{i=1}^{n} P_i x_i}{\sum_{i=1}^{n} P_i}$$

$$y = \frac{\sum_{i=1}^{n} P_i y_i}{\sum_{i=1}^{n} P_i}$$

Table 1. The data used in the calculation of the Fréchet distance between migrations of main crop caloric yield (i.e. maize, rice, wheat, and soybean) and population density.

Variables (unit)	Spatial and temporal resolution	Description and data sources	References
Crop-specific harvested area (ha)	0.083°; during 1997–2003	M3-Crops Data (www.earthstat.org/harvested-area-yield-175-crops/) was applied to identify where a crop of interest was grown.	Monfreda *et al* (2008)
Yield estimations for major crops (maize, rice, wheat, and soybean) (t ha^{-1})	0.5°; 1981–2016	Global Dataset of Historical Yield (GDHYv1.2 + v1.3; https://doi.pangaea.de/10.1594/PANGAEA.909132) was compiled from the crop harvested area around the year 2000 from the M3-Crops data, crop calendar data, satellite-derived crop-specific vegetation index, country yield statistics reported by the Food and Agriculture Organization of the United Nations (FAO), and production share by cropping season from the United States Department of Agriculture. This dataset was verified with global datasets of yields from M3-Crops, Spatial Production Allocation Model, another dataset available at the EarthStat, and subnational yield statistics for major crop-producing countries.	Iizumi *et al* (2014); Iizumi and Sakai (2020)
Population densities (number of persons km^{-2})	0.5°; years 2000, 2005, 2010, 2015, and 2020	Gridded Population of the World (GPWv4.11; https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals-rev11) was based on counts consistent with national censuses and population registers and spatially explicit administrative boundary data, and adjusted to match the 2015 Revision of the United Nation's World Population Prospects country totals and used an areal-weighting method which allocating population from census units into grid cells through a simple assumption that the population was an exclusive function of the land area of that grid cell. The database behind GPW likely represented the most comprehensive collection of census counts and other official population estimates by the administrative unit.	Doxsey-Whitfield *et al* (2015); CIESIN (2018)

where x and y were the longitude and latitude, in decimal degrees (°), of the center of gravity for the crop caloric yield (or population density), P_i was the crop caloric yield (or population density) of the ith grid cell, and x_i and y_i were the centers of gravity coordinates (°) of the ith grid cell. This calculation method of the center of gravity was widely used in analyzing migration flows as a function of the attributes of the locations (Zhang *et al* 2020b).

(c) For two curves given as continuous trajectories in a metric space $\mathbb{R}$, i.e. T_1: [0,1] → $\mathbb{R}$ and T_2: [0,1] → $\mathbb{R}$, the Fréchet distance d_F between T_1 and T_2 was defined as the infimum over all reparametrizations α and β of [0,1] of the maximum over time $t \in [0,1]$ of the distance in $\mathbb{R}$ between $T_1(\alpha(t))$ and $T_2(\beta(t))$:

$$d_F(T_1,T_2) = \underset{\alpha,\beta}{inf}\ \underset{t\in[0,1]}{max}\{d(T_1(\alpha(t)), T_2(\beta(t)))\}$$

where α and β of [0,1] were the continuous, non-decreasing, surjection functions with α, β: [0,1] → [0,1], d was the distance function of $\mathbb{R}$. As shown in the formula above, the Fréchet distance between the two curves indicated the length of the minimum distance sufficient for traversing both curves. This method considered the location and order of the points along the two curves which may have different lengths and was used to measure the similarity between migration trajectories of the crop caloric yield and population density in this study (Fréchet 1906, Alt and Godau 1995, Su *et al* 2020). A large Fréchet distance implied that the curves were not similar. For more details of the formulas of the Fréchet distance, please see Alt and Godau (1995).

(d) Stepwise regression analysis was the step-by-step iterative construction of multiple regression for modeling the relationship between one response variable (dependent variable; the Fréchet distance in this study) and several explanatory variables (independent variables; see table 2). This technique not only guaranteed the validity and importance of the

Table 2. Variables used in the stepwise regression analysis.

Variables (unit)	Description and data sources
Precipitation (mm a^{-1}), temperature (°C) and potential evapotranspiration (mm a^{-1})	These variables were used to indicate the climatic condition for countries/regions in this study. They were derived from CRU TS v4 (Climatic Research Unit gridded Time Series) spatial dataset (https://crudata.uea.acuk/cru/data/hrg/cru_ts_4.05/cruts.2103051243.v4.05/) by converting to time-series data for every country/region. Potential evapotranspiration was calculated using the Penman-Monteith formula. For more details please see Harris *et al* (2020).
[log] GDP (constant 2010 US$), [log] GDP per capita (constant 2010 US$), [log] land area ($km^2$), [log] land area under cereal production (ha), cereal yield (kg ha^{-1}), [log] total population, population density (people km^{-2} of land area), [log] rural population, [log] urban population, [log] refugee population by country or territory of origin, and mean annual exposure of $PM_{2.5}$ air pollution (micrograms m^{-3})	These variables were used to indicate the socioeconomic status of countries/regions in this study (Weinzettel *et al* 2013). They were extracted from the online database of World Development Indicators (https://data.worldbank.org/indicator) which was the primary World Bank collection of development indicators from 1960 for most countries and world regions, compiled from officially recognized international sources (e.g. World Bank national accounts data, FAO data). No temporal interpolation and extrapolation methods were used to fill the missing values in these time-series data. Logarithm transformation was applied for part of variables (labeled by '[log]') following previous literature.

Note: annual mean of the variables during the period 2000–2015 were included in regressions.

chosen variables but also reduced additional errors introduced by the redundant variables. In this study, bidirectional elimination was applied to obtain the optimal model by testing for independent variables to be included or excluded at each step (all of the climatic variables were eliminated in this process for each model). Ten-fold cross-validation was also used to train the model and obtain the best one with minimized root mean squared error. The regression coefficient and importance of variables in the final model were accessed using the unstandardized coefficients and the absolute values of the standardized coefficients, respectively. Moreover, simple linear regression analysis was applied to test the explanatory power of major variables derived from the stepwise regression analysis. The formula for linear regression was:

$$y = \beta_0 + \beta_1 x$$

where y was the dependent variable, x was the independent variable, β_0 was the y-intercept (constant term), and β_1 was the regression coefficient for the independent variable.

All countries/regions were classified into three categories by their economic situation: least developed countries (LDCs), developing countries excluded the LDCs, and developed countries. Data sourced from the World Economic Situation and Prospects 2021 report produced by the United Nations Department of Economic and Social Affairs (www.un.org/development/desa/dpad/publication/world-economic-situation-and-prospects-2021/). Please notice that only 140 countries/regions were remained in this analysis due to these countries/regions had continuous production of main crops and available climatic and socioeconomic data during 2000–2015. R (version 3.5.1; www.r-project.org/) was used to pretreat data to the same spatiotemporal resolution and conduct the calculations of the Fréchet distance and the stepwise regression analysis. The main R packages used in these calculation procedures were 'raster', 'ncdf4', 'rgeos', 'SDMTools', 'kmlShape', 'car', 'MASS', 'leaps', 'rworldmap', and 'ggpubr'. In the later part of the text, the term 'migration of main crop caloric yield' indicated the trajectory change in the location of the total caloric yield of the four main crops, and 'Fréchet distance' indicated the Fréchet distance between migrations of main crop caloric yield and population density.

3. Results

3.1. Migration of the gravity center of the main crop caloric yield and population density for countries/regions in the early 21st century

For spatiotemporal distributions of caloric yield for main crops and population density, please see supplementary material figure S1 (available online at stacks.iop.org/ERL/16/074045/mmedia) for more details. Then from supplementary material figure S2, for countries/regions with the higher Fréchet distance, the gravity centers for the main crop caloric yield and population density varied from country to country. Taking Australia as an example, there is almost no change for the gravity centers of the population density in the early 21st century; however, the gravity centers of the main crop caloric yield changed but had close locations for the 2000 (28.29°S, 143.13°E) and 2015 (29.37°S, 142.91°E). For Indonesia, the gravity centers of the population density shifted eastward from (4.98°S, 109.38°E) in 2000 to (4.85°S, 109.79°E) in 2015. However, the

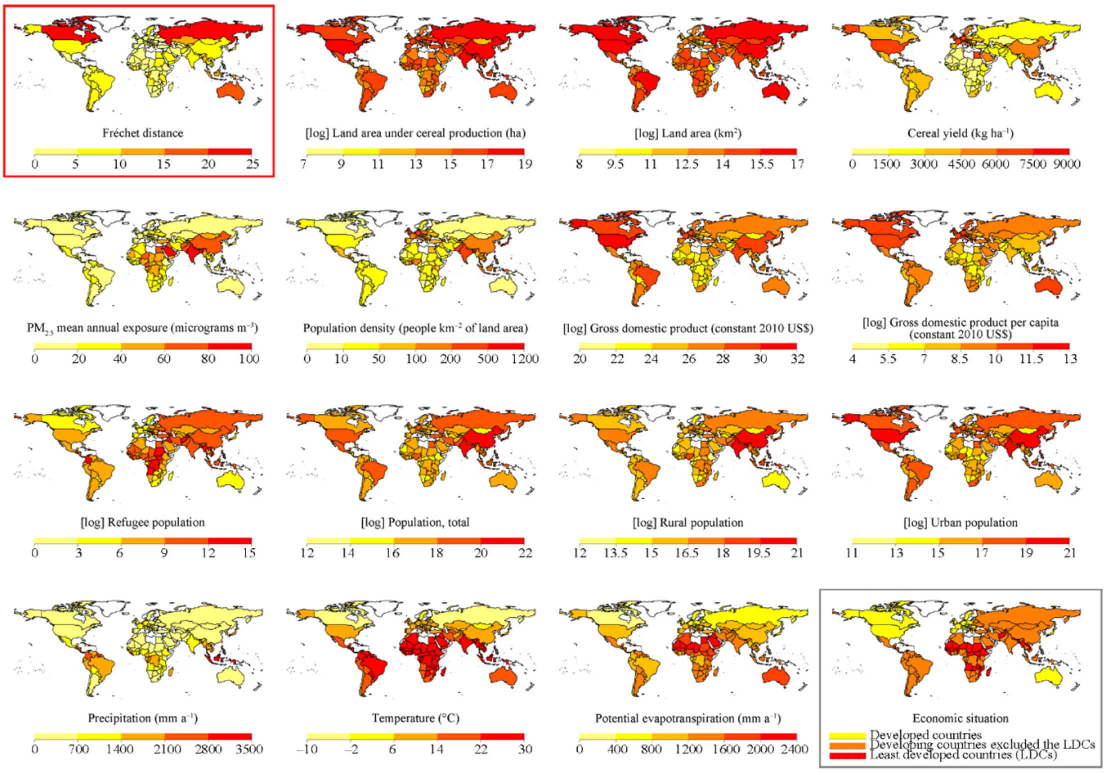

Figure 1. Fréchet distance between migrations of main crop caloric yield (i.e. maize, rice, wheat, and soybean) and population density for the years 2000, 2005, 2010, and 2015 (indicating by a red border), related climatic and socioeconomic indicators (annual mean during 2000–2015; see table 2), and categories of countries/regions according to their economic situation (indicating by a gray border) across the world. White in these maps indicates no main crop production or no data.

gravity centers of the main crop caloric yield westward migrated from (1.06°S, 106.65°E) in 2000 to (1.55°S, 104.97°E) in 2015. In general, most of these countries/regions with the higher Fréchet distance showed different migration directions for the gravity centers of the main crop caloric yield and population density (e.g. Canada), or stable population distribution but a varied pattern of crop caloric yield (e.g. Chile). For countries/regions with the lower Fréchet distance from supplementary material figure S2, most of these countries/regions were observed to slightly migrate for the gravity centers of the main crop caloric yield and population density (e.g. Algeria, Tajikistan), or unchanged pattern of crop caloric yield or population density distribution but slightly varied another migration (e.g. Georgia).

3.2. Spatial distribution of Fréchet distance and related socioeconomic variables in the early 21st century

From figure 1, (a) Russia, Canada, Australia, and Indonesia show the highest Fréchet distance between migrations of main crop caloric yield and population density. Some countries/regions such as China, Brazil, Argentina, Kazakhstan, and Saudi Arabia were close behind. Most countries/regions in Europe and Africa presented lower Fréchet distance. (b) USA, Japan, South Korea, Egypt, and some Eastern European countries led the cereal yield across the world. (c) For $PM_{2.5}$ (particulate matter with an aerodynamic diameter less than or equal to 2.5 μm) mean annual exposure, some Asian and African countries ranked first. (d) Most of the developed countries and some important developing countries (e.g. China, India, and Brazil) had the higher GDP; however, the different spatial pattern of GDP per capita was observed across the world. Countries/regions with higher GDP per capita were mainly distributed in Western Europe, Northern America, Australia, and Japan. Countries/regions with lower GDP per capita were mainly located in southern and southeastern Asia, the Arab world, and sub-Saharan Africa. (e) Canada, Australia, and some countries in Eastern Europe and Southern Africa had the lowest values of the refugee population. In addition, the descriptions of some variables that were like common sense (e.g. land area, total population, and temperature) were not given here.

3.3. Relationship between Fréchet distance and related climatic and socioeconomic variables in the early 21st century

Models based on related climatic and socioeconomic variables can explain 39%–93% of the variability in the Fréchet distance during 2000–2015 for different categories of countries/regions (table 3). The importance of each variable in each model was further tested and ranked by standardized coefficients. Among the

Table 3. Stepwise regression models which built in this study.

Explanatory power of major variables ($^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$)									
Model	Stepwise regression equation (and variable importance rank as superscripts)	Number of countries/regions	p-value	Root mean square error	R^2	GDP	Land area	Population	Land area under cereal production
World	Fréchet distance = –16.087–0.006 × log(land area under cereal production)^9 + 2.091 × log(land area)^4–0.00001 × cereal yield^8 –0.026 × $PM_{2.5}$ annual exposure^6 + 0.004 × population density^5–17.570 × log(GDP)^1 + 17.964 × log(GDP per capita)^2 + 0.008 × log(refugee population)^7 + 17.021 × log (population)^3	140	<0.001	3.18	0.43	0.16^{***}	0.37^{***}	0.13^{***}	0.16^{***}
Developed countries	Fréchet distance = –2.099–2.171 × log(land area under cereal production)^2 + 5.014 × log (land area)^1–1.542 × log(rural population)^3	28	<0.001	3.20	0.93	0.15^{*}	0.56^{***}	0.09	0.21^{*}
Developing countries	Fréchet distance = –13.179 + 1.332 × log(land area)^1	112	<0.001	2.91	0.39	0.20^{***}	0.33^{***}	0.15^{***}	0.15^{***}
—Developing countries excluded the LDCs	Fréchet distance = –15.785 + 1.549 × log(land area)^1	77	<0.001	3.11	0.45	0.21^{***}	0.39^{***}	0.14^{***}	0.15^{***}
—Least developed countries (LDCs)	Fréchet distance = –8.451 + 0.974 × log(land area under cereal production)^1–0.197 × log(land area)^3 + 0.0002 × cereal yield^5 + 0.012 × $PM_{2.5}$ annual exposure^4–0.005 × population density^2	35	<0.001	2.01	0.57	0.23^{**}	0.13^{*}	0.17^{*}	0.18^{*}

Note: for categories of countries/regions please see figure 1. In stepwise regression models, the p-values of independent variables were not shown because they were not the basis of variable elimination.

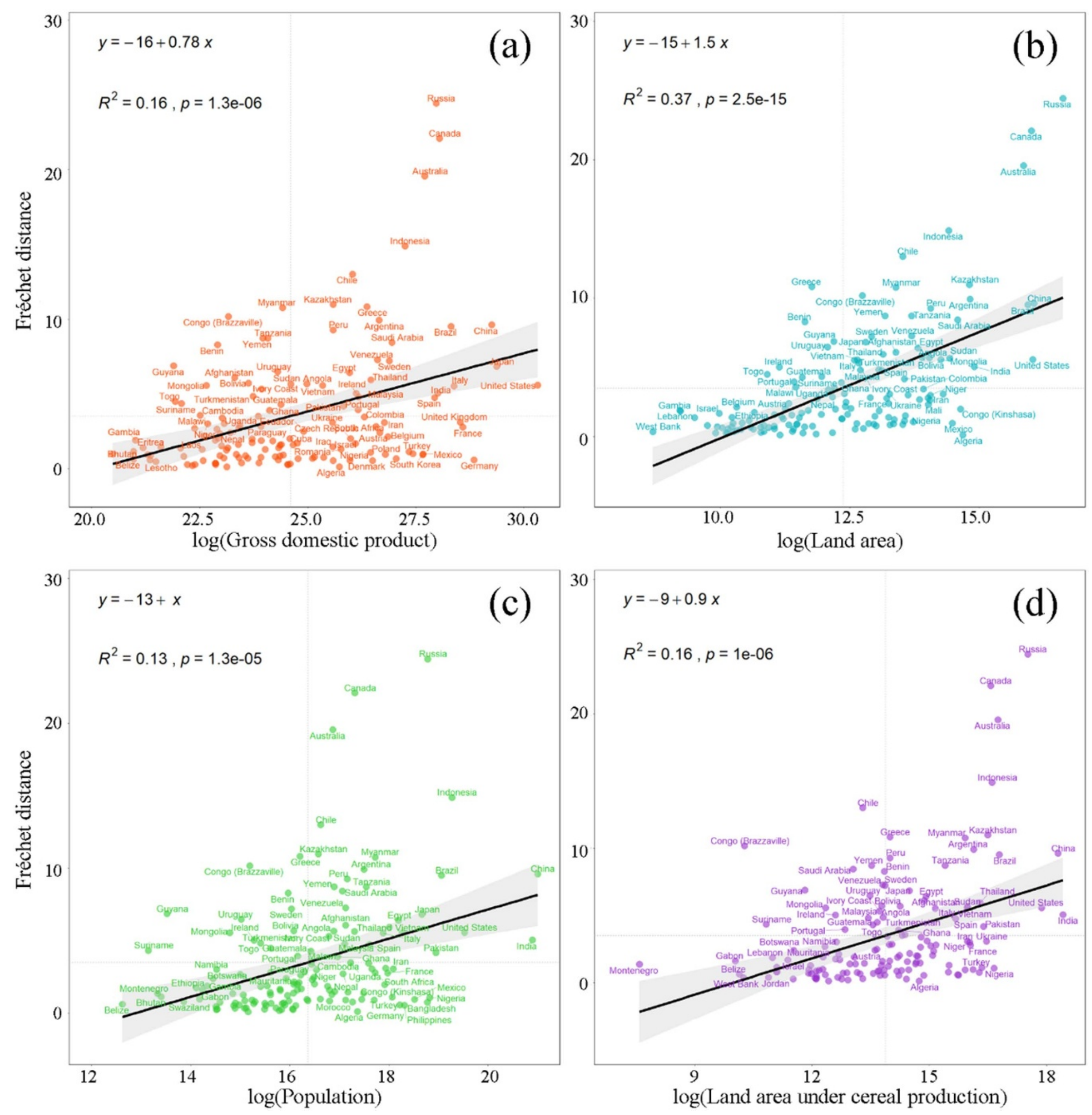

Figure 2. Correlation relationship between Fréchet distance between migrations of main crop caloric yield (i.e. maize, rice, wheat, and soybean) and population density and four major variables (i.e. (a) gross domestic product, (b) land area, (c) population, and (d) land area under cereal production) during 2000–2015 across the world. The data points represent different countries/regions (some data points are unlabeled due to overlapping); the vertical and horizontal lines indicate the means of major variable and Fréchet distance, respectively; and the black line and associated grey shadow indicate the regression line and its 95% confidence interval, respectively.

selected variables, log(land area) was always included in regression models and had higher importance. It was the strongest single predictor of Fréchet distance for most of the categories of countries/regions, explaining 13%–56% of the variability in Fréchet distance (table 3 and figure 2). And log(GDP) and log(land area under cereal production) also explained 15%–23% and 15%–21% of the variability in Fréchet distance, respectively (figures not shown). Furthermore, other variables such as log(refugee population) and $PM_{2.5}$ annual exposure were also included in models for all countries/regions and least developed countries. But these variables had lower importance in impacting the Fréchet distance. For the model for all countries/regions (n = 140), the higher the log(land area) and log(GDP per capita), the larger the Fréchet distance. For the model for least developed countries (n = 35), increasing population density caused decreasing Fréchet distance. Most of these countries/regions were in sub-Saharan Africa, and Western and Southeastern Asia (see figure 1).

4. Discussion

4.1. Causes of different Fréchet distance for countries/regions

In this study, the Fréchet distance is determined by different migration patterns of crop caloric yield and population density, which affecting by different conditions of socioeconomic, political, and environmental factors for countries/regions (Castelli 2018, Sloat et al 2020). Jackson et al (2019) indicated that between 2000 and 2010, distribution of maize increased in northeastern China, southern India, Eastern Europe, southeastern Brazil, and eastern United State; distribution of rice increased in southeastern Asia but decreased in southern China, southern Japan, and eastern India; distribution of soybean increased in Eastern Europe, southern Brazil, and eastern United State; and distribution of wheat mainly showed a decreasing trend in Western Europe and eastern and central United State but showed an increasing trend in southern Australia. For internal migration intensities of population, Bell *et al* (2018) indicated that progressive decline happened in a wide range of countries since the 1980s, including Russia, Brazil, Japan, and South Korea. However, Europe had observed an upward trend, whereas others had experienced stable migration intensities.

Taking Australia as an example, crop production in Australia is mainly conducted under rain-fed conditions and has become much more variable from season to season since about 2000 due to the increased variability in climate and inappropriate application of nitrogen fertilizers (Anderson *et al* 2016). Population redistribution in Australia is dominated by the intensity of internal migration, which has declined since the early 1980s for both local and long-distance moves (Kalemba *et al* 2020). The combination of these two sides (and large land area that ranked 6th around the world) causes the high Fréchet distance for Australia. For some developed European countries, suburbanization becomes the most widespread mode of expansion for a long phase (Salvati *et al* 2013). This makes the population concentrate near the crop production areas and causes the low Fréchet distance.

4.2. The role of socioeconomic variables on agriculture development, population variation, and their Fréchet distance

In this study, 11 socioeconomic variables and three climatic variables are considered in regression analysis. These variables are identified as important factors affecting agriculture development and population variation. Some main results are indicated here (table 3):

(a) In addition to the log(land area), log(GDP) has the highest explanatory power to the variability in Fréchet distance in most models. For all countries/regions, positive relationships between Fréchet distance and log(land area) and log(GDP per capita) are shown. Higher GDP (or GDP per capita) always means the higher capacity of crop production, transport, storage, and consumption, as well as the higher capacity of countries to invest in agriculture improvement (Hazell and Wood 2008). It also leads to the declined share of agriculture in the total GDP that agricultural population diversifies into the manufacturing and service sectors. It eases the conflict of land resources between humans and crops to enhance the agricultural potential and develop industrial agriculture (Hazell and Wood 2008, Kremen *et al* 2012). Moreover, countries/regions with high GDP (or GDP per capita) can design policies to develop small- and medium-sized cities and reduce regional inequality between the coastal and the inland areas for reducing long-distance and large-scale population migration and improving the level of urbanization (Satterthwaite *et al* 2010, Zhang *et al* 2020b). Relatively, for countries/regions with lower GDP (or GDP per capita), population growth leads to expansion of the cropped area around or close to urban centers and smaller farm sizes, which causes the close connection of the migrations for crop caloric yield and population density (Hazell and Wood 2008, Satterthwaite *et al* 2010, Ricker-Gilbert *et al* 2014). In regions across Africa and Asia that having low GDP (or GDP per capita), the problem of economic development is expected to worsen due to declined agricultural productivity and rapid population growth (Hanjra *et al* 2013).

(b) Log(refugee population) and $PM_{2.5}$ annual exposure are included in the regression model for all countries/regions; $PM_{2.5}$ annual exposure is also included in the regression model for least developed countries. In some case studies, a 10% reduction in yields of maize and wheat leads an additional 2% of the population to become refugees and migrate in Mexico (Feng *et al* 2010); in China and Italy, reduction of $PM_{2.5}$ can contribute to the increased yields of wheat and maize and decreased net internal migration of population (e.g. Zhou *et al* 2018, Germani *et al* 2021).

(c) Negative relationship between Fréchet distance and population density is shown for the least developed countries. Coupled relationship between increasing grain yield and agricultural labor is mainly concentrated in the underdeveloped farming areas (Ge *et al* 2018). In Malawi, areas of higher population density are associated with strengthen agricultural intensification and increasing maize yields (Ricker-Gilbert *et al* 2014).

(d) Although climate and its variability are identified to affect the distribution of agricultural production and cause about a third of the annual variability in agricultural yields (Porfirio *et al* 2017); and the population is also linked to climate through adaptation and mitigation (Stephenson *et al* 2010). However, these climatic variables are not included in regression models. Besides, other variables included in regression models are not discussed here due to they are directly linked with agricultural production or population.

4.3. Limitations of this study

Because the calculation is based on imperfect data and methods, this study has several potential limitations as following: (a) only four major crops (i.e. maize, rice, wheat, and soybean) are considered in this study. However, some kinds of crops such as cassava may be largely grown and produced to provide calories in tropical regions (especially Africa and South America) (Zhang and Yu 2020). Countries/regions in tropical regions may show different migration trajectories of crop caloric yield and population density due to their limited major crop production. (b) The yields available in the GDHY are model estimates and are influenced by misreporting or lacking in agricultural census statistics and use of time-constant data for the crop calendar, the share of production by cropping system, and harvested area map (Iizumi *et al* 2014, Iizumi and Sakai 2020). These lead to limited reliability of the modeled yields in minor crop-producing countries/regions (Iizumi *et al* 2014, Iizumi and Sakai 2020). (c) In the data of GPWv4.11, the statistical unit of census data varies from country to country. Using areal-weighting as the disaggregation method may vary the precision of estimates: for countries/regions where the administration unit of input data are large or the availability and quality of census population data are limited, the precision of population estimates for individual grid cells within those countries/regions is degraded (Doxsey-Whitfield *et al* 2015). (d) Conclusions are sensitive to the choice of crop yield and population datasets. For example, Tatem *et al* (2011) uses four different global gridded population datasets (i.e. GRUMP, LandScan, UNEP Global Population Databases, and GPW) to estimate populations at risk of *P. falciparum* malaria and indicates that none of the datasets is consistently more accurate than the others; differences among these datasets are sourced from the methods, input resolution, and date of the census data and mainly show in the low-income countries/regions. Similarly, different spatial resolutions of datasets could also lead to different conclusions. Better estimates may be produced due to the size of the selected country/region that is larger than the given resolution (i.e. $0.5° \times 0.5°$; $\sim$55 km $\times$ 55 km) (Doxsey-Whitfield *et al* 2015). Also, some limitations are related to the boundary data used to generate the migration of the gravity center. Moreover, (e) to ensure the consistency of data sources in the regression analysis, this study only considers the variables sourced from the databases of World Development Indicators and Climatic Research Unit. Some other indicators such as the Fragile States Index (https://fragilestatesindex.org/data/) which measures conflict in countries/regions may have a strong correlation with the Fréchet distance. There is still room for improvements to limit these uncertainties.

5. Conclusions

Over the past two decades, abundant spatial changes in the distributions of crop production and human population have taken place in most countries/regions. Knowledge about these changes is fundamental for tackling many global challenges. Aiming at food security and urbanization process under resource constraints, our findings suggest that similarity between migrations of main crop caloric yield and population density (measured by the Fréchet distance) is mainly explained by log(land area), log(GDP), and log(land area under cereal production) for countries/regions with different economic development stages. Log(GDP) is significantly positively correlated with the Fréchet distance, indicating the high level of national economic development may diverse the migrations of main crop caloric yield and population density.

This study may provide new opinions on the interlinkage of distributions of crop production and population. In further studies, the relationship of crop production, population, and socioeconomic development is needed to be analyzed afresh due to the changing global population and declining fertility rate in many countries, which would have positive implications for the environment and climate change, but possible negative implications for economic growth and social stability in parts of the world (Vollset *et al* 2020). For future policy options for different countries/regions with different economic development stages, variations in important socioeconomic indicators should be paid close attention to optimize the agricultural development and population redistribution (i.e. lessen the Fréchet distance).

Acknowledgments

This study was funded by the National Natural Science Foundation of China (Nos. 41961124006, 41730645), the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDA20040301), and the '111 Project' of China (No. B12007). The authors are very grateful to the editor and two anonymous reviewers for their valuable comments that have significantly improved the paper. The authors thank Mr. Shouhua Xu at the Institute of Soil and Water Conservation, Northwest A&F University for providing suggestions on the manuscript.

Conflict of interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

References

Alt H and Godau M 1995 Computing the Fréchet distance between two polygonal curves *Int. J. Comput. Geom. Appl.* **5** 75–91

Anderson K 2010 Globalization's effects on world agricultural trade, 1960–2050 *Phil. Trans. R. Soc.* B **365** 3007–21

Anderson W, Stephens D and Siddique K 2016 Dryland agriculture in Australia: experiences and innovations M Farooq and K Siddique ed *Innovations in Dryland Agriculture* (Berlin: Springer) 299–319

Bell M, Charles-Edwards E, Bernard A and Ueffing P 2018 Global trends in internal migration T Champion, T Cooke and I Shuttleworth ed *Internal Migration in the Developed World: Are We Becoming Less Mobile* (London: Routledge) 76–97

Cassidy E, West P, Gerber J and Foley J 2013 Redefining agricultural yields: from tonnes to people nourished per hectare *Environ. Res. Lett.* **8** 034015

Castelli F 2018 Drivers of migration: why do people move? *J. Travel Med.* **25** 1–7

Chen J 2007 Rapid urbanization in China: a real challenge to soil protection and food security *Catena* **69** 1–15

CIESIN, 2018. Gridded population of the world, version 4: population density adjusted to match 2015 revision UN WPP country totals, revision 11 (Palisades, NY: US NASA Socioeconomic Data and Applications Center (SEDAC)) (https://doi.org/10.7927/H4F47M65)

De Sherbinin A *et al* 2011 Preparing for resettlement associated with climate change *Science* **334** 456–7

Doxsey-Whitfield E, MacManus K, Adamo S, Pistolesi L, Squires J, Borkovska O and Baptista S 2015 Taking advantage of the improved availability of census data: a first look at the Gridded Population of the World, version 4 *Papers Appl. Geogr.* **1** 226–34

FAO 2020 FAOSTAT: agriculture (Rome: Food and Agricultural Organizations of the United Nations (available at: www.fao.org/corp/statistics/en/) (7 October 2020)

Feng S, Krueger A and Oppenheimer M 2010 Linkages among climate change, crop yields and Mexico–US cross-border migration *Proc. Natl Acad. Sci. USA* **107** 14257–62

Fréchet M 1906 Sur quelques points du calcul fonctionnel *Rend. Circ. Mat. Palermo* **22** 1–72

Ge D, Long H, Zhang Y and Tu S 2018 Analysis of the coupled relationship between grain yields and agricultural labor changes in China *J. Geog. Sci.* **28** 93–108

Germani A, Scaramozzino P, Castaldo A and Talamo G 2021 Does air pollution influence internal migration? An empirical investigation on Italian provinces *Environ. Sci. Policy* **120** 11–20

Hanjra M, Ferede T, Blackwell J, Jackson T and Abbas A 2013 Global food security: facts, issues, interventions and public policy implications *Global Food Security: Emerging Issues and Economic Implications* ed M Hanjra (New York: Nova Science Publishers) pp 1–35

Harris I, Osborn T, Jones P and Lister D 2020 Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset *Sci. Data* **7** 109

Hazell P and Wood S 2008 Drivers of change in global agriculture *Phil. Trans. R. Soc.* B **363** 495–515

Iizumi T and Sakai T 2020 The global dataset of historical yields for major crops 1981–2016 *Sci. Data* **7** 97

Iizumi T, Yokozawa M, Sakurai G, Travasso M, Romanenkov V, Oettli P, Newby T, Ishigooka Y and Furuya J 2014 Historical changes in global yields: major cereal and legume crops from 1982 to 2006 *Global Ecol. Biogeogr.* **23** 346–57

Jackson N, Konar M, Debaere P and Estes L 2019 Probabilistic global maps of crop-specific areas from 1961 to 2014 *Environ. Res. Lett.* **14** 094023

Kalemba S, Bernard A, Charles-Edwards E and Corcoran J 2020 Decline in internal migration levels in Australia: compositional or behavioural effect? *Popul. Space Place* e2341

Kremen C, Iles A and Bacon C 2012 Diversified farming systems: an agroecological, systems-based alternative to modern industrial agriculture *Ecol. Soc.* **17** 44

Liu Y and Yamauchi F 2014 Population density, migration, and the returns to human capital and land: insights from Indonesia *Food Policy* **48** 182–93

Monfreda C, Ramankutty N and Foley J 2008 Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000 *Glob. Biogeochem. Cycles* **22** GB1022

Najafi E, Devineni N, Khanbilvardi R and Kogan F 2018 Understanding the changes in global crop yields through changes in climate and technology *Earth's Future* **6** 410–27

Ning X, Zhang L, Qin Y and Liu K 2019 Temporal-spatial distribution of suitable areas for major food crops in China over 60 years *Adv. Earth Sci.* **34** 191–201 (in Chinese with English abstract)

Porfirio L, Newth D, Harman I, Finnigan J and Cai Y 2017 Patterns of crop cover under future climates *Ambio* **46** 265–76

Ricker-Gilbert J, Jumbe C and Chamberlin J 2014 How does population density influence agricultural intensification and productivity? Evidence from Malawi *Food Policy* **48** 114–28

Salvati L, Morelli V, Rontos K and Sabbi A 2013 Latent exurban development: city expansion along the rural-to-urban gradient in growing and declining regions of Southern Europe *Urban Geogr.* **34** 376–94

Satterthwaite D, McGranahan G and Tacoli C 2010 Urbanization and its implications for food and farming *Phil. Trans. R. Soc.* B **365** 2809–20

Seyler S, Kumar A, Thorpe M and Beckstein O 2015 Path similarity analysis: a method for quantifying macromolecular pathways *PLoS Comput. Biol.* **11** e1004568

Sloat L, Davis S, Gerber J, Moore F, Ray D, West P and Mueller N 2020 Climate adaptation by crop migration *Nat. Commun.* **11** 1243

Stephenson J, Newman K and Mayhew S 2010 Population dynamics and climate change: what are the links? *J. Public Health* **32** 150–6

Su H, Liu S, Zheng B, Zhou X and Zheng K 2020 A survey of trajectory distance measures and performance evaluation *VLDB J.* **29** 3–32

Tatem A, Campiz N, Gething P, Snow R and Linard C 2011 The effects of spatial population dataset choice on estimates of population at risk of disease *Popul. Health Metr.* **9** 4

Vollset S *et al* 2020 Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study *Lancet* **396** 1285–306

Weinzettel J, Hertwich E, Peters G, Steen-Olsen K and Galli A 2013 Affluence drives the global displacement of land use *Glob. Environ. Change* **23** 433–8

Xia H, Adamo S, De Sherbinin A and Jones B 2019 The influence of environmental change (crops and water) on population redistribution in Mexico and Ethiopia *Appl. Sci.* **9** 5219

Yang J, Wang H, Jin S, Chen K, Riedinger J and Peng C 2016 Migration, local off-farm employment, and agricultural production efficiency: evidence from China *J. Product. Anal.* **45** 247–59

Zhang S, Deng W, Peng L, Zhou P and Liu Y 2020a Has rural migration weakened agricultural cultivation? Evidence from the mountains of southwest China *Agriculture* **10** 63

Zhang X, Wang W, Harris R and Leckie G 2020b Analysing inter-provincial urban migration flows in China: a new multilevel gravity model approach *Migr. Stud.* **8** 19–42

Zhang Y, Niu H and Yu Q 2021 Impacts of climate change and increasing carbon dioxide levels on yield changes of major crops in suitable planting areas in China by the 2050s *Ecol. Indic.* **125** 107588

Zhang Y, Wang Y and Niu H 2017 Spatio-temporal variations in the areas suitable for the cultivation of rice and maize in China under future climate scenarios *Sci. Total Environ.* **601–602** 518–31

Zhang Y and Yu Q 2020 Identification of current research intensity and influence factors of agricultural nitrogen loss from cropping systems *J. Cleaner Prod.* **276** 123308

Zhou L, Chen X and Tian X 2018 The impact of fine particulate matter ($PM_{2.5}$) on China's agricultural production from 2001 to 2010 *J. Cleaner Prod.* **178** 133–41