Migration surge under the context of climate change: a case study of China

Haoxi Chen, Stéphane Goutte

Migration surge under the context of climate change: a case study of China

Haoxi CHEN[*], Stephane GOUTTE[†]

January 2024

Abstract

Climate change has become one of the principal global concerns, with an expected increase in the frequency of extreme weather events, rising temperatures, and shifts in precipitation patterns, all of which are likely to profoundly impact agriculture, human health, and the economy. In China, the swift process of urbanization, coupled with climate change, has heightened the complexity of migration patterns. The intricate interplay between socio-economic and environmental changes compels people to consider migration as a strategy for adaptation. Our research reveals a complex interconnection between climate change and the patterns of population influx and outflow. This study aims to explore the effects of climate change on migratory dynamics to inform policy-making for nations grappling with the dual challenges of economic growth and environmental sustainability.

Keywords: Climate change; Migration patterns; China.

Ethics approval and consent to participate: All procedures were approved by the UMI-SOURCE at the Paris Saclay University.
Consent for publication: No applicable
Competing Interests: The authors declare that they have no competing interests.
Author contributions: Haoxi CHEN and Stéphane GOUTTE conceived of the presented idea, developed the theory, and performed the computations. Stéphane GOUTTE verified the analytical methods and supervised the findings of this work. Haoxi CHEN contributed to the final version of the manuscript.
Funding: No applicable
Availability of data and materials: Despite having authorization to analyze and report on the migration data, contractual obligations and privacy concerns prevent the authors from releasing the raw dataset. Interested researchers may seek access by contacting https://www.geodata.cn/wjw/ and complying with the necessary agreements. Climate data are sourced from the Copernicus Climate Change Service (https://cds.climate.copernicus.eu/cdsapp/home), while disaster data are obtained from EM-DAT (https://www.emdat.be/). Pollution and socio-economic data are derived from published yearbooks.
Word count: 7173

[*]UMI SOURCE
[†]UMI SOURCE and Climate Economic Chair Paris Dauphine

1 Introduction

Climate change has emerged as a major concern for the international community. Climate variability is predicated to bring increased frequency of extreme weather events, elevated temperatures, melting ice caps and changing precipitation pattern (IPCC, 2007). An increasing body of research shows that the social and economic costs of climate change are very large, impacting agriculture, mortality, labor productivity, economic growth, conflict and migration (Dell et al., 2014; Deschênes and Greenstone, 2007; Lobell et al., 2008).

According to the United Nations Department of economic and social affairs' report in 2022, the number of international migrants has been steadily rising from 173 million in 2000 to 281 million in 2020, constituting 2.8% and 3.6% of the global population respectively. In 2020, women accounted for 135 million of migrants (3.5% of the world's female population) while men accounted for 146 million (3.7% of the global male population) (UN, 2022).

If climate variability processes continue, sustaining livehoods may become increasingly unfeasible, leading to a rising number of people requiring permanent relocation(Marchiori et al., 2012; Swim et al., 2011). The World Bank projects that weather-related extreme events could result in greater than 140 million additional people displaced within their own countries by 2050 (Rigaud et al., 2018). Importantly, it must be emphasised that the scale of internal migration is larger than that of international migration. Regardless of the determinant of migration, people are more sensitive to socioeconomic differences within their own country than among different countries (Coniglio and Pesce, 2015). Therefore, climate variability gives a stronger impact on internal migration compared to international migration (Beine et al., 2014).

1.1 Related litterature

Migration, as a social phenomenon, is closely intertwined with the challenges given by climate change. Climate change is always regarded as a stressor that intensifies migration pressure in vulnerable regions (Gemenne, 2011a). Facing the pressures brought by environmental variability, Reuveny identified three ways in which people can adapt to environmental challenges: staying in place and doing nothing, accepting the costs; staying in place and mitigating the changes; or leaving the affected area. Every choice is influenced by individual's available resources and his/her foresight. Comparatively, the costs of the other two choices are lower than migration (Reuveny, 2007).

Migration as an adaption strategy, can alleviate the pressure of population and natural resources, reduce risks and provide improved living conditions. Migration can significantly mitigate the potential impacts of future shocks (da haan, 2000). Figure 1 potentially depicts the mechanism by which climatic-environmental pressure precipitates migratory movements. Within this dynamic, environmental pressure, induced by climate change such as extreme weather occurrences and natural disasters, may result in the scarcity of exploitable resources. Under these stress conditions, individuals might opt to relocate to regions characterized by richer resource availability and diminished environmental pressure, in pursuit of enhanced prospects for survival and development. Such migratory process persist until a new, relatively stable state of equilibrium is achieved in two regions.But migration is costly in both financial and sociological/psychological terms because individuals tend to develop strong personal connections with their home location and its people throughout their lives (Koubi et al., 2016; Lewicka, 2011). Only when an environmental event has a profound impact on personal life, and personal adaptation or mitigation measures are failing, the option of migration becomes a consideration (Speare, 1974).

Environment variability, as a form of pressure, affects individuals differently, and people respond to it also in diverse manners. The impact of environmental events on migration is context-dependent and influenced by several factors, including household characteristics, socioeconomic status and political condition, etc. (Black et al., 2011; Hunter et al., 2015). A study of rural families in Salvador indicated that people tend to choose the alternative strategies to reduce damages when the

[1] Source:Haoxi CHEN

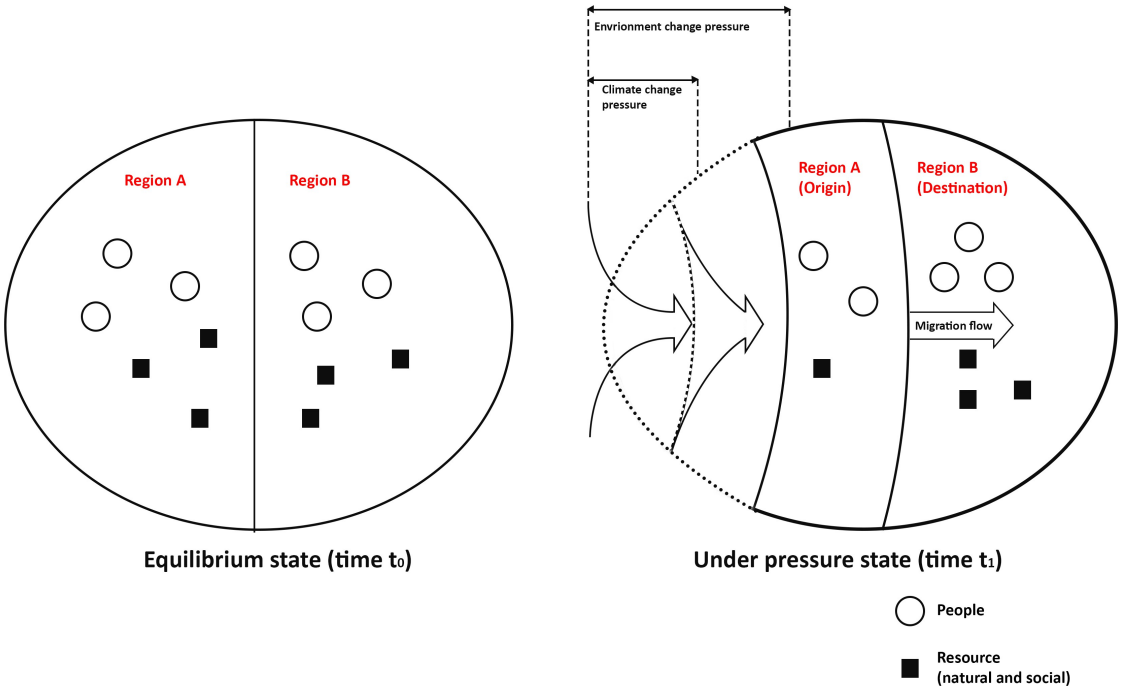

Figure 1: Climatic-environmental pressure and migration[1]

environmental shock is temporary, even if the shock is highly destructive (Halliday, 2006). Only when environmental variability poses a long-term risk to personal wellbeing, individual will prefer the higher cost approach to solve the risk fundamentally, for example, migration.

Migration activity is generally associated with the concept of vulnerability, which is a function of exposure and adaptive capacity in a particular time and place and in relation to a specific climatic stimulus (Gemenne, 2011b; Scheffran et al., 2012). It is important to note that climate change can act as a source of pressure, triggering migration activities, the resulting pressure doesn't affect everyone uniformly. From an individual's perspective, the decision to migrate is a subjective process. Migration decisions may be influenced more by perceptions of environmental problems rather than by specific identified environmental events (Bylander, 2015; Koubi et al., 2016). For example, Perch-nielsen's research on migration caused by sea level rise and flood revealed that floods are unlikely to be the principal mechanism by which climate change triggers mass migration(L. Perch-Nielsen et al., 2008). Migration reasons often stem from a wide range climate shocks across different locations and timeframes (Renaud et al., 2011). Unfortunately, up to now, we haven't evidenced a research comprehensively analyzes the relationship between this wide spectrum of climate shocks and migration activity.

Climate change also exerts asymmetrically on individuals. Firstly, the influence given by climate change doesn't operate uniformly across every aspect of personal life (Coniglio and Pesce, 2015). An individual's life can be regarded as a multivariable function encompassing gender, age, education, financial asset, family relationships, etc. Given the varying weights weight of these factors in personal life and the unequal impact of climate change pressure on them, an individual's decision on migration will differ significantly. Individuals always adapt themselves to long-term environmental pressure. For instance, Agder has demonstrated Vietnam society's adaption to prolonged environmental pressure (Koubi, 2016). Differences in strategic choices are not only due to individual circumstances, but also due to the nature and intensity of climate shock. This explains why, following the climate shock, some people migrated while others did not, in spite of experiencing similar environmental conditions (McLeman and Smit, 2006).

In the broader context of society, climate change gives a necessary impact on migration decisions by influencing the market dynamics. Marichiori highlighted how climate shocks affect migration decisions by influencing wage disparities between potential migrant origins and destinations (Mar-

Table 1: Overview of selected studies on Climatic-environmental change and migration

Author (Year)	**Research Topic**	**Methodology**	**Key Findings**
Hailliday, T.(2006)	Rural household migration in El Salvador	Panel data regression	When environmental impacts are temporary, even if they are highly destructive, individuals might opt for alternative adaptation strategies to mitigate losses.
L. Perch-Nielsen, S., et al. (2008)	Migration related to sea level rise and flooding	Box-and-arrow conceptual model	Floods are unlikely to be the principal mechanism triggering mass migration due to climate change.
Koubi, V.,et al. (2016)	Vietnam's societal adaptive capacity to climate change	Logistic regression	The differences in migration strategic choices arise not only from individual circumstances but also from the form and intensity of climate impacts.
Marchiori, L.,et al.(2011)	Weather anomalies on migration in sub-Saharan Africa	Two stage regression	Climate shocks affect the wage differences between places of origin and potential destinations, thereby influencing migration decisions.
Landry, C.E., et al. (2007)	Migration following Hurricane Katrina in the United States	Logistic regression	After the occurrence of the hurricane, a substantial number of impoverished African Americans were unable to leave New Orleans, and later, it was this poorest group that could not return to New Orleans.
Henry, S.,et al. (2004)	MIgration in Burkina Faso and Mali	Binary and multinomial logistic regression	Food scarcity during drought leads to increased prices, forcing people to spend more money on their basic needs rather than on long-distance migration.

chiori et al., 2012). The impact of climate change on productivity and the endowment of different production factors is asymmetric, leading to asymmetrical effects on production structures and factor returns. Individuals or families with great resources, for example, substantial fixed assets, their mobility will be reduced. Conversely, the extremely poor population in developing countries also faces limited mobility, because they are least able to afford the cost of migration. Studies in Burkina Faso and Mali, where droughts occurred during the 1970s and 1980s, revealed a decrease in international, long-distance migration (Henry et al., 2004). These West African studies suggest that food scarcity during drought increases prices, forcing people to allocate more funds towards meeting basic needs rather than investing in long-distance migration (Black et al., 2013). Moreover, studies of Hurricane Katrina's impact in the USA demonstrate that a considerable number of poorer, black residents of New Orleans were unable to evacuate the affected area. Over time, it become apparent that these poorer black residents were the least able to return to their homes and communities (Landry et al., 2007). These examples show how climate change-induced market influences can have complex and differential effects on migration patterns. It always exacerbates existing inequalities and disparities within society. Figure 2 delineates the process through which climate change engenders migration, by exerting extensive impacts on both societal and individual levels. Climate change fosters motivations for migration by affecting societies and individuals in various forms; however, the realisation of migratory activities ultimately depends by bilateral immigration policies, the attraction of resources at the destination, and whether the resources at an individual's disposal are sufficient to accomplish the migration activity.

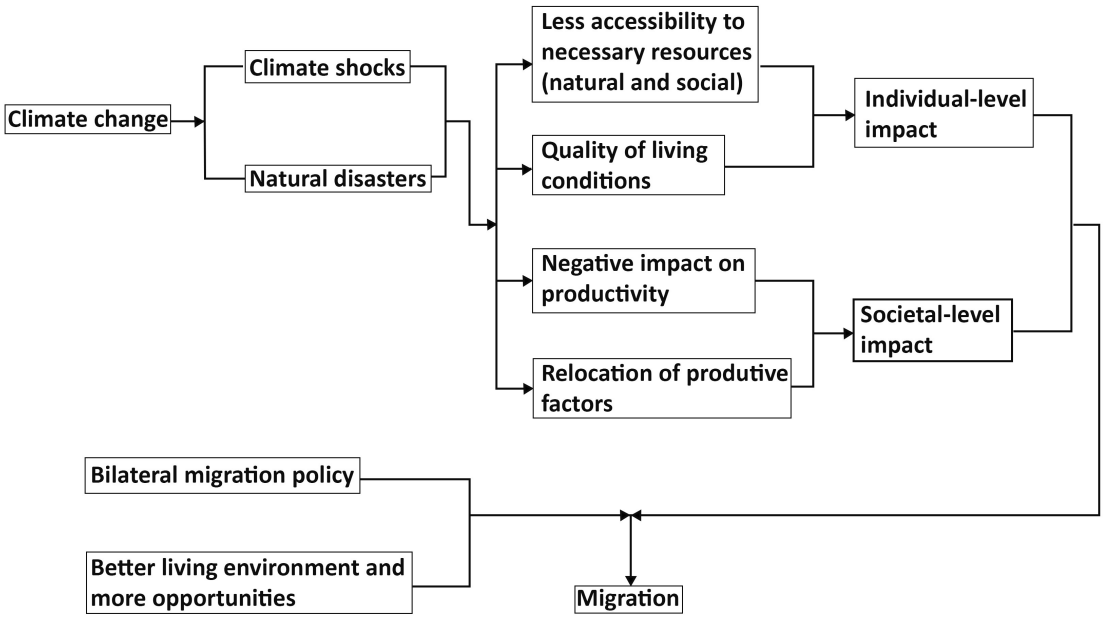

Figure 2: Process of migration in the context of climate change[2]

In most literature on migration studies, the push-pull theory remains central to explain the motivating factors behind migration. But this theory emphasises the motivation associated to social factors or economic factors. Environmental factors are rarely mentioned. The process of push and pull is typically categorised into 3 parts: (1) Origin-related factors, including political instability, conflict, high population growth rate, and lack of access to necessary resources; (2) Destination-related factors, including political stability, wage differentials, and access to necessary resources; (3) Intervention factors that either promote or restrict migration, mainly involving migration costs and migration policies (Arango, 2004; Black et al., 2013; Boyle et al., 2014; Zolberg and Benda, 2001). In fact, it is challenging to pinpoint the exact significance of each factor in the migration process, as their relevance may vary depending on the specific circumstances. However, climate variability can give the impact on all three of those components. The role of climate variability in migration is far more intricate than what we might envision. As a social phenomenon, migration simply manifests as an outcome or a result. Its operation logic is primarily displayed at the individual level, encompassing the overlapping interplay of various factors from both the natural and social realms. This multi-causal migration pattern has been widely recognised (Wolpert, 1966). It has to point out that climate variability plays a participatory in each part of this complex pattern.

Environmental factors have been considered in the classical framework of migration studies (Black et al., 2013; Henry et al., 2004; Wolpert, 1966), but they haven't held a central position for long. Only in recent years, climate change, as the fundamental cause of environmental change, has been addressed in migration research. However, Climate change introduces uncertainty into the equation. This uncertainty comes from the unpredictability of climate models' predictions, especially when considering different carbon emission scenarios, which increases the complexity of scientific investigation.

In recent years, two representative research frameworks have emerged to explain the migration activity in the background of environment variability. The first category seeks to explicitly combine and to distinguish environmental factors from environmental change factors within existing migration structures and behavioural drivers (Black et al., 2011; Meze-Hausken, 2000). Because migration is viewed both as a response to environmental change and as a contributing factor to environmental change. The second category aims to analyse migrants' coping mechanism of migrants in the face of environment variability by structuring the migration process. For example, Marchiori analyzed the impact of climate change on migration by establishing two channels (Marchiori et al., 2012). Despite these advancements, climate variability as a source of environment variability has

[2]Source: Haoxi CHEN

not been fully integrated into migration studies, which is a worthy oversight.

1.2 Chinese case

China, the country with the world's second-largest population, is also a nation with a diverse climate and complex geographical environment. It has undergone a large-scale and rapid urbanization process in recent decades, unparalleled by any other nation. Rapid urbanization and the resultant spatial disparities have created significant push-pull forces across different regions, resulting in a large and diverse migrant population within China. The country has consistently maintained a significant number of internal migrants in recent years. In 2020, the number of internal migrants was 375.88 million, which accounts for 26.62% of China's population (NHFPC, 2018). Furthermore, due to the dual urban-rural structure in China, it exists immense disparities in economic conditions and living environments between urban and rural areas. Constraints imposed by the household registration system have led to the emergence of various unique migration patterns in China (Hung 2022). The climate change and environmental stress brought about by rapid urbanization have intensified, adding a layer of specificity to migration research in China.

In recent years, research on migration in China has primarily focused on exploring the patterns of different migrant groups. Chen explored the living patterns of francophone African immigrants in Guangzhou in China (Chen et al., 2020). Liu discussed the preferences of environment-related residential of the immigrants in China (Liu et al., 2022). These studies are of course very important to the development of migration studies. But so far, we have not found a single work attempting to link climate change to internal migration in China. Climate change has not been considered and analyzed as an important pressure for migration.

This paper endeavours to understand the complexity of the relationship between climate variability and internal migration in China. It seeks to address the limitations of existing research and delve into the role of climate change in migration, within the context of China's rapid urbanisation and climate challenges. Recognising that internal migration is more sensitive to subtle differences within societies than international migration. The aim of this paper is to ascertain the role of climatic factors in the migration process and their patterns of influence.

The next section will outline the data used in this study, following by a detailed research methodology and results discussion. Finally, we will develop our findings and highlight avenues for future research in this critical area.

2 Data

We employ two primary datasets to develop this research, migration-related data and environmental and climatic data.

2.1 Migration related dataset

The migration related dataset is mainly sourced from the China Migrants Dynamic Survey (CMDS) 2017. CMDS is a significant social survey initiated by the National Health Commission (NHC) since 2009. The survey conducts a comprehensive annual nationwide sampling of population flows, covering 31 provinces, municipalities, autonomous regions and Xinjiang Production and Construction Corps within mainland China. It extensively captures migrant-related information, including personal level demographics, mobility patterns, financial particulars, habitation data, etc. For the purpose of our study, we collected data pertaining to population mobility. This data, originally focused on individual-level information, was restructured to analyze urban migration patterns. Consequently, we were able to ascertain the annual counts of migrants moving in and out of each city from 2005 to 2017. Furthermore, provincial statistical yearbooks are also employed to enrich this data.

2.2 Environmental and climatic dataset

The environmental and climatic dataset comprises three distinct components: climate data, natural disaster data and pollution data.

The climate data, including temperature and precipitation, is sourced from the Copernicus Climate Change Service[3]. Natural disaster data is collected from The International Disaster Database (EM-DAT) established by the Centre for Research on the Epidemiology of Disasters. EM-DAT achieves the occurrence and impacts of over 26,000 mass disasters worldwide from 1900 to the present day, compiled from various sources, including UN agencies, non-governmental organizations, reinsurance companies, research institutes, and press agencies.

Then, the pollution data segregated into two parts: per capita carbon emission data and city level PM2.5 data. Per capita carbon emission data is sourced from the China Energy Statistical Yearbook, which include the carbon emission data from over 300 Chinese cities, with the exception of Tibet's data. Concurrently, the PM2.5 data originates from the Atmospheric Composition Analysis Group at Dalhousie University, Canada. It combines global models, satellite observations, and air quality monitor data to develop estimates of on-the-ground PM2.5 levels.

It is important to mention that, owing to the unavailability of pertinent data related to Tibet, this region has been excluded from the study. We have organized all the raw data and conducted a data pivot centred around cities from 2005 to 2017, resulting in a new dataset for analysis. The specifics of this data are outlined in the table below.

[3]Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), (date of access), https://cds.climate.copernicus.eu/cdsapp/home

Table 2: Summary variables

Variable	Full name	Max	Min	Mean
Climate related variables				
Cata	Total number of natural disasters of the province for current year	0	17	3.493
Cata1	Total number of natural disasters of the province for the previous years	0	17	3.222
Cata5	Total number of natural disasters of the province for the previous 5 years	0	51	12.56
Prec	Annual precipitation of the city for current year	26.85	2743.3	971.57
Prec1	Annual precipitation of the city for the previous year	26.85	2743.3	967.37
Prec5	Mean annual precipitation of the city for the previous 5 years	46.34	2231.44	943.02
Temp	Annual average temperature of the city for current year	-3.06	26.38	14.03
Temp1	Annual average temperature of the city for the previous year	-3.056	26.376	13.983
Temp5	Mean annual average temperature of the city for the previous 5 years	-2.54	26.04	13.934
Geographical label				
City	City code	-	-	-
Province	Province code	-	-	-
Year	Year	2005	2017	-
Migration data				
Inflow	Total number of immigrants of the city for current years	1	1057	35.74
Outflow	Total number of emigrants of the city for current year	0	1143	34.62
Polluction related variables				
Carbon	Carbon emission per capita of the city for current year	2.288	36.232	9.119
Carbon1	Carbon emission per capita of the city for the previous year	1.903	32.464	8.717
Carbon5	Mean carbon emission per capita of the city for the previous 5 years	1.548	31.918	7.757
PM	Annual average PM2.5 concentration of the city for current year	2.057	108.526	44.961
PM1	Annual average PM2.5 concentration of the city for the previous year	2.057	108.526	45.4
PM5	Mean annual average PM2.5 concentration of the city for the previous 5 years	2.295	96.501	45.187
Socio-economic related variables				
Gdp	GDP per capita of the city for current year(CYN)	2755	215488	37365
Gdp1	GDP per capita of the city for the previous year(CYN)	2394	215488	33992
Gdp5	Mean GDP per capita of the city for the previous 5 years (CYN)	1957	200442	27527

3 Methodology

Two approaches have been performed to access the influence of climate change on migration patterns: K-means clustering analysis based on principal component analysis and feature importance analysis based on machine-learning model[4].

3.1 K-means clustering based on principal component analysis

Migration research typically involves substantial amounts of data and intricate variable relationships; therefore, principal component analysis is utilized to consolidate multiple variables into a few principal components. This approach enables the identification of the main characteristics and behavioural patterns of migrant groups through component loadings of the principal components, leading to a more accurate understanding of their actions, preferences, and needs. K-means clustering can further subdivide migrant populations, identify groups with similar characteristics and behaviours, and reveal underlying patterns and trends in migration behaviours against the backdrop of climate change. The identification, prediction, and analysis of migration behaviour patterns under the context of climate change are also the core content of this study.

The determination of the optimal number of clusters is a crucial step in k-means cluster analysis. This article employs two commonly used methods to ascertain the optimal number of clusters: the Elbow Method and the Silhouette Coefficient Method. The Elbow Method involves assessing the impact of the number of clusters (k) on the quality of clustering and identifying the point where increasing the number of clusters no longer significantly alters the sum of squared errors, thereby determining the optimal number of clusters. Then we use the Silhouette Coefficient to validate the results obtained from the Elbow Method. The Silhouette Coefficient also evaluates the reasonableness of the assignment of data points to each cluster by combining measures of intra-cluster similarity and inter-cluster dissimilarity. Its values range from -1 to 1, with values closer to 1 indicating more effective clustering.

Details on the specific steps of Principal Component Analysis and K-means clustering can be found in the Appendix.

According to the Elbow Method, the optimal number of clusters is identified as 4. However, the Silhouette Coefficient suggests that k=2, 3, or 4 could all be considered as optimal cluster numbers. Given that clustering into 2 or 3 groups does not provide sufficiently precise subdivisions for effective analysis, we opt to use k=4 in the subsequent chapters.

3.2 XGBoost-Shapley additive explanation machine-learning model

Migration behaviour is a multifaceted and complex system. Many researchers have employed the gravity model to assess the influence of various factors on migration behaviours, implying a general linearity between migration behaviour and its influencing factors. Nonetheless, the true nature of migration behaviour encompasses a complex interplay of linearity and nonlinearity, imbuing it with a multitude of characteristics. In response to the needs of migration research, there is a necessity to find a method that can not only determine the importance of features but also analyze the linear and nonlinear relationships between features and outcomes. This approach can more accurately simulate the genesis of migration decisions and behaviours. XGBoost, as an efficient machine learning algorithm, has demonstrated exceptional performance in regression and classification problems, making it an indispensable tool for this study.

XGBoost, proposed by Chen(2016), is a form of the gradient boosting decision tree algorithm, regarded as one of the most efficient algorithms in supervised learning. It is an optimized distributed gradient boosting library, comprising an ensemble model of numerous decision trees, each striving to correct the errors of the preceding one. XGBoost counteracts overfitting by incorporating

[4]In this study, the K-means cluster based on principal component analysis is implemented using "cluster" package in R and utilizes QGIS for visualization. The feature importance analysis is carried out using sklearn library, shap library and xgboost library in Python , with analysis facilitated by tools from matplotlib library.

regularization, the specific approach is as follows:

$$obj(\theta) = L(\theta) + \Omega(\theta) \tag{1}$$

$L(\theta)$ represents a loss function, and $\Omega(\theta)$ is a regularization function that helps to avoid overfitting in the model. It can be expressed as follows:

$$\Omega(f_x) = \gamma T + \frac{1}{2}\lambda \|W\|^2 \tag{2}$$

,where T denotes the number of leaf nodes in the decision tree and W represents the node's weights.

Since each city encompasses both aspects of immigration (inflow) and emigration (outflow), they are modeled separately. Subsequently, all observations are randomly divided into a training set and a validation set, with proportions of 75% and 25%, respectively. To evaluate the performance of the XGBoost models, the Mean Squared Error (MSE) metric will be used for measurement.

$$MSE = \frac{1}{n}\sum_{q=1}^{n}(y_q - \hat{y_q})^2 \tag{3}$$

,where n represents the total count of data points, y_q refers to the actual observation value of the q^{th} data point in the training set, and $\hat{y}_q$ denotes the observation value in the validation set.

Although machine learning models can offer superior accuracy, their results are generally more challenging to interpret compared to statistical models(Louhichi et al., 2023). To better elucidate the results of the XGBoost model, we have incorporated Shapley values to explain the relative importance of each feature in individual predictions. Shapley values, based on game theory, are a technique that has seen widespread application in recent years for interpreting results in machine learning(Merrick & Taly, 2020). This approach provides a quantifiable measure of each feature's contribution to the prediction, thereby enhancing the interpretability of complex machine learning models like XGBoost.

To calculate the Shapley value for feature i, one initially generates all possible coalitions S of n features excluding feature i. The "marginal contribution" of feature i is ascertained by calculating the difference between the outcomes of the feature function v for the feature set N (the complete set of features) and the set S. To ensure fairness, this calculation process is replicated across all possible feature combinations. By aggregating and averaging all the marginal contributions, a fair measure of feature i's contribution to the model's prediction is obtained, which is the Shapley value for feature i. This entire process can be articulated through the following equation:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!}(v(S \cup \{x_i\}) - v(S)) \tag{4}$$

In this equation, N is the complete set of features, S denotes all possible feature combinations that exclude feature i, v is the value function of the feature set, x_i is the value of feature i, and n is the total number of features.

An XGBoost model that utilizes Shapley values can provide insights into the relative contributions of different features in the processes of immigration inflow and emigration outflow. It also allows for the delineation of both linear and nonlinear relationships between various features and the output, making the interpretation of machine learning models more accessible and comprehensible.

4 Empirical analysis

4.1 Exploratory data analysis

Initially, we mapped the raw migration data onto a map for different years, representing cities with points and migration routes with lines. The brightness of the lines indicates the density of the migration flow: the brighter the line, the more migrants it represents. It's clearly observable that the overall scale of migration is rapidly increasing, especially in Southeastern China where the scale of migration flows is particularly massive. This side-by-side comparison reflects an undeniable reality: the significant increase in the number of migrants is concurrent with rapid economic development.

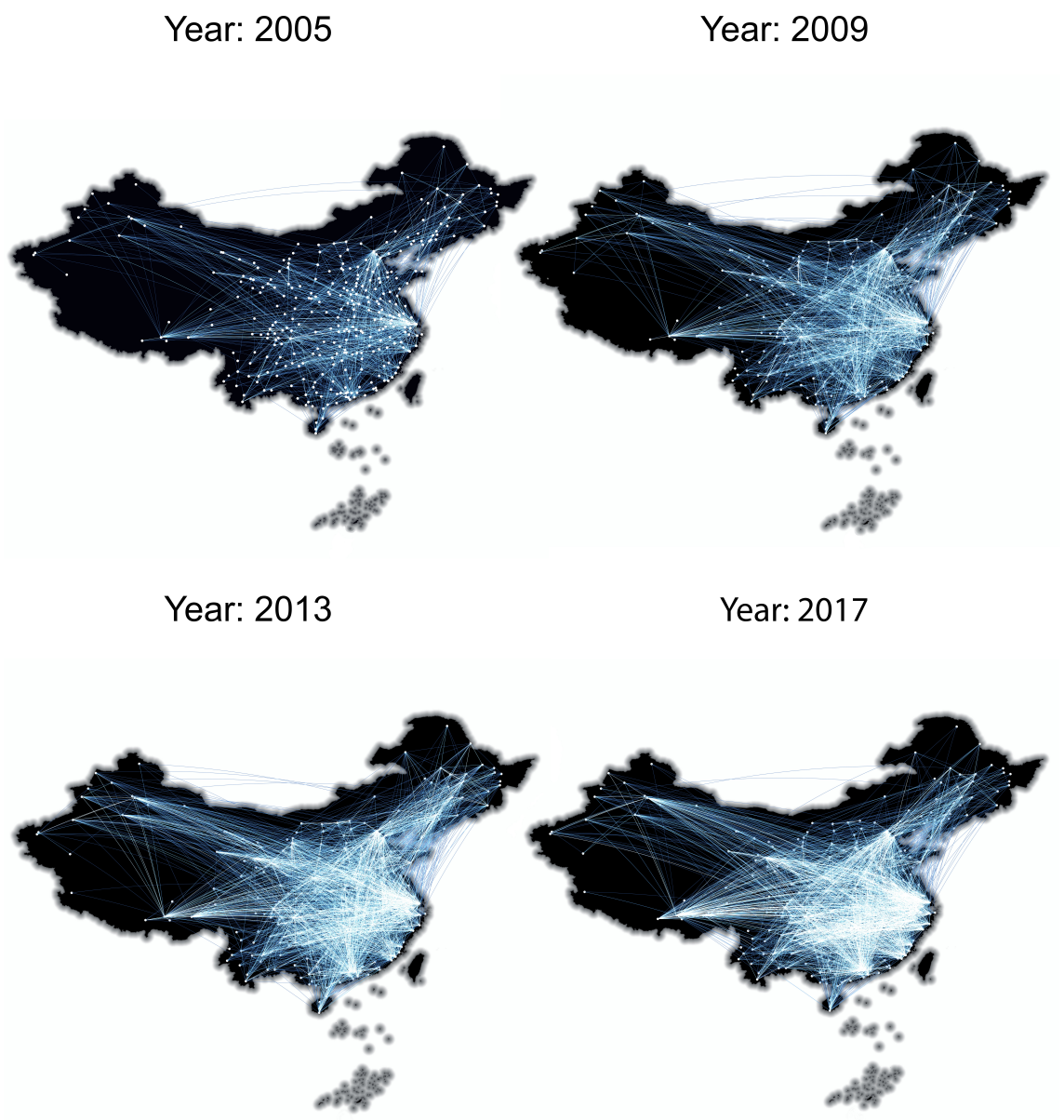

Figure 3: Internal migration pathway

Then we integrated data spanning multiple years and created figure 4, depicting the scale of migration inflows and outflows. It shows that, apart from regions with unique geographical features

like the Qinghai-Tibet Plateau, mountainous areas, and deserts, most areas in China have experienced both immigration and emigration. The data also identifies that the Central and Western regions of China serve as sources of emigration, while the Eastern and Northern coastal regions have become major destinations for immigrants. This pattern underscores the significant internal migration trends within China.

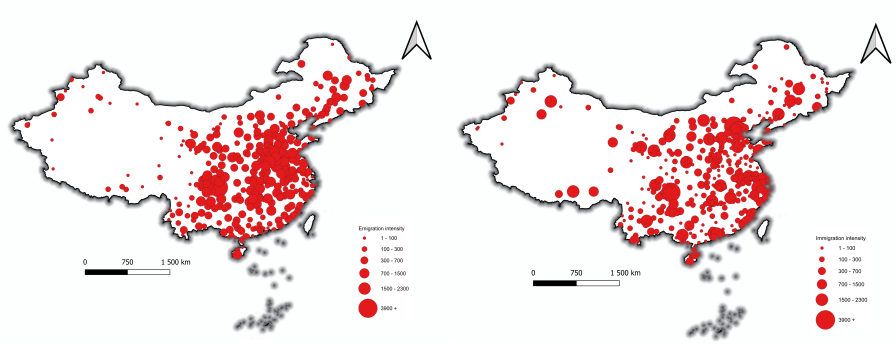

Figure 4: Cumulative immigration and emigration intensity in China from 2005 to 2017

4.2 K-means clustering analysis

After obtaining the results of the principal component analysis, the first principal component and the second principal component have been selected as the new indicators for migration impact. The first principal component encompasses variables related to temperature, precipitation, disasters, and partially to carbon emissions. The second principal component mainly includes variables related to the economy, carbon emissions, and pollution. The specific loadings for the first and second principal components are shown in the table below.

Table 3: Main factors' loading values of the 1st and 2nd principal component

Variables	**PC1**	**PC2**
Mean a. precipitation of the city for the previous 5 years	0.3261	0.0546
A. a. temperature of the city for the previous year	0.3228	0.0657
A. a. temperature of the city for current year	0.3227	0.0694
Mean a. a. temperature of the city for the previous 5 years	0.3226	0.0607
A. precipitation of the city for the previous year	0.3153	0.0735
A. precipitation of the city for current year	0.3132	0.0690
Carbon emission per capita of the city for current year	-0.2745	0.2105
Carbon emission per capita of the city for the previous year	-0.2699	0.2267
Total number of natural disasters of the province for the previous 5 years	0.2671	0.1198
Mean carbon emission per capita of the city for the previous 5 years	-0.2558	0.2555
Total number of natural disasters of the province for the previous years	0.2230	0.1457
Total number of natural disasters of the province for current year	0.2161	0.1345
GDP per capita of the city for the previous year	-0.0265	0.5024
Mean GDP per capita of the city for the previous 5 years	-0.0245	0.5019
GDP per capita of the city for current year	-0.0224	0.4961
Mean a. a. PM2.5 concentration of the city for the previous 5 years	0.01790	0.0816
A. a. PM2.5 concentration of the city for the previous year	0.0069	0.0436
A. a. PM2.5 concentration of the city for current year	-0.0014	0.0184

After that, based on the results of the principal component analysis, clustering is performed for each city. Figure 5 illustrates the clustering results projected onto a map.

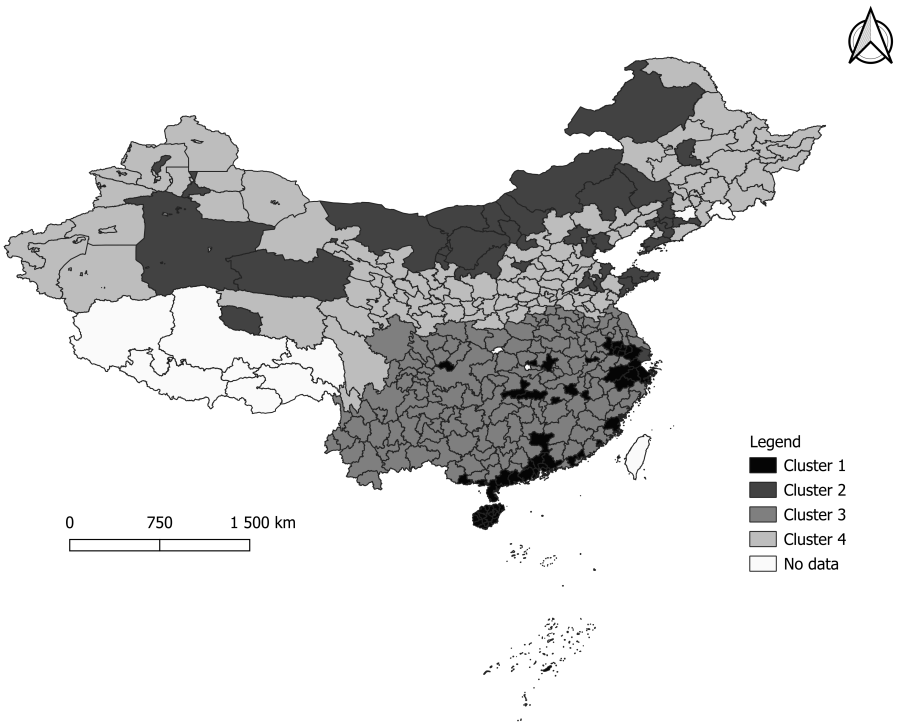

Figure 5: Cluster segmentation for k=4

According to the cluster analysis results, cities in China can be distinctly categorized into four major groups. The first cluster includes cities characterized by higher temperatures, greater precipitation, and more frequent natural disasters. These cities also tend to have higher economic levels and relatively higher per capita carbon emissions and pollution. They are commonly viewed as the most economically vibrant, wealthiest, and most urbanized regions in China. The second cluster mainly consists of cities in northern China, featuring lower temperatures, greater precipitation, and fewer natural disasters, but still maintaining high economic development levels and relatively high per capita carbon emissions and pollution. These include cities with high economic development, such as Beijing, Tianjin, and some cities in Shandong province, or those rich in mineral resources, like cities in Inner Mongolia autonomous region, Xinjiang uygur autonomous region, and Qinghai province. The third cluster encompasses cities with higher temperatures, more precipitation, and frequent natural disasters, but with relatively underdeveloped economies and lower per capita carbon emissions and pollution. These cities are widely spread across the inland areas of southern China. Finally, the fourth cluster includes cities located in the economically less developed and mineral-resource-scarce areas of northern China. These cities are characterized by lower temperatures, less precipitation, fewer natural disasters, and comparatively lower economic levels and per capita carbon emissions and pollution.

Table 4: Composition of 4 clusters

	Cluster 1		**Cluster 2**		**Cluster 3**		**Cluster 4**	
Cluster center	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
	3.312107	1.736	-3.495	2.788	1.688	-0.884	-2.378	-0.943
Geographical distribution	The cities in the eastern coastal region, as well as the capital cities of some southern provinces are included in this context.		Some cities in northern China are included, comprising Beijing, Tianjin, the majority of cities in Inner Mongolia autonomous region, as well as certain cities in Shandong province, Liaoning province, Xinjiang uygur antonomous region, and Qinghai province.		Most cities in southern region of China.		Most cities in northern region of China.	

Table 5: Migration intensity distribution of 4 clusters

Yearly migration numbers	0-30		30-60		60-90		90-120		120-150		150+	
	Outflow	Inflow	Outflow	Inflow	Outflow	Inflow	Outflow	Inflow	Outflow	Inflow	Outflow	Inflow
Cluster 1	80.042%	44.374%	11.253%	16.348%	4.883%	9.766%	1.911%	9.979%	1.0616%	5.308%	0.849%	14.225%
Cluster 2	78.926%	51.033%	14.669%	13.017%	2.893%	9.917%	1.860%	4.959%	1.240%	4.545%	0.413%	16.529%
Cluster 3	49.325%	81.104%	17.485%	10.736%	12.331%	3.553%	5.031%	1.472%	2.454%	0.736%	3.374%	2.393%
Cluster 4	59.329%	83.841%	25.917%	8.977%	8.821%	2.571%	3.513%	1.015%	1.717%	0.781%	0.703%	2.810%

Tables 5 present a detailed distribution of the inflow and outflow of migrants across each identified cluster. It is crucial to highlight that yearly migrations involving 0-30 individuals are not directly associated to the environmental and climate factors under examination in this paper. Such migrations are often influenced by personal choices, for example, individual career moves, emotional transitions, or family events, which do not have a direct correlation with the climatic, economic, or environmental elements central to this study. Consequently, these random migration fall outside the purview of the analysis.

The cities within the first cluster predominantly experience annual out-migration numbers ranging between 30-90 people, with the remainder accounting for about 3.82%. The inflow of migrants is significantly distributed across each segment, with the proportion being the largest among all clusters. This indicates that these cities are net inflow regions for migrants. In cities of the second cluster, yearly out-migration falls within the 30-90 range, with other segments constituting 3.52%. The inflow is well-distributed across each range, though slightly less than the first cluster, suggesting these cities also serve as net migrant inflow areas.

Cities in the third cluster exhibit a wide distribution of out-migration across all segments, while in-migration is mainly concentrated in the 30-90 range, with other ranges including only 4.60%. These cities are characterized by a net outflow of population. Finally, cities in the fourth cluster mainly see in-migration numbers between 30-120, and out-migration numbers predominantly in the 30-60 range. These cities constitute the least active segment in terms of migration, exhibiting minimal levels of both emigration and immigration.

4.3 XGboost-Shaply model analysis

For each city, considering two migration variables – inflow and outflow – we have developed an XGBoost model for in-migration (inflow XGBoost model) and another for out-migration (outflow XGBoost model). These models are utilized to analyze how different features influence both the inflow and outflow of migration in the same city.

Table 6: Variable contribution in emigration

Variable	**Mean Abs Shapley Value**	**Contribution**	**Rank**
PM5	0.280	18.100%	1
Carbon5	0.170	10.985%	2
Cata	0.166	10.719%	3
Temp5	0.118	7.642%	4
Prec5	0.094	6.087%	5
Cata1	0.087	5.604%	6
Gdp5	0.084	5.449%	7
Gdp	0.0823	5.314%	8
Carbon	0.075	4.816%	9
Cata5	0.056	3.626%	10
Temp	0.054	3.482%	11
PM	0.051	3.288%	12
Prec	0.047	3.062%	13
Prec1	0.042	2.686%	14
PM1	0.039	2.499%	15
Gdp1	0.035	2.275%	16
Temp1	0.034	2.193%	17
Carbon1	0.037	2.172%	18

MSE: 0.7172

Table 7: Variable contribution in immigration

Variable	Mean Abs SHAP Value	Contribution	Rank
Gdp	0.225	17.872%	1
Gdp1	0.153	12.147%	2
PM5	0.092	7.309%	3
Temp	0.089	7.094%	4
Prec5	0.083	6.608%	5
Gdp5	0.078	6.176%	6
Cata5	0.076	6.058%	7
Temp5	0.072	5.709%	8
PM1	0.055	4.329%	9
Cata	0.051	4.048%	10
Carbon	0.051	4.035%	11
Carbon1	0.040	3.188%	12
Temp1	0.039	3.118%	13
Prec	0.037	2.906%	14
PM	0.035	2.806%	15
Carbon5	0.033	2.651%	16
Prec1	0.0256	2.027%	17
Cata1	0.024	1.919%	18

MSE: 0.5681

Tables 6 and 7 display the mean absolute Shapley values for each feature, as well as their percentage contribution for both inflow model and outflow model. As for outflow model, the top five variables most significantly impacting the out-migration number are: mean annual average PM2.5 concentration of the city for the previous 5 years, mean carbon emission per capita of the city for the previous 5 years, total number of natural disasters in the province for current year, mean annual average temperature of the city for the previous 5 years, and mean annual precipitation of the city for the previous 5 years. Overall, climate-related variables contribute 45.08% to the migration outflow, ranking first. Pollution-related variables contribute a total of 41.882%, ranking second, and economic factors account for 13.038%.

According to inflow model, the five most influential variables on the in-migration number are: GDP per capita of the city for current year, GDP per capita of the city for the previous year, mean annual average PM2.5 concentration of the city for the previous 5 years, annual average temperature of the city for current year, and mean annual precipitation of the city for the previous 5 years. The total contribution of these factors exceeds 50%. Overall, climate-related variables have the highest contribution to migration inflow, accounting for 39.487%. Economic factors follow closely, constituting 36.195%, while pollution-related variables contribute 24.318%.

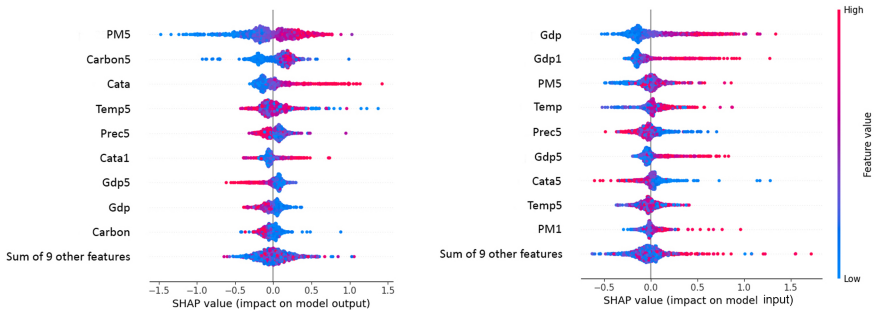

Figure 6: Beeswarm plots of outflow model and inflow model

For both the inflow model and the outflow model, beeswarm plots have been created. These plots visually correlate the Shapley values of each feature with the output values of the inflow and out-

flow models, demonstrating how each feature influences the model results. The positive or negative Shapley values indicate the extent to which a feature positively or negatively impacts the output for an individual instance. This visualization effectively highlights the relative importance and directional influence of each feature within the models.

In the outflow model, the mean annual average PM2.5 concentration of the city for the previous 5 years, the total number of natural disasters in the province for current year, and the total number of natural disasters in the province for the previous year show a clear positive correlation with their impact in emigration number. In contrast, the mean GDP per capita of the city for the previous 5 years and the GDP per capita of the city for current year exhibit a significant negative correlation with their influence in out-flow migration. Meanwhile, the mean carbon emission per capita of the city for the previous 5 years, the mean annual average temperature of the city for the previous 5 years, the mean annual precipitation of the city for the previous 5 years, the carbon emission per capita of the city for current year, and the carbon emission per capita of the city for the previous year demonstrate a nonlinear relationship with the out-migration numbers.

In the inflow model, the economic variables - GDP per capita of the city for current year, GDP per capita of the city for the previous year, and Mean GDP per capita of the city for the previous 5 years - all show a positive correlation with their impact in the number of migrants moving in. Conversely, the total number of natural disasters in the province for the previous 5 years exhibits a significant negative correlation with its influence in the number of immigrants. Meanwhile, the mean annual average PM2.5 concentration of the city for the previous 5 years, the annual average temperature of the city for current year, the mean annual precipitation of the city for the previous 5 years, the mean annual average temperature of the city for the previous 5 years, and the annual average PM2.5 concentration of the city for the previous year demonstrate a nonlinear relationship with the in-migration numbers.

Both the outflow and inflow models exhibit numerous features that have significant nonlinear impacts on the output results. Consequently, we have created Shapley dependence plots for the nine most crucial features to observe in greater detail the relationship between these features and the in/out migration flows. A Shapley dependence plot delineates the feature value on the x-axis and the corresponding Shapley value on the y-axis. These plots offer more comprehensive insights compared to merely examining the importance contributions. They provide a nuanced understanding of how each feature value influences the prediction, revealing the complexity and nature of their relationships with migration patterns.

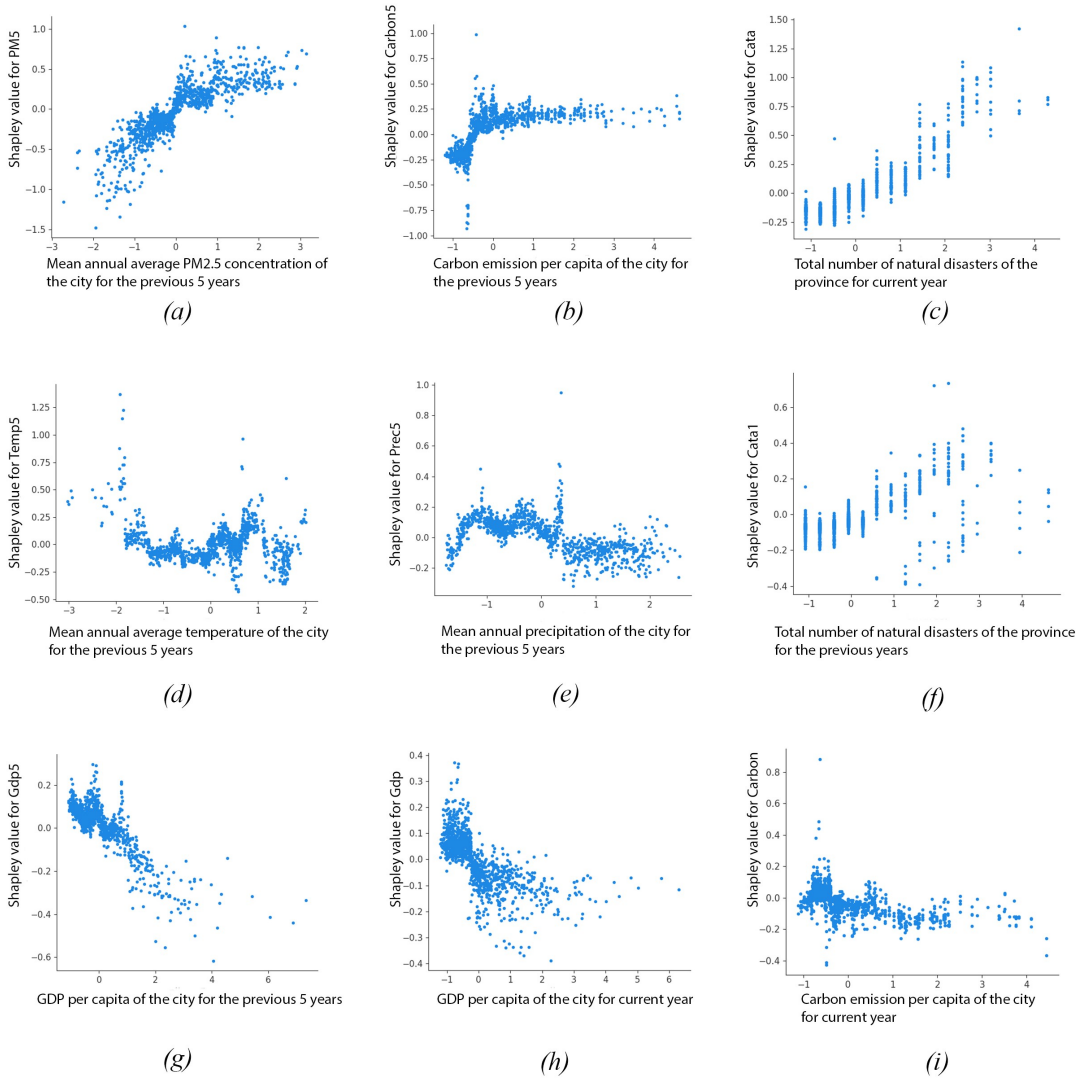

Figure 7: Shapley dependence plots of outflow model

In figure 7, apart from the variables for which a linear relationship can be directly inferred, many features exhibit nonlinear relationships with their impact in the output results. Figure 7-b illustrates a distinct logarithmic function relationship between the mean carbon emission per capita of the city for the previous 5 years and its influence in the emigration volume. As this feature increases, there is an initial rapid rise in its impact on out flow migration numbers, followed by a plateau at a stable level. The mean annual average temperature of the city for the previous 5 years, the mean annual precipitation of the city for the previous 5 years, and the carbon emission per capita of the city for current year exhibit a cyclical function relationship with their influence in emigration, characterized by multiple peaks. This pattern indicates a complex interplay of climatic and environmental factors with emigration volume, where the influence of these variables on emigration is not linear but fluctuates, reflecting various underlying dynamics.

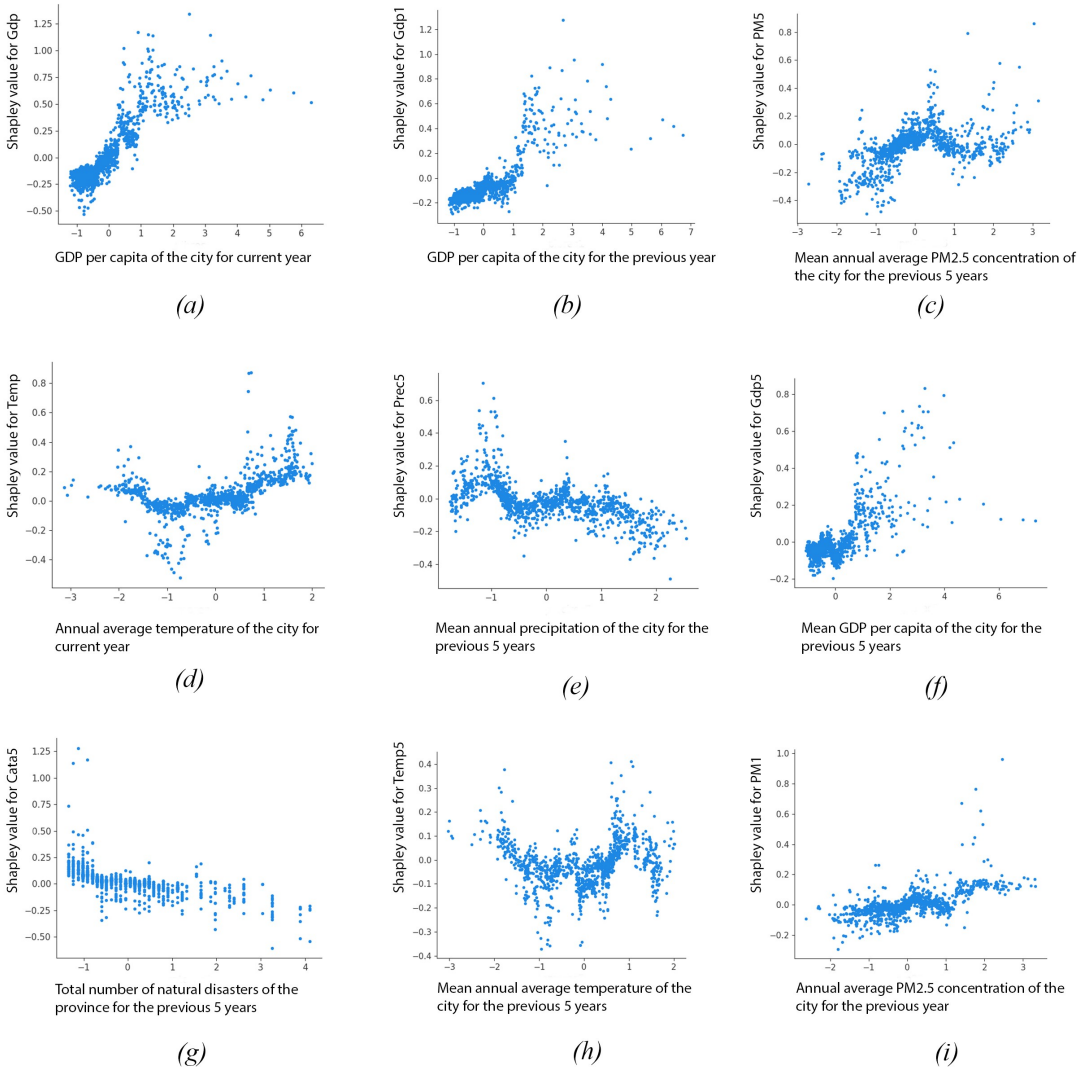

Figure 8: Shapley dependence plots of outflow model

For the inflow model, the mean annual average PM2.5 concentration of the city for the previous 5 years, the annual average temperature of the city for current year, and the annual average PM2.5 concentration of the city for the previous year exhibit a polynomial function relationship with their impact in the number of immigrants. As the mean annual average PM2.5 concentration of the city for the previous 5 years increases, its impact on migration initially surges, then sharply declines, after that gradually recovering. The relationship between the annual average PM2.5 concentration of the city for the previous year and its impact in the in-migration number demonstrates a similar trend, though it is relatively more gradual. With the increase in the annual average temperature of the city for current year, its influence on the number of incoming migrants initially decreases and then gradually increases. The mean annual precipitation of the city for the previous 5 years and the mean annual average temperature of the city for the previous 5 years also display a cyclical function relationship with their impact in in-migrant volume, characterized by multiple peaks. These patterns suggest a complex, non-linear interplay between climatic-environmental factors and migrant numbers, with varying influences at different levels of these variables.

5 Discussion

Migration, employed as an adaptive strategy to respond to challenges in regions vulnerable to climate and environment change, has exhibited various patterns and intensities across China. Significant distinctiveness is observable both in the apparent migration trends and in the underlying

drivers propelling these movements. These unique characteristics not only reflect the geographical diversity and social complexity of China but also unveil the creativity and resilience of humans in adapting to climatic and environmental changes.

Initially, cluster analysis categorized Chinese cities into four principal categories: The first category includes cities characterized by hot climates, abundant rainfall, frequent natural disasters, more developed economies, and higher levels of per capita carbon emissions and pollution. The second category consists of cities with cold climates, scarce precipitation, fewer natural disasters, developed economies, and similarly higher levels of per capita carbon emissions and pollution. The third category is composed cities with hot climates, plentiful rainfall, numerous natural disasters, but relatively underdeveloped economies and lower levels of per capita carbon emissions and pollution. Lastly, the fourth category of cities typically experience cold climates, less rainfall, fewer natural disasters, lower economic levels, and comparatively lower per capita carbon emissions and pollution levels.

The first and second categories of cities are definitive net population inflow areas, exhibiting much higher in-migration intensities compared to the third and fourth categories, yet their out-migration intensities are significantly lower. With China's rapid urbanization, economic factors often become the primary consideration for migrants when the number of migrants increases swiftly. While the first and second categories of cities are similar in economic levels and carbon emissions, the number of migrants moving out of the first category is slightly higher than that of the second. Similarly, among the third and fourth categories, which have comparable economic levels, the third category exhibits a higher out-migration intensity compared to the fourth. This indirectly indicates that when economic levels are similar, climatic factors such as temperature and precipitation do indeed influence migration patterns.

As definitive net outflow migration cities, the third and fourth categories reveal that even without the impacts of climate change and natural disasters, people in economically disadvantaged areas tend to actively seek better living conditions and development opportunities. However, climate change and environmental pressures are not without effect on these cities. A notable observation is that the out-migration intensity in each quantity segment for the third category cities is consistently higher than that of the fourth category, yet the in-migration intensity is similar between the two. This indicates that Climate and environmental pressures notably affect the third category of cities in southern China, characterized by higher temperatures, increased rainfall, and frequent natural disasters.

As for cities, the impact of climate change and environmental pressures is significant. According to the result of XGBoost outflow model, among the top five most influential factors, three are related to climate. Climate-related variables comprise a major proportion, indicating that climate change indeed affects population migration to a certain extent. The impact of climate change may also indirectly influence emigration through socio-economic factors.

It should be noted that in the case of China, the average PM2.5 levels and per capita carbon emissions over the previous five years have provided significant contributions to the outflow model's output. The pollution resulting from rapid urbanization and industrialization has necessary implications for health, quality of life, and related economic losses. Air pollution, linked to various health issues such as respiratory diseases, strokes, and cancer, directly increases individuals' medical and living costs(Lee et al., 2018; Turner et al., 2020). On a micro level, individuals exposed to polluted environments over the long term may incur higher health expenses. Moreover, areas with high pollution levels are less likely to attract business startups or investments due to higher environmental compliance costs and risks, potentially leading to a reduction in job opportunities(Khan & Ozturk, 2020; Wang et al., 2021). When the cost of living for individuals rises sharply, but their economic benefits do not significantly increase, the process of migration is likely to be triggered rapidly.

According to the results of the XGBoost inflow model, economic factors hold absolute importance among the top five most significant factors. At the individual level, the pursuit of economic benefits is a important motivator for domestic migration within China. The positive correlation between

per capita GDP and in-migration volume clearly indicates that higher economic levels make cities more attractive to migrants. However, climate and pollution factors also play significant roles in this process. The substantial contribution of climate-related variables to in-migration numbers underscores the importance of climate factors in the migration process. Compared to the high contribution of pollution variables in the outflow model, their contribution in the inflow model is significantly lower. On one hand, economic benefits sufficient to offset the personal costs of pollution can facilitate effective migration. On the other hand, cities with good economic conditions indeed have greater capacity for environmental management and need to reduce the pressures of climate change and environmental factors to attract more investment and to create more job opportunities.

6 Conclusion

In the study of the drivers of internal population migration in China, the significant role of climate change must be emphasized. It directly influences migration decisions through increasingly frequent natural disasters and exhibits a clear linear relationship with migration numbers. Climate factors such as precipitation and temperature further impact migration flows through cyclical or polynomial patterns. Additionally, climate change shapes migration trends indirectly by influencing the socio-economic structure, a phenomenon that is particularly evident in the dynamics of migration within China.

In summary, the following conclusions can be drawn:

- Climate change is a key factor influencing current and future population migration in China, affecting migration patterns and dynamics through both direct and indirect pathways.

- While economic factors remain the main drivers of population migration, environmental pressures related to climate change plays a significant role.

- An increase in pollution levels is associated with a rise in net out-migration, highlighting the importance of environmental quality in a city's attractiveness.

- Economic incentives and environmental quality together shape migration flows between Chinese cities. As environmental governance capabilities improve, economically developed cities might better balance these factors to attract more migrants.

In formulating future planning and migration policies, these findings emphasize the importance of considering the impacts of climate change and maintaining a balance between economic growth and environmental quality. Ultimately, this will contribute to creating more resilient and adaptive environments, better equipped to handle the challenges posed by climate change.

References

[1] Adger, W.N., Kelly, P.M., Winkels, A., Huy, L.Q., Locke, C., 2002. Migration, Remittances, Livelihood Trajectories, and Social Resilience. AMBIO J. Hum. Environ. 31, 358–366. https://doi.org/10.1579/0044-7447-31.4.358

[2] Arango, J., 2004. Theories of International Migration, in: International Migration in the New Millennium. Routledge.

[3] Beine, M., Bertoli, S., Fernandes-Huertas Moraga, J., 2014. A practitioners' guide to gravity models of international migration. FERDI Policy Brief.

[4] Black, R., Adger, W.N., Arnell, N.W., Dercon, S., Geddes, A., Thomas, D., 2011. The effect of environmental change on human migration. Glob. Environ. Change, Migration and Global Environmental Change – Review of Drivers of Migration 21, S3–S11. https://doi.org/10.1016/j.gloenvcha.2011.10.001

[5] Black, R., Kniveton, D., Schmidt-Verkerk, K., 2013. Migration and Climate Change: Toward an Integrated Assessment of Sensitivity, in: Faist, T., Schade, J. (Eds.), Disentangling Migration and Climate Change: Methodologies, Political Discourses and Human Rights. Springer Netherlands, Dordrecht, pp. 29–53. https://doi.org/10.1007/978-94-007-6208-4_2

[6] Boyle, P., Halfacree, K., Robinson, V., 2014. Exploring Contemporary Migration. Routledge.

[7] Bylander, M., 2015. Depending on the Sky: Environmental Distress, Migration, and Coping in Rural Cambodia. Int. Migr. 53, 135–147. https://doi.org/10.1111/imig.12087

[8] Chen, H., Yuan, Z., Cai, X., 2020. The daily life and place identity of the Francophone population in Guangzhou. Sci. Geogr. Sin. 40, 2027–2036. https://doi.org/10.13249/j.cnki.sgs.2020.12.009

[9] Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[10] Coniglio, N.D., Pesce, G., 2015. Climate variability and international migration: an empirical analysis. Environ. Dev. Econ. 20, 434–468.

[11] Dell, M., Jones, B.F., Olken, B.A., 2014. What Do We Learn from the Weather? The New Climate-Economy Literature. J. Econ. Lit. 52, 740–798. https://doi.org/10.1257/jel.52.3.740

[12] Deschênes, O., Greenstone, M., 2007. The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather. Am. Econ. Rev. 97, 354–385. https://doi.org/10.1257/aer.97.1.354

[13] Gemenne, F., 2011a. Why the numbers don't add up: A review of estimates and predictions of people displaced by environmental changes. Glob. Environ. Change, Migration and Global Environmental Change – Review of Drivers of Migration 21, S41–S49. https://doi.org/10.1016/j.gloenvcha.2011.09.005

[14] Gemenne, F., 2011b. Climate-induced population displacements in a 4°C+ world. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 369, 182–195.
https://doi.org/10.1098/rsta.2010.0287

[15] Halliday, T., 2006. Migration, Risk, and Liquidity Constraints in El Salvador. Econ. Dev. Cult. Change 54, 893–925. https://doi.org/10.1086/503584

[16] Henry, S., Schoumaker, B., Beauchemin, C., 2004. The Impact of Rainfall on the First Out-Migration: A Multi-level Event-History Analysis in Burkina Faso. Popul. Environ. 25, 423–460.
https://doi.org/10.1023/B:POEN.0000036928.17696.e8

[17] Hung, Jason. 2022. 'Hukou System Influencing the Structural, Institutional Inequalities in China: The Multifaceted Disadvantages Rural Hukou Holders Face'. Social Sciences 11(5): 194.
https://doi.org10.3390/socsci11050194

[18] Hunter, L.M., Luna, J.K., Norton, R.M., 2015. Environmental Dimensions of Migration. Annu. Rev. Sociol. 41, 377–397.
https://doi.org/10.1146/annurev-soc-073014-112223

[19] IPCC, 2007. Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC. Cambridge University Press.

[20] Khan, M. A., Ozturk, I. (2020). Examining foreign direct investment and environmental pollution linkage in Asia. Environmental Science and Pollution Research, 27(7), 7244–7255. https://doi.org/10.1007/s11356-019-07387-x

[21] Koubi, V., Spilker, G., Schaffer, L., Bernauer, T., 2016. Environmental Stressors and Migration: Evidence from Vietnam. World Dev. 79, 197–210. https://doi.org/10.1016/j.worlddev.2015.11.016

[22] L. Perch-Nielsen, S., B. Bättig, M., Imboden, D., 2008. Exploring the link between climate change and migration. Clim. Change 91, 375–393.
https://doi.org/10.1007/s10584-008-9416-y

[23] Landry, C.E., Bin, O., Hindsley, P., Whitehead, J.C., Wilson, K., 2007. Going Home: Evacuation-Migration Decisions of Hurricane Katrina Survivors. South. Econ. J. 74, 326–343. https://doi.org/10.1002/j.2325-8012.2007.tb00841.x

[24] Lee, K. K., Miller, M. R., Shah, A. S. V. (2018). Air Pollution and Stroke. Journal of Stroke, 20(1), 2–11. https://doi.org/10.5853/jos.2017.02894

[25] Lewicka, M., 2011. On the Varieties of People's Relationships With Places: Hummon's Typology Revisited. Environ. Behav. 43, 676–709. https://doi.org/10.1177/0013916510364917

[26] Liu, R., Greene, R., Yu, Y., Lv, H., 2022. Are migration and settlement environment-driven? Environment-related residential preferences of migrants in China. J. Clean. Prod. 377, 134263. https://doi.org/10.1016/j.jclepro.2022.134263

[27] Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing Climate Change Adaptation Needs for Food Security in 2030. Science 319, 607–610. https://doi.org/10.1126/science.1152339

[28] Louhichi, M., Nesmaoui, R., Mbarek, M., Lazaar, M. (2023). Shapley Values for Explaining the Black Box Nature of Machine Learning Model Clustering. Procedia Computer Science, 220, 806–811. https://doi.org/10.1016/j.procs.2023.03.107

[29] Marchiori, L., Maystadt, J.-F., Schumacher, I., 2012. The impact of weather anomalies on migration in sub-Saharan Africa. J. Environ. Econ. Manag. 63, 355–374. https://doi.org/10.1016/j.jeem.2012.02.001

[30] McLeman, R., Smit, B., 2006. Migration as an Adaptation to Climate Change. Clim. Change 76, 31–53. https://doi.org/10.1007/s10584-005-9000-7

[31] Merrick, L., Taly, A. (2020). The Explanation Game: Explaining Machine Learning Models Using Shapley Values. In A. Holzinger, P. Kieseberg, A. M. Tjoa, & E. Weippl (Eds.), Machine Learning and Knowledge Extraction (pp. 17–38). Springer International Publishing. https://doi.org/10.1007/978-3-030-57321-8_2

[32] Meze-Hausken, E., 2000. Migration caused by climate change: how vulnerable are people inn dryland areas? Mitig. Adapt. Strateg. Glob. Change 5, 379–406. https://doi.org/10.1023/A:1026570529614

[33] National Health and Family Planning Commission of China, 2018. China's Migrant Population Development Report.

[34] National Bureau of Statistics of China, 2021. The 7th population census data report 4–5.

[35] Renaud, F.G., Dun, O., Warner, K., Bogardi, J., 2011. A Decision Framework for Environmentally Induced Migration. Int. Migr. 49, e5–e29. https://doi.org/10.1111/j.1468-2435.2010.00678.x

[36] Reuveny, R., 2007. Climate change-induced migration and violent conflict. Polit. Geogr., Climate Change and Conflict 26, 656–673. https://doi.org/10.1016/j.polgeo.2007.05.001

[37] Rigaud, K.K., de Sherbinin, A., Jones, B., Bergmann, J., Clement, V., Ober, K., Schewe, J., Adamo, S., McCusker, B., Heuser, S., Midgley, A., 2018. Groundswell.: Preparing for Internal Climate Migration.

[38] Scheffran, J., Marmer, E., Sow, P., 2012. Migration as a contribution to resilience and innovation in climate adaptation: Social networks and co-development in Northwest Africa. Appl. Geogr., The Health Impacts of Global Climate Change: A Geographic Perspective 33, 119–127. https://doi.org/10.1016/j.apgeog.2011.10.002

[39] Speare, A., 1974. Residential satisfaction as an intervening variable in residential mobility. Demography 11, 173–188. https://doi.org/10.2307/2060556

[40] Swim, J.K., Stern, P.C., Doherty, T.J., Clayton, S., Reser, J.P., Weber, E.U., Gifford, R., Howard, G.S., 2011. Psychology's contributions to understanding and addressing global climate change. Am. Psychol. 66, 241–250. https://doi.org/10.1037/a0023220

[41] Turner, M. C., Andersen, Z. J., Baccarelli, A., Diver, W. R., Gapstur, S. M., Pope III, C. A., Prada, D., Samet, J., Thurston, G., Cohen, A. (2020). Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. CA: A Cancer Journal for Clinicians, 70(6), 460–479. https://doi.org/10.3322/caac.21632

[42] UN, D., 2022. Policy Brief No. 133: Migration Trends and Families — Department of Economic and Social Affairs [WWW Document]. URL https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-no-133-migration-trends-and-families/ (accessed 7.18.23).

[43] Wang, L., Dai, Y., Kong, D. (2021). Air pollution and employee treatment. Journal of Corporate Finance, 70, 102067. https://doi.org/10.1016/j.jcorpfin.2021.102067

[44] Wolpert, J., 1966. Migration as an Adjustment to Environmental Stress. J. Soc. Issues 22, 92–102.

https://doi.org/10.1111/j.1540-4560.1966.tb00552.x

[45] Zolberg, A.R., Benda, P.M., 2001. Global Migrants, Global Refugees: Problems and Solutions. Berghahn Books.

A Appendix

K-means clustering based on principal component analysis

Above all, the data has been standardised to perform principal component analysis. Principal component analysis (PCA) is a multivariate statistical technique typically employed to reduce the dimensionality of several interrelated variables while retaining the maximum information features of the original data. Assuming a dataset C with a dimension of n, PCA can project the n-dimensional features onto the new k-dimensional space. This k-dimensional space constitutes a new orthogonal feature, also referred to as the principal component. The new k-dimensional features are reconstructed based on the initial n-dimensional features.

In mathematics, PCA can be expressed as a collection of standardised n-dimensional data, X_1, X_2,... X_n, then i^{th} principal component Z_i can be written as a linear combination of the original variables as formula 5,

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{in}X_n \tag{5}$$

Then the i principal components are weighed and summed to obtain the final evaluation. The weight corresponds to the variance contribution ratio of each principal component. After obtained the first and the second principal component, a new dataset has been generated. This dataset is comprised of the first, second principal component and normalised migration flow. Regarding this dataset, the impact of input feature on migration flow is processed using K-means clustering.

K-means clustering is a commonly used unsupervised machine learning algorithm with crucial applications in pattern recognition, optimization, image processing, and other research (Ahmed et al., 2020). The foundational principles of the K-means algorithm is straightforward. A given dataset is divided into K clusters according to the Euclidean distance. The aim is to have the points within a cluster be as closely connected as possible and to maximise the distance between the clusters. It mainly involves the following processes,

(1) To determine the number of clusters (k) using the elbow method, the algorithm randomly selects q point from the dataset as the initial cluster centres Q_j, (j= 1,2,..., k).

(2) For each data point x_i(i=1...n), the algorithm computes the distance between the data point and cluster centre Q_j, subsequently assigning x_i to the nearest cluster. Let C_i denote the i^{th} clusters to which data point x_i is allocated. The minimum distance from x_i to C_i can be articulated as 6:

$$C_i = \arg\min \|x_i - Q_j\|^2 \tag{6}$$

(3) For each cluster, the algorithm calculates the mean value of all points contained within. The mean value will be the new cluster centre, which can be represented as 7:

$$Q_j = \frac{1}{N_j} \sum x_i \tag{7}$$

,where N_j is the number of points belonging to cluster j. Step (2) and (3) are reiterated until there is no significant change in cluster centres anymore. The algorithm will then conclude automatically.

The essence of the K-means algorithm is to minimise the cumulative of the distance between each point and its centre cluster. In other words, it seeks to reduce the Sum of Squared Errors (SSE).

$$SSE = \sum j \sum i \|x_i - Q_j\|^2 \tag{8}$$

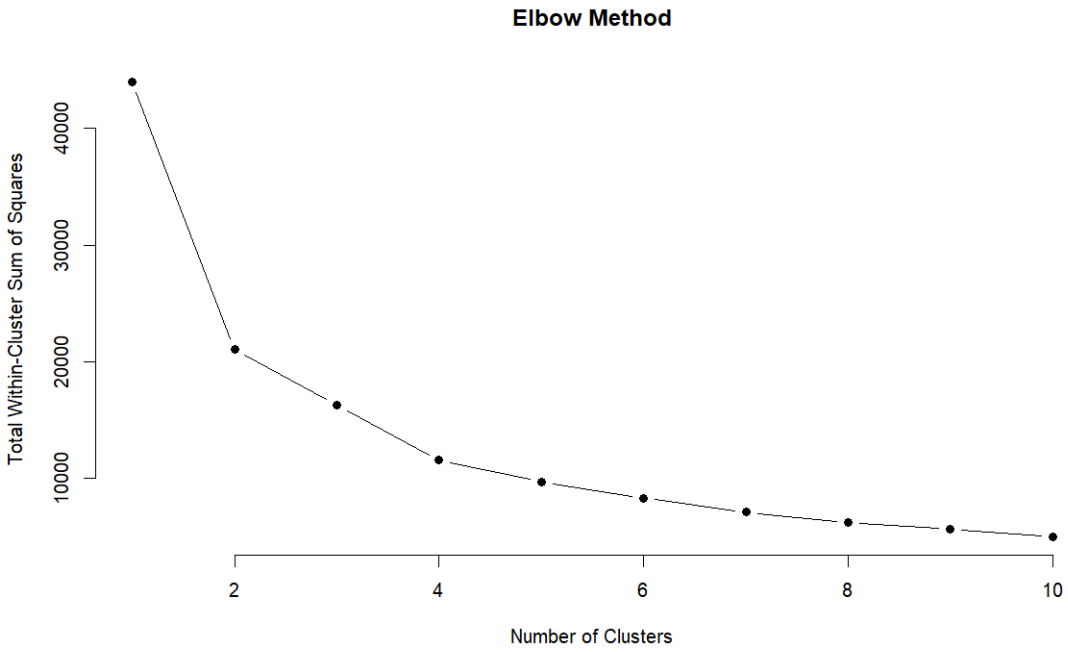

Figure 9: Elbow method graph

Table 8: Silhouette coefficient of cluster numbers

Cluster number	**Silhouette Coefficient**
1	0
2	0.4797
3	0.4493
4	0.4025
5	0.3738
6	0.3650
7	0.3668
8	0.3643
9	0.3418
10	0.3615