Simulation

Applications

Simulation: Transactions of the Society for Modeling and Simulation International
2021, Vol. 97(4) 267–285
© The Author(s) 2021
DOI: 10.1177/0037549720975128
journals.sagepub.com/home/sim
SAGE SCS

Understanding the dynamics of inter-provincial migration in the Mekong Delta, Vietnam: an agent-based modeling study

Hung Khanh Nguyen[1], Raymond Chiong[1], Manuel Chica[1,2] and Richard H Middleton[1]

Abstract

Recent large-scale migration flows from rural areas of the Mekong Delta (MKD) to larger cities in the South-East (SE) region of Vietnam have created the largest migration corridor in the country. This migration trend has further contributed to greater rural–urban disparities and widened the development gap between regions. In this study, our aim is to understand the migration dynamics and determine the most critical factors affecting the behavior of migrants in the MKD region. We present an agent-based model and incorporate the Theory of Planned Behavior to effectively break down migration intention into related components and contributing factors. A genetic algorithm is used for automated calibration and sensitivity analysis of model parameters, in order to validate our agent-based model. We further explore the migration behavior of people in certain demographic groups and delineate migration flows across cities and provinces from the MKD to the SE region.

Keywords

Agent-based modeling, migration, genetic algorithm, Mekong Delta

1. Introduction

Internal migration is a standard and prominent trend in many developing countries, especially those participating in globalization and economic integration.[1] Vietnam introduced its ‘‘economic renovations’’ in 1986, and has since become increasingly integrated into world markets. This economic integration, however, has created unbalanced development states and living standards among regions or provinces across the country. The inequalities have been considered the fundamental drivers of internal migration in Vietnam during the last decades.[2–4] Internal migration has further contributed to widening gaps between areas of origin and areas of destination, and consequently greater regional and provincial disparities.

The South-East (SE) region of Vietnam has been the most economically developed, with the highest per capita income and a higher level of living standards during the last two decades. This SE region and the triangle of a city and two provinces, comprising Ho Chi Minh City, Binh Duong, and Dong Nai, have been the primary migrant-receiving regions in Vietnam. The Mekong Delta (MKD), on the other hand, is the main migrant-sending region of the country. Most of the provinces in the MKD region have had major concerns about economic and social issues as well as their vulnerabilities to the increasing impacts of climate change in recent years. The fact that the two regions are close to each other further explains why the flow of migrants from the MKD to the SE region has been recognized as the largest migration corridor in Vietnam.[5,6]

According to national migration surveys,[6–8] there are four main groups of migration determinants, which are economic motivation, education pursuit, family-related factors, and other reasons. Many empirical studies on migration in Vietnam are in agreement with the survey

[1]School of Electrical Engineering and Computing, The University of Newcastle, Australia
[2]Andalusian Research Institute DaSCI, University of Granada, Spain

Corresponding author:
Raymond Chiong, School of Electrical Engineering and Computing, The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia.
Email: Raymond.Chiong@newcastle.edu.au

results and point out a range of factors influencing migration decisions at the national level. However, there are only a limited number of migration studies and reports focusing on the MKD region. In addition, to the best of our knowledge, no previous work related to migration in Vietnam has clearly specified how distinct components and relevant contributing factors form migration intention that leads to the actual migration behavior in the decision-making process of the migrants.

In this paper, we aim to study and understand the dynamics of individuals' migration decision-making behavior in the MKD region. We focus on the patterns of inter-provincial migration flows within the MKD region and from the MKD to the SE region. We want to identify the most critical components forming the migration behavior, as well as examine the impacts of different socio-economic and environmental factors that consequently lead to the migration decision of people in the MKD region. Additionally, we are interested in comparing the patterns of migration behavior among people in certain demographic groups, and in quantifying the migrant flows across cities and provinces in the migration corridor from the MKD to the SE region.

To do so, we adopt an agent-based modeling (ABM) approach,[9,10] which has become increasingly popular in demographic research in recent years.[11] ABM has several advantages over traditional empirical and statistical methods used in the migration literature.[12,13] For example, ABM has the ability to explicitly simulate autonomous decision-making and incorporate higher degrees of heterogeneity in human society.[14] ABM can also be efficiently used to model the impacts of social networks and social interactions,[15] which are considered critical components to explain the emergence of migration patterns.

We integrate the Theory of Planned Behavior (TPB), which is an established theory from social psychology,[16] into our proposed agent-based model. The TPB offers a behavioral heuristic approach, which is suitable for deliberating migration decisions that involve the inclusion of different background factors, the impact of peer influence, and the role of uncertainty. The TPB can also be utilized to effectively break down the cognition process of individual migration behavior into separate components, and subsequently be embedded in the migrants' decision rules.

Finally, we use a genetic algorithm (GA) to carry out automated calibration, parameter exploration, and sensitivity analysis, in the process of validating our agent-based model. The GA[17] has been proven to be a useful tool for automated calibration of different kinds of non-linear models, including agent-based models.[18–21] The GA also has advantages in terms of its capability to conduct automated sensitivity analysis[18] and explore wider ranges of parameter settings.[22]

The rest of this paper is organized into six main sections. In Section 2, we provide the background information about internal migration and the main migration determinants in Vietnam and the MKD region. The TPB and migration-based ABM-related work are described in Section 3. We then discuss our proposed agent-based model and the data collection process in Section 4. After that, the GA for model calibration and experimental setup is explained in Section 5. Experimental results are presented in Section 6. Finally, we draw conclusions and highlight future research directions in Section 7.

2. Background

2.1. Internal migration in Vietnam and the Mekong Delta

Vietnam is geographically divided into six administrative regions, namely the Northern Midlands and Mountains, the Red River Delta, the North and South Central Coast, the Central Highlands, the SE, and the MKD region. The country officially introduced economic renovations in 1986, moving from a centrally planned economy with public ownership of production toward a market economy. The transformation has not only led to significant economic growth and poverty reduction, but also changed the patterns of internal migration.[2,3] While large urban cities, including Ho Chi Minh City, Hanoi, and their surrounding provinces, have received high levels of industrial capital, other rural areas, such as the MKD, have lagged behind. These disparities have triggered a substantial flow of rural-to-urban migration, and shaped the dynamics of both regional and inter-provincial migration in Vietnam.[4]

The variations of regional migration in Vietnam from 2005 to 2017 are depicted in Figure 1. As we can see from the figure, the SE is the primary migrant-receiving region in the country, with a clear difference between the in-migration and out-migration flows. This region attracts more than 230,000 people on average, which is four times higher than the second most favored region. The MKD, on the other hand, has the highest negative net-migration rate and is the main migrant-sending region of Vietnam in terms of the number of net-migrants. Over the 13-year period from 2005 to 2017, slightly more than 25,000 on average entered the MKD region, whereas more than 110,000 people left the MKD region annually.

Large variations in migration have also been observed at the provincial level in Vietnam in terms of the average of migration rates (per thousand population) from 2005 to 2017, as shown in Table 1. The MKD region is made up of 12 provinces and Can Tho City, which is regarded as its regional center. It is worth noting that all the 12 provinces and one city in the MKD region have negative net-migration rates, indicating that they are all migrant-sending areas. In fact, most of the provinces with the highest negative average net-migration rates in Vietnam during the period under study are located in the MKD region.

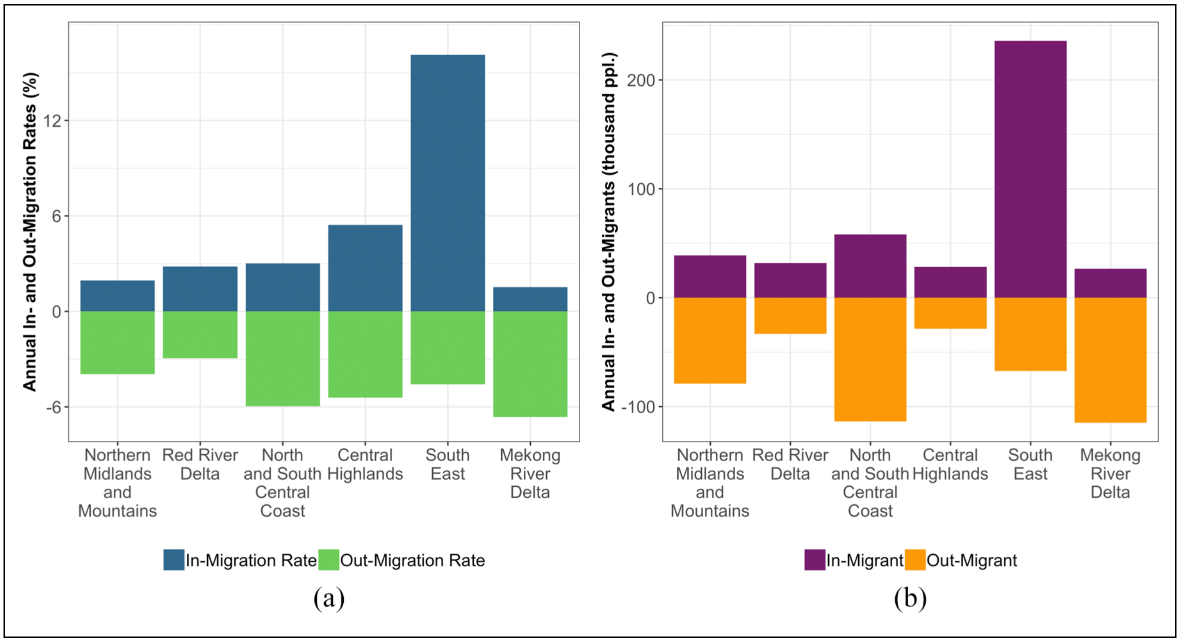

Figure 1. Differential of migration patterns among regions in Vietnam over a 13-year period between 2005 and 2017: (a) differential of in- and out-migration rates; (b) differential of in- and out-migrants.

Table 1. Average migration rates across cities and provinces in the MKD and the SE region from 2005 to 2017.

Province	Net-migr	Out-migr	In-migr
Mekong Delta			
Ca Mau	− 9.15	− 10.88	1.73
Bac Lieu	− 6.95	− 8.75	1.80
An Giang	− 6.76	− 9.11	2.35
Soc Trang	− 6.58	− 8.84	2.27
Ben Tre	− 6.52	− 10.02	3.50
Dong Thap	− 5.93	− 8.57	2.65
Hau Giang	− 5.34	− 9.35	4.01
Kien Giang	− 5.33	− 9.09	3.75
Vinh Long	− 4.18	− 8.79	4.61
Long An	− 3.49	− 7.71	4.23
Tra Vinh	− 2.65	− 7.38	4.73
Can Tho	− 1.36	− 8.21	6.87
Tien Giang	− 0.98	− 7.24	6.26
South-East			
Binh Duong	40.46	− 13.41	53.88
Ho Chi Minh City	12.44	− 7.26	19.72
Dong Nai	8.27	− 7.95	16.24
Vung Tau	2.51	− 7.02	9.53
Binh Phuoc	− 2.37	− 9.08	6.71
Tay Ninh	− 2.77	− 5.97	3.20

In contrast, the SE region, which covers Ho Chi Minh City and five provinces, has high positive net-migration rates. More specifically, two of the provinces, Binh Duong and Dong Nai, as well as Ho Chi Minh City, are those with the highest positive net-migration rates in Vietnam from 2005 to 2017. In terms of the average number of net-migrants in the same period, Hanoi, which is the capital city of Vietnam, is placed fourth below these three in the SE region.

2.2. Determinants of migration

According to internal migration surveys in Vietnam,[6–8] there are four main groups of migration determining factors: (a) economic motivation; (b) education pursuit; (c) family-related factors (marriage, being close to family); and (d) others (environmental impact, medical treatment).

2.2.1. Economic motivations. Migration surveys and empirical studies have shown that economic development is the most critical factor for migration in Vietnam.[3,4,23,24] Most migrants chose to move for economic reasons, including those who moved to look for employment and increase their incomes.

Nguyen-Hoang and McPeak,[25] Kim Anh et al.,[4] and Coxhead et al.[24] reported that foreign direct investment (FDI) and industrial zones, which have been distributed unequally throughout the country, underlie the impetus for internal migration in Vietnam for better employment opportunities. The 2015 internal migration survey indicated that FDI companies and businesses in the private sector are one of the main sources of employment for migrants.[6] The SE region had received FDI seven times more than the MKD region at the end of 2017. This clearly explains why the SE region has been cited as having the most working migrants from the MKD region.

The expected income differential between the origin and destination is also found to be essential for migration.[24] Based on the analysis of data from different surveys in the population and housing census of Vietnam, Kim Anh et al.[4] and Phan and Coxhead[3] suggested that provinces with high monthly income per capita are more likely to have higher rates of in-migration.

2.2.2. Education pursuit. Education has always been cited as one of the popular determinants by migrants in Vietnam.[5,7] The proportion of migrants who moved because of study purposes increased from 4.5% in 2004[7] to more than five times higher, 23.4% in 2015.[6] The increase in the education incentive among migrants, which is found in all regions of the country, reflects the fact that education has become more important for accessing well-paid employment in Vietnam.

2.2.3. Social network and family-related factors. The social network of migrants is found to be one of the sources of assistance in helping migrants adapt to their new living environment.[6,7] More than 60% of migrants responded in the 2015 migration survey[6] that they have families, relatives, and friends from their place of origin currently living in the place of destination. These personal relationships help reduce the risks associated with migration, save costs, and link the migrants to the place of destination.

2.2.4. Household income. The International Organization for Migration found a correlation between household income and the probability of migration in the MKD region. A survey of more than 1000 households showed that the migrants usually come from households that have a lower income, while non-migrants have better housing and are more well off.[26] Coxhead et al.[24] also indicated that high income appears to discourage people from migrating across provinces, since the gain from migration might not be sufficiently attractive. Entzinger and Scholten[26] studied the probability of migration at each income level, and showed that there is a sharp decline in the probability of migration when household income increases.

2.2.5. Environmental impacts. In addition to the socio-economic determinants of migration, the natural environment is increasingly recognized as influencing internal migration trends in Vietnam, and especially in the MKD region. Approximately 4.5% of migrants from the MKD region indicated that they moved for a more suitable natural environment.[26] Extreme weather events appear to have contributed to the key migration corridor of Vietnam between rural areas in the MKD and larger cities in the SE region.

3. Theory and related work

3.1. Theory of Planned Behavior

The TPB was originally proposed by Ajzen.[16] This psychological theory indicates that a particular behavior is explained by an intention, which is the result of a decision process, comprised of three core components, namely behavioral attitude (BA), subjective norm (SN), and perceived behavioral control (PBC).

In the context of migration, three types of beliefs – behavioral beliefs, normative beliefs, and control beliefs – determine the intention to migrate. According to the TPB, the beliefs people hold toward a migration option are influenced by a combination of different factors, including demographic characteristics as well as socio-economic and environmental factors (climate change, natural disaster). These background factors affect the formation of the beliefs, and indirectly shape the individual's migration intention and behavior.[16,27]

Firstly, behavioral beliefs are considered the motives or reasons behind people's decision to migrate. The evaluation of these beliefs on different outcomes that migration yields forms the BA toward migration. These behavioral beliefs are normally weighted by the subjective values that each person assigns to the outcomes.

Secondly, the normative belief of an individual is their perception of the social/normative pressure or how their peers make choices related to migration. Support from peers and their destination preferences can influence how a person perceives certain migration options and subsequently how willing they are to follow the choice.[28] The normative belief governs the SN for migration.

Thirdly, the control belief determines the PBC, which is the final component of the decision-making process to predict one's migration intention. The PBC is an individual's perception of their capability to take advantage of facilitators and to remove barriers to undertake an actual migration action.

The BA, SN, and PBC are indicators of migration intention. The stronger these components are, the more likely an individual would be to undertake the migration behavior. In addition, Ajzen[16] introduced the concept of actual behavioral control to improve the prediction of a certain behavior (i.e., migration in our case), which partly depends on factors that a person does not have complete control of. Ajzen[16] and Fishbein and Ajzen[27] used the PBC as a proxy to measure the actual behavioral control.

3.2. Agent-based migration models

There has been growing interest in the integration of the TPB with ABM to study migration behavior.[29,30] Kniveton et al.[28,31] developed an agent-based model on the basis of the TPB to study climate change-driven migration in Burkina Faso. In their model, the BA toward migration is defined as the probability of an individual with certain demographic characteristics. The SN is a function to assign the value of each of the migration options to an agent on the basis of their peers' most recent migration decisions. The PBC is computed from the assessment of whether the agent has assets and experience to undertake the migration action. The result is then compared to a random number between zero and one, and is converted to a binary outcome. If the PBC is zero, the agent does not develop a migration intention. An intention score is calculated for each of the migration options, and the one with the highest score is chosen.

Smith[32] subsequently developed another agent-based model, which is largely based on the decision model introduced in the work of Kniveton et al.,[28] to study rainfall-induced migration in Tanzania. Smith[32] designed the model to accommodate a less data-driven and more heuristic case study-based approach. In the model, while one's migration intention is driven by the characteristics of their age, gender, migration experience, and social network, it is also mediated by the household's ability to finance the move. The migration intention value of each agent is then compared with a pre-calculated threshold for migration, which is determined through survey data.

Willekens[12] and Klabunde et al.[33] studied multi-stage stochastic process models drawn from the TPB to understand international migration behavior. They extended the TPB into a process theory to account for the sequential nature of the decision process with three main stages. First, agents develop their beliefs that subsequently determine the intention to migrate. Each agent then moves to the planning and preparation phase, in which they consider the actual control over migration. A person might leave or stay in the country in the last stage. Willekens[12] and Klabunde et al.[33] emphasized the stochastic process of transitions between stages. For example, in Willekens's work,[12] the age at which an individual considers leaving the country and the duration they stay in each stage of the migration process are dependent on random factors drawn from an exponential waiting time distribution.

More recently, Nguyen et al.[34,35] integrated the TPB into their proposed agent-based model to explore the dynamics of migration flows across the MKD region. Their model was calibrated with actual data of in-, out-, and net-migration rates of the city and provinces in the region. In their initial work, the model was manually calibrated with three parameters, namely the BA, SN, and PBC. Nguyen et al.[35] later refined the model to improve the cognition of the migration decision process and implemented a GA to perform automated model calibration and sensitivity analysis with three additional parameters related to socio-economic attributes.

In our current work, we extend the agent-based model of Nguyen et al. with a comprehensive set of empirical data to determine the most critical components and factors that would affect the final migration decision of migrants in the MKD region. We take advantage of the GA to systematically validate our agent-based model with thorough automated calibration, model parameter exploration, and sensitivity analysis activities. Based on the richness of the output data generated from the agent-based model, migration flows across the city and provinces of the MKD to the SE region can be delineated.

4. Agent-based modeling of inter-provincial migration

4.1. Design of the agent-based model

The main entities of our agent-based model include *province* and *person* agents. Each province agent stores information about their population as well as socio-economic and environmental factors. Each person agent, which resides in a province agent, is classified into one of five quintile income groups, earning a certain income and bearing a certain living cost. Each person agent also has distinct views on how different socio-economic and environmental factors affect their migration decision-making process.

Each person agent makes a decision in two stages, including an assessment of migration intention of different provinces and the development of their own behavior toward migrating or staying. The individual migration decision process, which is adapted from Kniveton et al.'s work,[28] is shown in Figure 2. The migration intention assessment is based on the TPB framework.

In the migration decision process, a person agent initially computes an intention score for each destination, including the province in which the person agent currently resides. The migration intention, I, has three core components: BA, SN, and PBC. Agent i performs the intention calculation, $I_{i,j}(t)$, for each province agent j at time t as per the following equation:

$$I_{i,j}(t) = \alpha^1 BA_{i,j}(t) + \alpha^2 SN_{i,j}(t) + \alpha^3 PBC_{i,j}(t) \qquad (1)$$

4.1.1. Behavioral attitude.

The BA of a person agent toward the migration assessment is assumed to be an outcome of conscious calculus affected by different socio-economic levels and environmental impacts of each province agent. $BA_{i,j}(t)$, is calculated based on the following function[36]:

$$\begin{aligned} BA_{i,j}(t) = {} & \hat{\beta}_i^1 \overline{emp}_{i,j}(t) + \hat{\beta}_i^2 \overline{inc}_{i,j}(t) + \\ & \hat{\beta}_i^3 \overline{edu}_{i,j}(t) + \hat{\beta}_i^4 \overline{env}_{i,j}(t) \end{aligned} \qquad (2)$$

Here, person agent i considers the difference of four attributes, including employment prospect $\overline{emp}_{i,j}(t)$, potential income $\overline{inc}_{i,j}(t)$, education opportunity $\overline{edu}_{i,j}(t)$, and environment impact $\overline{env}_{i,j}(t)$, between the province in

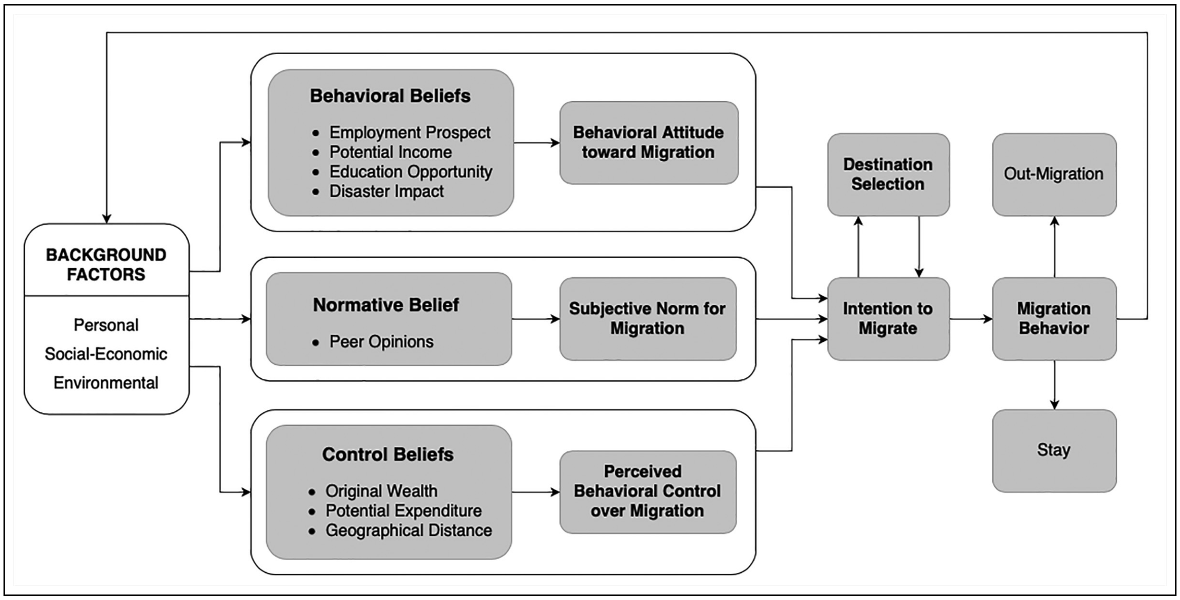

Figure 2. Individual cognition of the migration decision process.

which the agent is currently located and another province agent j. Values of $\overline{emp}$, $\overline{inc}$, and $\overline{edu}$ of each province are normalized in the range $[0, 1]$ among all province agents at the time the person agent makes the migration assessment. The attributes $\overline{env}$ are inversely transformed with the intent that a province with less extreme weather impact has a higher value. It is presumed that all province attributes are available without any error for the person agent during their migration assessment.

Since there is no previous empirical study exploring the relative weight of different socio-economic attributes and environmental impacts affecting individual migration decisions in the context of the MKD region, each parameter $\hat{\beta}_i^n$ in Equation (2) is randomly assigned for individual person agent i. Each parameter $\hat{\beta}_i^n$ is initially drawn from a uniform distribution in the corresponding range of $[0, \beta^n]$, and then adapted to satisfy the constraint: $\hat{\beta}_i^1 + \hat{\beta}_i^2 + \hat{\beta}_i^3 + \hat{\beta}_i^4 = 1$. The unique combination of $\hat{\beta}$ defines the heterogeneity of agents in their perception toward different attributes that impact the migration decision-making process.

The environmental impact, $env_j(t)$, of province agent j at time t is computed based on two components, namely the intensity of extreme weather events and the vulnerability of agent j to each event, vul_j. The intensity of climatic hazard component is produced by a Poisson distribution function $P(x, \lambda)$, where x is the observed occurrence of hazards and λ is the expected number of hazard events in a given time interval[20]:

$$env_{j,t} = P\left(\overline{hazards_j} + 1, hazards_j + t/steps\right) vul_j \qquad (3)$$

Here, $\overline{hazards_j}$ is the rounded up value of the average number of extreme weather events, $hazards_j$, that occurred in the province, and $t/steps$ provides an increasing trend for the intensity of hazard occurrences as t reaches max_{steps}.

4.1.2. Subjective norm. $SN_i(t)$ is updated at the same time as the BA is calculated. In this model, an individual agent is modeled in such a way that they might adapt their behavior according to their neighbor's final migration decision:

$$SN_i(t) = \hat{\gamma}_i \frac{\sum migr_i(t)}{\sum neig_i(t)} \qquad (4)$$

Fundamentally, the SN is calculated as the proportion of agent i's neighbors who have migrated.[33] It reflects the assistance of these neighbors on agent i's decision to migrate to a new destination. The value of $\hat{\gamma}_i$ is allocated randomly to each person agent, following a uniform distribution between $[0, \gamma]$. In our model, the neighbors of agent i are those located within i's neighborhood area, determined by a fixed range parameter, u.

4.1.3. Perceived behavioral control. The $PBC_{i,j}^k(t)$ of person agent i in income quintile group k incorporates their current income as both a facilitator and a barrier. Another barrier that person agent i considers toward province agent j is the migration cost. In this model, the migration cost is determined by living expenditure $exp_{i,j}(t)$ in the destination, province j. The geographical distance between provinces, $dis_{i,j}$, is also a critical determinant of migration cost, which includes the cost of transportation, as well as the psychological cost[3]:

$$PBC_{i,j}^{k}(t) = \frac{\frac{inc_i^k(t)}{inc_i^5(t)}\left(1 - \hat{\delta}_i^1 \frac{inc_i^k(t)}{inc_i^5(t)}\right)}{\left(1 + \hat{\delta}_i^2 \overline{exp}_{i,j}^{k}(t)\right)\left(1 + \hat{\delta}_i^3 \overline{dis}_{i,j}\right)} \quad (5)$$

Here, $inc_i^k(t)$ is the actual income value of person agent i in the corresponding quintile group k. However, the income of person agent i is normalized in the range $[0, 1]$ by dividing by the highest income within the same province, $inc_i^5(t)$. Values of $\overline{exp}_{i,j}^{k}(t)$ of each person agent in each quintile group are normalized among all province agents at each time t. Here, $\overline{dis}_{i,j}$ is computed as the normalized distance between the centroids of two province agents. Each parameter $\hat{\delta}_i^n$ is assigned to person agent i following a uniform distribution between $[0, \delta^n]$, to reflect the heterogeneity among person agents' perceptions on factors that facilitate or hinder their migration option.

By calculating the intention score using Equation (1), each person agent stores the intention values for the corresponding provinces. The person agent then develops their final migration behavior through a destination selection process and decides whether to migrate or stay. Every agent is modeled to only consider the list of provinces with higher intention values than the value yielded for their current province. The agent will identify the highest intention score in the list and compare this with a random number $\in [0, 1]$. If the number generated is less than the intention score, the agent migrates to the corresponding province. Otherwise, the following highest scoring province will be considered.[31,37,38] If none of the preferred migration provinces is chosen, the agent will stay at their current location. This stochastic process prevents a disproportionate in-migration flow to a specific destination with a marginally higher score and accounts for uncertainties in the actual migration behavior.[12]

If an agent decides to migrate, they inform the surrounding agents of their decision. Those neighboring agents then update the number of migrants in their network, $migr_i(t)$, with an increase of one. The agent who migrates to another province is randomly located within the geographical boundary of that province. This person agent establishes a new neighborhood and also notifies the other agents residing in the area. It is assumed that the agent will remain in the same income quintile group and have similar living expenditures.

At the end of each simulation step, the province agent updates their population, which is affected by both the natural population growth rate and the number of net-migrants. The person agent is modeled to remain alive during the simulation runs, since a net population growth rate is applied. The numbers of in-ward and out-ward migrants in each province agent, which are endogenously generated from the model, determine the flow of net-migrants.

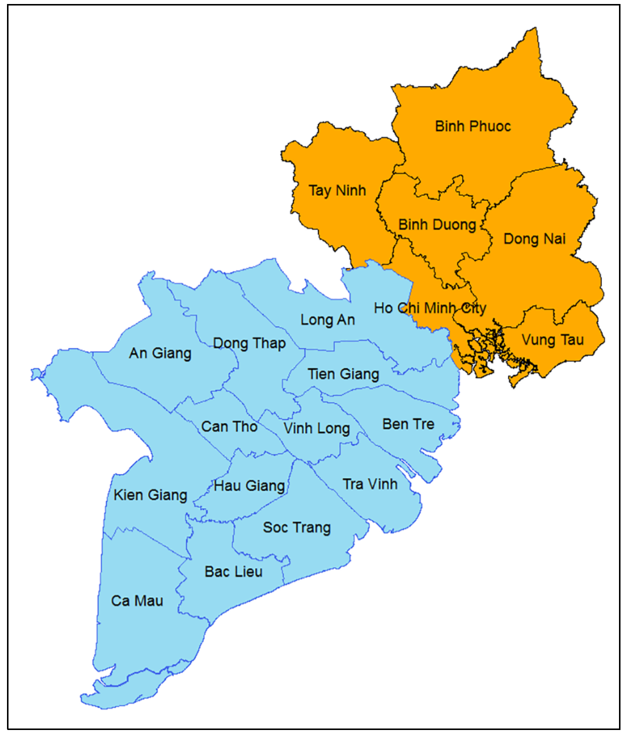

Figure 3. Locations and geographical boundaries of cities and provinces in the MKD and SE regions.

4.2. Data collection and attributes

There are two cities and 17 provinces in the MKD and SE regions. Their locations and geographic boundaries are shown in Figure 3, extracted from Vietnam's provincial-level map.[39] Islands of the provinces are not included in this map.

We explored and collected input data related to these two regions from different sources, which include the General Statistics Office of Vietnam[40] and Vietnam's Household Living Standard Survey datasets. Specific data for each city or province is in yearly time-series for a 13-year period between 2005 and 2017. Missing data of certain years was assumed to be the average of corresponding values in the previous and later years.

Based on the review of migration determinants and the availability of data, we included different factors contributing to the socio-economic attributes that migrants in the MKD region have considered in their decision-making process. The list of attributes and relevant factors can be found in Table 2.

As discussed, employment opportunities are among the most important reasons for the migrants. We consider five indicators representing the employment prospects of each province:

Table 2. Socio-economic attributes (development levels, environmental impacts) and the contributing factors.

Attribute	Factors
.emp	Number of FDI projects (*fdi*) Number of businesses (*biz*) Number of freight traffic and logistic activities (*log*) Number of agricultural farms (*agr*) Percentage of employed workers at 15 years of age and above (*opp*)
inc	Average monthly income per capita Monthly income per capita by income quintile
edu	Number of pupils in general education Number of teachers in general education Number of students in universities and colleges Number of lecturers in universities and colleges
exp	Monthly living expenditure per capita by income quintile by region Spatial cost of living index
env	Number of occurred natural hazards Total number of fatalities and damaged housing caused by natural hazards

FDI: foreign direct investment.

$$\overline{emp}_{i,j}(t) = \hat{\theta}^1\overline{fdi}_{i,j}(t) + \hat{\theta}^2\overline{biz}_{i,j}(t) + \hat{\theta}^3\overline{log}_{i,j}(t) + \hat{\theta}^4\overline{agr}_{i,j}(t) + \hat{\theta}^5\overline{opp}_{i,j}(t) \quad (6)$$

Here, $\hat{\theta}$ is the weighted value for each of the five θ parameters. We used $\theta \in [0, 1]$ as calibration parameters of our agent-based migration model. Employment opportunities generated from the FDI companies, represented by θ^1, are considered the main impetus for internal migration in Vietnam as well as migration flows from the MKD to the SE region.[4,24,25] We set $\theta^1 = 1$ to reflect the fact that among the relevant indicators, employment sources from FDI enterprises, $\overline{fdi}$, account for the highest contribution to the employment prospect, $\overline{emp}$.

The other sources of employment prospects are represented by the number of businesses, the amount of freight traffic (logistics activities), the number of farms (agricultural activities), and the percentage of employed workers over 15 years of age (employment opportunities). In fact, we had tested the employment prospect parameter with more contributing factors than those listed in Table 2. Factors that were removed after the parameter exploration process include the percentage of trained employed workers, the unemployment rate, and the underemployment rate by province. The weighted values of these factors were close to zero. Removal of these factors did not cause significant changes to the dynamics of model outcomes.

We assumed that four factors, namely the numbers of pupils and teachers of general education, and the numbers of students and lecturers in universities and colleges, contribute equally to the education attribute, *edu*. For the income attribute, *inc*, it is important to note that while we have used the general average monthly income per capita to calculate *BA* in Equation (2), we actually computed the *PBC* in Equation (5) with data of monthly income per capita by income quintile. We computed the provincial-level data of monthly living expenditure by income quintile based on the relevant regional-level data and the spatial cost of living index, which are available across cities and provinces in Vietnam.

Data related to the environmental impacts was collected from the national disaster database of Vietnam available in DesInvetar - Disaster Loss Databases.[41] The average frequency of climatic hazards in a year, $hazards_j$, was calculated as the average of all hazards occurred from 1989 to 2015. The vulnerability index, $vul_j \in [0, 1]$, of each province was accounted for by the number of dead, injured, and missing people and the number of damaged and destroyed houses accumulated by the occurrences of all climatic hazards in the 27-year period.

5. Experimental setup

5.1. Initialization

We implemented our agent-based model in Java using the Multi-Agent Simulator of Neighborhoods (MASON) framework.[42] Geographic boundaries and locations of the cities and provinces were extracted in the shapefile format through geoMASON.[43]

We initialized the model with 3340 person agents distributed across one city and 12 provinces in the MKD region, representing its total estimated population of 16,700,000 people at the end of 2004.[40] Person agents were populated into each province agent according to the population size of the corresponding province. The location of each person agent was assigned randomly in the given province, and parameter u was set so that the neighborhood of each person agent was defined within a 10 km radius from their location.[33] We did not populate province agents representing the city and provinces in the SE region with residents, since our focus in this study was on the pattern of migration flows in the MKD region.

Each person agent was initially categorized into one of five income quintiles. Every agent has their own perception of how different socio-economic and environmental attributes impact on their migration decision-making process. The weights of these attributes were initialized independently for each person agent, and kept constant during a simulation run but changed for different Monte Carlo (MC) trials.

Every simulation run was for a period of 13 years, from 2005 to 2017, based on the input data we have. We ran each simulation for 156 steps, with each step representing one month. The environmental impact attribute, *env*, was updated endogenously for every month. Since the data of province agents' attributes, including *emp*, *inc*, *pov*, *edu*,

Table 3. A set of 15 calibration parameters in the agent-based migration model.

Parameter	Description	Equation
α^1	Parameter of *BA*	Equation (1)
α^2	Parameter of *SN*	
α^3	Parameter of *PBC*	
β^1	Weight of employment prospect	Equation (2)
β^2	Weight of potential income	
β^3	Weight of education opportunity	
β^4	Weight of environmental impact	
γ	Weight of subjective norm	Equation (4)
δ^1	Weight of original wealth	Equation (5)
δ^2	Weight of potential expenditure	
δ^3	Weight of distance	
θ^2	Source from general business	Equation (6)
θ^3	Source from agriculture	
θ^4	Source from logistics	
θ^5	Employment rate	

and *exp*, is only available annually, it was assumed that those attributes have the same values during the 12 months of each year.

Each person agent was assumed to make migration decisions twice a year: in June and December. Each province agent would update its population based on the actual natural population growth rate and the number of net-migrants at the end of each year. Reported results for every province were calculated annually by averaging a total of 40 independent MC trials.

5.2. Automated calibration using a genetic algorithm

Automated calibration is a computationally intensive process that aims to fit simulated data to real-world data.[20,44,45] In this study, we developed a GA[17] to automate the calibration process of our model parameters. The GA, in a nutshell, involves the evolution of a population of solutions, each of which represents a set of model parameters in this context. These solutions are iteratively evolved until the best possible set is found. The GA has been proven to be a useful tool for automated calibration in different kinds of non-linear models, including agent-based models.[18–20,46]

The evaluation process entails iteratively running the model and carrying out parameter tuning in order to identify a set of parameter settings that best match the reference data.[45] We used the migration data at hand to discover the best values for all calibration parameters. Originally, there were 16 parameters. After some preliminary runs of the GA and exploration of the calibration output, a parameter related to the poverty rate was removed, as it did not contribute significantly to explaining the patterns of migration flows in the MKD region. Table 3 shows the final set of 15 calibration parameters.

Our GA was implemented using the Evolutionary Computation Java (ECJ) library.[47] The ECJ toolkit is a Java-based evolutionary computation package that can easily be integrated with the MASON framework. Since conducting automated calibration is a computationally expensive process, we used multi-threading to conduct our experiments. This multi-threading technique enables the evaluation with different MC runs to be executed in a multi-threaded fashion on high-performance computing facilities.

Detailed settings of the GA are as follows. We implemented a generational GA, where every generation of the population replaces the previous one.[17] The GA has a population size of 100. All the calibration parameters, as can be seen in Table 3, were initialized with random values following a uniform distribution within the range $[0, 1]$.

A three-tournament selection mechanism[48] with weak elitism was applied, which means the best solution is always preserved through every generation. We used the simulated binary crossover operator and polynomial mutation,[49] with a crossover probability of $p_c = 1$ and a mutation probability of $p_m = 0.2$. We adopted standard values for the distribution indexes[47,49] for the crossover and mutation operators (i.e., $\eta_c = 20$ and $\eta_m = 20$). Each GA run for calibrating the migration model ended after 500 evaluations.

We repeated each GA calibration model 20 times (with different seeds), because the GA itself is non-deterministic. A total of 1,000,000 combinations of solutions were assessed during the automated calibration process. At the end of all the runs, the results were returned.

The objective function of the GA, in our case, measures the deviation or error of the model outputs with the actual average of in-, out-, and net-migration rates of each province and city in the MKD region from 2005 to 2017 (shown in Table 1). The GA calibration process then identifies a set of parameters such that the error measure, ϵ, is minimized. We adopted the L^2 or Euclidean distance, which is equivalent to either the mean square error or root mean square error. The Euclidean distance is computed by the following:

$$L^2 = \sqrt{\sum_{i=1}^{3} |V_M(i) - V_R(i)|^2} \qquad (7)$$

where V_M and V_R are the vectors of three migration flow rates generated from the simulation model and attained from historical data in each province.

Due to the stochastic nature of the model, the value of L^2 was calculated as the average of fitting errors in all 40 MC runs. The final model error measure ϵ was the sum of 13 L^2 values calculated corresponding to one city and 12 provinces in the MKD region.

Table 4. Parameters calibrated by the GA.

Component	Calibrated parameter values				
Behavioral attitude	α^1 0.0013	β^1 0.9452	β^2 0.9953	β^3 0.2512	β^4 0.1069
Subjective norm	α^2 0.008	γ 0.6956			
Perceived behavioral control	α^3 0.0078	δ^1 0.4077	δ^2 0.2921	δ^3 0.8107	
Employment attribute	θ^2 0.1419	θ^3 0.5421	θ^4 0.0585	θ^5 0.5927	

6. Experimental results

6.1. Model calibration results

GA calibration of the agent-based migration model reported an Euclidean mean value of 50.6625 and a standard deviation of 10.682. Table 4 shows a set of 15 parameters calibrated by the GA. This set of parameter values returns the smallest error measure of $\epsilon = 37.9587$, which is 79.468% similar to the reference data, and therefore has been chosen to run the final model.

We also examined the normality of the results with the Shapiro–Wilk test (see Appendix A). In most cases, the p-value is greater than 0.05, implying that the distribution of the results is not significantly different from a normal distribution. Thus, we can assume that the relevant results are normally distributed. In cases where p-values < 0.05, we checked the normality visually using Q-Q plots (see Appendix B). We found that the simulation outputs in these cases are not significantly different from the normal distribution either.

In Figure 4, we show the respective mean and standard deviation of migration flow rates of each province between 2005 and 2017 over 40 independent MC trials as outputs of the model with the 15 calibrated parameters. The figure depicts three bar charts comparing the simulation results and actual migration dynamics across the city and provinces in the MKD region over the 13-year period. Based on the assumption of normality of the results, a 99% confidence interval is incorporated with the simulated mean value of each province. The order of the city and provinces is arranged by real values of the corresponding net-migration flow rates.

From Figure 4, we see that most of the observed values of in-, out-, and net-migration rates across the city and provinces in the MD region are within the 99% confidence interval. On the net-migration chart, it is clear that the simulation model has produced very similar patterns to the real data, capturing both the lowest and largest negative net-migration rates in the group of provinces positioned at

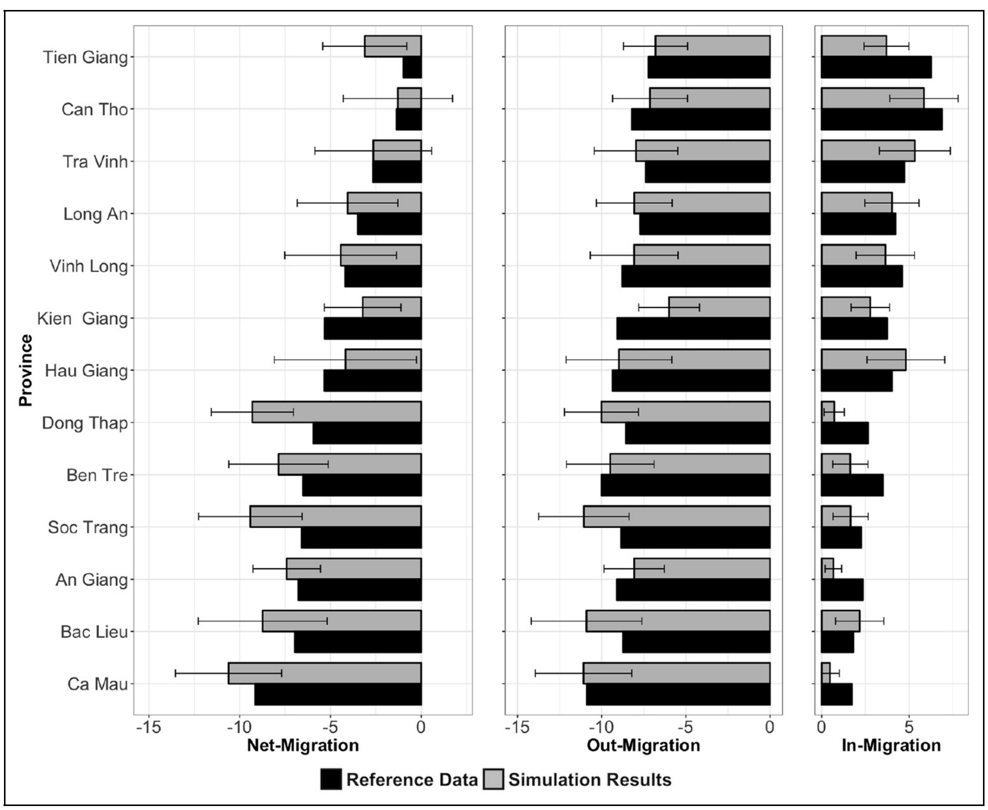

Figure 4. Comparison of simulated results and actual data of the average net-, out-, and in-migration rates of provinces in the MKD region from 2005 to 2017.

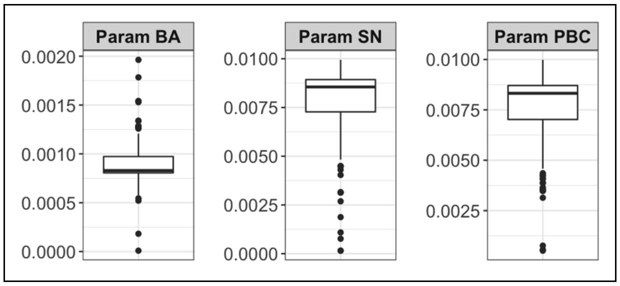

Figure 5. Boxplots showing the values of α^1, α^2, and α^3 from the best GA calibration solutions.

the top and bottom half of the chart. The only exception is the Dong Thap province.

The automated calibration also leads to a good fit on the actual out-migration rates. From the middle chart, the observed out-migration data of 12 out of 13 places are within the 99% confidence interval of the mean results generated from the agent-based model. On the in-migration chart, we observe that the simulation model is also able to replicate the dynamics of migration flows across the MKD region. There are only five provinces in which the in-migration rates are outside the intervals without any significant differences from the reference values.

6.2. Model parameter exploration

Alongside calibration, parameter exploration is another crucial ingredient for model testing and verification.[45] The GA has the advantage of being able to generate a large set of solutions to a problem. This enables the exploration of the already evaluated set of parameters and allows the examination of the distribution of these parameters.[20,44.] By analyzing the distribution of the model variables and parameters, we can achieve simpler and easier to understand model settings.[20]

The range of calibrated parameter values helps us to better understand the impact of these parameters on the final decision of migrants in the MKD region. Here, we focused on 100 sets of solutions that have produced the smallest values of error measure ϵ. Boxplots in Figures 5–7 represent the distributions of 11 of the calibration parameters obtained from the best solution sets and the relevant weighted values of these parameters.

The three boxplots in Figure 5 indicate that good values of α^1 (Param BA), α^2 (Param SN), and α^3 (Param PBC) are significantly small compared to the initial maximum value of 1 we set to implement the GA. This is expected, since the values of the three parameters are utilized to calculate the behavioral intention (Equation (1)), which determines the probability of people deciding to migrate. According to Table 1, the migration flow rates (number of migrants per 1000 population) are comparatively small.

In Figure 6, the top four boxplots display the range of good values of four β calibration parameters, while the bottom boxplots show the range of weighted values of these parameters. It is worth noting that the weights of these $\hat{\beta}$ attributes were initially drawn from a uniform distribution between 0 and the corresponding β, and then were weighted to calculate the BA.

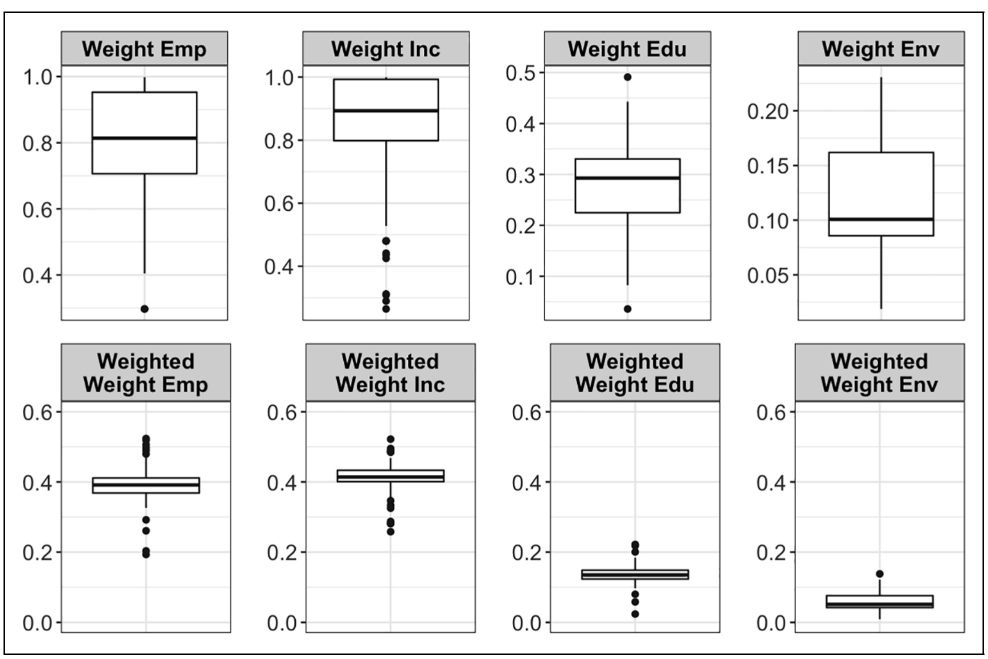

Figure 6. Boxplots showing the values of β^1, β^2, β^3, and β^4 from the best GA calibration solutions and weighted values of the relevant parameters.

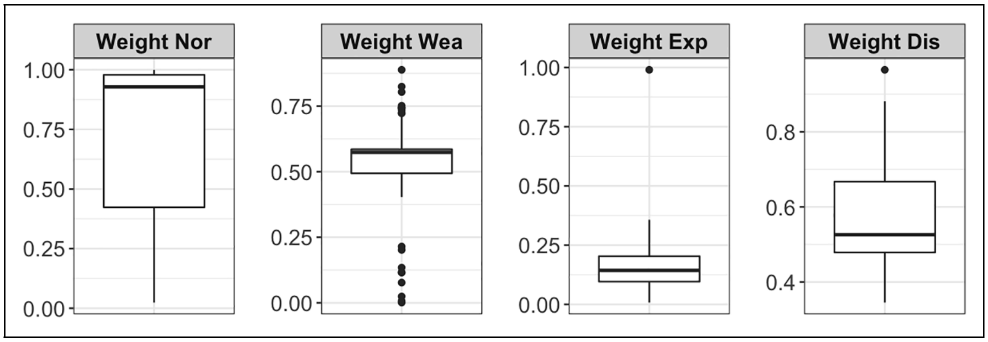

Figure 7. Boxplots showing the values of γ, δ^1, δ^2, and δ^3 from the best GA calibration solutions.

From the bottom boxplots, we see that the results generally support the literature on determinants of migration in Vietnam and the MKD region. Among the good weighted β values, the median weight of employment prospect attribute β^1 is the second highest at approximately 0.39133. The result is in line with the fact that the employment-related factor has been one of the leading reasons for people in the MKD region to migrate.[6,24] For the potential income attribute, the median is 0.4135, which is the highest among the β values. This result is also in agreement with the literature, indicating that the expected income differential between the origin and destination is crucial for migration in the MKD region.[24]

In terms of the education opportunity attribute, its corresponding boxplot also shows the distribution of good values of the weighted β^3 parameter. The median value is close to 0.1362, which indicates that the education-related factor accounts for more than 13% among the four migration determinants. The result is adequately justified by the fact that there is an increasing proportion of MKD migrants, from 4.5% in 2004 to 23.4% in 2015, migrating to Ho Chi Minh City in order to have access to higher educational institutions.[6]

For the environmental impact attribute, most of the weighted values of β^4 are close to 0.05. This result suggests that the impact of climatic hazards is relatively small (approximately 5%). The result, generally speaking, supports the recent survey in which 4.5% of migrants from the MKD region chose to move to a new place with a more suitable natural environment.[26]

Figure 7 displays the distribution of good values of different calibration parameters utilized in the calculation of Equations (4) and (5). In the first boxplot, we see that the median value of γ is 0.928, which indicates that people in the MKD region have a large variance in their perception regarding the influence of neighbors on the migration decision. From the other three boxplots, we see that most of the values of the δ parameters are larger than zero. For the original wealth and potential expenditure, the higher the values of δ^1 and δ^2, the less likely it is that people in higher income quintiles would move to a different place. For the weight of distance δ^3, the median value is high at 0.526. The result is again in agreement with findings from the study of Phan and Coxhead[3] that geographical distance is a critical determinant in the decision of migrants from the MKD region.

6.3. Sensitivity analysis

6.3.1. Univariate sensitivity analysis. We chose to run a sensitivity analysis following the one-factor-at-a-time methodology,[50] which modifies each parameter separately and keeps the other parameters fixed to their original values. Specifically, we conducted sensitivity analysis on the parameters of *BA* (α^1), *SN* (α^2), and *PBC* (α^3). We decided to use calibrated values of the three parameters as the baseline and varied these values from -100% to $+100\%$ with a fixed step of 25%. The other parameters were set to their corresponding calibrated values, as shown in Table 4. Results of this univariate sensitivity analysis are presented in Figure 8.

The heatmap indicates that varying the calibrated values of *BA* (α^1) causes the most significant change in the model results. Doubling the calibrated value of parameter α^1 leads to the largest error ($\epsilon = 266.73$) in comparison with the errors generated by doubling the values of parameters α^2 ($\epsilon = 40.97$) and α^3 ($\epsilon = 60.95$). We find a similar pattern when values of the three α parameters were each set to 0. Overall, changing the value of parameter *BA* results in the highest error value ($\epsilon = 167.17$), which is four times and nearly three times larger than the errors produced by changing the values of parameters *SN* and *PBC*, respectively. On the other hand, the results suggest that parameter *SN* has the least impact in determining intention among migrants in the MKD region. Changing the parameter values of *SN* does not yield significant differences ($\epsilon = 37.9587$).

6.3.2. Multivariate sensitivity analysis. We also employed the GA to perform multivariate sensitivity analysis, following Stonedahl and Wilensky's[18] approach, by modifying a

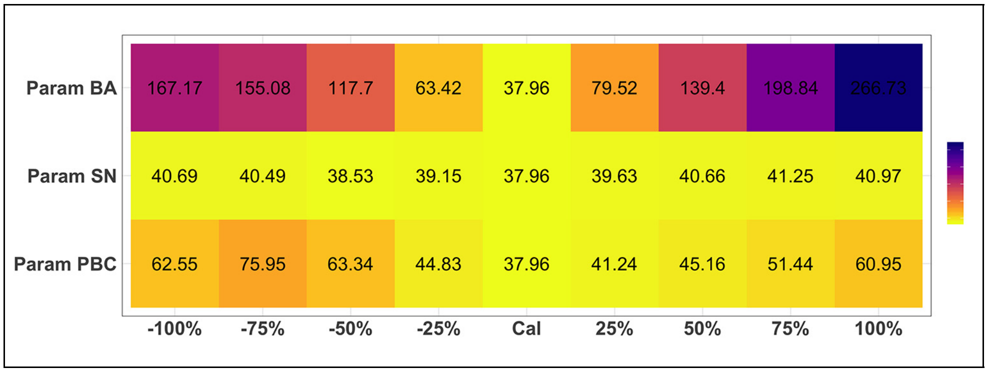

Figure 8. A heatmap with sensitivity analysis results of the three parameters α^1, α^2, and α^3.

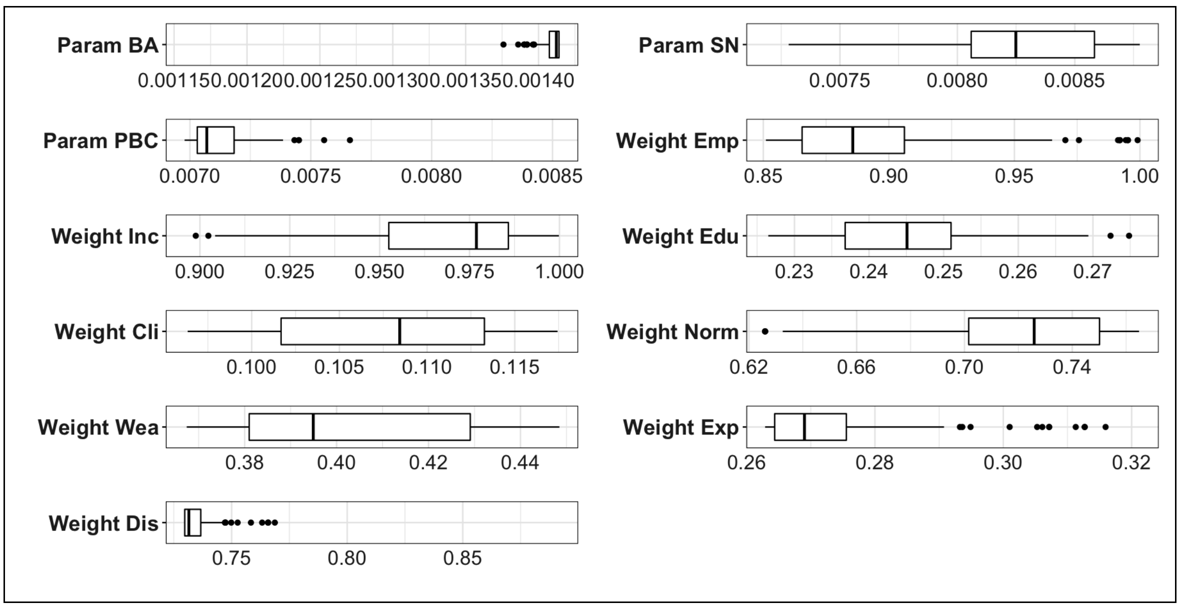

Figure 9. Boxplots showing values of all parameters related to *BA*, *SN*, and *PBC* obtained from the worst GA calibration solutions.

group of parameters.[50] This approach uses the GA to maximize (rather than minimize) the error measure, ϵ. The search space is constrained to a limited range within $\pm 10\%$ of the calibrated parameter values.[18,44,46] The upper bound, which we set to 1, was applied to the variations in the weights of employment prospect and potential income.

We conducted this series of sensitivity analysis on four groups of parameters (see Table 4). In each analysis, values of the related parameters were varied within the predefined range ($\pm 10\%$), while values of the remaining parameters were fixed. Results of the analysis can be seen in Figure 9, where 100 worst sets of parameters with error values ϵ ranging from 233.29 to 344.76 are included. In the figure, the outer box defines the parameter range, while the inner boxplots show the distributions of relevant parameters obtained from the worst solutions.

From the figure, we see that different GA runs found clear trends in different groups of parameter settings. The GA consistently explores higher values for parameter *BA*, while it selects low values for parameter *PBC* and the weight of distance. This also means that the agent-based migration model is particularly sensitive to these factors. The other parameter values are found to be relatively scattered throughout their ranges, indicating that it is not necessary for these parameters to be assigned a specific value in order to achieve large errors.

We also examined the distributions of error measurement ϵ yielded from the multivariate sensitivity analysis by the four groups of parameter settings. The summaries

Table 5. Statistical summaries of groups of parameters obtained from the worst GA calibration solutions.

Component	Min	1st Quin.	Median	Mean	3rd Quin.	Max
Behavioral attitude	50.72	64.85	79.66	92.67	112.57	285.74
Subjective norm	44.06	59.53	76.66	84.73	103.04	242.76
Perceived behavioral control	49.69	62.84	78.36	91.99	110.10	285.12
All	52.88	75.90	96.50	115.20	143.39	344.76

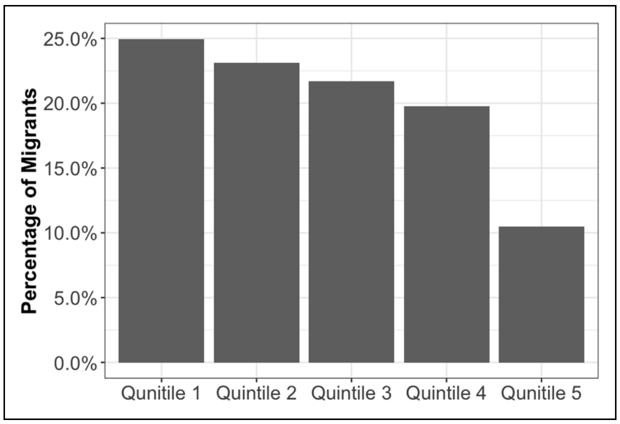

Figure 10. Percentage of the total number of migrants from the MKD region by income quintile.

are shown in Table 5. We found that the error generated from the group of *SN* parameters is not significantly smaller than errors produced from the groups of *BA* and *PBC* in any of the five quintiles. This points to the importance of the *SN* component in forming the migration intention, especially with the combination of the two relevant parameters.

6.4. Analysis of the migration dynamics

Based on the set of calibrated parameters with the smallest error measure, we further explored the migration dynamics in the MKD using our agent-based model. We first examined the proportions of migrants in the MKD region by income quintiles, which are shown in Figure 10. The simulation results indicate that people in higher income quintile groups tend not to migrate. The proportions of migrants among the first four income groups decline slowly from 25% in the first income quintile to approximately 20% in the fourth income quintile. The percentage of migrants from the last quintile (highest income) drops sharply to 13.2% of the total number of migrants. These results together with the exploration of the calibrated values of two δ^1 and δ^2 parameters advocate the empirical findings of Entzinger and Scholten[26] and Coxhead et al.,[24] in that high income and high expenditure in the destination discourage people from moving, since the new destination might not provide as much benefit compared to their current living conditions.

Secondly, we wanted to identify the most preferred destinations in the SE region among the migrants from the MKD region. The left-hand bar graph in Figure 11 shows the simulation results regarding the proportions of in-migrants toward the six potential destinations during the 13-year period. The right-hand bar chart displays the actual data of the percentages of in-migration flows to the same destinations but from all regions across Vietnam. There is no available provincial-level data with respect to migration flows from the MKD region, specifically to the SE region. Nevertheless, we find that the simulation results are able to replicate the dynamics of in-migration flows toward the city and provinces in the SE region. Ho Chi Minh City and the Binh Duong and Dong Nai provinces are the most popular destinations in the SE among the migrants either from the MKD region (from the simulation results) or from all regions across Vietnam (from real data).

Figure 12(a) shows the simulated migration flows across the MKD and the SE region and the annual average number of in-migrants to each city and province between 2005 and 2017. We clearly see that Ho Chi Minh City and the Binh Duong province are two of the most attractive destinations, receiving on average almost 63,000 and 30,000 people migrating from the neighboring region each year, respectively. In the MKD, most provinces receive less than 5000 people each year. The least preferred destination among the migrants is Ca Mau, which is located in the southernmost part of Vietnam, with only a little more than 520 people on average moving here each year.

Figure 12(b) highlights the in-migration flows of Ho Chi Minh City as the most favored destination among people in the MKD region. Main sources of migration toward Ho Chi Minh City are from Tien Giang, Long An, An Giang, and Dong Thap. Similar patterns can also be observed in the in-migrations flow into the Binh Duong province.

7. Conclusion and future work

In this paper, we have studied the dynamics of migration decisions of people in the MKD region. We took advantage of the TPB, which is a well-known theory derived

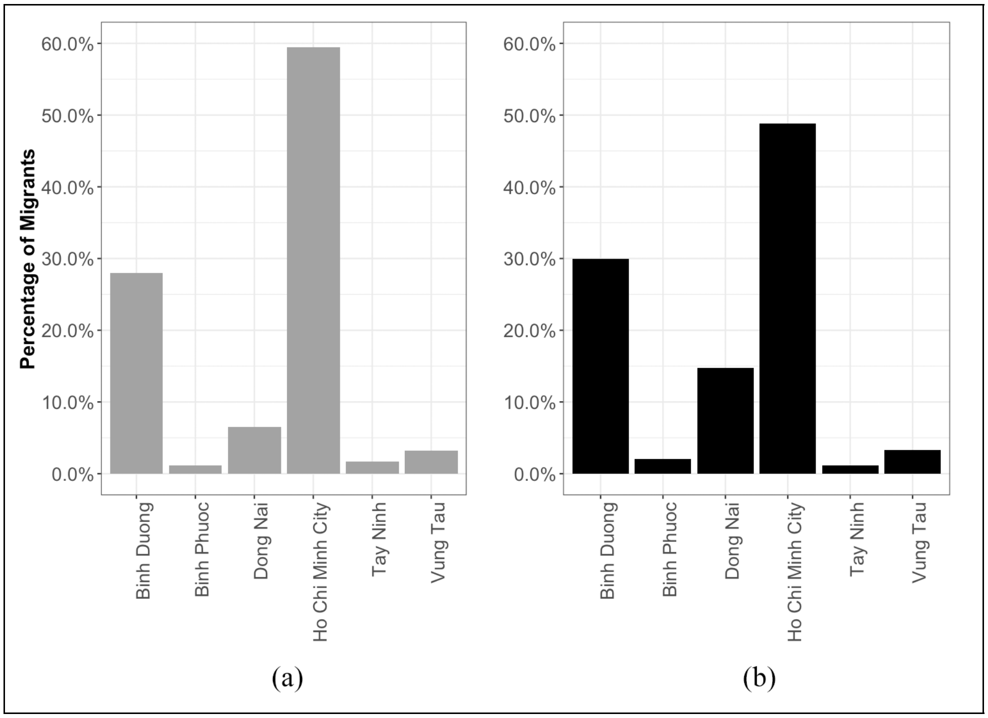

Figure 11. Comparison of in-migration flows to the SE region in a 13-year period from 2005 to 2017: (a) simulation results of in-migration flows to the SE from the MKD region; (b) real data of in-migration flows to the SE from other regions in Vietnam.

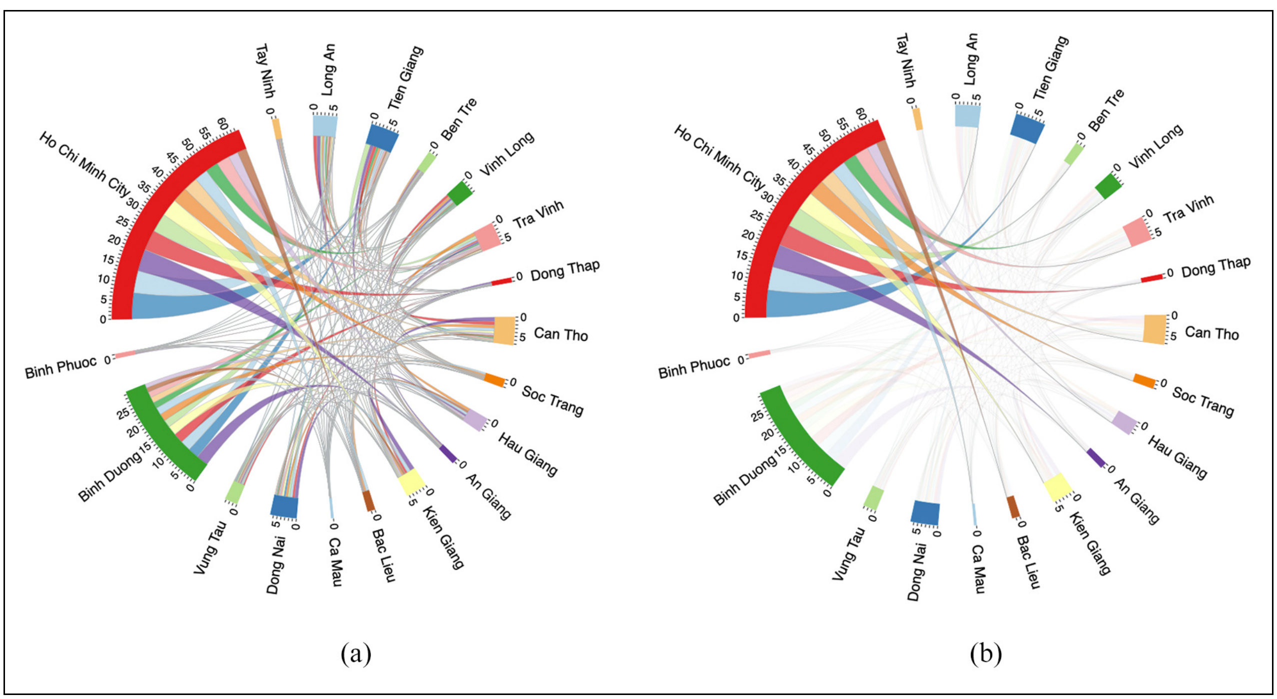

Figure 12. Provincial-level in-migration flows from the MKD to the SE region: (a) flows across all cities and provinces; (b) flows into Ho Chi Minh City as a main migration destination.

from social psychology,[16] to effectively break down the cognition process of individual migration behavior into different components that subsequently allow the inclusion of many different background factors that migrants would consider. To the best of our knowledge, this is the first time that behaviors of migrants have been separated into distinct elements, including the *BA*, *SN*, and *PBC*, in the literature of internal migration in Vietnam.

We found that the *BA* component has the largest contribution in forming the migration intention of people in the MKD region. The *BA* toward migration in this study was the evaluation of four socio-economic and environmental impact factors across potential destinations. Our results were in agreement with the existing literature indicating that economic reasons, which include employment prospects and potential income, are by far the most important factors.[3,4,24,51] The two account for more than 81% of the reasons that people in the MKD region consider when deciding where to migrate to, among the four factors investigated. Education opportunities and environmental impacts were also found to contribute to the migrants' attitude toward migration, with smaller proportions at 13% and 5%, respectively.

The *PBC* was identified as the second most important component in predicting the migration intention and behavior of people in the MKD region. In this study, we mainly focused on an individual's perception of their capability to remove barriers in order to take an actual migration action. We found that people in higher income quintile groups or wealthier people are less likely to migrate. The results generally support findings from the relevant literature indicating that there is some correlation between income and the probability of migration in the MKD region.[24,26] Potential expenditure in the destination and geographical distance were also identified as two of the determinants that discourage people from migrating.

The *SN* has the smallest impact among the three main elements in determining the intention to migrate. However, the *SN* is considered a relatively crucial component, especially with the involvement of both the individual perception of a migrant network's support and the weight of *SN* toward the migration intention. Through the multivariate sensitivity analysis, we found that the combination of these two factors yields greater effects on the final decision of migrants in the MKD region than each factor alone.

Klabunde and Willekens[29] recently indicated that the application of ABM, which has certain advantages over traditional empirical and statistical approaches, would continue to increase and establish a new generation of migration models in the near future. This study is in line with the current research trend of applying agent-based computation models to understand the dynamics of migration behavior.[11] We believe it is the first time ABM has been used to explain the migration decision process and determine the importance of different socio-economic and environmental factors affecting the behavior of migrants in the context of Vietnam and the MKD region specifically. Based on the richness of the output data generated from the agent-based model, we further delineated the migration flows across cities and provinces from the MKD to the SE region.

Klabunde and Willekens[29] also indicated that the validation process in most agent-based migration studies was performed at rudimentary levels and advocated for more work along this line of research. The fact that we implemented a systematic approach to validate our agent-based model further addresses the issue. We used a GA to conduct automated calibration, parameter exploration, and sensitivity analysis as the main validation tools.

This migration study, with the focus on the dynamics of migration in the MKD region, serves as the basis for future work related to a comprehensive internal migration study in Vietnam with seven socio-economic regions. The recent 2015 national migration survey[6] pointed out the variations in perceptions of how people in different regions view a range of factors in their final migration decision. We would like to replicate the migration flows across cities and provinces and further understand the differences regarding the migration behavior of people in distinct regions in Vietnam. Insights and findings from our future study will contribute to the design and evaluation of labor and social policies, including migration decisions.

The natural environment is increasingly recognized as influencing internal migration trends in Vietnam, and especially in the MKD region. Rapid-onset events have contributed to the migration corridor between rural areas in the MKD and more prosperous destinations in the SE region. The impacts of slow-onset events, such as salinization and sea-level rise, are expected to grow and might become a major challenge to the livelihood of people living in the MKD in the future.[26,52] We plan to incorporate relevant data into our future model, in order to explore the correlation between climate change and the dynamic large-scale migration flows in the MKD region with the evolution of time. Predictive model outcomes in "what-if" scenarios could assist local authorities in implementing effective relocation projects to adapt to climate change in the region.

Acknowledgements

The first author would like to acknowledge the support of a scholarship from the Australian Research Council (ARC)'s Industrial Transformation Training Centre to study a PhD degree in Computer Science at the University of Newcastle, Australia.

Funding

This work was supported by the ARC Industrial Transformation Training Centre (IC140100032). The third author was supported by the Spanish Ministry of Science, Andalusian Government, the National Agency for Research Funding AEI, and the ERDF (EU) under grants EXASOCO (PGC2018-101216-B-I00), AIMAR (A-TIC-284-UGR18), and SIMARK (PY18-4475), and the Ramón y Cajal program (RYC-2016-19800).

ORCID iD

Raymond Chiong https://orcid.org/0000-0002-8285-1903

References

1. Lucas RE. Internal migration in developing countries. *Handbk Populat Family Econ* 1997; 1: 721–798.
2. Marx V and Fleischer K. *Internal migration: opportunities and challenges for socio-economic development in Vietnam*. United Nations Vietnam, 2010.
3. Phan D and Coxhead I. Inter-provincial migration and inequality during Vietnam's transition. *J Develop Econ* 2010; 91: 100–112.
4. Kim Anh LT, Hoang Vu L, Bonfoh B, et al. An analysis of interprovincial migration in Vietnam from 1989 to 2009. *Global Health Action* 2012; 5: 9334.
5. GSO Vietnam. The 2014 Vietnam intercensal population and housing survey: migration and urbanization in Vietnam. Technical report, Vietnam News Agency Publishing House, 2015.
6. GSO Vietnam. *The 2015 national internal migration survey: major findings*. Vietnam News Agency Publishing House, 2016, pp.25–36.
7. GSO Vietnam. The 2004 Vietnam migration survey: major findings. Technical report, Statistical Publishing House, 2005.
8. GSO Vietnam. The 2009 Vietnam population and housing census: major findings. Technical report, Central Population and Housing Census Steering Committee, 2010.
9. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. *Proc Natl Acad Sci* 2002; 99: 7280–7287.
10. Macal CM and North MJ. Tutorial on agent-based modelling and simulation. *J Simulat* 2010; 4: 151–162.
11. Billari FC and Prskawetz A. *Agent-based computational demography: using simulation to improve our understanding of demographic behaviour*. Berlin: Springer, 2012.
12. Willekens F. The decision to emigrate: a simulation model based on the theory of planned behaviour. In: Grow A and Bavel JV (eds) *Agent-based modelling in population studies*. Cham, Switzerland: Springer International Publishing, 2017, pp.257–299.
13. Thober J, Schwarz N and Hermans K. Agent-based modeling of environment-migration linkages: a review. *Ecol Soc* 2018; 23(2): 41.
14. Gilbert N. *Agent-based models*. California: SAGE Publications, Inc., 2008.
15. Epstein JM. Remarks on the foundations of agent-based generative social science. *Handbk Cmputat Econ* 2006; 2: 1585–1604.
16. Ajzen I. The theory of planned behavior. *Organiz Behav Hum Decis Proc* 1991; 50: 179–211.
17. Goldberg DE and Holland JH. Genetic algorithms and machine learning. *Mach Learn* 1988; 3: 95–99.
18. Stonedahl F and Wilensky U. Evolutionary robustness checking in the artificial Anasazi model. In: *AAAI fall symposium: complex adaptive systems* (eds M Hadzikadic and T Carmichael), Arlington, VA, 11–13 November 2010, p. 148. California: AAAI – Association for the Advancement of Artificial Intelligence.
19. Stonedahl F and Wilensky U. Finding forms of flocking: Evolutionary search in ABM parameter-spaces. In: *international workshop on multi-agent systems and agent-based simulation* (eds T Bosse, A Geller and CM Jonker), Toronto, Canada, 11 May 2010, pp. 166. Berlin: Springer.
20. Chica M, Barranquero J, Kajdanowicz T, et al. Multimodal optimization: an effective framework for model calibration. *Inform Sci* 2017; 375: 79–97.
21. Moya I, Chica M and Cordón Ó. A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization. *Decis Supp Syst* 2019; 124: 113111.
22. Calvez B and Hutzler G. Automatic tuning of agent-based models using genetic algorithms. In: *international workshop on multi-agent systems and agent-based simulation* (eds JS Sichman and L Antunes), Utrecht, The Netherlands, 25 July 2005, p. 189. Berlin: Springer.
23. Phuong T, Tam N, Nguyet T, et al. Determinants and impacts of migration in Vietnam. *Market Pol Poverty Reduct Vietnam* 2008; 1: 59–92.
24. Coxhead I, Cuong NV and Vu LH. Migration in Vietnam: new evidence from recent surveys. *World Bank Vietnam Dev Econ Discuss Pap* 2015; 2: 1–33.
25. Nguyen-Hoang P and McPeak J. Leaving or staying: Inter-provincial migration in Vietnam. *Asian Pacific Migrat J* 2010; 19: 473–500.
26. Entzinger H and Scholten P. Adapting to climate change through migration: a case study of the Vietnamese Mekong River Delta. Technical report, International Organization for Migration, Geneva, Switzerland, 2016.
27. Fishbein M and Ajzen I. *Predicting and changing behavior: the reasoned action approach*. New York: Psychology Press, 2011.
28. Kniveton D, Smith C and Wood S. Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environ Change* 2011; 21: 34–40.
29. Klabunde A and Willekens F. Decision-making in agent-based models of migration: state of the art and challenges. *Eur J Populat* 2016; 32: 73–97.
30. Muelder H and Filatova T. One theory-many formalizations: testing different code implementations of the theory of planned behaviour in energy agent-based models. *J Artif Soc Simulat* 2018; 21(4): 5.
31. Kniveton DR, Smith CD and Black R. Emerging migration flows in a changing climate in dryland Africa. *Nat Climate Change* 2012; 2: 444–447.
32. Smith CD. Modelling migration futures: development and testing of the rainfalls agent-based migration model–Tanzania. *Climate Dev* 2014; 6: 77–91.
33. Klabunde A, Zinn S, Willekens F, et al. Multistate modelling extended by behavioural rules: an application to migration. *Populat Stud* 2017; 71: 51–67.
34. Nguyen HK, Chiong R, Chica M, et al. Agent-based modeling of inter-provincial migration in the Mekong Delta, Vietnam: a data analytics approach. In: *proceedings of the*

IEEE conference on big data and analytics (ICBDA), Langkawi Island, Malaysia, 21–22 November 2018, p. 104. IEEE.
35. Nguyen HK, Chiong R, Chica M, et al. Agent-based modeling of migration dynamics in the Mekong Delta, Vietnam: automated calibration using a genetic algorithm. In: *2019 IEEE congress on evolutionary computation (CEC)*, Wellington, New Zealand, 10–13 June 2019, p. 172. IEEE.
36. Hassani-Mahmooei B and Parris BW. Climate change and internal migration patterns in Bangladesh: an agent-based model. *Environ Dev Econ* 2012; 17: 763–780.
37. Smith C, Wood S and Kniveton D. Agent based modelling of migration decision-making. In: *proceedings of the European workshop on multi-agent systems (EUMAS-2010)* (eds S Miles and J Moraitis), Paris, France, December 2010, p. 305.
38. Cai R and Oppenheimer M. An agent-based model of climate-induced agricultural labor migration. In: *the Agricultural and Applied Economics Association's 2013 AAEA annual meeting* (eds D Hennessy, J Edward Taylor, BE Roe and M Khanna), Washington, DC, 4–6 August 2013.
39. Global Administrative Areas. Vietnam provincial-level map, https://gadm.org (2012, accessed 14 May 2018).
40. GSO Vietnam. General statistics in Vietnam, http://www.gso.gov.vn/ (2018, accessed 15 December 2018).
41. United Nations International Strategy for Disaster Reduction. DesInventar - Disaster Loss Databases - Vietnam database, https://www.desinventar.net/DesInventar/profiletab.jsp?-countrycode=vnm (2018, accessed 14 May 2018).
42. Luke S, Cioffi-Revilla C, Panait L, et al. MASON: a multiagent simulation environment. *Simulation* 2005; 81: 517–527.
43. Sullivan K, Coletti M and Luke S. GeoMason: geospatial support for MASON. Technical report, Department of Computer Science, George Mason University, 2010.
44. Miller JH. Active nonlinear tests (ANTs) of complex simulation models. *Manag Sci* 1998; 44: 820–830.
45. Oliva R. Model calibration as a testing strategy for system dynamics models. *Eur J Oper Res* 2003; 151: 552–568.
46. Lee JS, Filatova T, Ligmann-Zielinska A, et al. The complexities of agent-based modeling output analysis. *J Artif Soc Soc Simulat* 2015; 18(4): 4.
47. Luke S. The ECJ owner's manual. *San Francisco, California, A user manual for the ECJ Evolutionary Computation Library*, 2010, pp.1–206.
48. Talbi EG. *Metaheuristics: from design to implementation*. New Jersey: John Wiley Sons, 2009.
49. Deb K and Agrawal RB. Simulated binary crossover for continuous search space. *Complex Syst* 1994; 9: 1–15.
50. Ten Broeke G, Van Voorn G and Ligtenberg A. Which sensitivity analysis method should I use for my agent-based model? *J Artif Soc Soc Simulat* 2016; 19(1): 5.
51. Dang A, Goldstein S and McNally J. Internal migration and development in Vietnam. *Int Migrat Rev* 1997; 31: 312–337.
52. Anh DN, Leonardelli I and Dipierri AA. Assessing the evidence: migration, environment and climate change in Vietnam. Technical report, International Organization for Migration, Geneva, Switzerland, 1916.

Author biographies

Hung Khanh Nguyen received his PhD degree in Computer Science from The University of Newcastle, Australia, in 2020, and an MSc degree in Information Technology from The University of Queensland, Australia, in 2013. Currently, he works as a teaching and administrative assistant at The University of Newcastle. He was previously a lecturer with the Foreign Trade University in Vietnam for 4 years.

Raymond Chiong is an associate professor at the University of Newcastle, Australia. He is also a guest research professor with the Centre for Modern Information Management at Huazhong University of Science and Technology, China, and a visiting scholar with the Department of Automation, Tsinghua University, China. He obtained his PhD degree from the University of Melbourne, Australia, and an MSc degree from the University of Birmingham, England. His research interests include evolutionary game theory, ABM, machine learning, and optimization. He is the Editor-in-Chief of the *Journal of Systems and Information Technology*. He is also an Editor for the *Engineering Applications of Artificial Intelligence* journal. To date, he has produced close to 200 refereed publications.

Manuel Chica is a senior researcher at the University of Granada (Computer Science and Artificial Intelligence). He is also the Chief AI Officer at ZIO Analytics, and a conjoint lecturer at the University of Newcastle, Australia. His current research interests include ABM, evolutionary game theory, single and multi-objective evolutionary optimization, data science, and complex systems.

Richard H Middleton completed his PhD in 1987 at the University of Newcastle, Australia. He was a research professor at the Hamilton Institute, The National University of Ireland, Maynooth, from May 2007 to 2011. Between 2015 and 2020, he was the Head of the School of Electrical Engineering and Computing at the University of Newcastle. Currently, he is the Director of the ARC Training Centre for Food and Beverage Supply Chain Optimisation. In 2011, he was the President of the IEEE Control Systems Society. He is a Fellow of the IEEE and of the IFAC, and his research interests include a broad range of control systems theory and applications, including communications systems, control of distributed systems, optimization, and systems biology.

Appendix A

Table 6. The Shapiro–Wilk test of migration flow rates of the one city and 12 provinces in the MKD region.

Province	Net-migration		Out-migration		In-migration	
	W	p	W	p	W	p
An Giang	0.97	0.42	0.94	0.05	0.93	0.02
Bac Lieu	0.95	0.06	0.98	0.68	0.97	0.42
Ben Tre	0.95	0.06	0.95	0.09	0.97	0.47
Ca Mau	0.98	0.77	0.97	0.29	0.90	0.00
Can Tho	0.99	0.96	0.99	0.96	0.97	0.30
Dong Thap	0.99	0.99	0.98	0.84	0.89	0.00
Hau Giang	0.95	0.08	0.97	0.30	0.96	0.23
Kien Giang	0.96	0.14	0.97	0.47	0.97	0.50
Long An	0.97	0.34	0.96	0.22	0.97	0.46
Soc Trang	0.97	0.44	0.95	0.08	0.97	0.37
Tien Giang	0.96	0.12	0.93	0.02	0.97	0.37
Tra Vinh	0.97	0.35	0.94	0.04	0.95	0.09
Vinh Long	0.98	0.84	0.99	0.99	0.95	0.11

Appendix B

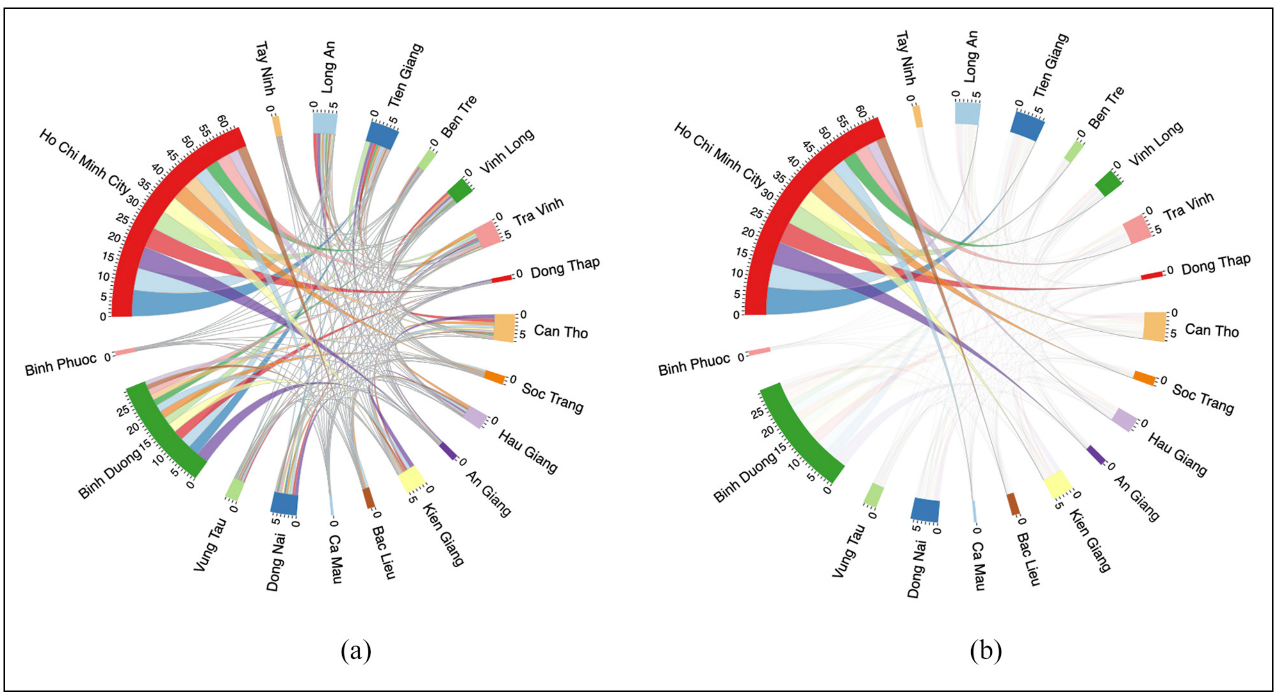

Figure 13. Q-Q plots of the migration flow rates of several provinces in the MKD region.