Journal Title
XX(X):1–19

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

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 dynamics of the migration decision process and determine the most critical factors affecting the behaviour of migrants in the MKD region. We develop an agent-based model and incorporate the Theory of Planned Behaviour 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 behaviour 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, and especially of those participating in globalization and economic integration[1]. Vietnam introduced its economic renovations in 1986, and has since become increasingly integrated into the world markets. This economic integration, however, created unbalanced states of development and living standards among regions or provinces across the country. The inequalities were considered the fundamental driver 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), which has been known to be the most economically developed, with the highest per capita income and higher level of living standard, is the primary migrant-receiving region in Vietnam. The Mekong Delta (MKD), conversely, is the main migration-sending region of the country. Most of the provinces in the MKD region have 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 together 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 results and point out a range of factors influencing migration decisions at the national level. However, there are only a limited number of migration research 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

[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, Australia.
Email: Raymond.Chiong@newcastle.edu.au

form migration intention that leads to the actual migration behaviour in the decision process of the migrants.

In this paper, we aim to study and understand the dynamics of individuals' migration decision-making behaviour 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 identify the most critical components forming the migration behaviour, 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. In addition, we are interested in comparing the patterns of migration behaviour 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 recently become popular in demographic research[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 Behaviour (TPB), which is an established theory from social psychology[16], into our proposed agent-based migration model. The TPB offers a behavioural heuristic approach, which is suitable for deliberate 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 behaviour 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 of the dynamics of internal migration and main migration determinants in Vietnam and the MKD region. Theory of Planned Behaviour and related works are described in Section 3. We then describe our proposed agent-based migration model and the MKD case study in Section 4. After that, the experimental setup and GA for model calibration in Section are explained in Section 5. We presented and discussed the results in Section 6. Finally, we draw a conclusion and future research directions in Section 7.

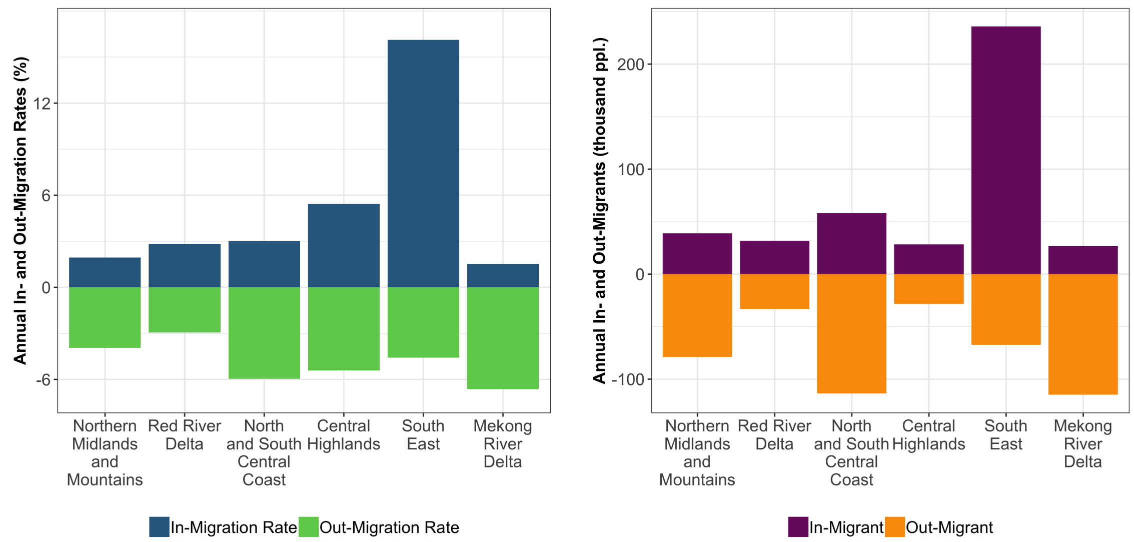
2 Background

2.1 Dynamics of internal migration in Vietnam and the Mekong Delta

Vietnam officially introduced Economic Renovations in 1986, moving from a centrally planned economy with public ownership of production towards a market economy. The transformation has not only led to significant economic growth and poverty but also changed the patterns of internal migration[2,3]. While large urban cities, including Ho Chi Minh City and Hanoi as well as surrounding provinces, have received large levels of industrial capital, other rural areas such as the MKD region 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 the regional migration in Vietnam from 2005 to 2017 are depicted in Figure 1. The SE is the primary migrant-receiving region of 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 favourite region. The MKD, on the other hand, has the highest negative net-migration rate and is the main migration sending region of Vietnam in term 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) in 13 years from 2005 to 2017 (Table 1). Two out of six cities and provinces in the SE region, including Ho Chi Minh City, and Binh Duong province are among those with the highest positive net-migration rates. It is also worth noting that all provinces and cities in the MKD region have negative net-migration rates, indicating that they are all migrant-sending areas. Most of the provinces with the highest negative average net-migration rates in Vietnam in the period are concentrated in the MKD region.



(a) Differential of in- and out-migration rates **(b)** Differential of in- and out-migrants

Figure 1. Differential of migration patterns among regions in Vietnam in 13-year period between 2005 and 2017

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

2.2 *Determinants of migration*

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

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 choose to move for economic reasons, including those who move to look for employment and increase their incomes.

Nguyen-Hoang and McPeak[25], Kim Anh et al.[4], and Coxhead et al.[24] claimed that foreign direct investment (FDI) and industrial zones, which have been distributed unequally throughout the country, underlie the impetus for the internal migration in Vietnam for better employment opportunities. The 2015 internal migration survey indicated that FDI companies and business in the private sector are one of the main sources of employment for migrants[6]. The SE region has received FDI seven times more than in the MKD region at the end of 2017. At the provincial level, four out of five cities and provinces in the SE attracted the most FDI capital in the country. This clearly explains why the SE region has been cited with 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.

Education pursuit. Education has always been cited as one of the popular determinants by the migrants in Vietnam[5,7]. The proportion of migrants who moved because of study purposes has 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.

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, 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.

Household income. The IOM found a correlation between household income and the probability of migration in the MKD region. A survey of more than 1,000 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] depicted the probability of migration at each income level with a sharp decline in the probability of migration when household income increases.

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 Behaviour

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

Three types of beliefs comprising behavioural beliefs, normative beliefs and control beliefs, determine an intention to migrate. According to the TPB, the beliefs people hold towards a migration option are influenced by a combination of different factors including demographic characteristics, social-economic and environmental factors (climate change, natural disaster). These background factors affect the formation of the beliefs, and, indirectly shape the individual migration intention and behaviour[16,27].

Firstly, behavioural beliefs are considered the motives or reasons that people choose to migrate. The evaluation of these beliefs in different outcomes that migration yields, are the bases to form the BA toward migration. These behavioural 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 the choices related to migration. Supports from peers and their destination preferences can influence how a person perceives certain migration option and is subsequently willing 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 the migration intention. The PBC is the individual's perception of their capability to take advantage of facilitators and to remove barriers to make an actual migration action.

The BA, SN and PBC are indicators of the migration intention. The stronger the components are, the higher the possibility the individual undertakes the migration behaviour. In addition, Ajzen[16] introduced the concept of actual behavioural control to improve the prediction of a certain behaviour, i.e., migration, which partly depends on factors that a person does not have complete control. Ajzen[16] and Fishbein and Ajzen[27] used the PBC as a proxy to measure the actual behavioural control.

3.2 Agent-based migration models

There has been a growing interest in the integration of the TPB with ABM to study migration behaviours[29,30]. Kniveton et al.[28,31] firstly developed an agent-based model on the basis of the TPB to study climate change-driven migration in Burkina Faso. The attitude toward migration is defined as the probability of an individual with certain demographic characteristics. The SN is a function to assign the values 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. Agents perform the intention score for each of the migration options, and the one with the highest score is chosen.

Smith[32] developed an 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 individual migration intention was driven by the characteristics of age, gender, migration experience and the social network, it is also be 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 also determined through survey data.

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

Nguyen et al.[34,35] recently integrated the TPB into their proposed agent-based model to explore the dynamics of migration flows across the MKD region. Both models were calibrated with the actual data of in-, out-, and net-migration rates of cities and provinces in the region. The prior model was manually calibrated with three parameters, including BA, SN and PBC. Nguyen et al.[35] refined the previous 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 relating to socio-economic attributes. In this paper, we extend the agent-based model of Nguyen et al. with relevant empirical data to determine the most critical components and factors that would affect the final migration decision of migrants in the MKD region.

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, 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 behaviour towards 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 the person agent currently resides. The migration intention, I, has three core components: behavioural attitude BA, subjective norms SN, and perceived behavioural control PBC. Agent i performs the intention calculation, $I_{i,j}(t)$, for each province agent j at time t as per the following equation:

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

Behavioural attitude. The BA of a person agent towards the migration assessment is assumed to be an outcome of a 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]:

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

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 the agent is currently allocated and another province agent j. Values of $\overline{emp}$, $\overline{inc}$, and $\overline{edu}$ of each province are normalized in the range of $[0, 1]$ 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 in 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

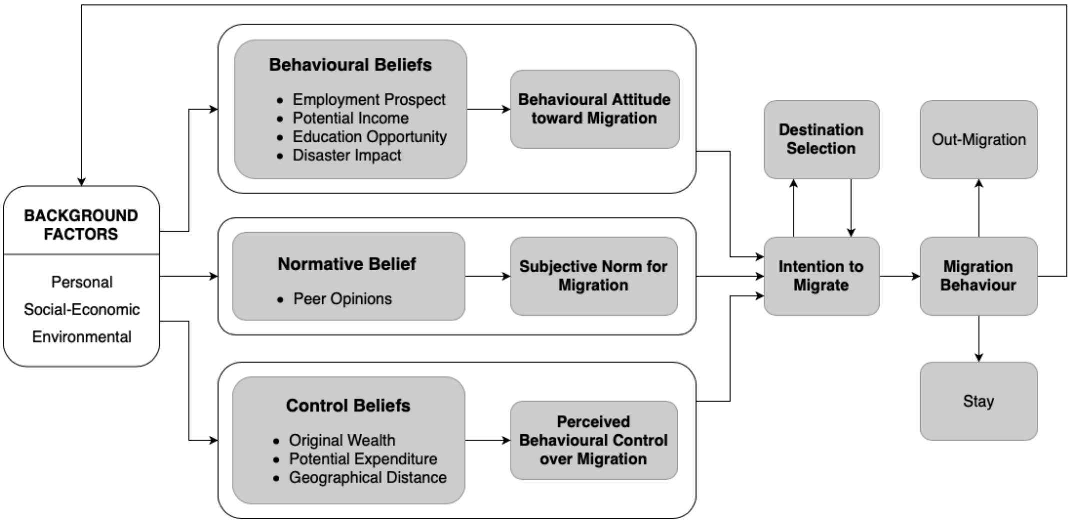


Figure 2. Individual cognition of the migration decision process

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 towards 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 including 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 $P(x, \lambda)$. x is the observed occurrence of hazards, and λ is the expected number of hazard events in a given time interval[36].

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

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

Subjective norm. $SN_i(t)$ is updated at the same time as the BA is calculated. In this model, an individual agent is also modelled that they might adapt their behaviour accordingly to their neighbour's final migration decision.

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

The SN is basically calculated as the proportion of agent i's neighbours who have migrated[33]. It reflects the assistance of these neighbours 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 neighbours of agent i are those located within i's neighbourhood area, determined by a fixed range parameter, u.

Perceived behavioural control. The $PBC_{i,j}^k(t)$ that 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 towards province agent j is the migration cost. In this model, the migration cost is determined by the living expenditure $exp_{i,j}(t)$ in the destination, province j. Lastly, 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]:

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

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 of $[0, 1]$ by dividing 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. $\overline{dis}_{i,j}$ is computed as the normalized distance between the centroids of two province agents. Each parameter $\hat{\delta}_i^n$ is assigned to each person agent i following a uniform distribution between $[0, \delta^n]$ to reflect the heterogeneity among person agents' perception on factors that facilitate or hinder their migration option.

By calculating the intention score following Equation 1, each person agent stores the intention values of the corresponding provinces. The person agent then develops their final migration behaviour through a destination selection process and decides whether to migrate or stay. The agent is modelled to only consider the list of provinces with higher intention values than the value yielded from 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 properly account for the uncertainties in the actual migration behaviour[12].

If an agent decides to migrate, they inform their choice to the current surrounding agents. Those neighbouring agents update their number of migrants in their network, $migr_i(t)$, with an increase of one. The agent who migrates to another province is randomly located in the geographical boundary of that province. This person agent establishes a new neighbourhood and notifies the other agents residing in this area of a new neighbour. It is also assumed that the agent remains in the same income quintile group and having similar living expenditure.

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 modelled to remain alive during the simulation runs, since a net population growth rate is applied. The number of in-ward and out-ward migrants in each province agent, which are endogenously generated from the model, prescribe the flow of net-migrants.

4.2 Case study of migration in the Mekong Delta

There are 19 cities and provinces in the MKD and the SE region. Among them, 13 cities and provinces are from the MKD, and the remaining are located in the SE region. Their locations and geographic boundaries, which were extracted from Vietnam's provincial-level map[39] are shown in Figure 3. Islands of the selected provinces and cities are not included.

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

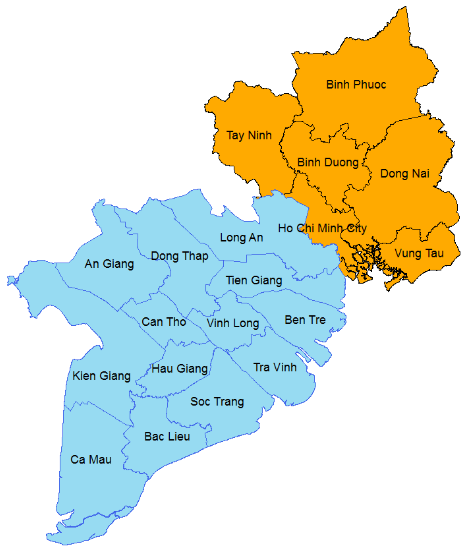


Figure 3. Locations and geographical boundaries of cities and provinces in the MKD and SE region

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

Table 2. Socio-economic development level, environmental impacts attributes and the contributing factors

Attribute	Factors
emp	Number of FDI projects (fdi) Number of business (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 of 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

As discussed, employment opportunity is the most important reasons among migrants in Vietnam. We consider five indications representing the employment prospects for each province.

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

$\hat{\theta}$ are the weighted values of the five θ parameters. $\theta \in [0,1]$ are used as calibration parameters of the agent-based migration model. The employment opportunities generated from the FDI companies, which is represented by θ^1, are considered the main impetus for the internal migration in Vietnam and specifically for the migration flow from the MKD to the SE region[4,24,25]. We choose to fix $\theta^1 = 1$ to maintain the persistence of the calibration results related to these θ values.

We assumed that three factors, which are the number of pupils of general education, and the number of students and lecturers in universities and colleges, contribute equally to account for the education attribute, edu. For the income attribute, inc, it is important to note that while we use the general average monthly income per capita to calculate BA in Equation 2, we compute the PBC in Equation 5 with the 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 is available across cities and provinces in Vietnam.

The environmental impacts data 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, $hazards_j$, is calculated as the average of all hazards occurred from 1989 to 2015. The vulnerability index, $vul_j \in [0,1]$, of each province is accounted by the number of dead, injured, missing people and the number of damaged and destroyed houses accumulated by the occurrence 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 MASON framework[42]. The geographic boundary and location of provinces and cities were extracted in the shapefile format using geoMASON[43].

The model was initialized with 3,340 person agents distributed across 13 provinces and cities in the MKD region, representing the 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 province. The location of each person agent was assigned randomly in the corresponding region, and parameter u was set so that the neighbourhood of each person agent is defined within 10km from their location. We did not initially populate province agents representing six cities 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. Each person agent also had their own perception of how different socio-economic and environmental attributes impact their migration decision-making process. The weights of these attributes were initialized independently for each person agent and kept constant during the run but changed for different Monte-Carlo (MC) trials.

We ran the simulation for 156 steps, with each step representing one month. The entire simulation was for a 13-year period, from 2005 to 2017. Since the data of province agents' attributes, including emp, inc, pov, edu and exp, is only available annually, it was assumed that those attributes have the same values during the 12 months of the year. The environmental impact attribute, env, was updated endogenously for every month.

Each person agent was assumed to make migration decisions twice a year: in June and December. Each province agent will also 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 Genetic algorithm of automated calibration

We implement automated calibration as one of the main validation tools[20,44,45] with the use of GA[17]. Automated calibration is a computationally intensive process that aims to fit simulated data to real-world data. The evaluation of the model fitting is done by iteratively running the computational model, and then tuning the parameters in order to identify a set of parameters that best match the reference data[45]. In this study, we discover the best values for all calibration parameters to obtain the best fit with real migration data at hand. Table 3 shows the set of 15 parameters, which are in the range $[0,1]$, to be calibrated.

A GA, in a nutshell, involves the evolution of a population of solutions, each of which represents a set of model's parameter. These solutions are iteratively evolved until the best possible set are found for a given modeling goal. GA has been proved to be a useful tool for an automated calibration

Table 3. Set of 15 calibration parameters in the agent-based migration model in the context of MKD region

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	

process in different kinds of non-linear models, including agent-based models[18–20,46].

In this study, we used the ECJ library[47] to implement a GA. ECJ is a Java-based evolutionary computation package that can be utilized to integrate with the MASON framework. Since conducting the automated calibration of an agent-based model is a computationally expensive process, we coded the model with multi-threading technique so that the whole process can be executed in multi-threads over High Performance Computing.

We implemented a standard generational GA, where every generation of the population replaces the previous one[17]. The generational GA has a population of 100 real-coded chromosomes comprised of the 15 parameters defined in Table 3. All calibration parameters were initialized with random values following uniform distribution within the range $[0, 1]$.

A 3-tournament selection mechanism[48] with weak elitism was applied, which means the best parent 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 also used 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 also ran the overall GA calibration model 20 times per model (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 GA calibration method returns the fitness results.

The fitness 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 then identifies a set of calibration 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:

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

where V_M and V_R are the vectors of three migration flow rates generated from the simulation model and attained from the historical data in each province. Due to the stochastic nature of the model, the value of L^2 is calculated as the average of fitting errors in all 40 MC runs. The final model error measure ϵ is the sum of 13 calculated L^2 values accordingly to the 13 provinces and cities in the MKD region.

6 Experimental results

6.1 Model calibration results

The GA calibration of the agent-based migration model ends with a Euclidean mean value of 50.6625 and a standard deviation of 10.682. Table 4 shows the list of 15 calibrated parameters using the standard generational GA for the agent-based migration model. This set of parameter values returns the smallest error measure of $\epsilon = 37.9587$, which is 79.468% close to the reference data and therefore, was chosen to run the final model.

Table 4. Calibrated parameters using the GA

Component	Calibrated Parameter Values				
Behavioural Attitude	α^1	β^1	β^2	β^3	β^4
	0.0013	0.9452	0.9953	0.2512	0.1069
Subjective Norm	α^2	γ			
	0.008	0.6956			
Perceived Behavioural Control	α^3	δ^1	δ^2	δ^3	
	0.0078	0.4077	0.2921	0.8107	
Employment Attribute	θ^2	θ^3	θ^4	θ^5	
	0.1419	0.5421	0.0585	0.5927	

We examined the normality of the results with Shapiro-Wilk's test, shown in Table 6 in Appendix A. In most

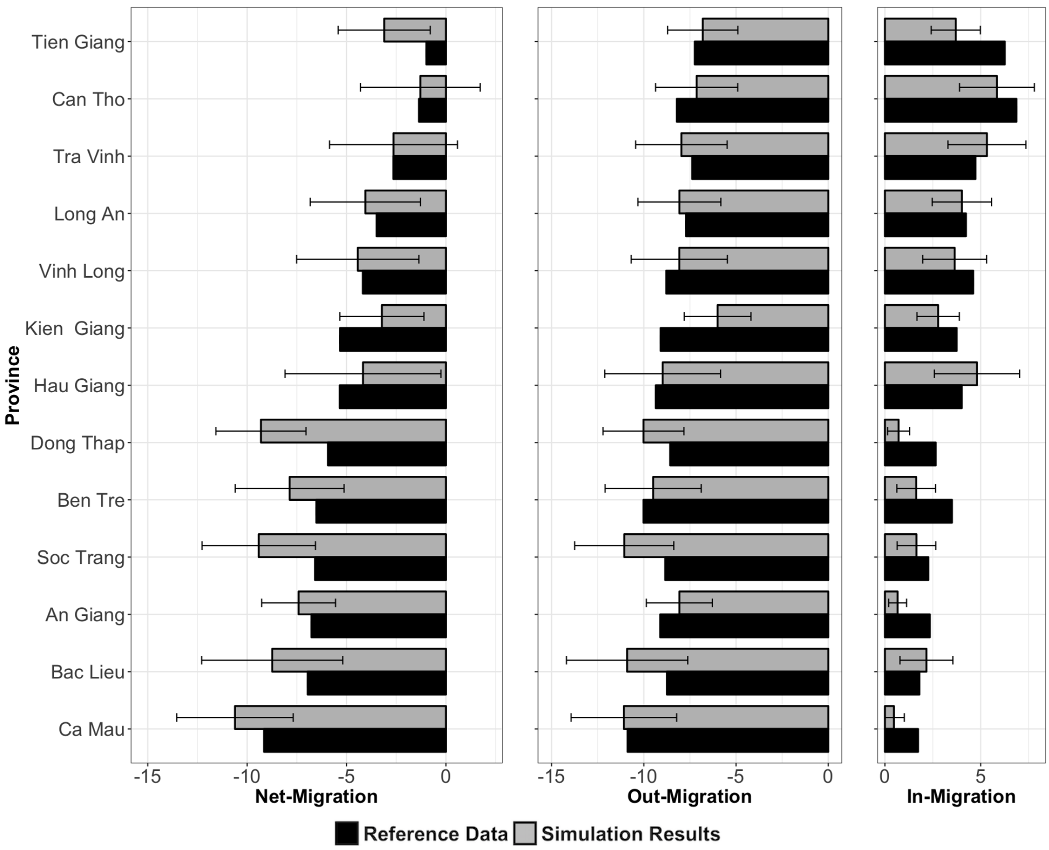


Figure 4. Comparison of simulated results and actual data of the average net-, out-, and in-migration rates of provinces in the MKD region from 2005 to 2017

cases, the p-value > 0.05 implies that the distribution of the result is not significantly different from the normal distribution. Therefore, we can assume that the relevant results are normally distributed. In several cases, where p-values < 0.05, the normality is visually checked with Q-Q plots. We found that the simulation outputs in these cases are not significantly different from the normal distribution too.

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

Figure 4 indicates that most of the observed values of in-, out- and net-migration rates across the cities 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 the top and bottom half of the chart. The only exception is Dong Thap province.

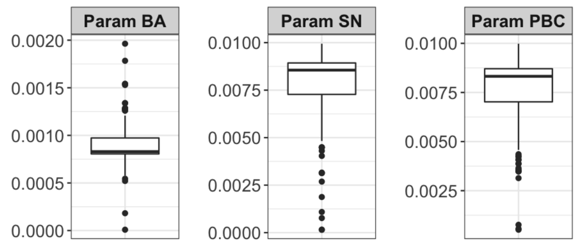


Figure 5. Boxplots showing the values of α^1, α^2 and α^3 from the best GA calibration solutions

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 provinces and cities are within the 99% confidence interval of the mean results generated from the agent-based model. On the in-migration chart, the simulation model is also able to replicate the dynamics of migration flow across the MKD region. There are only five provinces in which the in-migration rates are outside the intervals without any significant difference from the reference values.

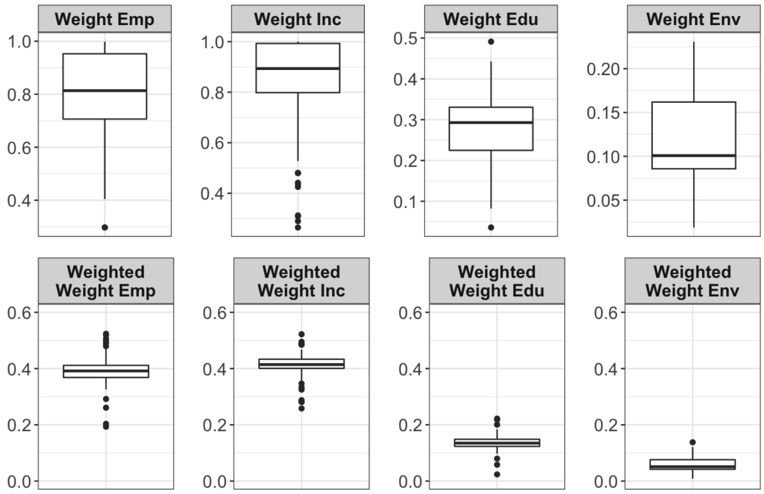


Figure 6. Boxplots showing the values of β^1, β^2, β^3 and β^4 from the best GA calibration solutions and the weighted values of relevant parameters

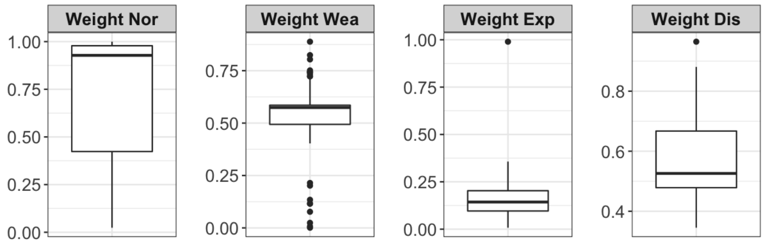


Figure 7. Boxplots showing the values of γ, δ^1, δ^2 and δ^3 from the best GA calibration solutions

6.2 Model parameter exploration

Together with calibration, parameter exploration is a crucial ingredient for model testing and verification[45]. The GA has the advantage of generating 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, the modeller can move forward to a simpler and easier to understand model settings[20].

We observed the ranges of calibrated parameter values to better understand the impact of these parameters on the final decision of migrants in the MKD region. We focused on the 100 sets of solutions that produce the smallest values of error measure ϵ. Boxplots in Figure 5, Figure 6 and Figure 7 represent the distribution of 11 among total 15 calibration parameters obtained from the best solutions 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, 1, setup to implement the GA. This is expected since the values of the three parameters are utilized to calculate the behavioural intention (Equation 1), which determines the probability that people decide to migrate. According to Table 1, the migration flow rates (number of migrants per 1,000 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 the weighted values of these parameters. It is worthy to note 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.

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 values among the β values. This result is also in agreement with the literature indicating the expected income differential between the origin and destination is crucial for migration in the MKD region[24].

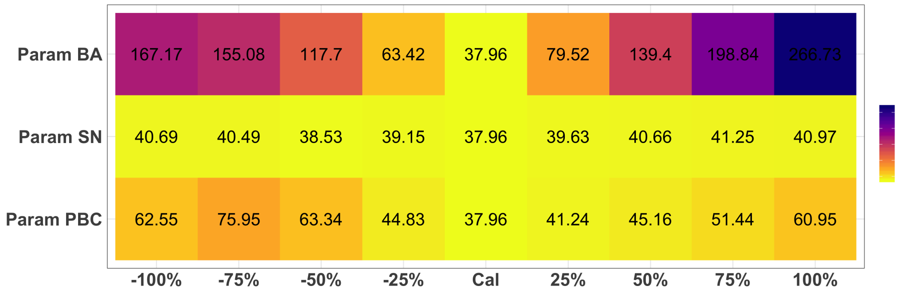


Figure 8. Heatmap indicating the sensitivity analysis on three parameters α^1, α^2, α^3

The third boxplot indicates the distribution of good values of the weighted β^3 parameter, representing education opportunity attribute. The median value is close to 0.1362, which can be assumed that 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 parameters are close to 0.05. This result suggests there is a relatively small impact of the climatic hazards, approximately 5% contributing to the attitude of person agents towards the internal migration. The result, in general, supports the recent survey that 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 Equation 4 and Equation 5. The median value of γ, which is 0.928, indicates that people in the MKD region have a large variance in their perception regarding the influence of neighbours on the migration decision.

In the remaining three boxplots, we see that most of the values of the three δ parameters are larger than zero. For the original wealth and the potential expenditure, the higher the value of δ^1 and δ^2, the less likely that people in high income quintile would move to a different destination. 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 the geographical distance is a critical determinant in the decision of the migrants from the MKD region.

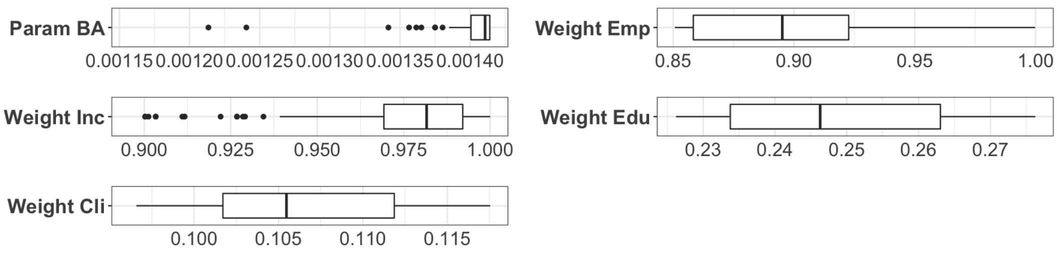
6.3 Sensitivity analysis

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

The heatmap indicates that varying the calibrated values of the parameter BA (α^1) causes the most significant change in the model result. 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 value of parameter α^2 ($\epsilon = 40.97$) and α^3 ($\epsilon = 60.95$). We find a similar pattern in the case when values of three α parameters are each set to 0. The change of parameter of BA results in the highest value of error ($\epsilon = 167.17$), which is four times and nearly three times larger than the error produced by the change in the parameter of SN and PBC. Besides, the results suggest the SN has the least impact in determining the intention among the migrants in the MKD region. Changing SN parameter values does not yield significant differences from the smallest $\epsilon = 37.9587$.

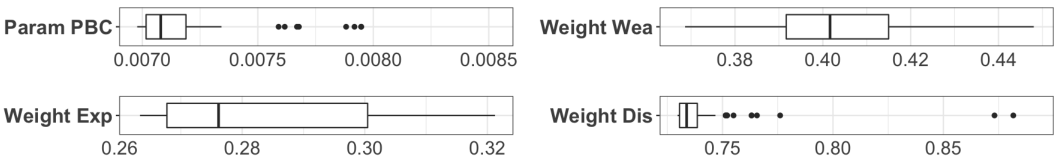
Multivariate sensitivity analysis. We also employed GA to perform sensitivity analysis following Stonedahl and Wilensky[18]'s approach by modifying a group of parameters[50]. This approach is a multivariate sensitivity analysis using the GA to maximize (rather than minimize) the error measure ϵ. The search space is constrained to a limited range within $\pm 10\%$ of the calibrated parameter values[18,44,46]. The upper bound, which is set to 1, was applied to the variations in the weights of employment prospect and potential income.

We conducted sensitivity analysis on four groups of parameters (indicated in Table 4) including those parameters related to BA, SN, PBC and all three prior groups. In each sensitivity analysis, values of related parameters are varied within the pre-defined range ($\pm 10\%$) while the values of

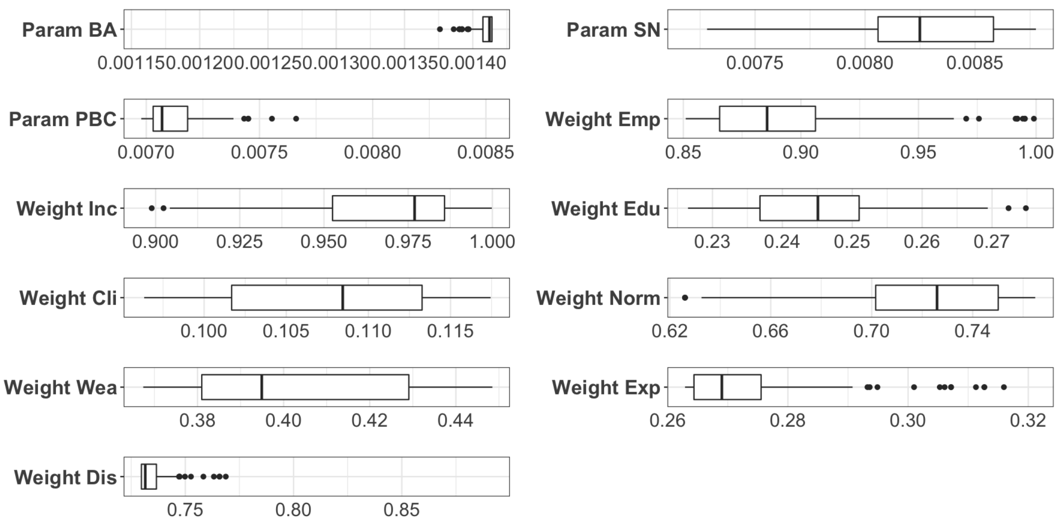


(a) Parameters related to the BA

(b) Parameters related to the SN



(c) Parameters related to the PBC



(d) Parameters belonging to three above groups

Figure 9. Boxplots showing values of different groups of parameters obtained from the worst GA calibration solutions

remaining parameters are fixed. 100 worst sets of parameters were chosen in each sensitivity analysis to examine and plot the graphs. Results for the analysis can be seen in corresponding groups of boxplots in Figure 9. The outer box defines the parameter ranges while the inner boxplots indicate the distribution of the relevant parameters obtained from the worst solutions.

The different GA searches found clear trends in different groups of parameter settings. The GA consistently explores high values for the parameter of BA, while it selects low values for the parameter of PBC and the weight of distance. This also means the agent-based migration model is particularly sensitive to these factors. The other parameters' values are found 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 the error measurement ϵ that are yielded from the multivariate sensitivity analysis by four groups of parameter settings. The summaries are displayed in Table 5. We found that the error generated from the group of SN parameters is not significantly smaller than the errors produced from the groups of BA and PBC in all five quintiles. This result denotes the importance of the SN component in forming the migration intention, especially with the combination of the two relevant parameters.

Table 5. Statistical summary of group of parameters obtained from all worst GA calibration solutions

Component	Min	1st Quin.	Median	Mean	3rd Quin.	Max
Behavioural 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 Behavioural 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

6.4 Analysis of the migration dynamics

Based on the final run of the agent-based model with the set of calibration parameters that generated the smallest error measure ϵ, we further explored relevant results and validated our findings. We firstly examined the proportion of migrants in the MKD region by income quintile, which is shown in Figure 10.

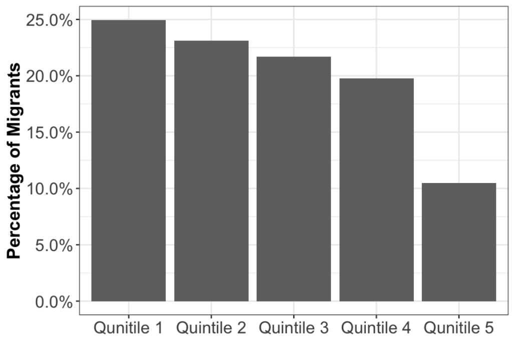


Figure 10. Percentage of total number of migrants from the MKD region by income quintiles

The simulation result shows that people in higher income quintile groups tend not to migrate. The proportions of migration 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 migrants. This result together with the exploration of the calibrated values of two δ^1 and δ^2 parameters advocate the empirical findings by Entzinger and Scholten[26] and Coxhead et al.[24] that high income and high expenditure in the destination discourages people from moving, since migration might not be as beneficial as their current living conditions are.

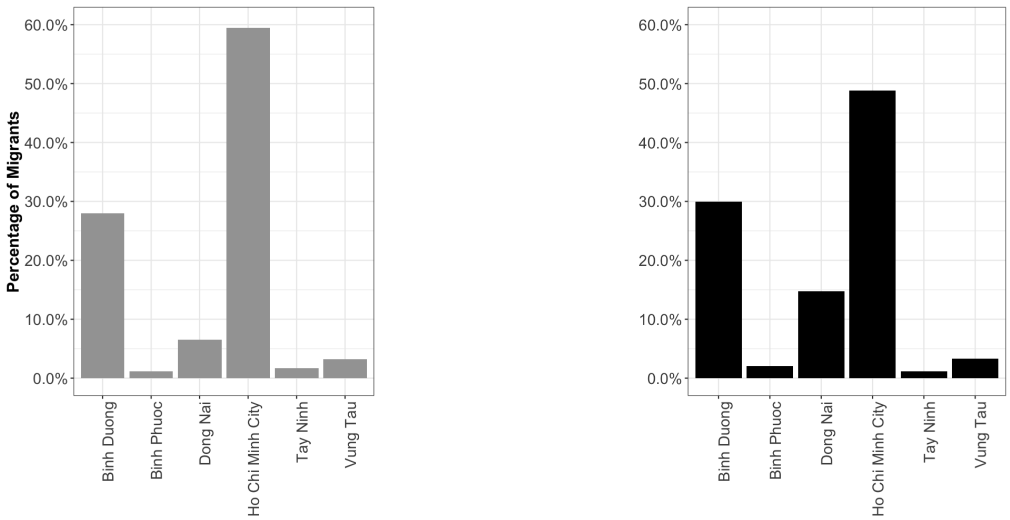
Secondly, we determined the most preferred destinations in the SE region among the migrants from the MKD region. The left bar graph in Figure 11 shows the simulation results regarding the proportion of in-migrants toward the six potential destinations during the 13-year period. The right bar chart displays the actual data of the percentage of in-migration flows to the same destinations but from all regions across Vietnam. There is no available provincial-level data with respect to the migration flows from the MKD region, specifically to the SE region. However, we find that the simulation results are able to replicate the dynamic of in-migration flows towards the cities and provinces in the SE region. Ho Chi Minh City, Binh Duong and Dong Nai province 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).

The top chord diagram in Figure 12 shows the simulated migration flows across the MKD and the SE region and the annual average number of in-migrants in each city and province between 2005 and 2017. We clearly see that Ho Chi Minh City and Binh Duong province are the most two attractive destinations, receiving on average almost 63,000 and 30,000 people migrating from the neighbouring region each year. In the MKD region, most cities and provinces receive less than 5,000 people, which are considered as intra-regional migrants, in each year. The least preferred destination among the migrants is Ca Mau, which is the southernmost of Vietnam's provinces, with only more than 520 people on average moving to each year.

The bottom diagram in Figure 12 highlights the in-migration flows of Ho Chi Minh city as the most favourite destination among people in the MKD region. Main sources of migration towards Ho Chi Minh City are from Tien Giang, Long An, An Giang and Dong Thap province. Similar patterns are also found in the in-migrations flow into Binh Duong province.

7 Conclusion and future work

In this paper, we studied the dynamics of migration decision of people in the MKD region. We took advantage of the TPB, which is a well-known theory derived from social psychology[16] to effectively break down the cognition process of individual migration behaviour into different components that subsequently allows the inclusion of many different background factors that the migrants consider. To the best of our knowledge, among an extensive literature on the internal migration in Vietnam, it is the first time the behaviour of the migrants has been separated into distinct elements including the BA, the SN and the PBC.



(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 11. Comparison of in-migration flows to the SE region in 13 years period from 2005 - 2017

We found that the BA component has the largest contribution in forming the migration intention of those people in the MKD region. The behavioural attitude towards migration in this study is the evaluation of four socio-economic and environmental impact factors across potential destinations. Our results are in agreement with the existing literature indicating that the economic reasons which include employment prospect and potential income are by far the most important factors[3,4,24,51]. The two, among the four factors, together account for more than 81% of the attitude that the people in the MKD region consider where to migrate to. Education opportunity and environmental impact factor are also found to contribute to the migrant's attitude towards migration with smaller proportion at 13% and 5% respectively.

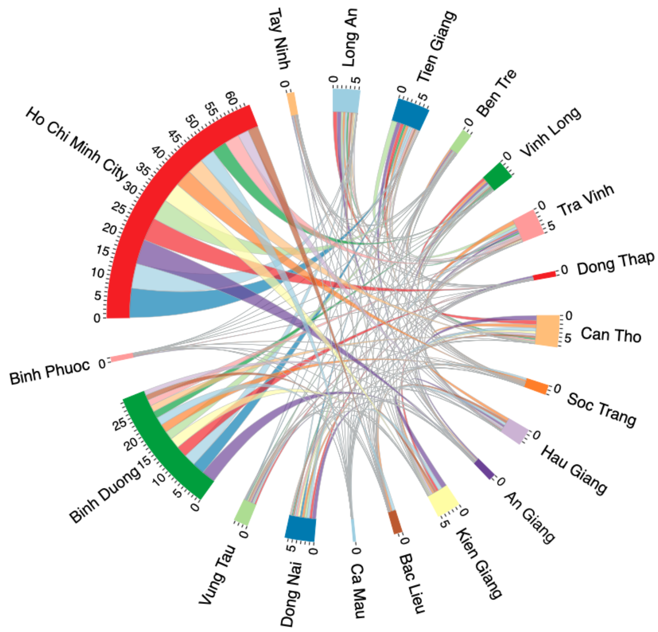
The PBC is the second most important component in predicting the migration intention and behaviour of people in the MKD region. In this study, we mainly focused on the individual's perception of their capability to remove barriers to make an actual migration action. We found that the people in higher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

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

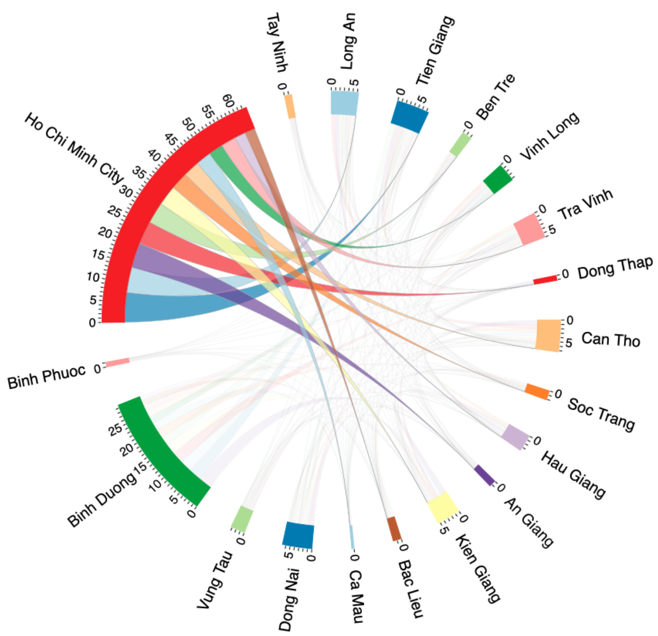
Klabunde and Willekens[29] recently indicated that the application of ABM, which has advantages over tradition empirical and statistical approaches, would continue to increase and establish a new generation of migration models in the future. This study is in line with the current research trend of applying agent-based computation models to understand the dynamics of migration behaviour[11]. We believe it is the first time ABM has been used to explain the migration decision process and justify the proportion of different socio-economic and environmental factors affecting the behaviour 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 delineate 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 had been performed at rudimentary levels and advocated more works in this matter. The fact that we implemented a systematic approach to validate our agent-based model further addresses the issue. We used a GA to conducted automated calibration, parameter exploration and sensitivity analysis as the main validation tools.

This migration work with the focus on the dynamics of migration in the MKD region serves as the basis for future



(a) Flows across all cities and provinces



(b) Flows into Ho Chi Minh city as main migration destination

Figure 12. Provincial-level in-migration flows from the MKD to the SE region

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 towards ranges of factors in their final migration decision. We aim to replicate the migration flows across cities and provinces and further understand the differences regarding the migration behaviour of people in distinct regions in Vietnam. Insights and findings from the future study will contribute to the design and evaluation of labour 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 the rural areas in MKD and large cities 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 the future model with relevant data aiming to explore the correlation between climate change, and the dynamic large-scale migration flows in the MKD region. Predictive outcomes of the model could assist local authorities in implementing effective relocation projects to adapt to climate change in the region.

References

[1] Lucas RE. Internal migration in developing countries. *Handbook of population and family economics* 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. *Journal of Development Economics* 2010; 91(1): 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(1): 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. In *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. *Proceedings of the national academy of sciences* 2002; 99(suppl 3): 7280–7287.

[10] Macal CM and North MJ. Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.* IEEE, pp. 14–pp.

[11] Billari FC and Prskawetz A. *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour*. Springer Science & Business Media, 2012.

[12] Willekens F. The decision to emigrate: A simulation model based on the theory of planned behaviour. In *Agent-Based Modelling in Population Studies*. Springer, 2017. pp. 257–299.

[13] Thober J, Schwarz N and Hermans K. Agent-based modeling of environment-migration linkages: A review. *Ecology and society* 2018; 23(2).

[14] Gilbert N. *Agent-based models*. SAGE Publications, Inc., 2008.

[15] Epstein JM. Remarks on the foundations of agent-based generative social science. *Handbook of computational economics* 2006; 2: 1585–1604.

[16] Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 1991; 50(2): 179–211.

[17] Goldberg DE and Holland JH. Genetic algorithms and machine learning. *Machine learning* 1988; 3(2): 95–99.

[18] Stonedahl F and Wilensky U. Evolutionary robustness checking in the artificial anasazi model. In *AAAI Fall Symposium: Complex Adaptive Systems*. pp. 120–129.

[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*. Springer, pp. 61–75.

[20] Chica M, Barranquero J, Kajdanowicz T et al. Multimodal optimization: An effective framework for model calibration. *Information Sciences* 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. *Decision Support Systems* 2019; : 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*. Springer, pp. 41–57.

[23] Phuong T, Tam N, Nguyet T et al. Determinants and impacts of migration in Vietnam. *Market, Policy and Poverty Reduction in Vietnam* 2008; : 59–92.

[24] Coxhead I, Cuong NV and Vu LH. Migration in Vietnam: New evidence from recent surveys. *World Bank: Vietnam Development Economics Discussion Papers* 2015; (2).

[25] Nguyen-Hoang P and McPeak J. Leaving or staying: Inter-provincial migration in Vietnam. *Asian and Pacific Migration Journal* 2010; 19(4): 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*. 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 Environmental 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. *European Journal of Population* 2016; 32(1): 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. *Journal of Artificial Societies and Social Simulation* 2018; .

[31] Kniveton DR, Smith CD and Black R. Emerging migration flows in a changing climate in dryland Africa. *Nature Climate Change* 2012; 2(6): 444–447.

[32] Smith CD. Modelling migration futures: Development and testing of the rainfalls agent-based migration model–Tanzania. *Climate and Development* 2014; 6(1): 77–91.

[33] Klabunde A, Zinn S, Willekens F et al. Multistate modelling extended by behavioural rules: An application to migration. *Population Studies* 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)*. IEEE, pp. 27–32.

[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)*. IEEE.

[36] Hassani-Mahmooei B and Parris BW. Climate change and internal migration patterns in Bangladesh: An agent-based model. *Environment and Development Economics* 2012; 17(6): 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)*.

[38] Cai R and Oppenheimer M. An agent-based model of climate-induced agricultural labor migration. Presented at the Agricultural and Applied Economics Associations 2013 AAEA Annual Meeting, Washington DC. August 46, 2013.

[39] Global Administrative Areas. Vietnam provincial–level map. `www.gadm.org`, 2012. Accessed: 2018-14-05.

[40] GSO Vietnam. General statistics in Vietnam. `http://www.gso.gov.vn/`, 2018. Accessed: 2018-12-15.

[41] United Nations International Strategy for Disaster Reducetion. DesInventar - Disaster Loss Databases - Vietnam database. `https://www.desinventar.net/DesInventar/profiletab.jsp?countrycode=vnm`, 2018. Accessed: 2018-14-05.

[42] Luke S, Cioffi-Revilla C, Panait L et al. MASON: A multiagent simulation environment. *SIMULATION* 2005; 81(7): 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. *Management Science* 1998; 44(6): 820–830.

[45] Oliva R. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research* 2003; 151(3): 552–568.

[46] Lee JS, Filatova T, Ligmann-Zielinska A et al. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation* 2015; 18(4).

[47] Luke S. The ECJ owners manual. *San Francisco, California, A user manual for the ECJ Evolutionary Computation Library* 2010; : 1–206.

[48] Talbi EG. *Metaheuristics: From Design to Implementation*. John Wiley & Sons, 2009.

[49] Deb K and Agrawal RB. Simulated binary crossover for continuous search space. *Complex Systems* 1994; 9(3): 1–15.

[50] Ten Broeke G, Van Voorn G and Ligtenberg A. Which sensitivity analysis method should I use for my agent-based model? *Journal of Artificial Societies and Social Simulation* 2016; 19(1).

[51] Dang A, Goldstein S and McNally J. Internal migration and development in Vietnam. *International Migration Review* 1997; 31(2): 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, 2016.

Appendix A

Table 6. Shapiro-Wilk's test of migration flow rates of the 13 cities and 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