

Agent-based Modeling of Migrant Workers Residential Dynamics within a Mega-city Region: the Case of Pearl River Delta, China

CINZIA LOSAVIO and JUSTE RAIMBAULT
UMR CNRS 8504 Géographie-cités

Abstract

This paper introduces an agent-based model of regional migration dynamics

Keywords : *Mingong ; Residential Dynamics ; Agent-based Modeling ; Zhujiang Delta Mega-city Region*

1 Introduction

Over the last three decades, rural-to-urban migrant-workers have been a driving force for China’s economy, raising attention on associated socio-economical issues. However, the importance of their economic diversity and social mobility has been poorly considered in the analysis of urban development strategy. We use an agent-based model to simulate residential dynamics of migrants in Pearl River Delta (PRD) mega city region, taking into account the full range of migrants’ socio-economical status and their evolution. Mega-city regions have become a new scale of Chinese State regulation, and PRD represent the most prosperous and dynamic one in term of migration waves, standing as an ideal unit of analysis. Our model unveils emergent patterns of dynamics, from micro behavior rules of discrete mobility choices. These choices are conditional to urban and economic environment, which evolution is controlled by meso-scale independent dynamics. The two scales are coupled through the dependence of discrete choice utilities to generalized accessibility that combines patch-level urban and economic context with a feedback of the dynamics themselves. This multi-scale aspect is crucial to distinguish endogenous from exogenous effects in regional migration patterns. We perform simulations to internally validate the model on synthetic data, by assessing statistical consistence and establishing phase diagrams across the parameter space. The application to the case study allows first to test how variation in socio-economic status yield more complex trajectories, and secondly to identify how the Party-State persist in controlling internal migration flows in a more sophisticated and strategically redefined way. Further work is directed towards a qualitative external validation of the model, by calibrating free parameters to reproduce meso-scale stylized facts, in order to guide interpretations of emergent outputs and potential policy applications.

1.1 Context

Mega-city Regions Mega-city regions (MCRs) are integrated sets of cities and their surrounding sub-urban hinterlands across which labour and capital can be reallocated at very low cost [Florida et al., 2008]. The expression “Mega-city Region” was first coined by Gottmann [Gottman, 1961] using the term megalopolis, that he defines as “urban area of several tens of millions of people, including several cities and major urban centers, and extending continuously over several 100 km”. It represents the new economic unit that emerges as metropolitan regions not only grow upward and become denser but also grow outward and into one another. These spaces result of the networking of a group of metropolitan areas deployed around very large cities and it is characterized by the “symbiosis between urban and rural areas”. Other characteristic are migration flows, density of connections, and regional migration patterns.

Pearl River Delta Since the gradual decentralization of the state power which occurred in the beginning of 1990 Mega-city regions have become a new scale of Chinese State regulation, and in particular the pearl River Delta, represent the most prosperous and dynamic one in term of migration waves. That's why we choose PRD as our unit of study. Since the Open Door Policy was implemented in 1978 the PRD launched a process of rapid economic and social transformation, becoming a global manufacturing region attracting an increasing number of migrant-workers from all over China. In fact, PRD was designed as an open economic zone in 1984, and was granted many 'one step ahead' policies to attract foreign capital, becoming the most important exporter center since economic reform. That results in an astonishing rise of population in the area. That today count more than 50 million people. If during the first year of the opening-up reforms the barycenter of the region was Guangzhou - the provincial capital, over the last decade PRD has become increasing polycentric. Taking PRD as the unit of our model we try to reproduce migrants residential patterns taking into account the full range of migrants' socio-economical status and their evolution.

Migrant Workers in PRD Even though migrants workers are generally considered and treated as a uniform category, previous research [Losavio, 2016], based on qualitative interviews and sociological observation, showed how considering their economical, cultural and human capital migrants workers are a very diversified social category. Especially 3 dimension are fundamental to understand migrant workers:

- Professional Dimension : industry, construction, private sector, services (will change not only their economical situation but which influence migrants' trajectory and the staying duration in the city and their residential choice.
- Residential Dimension : We could recognize two subcategories, with different socio-economic situations: (i) Temporary - migrants working in the construction sector, who are really mobile, changing city or province according to the labour demand; seasonal migrants workers, most of the time are peasants working as farmer and cultivating their land who come to the urban area to sell their product at the city price but without paying the running costs temporary migrants' trajectory are more difficult to detect since they move in a discretionary way ; (ii) Permanent : Who are living in the city for more than 6 months and are renting a room or apartment depending on their: Economic situation (revenues), Social network, Proximity to the work place
- Generational Dimension. *this dimension is not taken into account in the model for now, as simulated dynamics correspond to rather short time scales, between 10 and 20 years.*

All these sub-categories have different mobility patterns, that we try to simulate in our model.

Considering this diversity and translating it in qualitative stylized facts that correspond to precise patterns of synthetic data, this model aims at establishing a new perspective for understanding China's urban and regional mobility employing a more qualitative approach, specifying the mechanisms through which Party-State shape the parameters of migrants' choices.

1.2 Modeling Migrations

Modeling Rural-urban migrations [Todaro, 1969] classical equilibrium model

Modeling Rural-urban migrations in China Existing works in rural-urban migration modeling in China are mainly econometric studies, relying on census or on survey data. [Zhang and Zhao, 2013] estimate discrete choice models to study the trade-off between migration distance and earning difference. [Fan, 2005] shows that gravity-based models can explain well inter-provincial migratory patterns, implying an underlying strong dominant aggregation processes. The positive association between wage gap and migration rates was obtained from time-series analysis in [Zhang and Shunfeng, 2003]. An empirical study of intra-urban migrants residential dynamics is done by [Wu, 2006].

Towards an agent-based modeling approach To the best of our knowledge, there was no attempt in the literature before to focus on China’s migration issues from an agent-based perspective. The case of Mexico was tackled by [De Leon et al., 2007], but in the particular case of a border-town, and underlying processes are furthermore fundamentally different.

[Xie et al., 2007] : agent-based model to simulate the emergence of Urban Villages. [Silveira et al., 2006] : Ising model of rural-urban migration. [Fernandez et al., 2005] : study of population characteristics to establish the relevance of a future ABM.

The idea of applying complexity paradigms to rural-urban migration is far from new, as [Mabogunje, 1970] already theorized it in the frame of General System Theory, that for some is viewed as a precursor of complexity theories.

Following a logic of *Pattern-oriented modeling* [Grimm et al., 2005], combined with recent advances in multi-modeling [Cottineau et al., 2016], one can use agent-based models as powerful tools to test qualitative hypothesis, with a reasonable need for empirical data through toy-models or hybrid models.

2 Model

2.1 Rationale

The model is designed to include specific stylized facts and experiments. More precisely, a recent shift in socio-economic structure of migrating population was observed, including a rise of middle-income migrants and a relativisation of the role of *Hukou* in migration dynamics. The core of the model is thus centered on the exploration of the impact of a varying population economic structure for migrants on system dynamics, and the influence of government migration policies.

Ontology At a mesoscopic scale, i.e. for cities within the MCR, growth can be reasonably assumed as independent from migrants movements : they follow larger economic urban processes. Conditionally to such population and economic context, migration dynamics occur at the microscopic scale. We postulate a simple utility-maximization process.

2.2 Model Description

Setup The region is represented in the model by N patches, characterized by their population $P_i(t)$ and an economic structure $E_i^{(c)}(t)$ which consists in abstract variables representing potential number of jobs stratified by socio-economic classes c . The associated effective number of workers is denoted by $W_i^{(c)}(t)$. For the sake of simplicity, we assume a discrete number of classes. At initial time, the variables are initialized either following a synthetic data generation process (see below), or from real geographical data (abstracted and simplified to fit our context). Cells are grouped into $1 \leq k \leq K$ administrative cities that corresponding to the various centers of the MCR, on which aggregated population $\tilde{P}_k(t)$ and corresponding economic variables ($\tilde{E}_k^c(t)$) can be computed.

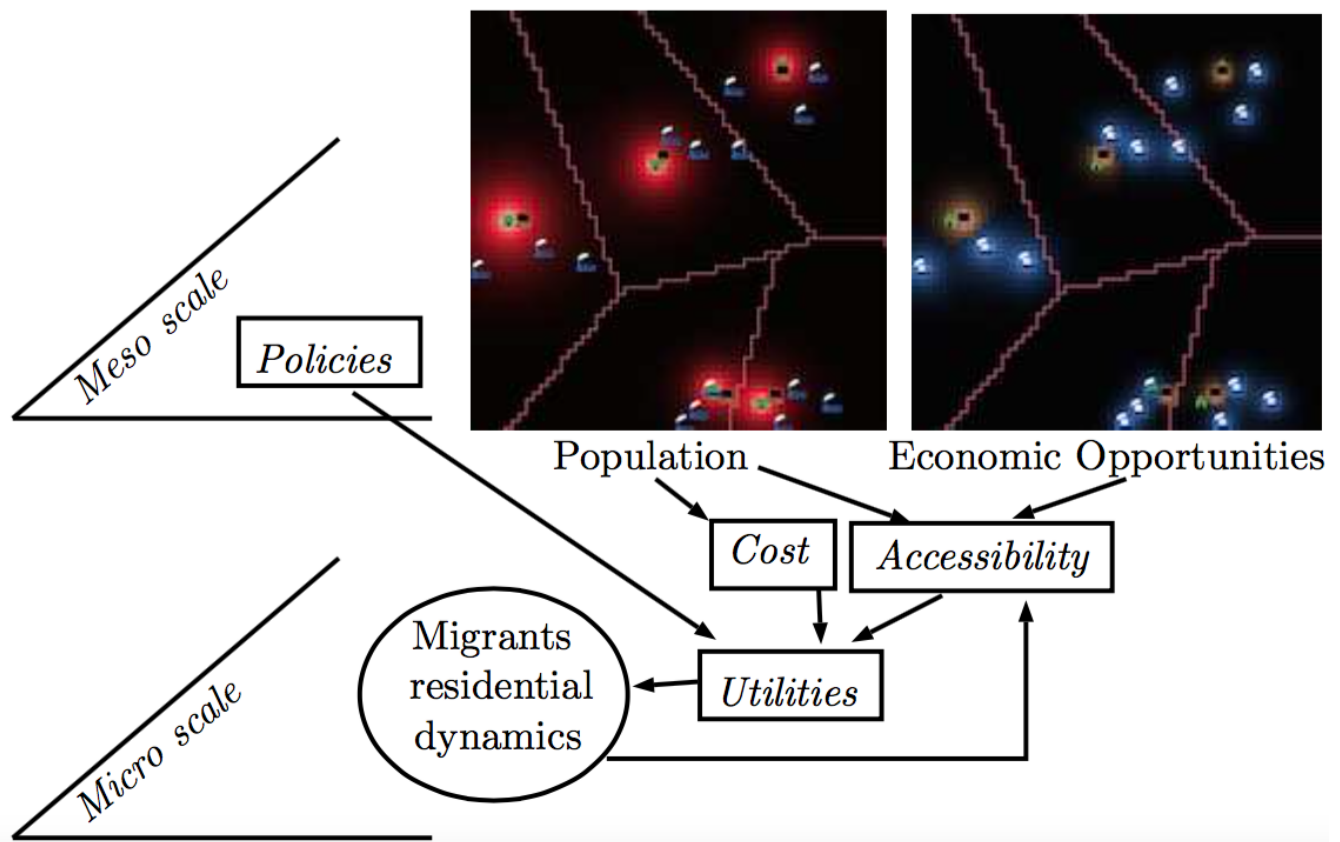


Figure 1: Multi-scale schema of processes included in the model

Agents An agent is a household of migrants, whose residence and job are located in cells (that can be different). Socio-economic structure of the population is captured by the distribution of wealth $g(w)$, which are then stratified into categories. At a given time, the utility difference between not moving and moving to cell j from cell i , for a category c is given by

$$\Delta U_{i,j}^{(c)}(t) = \frac{Z_j^{(c)} - Z_i^{(c)}}{Z_0} + \gamma \cdot \frac{C_i^{(c)} - C_j^{(c)}}{C_0} - u_i^{(c)} - h_j^{(c)}$$

where $Z_i^{(c)}$ is generalized accessibility given by $Z_i^{(c)} = P_i \cdot \sum_k \left[E_k^{(c)} - W_k^{(c)} \right] \cdot \exp\left(\frac{-d_{ij}}{d_0}\right)$, with d_{ij} effective travel distance¹ and d_0 commuting characteristic distance ; the parameter γ is the ratio giving the relative importance of life cost compared to accessibility in the migration decisions ; $C_i^{(c)}$ is the cost of life which is a function of cell and city variables, that will be taken as $C_i^{(c)} \propto P_i^{\alpha_0} \cdot \tilde{P}_i^{\alpha_1}$; $u_i^{(c)}$ a baseline aversion to move and $h_j^{(c)}$ an exogenous variable corresponding to regulation policies ; Z_0 and C_0 dimensioning parameters.

Temporal Evolution At each time step, the system evolves sequentially according to the following rules :

1. Cities mesoscopic variables $\tilde{P}_k(t)$ and $\tilde{E}_k^c(t)$ are deterministically updated. Population follows the very simple assumption of the expectancy of a Gibrat's law, i.e. $\tilde{P}_k(t) = g_k \cdot \tilde{P}_k(t-1)$. Economy follows a scaling law of population: $\tilde{E}_k^{(c)}(t) = E_k^{(c,0)} \cdot \left(\frac{\tilde{P}_k(t)}{\tilde{P}_0}\right)^{\gamma_k^{(c)}}$.
2. Patches variables are updated conditionally to the new aggregated values. For population, we adapt the aggregation-diffusion process detailed in [Raimbault, 2016]: a density ceil is introduced and diffusion is replaced by spatial noise. More precisely, let $\Delta \tilde{P}_k(t) = \tilde{P}_k(t) - \tilde{P}_k(t-1)$.
3. A number M of new migrants, enters the region. They lean on social network to first settle in the city and agglomerate following a preferential attachment by place of origin.
4. Migration occur following a discrete choice dynamics : the probability to move to cell j is given by

$$\mathbb{P}[i \rightarrow j|c] = \frac{\exp\left(\beta \cdot U_j^{(c)}\right)}{\sum_k \exp\left(\beta \cdot U_k^{(c)}\right) + \exp\left(U_{stay,i}^{(c)}\right)}$$

what simplifies into a reduced form, with $\beta' = \frac{\beta}{Z_0}$, $\gamma' = \frac{\gamma}{Z_0 C_0}$ and \tilde{u}, \tilde{h} accordingly rescaled variables, using the above utility expression :

$$\mathbb{P}[i \rightarrow j|c] = \frac{\exp\left(\beta' \cdot \left[\Delta Z_{i,j}^{(c)} - \gamma' \cdot \Delta C_{i,j}^{(c)} - \tilde{u}_i^{(c)} - \tilde{h}_j^{(c)}\right]\right)}{1 + \sum_k \exp\left(\beta' \cdot \left[\Delta Z_{i,k}^{(c)} - \gamma' \cdot \Delta C_{i,k}^{(c)} - \tilde{u}_i^{(c)} - \tilde{h}_k^{(c)}\right]\right)}$$

Residential movement is drawn randomly according to these probabilities, and jobs are chosen around new residence following an exponentially decreasing probability.

¹as the model does not focus on the role of transportation, we take euclidian distance, and d_0 captures typical commuting distance in both public transportation or car. A more complex model could include an explicit transportation network and modal choice depending on socio-economic category.

Table 1: Summary of main parameters

Parameter	Name	Values	Process
γ	Cost/Accessibility ratio	$\log \gamma \in [5; 8]$	Mobility
u_0	Move aversion	$u_0 \in [0; 5]$	Risk aversion
β	Discrete Choices	$\beta \in]0; +\infty[$	Determinism
g_w	Income Growth	$g_w \in [0; 1]$	Wealth Increase
d_0	Accessibility Decay	$d_0 \in]0; +\infty[$	Accessibility
σ	Wealth dispersion	$\sigma \in [0.1; 1.0]$	Economic Inequalities

5. Migrants update their wealth and eventually economic category, according to an abstract “quality of place” that we associate to per-capita GDP which follows a scaling law of population

$$w = w \cdot \left(1 + g_w \frac{P_i}{\max P_j} \cdot \frac{\tilde{P}_k}{\max \tilde{P}_l} \right)$$

Indicators The outcome of the model has many dimensions, and is measured through the following indicators :

- Total migrants wealth gain
- Total migrants social mobility
- Cumulated utility difference in migrations: when someone migrates, it corresponds to a given utility gain ΔU : this indicator is $\sum \Delta U^{(c)}$ on all realized migrations for each category.
- Inequalities are captured by the final ratio between socio-economic categories

Synthetic Data Generation A synthetic mega-city region is generated following simple stylized facts for systems of cities. A polycentric structure, each center having an exponential density decay, is created with random coordinates for centers. Population of centers follow a rank-size law of slope 1.2.

Parameters We summarise main parameters in table 1.

3 Results

3.1 Model Validation and Baseline Exploration

Internal validation statistical consistence : cf Fig. 2 ; system trajectories ; path-dependency.

External validation : baseline behavior We first establish the behavior of the model on the simplest baseline situation, to which will be compared more complex parameter configurations. The baseline assumes no city growth ($g_k = 1$), no external migration ($M = 0$), no income growth (g_w), u_i is fixed at a constant value U_0 and we have only one economic category. Fig. 3 show indicator values across the parameter space.

We extract the following stylized facts from model behavior :

- As expected, we observe a phase transition in number of migrations for increasing β ; the transition points increases when u_0 decreases, until the model follows a change in qualitative behavior (bifurcation ?) : increasing β first produces a peak in migrations before they fall down again. The

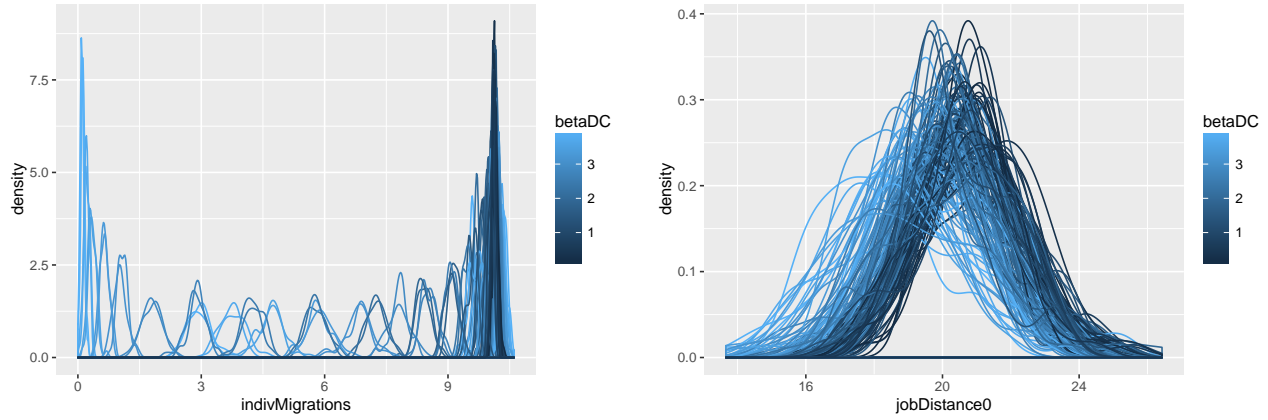


Figure 2: Statistical distributions of number of individual migrations and wealth gain over 50 runs, for some example parameter points

increase above 1 migration per migrant and per time step is possible by the disappearance of some of them due to a saturation of jobs. *When migrants have a high propensity to move, the spatial repartition of jobs becomes suboptimal in intermediate regimes of stochasticity, corresponding to a regime where congestion dominates.*

- The previous observation is consistent with the behavior of the average distance to job (this plot is quite noisy as it depends on spatial configuration, stabilizing it would thus need supplementary repetitions, as the exploration for sensitivity of meta-parameters) : it always decreases with β , but individual move repulsion changes the shape. Low values give a linear decrease of job distance with β , whereas high values produce a saturating curve that becomes constant. *The congestion regime corresponds to a linear decrease of job distance with randomness, meaning that social determinism creates spatial inequalities.*
- Not sensitive to γ : the sensitivity to ratio cost/accessibility is very low , what may be due to the dependence of dynamics to outliers of utility distributions : its central shape changes a lot with it but not extreme values. *Changing the relative importance of accessibility does not affect much the aggregated dynamics : an increased gain in mobility produced by policies for examples will have no effect on migrations patterns.*
- Temporal Dynamics : Fig. 4. The dependence of dynamics to parameter is much richer ; we observe a transition from a maximum, to a min, and then a max again. In configurations averse to move, time has a positive feedback on migrations ; intermediate situations show a negative feedback (progressive saturation in time) ; to switch again to a positive feedback in most mobile situations. *Configurations with an intermediate value of move aversion (in which real configurations fall) yield a negative feedback effect to time, witnessing a progressive saturation. In a “U-shape” manner, very mobile or very fixed configurations yield positive feedback of time.*

3.2 Stylized Facts

Role of economic categories

- *Adding categorisation does not change the qualitative behavior of the model. The lower category seems more vulnerable to spatial inequalities created by social determinism.*

Role of income growth

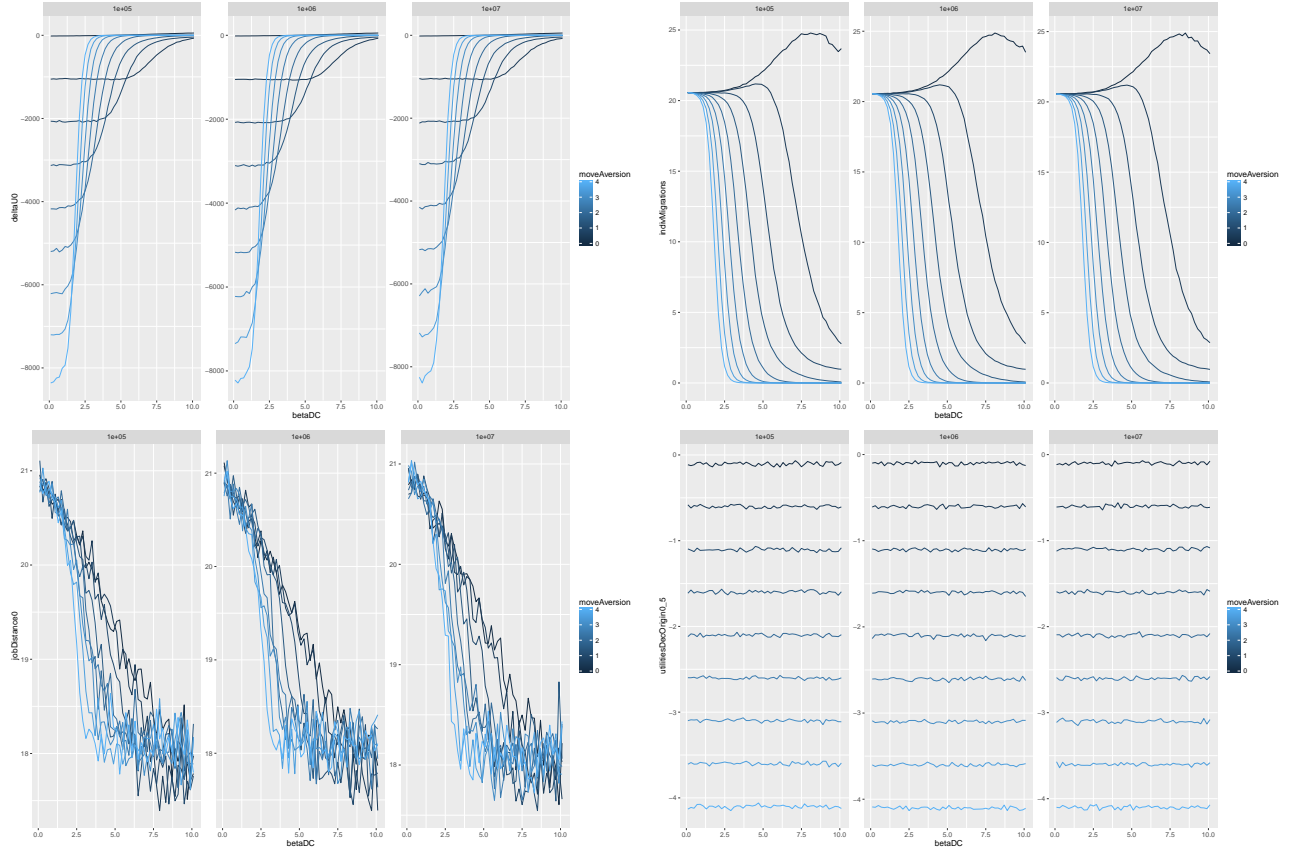


Figure 3: Model behavior, averaged over 50 runs, for a baseline configuration.

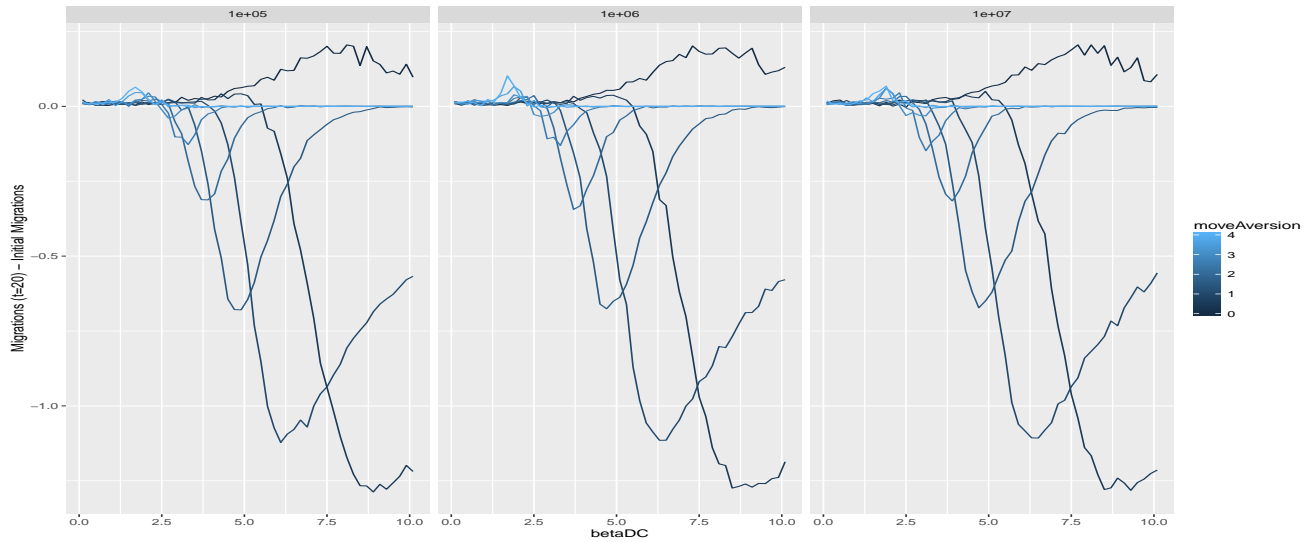


Figure 4: Difference in number of migrations between initial and final state of the model (relatively robust to final time).

Role of accessibility decay

Sensitivity to Urban Form

3.3 Application : Real Configuration

Stylized Pearl River Delta configuration The model is parametrized on a simplified configuration, with the sufficient information at the relevant scale: even if population data for example can be used with a very high resolution, it makes no sense regarding model ontology to include too precise GIS data as the city growth mechanisms are approximated processes. An correspondence between population and employment data is furthermore desirable, the former being available only at the scale of cities.

Comparing regulation policies Governments at different levels (national directives, regional policies, and municipal implementation) try to regulate migration patterns. A point-based mechanism, a sort of “social score”, has been tested for regulating urban Hukou. We stylize such policies by emphasizing their spatial aspect : across municipalities, there exist different incentives or dissuasions to migrate for some categories of migrants.

Experience Plan Concrete qualitative questions that can be asked to the model are for example:

- what is the impact of varying wealth distribution shape and width on system dynamics/migrant trajectories ?
- what is the impact of various spatial distribution of $h_j^{(c)}$, i.e. different government policies ? We stylize existing policies that qualitative studies have unveiled.
- Comparison with and without Hong-Kong and Macao

3.4 Summary of Stylized facts

From the baseline behavior we learn that :

- When migrants have a high propensity to move, the spatial repartition of jobs becomes suboptimal in intermediate regimes of stochasticity, corresponding to a regime where congestion dominates. More precisely, we use a “Random Utility” specification for migrants’ decisions, that includes a parameter β that represents a “level of randomness” (or stochasticity) : when it goes to 0, migrations are totally random and without any sense, when it goes to infinity, migrants follow deterministically the best choice between staying and going to the best migration patch (depending on the sign of ΔU). For intermediate values of β , we observe this congestion phenomenon.
- The congestion regime corresponds to a linear decrease of job distance with randomness, meaning that social determinism creates spatial inequalities. It means that proximity to job is correlated with more congestion, but there is no direct causality between the two, both are outcome of model processes.
- Changing the relative importance of accessibility does not affect much the aggregated dynamics: an increased gain in mobility produced by policies such as individual transportation subsidies will have no effect on migrations patterns. Accessibility is a generic notion to capture patterns of accessibility to urban functions and amenities ; in our case we use generalized Hansen accessibility, which can be understood in a simplified way number of jobs available to the population within a spatial range.

- Configurations with an intermediate value of move aversion (in which real configurations fall) yield a negative feedback effect to time, witnessing a progressive saturation. In a “U-shape” manner, very mobile or very fixed configurations yield positive feedback of time.

Comparing baseline configurations with one and two economic categories :

- Adding categorisation does not change the qualitative behavior of the model. It does not mean that categorization (based on professional qualifications, education, and economic situation) does not influence the migration behavior. In practice for example the points-based hukou system favoring “talents” – people with more qualifications, with stable job and economic situation etc. – can have a significant influence on migration behavior. It changes the migration behavior for individuals, but not the global qualitative picture of the previous stylized facts. The lower category seems more vulnerable to spatial inequalities created by social determinism.

Concerning the influence of economic parameters, namely income inequality and income growth, we find that :

-

The introduction of real world data slightly changes conclusions :

-

Finally, we compare stylized policies :

-

References

- [Cottineau et al., 2016] Cottineau, C., Rey-Coyrehourcq, S., and Reuillon, R. (2016). Back to the future... of multimodelling. In *RGS 2016, Session Geocomputation : the next 20 years*.
- [De Leon et al., 2007] De Leon, F., Felsen, M., and Wilensky, U. (2007). Netlogo urban suite-tijuana bordertowns model. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL*.
- [Fan, 2005] Fan, C. C. (2005). Modeling interprovincial migration in china, 1985-2000. *Eurasian Geography and Economics*, 46(3):165–184.
- [Fernandez et al., 2005] Fernandez, L. E., Brown, D. G., Marans, R. W., and Nassauer, J. I. (2005). Characterizing location preferences in an exurban population: implications for agent-based modeling. *Environment and Planning B: Planning and Design*, 32(6):799–820.
- [Florida et al., 2008] Florida, R., Gulden, T., and Mellander, C. (2008). The rise of the mega-region. *Cambridge Journal of Regions, Economy and Society*, 1(3):459–476.
- [Gottman, 1961] Gottman, J. (1961). Megalopolis. *Twentieth Century Fund*.
- [Grimm et al., 2005] Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., and DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *science*, 310(5750):987–991.
- [Losavio, 2016] Losavio, C. (2016). Comment analyser l’intégration des ouvriers-migrants dans la chine d’aujourd’hui. le cas de wuhan. *Institut d’études Politiques de Paris. Ecole Doctorale de Science Po*.

- [Mabogunje, 1970] Mabogunje, A. L. (1970). Systems approach to a theory of rural-urban migration. *Geographical analysis*, 2(1):1–18.
- [Raimbault, 2016] Raimbault, J. (2016). Calibration of a spatialized urban growth model. *Working Paper, draft at <https://github.com/JusteRaimbault/CityNetwork/tree/master/Docs/Papers/Density>*.
- [Silveira et al., 2006] Silveira, J. J., Espíndola, A. L., and Penna, T. (2006). Agent-based model to rural–urban migration analysis. *Physica A: Statistical Mechanics and its Applications*, 364:445–456.
- [Todaro, 1969] Todaro, M. P. (1969). A model of labor migration and urban unemployment in less developed countries. *The American economic review*, 59(1):138–148.
- [Wu, 2006] Wu, W. (2006). Migrant intra-urban residential mobility in urban china. *Housing Studies*, 21(5):745–765.
- [Xie et al., 2007] Xie, Y., Batty, M., and Zhao, K. (2007). Simulating emergent urban form using agent-based modeling: Desakota in the suzhou-wuxian region in china. *Annals of the Association of American Geographers*, 97(3):477–495.
- [Zhang and Zhao, 2013] Zhang, J. and Zhao, Z. (2013). Measuring the income-distance tradeoff for rural–urban migrants in china.
- [Zhang and Shunfeng, 2003] Zhang, K. H. and Shunfeng, S. (2003). Rural–urban migration and urbanization in china: Evidence from time-series and cross-section analyses. *China Economic Review*, 14(4):386–400.