Ikumo Isono ed., *The Need to Upgrade the Geographical Simulation Model (IDE-GSM)*,
ERIA Collaborative/Support Research Report, IDE-JETRO, 2021

Chapter 2
Possible directions for the development of the IDE-GSM and difficulties expected

Satoru Kumagai*

Abstract

This chapter reviews the history of expanding the Geographical Simulation Model (IDE-GSM), a computational general equilibrium model based on spatial economics. We show the possible direction of future development of the model and some expected difficulties in its future development. We then discuss the fundamental trade-off in the simulation model: the balance between the realism of the simulation and the increased amount of data and parameters required. To remedy this trade-off, we need to develop tools to compile the necessary dataset and calibrate the parameters endogenously and automatically.

Keywords: Spatial Economics, CGE model, simulation analysis.

Introduction

The IDE-GSM (Geographical Simulation Model), which has been developed by the Institute of Developing Economies, JETRO (IDE-JETRO) since 2007, is a computational

* Director, Economic Geography Studies Group, Development Studies Center, Institute of Developing Economies, JETRO (IDE-JETRO).

general equilibrium (CGE) model that predicts global industrial and population agglomerations at a sub-national (state or provincial) level (Kumagai et al. 2013). The IDE-GSM has been utilized by various international organizations, such as the Economic Research Institute for ASEAN and East Asia (ERIA), the World Bank, and the Asian Development Bank (ADB), to plan and analyze economic corridors, transportation infrastructure development plans, and the effects of free trade agreements. In general, analyzing the economic effects of transportation infrastructure requires a great deal of time and money, as data must be prepared on population and gross domestic product (GDP) by industry for each region and a wide range of information must be collected, such as transport volume by route. The IDE-GSM is suitable for the economic impact analysis of international transportation infrastructure development because it can perform simulation analysis with a relatively small amount of data.

In this chapter, we discuss possible directions for the expansion of the IDE-GSM and expected difficulties. While there are some possible directions of expansion for the model, each direction of expansion comes with its own set of difficulties. These difficulties are unavoidable because every simulation model has a fundamental trade-off: the simulation's reality versus the required data/parameters. The IDE-GSM, which can be simulated using a relatively small dataset, is no exception. The remainder of this chapter is organized as follows. Section 2 briefly describes the structure of the IDE-GSM. Section 3 discusses possible directions for expansion of the model and anticipated difficulties. Section 4 describes the fundamental trade-offs in the simulation model and how they can be resolved. Finally, Section 5 concludes the chapter.

2. IDE-GSM

2.1. Basic structure of the IDE-GSM

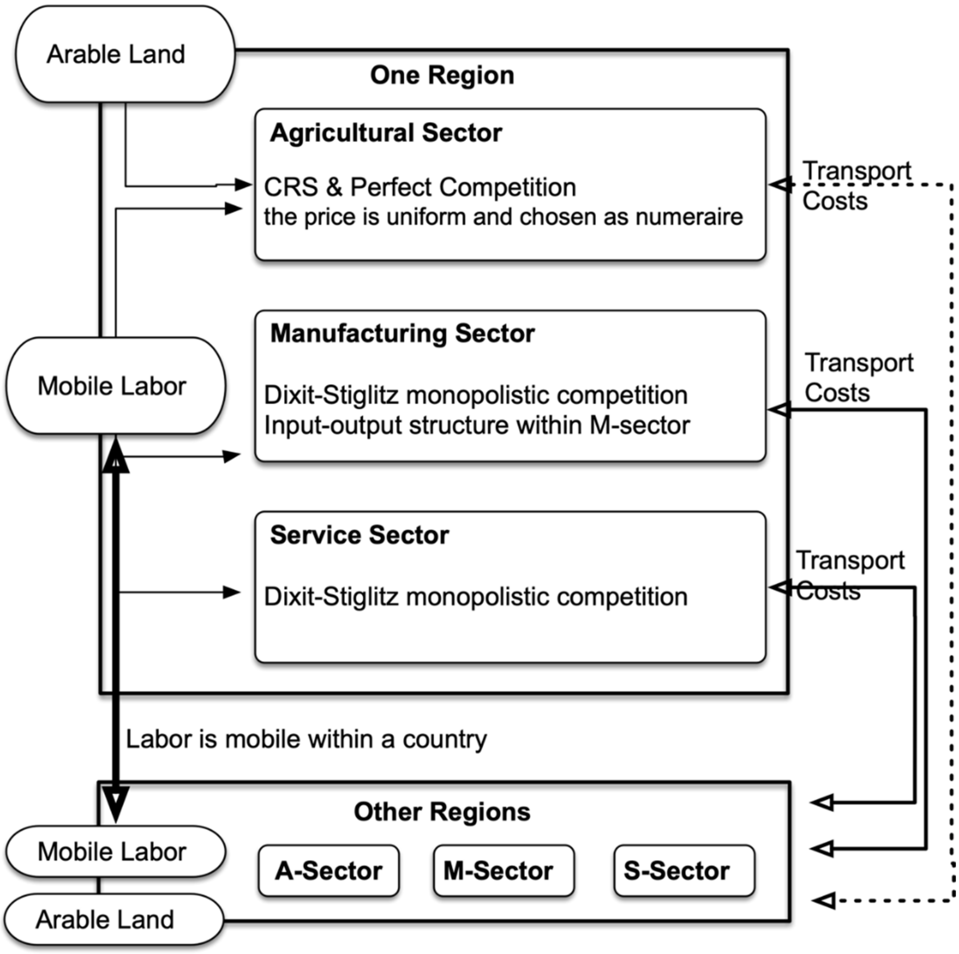
The IDE-GSM is based on spatial economics or new economic geography (NEG). Theoretically or empirically, these simulation models tend to be very complicated and difficult to solve mathematically; thus, NEG studies frequently use numerical simulations. The basic model, the Core-Periphery (CP) model by Krugman (1991), also uses numerical solutions to show a set of spatial economic models' fundamental characteristics. The basic CP model is a two-location/two-goods model, setting one good (typically assumed as agricultural goods) as numeraire, which is produced by a constant return to scale technology and incurring zero transport costs, whereas the other good is produced by an increasing return to scale technology (typically assumed as manufactured goods) and incurring non-zero and positive transport costs.

The beauty of the CP model with many locations is its simplicity. Despite its simplicity, we can derive rich implications for the real world. Indeed, the IDE-GSM started as a branch of the CP model in many locations, except that the geography is not the "racetrack" (Krugman 1993), but a realistic network of cities that replicates the real world around the Mekong region.

The IDE-GSM was developed based on this CP model, with two main objectives: (1) to simulate the dynamics of the locations of populations and industries in East Asia over the long term and (2) to analyze the impact of specific transport and trade facilitation measures (TTFMs) on regional economies at sub-national levels. There are more than 3,000 regions in our simulation model, along with two endowments: labor and land. Labor is mobile within a country, but it is prohibited from migrating to other countries. The land is unequally spread in all regions and is jointly owned by all the laborers of the region.

Figure 2-1 shows the structure of the IDE-GSM model. All products in the three sectors are tradable. Transport costs are supposed to be of the iceberg type; if one unit of product is sent from one region to another, the unit with less than one portion arrives. The supplier sets a higher price depending on the lost portion. The increase in price compared with the producer's price is regarded as the transport cost. Transport costs within the same region are considered negligible.

Figure 2-1: Structure of the model

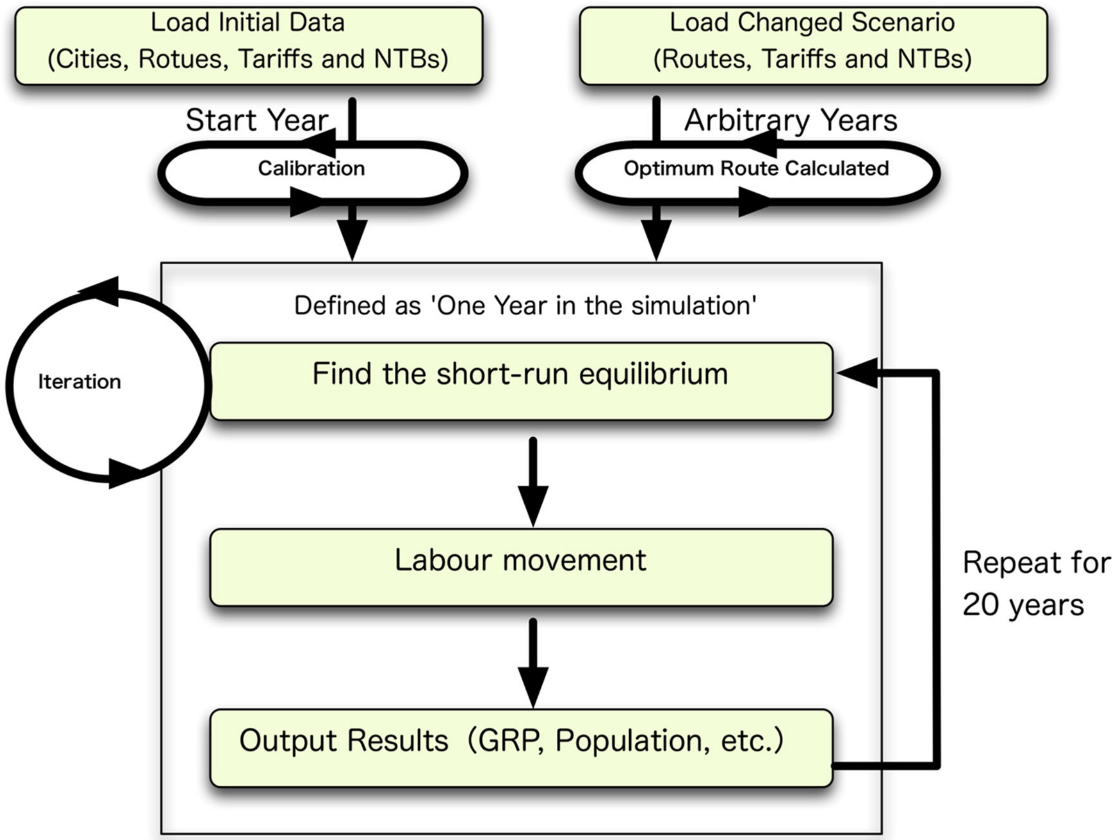


Source: Authors.

Note: CRS: constant returns to scale.

The simulation procedures are shown in Figure 2-2. First, with distributions of employment and regional GDP by sector and region according to the actual data, a short-run equilibrium is obtained. Observing the achieved equilibrium, workers migrate among regions and industries according to differences in the real wages; workers move to sectors that offer higher real wage rates in the same region and move to regions that offer higher real wages within the same country. Thus, we obtain a new distribution of workers and economic activities. With this new distribution and predicted population growth, the next short-run equilibrium is obtained for the following year, and we observe migration again. These computations are typically repeated for 20 years, for example, from 2010 to 2030.

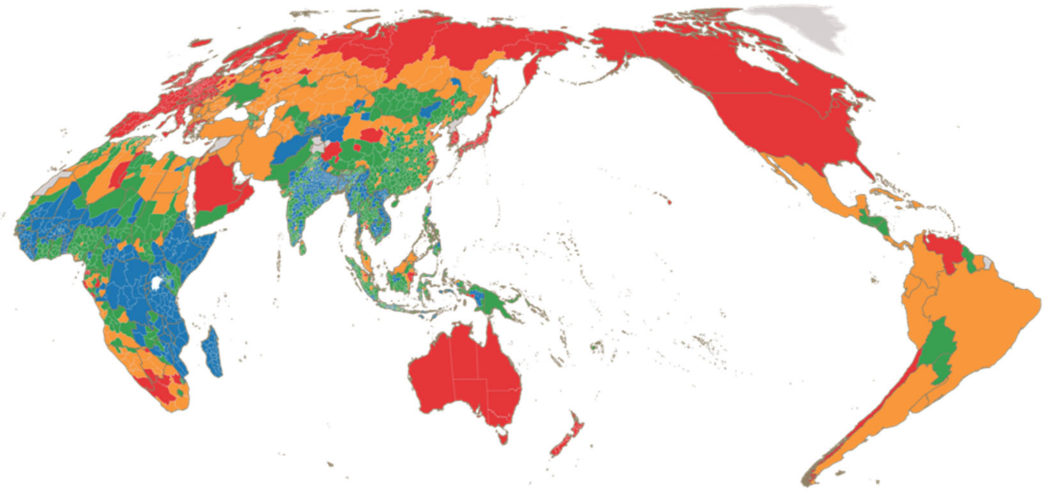
Figure 2-2: Procedures of the simulation



Source: Authors.

Primarily based on official statistics, we derive the gross regional product (GRP) in 2010 for the agriculture, mining, services, and five manufacturing sectors. The five manufacturing sectors are food processing, garments and textiles, electronics and electric appliances (E&E), automotive, and other manufacturing. Population and area of arable land for each region are compiled from official statistical sources. Figure 2-3 shows the GRP per capita for each region in 2010.

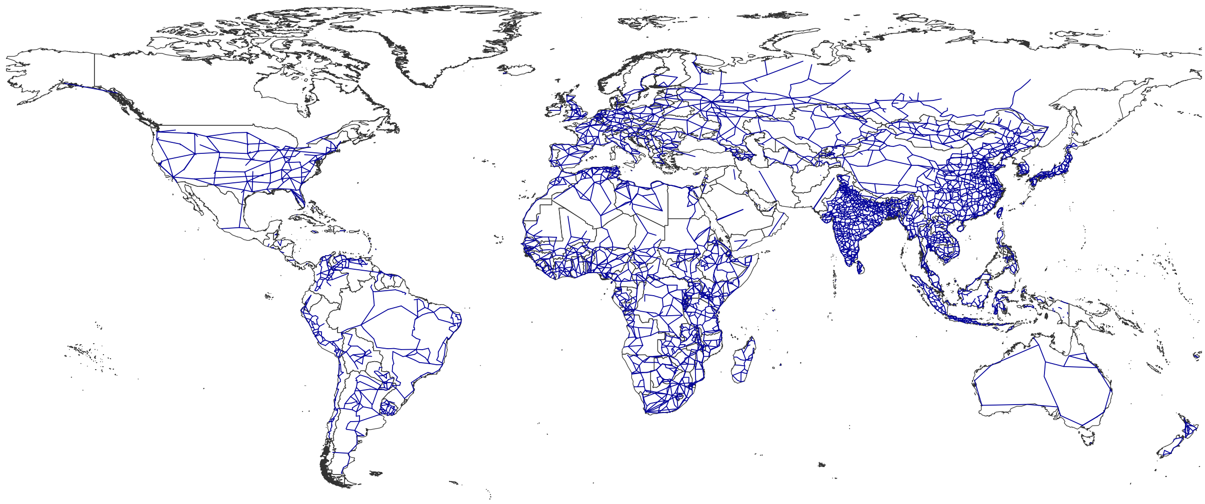
Figure 2-3: GRP per capita in East Asia, 2010



Source: Authors.

More than 17,000 routes are included in the simulation (land: 11,500; sea: 1,300; air: 2,500; and railway: 1,650). The route data comprise start city, end city, the distance between the cities, the vehicle's speed running on the route, etc. The land routes between cities are based mainly on the "Asian Highway" database of the United Nations Economic and Social Commission for Asia and the Pacific. The actual road distances between cities are used: if the road distances are not available, then the distances between cities in a straight line are employed. Figure 2-4 shows the land route networks incorporated in the IDE-GSM. The data on air and sea routes are compiled from Nihon Kaiun Shukaijo (1983) and the dataset assembled by the Logistics Institute-Asia Pacific team, and 1,300 sea routes and 2,500 air routes are selectively included in the model. The railway data are adopted from various sources such as maps and the official websites of railway companies.

Figure 2-4: Land route network data in the IDE-GSM



Source: Authors.

2.2. History of expansion of the IDE-GSM

Version 1.0

We launched the development of the IDE-GSM in 2007. At the end of FY2007, the first-generation GSM was completed. The IDE-GSM 1.0 initially covered eight countries in continental Southeast Asia (Singapore, Peninsular Malaysia, Thailand, Cambodia, Laos, Myanmar, and Vietnam) and was then extended to Yunnan Province in China, Guangdong and Guangxi provinces of China, Hong Kong and Macao, Bangladesh, and nine states in India (Arunachal Pradesh, Assam, Manipur, Mizoram, Nagaland, Sikkim, Tripura, and West Bengal). The economic sector is divided into agriculture, manufacturing, and services. The mode of transport considered was truck only. Intermediate inputs are not incorporated in this model, and the model is a simple extension of Krugman's core-periphery model.

Version 2.0

Version 2.0 of the IDE-GSM was developed in FY2008 and incorporates intermediate inputs. The economic model is very similar to the model in Chapter 14 of Fujita, Krugman, and Venables (1999). The manufacturing sector is divided into five sectors (automotive, E&E, textiles and garments, food processing, and other manufacturing), so the number of economic sectors is now seven. Dividing the manufacturing sector into five enables various analyses that were not possible in the first-generation model. For instance, we can now predict more precisely which industry is located where. We also estimate the impact of a specific infrastructure project, such as the East-West Economic Corridor, on a specific industry in a specific location, such as the automotive industry in Bangkok.

On the other hand, we need to recompile the dataset for seven economic sectors and specify various industry-specific parameters, such as transport costs, substitution elasticity, labor inputs, and consumption shares.

Version 3.0

Version 3.0 of the IDE-GSM was developed in FY2009. GSM version 3.0 adopts a modal mix that minimizes the total transport costs, considering monetary and time costs. To incorporate realistic modal choices, sea and air routes are added to the route data. The calculation of transport costs was significantly refined. The earlier version of the model treats transport costs merely as an increase in time and distance. In Version 3.0, the transport costs are calculated as a percentage of the value of a good per 20-ft. container, considering both the time and monetary costs. The parameter of the elasticity of substation by industry is estimated using the International Input-Output Table for Asia. Formally, these parameters were adopted from various studies, typically Hummels (1999). Geographical coverage has also been significantly extended. Indonesia, the Philippines, Sabah, Sarawak in Malaysia, Brunei, and nine states in India (Andhra Pradesh, Bihar, Chhattisgarh, Jharkhand, Karnataka, Kerala, Maharashtra, Orissa, and Tamil Nadu) were added to the dataset. To address the problems associated with the inclusion of some extremely high-income regions with natural resources (Bontang, Indonesia, typically), the mining sector is added as an alternative to the agriculture sector. The mining sector is treated the same as the agricultural sector, which uses land as an input; all the population shares the land's rent within a region. The labor (land) share of the income is set lower (higher) than that of the agricultural sector.

Version 4.0

Version 4.0 of the IDE-GSM was developed in FY2010. The geographical coverage was again extended significantly. All states in India and all provinces in China are included in the dataset. Another significant expansion of the model is the inclusion of policy and cultural barriers (PCBs) (Kumagai and Isono 2011). Currently, PCB costs are imposed in transactions between regions in different countries. The introduction of PCBs significantly reduced the economic impact of TTFMs compared to the earlier version of GSM. In addition to these extensions, another mode of transport, railway, is incorporated. The data for railways are not fully included, but we can now simulate the economic impact of the Singapore-Kunming Rail Link (SKRL) development. The rest of the world (ROW) is also included. ROW is now represented by three regions: Japan, the EU, and the US.

Version 5.0

Version 5.0 of the IDE-GSM was developed in FY2011, with the geographical coverage further extended. Japan, a part of the ROW in GSM 4.0, is divided into 47 prefectures and included in the simulation. Korea and Taiwan were also included. Sixteen countries are now covered in the simulation. Several refinements of the parameters are ongoing:

1) The immigration parameters, which are very important for determining the speed of adjustment of location choice but were temporarily set so far, are estimated using data.
2) Tariffs are carved out from other policy and cultural barriers.
3) More realistic traffic controls, such as "transit trade," one-way routes (critical to the realization of circular maritime transport routes), and "cargo-only" transport mode (not for passenger transport), can now be realized through simulations.

Version 5.1

Version 5.1 of the IDE-GSM was developed in FY2014. In Version 5.1, the calculation of rents for agricultural land was refined. In addition, the formula for imposing non-tariff barriers (NTBs) was improved more realistically.

Version 6.0

Version 6.0 of the IDE-GSM was developed in FY2015. High-speed rail (HSR) has been introduced as a new transportation mode. High-speed rail was set up as a passenger-only mode, which is faster, but simultaneously more costly, than the conventional railway. In Version 6.0, domestic NTBs have been set for transactions across administrative boundaries within a country for large countries such as India and China. In addition, the NTBs can be changed city by city, making it possible to simulate special economic zones (SEZs) for designated cities. Furthermore, an extension was made to simulate the Border Economic Zone (BEZ) assumed for Thailand.

Version 7.0

Version 7.0 of the IDE-GSM was developed in FY2016. In version 7.0, the mining sector has been separated from the agricultural sector. Previously, the mining sector was identical to the agricultural sector. Mining producing regions were treated as a special kind of agricultural sector by setting the labor input share lower. In addition, bilateral NTBs were set by calibration, giving the world trade matrix's real values. In the past, NTBs were unilaterally provided as exogenous variables for each country.

Version 7.1

Version 7.1 of the IDE-GSM was developed in FY2017. The geo-economic data for Europe and Africa were composed of more detailed administrative divisions and incorporated into the dataset for simulation. This allowed us to cover a large part of the world at a sub-national level, except for South America.

Version 7.2

Version 7.2 of the IDE-GSM was developed in FY2018. Version 7.2 migrated to OpenJDK 12 and NetBeans 11.0. Version 7.2, which is tuned to simulate a longer period, up to 2050, by gradually reducing the exogenous technological growth rate beyond 2023.

Version 8.0

Version 8.0 of the IDE-GSM was developed in FY2019. The most significant change in Version 8.0 is that the formula for the trade value has been changed to calculate the trade volume (value) much closer to reality. This makes it possible to calculate the changes in traffic volume for each route more accurately. Before this version, transport volume was treated as a reference value associated with economic effects. In terms of the program, the lack of memory was apparent as the amount of data increased; thus, the garbage collection process was revised and succeeded in coping with the lack of memory.

Version 9.0

Version 9.0 of the IDE-GSM was developed in FY2020. The Input-Output (IO) tables in the previous versions of the IDE-GSM included intermediate inputs only on the diagonal cell, and inputs from other industries were not assumed; Version 9.0 incorporated the IO tables for the manufacturing and service sectors that included intermediate inputs for each other. This improvement makes the simulation results more realistic. For instance, a reduction in transaction costs for the service sector affects the manufacturing sector.

Improvements were made to the program so that price and price index calculations for different industries can be performed in parallel without causing problems.

In Version 9.0, the route data for railways were duplicated, and passenger and cargo traffic were separated. This makes it possible to set different waiting times (frequency of operation) for passenger and freight trains.

3. Possible direction for development and difficulties expected

Future directions for extending the IDE-GSM include 1) a finer level of administrative division, 2) a finer industrial sector, 3) heterogeneity in firms/labor/products, and 4) congestion/capacity constraints in routes. Each of these directions has inherent difficulties.

3.1. Finer level of administrative divisions

Finer geographic units are desirable for more accurate simulations. However, adopting finer geographic units may lead to an explosive increase in necessary regional economic and route data. As a practical matter, there is a limit to the administrative division level for which each country provides geo-economic data. In many countries, geo-economic data are available for a one-level lower administrative division than the country-level, and economic data are rarely available for more detailed administrative divisions.

The more detailed the administrative divisions, the more detailed the route data that connect the administrative divisions. Route data are rarely provided officially, even more so than economic data. Therefore, it is necessary to create data from map information. The more detailed the administrative division, the more detailed the corresponding route data need to be, and the more time is needed to create it.

3.2. Finer industrial sector

For more accurate simulations, a more detailed industrial classification would be preferable. The IDE-GSM currently has four sectors: agriculture, mining, manufacturing, and services. The manufacturing sector is further divided into five sub-sectors: automobiles, electronics and electrical appliances, textiles and clothing, food processing, and other manufacturing.

The more detailed the industrial sectors are, the more difficult it becomes to obtain GDP data by industry, particularly at the sub-national level. Many countries provide data at the sub-national level with "manufacturing" as one sector, while the service sector is often subdivided into several sectors.

A more refined definition of industrial sectors also creates the need to successfully set appropriate parameters to reproduce each industry's nature in the simulation. For example, it is not difficult to differentiate the food industry's automobile industry using a few parameters. However, setting the automobile and motorcycle industries as separate industries requires more precisely estimated parameters.

3.3. Heterogeneity in firms/labor/products

More accurate simulations should introduce heterogeneity in factors such as firms, workers, and products. For example, there are firms with higher productivity than others, and they behave differently. For workers, unskilled and skilled labor have different wages and mobility.

The introduction of such heterogeneity may also increase the amount of data required, particularly if geo-economic data are required at the sub-national level. For each region, data on firms' and workers' quality are needed to assume the distribution of firm productivity and share of skilled/unskilled workers. In countries where sub-national-level data are generally unavailable, it is not easy to set such data to plausible values.

3.4. Congestion/capacity constraint in routes

To make the simulation more accurate concerning transportation, congestion and the upper limit of each route's capacity should be calculated endogenously in the simulation, and the route selection should be changed dynamically. For example, if traffic is concentrated on one route, congestion will occur there, and the average speed will decrease. If this congestion is considered, another route may become optimal. For railroads, airlines, and ports, it is more realistic to assume an upper limit on the number of containers transported in one year. If the volume of traffic in a given year exceeds the upper limit of capacity, it would be desirable to establish a system that allows excess logistics to flow to other routes.

However, to realize such a dynamic route selection based on the traffic volume, it is necessary to deal with the increase in computation volume. The computation to select the best route-mode combinations for each industry-origin-destination combination is so time-consuming that the graphics processing unit (GPU) is utilized to perform the computation. Suppose endogenous congestion and capacity constraints are considered; in this case, computational capacity is likely to become an issue, as multiple traffic calculations and recalculations of optimal routes will be required for every period.

3.5. An example of difficulties associated with expansion of the model

Here, we introduce an example of the difficulties associated with an expansion of the model, the inclusion of a realistic IO table in the model, performed in version 9.0 of the IDE-GSM. We compare the simulated economic impacts of the Southern Economic Corridor (SEC), spanning from Vung Tau, Vietnam to Dawei, Myanmar, through Phnom Penh, Cambodia and Bangkok, Thailand.

The inclusion of the complex IO structure among manufacturing sectors and services in version 9.0 of the IDE-GSM causes some changes in the simulation results. The IDE-GSM 8.0 also contains an IO structure, but it is limited within its sector; that is, the automotive sector uses the intermediate input only from the automotive sector.

Table 2-1 shows the differences in the simulated economic impacts from the SEC by versions 8.0 and 9.0. The economic impacts are compared for the year 2030, 10 years after the completion of the SEC. SEC's total economic impact is 782 million USD by version 8.0, while it is 2,498 million USD, 3.2 times larger by 9.0. By economic sector, the economic impact of version 9.0 is significantly larger in the automotive sector in Thailand, while the economic impact for the food processing industry in Myanmar becomes negative. These results come from a complicated interaction of comparative advantage changes among industrial sectors, taking into account the IO structure.

Table 2-1: Simulated economic impacts from the SEC by version 8.0 and 9.0 (2030, compared with baseline, million USD)

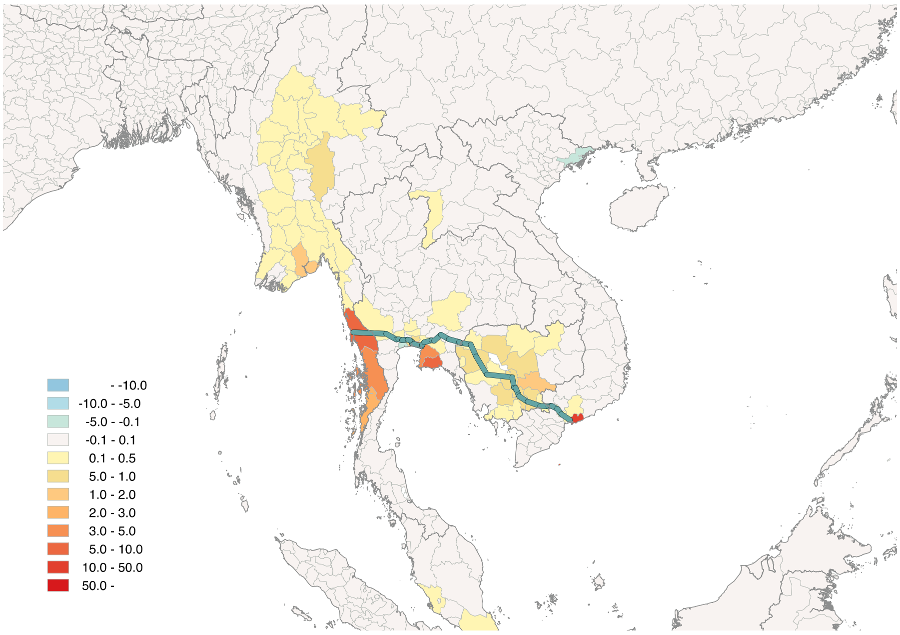
Version 8.0	Agriculture	Automotive	E&E	Textile	Food Proc.	Oth. Mfg.	Services	Mining	Real GDP
Thailand	6	3	-1	45	18	-7	303	0	366
Cambodia	7	0	0	27	10	0	166	0	210
Laos	0	0	0	0	0	0	0	0	1
Myanmar	-1	0	0	0	24	0	1	0	24
Vietnam	6	3	7	54	31	31	12	0	143
ASEAN10	18	6	6	125	84	25	482	0	745
World	23	4	7	104	74	25	545	1	782

Version 9.0	Agriculture	Automotive	E&E	Textile	Food Proc.	Oth. Mfg.	Services	Mining	Real GDP
Thailand	7	36	24	78	46	102	822	0	1,115
Cambodia	8	1	1	3	7	6	582	0	608
Laos	0	0	0	0	0	0	2	0	2
Myanmar	1	1	-1	-1	-44	-2	104	0	58
Vietnam	5	0	0	58	-4	-41	270	1	289
ASEAN10	22	44	31	138	6	75	1,771	1	2,089
World	34	70	32	93	-11	137	2,139	4	2,498

Source: Calculated by IDE-GSM.

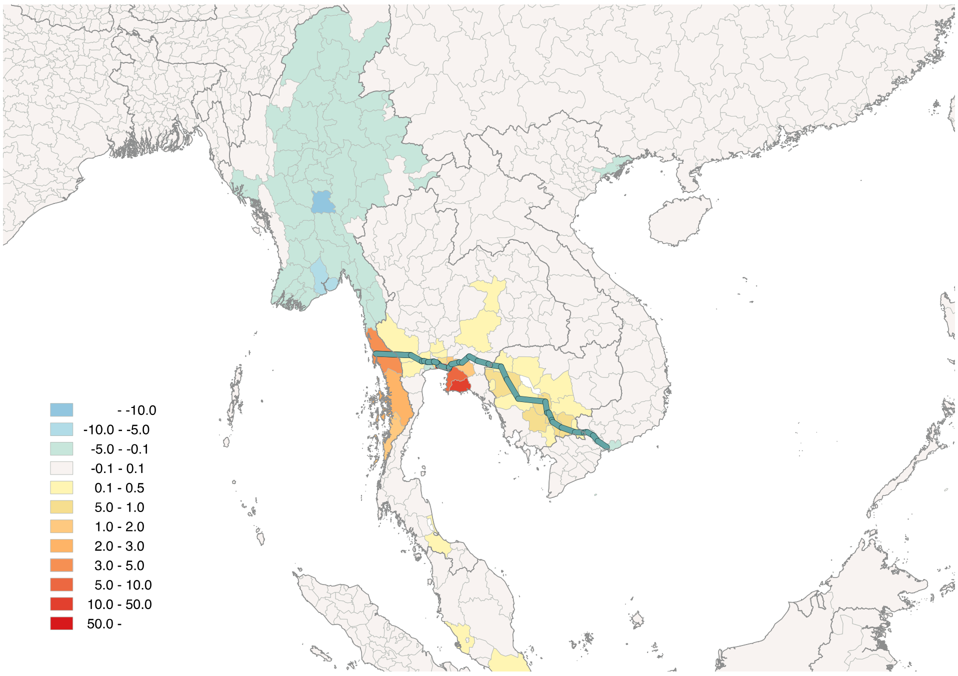
Figures 2-5a and 2-5b show the distributions of economic impacts for the food processing sector by the IDE-GSM versions 8.0 and 9.0. In version 8.0, the economic impacts for the food processing sector in Myanmar are dispersed all over the country. In version 9.0, the economic impacts for the sector in Myanmar were concentrated in the southern part of the country and negatively impacted other regions. An interpretation of this result is that, in version 9.0, the food processing sector has more IO relations with its own and other sectors, making the southern part of Myanmar more advantageous than other parts of the country.

Figure 2-5a: Simulated economic impacts from the SEC for food processing sector by version 8.0 (2030, compared with baseline, million USD)



Source: Calculated by IDE-GSM 8.0.

Figure 2-5b: Simulated economic impacts from the SEC for food processing sector by version 9.0 (2030, compared with baseline, million USD)



Source: Calculated by IDE-GSM 9.0.

In version 9.0 of the IDE-GSM, the IO table is taken from Thailand's data for 2005 and universally applied to all regions in all countries. It is desirable to introduce designated IO tables for all regions, but it is impossible to compile different IO tables for all regions. It is also desirable to adopt a different IO table if one region has a higher income per capita than 10 years ago. However, it is also difficult to automatically compile different IO tables for different income levels.

4. A fundamental trade-off in geographical simulation

As mentioned above, there is a fundamental trade-off between making the simulation model more realistic and increasing the number of required data and parameters, which makes it more challenging to obtain the correct data and plausible parameters. Thus, the simulation results are not always closer to reality. Because the IDE-GSM includes many developing countries as regions of analysis, the model has been extended to consider that the data and parameters required for the simulation should not increase significantly.

This trade-off between the realism of the simulation and the increase in the amount of data and parameters required is so fundamental that it is challenging to address. To overcome this constraint, it is necessary to consider how to create data and parameters with high accuracy, and several attempts are still being made by the IDE-GSM team.

The IDE-GSM produces geo-economic data at the sub-national level according to the statistical data obtained from each country's national statistical offices. However, many developing countries do not have access to economic data at the sub-national level.

For instance, official data on GDP by region are difficult to obtain in many African countries. An alternative way to estimate economic activities in developing economies is to use the strength and distribution of nighttime lights. It is now widely known that the strength of nighttime lights and economic activity are firmly correlated. Usually, the relationship between these two phenomena is determined by coefficients derived from regression analyses using "actual" data and nighttime light satellite imagery (Ghosh et al. 2010). In addition to nighttime light, we also consider land use to estimate the agricultural GDP of each region. Here, we use nighttime lights and land use to estimate the GDP distribution at the district level, taking the national GDP as given.

Besides the dataset at a more refined administrative district level, we also need to obtain various parameters associated with a finer industrial sector or a larger number of transport modes. So far, we have adopted some parameters, such as the degree of differentiation in goods, country-level NTBs, and transport parameters from different studies outside. On the other hand, we calibrate some parameters, such as bilateral NTBs, to replicate the actual bilateral trade volumes. To introduce a finer level of the industrial sector and more transport modes, we need to calibrate and estimate more parameters inside the model.

5. Conclusion

In this chapter, we reviewed the history of the expansion of the IDE-GSM. We also showed the possible direction of future development of the model and expected difficulties. We then discussed a fundamental trade-off in the simulation model; that is, the trade-off between the realism of the simulation and the increase in the amount of data and parameters required. To remedy this trade-off, we need to develop tools to compile the necessary dataset and calibrate the parameters endogenously automatically.

References

Fujita, M., P. Krugman, and A. J. Venables (1999) *The spatial economy: Cities, regions, and international trade,* MIT Press.

Ghosh, T., R. L. Powell, C. D. Elvidge, K. E. Baugh, P. C. Sutton, and S. Anderson (2010) “Shedding light on the global distribution of economic activity”, *The Open Geography Journal*, 2010(3), 148–161.

Hummels, D. (1999) “Toward a geography of trade costs”, GTAP working paper No. 17.

Krugman, P. (1991) “Increasing returns and economic geography”, *Journal of Political Economy*, 99, 483–99.

Krugman, P. (1993) “On the number and location of cities”, *European Economic Review*, 37 (2-3), 293–298.

Kumagai, S., K. Hayakawa, I. Isono, S. Keola, and K. Tsubota (2013) “Geographical simulation analysis for logistics enhancement in Asia”, *Economic Modelling*, 34, 145–153.

Kumagai, S., and I. Isono (2011) “Economic impacts of enhanced ASEAN–India connectivity: Simulation results from IDE/ERIA-GSM” in Kimura, F. and S.

Umezaki (eds.), *ASEAN–India connectivity: The comprehensive Asia development plan, Phase II*, ERIA Research Project Report 2010-7, ERIA, 243–307.

Nihon Kaiun Shukaijo (1983) *Distance tables for world shipping (Eighth edition)*, Nihon Kaiun Shukaijo.