Simulating the structure and localization of activities for decision making and freight modeling: the SIMETAB model.

Mathieu Gardrat, Marc Serouge, Florence Toilier, Jesus Gonzalez-Feliu

12

SIMULATING THE STRUCTURE AND LOCALIZATION OF ACTIVITIES FOR DECISION MAKING AND FREIGHT MODELING: THE SIMETAB MODEL

Mathieu Gardrat, LET - University of Lyon, France
Marc Serouge, LET - University of Lyon, France
Florence Toilier, LET - University of Lyon, France
Jesus Gonzalez-Feliu, LET - University of Lyon, France

KEYWORDS: urban modeling, urban typology, urban goods movements

ABSTRACT

This paper aims to introduce the methodology of the SIMETAB model which attempts to simulate the economic structure of an area, and then produce the inputs needed by many urban goods simulation models. Developed in the SILOGUES[1] project, the proposed model aims at understanding the economic structure of the city, in order to simulate the urban goods movements depending of the activity using an existing mode. This model's main function is to simulate the number of establishments by category of size and field of activity for a given zone, or to build evolution scenarios from a starting situation: starting from a known structure (or simulated through SIMETAB), it is possible to make the economic structure of a city evolve. This feature allows the decision makers to project an alteration of the urban framework in the future, from basic data such as population and employment statistics.

[1]"SImuler la LOGistique Urbaine dans son Environnement économique et Spatial", a LET project which main target is to simulate the urban goods transport in its economic and spatial environment through several simulation models (ex: FRETURB, SIMETAB…)

To INDEX

INTRODUCTION

The knowledge of the economic structure of cities is a critical point in the decision-making process related to urban goods and urban planning interactions. The structure of the activities defines the logistics behavior of an area. This reflection finds its source in the FRETURB urban goods models (Routhier and Toilier, 2007). As urban goods movement models need to be precisely calibrated on the structure of the studied area, the knowledge on the activity of a given area becomes critical for decision makers. This problem is moderately relevant in France where a national establishment database is easily available (i.e. SIRENE files, which are censorial databases of each town's economic activities). In this case decision makers have a powerful tool of knowledge and high precision. However, in other countries this sort of database is sometime not available publicly or even existing.

Modeling the economic activities of a city is fundamental when it comes to describing the urban goods transport behaviors. It was proven thanks to the urban goods transport surveys carried out in France between 1994 and 1997 in three different urban areas that the type of activity of an establishments defines its urban goods movements behavior (Patier and Routhier, 2009). Homogeneous strata of activities showed for example the same ratios in terms of generated movements for one employee (a warehouse generates approximately ten goods movements per employee, whereas a tertiary office only generates 0.3 movement per employee), whether it was in a small or a large city. This is why the SIMETAB model, by describing the activities implanted inside urban areas, can be an input for urban goods movements estimation, through activity based models or even simple ratios.

Another critical point brought up by this problem is the evolution of the structure of activity in a city. Indeed, it is hard to determine evolution scenarios with basic data such as the population and the employment. Destruction, creation and relocations of establishments are fundamental information for the transport and the urban planning process. Data can be found on this subject (ex: INSEE[2] databases in France), but are not relevant in prospective approach. As the work carried out on this particular subject is not very advanced, this point will not be mentioned in the paper.

We can however tackle shortly the hypothesis behind the evolution of urban activities. The main assumption of this model is that a given French urban area, through its changes in population and employment, is likely to evolve into another type of urban area already described in the classification used to calibrate the model. This hypothesis is the key to model future changes in activities inside French urban areas. Since data were only collected in continental France, we can only assume that once the model is fully calibrated, it will only be applicable to the continental French territory without adaptations needed. Furthermore the verification of this hypothesis would need a temporal analysis based on extensive data collected over years to infirm or confirm this hypothesis. This work is not yet fully completed so we will not be able to communicate on the capacity of the model to describe evolutions of urban structures over time.

GENERAL METHODOLOGY

The aim of this paper is to present a methodology to understand the territorial dynamics and urban systems (Antoni, 2011), taking into account their characteristics in order to build a typology of different urbanized territories. This work is essentially linked to the experience of the LET in terms of knowledge of the economic and functional behavior of the

[2] The INSEE is the national institute for statistics and economic studies in France

establishments in the urban environment (noticeably the work of stratification realized for the French surveys on urban goods movements and FRETURB; Patier and Routhier, 2009). Thanks to INSEE, IGN[3], and LET[4] data we were able to detail the activities of continental France at the "commune[5]" level, with data concerning the population, the employment (number and structure), the number of establishments, their size and field of activity, the surface area, and other geographical data (such as the coordinates of each city). These data were then used to calibrate the model.

We will first describe the methodology used to build the methodology, then how were built the various other components of the model. The third part is an evaluation of the results obtained through a series of test, to finish with a reflection on the possible further improvements of the model.

The main simulation steps defined for this model are the following:

- Generation of the number of establishments for each zone starting from easily available data such as the population, the employment, the surface, etc.
- Establishment definition, by giving to each establishment a size and a field of activity.
- Construction of evolution scenarios of the economic structure of the given area (number of establishments of each type in each spatial unit).

The next figure synthesizes the modeling process of the economic structure of an area.

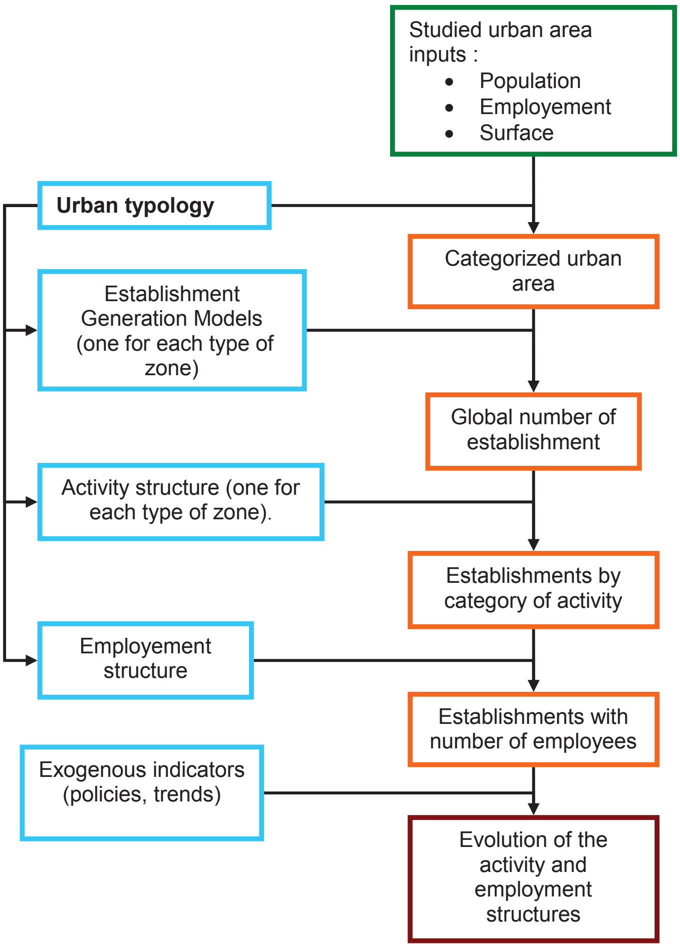

Figure 1 Simulation process of the SIMETAB model

[3] The IGN is the national geographic institute in France
[4] "Laboratoire d'Economie des Transport" in Lyon
[5] A commune is the basic geographical and administrative fragmentation in France, that can assimilated to a town. During the rest of this paper, this term will be used to describe the studied zones

To INDEX

The core of the methodology is the definition of the typology of the urban area through the characteristics available at a national level. The construction of this typology is in fact based on the analysis of the 36 000 French communes (the basic administrative fragmentation in France for cities). This division of the French territory is both political and historical and does not necessarily reflects a functional or land-use homogeneity. The reason why it was chosen as a basic unit is because it is a standard and data collection on this scale is the most efficient and complete.

Through a PCA and an optimized k-means algorithm, the number of urban categories is firstly defined, each one corresponding to a precise type of urban zone, for example: highly residential, lower density area, high tertiary activity, among others. This typology is on of the main inputs of the SIMETAB model, along with the basic data on the studied area (population, employment, surface of each zone). A multiple discriminant analysis is used to classify the zones of the studied area and to allocate a type on each one. Then, for each type of urban zone, multiple linear models are fitted to determine the total number of establishment of each zone. Those linear regression models, along with the multiple discriminant analysis were calibrated using R (Venables *et al.*, 2011). The corresponding activity of each establishment is then defined through the economic structure of each type of zone defined in the typology. The definition of the size of each establishment is the final steps in the modeling process. By breaking down the global number of employee of each zone through statistical data on the employment structure (defined in the typology), each establishment is allocated with a precise number of employees, therefore rendering the economic and employment structure of the studied area.

It is fundamental not to take only into account the characteristics of the commune itself, but also its local environment, taking the example of a 25000 inhabitants city: is it a small rural city, or is comprised in a much larger conurbation. For his reason, basic data related to the conurbation have at least to be included in the construction of the urban typology. Indeed, the dynamics involved in large urban agglomerations (such as the metropolization) are different than the ones involved in smaller cities (Mignot and Villareal Gonzales, 2004). Concentration, segregation, urban sprawl have to be taken into account in this urban typology.

To build the categories in the typology, includes basic data but at different spatial levels. Each zone defined in the database is included in another “macro-zone” defined by aggregated data on the zone. This means that one zone is characterized by its own characteristics (population, surface, employment…) but also by the characteristics of its local environment, meaning the population, employment, number of communes, and surface of the studied urban area. These indications allow to take into account the urban dynamics for very large conurbations, such as Paris, Lyon, Marseilles, etc., and also to refine the typology.

This typology is first built by categorizing every zone (commune) by taking into account the characteristics of the zone itself and of the urban area related to this zone. The communes are classified in 10 categories of population, ranging from the less populated (from 0 to 250 inhabitants) to the highest (starting from 100 000 inhabitants). The urban areas (as defined by the INSEE) are also classified in 11 categories of population: the first category (and smallest in population) being the communes that are not included in an urban area and considered as rural; the last category being urban areas including more than 10 000 000 inhabitants, namely Paris for this particular set of data.

The combination of these two categories provides with precision the characteristics of the commune and its environment. Hence, by merging both classifications, we can first define 96 types of communes only through the population characteristics. Note that there are not 110 categories due to the inexistence of some combinations. The classification is however not sufficient to obtain enough disaggregation and pertinence. That is why a PCA is performed for each of the 96 categories, in order to understand the existing differences within group, and to build more homogenous groups.

During this process the variables taken into account were: the population, the employment, the surface, the percentage of active people and retired people of the zone, the population of the urban area, the employment of the urban area, the number of communes in the urban area, the distance of the commune to the center of the urban area.

Thanks to this process we are then able to build coordinate tables for each of the 96 categories previously defined. Each category is then divided into several sub-categories in order to improve the homogeneity of the final categories. This new classification is made by using a k-means clustering that maximizes the inter-class distances and that minimizes the intra class distances. The final result of the division process is the generation of 219 homogenous classes of communes.

CALIBRATION PROCESS

The model is calibrated by knowing elementary factors of the commune which were described previously, but also by knowing the exact economic structure of the studied zone: we are hence in possession of the number of establishments of the zone, their precise activity and their category of staff (number of employees). The type of activity is classified in 700 categories. We however chose to aggregate these activities into 17 types to facilitate the data processing and the readability of the model. We will now see how a small set of communes is classified in the typology at the beginning of the modeling process.

Although we determined a refined typology for all the communes, it is hard to define in which group a hypothetical new commune would be classified. This question is pertinent when it comes to a city we don't know extensive data about. In order to retrieve in which category the commune is, a multiple discriminant analysis is used to classify the communes into the most pertinent class. The variables included in this process of classification are basic, easily available and consist in:

- the population of the commune,
- the number of employees,
- the surface of the commune,
- the number of communes of the conurbation,
- the distance to the center of the conurbation,
- the employment density of the commune,
- the population density of the commune,
- the number of employees of the conurbation,
- the population of the conurbation.

The success rate of the discriminant analysis is 60%, where the exact category of the commune is found amongst the 219 categories. The rest of the time the process sends back a very similar category to the real category of the commune.

The second phase consists in modeling the number of establishments of each zone through a few easily available variables, whatever their type of activity or size. For this purpose, regressions were fitted to generate the global number of establishments of each type of zone. In order to determine which variables are to be chosen, an analysis of variance is performed for all the types of communes defined in the typology. It appears that 3 variables are suitable for the construction of the regressions:

- the employment,
- the surface,
- the population.

To INDEX

For each category of commune, an analysis of variance was performed in order to determine which variables were significant. Out of the 219 tests of variance carried out, the next table indicates the number of times the variables were considered as significant in the determination of the number of establishments of a commune.

Table 1 Frequency for which the variables where considered significant

Variable	Frequency
Number of employees of the commune	159
Surface of the commune	126
Population of the commune	96
Distance to center	15
Number of employees of the conurbation	12
Number of communes of the conurbation	12
Population of the conurbation	6

The results are not surprising, and we can notice a large gap between the first 3 categories and the rest. The regressions were thus built with these 3 significant variables. Having 219 types of zones, the same number of regressions was created. These regressions were built on the totality of the statistical population of each category.

The economic structure of every type of commune is then determined through the average economic structure of all the communes included in the same category. The hypothesis underlying this method is that the classes are homogenous enough to apply an average structure. The variance is in fact quite low in most of the categories, except for the industrial activities (considered as specific activities).

The same methodology is used to affect a staff for each establishment, in the various categories of size defined by the INSEE. Thus, for each class of commune and activity, a structure of size is determined and applied to distribute the number n of establishments of the activity j into the class x of size. This leads to build probability tables used to distribute the establishments in the proper category of activity and size. There are as many tables as there are categories in the typology (in this case 219). The modeling process draws the establishments thanks to this sort of tables. The limit of this process is that it is not directly linked to the real employment of the commune defined in the typology. Thus corrections have to be made in order to increase or decrease the number of employees affected to each establishment to fit with the reality, consequently affecting the size of the establishments of the zone.

RESULTS

In order to analyze the robustness of the model we will now proceed to an evaluation of a simulation batch that was carried out on the 232 French urban areas that include more than 5 communes.

The final results concerning the number of establishments are quite satisfying as 80% of the results are between a deviation of -20% and 20%, as shown in the next graph. The overestimation (more than 50%) of the number of establishments mainly concerns the smallest urban areas. To find this result, we compared the number of establishments found for each of the 232 modeled urban areas to the actual number of establishments in each zone, thus testing if the model fits the reality. The next graph shows the distribution of deviations found for the 232 studied urban areas.

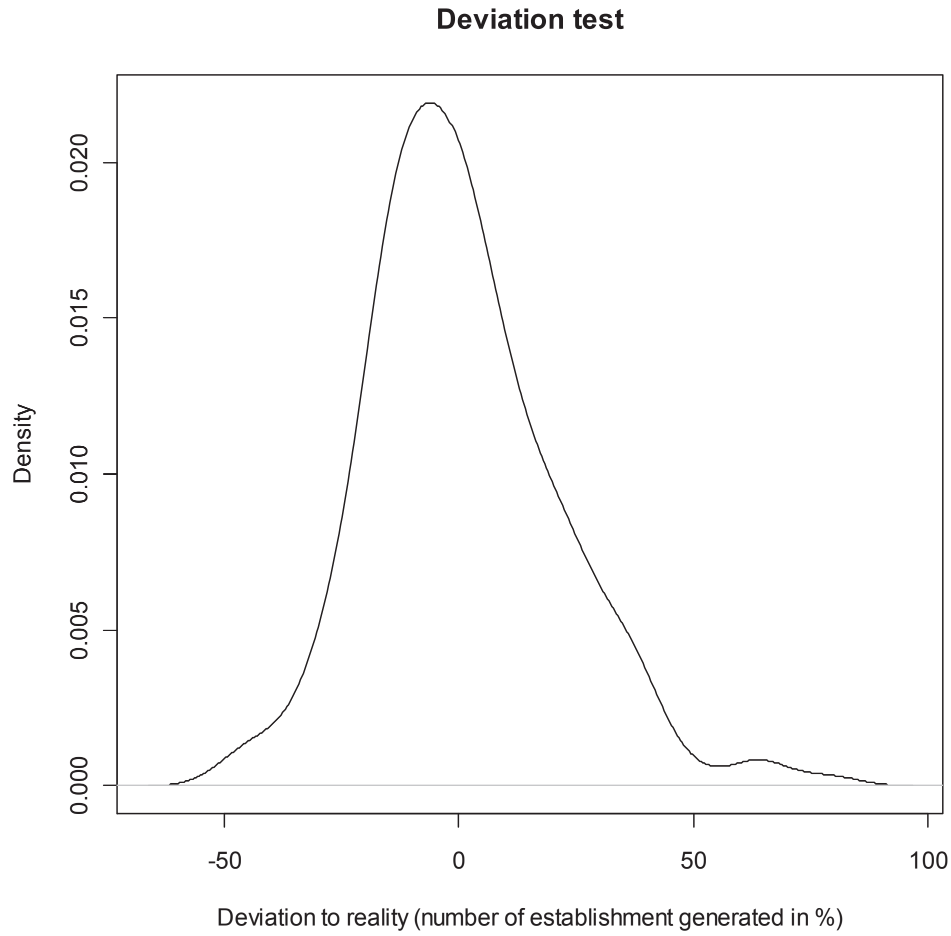

Figure 2 Deviation rates comparing the real situation and the establishments generated by the model (for all categories of communes)

The deviation found for the repartition of the establishments in their category of activity is satisfying. The difference is larger for industrial activities that are linked to other factors as the population. For others, the results are acceptable, though there is a real need for recalibration, or even the use of another method of modeling. The distribution of the deviation to the reality for each type of activity is shown in Table 3.

Industrial and agricultural activities (1 to 7) have the largest deviations. The mention “inf” indicates that some values produced were infinite, due to the fact that the real number of establishments of a certain activity was equal to 0, and that the model generated establishments in this category. The results are more acceptable for categories 9 to 17 except for the 12, 14 and 15 activities. The employment structure deals with the same problem, as the repartition of the size of the establishments in each activity is defined through the average structure of the category built previously. The graph in Figure 3 shows the deviation for the number of employees generated for each commune.

To INDEX

Table 2 Deviation rates comparing the real situation to the number of establishments generated by the model (in percentage of deviation)

	Type of activity	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
1	Agriculture, fishing and Forestry	-73.0	5.1	59.4	106.2	150.6	1176.0
2	Food, drinks and tobacco industry	-81.0	-18.3	0.0	2.3	20.0	121.7
3	Coking, refining	-100.0	-100.0	-66.7	Inf	0.0	Inf
4	Electric, electronic, IT and machinery industry	-100.0	-38.1	-7.6	Inf	28.9	Inf
5	Transport equipment industry	-100.0	-50.0	0.0	Inf	100.0	Inf
6	Other industrial products	-75.3	-17.9	0.0	5.1	21.8	160.6
7	Extractive, energy, water, waste and cleanup industry	-75.0	-18.5	4.2	12.3	30.0	500.0
8	Construction	-48.4	-16.1	4.7	9.4	22.9	123.5
9	Shops; cars and motocycles repear	-64.4	-18.8	-5.7	-3.9	8.7	126.1
10	Transport and storage	-66.7	-14.9	4.7	6.8	25.1	102.8
11	Restaurants and hosting	-72.4	-15.0	4.6	7.7	21.5	197.0
12	Information and communication	-69.6	-22.1	-5.5	Inf	19.4	Inf
13	Financial and Insurance activities	-85.7	-20.3	-5.4	-0.2	13.0	250.0
14	Real-estate activites	-71.3	-14.3	4.0	13.4	25.0	600.0
15	Sciences and techniques (including administration and support)	-52.4	-16.3	-2.6	4.0	14.9	205.8
16	Public administration, teaching, healthcare and social action	-69.8	-17.6	-4.8	-4.4	7.2	84.7
17	Other services activities	-59.6	-17.4	-5.7	-4.0	6.1	72.4

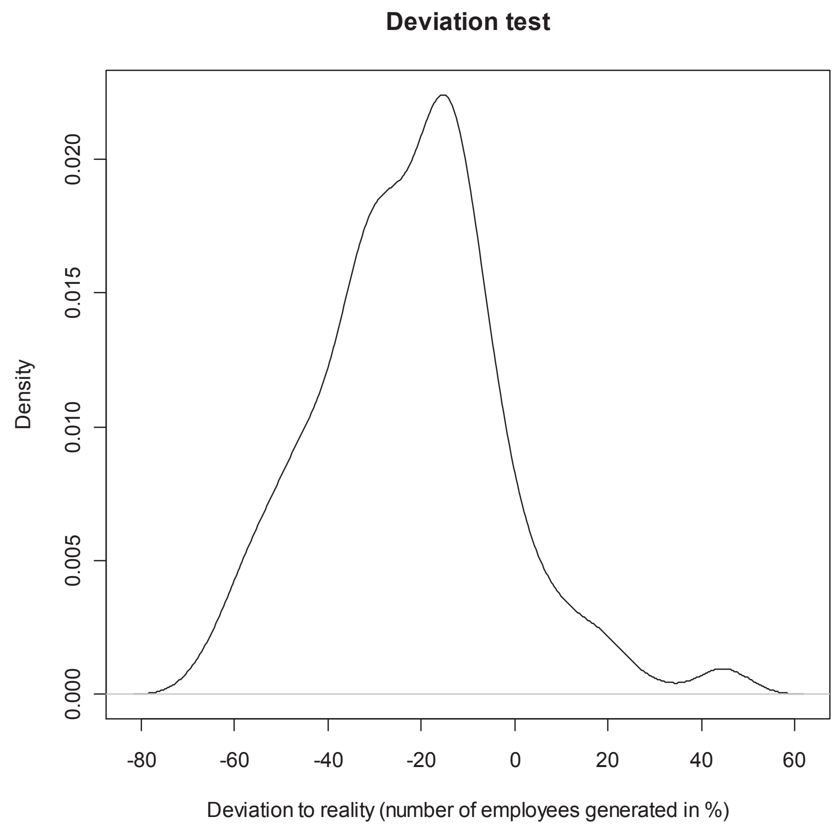

Figure 3 Deviation rates comparing the real global number of employees and the employees generated by the model (using the middle class staff), all categories of activities and communes included

It appears clearly that the number of employees (and possibly the size of the establishments), is underestimated by 20% (average -20.44, median -22.11). The cause of this deviation is mainly due to the fact that for each class of size, we chose to affect the average number of employees calculated between the extremes of the class as indicated in the next table produced by the INSEE:

Table 3 Table representing the number of employees for each class of size (source: INSEE)

Class Code	Minimum number of employees of the class	Maximum number of employees of the class	Middle class number of employees
0	1	1	1
1	2	3	2
2	4	6	5
3	7	10	8
11	11	20	16
12	21	50	36
21	51	100	76
22	101	200	150
31	201	250	226
32	251	500	376
41	501	1000	750
42	1001	2000	1500
51	2001	5000	3500
52	5001	10000	7500
53	10001	15000	12500

However, by using the upper extreme of each class of staff size, we can notice that the distribution of deviations is almost centered on 0 (average -0.55, median 2.25), as indicated in the next graph

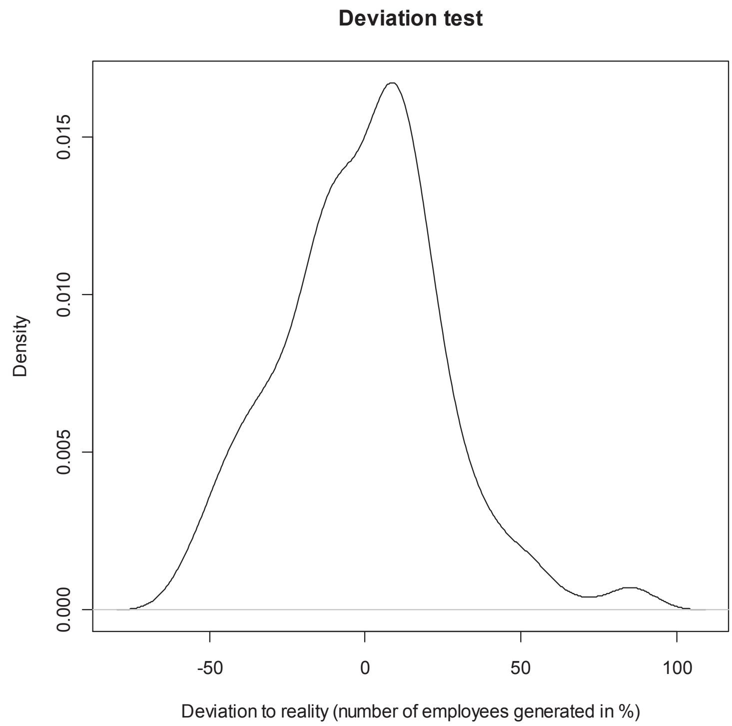

Figure 4 Deviation rates comparing the real global number of employees and the employees generated by the model (using the upper class staff), all categories of activities and communes included

The results of this model, if not completely satisfying for now (especially concerning the employment structure), make way for promising improvements concerning the structure of the activity and the employment generation.

FURTHER IMPROVMENTS

The main improvements that have to be done mainly concern the modeling process of the economic structure (distribution of the types of activity). As seen in the previous part the main difficulty met, was the definition of the specific activities of the city that do not support directly the urban population needs (industry and agriculture).

We differentiate common and specific activities as followed:

- Common activities (Beguin, 1995) are characterized by their direct contact with the population with the underlying purpose of sustaining and responding to the needs of the populations, such as supermarkets, bakeries, clothes shops, etc.
- Specific activities (Manzagol, 1995) are disconnected from the populations (except for the employment) and tend to communicate exclusively with other companies. Their localization and existence is due to multiple vectors, which some are not linked to the urban morphology (history, tradition…).

For the first category, the SIMETAB model can easily determine, the number, size and type of establishments for several reasons:

- These activities are linked to the population needs (food, clothing, leisure, etc.)
- They are usually equally spread across a territory (with rare exceptions), for they maximize their possibility to meet demand.
- Consequence of the last point; he rarest activities are concentrated in the largest urban zones, as the offer for the rarest products takes place where they have the best chances meet the demand, i.e. the largest cities.

Thus, if the rendering of the model is not perfect for this category, the results are more coherent compared to the simulations of specific activities.

This second category is in fact much more difficult to build as it seems complex to model the localization of a car factory, with only the employment, the population or other basic geographical data, as additional factors enter in the determination of the location of this type of activity (history, policies, etc.). Complementary data are needed to build a model for specific activities generation.

Though specific activities are less represented in the statistical structure in terms of number of establishments, their weight in the employment and in urban goods movements are non-negligible. The simulation of these activities is consequently fundamental for urban goods movements modeling, especially for suburban spaces that usually welcome these sorts of establishments.

Concerning the spatial distribution of the activities, the use of central place theory models could be used to create a localization module, were the establishment could be spread across the space, taking into account the density and the size of the modeled area, and also the road and public transport network. This module would mainly be used for common activities, as they are closely linked to the population.

Modeling the localization of specific activities might be the most complex task of the model. The model can however model the number of establishments with a relative robustness, when it comes to determine aggregated classes of specific activities (i.e. the specific activities are divided in not more than 10 categories). To attain this objective, complementary inputs should be integrated.

To INDEX

The next phase of improvement includes the calibration of a new urban typology through smaller geographical zones. The main disadvantage of the commune level is that it is too large to render the sharpness of the urban frame: a large commune such as Toulouse (approximately 449,000 inhabitants) is characterized by a variety of urban sub-spaces (residential, commercial, tertiary…) that the communal division cannot represent. In this context a finer division of the urban territory allows to make emerge more specific spaces, characterized by a unique economic structure.

This finer division also allows the generation of more data for calibration. The SIMETAB model, being essentially fed by statistics, finds its reliability in the data used to calibrate the model. By increasing the number of individuals in the statistical population, the process is refined to gain in precision.

Finally, the evolution of the economic structure is, as the simulation of specific activities, a difficult subject to tackle. The mechanisms behind the changes of the economic structure of a city are bound to a vast number of parameters that have to be clearly defined and linked together. Urban policies in terms of real-estate, investments and regulations as well as the characteristics of the population are (among others) key factors of the urban evolutions. The integration of this point will be linked to the construction of basic indicators on exogenous factors.

CONCLUSIONS

The SIMETAB model, by defining the urban category in which a given area is, can provide precise information on the activity of the studied zone and can also be an input for the FRETURB model, allowing a better understanding of the activities and subsequent transport behavior, thus helping the decision making process. Given that homogeneous strata of activities share the same urban goods behavior, the SIMETAB model can be a start for basic urban goods diagnosis through the ratios that were defined in France through surveys.

With an appropriate calibration process, this model will ease the use of the FRETURB model (further explanations in Routhier, Toilier, 2007) at an international level (reasonably European), by building an establishment database to substitute any other database related to the economic activity of a city. Also, the possibility of building evolution scenarios should be a powerful feature, allowing the prediction of goods movements and economic activity in several possible futures. Other works carried out on this model tend to generalize the simulation process on the whole urban framework and not only on its economic structure: households sizes and activities, attractiveness, accessibility, road network imprint, etc.

REFERENCES

Antoni, J.-P. (2011). Des éclairages pour modéliser la ville, in Modéliser la ville, Paris, Economica, 2011.

Beguin, H. (1995), « La localisation des activités banales », in BAILLY, A., FERRAS, R., PUMAIN, D.(dir), *Geography encyclopedia,* Paris, Economica, 1995.

Comi, A., P. Delle Site, F. Filippi and A. Nuzzolo (2012). Urban Freight Transport Demand Modelling: a State of the Art, Trasporti Europei/European Transport, Vol. 51, Paper n° 7, 1-17.

Gonzalez-Feliu, J. and J. L. Routhier (2012). Modeling Urban Goods Movement: How to be oriented with so many approaches? Procedia - Social and Behavioral Sciences, Vol. 39, 89-100.

Manzagol, C. (1995), « La localisation des activités spécifiques », in BAILLY, A., FERRAS, R., PUMAIN, D.(dir), *Geography encyclopedia,* Paris, Economica, 1995.
Mignot, D., Villarreal Gonzales, D. (2004), « Métropolisation, concentration et ségrégation dans les villes européennes et sud-américaines : une comparaison Costa-Rica, France et Mexique ». proceedings of the fifteenth ASRDLF conference, Bruxelles, 2004.
Patier, D. and J. L. Routhier (2009). How to improve the capture of urban goods movement data? In Bonnel, P., Lee-Gosselin, M., Zmud, J. and Madre, J.L. (Eds) (2008), Transport Survey Methods. Keeping Up With a Changing World. Emerald, Bingley.
Routhier, J. L. and F. Toilier (2007). FRETURB V3, a policy oriented software of modelling urban goods movement. Proceedings of the 11th World Conference on Transport Research—WCTR'07, 24–28 June 2007, Berkeley, USA (2007).
Venables, W. N., D. M. Smith, the R Development Core Team (2011), An Introduction to R, http://www.r-project.org/.