An integrated land use-transport model for the Paris Region (*SIMAURIF*) : Ten lessons learned after four years of development

Dany NGUYEN-LUONG

Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France (IAURIF)
15 rue Falguière, 75740 Paris Cedex 15, France.
dany.nguyen-luong@iaurif.org

Abstract : After four years of developing of an integrated land use-transport model for the Paris Region, we have learned a number of lessons. This paper presents ten of these lessons from the point of view of a practitioner, after presenting the background of the project, and then proposes a few perspectives.

Keywords: transport systems, land use modelling, integrated land use-transport model, LUTI, simulation, calibration, UrbanSim, Davisum, Metropolis

1. INTRODUCTION

SIMAURIF is an integrated land use-transport model for the Paris Region that has been developed over the last four years by the Institute for Urban Planning and Development of the Paris Ile-de-France Region (IAURIF), in collaboration with the THEMA laboratory of the University of Cergy-Pontoise. We completed the project in December 2007. It is based on two existing models : OPUS/UrbanSim - a land use model developed by the University of Washington, and Davisum-METROPOLIS - the transport model including the travel demand model developed by IAURIF. This project was funded via a grant by PREDIT (Inter-ministerial transport research and innovation programme), DREIF (Governmental department of transport in Paris region) and RFF (French Railways Network).

SIMAURIF is the first project using OPUS/UrbanSim in the world, outside the USA, to be completed. Although there are many urban areas in the world that are developing integrated land use-transport models using OPUS/UrbanSim (e.g., Seattle; Detroit; Honolulu; Houston; Phoenix, Arizona; San Francisco; Burlington, Vermont; Durham, North Carolina; Melbourne; El Paso, Texas; Amsterdam; Tel Aviv; Zurich; Pescara, Italy; Lyon, France), none of these cases has resulted in a completed model to date.

This paper is addressed to Metropolitan Planning Organizations as well as to research laboratories that are using or contemplating using OPUS/UrbanSim. The reader can be reassured : there are no mathematical formulas in this paper. We present the mains lessons of such a project from the point of view of a modeller-practitioner, not a researcher. So we are aware this paper is not very academic, but it will be quite objective and will reflect the reality of the daily practise. We hope it will be useful to the research community because it contains an abundance of practical advice and raises some legitimate concerns about developing integrated land use-transport models.

Before presenting the ten lessons, we present some background about the *SIMAURIF* project and the context of transport modelling in France.

2. BACKGROUND

In France, there are many classical transport models, of varying degrees of sophistication, which make use of commercial software such as PTV's Davisum or Citilab's Cube. On the other hand, there are no examples of integrated land use-transport models in France. By contrast, in the U.S., this kind of model is more common, in part, because of laws such as ISTEA and TEA-21, which require regional transport plans to take into account the interactions between land use and transport. So, in the U.S., one finds applications of land use model software like MEPLAN, DRAM/EMPAL, TRANUS, and, more and more, UrbanSim. In 2003, we decided to choose UrbanSim for our project. The main reason was that it contained interesting capabilities versus the other packages and above all, it was free (GNU licence). The aim was to develop an operational model that could test an isolated transport scheme or simulate transport policies or land use policies and show the effects on urban development or on mobility behaviour.

The second impetus for our project was that in France, in the field of socio-economic assessment of infrastructures, the cost-benefit method often does not satisfy decision makers, even though many scientists and engineers find this method to be robust and powerful. To most decision makers, terms such as "internal benefit-cost ratio," "time savings," or "externalities" of an infrastructure project have little meaning. So, our project had another objective : to allow one to measure the indirect effects of providing infrastructure, and not only the direct effects (e.g., time savings and environmental savings). What are these indirect effects? Residential attractiveness (e.g., the number of households that will move near an infrastructure project ten years after its opening), economic development (e.g., the number of jobs attracted around or along a new infrastructure project ten years after the project is built), and, finally, the increase in land prices or housing prices inducted by the new infrastructure. So we aimed to develop a model that would be able to quantify these indirect effects and to include them as new criteria, similar to an internal benefit-cost ratio in a multi-criteria approach.

Next, we present our ten lessons.

3. TEN LESSONS

Lesson 1 : From theory to practice : a difficult journey

A review of the literature in 2003 and again in 2007 concerning integrated land use-transport models showed that the literature about this kind of model is typically very theoretical (e.g., focusing on aspects such as localisation models, hedonic land price models, activity-based models, etc.), but nothing was found about practical development and applications. And even if an application is presented in a paper or in a report, no details are provided about the construction of the databases or the whole calibration, which are usually the most difficult steps. Actually, in the beginning of such a project, most people underestimate the difficulties involved in developing these models. The more you advance in the project, the more you find out about complexities that you did not foresee in the beginning. So, the first lesson is : if you underestimate the difficulties in a project, you will be obliged, by the end of the project, to simplify or to sacrifice some tasks and/or to take short cuts because of the deadlines.

Lesson 2 : A methodology to follow

We have developed a methodology that should gain widespread acceptance. This methodology is based on the following principles : Take into account environmental constraints (e.g., flood zones, historic preservation areas, etc.) and political constraints. Maintain a consistency between inputs and outputs of the different models. Do not hesitate to adapt the data structure of OPUS/UrbanSim to the situation at hand in the study. Exploit, to the maximum extent possible, fine-grained geographic data, thanks to a geographical information system (GIS). Avoid simulating data or synthesizing missing data that will be used for calibration. Make a descriptive analysis of the data before using it for modelling. Build a single calibration database for all the models. Keep in mind that after the independent estimation of each sub-model, there will a step of assembling all the models together and calibrating the whole model that will be tedious and very time-consuming (e.g., a simulation lasts seven hours on a standard PC and you should run the model at least one hundred times). Last, distinguish clearly the calibration period and the simulation/application period.

Lesson 3 : It's worth building a huge database at the grid level

The saying "Garbage in, garbage out" is well known. It's interesting to note that almost all modelling studies conducted by researchers present extraordinary results, despite the fact that the data used in the models are barely described, and/or the quality of the data is often mediocre. In our project, we have, for the first time in France, built a huge database in a grid of 50,000 cells covering the entire Paris Region, with each cell being a 500 meter by 500 meter square. But is this grid relevant, and if so, is it worth the trouble, given that its construction is very time consuming? We answer, "yes," except in the case of the land price model. We were unable to estimate a land price model at the cell level, due to the large size of the study area (about 12,000 sq km). On the other hand, the grid allowed us to incorporate two interesting elements : the possibility to take into account the notion of a neighbourhood (each cell interacts with its eight neighbours) and the possibility of outputting results at the cell level (around stations, for example). Of course, the reliability of the results at that level remains to be shown due to the margins of error.

Lesson 4 : Explanative model versus predictive model : finding the compromise

Our experience with the project shows that you are better off building a simple model, with few variables (ten at the most), than building a very sophisticated model, with plenty of variables. This is especially true if you included simulated variables, perhaps because you felt they were important, but, since they were not available, you decided to create them by estimating another sub-model. The problem is that, once you've estimated this explanative model, you are likely to find out it is impossible to use it in a simulation. In our project, we made the mistake of spending too much time in the estimation of each model, and, in the end, we were unable to implement the models in UrbanSim, so we had to go back and re-estimate simpler models.

In the famous novel by Michel Tournier (1967) entitled, *Vendredi ou les Limbes du Pacifique* ("Friday"), there is a passage where, after reading the Bible, Robinson decides to join Chili by building a boat he calls "The Escape". Robinson devotes all his energy and imagination to building this boat. But, the day the boat is finished, Robinson suddenly realizes the boat will never float because it's too heavy. The same lesson applies to modelling : if your model is too sophisticated, you won't be able to apply it in forecasting because the variables which were chosen for estimation cannot be forecasted.

Lesson 5 : UrbanSim, which was designed for American metropolitan areas, has to be converted to the French context

Our project has shown that it's difficult to convert a model which simulates urban development in the North American context to another very different context. UrbanSim was designed to simulate American metropolitan areas, which are typically characterised by an uncontrolled urbanization, which spreads over a big area, but results in relatively low urban densities and is characterized by a transport system, dominated by roads with relatively high speed and capacity. In France, by contrast, city development is more controlled by government plans and intervention, which is characterized by an efficient organization of space and not so much by adding to the road network. For example, in the data structure of UrbanSim, there is no place to integrate public transport variables. So we had to work hard to adapt UrbanSim to the case of the Paris Region, sometimes resulting in simplifications that were not intellectually satisfying.

Lesson 6 : Developing an interface to manage the interaction between the transport model and the land use model : a partially reached objective

This part of *SIMAURIF* is easily understandable. But here again, it's the passage from theory to practise which is complex. We wanted to automate the following steps : execution of the classic "four-step" transport model, exporting the results of the transport model into the UrbanSim input database, running the four components of UrbanSim, exporting the results in the UrbanSim database to the trip generation step of the travel demand model, then running this cycle each year during the period of simulation, by modifying the road and public transport networks every three years (all with the push of a button!). We concede that we have only partially reached this objective. We had developed, during the second year of the project, a prototype to interface between UrbanSim 2 and Davisum-Metropolis. But the last year, we had to abandon this interface because of the re-engineering of UrbanSim that was part of its Version 4 update. So the current interface has been simplified to its simplest form : a script written in Python language, which takes into account every three years the supply of transport and only for off-peak/unconstrained conditions (i.e., without assigning the origin-destination matrices). The transport variables are computed externally and are then formatted into text files which are read by the script.

Lesson 7 : It's vital to have an excellent knowledge of the study area (development, demography, economy, transport, environment)

To make the project successful, it's necessary to have a global vision of the interactions between transport and urbanization, a good understanding of the development process occurring in the study area, a "perfect" knowledge of the urban projects and transports projects, an excellent knowledge of the geography of the employment and population, a "perfect" knowledge of the land use evolution, and, finally, a good understanding of the game theory that governs the actors or stakeholders in the system. In other words, the risk in conducting a project such as this is to think that it is simply an exercise in econometrics and computer science, and that you could simply estimate the logit models for Paris Region in the same way that you would for San Francisco.

Lesson 8 : *SIMAURIF* allows one to measure some indirect effects (e.g., residential attractiveness and economic development caused by an infrastructure), but not yet the housing prices.

SIMAURIF was applied to assess the effects of the North Tangential, a proposed tram line in a northern suburb of Paris, which is scheduled to open in 2016. For the first time in France, an integrated model was used to forecast the spatial distribution of population and jobs at the cell level for a horizon year of 2026, and to distinguish a difference in the distribution of populations and jobs between the scenario with and the scenario without the infrastructure. The GIS allows one to create maps, once grid cells results have been transferred, and to show interesting results around the stations. On the other hand, *SIMAURIF* was not able to simulate housing prices in 2026 at the cell level, but only at an aggregate level. So a tool like *SIMAURIF* can be used in a multi-criteria analysis to evaluate the indirect socio-economic effects of an infrastructure project.

Lesson 9 : UrbanSim, which is a product of the research world, is so difficult to master that we do not recommend it for use in France if you have to get the model up and running in a short time frame

The latest version of OPUS/UrbanSim, Version 4, has a very different logic from that used in versions 2 and 3. In short, Version 4 is about scripting and macrolanguage, whereas the previous versions took the approach of a "black box" or hidden system. For a new user, the "black box" approach is preferable, since it relies on pre-established and inter-connected models, which you just have to estimate, without worrying about additional implementation issues. In earlier versions of the software, the practitioner could concentrate on his job, which is modelling, not implementation. In the beginning of the project, we were really enthusiastic about the choice of UrbanSim : the architecture looked really interesting and relevant. It was comprised of four clear blocks : a residential location choice model, an employment location choice model, a land price model, and a development type choice model. But the shift to a new version in 2007 meant that we had to adapt and simplify all the databases we had built in the previous version. Moreover, the documentation wasn't updated at the same speed as the software. It appeared, based on various messages sent to the UrbanSim users' mailing list, that this new version of UrbanSim had really destabilized a number of current projects around the world, and that the users were having many difficulties because of purely software-specific problems.

Another important problem is that the interaction between the four main components of UrbanSim is not, in our opinion, clearly designed. Finally, the choice of the Python language for the new version was not, according to us, the best choice because it requires one either to learn this new language or to find a computer programmer, who would work on only this language.

Lesson 10 : It's necessary to put together an inter-disciplinary team and to allocate adequate resources

Classical transport modelling requires only abilities in modelling and applied mathematics. By contrast, modelling the interaction land use and transport requires one to have an additional skill set : econometrics, demography, economics, geography, urban planning, complex systems, mobility behaviour, and computer science. A project like *SIMAURIF* takes at least three years and a team of at least four people. So you have to put together a stable team for that period. If the objective is to result in an operational tool, you also have to make sure that the project is not dominated by econometrics experts or computer experts. The modeller has to remain the head of the project and limit the ambitions of his colleagues. It seems to us that the current UrbanSim development team at the University of Washington is now controlled more by econometrics experts and computer experts than by urban planners and modellers. As a result, OPUS/UrbanSim has become a project for computer science PhD students. But to our knowledge, the development team is working hard on addressing most of the concerns presented in this paper through new software enhancements (e.g., development of a graphical user interface and a tutorial).

4. PERSPECTIVES

In the short term, it seems worthwhile to apply *SIMAURIF* to other important transport planning studies in the Paris Region, such as the Arc-Express project. The database for the region is already built, the model is calibrated (even though the calibration could always use improvements), the transport model networks for the base scenario have already been built (one for every three years from 1999 to 2026). You just have to define the scenarios, build the corresponding modelling networks and run the model. Furthermore, we'll have to enhance the value of this project to the academic community and also to other potential users in the Paris Region.

In the mid term, it seems that if you want to get a fully customized piece of software and, above all, one that you can control easily without using Python, it is probably better to develop your own land use model. You might choose to keep the global architecture of UrbanSim (without dropping any sub-models, even the most difficult one, the Development Project Location Choice Model which aims to limit the capacity of the other location choice models) and the grid data structure, both of which are really quite relevant, and then implement your own model in a classical procedural language (e.g., Visual Basic, C). You can also program the travel demand model (the first three steps) and program easily the feedback between your own software and an assignment software package, such as Davisum-METROPOLIS.

In the long term, it would be interesting to connect the integrated model to an environmental assessment model (pollutants emission and dispersion model, noise propagation model) in order to calculate sustainable development indicators, and upstream, to couple a macro socio-economic model which can feed with projections of population and employments the totals controls tables for each year of the period of simulation.

Despite the difficulties encountered in this project and any shortcomings with the current *SIMAURIF* model, this project has allowed us to advance the state-of-the-practise in the modelling of the interaction between land use and transport in France.

For more information about *SIMAURIF*, three reports in French are available on http://www.iaurif.org (Ressources documentaires/Consultation en ligne : bibliothèque) or directly via the following links :

- *SIMAURIF*. Modèle dynamique de SIMulation de l'interAction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport final de la 2ème phase (2007)
http://www.iaurif.org/exl-doc/IA48950_OUV00005358.pdf

- *SIMAURIF*. Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport intermédiaire de la deuxième phase (2005)
http://www.iaurif.org/exl-doc/IA46029-2_OUV00002471.pdf

- *SIMAURIF*. Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport final de la 1ère année (2004)
http://www.iaurif.org/exl-doc/IA46029-1_OUV00001603.pdf

Acknowledgements
I wish to thank Mark Moran from the Metropolitan Washington Council of Governments (MWCOG) for his comments and corrections.