A survey on the calibration and validation of integrated land use and transportation models

Patrick Bonnel, Nicolas Coulombel, Emmanuel Prados, Peter Sturm, Elise Arnaud, Cédric Boittin, Louafi Bouzouina, Jorge Cabrera Delgado, Thomas Capelle, Jean Delons, et al.

A survey on the calibration and validation of integrated land use and transportation models

Patrick Bonnel[1], Nicolas Coulombel[2], Emmanuel Prados[3], Peter Sturm[3], Elise Arnaud[3], Cédric Boittin[4], Louafi Bouzouina[1], Jorge Cabrera Delgado[1], Thomas Capelle[3], Jean Delons[5], Laurent Gilquin[3], Vincent Hilaire[4], Matthieu de Lapparent[6], Dany Nguyen-Luong[7], Jean-Pierre Nicolas[1], Mathieu Saujot[8], Anthony Tschirhard[3], and Arthur Vidard[3]

[1]Laboratoire d'Economie des Transports, Ecole Nat. des Travaux Publics de l'Etat, Vaulx-en-Velin, France
[2]Laboratoire Ville Mobilité Transport, Université Paris Est, Marne la Vallée, France
[3]Laboratoire Jean Kuntzmann, INRIA Grenoble Rhône-Alpes et Université de Grenoble, France
[4]Laboratoire Systèmes et Transports, Université de Technologie de Belfort-Montbéliard, France
[5]Cofiroute, Sèvres, France
[6]Laboratoire Dynamiques Economiques et Sociales des Transports, IFSTTAR, Marne la Vallée, France
[7]Institut d'Aménagement et d'Urbanisme Île de France, Paris, France
[8]Institut du développement durable et des relations internationales, Sciences Po, Paris, France

1 Context and motivation

A fundamental goal of Land Use – Transport Interaction (LUTI) models is to capture the strong interplay between land use and transportation in metropolitan areas. Inherently, sector-specific models, transport and urban alike, cannot take this interaction into account and thus miss one side of the story. LUTI models aim to fill this gap, and ultimately to provide better decision helping tools for urban and regional long term planning. LUTI models first appeared in the 1960s. The complexity of LUTI models, combined with computational limitations at the time (see [3] for a discussion on these points), caused a pause in their development. However, interest in LUTI models has risen again in the 1990s and their number and complexity have been growing steadily ever since. This goes hand in hand with increasing expectations from end users as well as with new theoretical developments and a drastic increase in computational capacities, the latter enabling for instance the development of micro-simulation models.

Following the glossary set by David Simmonds Consultancy *et al.* [2], currently the two main families of LUTI models are spatial economics models (*e.g.* MEPLAN, TRANUS, PECAS, RUBMRIO, ...) and activity based models (*e.g.* UrbanSim, IRPUD, DELTA, ...). The former are essentially based on an equilibrium approach[1], which comes together with a relatively aggregate representation of space and of the economic agents and activities. The latter focus on the system dynamics and on the processes of change. They aim at more realism in how things are represented and are usually much more disaggregated. Micro-simulation is also widespread in this branch.

While spatial economics models may be mathematically more complex to handle, due to the requirement of reaching an equilibrium for a complex set of parameters and equations, activity-based models usually have more parameters and larger data requirements for their instantiation. Besides these differences, all LUTI models share several needs, among which – like for most models in general – a need for

[1]Several spatial-economic models are not exactly equilibrium models as they consider some dynamics in the land-use system. Typically, when there is an exogenous change in the system (new transport line, increase in population, etc.) they do not shift from one long-term equilibrium to another instantly: some friction factors ensure a smoother transition between the two steady states.

calibration (parameter estimation) methods in order to instantiate them and for validation approaches in order to justify their operational capacity. Both of these, parameter estimation and model validation, are complex issues; in this document we focus on numerical/computational aspects. The survey reported here is motivated by our conviction that the high complexity of current LUTI models requires the best possible numerical tools for calibration and validation. Developing and applying these tools is the primary objective set by the CITiES project funded by the French ANR (*Agence Nationale de la Recherche*). A starting point of this endeavor is a survey of existing state of the art reported in the LUTI literature. This document describes the scope of this ongoing study and some intermediate conclusions.

2 Methodology

Terminology. Model calibration is sometimes considered as the whole process of setting up the model structure: choice of theoretical model(s), of the spatial and economic disaggregation levels, choice of the functional form of the equations and of explanatory variables, and finally estimation of the model parameters. Although we are aware that in practice, this joint and usually iterative procedure is probably always necessary, in the scope of this study the term **calibration** specifically refers to the estimation of model parameters. Inputs to calibration primarily consist of **data** (observations made in the study area) but may also comprise **prior knowledge** on the parameters to estimate (such as typical values or expected ranges) or on expected model outputs.

Model **validation** is a vast question with less consensus on definition. Some consider a model to be validated if, once it is calibrated, the model outputs correctly match observed values. This is not usually a good practice, since the observed values are almost always used in the calibration. A model should be validated against its intended usage, which in the case of LUTI models typically involves simulations for a future horizon. The predictive capacity of the model may be assessed by confronting simulation results against actual observations that were not used for calibration; this is a standard approach in disciplines where experimentation is possible. Generally speaking, available data are split between a “training” set used for model calibration and a “test” set used purely for validation. Historical validation is the most frequent form of this approach: the model is calibrated for a given year (or time interval), and then tested against another year (or time interval). However, doing so may prove difficult for a LUTI model, considering that the sole data collection for the calibration step is already a very lengthy process. Furthermore, it is widely acknowledged that an appropriate usage of LUTI models may not be in interpreting model outcomes as accurate predictions, but rather as general trends. A more “behavioral” validation may thus be in order, including assessment by experts or sensitivity analysis. Finally, let us note that strictly speaking, one may never validate a model but only invalidate it, for instance by performing an uncertainty analysis (which aims to determine how the uncertainty on data used for calibration propagates to parameter values); if a model is found to be invalid, the reason may be that the model itself is inappropriate or that it is too disaggregated, or rather that the data are insufficient, etc.

Calibration. As explained above, calibration refers here to the estimation of model parameters and more specifically, to their estimation using numerical methods. This requires defining several elements: an overall estimation strategy (*e.g.* the sequence in which parameters are estimated), a mathematical formulation of the estimation problem, and actual estimation methods. Our survey analyzes the LUTI literature based on taxonomies for these components.

Concerning the overall estimation strategy, we adopt the taxonomy proposed by Abraham [1], which distinguishes: *piecewise* (estimation carried out module by module), *limited view* (black box calibration of the entire model), *simultaneous* (estimation of all model parameters, but unlike black box, consideration of outputs of individual modules), *sequential* (piecewise followed by simultaneous), *Bayesian sequential* (like sequential, using confidence intervals on parameters in addition).

As for the mathematical formulation of the estimation problem, we study if and how existing LUTI calibration approaches are based on an explicit formulation in terms of a numerical optimization problem, *i.e.* on the basis of one or multiple cost functions. We also examine if constraints or prior knowledge on parameters or model outputs are used in the formulation (*e.g. via* constrained optimization problems).

Finally, the actual estimation methods used are surveyed, ranging from *ad hoc* search methods over gradient descent, to more complex optimization methods.

Validation. Like for calibration, we scrutinize the LUTI literature in order to determine what kind of validation is carried out and how. As seen previously, the main forms of validation are: validation using only data used for calibration, validation against additional data (including "historical" validation), validation by experts, uncertainty analysis and sensitivity analysis.

3 Intermediate conclusions

While our systematic review of the LUTI literature is still underway, we can draw some intermediate conclusions based on our readings so far. The first obvious fact is that there is not one standard approach to calibration and validation. The model type (spatial-economic or activity based), available data, and the features of the software used, among other things, strongly determine the methods that are used. Other than that, there is no consensus on which objective function to use for calibration for instance, or what indicators to use when doing historical validation.

This being said, some general facts already emerge. First, piecewise calibration is still the dominant strategy. It is probably the simplest one to implement, but also rather suboptimal: modules are calibrated individually, without taking into account the interplay between them that makes up the entire model. A few exceptions exist though, such as the sequential-type calibration approach for MEPLAN proposed in [1]. Calibration methods vary depending on the model type, but the use of prior knowledge and of constraints on the parameter values is extremely rare. Regarding validation, historical validation is more and more frequent, and largely prevails compared to other potential forms of validation against additional data. Sensitivity analysis is not frequent, and when carried out, it is often under the form of test scenarios [4]. Last, uncertainty analysis starts to rise, but works on this topic remain seldom and this field is far from being mature.

Acknowledgment: This work is supported by the CITiES project funded by the *Agence Nationale de la Recherche* (grant ANR-12-MONU-0020).

References

[1] J.E. Abraham. *Parameter Estimation in Urban Models: Theory and Application to a Land Use Transport Interaction Model of the Sacramento, California Region*. Phd thesis, University of Calgary, Canada, 2000.

[2] David Simmonds Consultancy and Marcial Echenique and Partners Limited. Review of land-use/transport interaction models. Technical report, DETR, London, 1999.

[3] Douglass B. Lee. Requiem for large scale models. *Journal of the American Institute of Planners*, 39(3):163–178, 1973.

[4] T.W. Nicolai, L. Wang, K. Nagel, and P. Waddell. Coupling an urban simulation model with a travel model – a first sensitivity test. In *Computers in Urban Planning and Urban Management*, 2011.