Journal de la Société Française de Statistique
Vol. 158 No. 1 (2017)

Sensitivity Analysis and Optimisation of a Land Use and Transport Integrated Model

Titre: Analyse de sensibilité et optimisation d'un modèle transport-urbanisme

Laurent Gilquin[1], Thomas Capelle[1], Elise Arnaud[2] and Clémentine Prieur[2]

Abstract: Land Use and Transportation Integrated (LUTI) models have become a norm for representing the interactions between land use and the transportation of goods and people in a territory. Through the use of these models, urban planning policies and development scenarios can be evaluated. The calibration of LUTI models is a heavy task, involving gathering of massive amounts of data and the estimation of an important number of parameters. In this paper, the calibration of the open-source LUTI model Tranus is considered. Classical calibrations of Tranus rely on ad hoc econometric techniques and time-consuming trial and error procedures.Here, a two-step calibration that comprises global sensitivity analysis and optimisation is proposed. The sensitivity analysis presented herein is based on the replication method for the estimation of Sobol' indices and generalised to take into account multivariate outputs. The optimisation step is an iterative process combining stochastic and deterministic procedures. The proposed calibration procedure is applied to a study area in the State of Mississippi. Compared to a previous ad hoc procedure, this new approach results in a significant improvement of the adjustment factors of Tranus while reducing drastically the calibration time.

Résumé : Les modèles « transport-urbanisme » sont devenus une norme pour représenter les interactions entre l'usage des sols et le transport de marchandises et d'individus. Ces modèles sont principalement utilisés dans le cadre d'évaluations de politiques d'urbanisme et de scénarios de développement urbain. Le calage des modèles « transport-urbanisme » est une tâche difficile qui nécessite l'estimation d'un nombre important de paramètres. Dans ce papier, nous considérons le calage du modèle en libre accès Tranus. Une estimation classique des paramètres de Tranus repose à la fois sur des techniques ad hoc d'économétrie et sur des procédures de type essais-erreurs coûteuses en temps. Dans ce papier, nous proposons un calage en deux étapes comprenant une phase d'analyse de sensibilité globale et une phase d'optimisation itérative. La méthode d'analyse de sensibilité présentée ici est basée sur la méthode répliquée, estimant des indices de Sobol', et généralisée au cas de sorties multidimensionnelles. La phase d'optimisation est une procédure itérative combinant deux approches : une stochastique et une analytique. La méthode de calage est appliquée à la zone d'étude dans l'Etat du Mississippi. Par comparaison avec une précédente méthode de calage ad hoc, notre approche aboutit à une amélioration significative des facteurs d'ajustement de Tranus avec un temps de calage considérablement réduit.

Keywords: sensitivity analysis, optimisation, EGO, LUTI model

Mots-clés : analyse de sensibilité, optimisation, modèle « transport-urbanisme »

AMS 2000 subject classifications: 91B72, 62P20

[1] Inria Grenoble - Rhône-Alpes, Inovallée, 655 avenue de l'Europe, 38330 Montbonnot
Jean Kunzmann Laboratory, F-38000 Grenoble, France.
E-mail: laurent.gilquin@inria.fr

[2] Univ. Grenoble Alpes, Jean Kunzmann Laboratory, F-38000 Grenoble, France
CNRS, LJK, F-38000 Grenoble, France, Inria

1. Introduction

Land use and transport integrated (LUTI) models have received a regain of interest from researchers and urban planners during the last decade. Among the large number of available LUTI models, this paper focus on the model Tranus developed by de la Barra (1999). Tranus and other LUTI models aim to represent the deep interactions between travel behaviours and land use. Their scope of use ranges from urban metropolitan areas to regional level. LUTI modelling is mainly used to evaluate alternative planning scenarios, simulating their impact on land cover and travel demand. Instantiating a LUTI model requires the gathering of huge amounts of data and the estimation of several parameters to reproduce, as closely as possible, base-year observations (such as socio-economic surveys, transport data, etc ...) on the studied area. These models include systems of complex nonlinear equations. Analysing these systems is typically a hard task, particularly in the presence of parameters whose effects may be difficult to assess. Interactions between parameters of the model make that a small change in a parameter may result in large changes in the model outputs. In such cases, calibration plays a central role, as it helps estimating optimal values of these parameters, creating a robust model. The classical calibration approach of these models relies on econometric ad hoc procedures and trial and errors techniques. An exception can be found in Abraham and Hunt (2000), where an automatic calibration of the LUTI model MEPLAN is proposed.

Assessing sensitivity of the input parameters on the outputs during the calibration process is essential to reach a proper calibration of the model and ensure better predicting capabilities. As a matter of fact, in the context of traffic simulation Daamen et al. (2014) put forward the need to carry out a sensitivity analysis as a first step preceding the calibration of the model. Global sensitivity analysis methods are useful tools to quantify the influence of the model inputs on the outputs and detect potential interactions between them. Among the large number of available approaches, the variance based method introduced by Sobol' (1993) allows to calculate sensitivity indices called Sobol' indices. These indices are scalars between 0 and 1 that summarise the influence of each input or set of inputs: the higher is the index, the more influential is the input. First-order indices estimate the main effect from each input whereas higher-order indices estimate the corresponding importance of interactions between inputs. Various estimation procedures of these indices have been proposed in the literature (*cf.* Saltelli, 2002 for a survey). A first implementation of a sensitivity analysis on Tranus was performed in Dutta et al. (2012) using the "pick-freeze" estimation procedure introduced in Sobol' (1993). Unfortunately, this procedure requires a significant number of model evaluations that increases dramatically with the dimension of the input space. A solution to break this dependency lies in the use of replicated designs. Based on such designs, an estimation procedure for Sobol' indices was proposed in Mara and Joseph (2008) and further studied in Tissot and Prieur (2015). In this paper, an extension of the replication procedure to deal with multidimensional output is proposed. This generalisation is then applied to the land use and activity module of Tranus to select the most influential parameters based on main effects and second-order interactions.

Following the outcome of the sensitivity analysis, a two-stage iterative estimation of the influential parameters of the land use and activity module is proposed. First, a stochastic algorithm is applied to find optimal values of the influential parameters selected by the sensitivity analysis. Then, an analytical optimisation of an internal dispersion parameter is performed taking as inputs the previously optimised parameters. This second step is based on a careful investigation of the system of equations detailed in Capelle et al. (2015). This two-stage optimisation is iterated until an equilibrium is reached on the internal dispersion parameter.

The remainder of the paper is organised as follows. In Section 2, a detailed description of Tranus is provided, focusing on the land use and activity module and its principal variables relevant to this paper. Section 3 details the two main ingredients of the proposed calibration procedure from a methodological point of view: sensitivity analysis and the stochastic algorithm used in the two-stage iterative optimisation. In Section 4, the whole calibration procedure is detailed in the form of an algorithm. The proposed methodology is then applied to the study area of Mississippi. Finally, results are compared to those obtained with a classical ad hoc procedure.

2. Description of Tranus and problem statement

2.1. General structure of the model

This paper focus on the LUTI model Tranus. This type of software provides a framework for modelling land use and transportations in an integrated manner. It offers a flexible package to be used from urban and regional up to national scale. Tranus is based on the classical input-output model (*cf.* Leontief and Strout, 1963) and generalises it, adding the transportation layer on top of it.

The area of study is divided in spatial zones and economical sectors. The concept of sectors is more general than in the traditional definition. It may include the classical sectors in which the economy is divided (agriculture, manufacturing, mining, etc.), factors of production (capital, land and labour), population groups, employment, floorspace, land, energy, or any other that is relevant to the spatial system being represented.

Tranus combines a land use and activity module with a transportation module. Both of these modules are linked together and serve as input to each other, as illustrated in Figure 1. The spatial economic system is simulated by the activity model, representing the interactions of the various economical sectors in a specific time period. These interactions result in transportation demand that is afterwards affected to the network by the transportation module. In this way the movements of people or freight are explained as the results of the economic and spatial interaction between activities, the transport system and the real estate market. In return, the transport demand and the flux of goods influences the activities in the territory, affecting the access to transportation, the price of goods and ultimately the land rents. Both of these modules are based on classical discrete choice theory (*cf.* McFadden, 1973; McFadden and Train, 2000), input-output analysis (*cf.* Leontief, 1941), land choice (*cf.* Wilson, 1981), multi-modal path choice and trip assignment. A comprehensive review of transport modelling can be found in Ortúzar and Willumsen (2011).

The convergence is attained when both modules are in equilibrium. The land use and activity module iteratively equilibrates offer and demand, also computing the consumption costs and prices. This is done at current transportation costs. In the other hand, the transportation module assigns the transport demand to the network and computes the new transportation costs. This back and forth procedure iterates until a general equilibrium conditions is found. This condition is basically that neither land use, nor transport evolve anymore. Figure 1 illustrates this mechanism. As it is already well calibrated, no further details on the transportation module of Tranus are

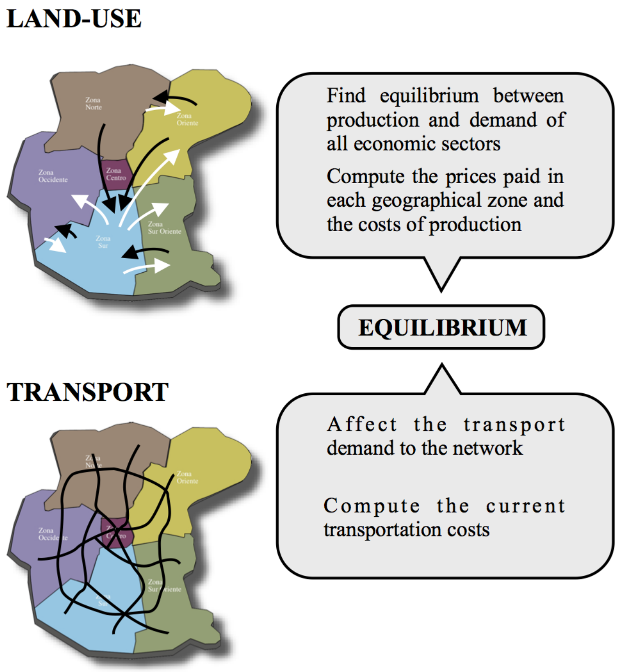


FIGURE 1. *Schematic overview of Tranus.*

provided. The procedure proposed in this paper focuses exclusively on the land use and activity module.

2.2. *The land use and activity module*

The land use and activity module's objective is to find an equilibrium between the production and demand of all economic sectors and zones of the modeled region. To attain the equilibrium, various parameters and functions are used to represent the behaviour of the different economic agents. Among these parameters are demand elasticities, attractiveness of geographical zones, technical coefficients, etc.

In the context of the present study, it is important to make the difference between two types of economic sectors; transportable and non-transportable sectors. The main difference between these two types of sectors is that transportable sectors can be consumed in a different place from where they were produced. As an example, the demand for coal from a metal industry can be satisfied by a mining industry located in another region. On the other hand, a typical non-transportable sector is floorspace: land is consumed where it is "produced". Transportable sectors induces transport demand, which ultimately influences transportation costs. Non-transportable sectors, on the other hand, do not require transportation. In Capelle et al. (2015) a detached calibration of Tranus was proposed, first calibrating non-transportable sectors, and building up the rest of the calibration over them. This upgraded version of Tranus will be considered in the remainder of the paper.

In the following, the main terminologies used in Tranus are introduced. The set of equations

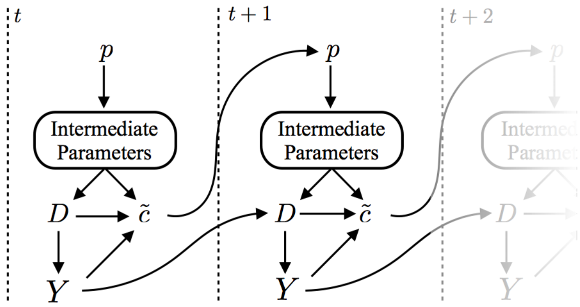


FIGURE 2. *Sketch of computations in the land use and activity module. t is the iteration index of the convergence.*

relevant to this paper can be found in the Appendix. For a complete description of Tranus, the reader may consult de la Barra (1999). The study area is divided in a set $\mathscr{Z}$ of spatial zones, and the economy is represented by a set $\mathscr{N}$ of economical sectors. Base-year data is denoted as $Y_0 = \{Y_{0i}^{n}\}_{n\in\mathscr{N},\, i\in\mathscr{Z}}$ and corresponds to the observed production for the base-year in each zone and for each economical sector. Sub-index 0 denotes the year of reference for the calibration.

The basis of the land use and activity model relies in four principal quantities:

— Productions: $Y = \{Y_i^n\}_{n\in\mathscr{N},\, i\in\mathscr{Z}}$ expresses how many "items" of an economic sector n are produced in a zone i.
— Demands: $D = \{D_i^{mn}\}_{(m,n)\in\mathscr{N}\times\mathscr{N},\, i\in\mathscr{Z}}$ expresses how many items of a sector n are demanded by the part of sector m located in zone i.
— Prices: $p = \{p_i^n\}_{n\in\mathscr{N},\, i\in\mathscr{Z}}$ defines the price of (one item of) sector n located in zone i.
— Costs: $\tilde{c} = \{\tilde{c}_i^n\}_{n\in\mathscr{N},\, i\in\mathscr{Z}}$ is the cost of consumption of sector n in zone i

Productions, demands, prices and costs are defined for each sector and each zone. To fix ideas, productions of a household sector represent the total households of a certain income level (rich for instance) in the corresponding zone. For housing type, production represents the available surface (square meters) in the zone of interest. Demands are tied to productions and represent how many units of the desired economic sector are consumed in a zone by another economic sector. The canonical example is demand for housing by households. Prices are very straightforward, representing the price of consuming one unit of the given economical sector in that zone. When considering e.g. households, prices represent salaries/incomes. Finally, costs are results of the model after the consumption chain has taken place, the equilibrium state is attained when costs are in perfect equilibrium with prices, and demand are in equilibrium with productions.

Productions, demands and prices form part of a dynamic system of equations. These equations depend on one another, and are linked by a list of equations that need to be computed one after another. This is detailed in de la Barra (1999). A graphical representation of this feedback is represented in Figure 2. For instance, demand induces production and vice-versa. The iteration scheme is as follows: prices of a current iteration (t) translate into intermediate variables (that will not be detailed here) which enables the computation of demand and consumption costs (noted as c in Figure 2). This is done based on the current transportation costs and disutilities. Once demand and costs are known, the current production is evaluated and fed back to compute a new set of prices, for a next iteration ($t+1$). The process is bottom-up, starting with land use prices and exogenous production and demand up to the production (destined for exportation outside the study area) and prices of transportable sectors. All the above computations are repeated until convergence is attained in productions X and prices p at the same time (convergence in these two sets of variables implies convergence in all others).

2.3. Problem statement

The calibration procedure proposed herein consists in adjusting the model parameters to be able to reproduce a base-year data in the study area. It is important to note that this calibration procedure is applied independently for each transportable sector n. This type of sectors are considered because it is where a sensitivity analysis is relevant, as the many interactions with the transportation costs and different parameters make calibration very complex. Also, for this particular sectors, modellers would like to identify the relevant parameters to improve their calibration. In Capelle et al. (2015), a decomposed calibration of Tranus is proposed, calibrating first the non-transportable sectors and later the transportable sectors. Starting from the results of Capelle et al. (2015), it appears that the calibration of the transportable sectors and the numerous parameters that are often ignored as their interactions are not fully understood, could take benefit from a sensitivity analysis. The calibration is made as following: once a sector is selected, the model parameters relative to this sector are adjusted through the proposed calibration procedure. The process is repeated for each transportable sector. The justification of this approach relies on a sector-wise decomposition of the system of equations of the land use and activity module.

Given initial values of the parameters for a transportable sector $n \in \mathscr{N}$, the land use and activity module estimates the adjustment parameters $h^n = (h_i^n)_{i \in \mathscr{Z}}$ of the utilities (*cf.* Appendix A, Equation (13)), known as shadow prices. In economics, the utility is defined as a measure of preference over sets of goods and services. For example, in the land use module, utility is used to measure the willingness of a household to relocate from a geographic zone to another. Still in economics, the shadow price is formally defined as the value of the Lagrange multiplier at the optimal solution of a constrained optimisation problem. However, in the present problem, the definition of the shadow price differs as it only represents an adjustment factor added to the price to ease the convergence of the land use module. Thus here, the shadow prices are price correcting additive factors that compensate the utilities to replicate the base-year production Y_0.

The following optimisation problem is solved to compute the shadow prices:

$$\hat{h}^n = \underset{h^n}{\operatorname{argmin}} \|Y(h^n) - Y_0\|^2, \tag{1}$$

where $\| \, . \, \|$ denotes the euclidean norm. In the remainder of the paper, the optimisation of the shadow prices is viewed as an internal process of the land use and activity model. Figure 3 gives a scheme of the inputs and outputs considered for each transportable sector n. The inputs parameters fall into three different categories:

— the parameter λ^n is involved in Equation (13) of Appendix A,
— the parameter β^n is involved in Equation (14) of Appendix A,
— parameters b_l^n, for $l \neq n, l \in \mathscr{N}$, are involved in Equation (15) of Appendix A.

The parameter λ^n represents the marginal utility of income; in common terms, it is the weight of prices in the utility function. β^n is the logit dispersion parameter of the location probabilities.

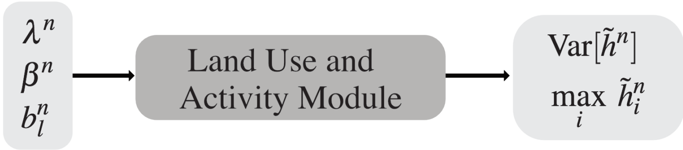


FIGURE 3. *Inputs and outputs of the land use and activity module for the sector n.*

Finally, the parameter b_l^n is the weight of sector l in the attractor for sector n (which are very important for transportable sectors).

The outputs considered are built upon a new quantity called normalised shadow prices. The normalised shadow price $\tilde{h}_i^n$ corresponds to the percentage of the price p_i^n corrected by the shadow price h_i^n, that is:

$$\tilde{h}_i^n = 100 \times \left| \frac{h_{n,i}}{p_{n,i}} \right|, \; i \in \mathscr{Z} \tag{2}$$

where $|\,.\,|$ denotes the absolute value function. Set $\tilde{h}^n = (\tilde{h}_i^n)_{i \in \mathscr{Z}}$ to be the vector of normalised shadow prices relative to the sector n. The two outputs considered are the following:

i) the variance of the normalised shadow prices: $\mathrm{Var}[\tilde{h}^n]$

ii) the maximum of the normalised shadow prices: $\max\limits_{i \in \mathscr{Z}} \tilde{h}_i^n$

For each sector n, a good calibration would be one that results in small values of the normalised shadow prices particularly in term of variance. Indeed, minimising the variance of the normalised shadow prices is a general consensus reached by both modellers and users of Tranus.

3. Calibration procedure's main tools

In this section, the two main methodological ingredients used in the calibration procedure are presented: the replication procedure for the sensitivity analysis and the EGO algorithm for the stochastic optimisation. Each tool is presented in a general framework nonspecific to Tranus. The algorithm summarising the proposed calibration procedure is presented in Section 4.1.

3.1. Global sensitivity analysis: replication procedure for multivariate outputs

For the sensitivity analysis, a generalisation of the replication procedure is proposed. This generalisation is based on the work of Gamboa et al. (2014) reviewed hereafter. Consider the following model:

$$f : \left\{ \begin{array}{ccc} \mathbb{R}^d & \rightarrow & \mathbb{R}^m \\ x = (x_1, \ldots, x_d) & \mapsto & z = f(x) \end{array} \right.$$

where f is the model, z the output vector, x the input vector, d the dimension of the input space and m the dimension of the output space.

Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. The uncertainty on the inputs is modeled by a random vector $X = (X_1, \ldots, X_d)$ whose components are independent. Denote Z the vector of random variables modelling the output vector:

$$Z = (Z_1, \ldots, Z_m) = f(X_1, \ldots, X_d).$$

Let $P_X = P_{X_1} \otimes \ldots \otimes P_{X_d}$ denote the probability distribution of X. Assume that $f \in \mathbb{L}^2(P_X)$ and that the covariance matrix of Z, denoted by Σ, is positive definite. Let u be a subset of $\{1, \ldots, d\}$ and denote $\sim u$ its complementary. Set $X_u = (X_i)_{i \in u}$ and $X_{\sim u} = (X_i)_{i \in \{1,\ldots,d\} \setminus u}$. Recall the following Hoeffding (1948) decomposition of f:

$$f(X) = f_0 + f_u(X_u) + f_{\sim u}(X_{\sim u}) + f_{u,\sim u}(X_u, X_{\sim u}), \tag{3}$$

where $f_0 = \mathrm{E}[Z]$, $f_u = \mathrm{E}[Z|X_u] - f_0$, $f_{\sim u} = \mathrm{E}[Z|X_{\sim u}] - f_0$ and $f_{u,\sim u} = Y - f_u - f_{\sim u} - f_0$. By taking the covariance matrix of each side of (3), due to orthogonality:

$$\Sigma = C_u + C_{\sim u} + C_{u,\sim u} \tag{4}$$

Let Id_m be the $m \times m$ identity matrix. Equation (4) can be projected on a scalar as follows:

$$\mathrm{Tr}(\mathrm{Id}_m \Sigma) = \mathrm{Tr}(\mathrm{Id}_m C_u) + \mathrm{Tr}(\mathrm{Id}_m C_{\sim u}) + \mathrm{Tr}(\mathrm{Id}_m C_{u,\sim u}) \tag{5}$$

where Tr denotes the trace operator. Following (5) and under the condition $\mathrm{Tr}(\Sigma) \neq 0$, the generalised Sobol' index is defined as follows:

$$S^u(f) = \frac{\mathrm{Tr}(C_u)}{\mathrm{Tr}(\Sigma)}. \tag{6}$$

The generalised Sobol' index $S^u(f)$ is a scalar between 0 and 1 that summarises the influence of inputs in u on the output Z. An index close to 1 means that the set u is influential. At the opposite, an index equal to 0 means that the set u is not correlated to the output Z. First-order indices estimate the main effect of each input whereas higher-order indices estimate the corresponding importance of interactions between inputs.

Remark 1. *When $u = (v, w)$ is a 2-subset of $\{1, \ldots, d\}$, the importance of the interaction between v and w is quantified by the second-order generalised Sobol' index defined by: $S^{(v,w)}(f) - S^v(f) - S^w(f)$.*

Classical estimation of $S^u(f)$ The classical estimation procedure for $S^u(f)$ is a generalisation of the one used in the univariate case (*cf.* Sobol', 1993). The procedure consists of a Monte-Carlo pick-freeze method. In the pick-freeze method, the Sobol index is viewed as the regression coefficient between the output of the model and its pick-freezed replication. This replication is obtained by holding the value of the variable of interest X_u (frozen variable) and by resampling the other variables $X_{\sim u}$ (picked variables).

Set $Z = f(X_u, X_{\sim u})$ and $Z^u = f(X_u, X'_{\sim u})$ where $X'_{\sim u}$ is an independent copy of $X_{\sim u}$. Let $N > 0$ be an integer and $Z_1, \ldots, Z_N$ (resp. $Z_1^u, \ldots, Z_N^u$) be N independent copies of Z (resp. Z^u) where:

$$Z_i = (Z_{i,1}, \ldots, Z_{i,m}),\ Z_i^u = (Z_{i,1}^u, \ldots, Z_{i,m}^u),\ \forall\, i \in \{1, \ldots, N\}.$$

As in Janon et al. (2014) and Monod et al. (2006), the following estimator of $S^u(f)$ is considered:

$$\widehat{S^u}(f) = \frac{\sum_{l=1}^{m}\left(\frac{1}{N}\sum_{i=1}^{N} Z_{i,l}Z^u_{i,l} - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{Z_{i,l}+Z^u_{i,l}}{2}\right)^2\right)}{\sum_{l=1}^{m}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{Z_{i,l}^2+(Z^u_{i,l})^2}{2} - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{Z_{i,l}+Z^u_{i,l}}{2}\right)^2\right)}. \tag{7}$$

Using this approach, estimating all first-order Sobol' indices requires $N(d+1)$ evaluations of the model by means of $d+1$ designs of experiments, each of size N. In the univariate case, the replication procedure introduced in Mara and Joseph (2008) allows to estimate all first-order indices with only two replicated designs, each of size N, resulting in a total of $2\times N$ evaluations of the model. Replicated designs are also referred as *plans based on permuted columns* in McKay (1995), Morris et al. (2006) and Morris et al. (2008). In these papers, an arbitrary number of r replications of the initial design is used to define an estimator of first-order indices. This estimator is of different nature from the one introduced in Mara and Joseph (2008) and further studied (asymptotic properties for first-order indices) and generalised in Tissot and Prieur (2015) to the estimation of closed second-order indices. An extension of the latter procedure, called replication procedure, to the case of multivariate output is proposed in the next paragraph. With this new approach, the number of model evaluations required to compute all first-order or second-order generalised Sobol' indices can be drastically reduced.

Replication procedure for $S^u(f)$ The replication procedure relies on the construction of two replicated designs of experiments **X** and **X'** defined as follows:

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & \dots & X_{1,j} & \dots & X_{1,d} \\ \vdots & & \vdots & & \vdots \\ X_{i,1} & \dots & X_{i,j} & \dots & X_{i,d} \\ \vdots & & \vdots & & \vdots \\ X_{N,1} & \dots & X_{N,j} & \dots & X_{N,d} \end{pmatrix}, \quad \mathbf{X'} = \begin{pmatrix} X'_{1,1} & \dots & X'_{1,j} & \dots & X'_{1,d} \\ \vdots & & \vdots & & \vdots \\ X'_{i,1} & \dots & X'_{i,j} & \dots & X'_{i,d} \\ \vdots & & \vdots & & \vdots \\ X'_{N,1} & \dots & X'_{N,j} & \dots & X'_{N,d} \end{pmatrix},$$

where $\forall\ k \in \{1,\dots,d\}$, $X_{1,k},\dots,X_{N,k}$ are N independent copies of X_k. For the estimation of first-order indices, **X** and **X'** are two replicated Latin Hypercubes. For the estimation of closed second-order indices, **X** and **X'** are two replicated randomised orthogonal arrays (*cf.* Tissot and Prieur, 2015). Denote **Z** and **Z'** the two arrays of model outputs associated to these two designs. Let Z_i and Z'_i denote their corresponding rows, $\forall\ i \in \{1,\dots,N\}$:

$$Z_i = f(X_{i,1},\dots,X_{i,d}) = (Z_{i,1},\dots,Z_{i,m}),$$
$$Z'_i = f(X'_{i,1},\dots,X'_{i,d}) = (Z'_{i,1},\dots,Z'_{i,m}).$$

The key point of the replication procedure consists in a "smart" rearrangement of the rows of **Z'** to mimic the pick-freeze method. The array resulting from this rearrangement corresponds to $\mathbf{Z^u}$. In the pick-freeze method, for each u, the evaluation of $\mathbf{Z^u}$ requires a new design of experiments. At the opposite, in the replication method $\mathbf{Z^u}$ requires no additional evaluations of the model. Let π denote the permutation used to re-arrange **Z'**, $\forall\ i \in \{1,\dots,N\}$:

$$Z^u_i = f(X'_{\pi(i),1},\dots,X'_{\pi(i),d}) = (Z'_{\pi(i),1},\dots,Z'_{\pi(i),m}).$$

Let $u = \{u_1, \dots, u_s\} \subset \{1, \dots, d\}$. From a design point of view, π is chosen to insure that:

$$X'_{\pi(i),u_j} = X_{i,u_j}, \ \forall\, j \in \{1, \dots, s\}, \forall\, i \in \{1, \dots, N\}.$$

thus insuring that both $\mathbf{Z}$ and $\mathbf{Z^u}$ are evaluated on the same coordinates indexed by u. Then, $S^u(f)$ is estimated using equation (7) with both $\mathbf{Z}$ and $\mathbf{Z^u}$. This extends the replication procedure to the estimation of generalised Sobol' indices. For the sake of brevity of this paper, the replication procedure is not further detailed. For a comprehensive description, the reader may consult Tissot and Prieur (2015).

To decide whether a parameter is influential or not, a threshold for both first- and second-order indices is fixed. If one of the estimated indices is higher than the threshold, the corresponding parameter is selected. In the application presented in Section 4, the threshold has been arbitrarily fixed at 0.1. This choice is based on the assumption that the "effective dimension" (*cf.* Caflisch et al., 1997) of the model is small (≤ 3).

3.2. Stochastic optimisation: EGO algorithm

The stochastic optimisation procedure presented in this section corresponds to the Efficient Global Optimisation (EGO) algorithm introduced by Mockus et al. (1978). The main idea underlying the EGO algorithm is to fit a response surface, often called metamodel, to data collected by evaluating the complex numerical model at a few points. The metamodel is then used in place of the numerical model to optimise the parameters. The metamodel used in the EGO algorithm is a Gaussian process defined as follows:

$$g : \left\{ \begin{array}{ccc} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ x = (x_1, \dots, x_d) & \mapsto & z = g(x) = \mu(x) + \varepsilon(x) \end{array} \right. ,$$

x are the parameters selected by means of the sensitivity analysis, z a scalar output of the numerical model, d the dimension of the input space, μ the model trend and ε is a centered stationary Gaussian process $\varepsilon(x) \sim \mathrm{N}(0, K_\chi)$. χ denotes the structure of the covariance matrix K_χ of ε. Let x^i, x^j denote two points of $\mathbb{R}^d$, $\chi = \{r, \theta, \sigma\}$ with $(K_\chi)_{i,j} = \sigma^2 r_\theta(x^i - x^j)$ where:
— $r_\theta(.)$ is the correlation function chosen here to be the Matèrn $5/2$ function,
— σ^2 is the variance of g,
— θ are the hyperparameters of r.
The parameters μ, σ and θ are estimated by maximum likelihood. In the following, Z denotes the random variable modelling the output z.

Expected Improvement Once the metamodel is fitted, it is used by the algorithm to search for a minimum candidate. The EGO algorithm uses a searching criterion called "expected improvement" that balances local and global search. Let x be a candidate point, the expected improvement evaluated at x writes as follows:

$$EI_\chi(x) = \mathrm{E}[\max(z_{min} - Z, 0)],$$

where z_{min} is the current minimum of the metamodel. A numerical expression of $EI_\chi(x)$ can be derived. Let $\widehat{Z}$ denote the *BLUP* (Best Linear Unbiased Predictor) estimator (*cf.* Jones et al., 1998) of Z and $\sigma_{\widehat{Z}}$ its standard deviation, the following expression for $EI_\chi(x)$ is obtained:

$$EI_\chi(x) = (z_{min} - \widehat{z}(x))\phi_{\mathcal{N}}\left(\frac{z_{min} - \widehat{z}(x)}{\sigma_{\widehat{z}}}\right) + \sigma_{\widehat{z}} f_{\mathcal{N}}\left(\frac{z_{min} - \widehat{z}(x)}{\sigma_{\widehat{z}}}\right),$$

where $\phi_{\mathcal{N}}$ is the normal cumulative distribution function and $f_{\mathcal{N}}$ is the normal probability density function. The first term of $EI_\chi(x)$ is a local minimum search term whereas the second term corresponds to a global search of uncertainty regions. The main steps of the EGO algorithm can be summarised as follows:

1. generate a design of experiments and evaluate the numerical model at these points,
2. fit the metamodel with both the design of experiments and the associated model outputs,
3. search a new evaluation point using the expected improvement criterion,
4. evaluate the numerical model at this new point and re-estimate the parameters of the metamodel (θ, σ),
5. repeat steps 3 to 5 until a stopping criterion is reached.

For the choice of the stopping criterion, one can look at the value of the expected improvement. A value close to zero indicates that the input space has been sufficiently explored. Hence, a lower bound on the expected improvement can be selected as the stopping criterion. In the application presented hereafter, the lower bound is set equal to 10^{-5}. As a result, the stopping criterion writes:

$$EI_\chi(x) \leq 10^{-5} \;\; .$$

To ensure that the EGO algorithm actually stops, the maximum number of iterations is fixed at 200. The two R packages "DiceOptim" and "DiceDesign" developed by Roustant et al. (2012) are used to implement the EGO algorithm.

4. Application to Tranus

In this section, an innovative calibration procedure for the land use and activity module of Tranus is detailed, under the form of an algorithm. The procedure includes both methods presented in Section 3. The methodology is then applied to the Mississippi region including the Chickasaw, Lee, Pontotoc, and Union (and consequently, the four largest towns of the area: Houston, Tupelo, Pontotoc, and New Albany).

4.1. Calibration procedure algorithm

Algorithm 1 summarises the proposed calibration procedure for the land use and activity module of Tranus. It is worth reminding that the calibration procedure is applied independently for each transportable sector n.

Once a transportable sector n is selected, the sensitivity analysis presented in Section 3.1 is performed on the parameters β^n and $b_{n,l}$, $l \neq n$. The outputs considered for the sensitivity analysis are both the variance and the maximum of the normalised shadow prices $\tilde{h}^n$ (Equation 2). The set of influential parameters selected is denoted by ρ^n.

Algorithm 1 Calibration procedure for the land use and activity module.

1: **for** each sector transportable n **do**
2: Set: $\lambda^{n(0)} \leftarrow \lambda_0^n$
3: Run sensitivity analysis with inputs: β^n, $\{b_l^n\}_{n \neq l}$ and outputs: $Var[\tilde{h}^n], \max_{i \in \mathscr{Z}} \tilde{h}_i^n$
4: Instantiate:
— $\rho^{n(0)} \leftarrow$ set of most influential parameters
— $k \leftarrow 1$
5: **while** $|\lambda^{n(k)} - \lambda^{n(k-1)}| \geq \varepsilon$ **do**
6: Given $\lambda^{n(k-1)}$, estimate $\rho^{n(k)}$ with the EGO algorithm
7: Given $\rho^{n(k)}$, estimate $\lambda^{n(k)}$ with the analytical optimisation
8: Return optimal values ρ^{n*} and λ^{n*}

Following the sensitivity analysis, an iterative optimisation is conducted. This optimisation comprises two stages. At iteration k, the EGO algorithm presented in Section 3.2 is applied to find optimal values of the set $\rho^{n(k)}$ given $\lambda^{n(k-1)}$. Then, an analytical optimisation of $\lambda^{n(k)}$ is performed taking as inputs the optimal values found for the set $\rho^{n(k)}$. This step is further described in the next section. The process is iterated until an equilibrium is reached on λ^n. Then, the optimal values ρ^{n*} and λ^{n*} are returned.

Remarks

1. One might wonder why the sensitivity analysis does not include the parameter λ^n. The main reason behind this choice is that the analytical optimisation provides a global optimum for λ^n. Hence, by automatically selecting the parameter λ^n, one can reduce the number of Sobol' indices to estimate and obtain a finer optimisation of the land use and activity module.
2. For the stochastic optimisation, only one of the two outputs is conserved to perform the EGO algorithm: the variance of the normalised shadow prices $Var[\tilde{h}^n]$. As already mentioned, minimising the variance of the normalised shadow prices over the maximum is a general consensus reached by both the modeller and users of Tranus.

4.2. Analytical optimisation procedure (obtaining the λ^n parameters)

Once optimal values for the selected parameters β^n and b_k^n are found, the value of the λ^n parameter is computed analytically. The basic idea is to start from a good guess of the logit dispersion parameters (β^n) and from there, find the optimal values of the marginal utilities (λ^n) to minimise the variance of the shadow prices for the corresponding economic sector. This methodology explicits the dependency of λ^n on β^n, showing that the optimal value of λ^n is a function of β^n. For the sake of clarity, the notation $\lambda^n, \lambda^{n(k)}$ and $\beta^n, \beta^{n(k)}$ are confounded in the remainder of the section.

The parameter λ^n is involved in the location probabilities equation (Equation 13 of Appendix A):

$$U_{ij}^n = \lambda^n (p_j^n + h_j^n) + t_{ij}^n \ , \ (i,j) \in \mathscr{Z}^2. \tag{8}$$

The optimal value of λ^n cannot be retrieved directly from Equation 8 as the quantity $(p^n_j + h^n_j)$ is estimated as a whole during the internal optimisation of the shadow prices (*cf.* Capelle et al., 2015). To overcome this problem, an auxiliary variable is introduced:

$$\phi^n_j = \lambda^n(p^n_j + h^n_j),\ j \in \mathscr{Z}.$$

With this new variable, Equation 8 can be rewritten as follows:

$$U^n_{ij} = \phi^n_j + t^n_{ij}\ ,\ (i,j) \in \mathscr{Z}^2. \tag{9}$$

Recall that the shadow prices are prices correcting additive factors that are calibrated to obtain a small variance. From Equation (9), one can express the optimal value of λ^n that minimises the variance of the shadow prices. Set $\phi^n = (\phi^n_j)_{j\in\mathscr{Z}}$ with all other parameters fixed, in particular the parameters ρ^n estimated with the EGO algorithm (*cf.* Section 3.2). The corresponding calibration problem writes:

$$\phi^{n*} = \underset{\phi^n}{\operatorname{argmin}} \quad \|Y(\phi^n) - Y_0\|^2\ . \tag{10}$$

It is worth recalling that $p^n = (p^n_j)_{j\in\mathscr{Z}}$ and $h^n = (h^n_j)_{j\in\mathscr{Z}}$ are the vectors of prices and shadow prices. Once the optimal value ϕ^{n*} is obtained, the equilibrium prices p^{n*} can be computed solving a linear system. Then, the shadow prices are expressed as follows:

$$h^n = \frac{\phi^{n*}}{\lambda^n} - p^{n*}\ .$$

From this, the following problem can be posed:

$$\min_{\lambda^n} \quad \operatorname{Var}\left[\frac{\phi^{n*}}{\lambda^n} - p^{n*}\right],$$

the analytical solution of which is $\lambda^{n*} = \dfrac{\operatorname{Var}(\phi^{n*})}{\operatorname{Cov}(\phi^{n*}, p^{n*})}$. Note that this value can be negative. The details of Problem (10) and how the prices are computed can be found in Capelle et al. (2015).

4.3. *Results*

This section presents the application of the proposed calibration procedure to a study area in the State of Mississippi. First, a brief insight of the area of study is given. Then, the results of both the sensitivity analysis and the iterative optimisation used in the calibration procedure are presented. These results are compared to those obtained with a former ad hoc procedure.

Area of study: The area of study is divided into 12 economical sectors and 103 spatial zones. The dataset was provided by Brian Morton [1]. In Table 1 are listed the different economical sectors of Tranus relative to the area of study. The proposed calibration procedure focus only on transportable sectors.

[1] Brian Morton is a senior research associate working at the Center for Urban and Regional Studies of North Carolina, Hickerson House, 108 Battle Lane, Campus Box 3410. The data were provided to us during an informal workshop taking place at INRIA on the topic of LUTI model calibration.

TABLE 1. *Economical sectors of Tranus for the study area of Mississippi.*

sector n	name	description	type
1	AFFHM(Agriculture,Fishing,...)	Business	Exogenous
2	Commercial	Business	Transportable
3	Other industries	Business	non-Transportable
4	Single person (15-64 years old)	Households	Transportable
5	Married couple with kids	Households	Transportable
6	Married couple without kids	Households	Transportable
7	Other families with kids	Households	Transportable
8	65 years and older	Households	Transportable
9	All other households	Households	Transportable
10	1-unit housing units	Floorspace	Land
11	Multiunit housing units	Floorspace	Land
12	Mobile homes	Floorspace	Land

TABLE 2. *Distributions of the* 12 *parameters,* $U(a,b)$ *stands for the uniform distribution with support* $[a,b]$.

parameters	labels	distributions
β^n	1	$U(2,10)$
$\{b_l^n\}_{l=1,\ldots,n,\ l\neq n}$	$2,\ldots,12$	$U(0,1)$

Results of the sensitivity analysis: A total of 7 sensitivity analyses are performed, one for each transportable sector. For each sensitivity analysis, the 12 following parameters are considered:

— The logit dispersion parameter β^n.

— The 11 parameters b_l^n, for all sectors $l \neq n$.

In Table 2 are listed the distribution of each parameter. These distributions were selected by expertise, as each model is different, a good range of a priori possible values needs to be explored. The Sobol' index S^k will refer to the parameter labeled by k. For instance, S^1 will always correspond to the logit dispersion parameter β^n of the sector n under analysis, and $S^2,\ldots,S^{12}$ the corresponding cross relations b_l^n parameters (without including the case $l=n$).

The outputs considered are the variance and the maximum of the normalised shadow prices, as introduced in Section 2.3. The approach proposed is to use the replication procedure presented in Section 3.1 to estimate first-order and second-order generalised Sobol' indices of these parameters. Asymptotic confidence intervals can be computed for first-order Sobol' indices (*cf.* Tissot and Prieur, 2015). Using Remark 1 page 97, bootstrap confidence intervals can be derived for second-order Sobol' indices.

Before presenting the main results, it seems important to illustrate the selection procedure of the influential parameters for a sector. Figures 4 and 5 show the results obtained for the estimation of first-order and second-order indices relative to Sector 4. The dashed line represents the threshold value used for selecting the influential parameters.

For the estimation of first-order indices, a size $N = 5\times 10^3$ was chosen for the two replicated Latin Hypercubes required by the replication procedure (Section 3.1). Looking at the results, the parameters β^4 and b_{10}^4 are the most influential (*cf.* Figure 4). Since the sum of the first-order indices is less than 75% it is interesting to study the second-order indices.

For the estimation of second-order indices, a size $N = 67^2$ was selected for the two replicated randomised orthogonal arrays required by the replication procedure. The two black points of

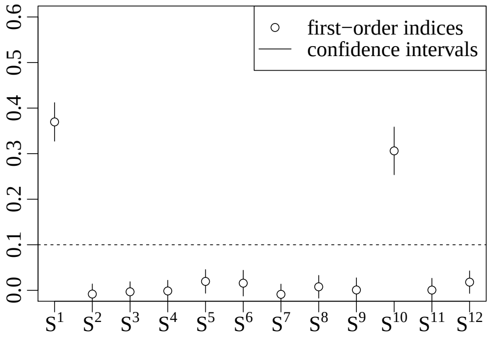


FIGURE 4. *First-order indices estimation for the 12 parameters of sector 4. S^1 corresponds to β^4, S^2 to b_1^4, S^3 to b_2^4 and so on.*

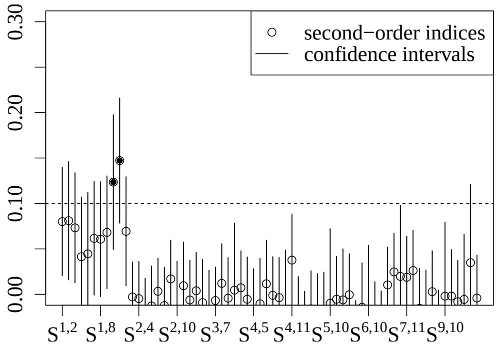


FIGURE 5. *Second-order indices estimation for the 12 parameters of sector 4. The second order indexes are given in the following order: $\{(1,2),\dots,(1,12),(2,3),\dots,(2,12),(3,4),\dots\}$.*

TABLE 3. *Most influential parameters selected by the sensitivity analysis based on main effects and second-order interactions.*

sector	first-order	second-order	selected parameters: ρ	variance explained (in percent)
2	β^2	none	β^2	33
4	β^4, b_{10}^4	$\beta^4 * b_{10}^4$, $\beta^4 * b_{11}^4$	β^4, b_{10}^4, b_{11}^4	95
5	β^5, b_{10}^5	$\beta^5 * b_{10}^5$, $\beta^5 * b_{11}^5$	β^5, b_{10}^5, b_{11}^5	89
6	β^6, b_{10}^6, b_{11}^6	$\beta^6 * b_{10}^6$	β^6, b_{10}^6, b_{11}^6	90
7	β^7, b_{10}^7	none	β^7, b_{10}^7	85
8	β^8, b_{10}^8	$\beta^8 * b_{10}^8$	β^8, b_{10}^8	89
9	β^9, b_{10}^9,	$\beta^9 * b_{10}^9$	β^9, b_{10}^9	93

Figure 5 correspond to the two most influential interactions: $\beta^4 * b_{10}^4$ and $\beta^4 * b_{11}^4$. The number of bootstrap replications used to compute the confidence intervals equals 1000.

In conclusion, only 3 of the 12 parameters of the sector 4 are significantly influential either directly by their main effects or through their second-order interactions: β^4, b_{10}^4 and b_{11}^4.

The same procedure is performed for the other transportable sectors. For each sector n, the set ρ^n that comprises the most influential parameters selected by the sensitivity analysis is listed in Table 3. The last column of the table gives the proportion of the model's variance explained by the selected parameters. This proportion is calculated by multiplying the sum of the generalised Sobol' indices of the first two columns by 100. Looking at the results, only 3 parameters appear to be overall the most influential: β^n, b_{10}^n and b_{11}^n, $n \in \{2,4,5,6,7,8,9\}$.

These results fall within our range of expectations. The parameter β^n is a dispersion parameter of a multinomial logit function (*cf.* Equation (14) of Annexe A). A slight variation of this parameter leads to a significant change in the calculation of the probabilities of localisation. Both parameters b_{10}^n and b_{11}^n act as weights in the attractiveness for sector n. These two parameters are more prone to be influential than others b_l^n as sectors 10 and 11 correspond to the two main floorspace types.

Looking at the last column of Table 3, for some sectors (in particular for sector 2) the variance is not fully explained by the parameters selected. A solution for further explaining this variance would be the use of Saltelli's procedure to estimate first-order and total-effect Sobol' indices (index measuring the contribution of an input through its main effect plus all its interactions). However, Saltelli's procedure would requires $N \times 14$ evaluations of the model for each sector. In comparison, the proposed approach require only $N \times 4$ evaluations of the model to estimate both first-order and closed second-order indices ($2 \times N$ for first-order indices and $2 \times q^2$, with $q \approx \sqrt{N}$ for closed second-order indices). Since the number of evaluations is a concern in this study, it has been decided to go for the replication procedure at the price of not using the total effect Sobol' indices.

A compromise would be to fit a metamodel and evaluate first-order and total effect Sobol' indices with the metamodel outputs. Such a methodology has been applied by Ge et al. (2014) to the traffic simulation model Aimsun. In their application, the use of a metamodel is particularly efficient as microscopic traffic simulators are known to be quite time-consuming. However, one may argue that the complexity of Tranus makes it hard to perform a metamodel based estimation of Sobol' indices.

An alternative to the calculation of Sobol' indices was proposed by Ge and Menendez (2014), *i.e.* to perform a sensitivity analysis prior to the calibration of the traffic model VISSIM. This approach called quasi-OTEE is a screening method based on the original Elementary Effects method introduced by Morris (1991). The computational cost of the quasi-OTEE method is $(m-n+1)*(m+n)/2$ where m is the number of Morris trajectories and n is the number of quasi-optimal trajectories selected among the m latter. The authors advise to choose m and n in the ranges $[500, 1000]$ and $[10, 20]$ respectively. With the minimal setting $m = 500$ and $n = 10$, the cost of the quasi-OTEE method equals 125 205 which is 6 times higher than the cost of the procedure proposed here to estimate first- and second-order indices ($2 \times 5000 + 2 \times 67^2$). Furthermore, as a screening method, the quasi-OTEE approach should be less precise than the calculation of Sobol' indices.

Results of the subsequent iterative optimisation: Following the results of the above sensitivity analysis, for each transportable sector n, the aim is to find the set of parameters (ρ^n, λ^n) minimising the variance of $\tilde{h}^n$. The initial value λ_0^n (Step 3 of Algorithm 1) instantiating the parameter λ^n is obtained by expertise. The results obtained both in terms of variance and maximum of the normalised shadow prices are compared to those obtained with a former ad hoc procedure. The number of initial evaluations performed to fit the metamodel for the set of parameters (ρ^n, λ^n) of each sector n is the following:

— 21 evaluations for sector 2,
— 51 evaluations for sectors 4 to 6.
— 41 evaluations for sectors 7 to 9.

The quality of the fitting is assessed by diagnostic plots (fitted values against response values, standardised residuals, Q-Q plots of standardised residuals) based on leave-one-out cross validation results (*cf.* Roustant et al., 2012). For each sector n, the set of evaluations includes the one obtained with the ad hoc procedure at the optimal set of parameters.

Table 4 summarises the results obtained with both the ad hoc procedure and the innovative iterative optimisation. For each sector n, ρ^{n*} and λ^{n*} denote the optimal values of the parameters

TABLE 4. *Variance and maximum of the normalised shadow prices $\tilde{h}_n$ obtained with both ad hoc procedures and the iterative optimisation.*

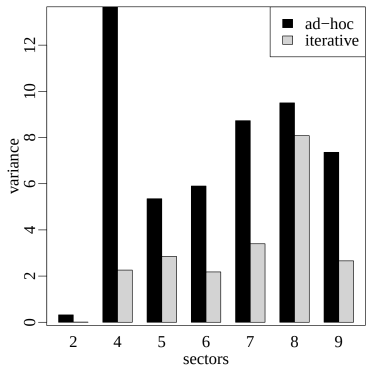
sector n	procedure	ρ^{n*}	λ^{n*}	variance $\tilde{h}_n$	max $\tilde{h}_n$	gain
2	ad hoc	2	0.005	0.32	2.95	98 %
	iterative	4.03	0.43	7×10^{-3}	0.11	
4	ad hoc	$(2,0,0)$	0.001	13.66	24.95	83 %
	iterative	$(6.49,0.38,0)$	0.001	2.26	7.63	
5	ad hoc	$(2,0,0)$	0.001	5.35	14.83	47 %
	iterative	$(2.50,0.02,0.79)$	-0.013	2.85	8.88	
6	ad hoc	$(2,0,0)$	0.001	5.90	16.65	63 %
	iterative	$(6.64,0.05,0.79)$	-0.003	2.18	7.72	
7	ad hoc	$(2,0)$	0.001	8.73	19.67	61 %
	iterative	$(9.17,1)$	0.001	3.40	8.23	
8	ad hoc	$(2,0)$	0.001	9.50	20.58	15 %
	iterative	$(5.72,0.97)$	0.001	8.08	15.03	
9	ad hoc	$(2,0)$	0.001	7.36	17.6	64 %
	iterative	$(9.29,0.95)$	0.001	2.66	6.82	

obtained at the end of both approaches. The column "gain" represents the improvement (in percent) of the variance obtained with the proposed iterative estimation relatively to the one obtained with the ad hoc procedure conducted by experts.

One can observe that the values of the variance and maximum of the normalised shadow prices obtained with the ad hoc procedure are heterogeneous. Furthermore, the value of the maximum is quite high for some sectors (up to 20% of the price). The results obtained with the proposed iterative optimisation are relatively homogeneous except for sectors 2 and 8. The discrepancy observed for these two sectors comes from the quality of their respective datasets. Indeed, the data relative to commercial business (sector 2) are easy to collect and thus of high quality and quantity. At the opposite, data relative to the 65 years and older households (sector 8) are quite complex to collect and often lack precision.

The main observation is that an improvement in terms of both variance and maximum of the normalised shadow prices is observed for all sectors when using the iterative approach. Figure 6 gives an illustration of this improvement. The black bars represent the values obtained with the ad hoc procedure, the grey bars those obtained with the proposed iterative approach. A significant diminution for both the variance and maximum criteria is observed. Furthermore and most importantly, the proposed iterative approach is drastically faster than the ad hoc procedure conducted by experts: the calibration procedure requires a few hours compared to several days (up to weeks) for the ad hoc procedure.

In the proposed calibration procedure, only the best set of parameters $(\rho^{n*}, \lambda^{n*})$ is conserved. Ciuffo and Azevedo (2014) have presented an alternative setting in which a metamodel is fitted and several best sets of parameters are selected. It is true that for complex systems such as LUTI models, the best solution of the EGO optimisation probably corresponds to only one of many combinations of the inputs that provide the model with a sufficient robustness. The method of Ciuffo and Azevedo (2014) allows to investigate the behavior of the model for various combinations and to derive uncertainty margins of the outputs. Adapting this methodology to the



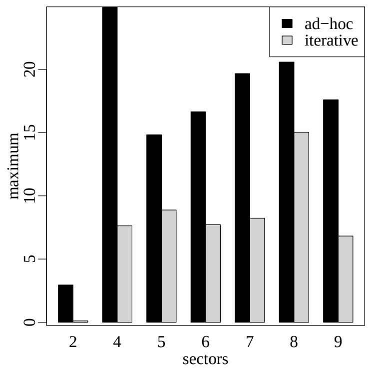


FIGURE 6. *Variance (left figure) and maximum (right figure) of the normalised shadow prices obtained with both the ad hoc procedure (referred to as ad hoc) and the proposed iterative optimisation (referred to as iterative).*

calibration procedure of Tranus proposed in this paper would be an interesting complementary work.

Conclusion

In this paper a calibration procedure combining a global sensitivity analysis and an iterative optimisation to calibrate the land use and activity module of the LUTI model Tranus has been proposed. The sensitivity analysis presented herein is a generalisation of the replication procedure to select the most influential parameters of the model when the output is multidimensional. The optimisation phase was carried on using a two-stage process combining stochastic (algorithm EGO) and deterministic procedures. The deterministic procedure exploited the system of equations of Tranus to derive an analytical solution for the marginal utilities and consequently obtain a finer optimisation of the land use and activity module.

An application to the study area of Mississippi was presented. The proposed methodology was compared to a former ad hoc calibration procedure in terms of variance and maximum of the normalised adjustment parameters (shadow prices). The results showed a significant improvement on both criteria reducing the value of the variance by a large margin with a drastic gain of time: that suggests that the proposed methodology is useful to improve the calibration of such models. The next step would consist in verifying if the optimal values found for the parameters ensure better predicting capabilities when evaluating alternative planning scenarios.

Acknowledgments

The authors thank the guest editor and the two reviewers for their helpful suggestions and comments which substantially improved the quality of this paper. The authors are also thankful to Brian J. Morton for providing high quality data and instantiating the Tranus model on the study area of Mississippi. This work is supported by the CITiES project funded by the Agence Nationale de la Recherche (grant ANR-12-MONU-0020).

References

Abraham, J. E. and Hunt, J. D. (2000). Parameter estimation strategies for large-scale urban models. *Transportation Research Record: Journal of the Transportation Research Board*, 1722(1):9–16.

Caflisch, R. E., Morokoff, W. J., and Owen, A. B. (1997). Valuation of mortgage backed securities using brownian bridges to reduce effective dimension. *J. Comput. Finance*, 1(1):27–46.

Capelle, T., Sturm, P., Vidard, A., and Morton, B. J. (2015). Formulating LUTI calibration as an optimisation problem: Example of tranus shadow price estimation. *Procedia Engineering*, 115:12–20.

Ciuffo, B. and Azevedo, C. L. (2014). A sensitivity-analysis-based approach for the calibration of traffic simulation models. *IEEE Trans. Intell. Transp. Syst.*, 15(3):1298–1309.

Daamen, W., Buisson, C., and Hoogendoorn, S. P. (2014). *Traffic Simulation and Data: Validation Methods and Applications*. Taylor & Francis Group.

de la Barra, T. (1999). Mathematical description of TRANUS. Technical report, Modelistica, Caracas, Venezuela.

Dutta, P., Saujot, M., Arnaud, E., Lefevre, B., and Prados, E. (2012). Uncertainty propagation and sensitivity analysis during calibration of tranus, an integrated land use and transport model. *International Conference on Urban, Regional Planning and Transportation*.

Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014). Sensitivity analysis for multidimensional and functional outputs. *Electron. J. Statist.*, 8(1):575–603.

Ge, Q., Ciuffo, B., and Menendez, M. (2014). An exploratory study of two efficient approaches for the sensitivity analysis of computationally expensive traffic simulation models. *IEEE Trans. Intell. Transp. Syst.*, 15(3):1288–1297.

Ge, Q. and Menendez, M. (2014). An efficient sensitivity analysis approach for computationally expensive microscopic traffic simulation models. *International Journal of Transportation*, 2(2):49–64.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19(3):293–325.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2014). Asymptotic normality and efficiency of two Sobol' index estimators. *ESAIM Probab. Stat.*, 18:342–364.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *J. Global Optim.*, 13:455–492.

Leontief, W. W. (1941). *The Structure of the American Economy 1919-1939*. New York: Oxford University Press, 2nd edition, 1951 edition.

Leontief, W. W. and Strout, A. (1963). *Multi-Regional Input-Output Analysis*. Structural Interdependence and Economic Development. London: Mcmillan.

Mara, T. A. and Joseph, O. R. (2008). Comparison of some efficient methods to evaluate the main effect of computer model factors. *Journal of Statistical Computation and Simulation*, 78(2):167–178.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. *Frontiers in Econometrics - Academic Press New York*, pages 105–142.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470.

McKay, M. D. (1995). Evaluating prediction uncertainty. Technical report, Los Alamos National Laboratory Report NUREG/CR- 6311, LA-12915-MS.

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). The application of bayesian methods for seeking the extremum. In Dixon, L. C. W. and Szego, G. P., editors, *Towards Global Optimisation, Vol. 2*, pages 117–129. North-Holand.

Monod, H., Naud, C., and Makowski, D. (2006). *Uncertainty and sensitivity analysis for crop models*, chapter 3, pages 55–100. Elsevier.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174.

Morris, M. D., Moore, L., and McKay, M. D. (2006). Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs. *J. Statist. Plann. Inference*, 136(9):3203–3220.

Morris, M. D., Moore, L., and McKay, M. D. (2008). Orthogonal arrays in the sensitivity analysis of computer models. *Technometrics*, 50:205–215.

Ortúzar, J. D. and Willumsen, L. G. (2011). *Modelling Transport*. Wiley, Fourth Edition edition.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *J. Stat. Softw*, 51(3).

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.*, 145(2):280–297.

Sobol', I. M. (1993). Sensitivity indices for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414.

Tissot, J. Y. and Prieur, C. (2015). A randomized orthogonal array-based procedure for the estimation of first- and second-order sobol' indices. *J. Statist. Comput. Simulation*, 85:1358–1381.

Wilson, A. G. (1981). *Optimization in locational and transport analysis*. Wiley & Sons, New York.

Appendix A: Tranus system of equations

Only a subset of model equations relevant to this paper is presented. The demand is computed for all combinations of zone i, demanding (consuming) sector m and demanded sector n:

$$D_i^{mn} = (Y_i^{*m} + Y_i^m)\, a_i^{mn} S_i^{mn} \tag{11}$$

$$D_i^n = D_i^{*n} + \sum_m D_i^{mn} \tag{12}$$

where Y_i^{*m} is the given exogenous production (for exports), Y_i^m the induced endogenous production obtained in the previous iteration (or initial values), and D_i^{*n} exogenous demand. D_i^n in (12) then gives the total demand for sector n in zone i. a_i^{mn} is a technical demand coefficient and S_i^{mn} is the substitution proportion of sector n when consumed by sector m on zone i.

In parallel to demand, one computes the **utility** of all pairs of production and consumption zones, i and j:

$$U_{ij}^n = \lambda^n (p_j^n + h_j^n) + t_{ij}^n \ . \tag{13}$$

Here, λ^n is the marginal utility of income for sector n and t_{ij}^n represents transport disutility. Since utilities and disutilities are difficult to model mathematically (they include subjective factors such as the value of time spent in transportation), Tranus incorporates adjustment parameters h_j^n, so-called shadow prices, amongst the model parameters to be estimated.

From utility, one computes the probability that the production of sector n demanded in zone i, is located in zone j. Every combination of n, i and j is computed:

$$Pr_{ij}^n = \frac{A_j^n exp\left(-\beta^n U_{ij}^n\right)}{\sum_k A_k^n exp\left(-\beta^n U_{ik}^n\right)} \ . \tag{14}$$

Here, k ranges over all zones. β^n is the dispersion parameter for the multinomial logit model expressed by the above equation. A_j^n represents attractiveness of zone j for sector n and is expressed as follows:

$$A_j^n = \sum_l b^{n,l} Z_j^l \tag{15}$$

where the $b_{n,l}$ are the relative weights of sector l in the attractive function of sector n and the Z_j^l are variables depending on prices and productions.

From these probabilities, new productions are then computed for every combination of sector n, production zone j and consumption zone i:

$$Y_{ij}^n = D_i^n Pr_{ij}^n \ . \tag{16}$$

Total production of sector n in zone j, is then:

$$Y_j^n = \sum_i Y_{ij}^n \tag{17}$$

$$= \sum_i D_i^n Pr_{ij}^n \ . \tag{18}$$

The set of prices are also governed by a set on non-linear equations, and are computed simultaneously to attain equilibrium.