



**HAL**  
open science

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

Laurent Gilquin, Thomas Capelle, Élise Arnaud, Clémentine Prieur

► **To cite this version:**

Laurent Gilquin, Thomas Capelle, Élise Arnaud, Clémentine Prieur. Sensitivity Analysis and Optimisation of a Land Use and Transport Integrated Model. *Journal de la Société Française de Statistique*, 2017, Computer Experiments, Uncertainty and Sensitivity Analysis, 158 (1), pp.90-110. hal-01291774

**HAL Id: hal-01291774**

**<https://inria.hal.science/hal-01291774>**

Submitted on 22 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Laurent Gilquin<sup>1</sup>, Thomas Capelle<sup>1</sup>, Elise Arnaud<sup>2</sup> and Clémentine Prieur<sup>2</sup>

**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, we consider the calibration of the open-source LUTI model Tranus. Classical calibrations of Tranus rely on ad-hoc econometric techniques and time-consuming trial and error procedures. Here, we propose a two steps calibration procedure comprise of global sensitivity analysis and optimisation. The sensitivity analysis presented 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. Our calibration method is applied to a North-Carolina urban area. Compared to a previous ad-hoc procedure, our 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. Notre méthode de calage est appliquée à la zone d'étude de Caroline du Nord. 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. Introduction

Land use and transport integrated (LUTI) models have received a regain of interest from researchers and urban planners the last decade. Among the large number of available LUTI models,

<sup>1</sup> 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](mailto:laurent.gilquin@inria.fr)

<sup>2</sup> Univ. Grenoble Alpes, Jean Kunzmann Laboratory, F-38000 Grenoble, France

CNRS, LJK, F-38000 Grenoble, France, Inria

we focus here 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 us to estimate 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. 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. An index close to 1 means that the set is influent. At the opposite, an index equals to 0 means that the set is not correlated to the output. First-order indices estimate the main effect from each input whereas higher-order indices estimate the corresponding order of interactions between inputs. Various estimation procedures of these indices have been proposed in the literature (see [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 depending on the input space dimension. 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 \(2014\)](#). We propose here an extension of the replication procedure to deal with multidimensional output. This generalisation is then applied to the land use and activity module of *Tranus* to select the most influent parameters based on main effects and second-order interactions.

Following the sensitivity analysis, we propose an estimation procedure of the influent parameters of the land use and activity module using a two-stage iterative optimisation. Firstly, a stochastic algorithm is applied to find optimal values of the influent parameters selected by the sensitivity analysis. Secondly, an analytical optimisation of an internal dispersion parameter is performed taking as inputs the previously optimised parameters. This two-stage optimisation is iterated until an equilibrium is reached on the internal dispersion parameter. This second step is based on a careful investigation of the system of equations detailed in [Capelle et al. \(2015\)](#).

The rest of the paper is organised as follows. In [Section 2](#), a detailed description of *Tranus* is provided. In particular, we focus on the land use and activity module detailing the principal variables relevant to this paper. [Section 3](#) details the two main ingredients of our calibration

procedure from a methodological point of view: sensitivity analysis and the stochastic algorithm used in the two-stage iterative optimisation. In Section 4, we detail our whole calibration procedure in the form of an algorithm. Our methodology is then applied to the study area of North Carolina. Our 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*

In this paper, we will consider *Tranus*, a land use and transportation integrated model (LUTI). This type of software provides a framework for modelling land use and transportations in an integrated manner. It provides a flexible package to be used from urban, regional up to national scale. *Tranus* takes the basis from the classical input-output model (see [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 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 (see [McFadden \(1973\)](#); [McFadden and Train \(2000\)](#)), input-output analysis (see [Leontief \(1941\)](#)), land choice (see [Wilson \(1981\)](#)), multi-modal path choice and trip assignment. A comprehensive review of transport modelling can be found in [Ortuzar and G. \(2011\)](#).

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 nor land use, nor transport evolves anymore. [Figure 1](#) illustrates this mechanic. As it is already well calibrated, no further details on the transportation module of *Tranus* are 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,

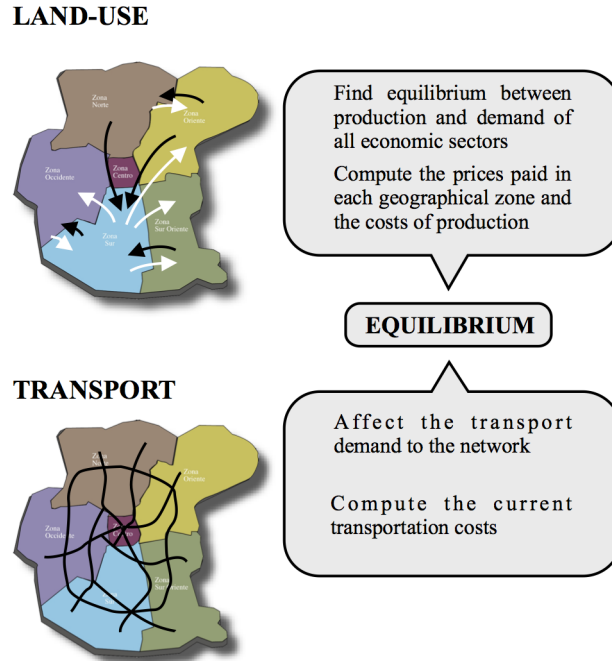


FIGURE 1. Schematic overview of *Tranus*.

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.

It is important for our paper 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 *generate flux*, that induces transport demand, which ultimately influence transportation costs. Non-transportable sectors, on the other hand, neither require transportation nor generate fluxes. In [Capelle et al. \(2015\)](#) we proposed a detached calibration of *Tranus*, calibrating first non-transportable sectors, and building up the rest of the calibration over them. We will consider this version of *Tranus* for the rest of the paper.

In the following, we introduce terminologies used in *Tranus* that are relevant to this paper. The set of equations relevant to this paper can be found in the Appendix. For a complete description of *Tranus*, see [de la Barra \(1999\)](#). We will consider a modelled territory or region divided in a set  $\mathcal{Z}$  of spatial zones, and the economy to be represented by a set  $\mathcal{N}$  of economical sectors. We will consider also that we have data  $Y_0 = \{Y_{0i}^n\}_{n \in \mathcal{N}, i \in \mathcal{Z}}$  corresponding to the observed production for a base year in each zone and for each economical sector. We denote as sub-index 0 this year of reference for the calibration.

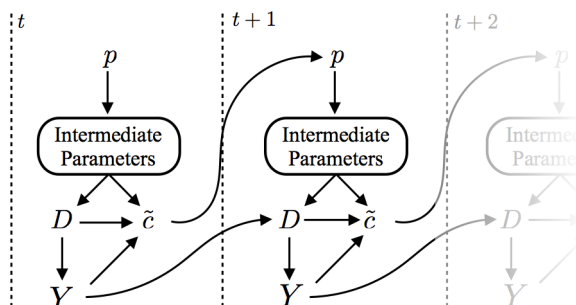


FIGURE 2. Sketch of computations in the land use and activity module.  $t$  is the iteration index of the convergence.

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

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

The computation of each of these quantities depends on one another and are linked by group of non-linear equations. The multiple dependencies of these variables are represented in Figure 2. The feedback is as follows, at current prices, after the computation of various intermediate variables (that will not be detailed here) we compute the demand and costs of consumption, finally computing the production. This production is not necessarily equal to the demand, and the costs of production are not necessarily equal to the prices, so we have to iterate these quantities until an equilibrium in production-demand and costs-prices is found. To do so, the costs of iteration  $t$  are fed as prices in iteration  $t + 1$ , and production in iteration  $t$  is compared to demand in iteration  $t + 1$ . All above computations are repeated until convergence is obtained in these two sets of variables. This process is done from bottom (land, exogenous prices) to top (transportable sectors).

### 2.3. Problem statement

The calibration procedure that we propose consists in adjusting the model parameters to be able to reproduce a base year’s data in the study area. It is important to note that our calibration procedure is applied independently for each sector  $n$  generating flux. Once a sector generating flux is selected, the model parameters relative to this sector are adjusted through our calibration procedure. The process is repeated for each sector generating flux. The justification of this approach relies on a sector-wise decomposition of the system of equations of the land use and activity module. Further details concerning this decomposition are provided in [Capelle et al. \(2015\)](#).

Given initial values of the parameters for a sector  $n \in \mathcal{N}$  generating flux, the land use and activity module estimates the adjustment parameters  $h^n = (h_i^n)_{i \in \mathcal{Z}}$  of the utilities (see Appendix A, (13)), known as shadow prices. The shadow prices are price correcting additive factors. They compensate the utilities to replicate the base year production  $Y_0$ . The following optimisation

problem is solved to compute the shadow prices:

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

where  $\| \cdot \|$  denotes the  $L^2$  norm. For the rest of the paper, this optimisation of the shadow price 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 sector  $n$  generating flux. The inputs parameters fall into different categories:

- parameter  $\lambda^n$  is involved in Equation (13) of Annexe A,
- parameters  $\beta^n$  is involved in Equation (14) of Annexe A,
- parameters  $b_l^n$ ,  $l \neq n, l \in \mathcal{N}$ , are involved in Equation (15) of Annexe A.

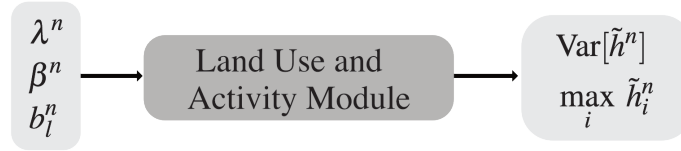
The outputs considered are built upon a new quantity called normalised shadow price. The normalised shadow price  $\tilde{h}_i^n$  corresponds to the percentage of the price  $p_i^n$  corrected by the shadow price  $h_i^n$ , that is:

$$\tilde{h}_i^n = 100 \times \left| \frac{h_{n,i}}{p_{n,i}} \right|, \quad i \in \mathcal{Z} \quad (2)$$

where  $| \cdot |$  denotes the absolute value function. We set  $\tilde{h}^n = (\tilde{h}_i^n)_{i \in \mathcal{Z}}$ , 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:  $\operatorname{Var}[\tilde{h}^n]$
- ii) the maximum of the normalised shadow prices:  $\max_{i \in \mathcal{Z}} \tilde{h}_i^n$

FIGURE 3. Inputs and outputs of the land use and activity module for the sector  $n$  generating flux.



For each sector  $n$  generating flux, a good calibration would be one that results in small values of the normalised shadow prices particularly in term of variance.<sup>1</sup>

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

### 3. Calibration procedure's main tools

#### 3.1. Global Sensitivity Analysis: replication procedure for multivariate outputs

In this section, we present a generalisation of the replication procedure, proposed in Mara and Joseph (2008) and deeply studied in Tissot and Prieur (2014), to the case of multidimensional

<sup>1</sup> Minimising the variance of the normalised shadow prices is a general consensus reached by both modellers and users of Tranus.

output. We first review the generalisation of the variance-based method introduced by Sobol' (1993) that relies on the estimation of generalised Sobol' indices (see Gamboa et al. (2014)).

Consider the following model:

$$f: \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{R}^m \\ x = (x_1, \dots, x_d) & \mapsto z = f(x) \end{cases}$$

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

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

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

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

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

$$\Sigma = C_u + C_{\sim u} + C_{u, \sim u} \quad (4)$$

Let  $\text{Id}_m$  be the  $m \times m$  identity matrix. Equation (4) can be projected on a scalar as follows:

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

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

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

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 influent. 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 from each input whereas higher-order indices estimate the corresponding order of interactions between inputs.

**Remark 1.** When  $u = (v, w)$  is a 2-subset of  $\{1, \dots, d\}$ , the influence 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 (see 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-frozen 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).

We 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  $Z_1, \dots, Z_N$  (resp.  $Z_1^u, \dots, Z_N^u$ ) be  $N$  independent copies of  $Z$  (resp.  $Z^u$ ) where:

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

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

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

Using this approach, estimating all first-order Sobol' indices requires  $N(d+1)$  evaluations of the model through  $d+1$  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 (2014) to the estimation of closed second-order indices. We propose here an extension of this procedure to the case of multivariate output. 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 method relies on the construction of two replicated designs of experiments  $\mathbf{X}$  and  $\mathbf{X}'$  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,  $\mathbf{X}$  and  $\mathbf{X}'$  are two replicated Latin Hypercubes. For the estimation of closed second-order indices,  $\mathbf{X}$  and  $\mathbf{X}'$  are two replicated randomised orthogonal arrays (see Tissot and Prieur (2014) for further details on the construction of  $\mathbf{X}$  and  $\mathbf{X}'$ ). Denote by  $\mathbf{Z}$  and  $\mathbf{Z}'$  the two

arrays of model outputs associated to these two designs. Let  $Z_i$  and  $Z'_i$  denote their corresponding rows,  $\forall i \in \{1, \dots, N\}$  we have:

$$Z_i = f(X_{i,1}, \dots, X_{i,d}) = (Z_{i,1}, \dots, Z_{i,m}),$$

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

The key point of the replication method consists in a “smart” rearrangement of the rows of  $\mathbf{Z}'$  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  $\pi$  denote the permutation used to re-arrange  $\mathbf{Z}'$ ,  $\forall i \in \{1, \dots, N\}$ :

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

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

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

thus insuring that both  $\mathbf{Z}$  and  $\mathbf{Z}^u$  are evaluated on the same  $u$  coordinates. We propose here to estimate  $S^u(f)$  using formula (7) with both  $\mathbf{Z}$  and  $\mathbf{Z}^u$ . In this fashion, we extend the replication procedure to the estimation of generalised Sobol' indices. For a concern of size of the paper, we choose to not further detail the replication procedure. The interested reader can find a detailed description in [Tissot and Prieur \(2014\)](#).

The selection of an influent parameter is made by fixing a threshold for both first- and second-order indices. If one of the estimated indices is higher than the threshold, the corresponding parameter is selected. For the application presented in Section 4, we subjectively fixed the threshold equals to 0.1. This choice is based on the assumption that the effective dimension (see [Caffisch et al. \(1997\)](#)) of the model is small ( $\leq 3$ ).

### 3.2. Stochastic optimisation: EGO algorithm

The stochastic optimisation procedure presented in this section corresponds to the Efficient Global Optimisation (EGO) algorithm introduced by [Jones et al. \(1998\)](#). The main idea underlying the EGO algorithm is to fit a response surface, often denoted by 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:

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

where  $x$  are the parameters selected with the sensitivity analysis,  $z$  a scalar output of the numerical model,  $d$  the dimension of the input space,  $\mu$  the model trend and  $\varepsilon$  is a centered stationary Gaussian process  $\varepsilon(x) \sim N(0, K_\chi)$ .  $\chi$  denotes the structure of the covariance matrix  $K_\chi$  of  $\varepsilon$ . 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:

- $r_\theta(\cdot)$  is the correlation function chosen here to be the Matérn 5/2 function,
- $\sigma^2$  is the variance of  $g$ ,
- $\theta$  are the hyperparameters of  $r$ .

The parameters  $\mu$ ,  $\sigma$  and  $\theta$  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:

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

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 Estimator) estimator (see 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 on 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 on this new point and re-estimate the parameters of the meta-model ( $\theta$ ,  $\sigma$ ),
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. Indeed, a value of the expected improvement close to zero indicates that the input space has been sufficiently explored. Thus, a lower bound on the expected improvement can be selected as the stopping criterion. Here, we set the lower bound equal to  $10^{-5}$ . Thus, the stopping criterion writes:

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

To ensure that the EGO algorithm finishes, we also fix a maximum number of iterations equal to 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, we detail our calibration procedure for the land use and activity module of Tranus in the form of an algorithm. The procedure includes both methods presented in Section 3. Our methodology is then applied to the study area of North Carolina.

#### 4.1. Calibration procedure algorithm

Algorithm 1 summarises our calibration procedure for the land use and activity module of Tranus. Recall that the calibration procedure is applied independently for each sector  $n$  generating flux.

---

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

---

- 1: **for** each sector  $n$  generating flux **do**
  - 2:   Set:  $\lambda^{n(0)} \leftarrow \lambda_0^n$
  - 3:   Run sensitivity analysis with inputs:  $\beta^n, \{b_l^n\}_{n \neq l}$  and outputs:  $Var[\tilde{h}^n], \max_{i \in \mathcal{Z}} \tilde{h}_i^n$
  - 4:   Instantiate:
    - $\rho^{n(0)} \leftarrow$  set of most influent 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  $\rho^{n*}$  and  $\lambda^{n*}$
- 

Once a sector  $n$  generating flux is selected, the sensitivity analysis presented in Section 3.1 is performed on the parameters  $\beta^n$  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 influent parameters selected is denoted by  $\rho^n$ .

Following the sensitivity analysis, an iterative optimisation is conducted. This optimisation is comprised of 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  $\lambda^n$ . At the end of the iterations optimal values  $\rho^{n*}$  and  $\lambda^{n*}$  are returned.

#### Remarks

1. One might wonder why the sensitivity analysis does not include the parameter  $\lambda^n$ . The main reason behind this choice is that the analytical optimisation provides a global optimum for  $\lambda^n$ . Hence by automatically selecting the parameter  $\lambda^n$ , we both 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, we chose to conserve only one output to perform the EGO algorithm: the variance of the normalised shadow prices:  $Var[\tilde{h}^n]$ .<sup>2</sup>

#### 4.2. Analytical optimisation procedure

Consider a sector  $n$  generating flux. In this section, we present an analytical procedure that estimates the parameter  $\lambda^n$ . For the sake of clarity, the notation  $\lambda^n, \lambda^{n(k)}$  and  $\beta^n, \beta^{n(k)}$  are confounded for the rest of the section.

---

<sup>2</sup> Minimising the variance of the normalised shadow prices over the maximum is a general consensus reached by both the modeller and users of Tranus.

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

$$U_{ij}^n = \lambda^n(p_j^n + h_j^n) + t_{ij}^n, \quad (i, j) \in \mathcal{L}^2. \quad (8)$$

The optimal value of  $\lambda^n$  cannot be retrieved directly from Equation 8 as the quantity  $(p_j^n + h_j^n)$  is estimated as a whole during the internal optimisation of the shadow prices (see Capelle et al. (2015) for further details). To overcome this problem, we introduce an auxiliary variable:

$$\phi_j^n = \lambda^n(p_j^n + h_j^n), \quad j \in \mathcal{L}.$$

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

$$U_{ij}^n = \phi_j^n + t_{ij}^n, \quad (i, j) \in \mathcal{L}^2. \quad (9)$$

Recall that the shadow prices are prices correcting additive factors that are calibrated to obtain a small variance. From Equation 9, we can express the optimal value of  $\lambda^n$  that minimises the variance of the shadow prices. We set  $\phi^n = (\phi_j^n)_{j \in \mathcal{L}}$  with all other parameters fixed, in particular the parameters  $p^n$  estimated with the EGO algorithm (see Section 3.2). The corresponding calibration problem writes:

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

Recall that we have  $p^n = (p_j^n)_{j \in \mathcal{L}}$  and  $h^n = (h_j^n)_{j \in \mathcal{L}}$  the vectors of prices and shadow prices. Once the optimal value  $\phi^{n*}$  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:

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

where the analytical solution equals  $\lambda^{n*} = \frac{\operatorname{Var}(\phi^{n*})}{\operatorname{Cov}(\phi^{n*}, p^{n*})}$ . Note that its 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

In this section, we propose an application of our calibration procedure to a study area in North Carolina. We first give a brief insight of the area of study. Then, we present the results of both the sensitivity analysis and the iterative optimisation used in our calibration procedure. The results are compared to those obtained with a former ad-hoc procedure.

**Study of area:** The area of study is the North Carolina State divided into 12 economical sectors and 103 spatial zones. The dataset was provided by Brian Morton<sup>3</sup>. In Table 1 are listed the different economical sectors of Tranus relative to the area of study. Our calibration procedure focus only on the sectors generating flux.

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

sector	name	description	generates flux
1	AFFHM(Agriculture,Fishing,...)	Business	no
2	Commercial	Business	yes
3	Other industries	Business	no
4	Single person (15-64 years old)	Households	yes
5	Married couple with kids	Households	yes
6	Married couple without kids	Households	yes
7	Other familie with kids	Households	yes
8	65 years and older	Households	yes
9	All other households	Households	yes
10	1-unit housing units	Floorspace	no
11	Multiunit housing units	Floorspace	no
12	Mobile homes	Floorspace	no

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

- The parameter  $\beta^n$ .
- The 11 parameters  $b_l^n, l \neq n$ .

In Table 2 are listed the distribution of each parameter. These distributions were selected by expertise. The Sobol' index  $S^k$  will refer to the parameter labeled by  $k$ . The outputs considered are

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

parameters	labels	distributions
$\beta^n$	1	$U(2,10)$
$b_l^n$	2, ..., 12	$U(0,1)$

the variance and the maximum of the normalised shadow prices 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 (see Tissot and Prieur (2014)). Using Remark 1, bootstrap confidence intervals can be derived for second-order Sobol' indices.

Before presenting the main results, we propose to illustrate the selection procedure of the influent parameters for a sector. Figures 4 and 5 shows 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 influent 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

<sup>3</sup> 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.

parameters  $\beta^4$  and  $b_{10}^4$  are the most influent. Since the sum of the first-order is less than 75% is it interesting to study the second-order indices.

For the estimation of second-order indices. A size  $N = 47^2$  was selected for the two replicated randomised orthogonal arrays required by the replication method. The two black points of Figure 5 correspond to the two most influent 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.

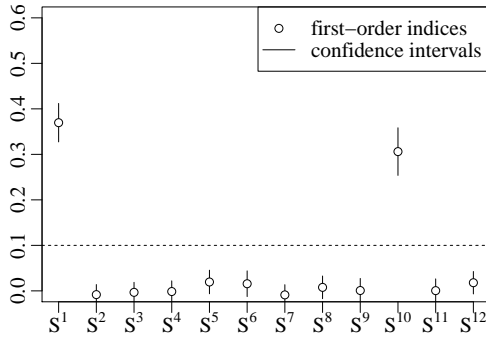


FIGURE 4. First-order indices estimation for the 12 parameters of sector 4

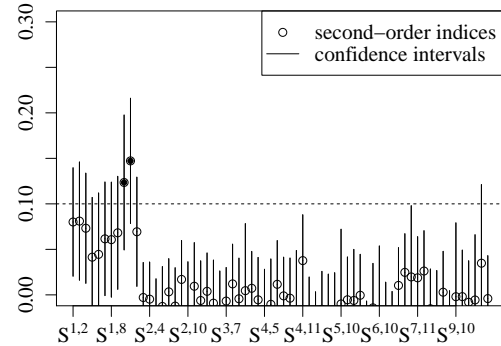


FIGURE 5. Second-order indices estimation for the 12 parameters of sector 4

As a conclusion, only 3 of the 12 parameters of the sector 4 are significantly influent either directly by their main effects or by their second-order interactions:  $\beta^4$ ,  $b_{10}^4$  and  $b_{11}^4$ .

The same procedure is performed for the other sectors generating flux. For each sector  $n$ , the set  $\rho^n$  comprised of the most influent parameters selected by the sensitivity analysis is listed in Table 3. The last column of the table gives the proportion of the 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 influent:  $\beta^n$ ,  $b_{10}^n$  and  $b_{11}^n$ ,  $n \in \{2, 4, 5, 6, 7, 8, 9\}$ .

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

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

These results fall within our range of expectation. The parameter  $\beta^n$  is a dispersion parameter of a multinomial logit function (see 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 influent than the other  $b_l^n$  since sectors 10 and 11 correspond to the two main floorspace.

**Remark 2.** 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. An alternative would consist in using Saltelli's procedure to estimate first-order, second-order and total-effect Sobol' indices (index measuring the contribution of an input through all its interactions). However, Saltelli's procedure would require  $N \times 25$  evaluations of the model for each sector generating flux. In comparison, our approach requires only  $N \times 4$  evaluations of the model to estimate both first-order and closed second-order indices. Since the number of evaluations was a concern for us, we decided to go for the replication procedure at the price of not using the total effect Sobol' indices.

**Results of the iterative optimisation** For each sector  $n$  generating flux, the aim is to find the set of parameters  $(\rho^n, \lambda^n)$  minimising the variance of  $\tilde{h}^n$ . The initial value  $\lambda_0^n$  (Step 3 of Algorithm 1) instantiating the parameter  $\lambda^n$  is obtained by expertise. The results obtained in terms of variance and maximum of the normalised shadow prices are compared to those obtained with a former ad-hoc procedure.

Table 4 summarises the results obtained with both the ad-hoc procedure and our iterative optimisation.  $\rho^{n*}$  and  $\lambda^{n*}$  denote the optimal values of the parameters obtained at the end of both approaches for each sector  $n$  generating flux. The column gain represents the improvement (in percentage) of the variance obtained with our iterative estimation relatively to the one obtained with the ad-hoc procedure conducted by experts.

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

sector $n$	procedure	$\rho^*$	$\lambda^*$	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	$7 \times 10^{-3}$	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	

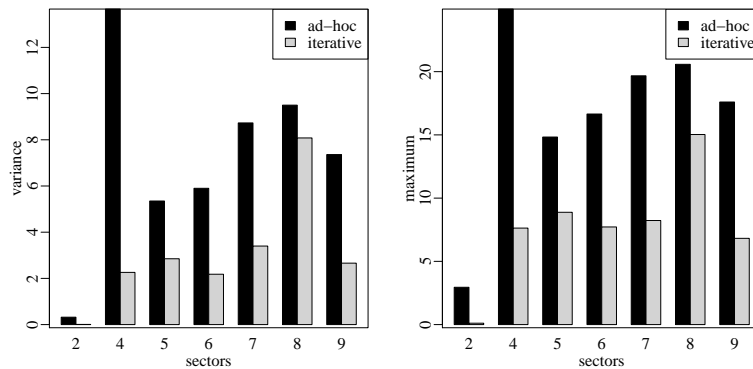
Looking at the results, we observe that the values of the variance and maximum of the normalised shadow prices obtained with the ad-hoc procedure are heterogeneous. Furthermore, the value of the maximum is quite high for some sectors (up to 20% of the price). The results obtained with our iterative optimisation 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 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 lacking of 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 while using our 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 our iterative approach. A significant diminution for both the variance and maximum criteria is observed.

FIGURE 6. Variance (left figure) and maximum (right figure) of the normalised shadow prices obtained with both the ad-hoc procedure (refers as ad-hoc) and our iterative optimisation (refers as iterative)



Furthermore and most importantly our approach is drastically faster than the ad-hoc procedure conducted by experts. Our calibration procedure requires a few hours compared to several days (up to weeks) for the ad-hoc procedure.

## Conclusion

In this paper we proposed 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. We presented as sensitivity analysis a generalisation of the replication procedure to select the most influent parameters of the model when the output is multidimensional. The optimisation phase was carried on using a two stages process combining stochastic and deterministic procedures. An application to the study area of North Carolina was presented. Our methodology was compared to a former ad-hoc calibration procedure in terms of variance and maximum of the normalised adjustment parameters (shadow prices). Our approach showed a significant improvement on both criteria reducing the value of the variance by a large margin with a drastic gain of time. These results have proved that our methodology is useful to outperform 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 are thankful to Brian J. Morton for providing high quality data and instantiating the Tranus model on the study area of North Carolina. 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.
- Caffisch, 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*, Volume 115:12–20.
- 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.
- 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*.
- Leontief, W. (1941). *The Structure of the American Economy 1919-1939*. New York: Oxford University Press, 2nd edition, 1951 edition.
- Leontief, 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.
- Monod, H., Naud, C., and Makowski, D. (2006). *Uncertainty and sensitivity analysis for crop models*, chapter 3, pages 55–100. Elsevier.
- 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.
- Ortuzar, J. d. D. and G., W. L. (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*, Volume 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. (2014). 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*. Chichester, 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  $n$ :

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

$$D_i^n = D_i^{*n} + \sum_m D_i^{mn} \quad (12)$$

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  $j$ :

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

Here,  $\lambda^n$  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, we compute 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:

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

Here,  $k$  ranges over all zones.  $\beta^n$  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:

$$A_j^n = \sum_l b_{n,l} Z_j^l \quad (15)$$

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  $i$ :

$$Y_{ij}^n = D_i^n Pr_{ij}^n . \quad (16)$$

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

$$Y_j^n = \sum_i Y_{ij}^n \quad (17)$$

$$= \sum_i D_i^n Pr_{ij}^n . \quad (18)$$

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