

# ANALYSE DE SENSIBILITÉ ET OPTIMISATION POUR LE CALAGE D'UN MODÈLE TRANSPORT-URBANISME

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

<sup>1</sup> *Inria Grenoble - Rhône-Alpes, Inovalée, 655 avenue de l'Europe, 38330 Montbonnot, prenom.nom@inria.fr*

<sup>2</sup> *LJK, Université de Grenoble, 51 rue des Mathématiques Campus de Saint Martin d'Hères, 38041 Grenoble cedex 09, prenom.nom@imag.fr*

**Résumé.** Le développement et l'utilisation de modèles intégrés 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 souvent utilisés comme outils d'aide à la décision pour des politiques de développement urbain et d'aménagement des sols. Le calage des modèles transport-urbanisme est une tâche longue et difficile qui met en jeu l'estimation d'un nombre important de paramètres. Nous proposons ici une méthode de calage du modèle en libre accès *Tranus*. La méthode proposée comprend deux étapes : une phase d'analyse de sensibilité et une phase d'optimisation. La phase d'analyse de sensibilité utilise une généralisation de la méthode répliquée au cas de sorties multidimensionnelles. La phase d'optimisation est une procédure itérative combinant deux approches : une stochastique (algorithme EGO) et une analytique. Nous proposons d'illustrer notre méthode de calage pour une instanciation de *Tranus* sur la zone d'étude de Caroline du Nord.

**Mots-clés.** modèle transport-urbanisme, analyse de sensibilité, optimisation, EGO.

**Abstract.** The development and the use of LUTI (Land Use and Transport Integrated) models have become a norm for representing the interactions between land use and the transportations of goods and people in a territory. These models are often used as decision-making tools for the evaluation of urban planning and environmental policies. The calibration of LUTI models is a difficult and time-consuming process requiring the estimation of a large number of parameters. We propose here a calibration procedure for the open-source LUTI model *Tranus*. This procedure is composed of two steps: sensitivity analysis and optimisation. The sensitivity analysis presented is a generalisation of the replication procedure to deal with multivariate outputs. The optimisation step is an iterative process combining stochastic (EGO algorithm) and deterministic approaches. We propose to illustrate our calibration procedure for an instantiation of *Tranus* on the urban area of North Carolina.

**Keywords.** LUTI model, sensitivity analysis, optimisation, EGO.

# 1 Contexte

Nous considérons ici le modèle transport-urbanisme *Tranus* développé par de la Barra (1999). *Tranus* est un logiciel offrant une structure intégrée pour modéliser l’usage des sols et le réseau de transport sur un territoire. *Tranus* comprend deux composantes principales : un module d’activités, usage des sols simulant l’activité économique d’un territoire, et un module de transport estimant la demande en transport et l’assignant aux modes de transport disponibles. Le territoire est découpé en zones spatiales et secteurs économiques. Les principaux secteurs économiques sont les types de ménages, les types de sols et/ou de logements et les types d’emplois. Il est important de distinguer deux catégories de secteurs économiques : les secteurs générant des flux et ceux qui n’en génèrent pas. Les secteurs générant des flux induisent une demande en transport qui influence les coûts de transport.

L’analyse des modèles transport-urbanisme est une tâche difficile mettant en jeu un nombre considérable de paramètres dont il est difficile d’identifier les effets. Les interactions existantes entre ces paramètres font qu’une petite variation d’un paramètre peut entraîner des changements importants dans les sorties du modèle. Dans ce cas, le calage joue un rôle central, permettant d’estimer les valeurs optimales des paramètres et d’obtenir un modèle robuste. Les approches classiques de calage utilisées s’appuient en majorité sur des procédures économétriques ad hoc.

La procédure de calage que nous proposons ici porte sur le module d’activité et d’usage des sols de *Tranus*. Cette procédure consiste à ajuster les paramètres du modèle afin de reproduire les données (productions, prix) d’une année de base sur une zone d’étude. Notre procédure est appliquée indépendamment pour chaque secteur économique générant des flux. Etant donné un jeu de valeurs initiales pour les paramètres d’un secteur générant des flux, le module d’activités et usage des sols procède à une optimisation interne de variables d’ajustement appelées “shadow prices”. Les sorties du module d’activité et d’usage des sols que nous considérons sont des quantités d’intérêt construites à partir des “shadow prices” : variance et maximum.

Notre procédure de calage comprend deux étapes principales : une phase d’analyse de sensibilité et une phase d’optimisation. La méthode d’analyse de sensibilité proposée ici est une généralisation de la méthode répliquée au cas de sorties multivariées. La phase d’optimisation est une procédure itérative combinant deux approches : la première, stochastique, correspond à l’algorithme EGO (Efficient Global Optimisation) introduit par Jones *et al.* (1998) et la deuxième est une formulation analytique de l’optimum pour un type de paramètre. Dans les deux sections qui suivent, nous présentons l’analyse de sensibilité puis l’algorithme général synthétisant notre procédure de calage. Nous introduisons la méthode d’analyse de sensibilité dans un cadre général non spécifique au module d’activités et d’usage des sols de *Tranus*.

## 2 Analyse de sensibilité pour sorties multivariées

Nous commençons par passer en revue une généralisation de la méthode basée sur la décomposition de la variance introduite par Sobol' (1993). Cette généralisation, proposée par Gamboa *et al.* (2014), s'applique au cas de sorties multivariées d'un modèle et introduit des indices de sensibilité appelés : indices de Sobol' généralisés.

Considérons le modèle suivant:

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

où  $f$  représente le modèle étudié,  $z$  le vecteur de sorties du modèle,  $x$  le vecteur de paramètres d'entrée du modèle,  $d$  la dimension de l'espace des paramètres d'entrée et  $m$  la dimension de l'espace des sorties du modèle. Notons  $(\Omega, \mathcal{A}, \mathbb{P})$  l'espace de probabilité. L'incertitude sur les paramètres d'entrée est modélisée par un vecteur aléatoire  $X = (X_1, \dots, X_d)$  de composantes indépendantes. Soit  $Z$  le vecteur de variables aléatoires modélisant le vecteur de sorties du modèle :

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

Notons  $P_X = P_{X_1} \otimes \dots \otimes P_{X_d}$  la distribution de  $X$ . Nous supposons que  $f \in \mathbb{L}^2(P_X)$  et que la matrice de covariance de  $Z$ , notée par  $\Sigma$ , est définie positive. Soit  $u$  un sous-ensemble de  $\{1, \dots, d\}$  et  $\sim u$  son complémentaire. Nous fixons  $X_u = (X_i)_{i \in u}$  et  $X_{\sim u} = (X_i)_{i \in \{1, \dots, d\} \setminus u}$ . Rappelons la décomposition fonctionnelle ANOVA de  $f$  introduite par Hoeffding (1943) :

$$f(X) = f_0 + f_u(X_u) + f_{\sim u}(X_{\sim u}) + f_{u, \sim u}(X_u, X_{\sim u}), \quad (1)$$

où  $f_0 = \mathbb{E}[Z]$ ,  $f_u = \mathbb{E}[Z|X_u] - f_0$ ,  $f_{\sim u} = \mathbb{E}[Z|X_{\sim u}] - f_0$  et  $f_{u, \sim u} = Y - f_u - f_{\sim u} - f_0$ . En prenant la matrice de covariance de chaque côté de l'équation (1), par orthogonalité nous obtenons l'équation suivante:

$$\Sigma = C_u + C_{\sim u} + C_{u, \sim u}. \quad (2)$$

Notons  $\text{Id}_m$  la matrice identité de taille  $m \times m$ . L'équation (2) peut être projetée comme suit:

$$\text{Tr}(\text{Id}_m \Sigma) = \text{Tr}(\text{Id}_m C_u) + \text{Tr}(\text{Id}_m C_{\sim u}) + \text{Tr}(\text{Id}_m C_{u, \sim u}), \quad (3)$$

où  $\text{Tr}$  correspond à l'opérateur trace. En utilisant (3) et sous la condition  $\text{Tr}(\Sigma) \neq 0$ , nous définissons l'indice de Sobol' généralisé :

$$S^u(f) = \frac{\text{Tr}(C_u)}{\text{Tr}(\Sigma)}. \quad (4)$$

L'indice de Sobol' généralisé  $S^u(f)$  est un scalaire compris entre 0 et 1 qui évalue l'influence de l'ensemble de paramètres d'entrée  $u$  sur le vecteur de sorties  $Z$ . Un indice proche de 1 signifie que l'ensemble  $u$  est influent. A l'opposé, un indice proche de 0 signifie que l'ensemble  $u$  n'est pas corrélé avec  $Z$ . Les indices d'ordre un estiment l'effet principal de chaque paramètre d'entrée tandis que les indices d'ordre supérieur estiment les interactions d'ordre correspondant entre les paramètres d'entrée.

**Méthode répliquée pour  $S^u(f)$**  La procédure classique pour estimer  $S^u(f)$  est une généralisation de celle proposée par Sobol’ (1993) dans le cas univarié. Cette procédure correspond à une méthode Monte-Carlo “pick-freeze”. Dans la méthode “pick-freeze”, l’indice de Sobol’ est interprété comme le coefficient de régression entre la sortie du modèle et sa réplique “pick-frozen”. La réplique est obtenue en fixant la valeur de la variable d’intérêt  $X_u$  (variable “frozen”) et en rééchantillonnant les autres variables  $X_{\sim u}$  (variables “picked”).

Avec cette procédure, estimer tous les indices généralisés d’ordre un requiert  $N(d + 1)$  évaluations du modèle à partir de  $d + 1$  plans d’expériences chacun contenant  $N$  points d’évaluation. Dans le cas univarié, la méthode répliquée proposée par Mara et Rakoto (2008) permet d’estimer tous les indices d’ordre un avec seulement deux plans répliqués, chacun de taille  $N$ , résultant en un total de  $2 \times N$  évaluations du modèle. Cette méthode a été étudiée plus précisément et étendue par Tissot et Prieur (2014) pour l’estimation des indices d’ordre deux.

Nous proposons ici une extension de cette procédure au cas de sorties multivariées. Avec cette nouvelle approche, le nombre d’évaluations du modèle requis pour évaluer tous les indices de Sobol’ généralisés d’ordre un ou d’ordre deux est considérablement réduit.

### 3 Procédure de calage

Algorithme 1 synthétise notre procédure de calage pour le module d’activités et d’usage des sols de Tranus. Cette procédure est appliquée indépendamment pour chaque secteur  $n$  générant des flux. Le paramètre  $\lambda^n$  correspond au paramètre optimisé par l’approche analytique dans la phase d’optimisation. Une fois un secteur  $n$  générant des flux sélectionné,

---

**Algorithm 1** Procédure de calage pour le module d’activité et d’usage des sols de Tranus.

---

- 1: **for** chaque secteur  $n$  générant des flux **do**
  - 2:   Initialiser :  $\lambda^{n(0)} \leftarrow \lambda_0^n$ .
  - 3:   Lancer l’analyse de sensibilité pour l’ensemble de paramètres du secteur  $n$ .
  - 4:   Initialiser :
    - $\rho^{n(0)} \leftarrow$  l’ensemble de paramètres identifiés comme les plus influents,
    - $k \leftarrow 1$ .
  - 5:   **while**  $|\lambda^{n(k)} - \lambda^{n(k-1)}| \geq \epsilon$  **do**
  - 6:     Etant donné  $\lambda^{n(k-1)}$ , estimer  $\rho^{n(k)}$  avec l’algorithme EGO.
  - 7:     Etant donné  $\rho^{n(k)}$ , estimer  $\lambda^{n(k)}$  avec la procédure analytique.
  - 8:   Retourner les valeurs optimales  $\rho^{n*}$  et  $\lambda^{n*}$ .
- 

l’analyse de sensibilité présentée dans la section 2 est appliquée aux paramètres du secteur  $n$ . Les sorties du module considérées sont la variance et le maximum des “shadow prices”. L’ensemble de paramètres identifiés comme les plus influents est noté  $\rho^n$ .

A la suite de l’analyse de sensibilité, une optimisation itérative est conduite. Cette optimisation se décompose en deux étapes. A l’itération  $k$ , étant donné  $\lambda^{n(k-1)}$ , l’algorithme EGO est appliqué pour trouver les valeurs optimales du set  $\rho^{n(k)}$ . Puis, une optimisation analytique du paramètre  $\lambda^{n(k)}$  est réalisée étant donné les valeurs optimales trouvées pour le set  $\rho^{n(k)}$ . Le processus est itéré jusqu’à obtenir un équilibre sur le paramètre  $\lambda^n$ .

L’analyse de sensibilité n’inclut pas le paramètre  $\lambda^n$ . La raison de ce choix est que la procédure analytique optimisant  $\lambda^n$  fournit une valeur optimale globale pour ce paramètre. Pour l’optimisation utilisant l’algorithme EGO nous avons choisi de ne prendre comme sortie du module que la variance des “shadow prices”. Minimiser la variance des “shadow prices” au lieu du maximum est un consensus général entre les utilisateurs de Tranus.

## 4 Application

Nous illustrons notre procédure de calage pour les paramètres liés à un secteur économique de Tranus générant des flux. La zone d’étude considérée est l’Etat de Caroline du Nord. Les données relatives à cette zone d’étude nous ont été fournies par Brian Morton <sup>1</sup>. Le secteur considéré est un type de ménage (célibataire d’âge compris entre 15 et 64 ans), associé à la valeur  $n = 4$ . Les paramètres du module d’activités et usage des sols associés à ce secteur sont au nombre de 12. Le nombre de “shadow prices” calculés par le module est de 103 (un pour chaque zone spatiale du territoire).

**Résultats de l’analyse de sensibilité** Nous appliquons la méthode répliquée pour le cas de sorties multivariées. Nous estimons avec cette méthode les indices généralisés de Sobol’ d’ordre un et deux relatifs aux 12 paramètres du secteur considéré. Les sorties du modèle sont la variance et le maximum des “shadow prices” relatifs à ce secteur.

Figures 1 and 2 illustrent les résultats obtenus pour l’estimation des indices généralisés d’ordre un et deux. Pour les indices d’ordre deux les intervalles de confiance sont obtenus par bootstrap (1000 répétitions). Une taille  $N = 5000$  a été choisie pour chacun des deux plans requis par la méthode répliquée. Les deux points noirs de la figure 2 correspondent aux deux interactions les plus influentes. Ainsi, pour le secteur considéré, seulement 3 des 12 paramètres relatifs à ce secteur sont influents soient par leurs effets principaux soit par leurs interactions d’ordre deux. Le set  $\rho^4$  est donc constitué de 3 paramètres.

**Résultats de l’optimisation itérative** L’optimisation itérative est ensuite appliquée pour trouver les valeurs optimales de  $\rho^4$  et  $\lambda^4$ . La valeur initiale  $\lambda_0^4$  est obtenue par expertise. Les résultats obtenus sont comparés à des précédents résultats obtenus avec une procédure économétrique ad-hoc. Table 1 donne les résultats obtenus avec chaque

---

<sup>1</sup>Brian Morton est un chercheur travaillant au centre de recherche “Center for Urban and Regional Studies” de Caroline du Nord North, Hickerson House, 108 Battle Lane, Campus Box 3410. Les données nous ont été fournies à l’occasion d’un groupe de travail portant sur le calage des modèles LUTI.

Figure 1: Indices d'ordre un.

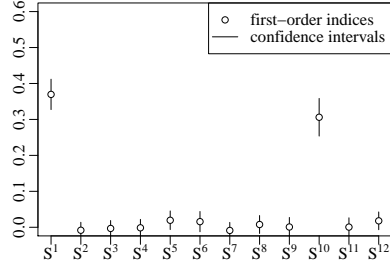
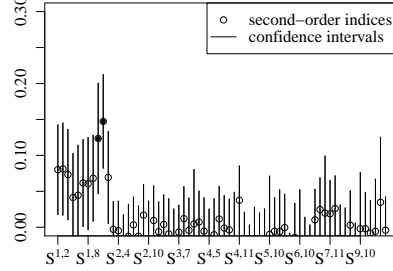


Figure 2: Indices d'ordre deux.



approche en terme de variance et maximum des “shadow prices”. On observe que notre algorithme itératif conduit à une diminution considérable à la fois de la variance et du maximum des “shadow prices”.

| méthode   | $\rho^{4*}$     | $\lambda^{4*}$ | variance SP | maximum SP |
|-----------|-----------------|----------------|-------------|------------|
| ad hoc    | (2, 0, 0)       | 0.001          | 13.66       | 24.86      |
| itérative | (6.49, 0.38, 0) | 0.001          | 2.26        | 7.63       |

Table 1: Variance et maximum des “shadow prices” (SP) obtenus par la procédure ad hoc et notre optimisation itérative.

## Bibliographie

- [1] de la Barra, T. (1999), *Mathematical description of TRANUS*, Modelistica, Caracas.
- [2] Gamboa, F et Janon, A. et Klein, T. et Lagnoux, A. (2014), Sensitivity analysis for multidimensional and functional outputs, *Electron. J. Statist.*, 8, 1, 575–603.
- [3] Hoeffding, W. F. (1948), A class of statistics with asymptotically normal distributions, *Annals of Mathematical Statistics*, 19, 3, 293–325.
- [4] Jones, D. R. et Schonlau, M. et Welch, W. J. (1998), Efficient Global Optimization of Expensive Black-Box Functions, *Journal of Global Optimization*, 13, 455–492.
- [5] Mara, T. A. et Rakoto, O. J. (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.
- [6] Sobol’, I. M. (1993), Sensitivity indices for nonlinear mathematical models, *Mathematical Modeling and Computational Experiment*, 1, 407–414.
- [7] Tissot, J. Y. et 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.