



HAL
open science

Calibration of the Tranus Land Use Module: Optimisation-Based Algorithms, their Validation, and Parameter Selection by Statistical Model Selection

Thomas Capelle, Peter Sturm, Arthur Vidard, Brian Morton

► **To cite this version:**

Thomas Capelle, Peter Sturm, Arthur Vidard, Brian Morton. Calibration of the Tranus Land Use Module: Optimisation-Based Algorithms, their Validation, and Parameter Selection by Statistical Model Selection. *Computers, Environment and Urban Systems*, 2019, 77, pp.101146:1-13. 10.1016/j.compenvurbsys.2017.04.009 . hal-01519654

HAL Id: hal-01519654

<https://inria.hal.science/hal-01519654v1>

Submitted on 9 May 2017

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.

Calibration of the Tranus Land Use Module: Optimisation-Based Algorithms, their Validation, and Parameter Selection by Statistical Model Selection

Thomas Capelle ^a, Peter Sturm ^a, Arthur Vidard ^a, Brian Morton ^b

^a*Inria and Univ. Grenoble Alpes, Lab. LJK, Grenoble, France*

^b*Center for Urban and Regional Studies, Univ. of North Carolina at Chapel Hill, USA*

Abstract

Instantiating land use and transport integrated models (LUTI modelling) is a complicated task, requiring substantial data collection, parameter estimation and expert analysis. In this work, we present a partial effort towards the automation of the calibration of Tranus, one of the most popular LUTI models. First, we give a detailed mathematical description of the activity module and the usual calibration approach. Secondly, we reformulate the estimation of the endogenous parameters called shadow prices as an optimisation problem. We also propose an optimisation algorithm for the calibration of the substitution submodel, setting a base for future fully integrated calibration. We analyse the case of transportable and non-transportable economic sectors and propose a detailed mathematical scheme for each case. We also discuss how to validate calibration results and propose to use synthetic data generated from real world problems in order to assess convergence properties and accuracy of calibration methods. Results of this methodology are presented for realistic scenarios. Finally, we propose a model selection scheme to reduce the number of shadow prices that need to be calibrated, with the aim of reducing the risk of overfitting to data.

Keywords: Land use, optimisation, transport, LUTI, model selection

1. Introduction

LUTI (land-use and transportation integrated) models aim at representing the complex interactions between land use and transportation offer and demand within a ter-

^{*}This document is a collaborative effort.

E-mail addresses: thomas.capelle@inria.fr (Thomas Capelle), peter.sturm@inria.fr (Peter Sturm), arthur.vidard@inria.fr (Arthur Vidard), bjmorton@email.unc.edu (Brian Morton).

ritory. They are mainly used to evaluate different alternative planning scenarios, by simulating their tendential impacts on patterns of land use and travel behaviour. Since the early 60's LUTI modelling has attracted researchers that aimed to model the complex economical relations in urban areas; a good overview of the evolution and history of LUTI modelling can be found in [1]. Setting up a LUTI model requires the estimation of several types of parameters to reproduce as closely as possible, observations gathered on the studied area (socio-economic data, transport surveys, etc.). The vast majority of available calibration approaches are semi-automatic, estimating one subset of parameters at a time, without a global integrated estimation. Automatic calibration of LUTI model is not a common practice; an exception has been proposed for the Meplan model [2].

We consider *Tranus* [3, 4], an open source LUTI model that is widely used. *Tranus* is a classical LUTI model, with two separated modules: the activity module and the transport module. The activity module, is an equilibrium type model based on micro-economic principles that balance the offer and demand of the different economical sectors that interact at each level. Economical sectors are considered in the broad sense, amongst them we have: land, goods, salaries, housing, transportation demand, etc. Also, the price paid for each economical sector has to be balanced with respect to offer and demand, thus there are two equilibria that have to be achieved, offer versus demand and (production) cost versus prices. The transportation module, computes the costs of transportation and assigns the demand to the network. Both modules interact back and forth until a general equilibrium is achieved.

The calibration process is usually done by an expert modeller who iteratively tunes a group of parameters to reproduce as closely as possible the observations gathered in the area of study. This process is usually done manually, with little to no automation, adjusting the different economical parameters (for example, the demand curves for different goods in a specific geographical zone). At the same time, *Tranus* computes internally a set of adjustment coefficients (called shadow prices in *Tranus*) that correct the utilities and account for un-modelled effects. These endogenous variables help the model achieve a better response and fit more precisely to the observed data.

In this paper we address several shortcomings of the classical approach of calibration used in *Tranus*. In our previous work [5] we proposed a first approach for the reformulation of the heuristic calibration algorithm used in the land use and activity module as an optimisation problem. In this paper we extend this approach by having a closer look at the inner loop that computes the shadow prices and propose an efficient methodology for their estimation by decoupling the calibration in smaller problems. To be able to do this, we have to carefully investigate the system of equations that are computed in the activity module. We also introduce auxiliary variables, which enables a closed form computation instead of an iterative one. This in turn makes it possible to use sophisticated numerical optimisation methods and opens the door to the simultaneous estimation of different parameter types of the model. The ultimate goal of this approach is to simultaneously calibrate the various parameters of *Tranus*' inner and outer loops. In this direction, this scheme gives valuable results for usually hard-to-calibrate parameters (so-called penalising factors).

We also present a detailed methodology for the construction of synthetic scenarios based on real calibrated study areas. These synthetic scenarios have a perfect fit

without the need of the so-called shadow prices (usually we set their value to zero), enabling us to validate our optimisation algorithms knowing the ground truth values of the shadow prices. A simple example is presented to understand the problematic of synthetic scenario generation and the corresponding equilibrium prices problem.

Finally, we question the rationale of usual calibration approaches for *Tranus* (and other LUTI models), which consist in estimating parameters for which the model reproduces observations exactly. In *Tranus*, this is achieved by enriching the underlying macro-economic model with the already mentioned auxiliary variables, the shadow prices. While this allows to correct for unavoidable un-modeled effects, it also bears the risk of over-parameterisation and overfitting. We propose a model selection scheme, aiming at a compromise between model complexity (here, number of shadow prices) and goodness of fit to observations, reducing the risk of overfitting and increasing the likelihood of achieving good predictions with a model. These issues are related to the question of model sensitivity; indeed sensitivity analyses are valuable tools to understand how uncertainties in a model's outputs can be apportioned to different sources of uncertainty, such as the uncertainty in calibrated model parameters.

2. Description of *Tranus*

Tranus is a land use and transportation integrated model (LUTI), providing a framework for modelling land use and transportations in an integrated manner. It can be used at urban, regional or even national scale. The area of study is divided in **spatial zones** and **economical sectors**; the basic concepts of the original input-output model (see [6]) have been generalised and given a spatial dimension. 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 two main modules: the **land use and activity** and the **transportation** modules.

The **land use and activity** module simulates a spatial economic system by modelling the locations of activities and the interactions between economic sectors for a specific time period. The location of activities is determined by the land use and activity module based on various factors, such as land prices/rents and transportation costs and disutilities (such as time spent in transportation), the latter being provided by the transportation module. The **transportation** module, on the other hand, dispatches the travel demand induced by the activity model (such as home-to-work and shopping trips) and assigns it to the transport supply. Based on this, transportation disutilities (such as caused by traffic congestion) are re-calculated and output by the transportation module. Both modules are thus linked together, serving both as input and output to each other. 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 turn, the accessibility that results from the transport system influences the location and interaction between activities, also affecting land rent. The two modules use discrete choice logit models [7, 8], linked together in a consistent way. This includes activity-location, land-choice, and multi-modal path choice and trip

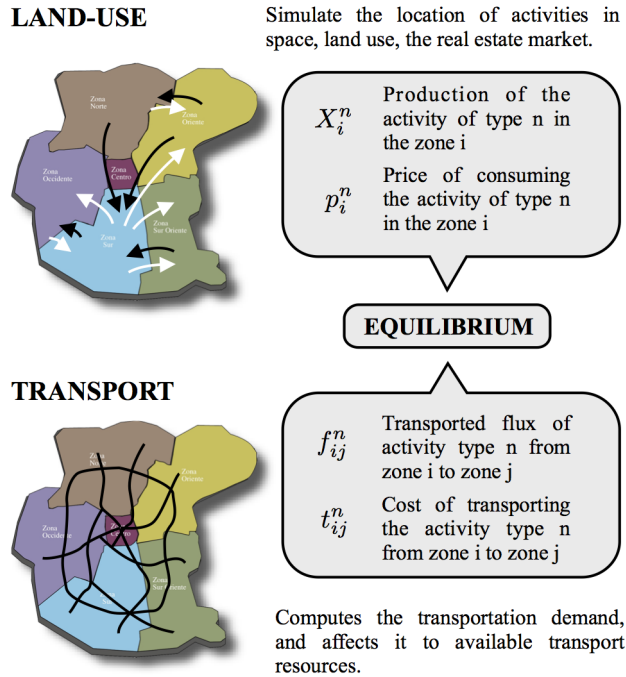


Figure 1: Schematic overview of Tranus.

assignment.

To attain a convergence status, Tranus runs both modules iteratively until an equilibrium is found. The land use and transportation modules need to reach their own respective equilibrium status. First, the land use module needs to achieve equilibrium between offer and demand, and equilibrium between the price paid and the cost of producing each economic sector. This is done at current transportation costs and disutilities. Secondly, the transportation module takes as input the transport demand and equilibrates the transportation network to satisfy the given demand. Both modules are run iteratively until a general equilibrium status is found. This is achieved when neither land use nor transportation, evolve anymore, as illustrated in Figure 1.

2.1. The land use and activity module

In this paper we only work with the land use and activity module, (from now on land use module). Our main goal is to improve this module and make the calibration of the parameters involved easier. We will consider the input needed (for the calibration of the land use and activity module) from the transport module as data readily available. This technique of “freezing” the transportation system is already used by Tranus modellers for the calibration of floorspace sectors and land. To do so, we will have to make the distinction 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 influences transportation costs. Non-transportable sectors, on the other hand, neither require transportation nor generate fluxes. Usually, three types of economic sectors are classified: land or floorspace, households and businesses. Land is usually composed of two or three types of residential floorspace (e.g. detached houses, apartments, mobile homes), and commercial floorspace of offices and stores. Households are usually classified by socio-economic level, based on income or the household composition. Business sectors comprise industries (whose output is mainly destined for exportation), services (schools, universities, recreational) and commerce. The standard approach for the consumption chain is as follows: Industry has a demand for labour (households) and service businesses. Households also consume services, and services also require labour, thus “consume households”. Finally, all businesses and households consume land. For instance, households will locate in residential zones, and the feedback of household and business “consumption” will induce home-to-work trips (see [9]). This process results in economic exchanges, sometimes inducing flux (transportable sectors) and sometimes in-place consumption (land). The offer and demand is equilibrated and a set of equilibrium prices for each economic sector is attained.

The land use module’s objective is to find an equilibrium between the production and demand of all economic sectors and zones of the modelled 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 following, we introduce the parts of the terminology, parameters and equations used in *Tranus* that are relevant to this paper. See [10] for a complete description.

- **Productions:** X_i^n expresses how many “items” of an economic sector n are present/produced in a zone i .
- **Demands:** D_i^m expresses how many items of a sector n are demanded by the part of sector m located in zone i .
- **Prices:** p_i^n defines the **price** of (one item of) sector n produced in zone i .

It is important to realise that “price” in the case of land, is the actual rent, whereas the price of a household represents salary.

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 [10]. 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 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,

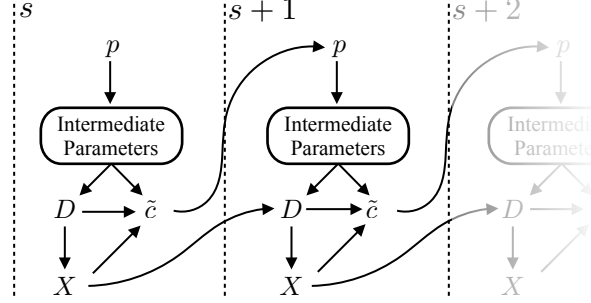


Figure 2: Sketch of computations in the land use module. The variables in this figure are p and c (Eqs. 2.9 and 2.8), demands D (Eq. 2.1) and productions (Eq. 2.5). The intermediate variables are: Pr_{ij}^n (Eq. 2.4), S_i^{mn} (Eq. 4.1) and a_i^{mn} (Eq. 3.6), for more details about the intermediate variables see table 1.

the current production is computed and fed back to compute a new set of prices, for a next iteration. The process is bottom-up, starting with land use prices and exogenous production (destined for exportation outside the study area) and demand up to the production 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).

In the following, we only show those model equations that are relevant to this paper. Demand is computed for all combinations of zone i , demanding (consuming) sector n and demanded sector m :

$$D_i^{mn} = (X_i^{*m} + X_i^m) a_i^{mn} S_i^{mn} \quad (2.1)$$

$$D_i^n = D_i^{*n} + \sum_m D_i^{mn} \quad (2.2)$$

where X_i^{*m} is the given exogenous production (for exports), X_i^m the induced endogenous production obtained in the previous iteration (or initial values), and D_i^{*n} exogenous demand. D_i^n in (2.2) 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 (explained later in more detail).

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

$$U_{ij}^n = \lambda^n (p_j^n + h_j^n) + t_{ij}^n . \quad (2.3)$$

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, 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:

$$Pr_{ij}^n = \frac{A_j^n e^{-\beta^n U_{ij}^n}}{\sum_l A_l^n e^{-\beta^n U_{il}^n}} . \quad (2.4)$$

Here, A_j^n represents attractiveness of zone j for sector n and β^n is the dispersion parameter for the multinomial logit model expressed by the above equation.

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

$$X_{ij}^n = D_i^n Pr_{ij}^n . \quad (2.5)$$

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

$$X_j^n = \sum_i X_{ij}^n \quad (2.6)$$

$$= \sum_i D_i^n Pr_{ij}^n . \quad (2.7)$$

Given the computed demand and production, consumption costs are computed as

$$\tilde{c}_i^n = \frac{\sum_j X_{ij}^n (p_j^n + tm_{ij}^n)}{D_i^n} \quad (2.8)$$

where tm_{ij}^n is the monetary cost of transporting one item of sector n from a production zone j to consumption zone i .

These finally determine the new prices:

$$p_i^m = VA_i^m + \sum_n a_i^{mn} S_i^{mn} \tilde{c}_i^n \quad (2.9)$$

where VA_i^m is value added by the production of an item of sector m in zone i , to the sum of values of the input items.

2.2. Calibration of shadow prices as done in *Tranus*

The calibration process consists in adjusting the model parameters to be able to reproduce a base year's data in the study area. Obtaining a good calibration is a long process, that is usually performed by experts and can take months. A mix of tools are used to estimate the various parameters of the model. Econometrical, ad-hoc procedures and interactive trial-and-error can be counted among the tools used by experts to obtain a good fitting model.

For the calibration phase, parameters are separated in three sets:

- i. Parameters that are computed externally using the appropriate data and econometrical techniques (for example demand coefficients and attractors).
- ii. The adjustment parameters h_j^n of the utilities (2.3), known as shadow prices.

- iii. The remaining parameters (for example the penalisation factors and logit dispersion parameters).

After computing the external parameters (set i), and giving initial values to set iii, the model iterates until convergence. The iteration process is constructed in such a way, that the shadow prices will be adjusted to force the productions to reproduce the observed productions X_0 in the study area. These variables will “try” to compensate for the other parameters to have a perfect fit; they act as correction terms to compensate for parts of the utility that are not represented by the model. One wants to make the values of the shadow prices as small as possible. This process of parameter calibration is done repeatedly until the expert modeller is satisfied with the parameters and the values of the shadow prices.

The computation of the shadow prices is automatically done as follows at the end of each iteration (cf. figure 2 and the above equations):

$$h_i^{n,s+1} = (h_i^{n,s} + p_i^{n,s}) \frac{X_i^{n,s}}{X_{0,i}^n} - p_i^{n,s+1} . \quad (2.10)$$

Based on the same approach as proposed by [11], the shadow prices for the next iteration $s + 1$ increase proportionally to the excess of computed, as compared to observed, productions.

2.3. Parameters to Calibrate

Table 1 presents the parameters to calibrate in the land use module, showing where in the computations they are involved, a brief description of their meaning and the type of calibration commonly used to obtain their value. It is important to note that even if many parameters are estimated externally with available data, very frequently some adjusting is done in Tranus afterwards.

Intermediate quantity	Equations	Parameter	Description	Type of Calibration
a_i^{mn}	(2.1, 2.9, 3.6)	min^{mn}	The minimum consumption	(i)
		max^{mn}	The maximum consumption	(i)
		δ^{mn}	Demand Elasticity	(i), (iii)
		h_i^n	Shadow price	(ii)
S_i^{mn}	(2.1, 4.1)	σ^n	Logit dispersion parameter	(iii)
		ω^{mn}	Penalising factor	(iii)
		h_i^n	Shadow price	(ii)
P_{ij}^n	(2.4, 2.7)	λ^n	Marginal utility of Income	(i), (iii)
		β^n	Logit dispersion parameter	(iii)
		h_i^n	Shadow price	(ii)

The three intermediate quantities exposed in table 1 are the drivers of the Tranus land use module, and as we can observe, the shadow prices (h_i^n) are involved in the computation of all three of them.

3. Proposed Calibration Approaches

Our main motivations are to replace the sequential calibration process outlined above by a process that rigorously estimates as many parameters as possible, taking into account all available constraints and assumptions in a systematic manner, to automatise as much as possible the calibration process, and to make it more reproducible. We believe that a natural way of achieving these goals is to explicitly formulate the calibration process in terms of a cost function (or possibly, as a multi-criteria decision problem) that is to be minimised or maximised, with respect to a set of constraints, when given. This is for example not directly the case in the existing approach, where the estimation of shadow prices and other parameters is done without definition of an explicit cost function. Formulating calibration via explicit cost functions enables to use the rich variety of optimisation algorithms existing in the literature and in numerical libraries.

A first step in this direction concerns the estimation of shadow prices, a second step deals with the automatic estimation of both shadow prices and other parameters; these two steps are described in the following.

3.1. Reformulating calibration of shadow prices as an optimisation problem

It is important to notice that calibration of the land use module involves the estimation of all the parameters of the model to make productions as close as possible to the observed base year data X_0 .

To reformulate the calibration as an optimisation problem, we must compute shadow prices that make the productions as similar as possible to the observed productions. This can be written as an optimisation problem:

$$\min_h \|X(h) - X_0\|^2 . \quad (3.1)$$

Here, h is a vector containing all shadow prices, X_0 the vector of observed productions, and $X(h)$ the vector of productions computed by the model, after convergence of the iterative process shown in figure 2. The dependency of these on the shadow prices is visible from equations (2.3) to (2.7). Likewise, each evaluation of the productions $X(h)$ involves the convergence of the dynamic system exposed in Figure 2. Each evaluation of the cost function involves the convergence of the dynamic system in productions as well as prices.

This double convergence problem can be avoided by including the prices amongst the variables to be optimised, instead of leaving them as endogenous variables. Moreover, one can compute directly productions that are in equilibrium for a given set of shadow prices and prices. To do so, we observe that the computation of demand and production involves a set of linear equations (2.1), (2.2), (2.5), and (2.7). If we reorganise these equations, knowing that only productions are needed in our cost function, we may only need to compute these. To do so, we substitute D_i^n in equation (2.5) using equations (2.1) and (2.2), giving:

$$X_{ij}^n = \left\{ D_i^{*n} + \sum_m (X_i^{*m} + X_i^m) a_i^{mn} S_i^{mn} \right\} Pr_{ij}^n . \quad (3.2)$$

Upon substituting this into (2.7), we obtain the following linear system in X_j^n :

$$X_j^n = \sum_i \left\{ D_i^{*n} + \sum_m X_i^{*m} a_i^{mn} S_i^{mn} \right\} Pr_{ij}^n + \sum_i \sum_m a_i^{mn} S_i^{mn} Pr_{ij}^n X_i^m . \quad (3.3)$$

If presented in matrix form, this correspond in general to a matrix of size $M \cdot N$ where M is the number of zones, and N is the number of sectors.

By construction, the solution of this linear system represents an equilibrium of production and demand: if one computed demands D from the obtained productions, and computed productions from them, the latter would be equal to the productions obtained in the first place.

The most usual optimisation methods require the computation of partial derivatives of the cost function [12]. This is still difficult for the cost function (3.1). Each evaluation of the productions involves solving a linear system of the type (3.3). An analytical solution seems out of reach even for models with few sectors and zones. Estimating the gradient numerically via finite differences, is possible but rather costly. It would require at least one evaluation of $X(h)$ per shadow price to estimate, each evaluation requiring the solution of the linear system (3.3). Moreover, even if productions computed this way are in equilibrium, the prices p still need to iterate until convergence is obtained. Indeed, convergence in prices is only obtained when consumption costs equal corresponding prices (cf. equations (2.8) and (2.9) as well as figure 2).

These remaining difficulties can be solved as follows. First, for a successful calibration, we want to have the computed productions equal to the observed base year productions X_0 . This correspond to the usual rationale for Tranus models ¹. Hence, we can simply impose this condition by replacing productions in the right hand side of (3.3), with the observed base year productions. This approach enables us to compute the productions directly, without the need to solve a linear system. Similarly, this simplifies the analytical computation of the cost function's derivatives.

To address the second problem (equilibrium of prices), we add the prices explicitly to the set of parameters to be optimised. We use the current values of prices, and compare them against the prices computed by the model in the next iteration, cf. (2.9). The difference between the current prices and the ones computed by the model through equations (2.3) to (2.9), is added to (3.1), in order to form a new cost function:

$$\min_{h,p} \|X(h, p, X_0) - X_0\|^2 + \|\hat{p}(h, p, X_0) - p\|^2 . \quad (3.4)$$

¹Achieving perfect equality between observed productions and productions generated by the model, is in general possible since there are as many shadow prices to adjust, as there are observed productions. In section 6.2, we discuss ideas for alternative rationales.

Here, \hat{p} is the vector of prices computed by the model using (2.9) and the notation $X(h, p, X_0)$ shows that modelled productions are computed as explained above by substituting observed productions X_0 into the right-hand side of (3.3).

The above cost function has a closed-form that permits us to compute the derivatives in h and p directly. No more iterations or waiting for convergence is required in this approach. The cost function (3.4) is of (non-linear) least squares type, meaning that any least squares optimisation approach can be used; in our work we apply the Levenberg-Marquardt method [13].

Let us also note that other choices than the L_2 norm would of course be possible to define the cost function of (3.4). We may also weight the two terms differently, in order to favour equilibrium in production over that in prices or vice-versa in cases where a global equilibrium cannot be reached.

So far, we have not used any specificities of activity sectors in the outlined approach. This is done in the following two sections, first for non-transportable and then for transportable sectors.

3.2. Land use sectors (non transportable sectors)

The classical non transportable sector is land. It is a very particular economical sector, it must be consumed where it is produced. Moreover, land does not consume other economical sectors and the amount of available land is fixed. For the calibration purpose, the prices for the land use sectors are known, this means the p_i^n variables for the calibration year are considered as input and do not need to be computed. This translates into a simplified set of equations for the computation of production of land. We have to detail two extra equations to understand how this enters our optimisation scheme. First, as land is non-transportable, the location probability (2.4) vanishes, so equation (3.3) can be re-written as:

$$X_i^n = D_i^{*n} + \sum_m (X_{0i}^{*m} + X_{0i}^m) a_i^{mn} S_i^{mn} . \quad (3.5)$$

We now detail the two variables a_i^{mn} (technical demand coefficients) and S_i^{mn} (substitution probabilities):

$$a_i^{mn} = \min^{mn} + (\max^{mn} - \min^{mn}) \exp(-\delta^{mn}(p_i^n + h_i^n)) . \quad (3.6)$$

Where \min^{mn} , \max^{mn} and δ^{mn} are parameters of the demand function estimated externally using data on rental prices. The substitution probability S_i^{mn} is only a function of shadow prices associated to the same zone i , see the next section for more details. As land prices are known, we can clearly see that the production of land X_i^n is only a function of the shadow prices of the same geographical zone i (there is no dependency on other zones). This leads to optimisation problems (one per zone) that are very small, with the number of variables equal to that of land use sectors. We can re-write the optimisation problem (3.4) as one optimisation problem for each geographical zone i :

$$\forall i \quad \min_{h_i^l} \|X_i^l(h_i^l, X_0^l) - X_{i0}^l\|^2 \quad (3.7)$$

where the superscript l stands for land sector. An example of the optimisation methodology for land sectors is presented in appendix Appendix B.

Once the optimisation is done for each geographical zone and the shadow prices for land use are computed, we can proceed to computing the optimal shadow prices of the transportable sectors. We will further exploit this feature of the model to obtain an automated calibration of the substitution parameters. Details on how to compute partial derivatives of the cost function, required by the optimisation algorithm, can be found in the appendix.

3.3. New approach for computing the induced production of transportable sectors

Transportable sectors, are economical sectors that consume (and can be consumed) in a different location from where they are produced. Households and commerce are examples of such sectors. For this type of sector, the technical coefficients a_i^{mn} are considered constant and in all practical usages of Tranus, the substitution probabilities $S_i^{mn} = 1$, i.e., there is no substitution considered in the transportable economic sectors (this is not a limitation of the model, rather it is common practice). Realising this implies that the total demand D_i^n is not a function of the prices or the shadow prices, and enables the computation of the total demand D_i^n (2.2) for each transportable sector n and geographical zone i only as a function of the base year data X_0 . In fact, this is done just after computing the land use sectors' shadow prices.

Thus, in the computation of the induced productions of a transportable sector n , X_j^n (cf. (2.7): $X_j^n = \sum_i D_i^n Pr_{ij}^n$), we only need to determine the values of the location probabilities Pr_{ij}^n . If we go back to the definition of the location probabilities (2.4) and the underlying utilities (2.3), we realise that the utility makes no distinction between the price and shadow price part, so if we set:

$$\phi_j^n = p_j^n + h_j^n \quad (3.8)$$

the location probability can be computed as a function of ϕ . Instead of posing the induced production as a function of (h, p, X_0) , we can look at the induced production $X(\phi, X_0)$ as a function of ϕ . Obtaining the optimal values of ϕ that minimise the difference between computed and observed productions, is the solution to the following problem:

$$\min_{\phi} \|X(\phi) - X_0\|^2 \quad (3.9)$$

Since the location probability Pr_{ij}^n is a function only of the ϕ^n variables for the same sector n , we get one optimisation problem for each economical sector n . Each of these optimisation problems is relatively simple and small in size, there are as many variables as geographical zones. The gradient of the cost function can be computed analytically using the well known derivatives of the logit probability Pr . We use the Levenberg-Marquardt method to solve each problem. Once all the optimal values of ϕ have been computed, we can compute the prices by solving the linear system (2.9) for prices. Doing so allows us to recover the shadow prices from ϕ , subtracting the prices from the respective optimal ϕ values.

One consideration that one has to deal with, is that the location probabilities Pr follow a logit formulation, so the utilities can only be identified up to a constant per

economical sector. This is a known property of logit models. As the prices are obtained from equation (2.9), this approach is considerably simpler and more stable than solving the double-objective optimisation approach proposed in (3.4), moreover, it exploits every little detail of the formulation of each function of the model. It also permits to calibrate incrementally, starting by the land use sectors and then obtaining the calibration in the transportable sectors. From the mathematical point of view it is also simpler, because the large optimisation problem in (3.4) is now decoupled into smaller optimisation problems, with fewer variables, allowing the modeller to finish the calibration of one set of variables before moving to the next stage.

To summarise, in the case of land use (non-transportable) sectors, there is one small optimisation problem to be solved for each geographical zone, whereas for the transportable sectors, we have one optimisation problem per economic sector. In the appendix Appendix B, we present a detailed application of our methodology to a small Tranus scenario. The example is solved in detail to give an insight on how the optimisation algorithm computes shadow prices and prices.

We encountered some numerical issues related to the fact that the location probability would vanish for large values of the utility function; this behaviour is explained in the appendix Appendix C.

4. Simultaneous estimation of shadow prices and land use substitution parameters

As stated in the introduction, one would like to have a simultaneous estimation of the whole set of parameters. In this section we present one step in this direction, we have constructed a two-phase algorithm that permits the estimation of the shadow prices and the substitution parameters at the same time. We have chosen the penalising factors in the substitution submodel (ω^{mn} in (4.1)) because these are very hard to calibrate parameters, as relevant data are not readily available. The scheme we propose exploits the fact that the substitution submodel is used for land sectors, where we have already a simplified computation of the productions, as explained in Section 3.2. Tranus models include a discrete choice submodel that represents the households' ability to choose among different types of residential buildings (i.e., floor space). That choice is captured in the "substitution model". The model is driven by the substitution probabilities:

$$S_i^{mn} = \frac{W_i^n \exp(-\omega^{mn} a_i^{mn} \cdot (p_i^n + h_i^n))}{\sum_{l \in K^m} W_i^l \exp(-\omega^{ml} a_i^{ml} \cdot (p_i^l + h_i^l))} . \quad (4.1)$$

Here, K^m represents the set of substitutes that sector m has access to, for example, for "rich" households m , this could be $K^m = \{\text{condos, detached houses}\}$. We will calibrate the penalising factors ω^{mn} , they represent a correcting multiplicative coefficient for household m consuming housing sector n . Using Tranus terminology, W_i^n is an "attractor", a parameter that represents attributes of floor space sector n other than cost (utility); it is specified (and potentially calibrated) for each zone in which sector n is present. From equation (3.6) we can see that the demand coefficient a_i^{mn} is also a func-

tion of the prices and shadow prices. It is important to remember that prices are known for land sectors.

Simultaneous calibration of shadow prices and penalising factors.

1. **Phase 1: estimating parameters' initial values with multinomial logistic regression.** The substitution model's parameters are estimated with multinomial logistic regression [7]. The data that are essential for estimation are household level observations on floor space consumption, housing expenditure, and the Tranus sector to which the household belongs. The dependent variable in a regression will be the choice of floor space sector, and the independent variable is the housing expenditure. The regressions are conducted separately for each household sector, and they yield estimates of ω^m for each combination of floor space sector n and household sector m .
2. **Phase 2: fine tuning the penalising factors.** The penalising factors estimated in Phase 1 probably still need to be fine tuned to reduce the differences between the predicted production of floor space and the observed production of floor space. Fine tuning probably would also be necessary to achieve reasonable values of the floor space sectors' shadow prices.

If we consider all of Tranus' parameters fixed except the parameters ω , and include these parameters in the optimisation problem presented in (3.4), we obtain the following cost function:

$$f(h, \omega) = \|X(X_0, h, \omega) - X_0\|^2 . \quad (4.2)$$

We would like to find the values of ω that reduce the corresponding shadow prices. We propose to solve the following equation:

$$\min_{\omega \in \Omega} f(h = 0, \omega) \quad (4.3)$$

where Ω is a set of bounds over the penalising factors ω . We used a conjugate-gradient algorithm to solve this problem, and the starting points are the values obtained from the Multinomial Logistic regression of Phase 1. If we call ω^* the solution of (4.3), then the final values for the shadow prices for the land use sectors are:

$$h^* = \arg \min_h f(h, \omega^*) .$$

The details and derivative estimation are presented in the appendix.

5. Generation of Synthetic Scenarios for Validation of Calibration Algorithms

The most usual methods for validating calibration results for LUTI models are to evaluate the goodness-of-fit to the data used for calibration and expert validation (assessment of the plausibility of calibrated model parameters by expert modelers). The latter is more qualitative, whereas the former gives a quantitative validation. However, a well-known issue in data assimilation is that goodness-of-fit to data used for model

calibration, is an imperfect indicator of calibration accuracy: even a perfect fit does not in general guarantee that calibrated parameter values are accurate (this issue is called “bogus calibration” in [14]; it is also related to the potential problem of model overfitting, see the following section). This is for instance the case, but not only, if there exist multiple solutions that represent a good fit, which is possible in general for non-linear models.

Additional, alternative, methods for validation are ideally applied to validate model calibrations. For the specific case of LUTI models for instance, Batty recommends to validate land use calibration results using data not used for this calibration phase, such as observations on trip distributions [14]. Another validation approach is to compare calibrated parameters against “ground truth” values, whenever these are available or can be obtained using other data and/or models. As for Tranus (and LUTI models in general), several parameter types are usually not directly observable in practice, meaning that such ground truth information is not accessible. In the following, we propose an approach to for generating plausible synthetic datasets with an associated ground truth, on the basis of which calibration algorithms may be assessed (typically, in terms of convergence properties and accuracy). The idea is to establish a synthetic model (with shadow prices set to an arbitrary value, we simply choose zero as this value), that is a perfect fit to synthetic data on observed productions. One way of doing so would be to invent an entirely artificial scenario. Instead, we aimed at obtaining a scenario that should be as close as possible to a real or realistic scenario. To do so, we “tweak” real observed productions so that they become perfectly consistent with a model with the chosen shadow prices. With the synthetic scenario thus obtained, we can now validate our calibration approaches, since we know the (synthetic) “ground truth” for the parameters to estimate: calibration is run on the generated synthetic data (possibly, by adding noise to these data), shadow prices (or analogously, other parameters) are estimated and then compared to the ground truth. This enables the validation of calibration approaches by means other than goodness-of-fit.

To generate such a synthetic “perfect fit” scenario, we have to solve a subproblem of the original calibration optimisation problem exposed in (3.4), where we do not consider the observed productions. We only need to obtain equilibrium in prices, and compute the values of the induced productions afterwards. To do so, we replace the consumption cost equation (2.8) in the prices (2.9), and by identifying the location probability as $Pr_{ij}^n = X_{ij}^n/D_i^n$, we obtain the following system:

$$p_i^m = VA_i^m + \underbrace{\sum_n a_i^{mn} S_i^{mn} \sum_j Pr_{ij}^n(h, p) \cdot (p_j^n + tm_{ij}^n)}_{\hat{p}(h, p)} . \quad (5.1)$$

The dependence of the location-choice probability makes this system hard to solve even for small models. Our approach to solve this fixed-point equation is to solve the following optimisation problem:

$$\min_p \|\hat{p}(h, p) - p\|^2 . \quad (5.2)$$

We have to make sure that the solution of (5.2) is a set of prices that are in equilibrium, that is for which $\hat{p} = p$.

After obtaining convergence in the prices, we compute the induced productions and then use them as observed productions in our synthetic scenario. This methodology produces a scenario where the optimal value of the shadow prices is zero (by construction) and that reproduces the (synthetic) productions perfectly. We could also set the shadow prices value to any other value than zero here. As mentioned, the generation of such a synthetic scenario enables us to test our calibration methodology and optimisation algorithms against a known optimal value (here, shadow prices equal to zero).

6. Results

6.1. Results of Shadow Price Calibration

We consider the example exposed in [5], a simple model that allows to illustrate the methodology for generating synthetic scenarios with perfect fit (with “ground truth” shadow prices equal to zero). We applied our approach to the *Example_C* model from the Tranus website², a small model with 3 zones and 5 sectors. First, we generated synthetic data from that model as described just above, with shadow prices $h_i^n = 0$ for each sector n and zone i . As expected, the cost function is zero at $h = 0$, and increases its value when we get away from the optimum. The cost function appears to be locally convex near the optimal value, cf. figure 3.

If we consider for example sector 1 and zone 1, we can plot a “slice” of the cost function (3.4) near the optimal value $h_1^1 = 0, p_1^1 = 2.676$ as shown in figure 3. Here we can observe that as the shadow price gets larger the cost increases up to a plateau state ($X_1^1(h) \rightarrow 0$). In the case of the price p , if we move away from the optimal value $p = 2.676$, the cost increases quadratically.

We tested the robustness of the optimisation scheme with 1,000 random initial sets of shadow price values; the optimisation procedure outlined in this paper always converged to the ground truth solution. The initial values of shadow prices in these random trials were generated from a uniform distribution in $[-10, 10]$, which is highly representative (prices are in the interval $[0, 4]$ and nearly all shadow prices of a model are in practice smaller than the corresponding prices).

We also applied the same procedure to a Tranus model for North-Carolina (North-Carolina-1 model, consisting of 102 zones and 12 economical sectors) modified by our synthetic data methodology (Section 5). After setting the desired value for the shadow prices to $h = 0$, we tried 10,000 random initial sets of shadow prices values; and the algorithm proved to converge to the correct shadow prices for every single starting point. As for the calibration procedure implemented in Tranus’ release, it failed to converge when starting values were too far away from the zero vector. We considered initial shadow prices uniformly distributed in the interval $[\epsilon \cdot -p_{max}, \epsilon \cdot p_{max}]$, where ϵ is a parameter in $[0, 1]$ and p_{max} is the maximum observed price. As ϵ increases, the initial shadow prices can take values further away from the optimal solution $h^* = 0$.

²<http://www.tranus.com/tranus-english>

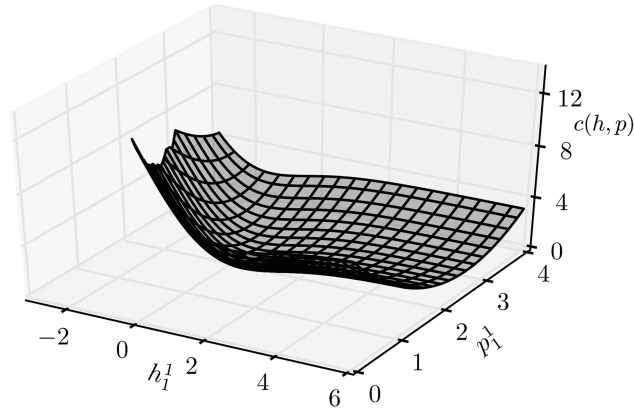


Figure 3: Plot of cost function for a given pair (h_1^1, p_1^1) near the optimal value $(0, 2.676)$.

These initial values are representative of the expected values of shadow prices as one would like that shadow prices do not exceed prices. Table 2 presents the convergence status for each value of ϵ (1,000 random values were taken for each ϵ). We observe that the iterative approach of Tranus fails to converge as the initial values get further away from the true solution.

Table 2: Comparison of calibration algorithms for the North-Carolina-1 model. Shown are the percentages of random trials for which the algorithms converged to the correct solution.

ϵ value:	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Tranus	100%	100%	63%	3%	0%	0%	0%	0%	0%	0%
Our algorithm	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

6.2. Towards a statistical model selection scheme

We consider here the application of statistical model selection in order to reduce the number of shadow prices needed to have a model reproduce the observed data. The classical approach is to iteratively modify the parameters until a perfect fit is achieved (with near zero cost function), when this is achieved the modeller will look at the values of the shadow prices as a quality test for the calibration. If the shadow prices are large, it means that the model has to compensate the various effects of the other economical parameters to attain a perfect fit. Thus, the modeller will tweak economical parameters (such as dispersion parameters) to maintain the perfect fit but with smaller shadow prices. A calibration will be completed when the model reproduces the observed data perfectly and the values of the shadow prices are small (for some economical sectors we will ask their variance to be small instead). As there are as many shadow prices as observations we are trying to fit (there is one shadow price per observed production) the risk of overfitting is possible, which will in turn undermine the predictive capabilities of the model.

What we propose, is sacrificing the perfect fit of the cost function, in order to lower the number of parameters to calibrate, particularly of shadow prices. To find the optimal trade-off between how many shadow prices we keep in the model and the desired value of the cost function is something that will have to be discussed with the community of modellers. What we propose here, is a simple model selection scheme that instead of having one shadow price per economical sector and zone, keeps only one shadow price per geographical zone. Doing so enables us to exploit the fact that we have independent optimisation problems for each geographical zone for land (non transportable sectors). In this case, for the North-Carolina-1 model, as there are only 3 economical sectors for land, we reduce by two thirds the complexity of the model. The calibration of the remaining third of shadow prices gave rise to a residual fit of the cost function of only 3% (ratio of residuals over observed productions).

Selecting the shadow prices to be kept in the model is easily done for the land sectors (non-transportable), because the prices (rents) are known. We propose the simplest model selection scheme: to keep just one shadow price per geographical zone, this is an arbitrary choice among multiple possible selection schemes. We achieved this, by computing the optimal shadow prices from (3.1) and setting to zero the “small” shadow prices. Followed by a re-calibration of the remaining shadow prices. One can adjust the threshold used to declare shadow prices as being small to a desired level in the cost function, thus keeping more or less shadow prices. We also see in Figure 4 that the values of the shadow prices relatively to the prices have not seen a large increase.

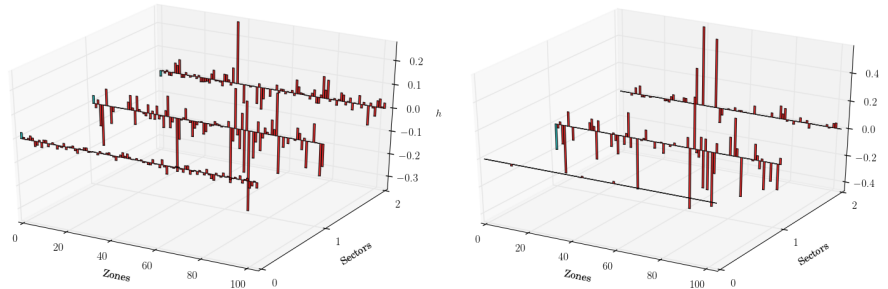


Figure 4: The graphs show ratios of calibrated shadow prices over prices. Left: when all shadow prices are estimated (in this case, the fit of computed to observed productions is perfect). Right: here only one third of shadow prices are estimated, the others are set to zero. The fit is not perfect but good (3%, see text). Note that the scales of the two graphs are different. One can observe that on the right-hand side, shadow price to price ratios are not much larger than on the left-hand side, another indicator that it is plausible to exclude many shadow prices from the calibration.

We are currently working on testing this methodology for two time periods, to determine if noise in the base year productions could really propagate to the shadow prices and undermine the predictive capabilities of the model.

6.3. Results of Estimation of Shadow Prices and Substitution Parameters

We applied this procedure (see Section 4) to a second real-scale LUTI model for North-Carolina (North-Carolina-2 model), with 38 zones, 3 floorspace and 9 other economic sectors. Figure 5 shows the shadow prices for all zones and floorspace sectors, after the two phases of our process, cf. section 4. After each phase, a global equilibrium of demand, production and prices, is achieved, however after the novel second phase, shadow prices are much smaller, meaning that the model represents reality much better (small ratios of shadow prices over prices is a crucial criterion used by practitioners to assess the quality of a Transus LUTI model).

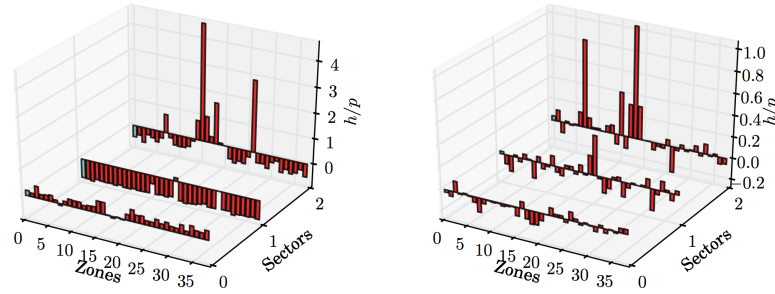


Figure 5: Ratios of shadow prices and prices after phase 1 (left) and 2 (right). Note the different scales of the graphs.

7. Conclusions and final remarks

The Transus LUTI framework is a very powerful tool and the modelling possibilities are endless. However, the complexity of such large scale models is something that can not be underestimated, making the calibration and utilisation of these tools very expensive. We have contributed with a reformulation of the land use module that simplifies the calibration process, exploiting the very basics of the mathematics that are behind the microeconomic models used, permitting the expert to incrementally calibrate the variables (from land use sectors to transportable sectors). The optimisation approach is more stable and clear than the classical approach, and enables the use of powerful optimisation algorithms currently available, solving the occasional non convergence issues of the previous approach.

The procedure exposed for generating synthetic data is simple and straightforward, enabling us to try and benchmark our methodologies. We are currently preparing a set of benchmarks of calibrated models.

The proposed methodology for reducing the number of shadow prices needs additional fine tuning, but is a first step in what we consider a promising direction. We believe that the model “as it is” with one shadow price per observation bears a risk of overfitting. Determining which shadow prices have to be removed may not be completely automisable, and the expert eye of the modeller has to have the last call. Building on

this work on semi-automatic parameter estimation, a sensitivity analysis approach for Tranus has been developed in [5] where the most influential parameters are automatically selected and then optimised with stochastic and deterministic procedures.

Finally, the simultaneous calibration of different parameter types is a potentially very powerful tool. The results that we have for the North Carolina models have proven to be useful and saved many trial and error sessions. We would like to apply this idea of simultaneous optimisation to other “hard” to calibrate parameters, and we are working with modellers to identify them. A fully integrated and automatic calibration is our dream.

Acknowledgment: This work is supported by the CITiES project (ANR-12-MONU-0020). We are grateful to the anonymous reviewers for their suggestions, which helped to improve this paper.

References

- [1] M. Wegener, Overview of land-use transport models, in: D. A. Hensher, K. Button (Eds.), *Transport Geography and Spatial Systems, Handbook in Transport*, Pergamon/Elsevier Science, Kidlington, UK, 2004, pp. 127–146.
- [2] J. E. Abraham, *Parameter estimation in urban models: Theory and application to a land use transport interaction model of the sacramento, california region*, PhD thesis, University of Calgary, Canada (2000).
- [3] T. de la Barra, Modelling regional energy use: a land use, transport and energy evaluation model, *Environment and Planning B: Planning and Design*, 9 (1982) 429–443.
- [4] T. de la Barra, *Integrated Land Use and Transport Modelling*, Cambridge University Press, 1989.
- [5] T. Capelle, P. Sturm, A. Vidard, B. J. Morton, Formulating LUTI calibration as an optimisation problem: Example of tranus shadow price estimation, *Procedia Engineering* Volume 115 (2015) 12–20.
- [6] W. Leontief, A. Strout, *Multi-Regional Input-Output Analysis, Structural Interdependence and Economic Development*, London: Mcmillan, 1963.
- [7] D. McFadden, Conditional logit analysis of qualitative choice behaviour, *Frontiers in Econometrics - Academic Press New York* (1973) 105–142.
- [8] D. McFadden, K. Train, Mixed MNL models for discrete response, *Journal of Applied Econometrics* 15 (2000) 447–470.
- [9] I. Lowry, *A model of metropolis*, Tech. Rep. Memorandum RM-4035-RC, The RAND Corporation, Santa Monica, California (1964).
- [10] T. de la Barra, *Mathematical description of TRANUS*, Tech. rep., Modelistica, Caracas, Venezuela (1999).
URL <http://www.tranus.com/tranus-english>

- [11] G. M. Hyman, The calibration of trip distribution models, *Environment and Planning* 1 (1969) 105–12.
- [12] J. Nocedal, S. Wright, *Numerical Optimization*, 2nd Edition, New York: Springer-Verlag, 2006.
- [13] K. Levenberg, A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics* 2 (1944) 164–168.
- [14] M. Batty, *Urban Modelling*, Cambridge University Press, 1976.

Appendix A. Detailed mathematical description

A.1. Land use economical sectors

If we consider land-use sectors, many simplifications are done to the equations of the model. Production is estimated using equation (3.5). The technical coefficient a_i^{mn} represents the demanded amount of sector n by sector m in zone i and is described by the equation (3.6). For the land use sectors, the consumption costs \tilde{c}_i^n are equal to the price paid: $p_i^n + h_i^n$, where the prices p_i^n are known. Then equation (4.1) is as follows:

$$S_i^{mn} = \frac{W_i^n \exp(-\omega^{mn} a_i^{mn} (p_i^n + h_i^n))}{\sum_{l \in K^m} W_i^l \exp(-\omega^{ml} a_i^{ml} (p_i^n + h_i^n))}. \quad (\text{A.1})$$

In the following, we will note $U_i^{mn} = -\omega^{mn} a_i^{mn} (p_i^n + h_i^n)$.

Derivative estimation:

Let us consider m and m' as consuming sectors, n as land use sector and $q \in K^m$

$$\frac{\partial a_i^{mn}}{\partial h_i^q} = \begin{cases} -\delta^{mn} (\max^{mn} - \min^{mn})^{mn} e^{-\delta^{mn} (p_i^n + h_i^n)} & q = n \\ 0 & q \neq n \end{cases} \quad (\text{A.2})$$

The well known logit derivatives for S_i^{mn} are:

$$\frac{\partial S_i^{mn}}{\partial h_i^q} = \begin{cases} \frac{\partial U_i^{mn}}{\partial h_i^n} [S_i^{mn} - (S_i^{mn})^2] & q = n \\ \frac{\partial U_i^{mq}}{\partial h_i^q} S_i^{mn} S_i^{mq} & q \neq n \end{cases} \quad (\text{A.3})$$

if $q \neq n$, then $\frac{\partial U_i^{mn}}{\partial h_i^q} = 0$, then for $q = n$:

$$\begin{aligned} \frac{\partial U_i^{mn}}{\partial h_i^n} &= -\omega^{mn} \left[\frac{\partial a_i^{mn}}{\partial h_i^n} (p_i^n + h_i^n) + a_i^{mn} \right] \\ &= -\omega^{mn} [a_i^{mn} - \delta^{mn} (\max^{mn} - \min^{mn}) (p_i^n + h_i^n) e^{-\delta^{mn} (p_i^n + h_i^n)}] \end{aligned} \quad (\text{A.4})$$

With these results, we can compute the gradient of the production function exposed in (3.5):

$$\begin{aligned} \frac{\partial X_i^n}{\partial h_k^q} &= \partial \frac{\sum_m (X_{0i}^m + X_{0i}^{*m}) a_i^{mn} S_i^{mn}}{\partial h_k^q} \\ &= \sum_m (X_{0i}^m + X_{0i}^{*m}) \frac{\partial}{\partial h_k^q} [a_i^{mn} S_i^{mn}] \\ &= \sum_m (X_{0i}^m + X_{0i}^{*m}) \left[\frac{\partial a_i^{mn}}{\partial h_k^q} S_i^{mn} + a_i^{mn} \frac{\partial S_i^{mn}}{\partial h_k^q} \right] \end{aligned} \quad (\text{A.5})$$

We can do the same computations for ω and get:

$$\begin{aligned}\frac{\partial X_i^n}{\partial \omega^{m'q}} &= \sum_{m \in K^n} (X_{0i}^m + X_{0i}^{*m}) a_i^{mn} \frac{\partial S_i^{mn}}{\partial \omega^{m'q}} \\ &= (X_{0i}^{m'} + X_{0i}^{*m'}) a_i^{m'n} \frac{\partial S_i^{m'n}}{\partial \omega^{m'q}}\end{aligned}\quad (\text{A.6})$$

where:

$$\frac{\partial S_i^{mn}}{\partial \omega_{mq}} = \begin{cases} -a_i^{mn} (p_i^n + h_i^n) [S_i^{mn} - (S_i^{mn})^2] & q = n \\ a_i^{mq} (p_i^q + h_i^q) S_i^{mn} S_i^{mq} & q \neq n \end{cases}$$

replacing (A.7) in Equation (A.6):

$$\frac{\partial X_i^n}{\partial \omega^{m'q}} = \begin{cases} -(X_{0i}^m + X_{0i}^{*m}) (a_i^{mn})^2 (p_i^n + h_i^n) [S_i^{mn} - (S_i^{mn})^2] & q = n \\ (X_{0i}^m + X_{0i}^{*m}) a_i^{mn} a_i^{mq} (p_i^q + h_i^q) S_i^{mn} S_i^{mq} & q \neq n \end{cases}$$

A.2. Transportable Sectors

The location probability derivatives follow the well-known logit derivatives:

$$\frac{\partial P_{ij}^n}{\partial h_k^n} = \begin{cases} -\lambda^n \beta^n [P_{ij}^n - (P_{ij}^n)^2] & k = j \\ \lambda^n \beta^n P_{ij}^n P_{ik}^n & k \neq j \end{cases}\quad (\text{A.7})$$

If we consider the optimisation function exposed in (3.4), we need to compute derivatives for \hat{p} . From equation (2.9) the derivatives are as follows:

$$\begin{aligned}\frac{\partial \hat{p}_i^n}{\partial h_k^q} &= \sum_n \frac{\partial a_i^{mn}}{\partial h_k^q} S_i^{mn} \tilde{c}_i^n + a_i^{mn} \frac{\partial S_i^{mn}}{\partial h_k^q} \tilde{c}_i^n + a_i^{mn} S_i^{mn} \frac{\partial \tilde{c}_i^n}{\partial h_k^q} \\ &= \frac{\partial a_i^{mq}}{\partial h_k^q} S_i^{mq} \tilde{c}_i^n + \sum_n a_i^{mn} \frac{\partial S_i^{mn}}{\partial h_k^q} \tilde{c}_i^n + a_i^{mn} S_i^{mn} \frac{\partial \tilde{c}_i^n}{\partial h_k^q}\end{aligned}$$

and the derivatives for equation (2.8):

$$\frac{\partial \tilde{c}_i^n}{\partial h_k^q} = \begin{cases} \sum_j \frac{\partial P_{ij}^n}{\partial h_k^q} (p_j^n + tm_{ij}^n) & \text{if } q = n \text{ is transportable} \\ 0 & \text{else} \end{cases}$$

Since $\frac{\partial P_{ij}^n}{\partial h_k^q} = 0$ for $q \neq n$:

$$\frac{\partial \tilde{c}_i^n}{\partial h_k^n} = -\lambda^n \beta^n (P_{ik}^n - (P_{ik}^n)^2) (p_k^n + tm_{ik}^n) + \lambda^n \beta^n \sum_{j \neq k} P_{ij}^n P_{ik}^n (p_j^n + tm_{ij}^n)$$

The utility function is linear in the prices and the shadow prices, so the derivative estimation where the utility is involved are the same for h_k^q and p_k^q . In this case, we have to compute differently for the prices:

$$\frac{\partial \tilde{c}_i^n}{\partial p_k^n} = \begin{cases} P_{ik}^n + \sum_j \frac{\partial P_{ij}^n}{\partial p_k^n} (p_j^n + tm_{ij}^n) & \text{if } n \text{ is transportable} \\ 1 & \text{else} \end{cases}$$

Finally, the partial derivatives for the complete cost function, are as follows:

$$\begin{aligned}
\frac{\partial f}{\partial h_k^n} &= \frac{\partial}{\partial h_k^n} [\|X - X_0\|^2 + \|\hat{p} - p\|^2] \\
&= \frac{\partial}{\partial h_k^n} \left[\sum_n \sum_j (X_j^n - X_{j0}^n)^2 + \sum_n \sum_j (\hat{p}_j^n - p_j^n)^2 \right] \\
&= 2 \left[\sum_n \sum_j (X_j^n - X_{j0}^n) \frac{\partial X_j^n}{\partial h_k^n} + \sum_n \sum_j (\hat{p}_j^n - p_j^n) \frac{\partial p_j^n}{\partial h_k^n} \right]
\end{aligned}$$

Appendix B. Optimisation algorithm applied to the Example C

In this section we will present a step by step computation of Tranus equations and our optimisation algorithms applied to a simple scenario. This is the basic scenario for testing Tranus functionality and it is readily available from the Tranus website.

The scenario Example C has 5 economical sectors, and 3 geographical zones. A brief description of the respective economical sectors is presented in table B.3.

Table B.3: Example C: Economical sectors description

Number	Name	Type
1	Basic Employment	Exogenous
2	Service Employment	Transportable
3	Low Income Household	Transportable
4	High Income Household	Transportable
5	Land	Non transportable

There is only one land use sector and this model doesn't have substitution between sectors. Traditionally, the substitution in Tranus is only used between land use sectors. In table B.4 we present a summary of the demand functions (a_i^{mn}), the values for the different parameters are estimated externally. Only elastic demand functions are considered for land consumption. This is also a standard practice in Tranus modelling.

Table B.4: Demand functions parameters

m	n	Min	Max	Elast. (δ^{mn})
1	3	1.998969	1.998969	0.0
1	4	1.248126	1.248126	0.0
1	5	0.004	0.01	-7e-01
2	3	1.609238	1.609238	0.0
2	4	1.448615	1.448615	0.0
2	5	0.003	0.009	-8e-01
3	2	0.1203459	0.1203459	0.0
3	5	0.003	0.008	-7e-01
4	2	0.1532743	0.1532743	0.0
4	5	0.005	0.012	-6e-01

We will compute the shadow prices for land use sectors first (sector 5), and then we will compute shadow prices and prices for the transportable sectors (sectors 2, 3 and 4).

B.1. Land (sector 5) shadow prices calibration

As there is only one floorspace sector, we can easily write explicitly the equations corresponding to the production of this sector. For this model exogenous demand is zero for all sectors, so from equation (3.5) we have:

$$X_i^5 = \sum_{k=1}^4 a_i^{k5} S_i^{k5} \hat{X}_i^k, \quad i = 1, 2, 3,$$

here $\hat{X}_i^k = X_i^k + X_i^{*k}$ (induced and exogenous production). The following table presents the base year's data, reproducing this data is the goal of our calibration. For the land sector (sector 5) we have to fit the productions to 66, 110 and 128 respectively.

Table B.5: Base year's data

Sector	Zone	ExogProd (X_0^*)	InducedPro (X_0)	Price
1	1	5000	0	0
1	2	800	0	0
1	3	1100	0	0
2	1	0	3500	0
2	2	0	700	0
2	3	0	900	0
3	1	0	4000	0
3	2	0	13000	0
3	3	0	5000	0
4	1	0	1500	0
4	2	0	3000	0
4	3	0	11500	0
5	1	0	66	2.5
5	2	0	110	1.2
5	3	0	128	1.8

As one can see from Table B.5, a sector has Exogenous or Induced Production, but not both. So from now on, we will drop the *hat* from \hat{X} and just write X . So the calibration of sector 5 can be written as 3 optimisation problems, one per geographical zone:

$$\min_{h_i^5} \|X_i^5(h_i^5) - X_{0i}^5\|^2, \quad i = 1, 2, 3 \quad (\text{B.1})$$

with only one variable per problem, the corresponding shadow price. In fact, the quantity X_i^5 can be simplified, as there is only one land sector, there is no substitution, hence $S_i^{mn} = 1$:

$$X_i^5 = \sum_{k=1}^4 a_i^{k5} X_i^k, \quad i = 1, 2, 3,$$

So, for instance the explicit equation for the production of land in zone 1 comes from:

$$\begin{aligned} X_1^5 &= 1500 \left(0.005 + 0.008 e^{\frac{-3}{5}(h_1^5+2.5)} \right) \\ &+ 3500 \left(0.003 + 0.006 e^{\frac{-4}{5}(h_1^5+2.5)} \right) \\ &+ 4000 \left(0.003 + 0.005 e^{\frac{-7}{10}(h_1^5+2.5)} \right) \\ &+ 5000 \left(0.004 + 0.006 e^{\frac{-7}{10}(h_1^5+2.5)} \right) \end{aligned}$$

The cost function (B.1) can be plotted, see figure B.6.

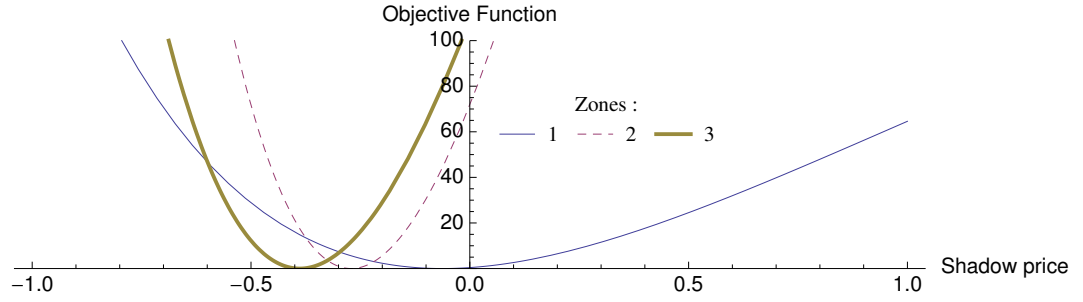


Figure B.6: Objective functions (B.1) near their optimal values $h^5 = (-0.062, -0.271, -0.388)$

B.1.1. Transportable sectors (2,3,4) shadow prices calibration

In this section we will present the methodology exposed in 3.3 to compute the prices and shadow prices of transportable sectors. As one can see from table B.5, we don't know the prices of sectors 1,2,3 and 4. Also, from table B.4 we can notice that for transportable sectors the demand functions are constant (or set to their min value). If we go back to the equations of how total demand is computed (2.2):

$$D_i^n = D_i^{*n} + \sum_m a_i^{mn} (X_{0i}^m + X_i^{*m})$$

we can completely compute demands as constants that do not depend on prices nor shadow prices, see table B.6.

Table B.6: Demands D_i^n per sector and zone.

Sector	Zone		
	1	2	3
2	711.29505	2024.3196	2364.38395
3	15627.178	2725.6418	3647.1801
4	11310.7825	2012.5313	2676.6921

From demands, we can compute productions X_j^n using equation (2.7) as follows:

$$X_j^n = \sum_i D_i^n Pr_{ij}^n$$

The logit Pr_{ij}^n comes from equation (2.4) and is a function of the location utility defined in equation (2.3): $U_{ij}^n = (p_j^n + h_j^n) + t_{ij}^n$. We will utilise the same notation as in section 3.3, let's define $\phi_j^n = h_j^n + p_j^n$. We will rewrite the utility function as $U_{ij}^n = \phi_j^n + t_{ij}^n$. The transportation disutilities t_{ij}^n come from the transportation module and they are given as input to the land use module (table B.7, left).

Table B.7: Left: (t_{ij}^n) Transportation disutilities per sector and pair of zones. Right: (tm_{ij}^n) Transportation costs per sector and pair of zones.

		Sector 2					Sector 2		
z/z		1	2	3	z/z	1	2	3	
1		0.386	1.286	1.555	1	0.377	1.255	1.365	
2		2.651	0.618	2.061	2	1.616	0.345	1.150	
3		2.742	2.411	0.723	3	1.593	1.051	0.315	
		Sector 3					Sector 3		
z/z		1	2	3	z/z	1	2	3	
1		0.564	1.879	1.898	1	0.133	0.680	0.444	
2		1.076	0.323	1.626	2	0.539	0.114	0.381	
3		1.121	1.462	0.336	3	0.423	0.443	0.127	
		Sector 4					Sector 4		
z/z		1	2	3	z/z	1	2	3	
1		0.765	3.693	2.550	1	0.567	1.889	2.343	
2		1.801	0.540	1.836	2	1.262	0.379	1.661	
3		1.220	2.438	0.366	3	1.829	1.459	0.438	

The location probabilities $Pr_{ij}^n(\phi_k^n : \forall k)$ are a function of our new variable ϕ , each location probability is dependent on the same sector and all zones' ϕ variable, this comes from the fact that Pr is a logit probability, and the denominator has all possible choices, hence, all other zones. We will denote $\phi^n = \{\phi_k^n\}_k$ the vector for all zones. The optimisation problem we must solve is as follow:

$$\min_{\phi^n} \|X_j^n(\phi^n) - X_{0j}^n\|^2 \quad n = 2, 3, 4 \quad (\text{B.2})$$

As for the non-transportable sectors, we have separated optimisation problems, but here instead, one per economical sector. However, these problems are larger, with as many variables as zones (3 for this example). The optimal solution is obtained using the Levenberg-Marquardt algorithm with analytic computation of the jacobian [13], the solution is presented in the following table (table B.8):

Table B.8: Optimal values of ϕ .

Sector	Zone		
	1	2	3
2	4.288	5.482	5.622
3	1.520	0.477	0.588
4	2.355	0.093	0.854

These values represent the sum of $p + h$, so to identify both parameters separately we must look at the equations that impose equilibrium in prices. These equations are

equation (2.8) and (2.9). First, in equation (2.8) we can replace X_{ij}^n/D_i^n by Pr_{ij}^n using equation (2.5). As we have found the optimal values of ϕ , the location probabilities are determined by these values. So prices have to solve the following linear system:

$$p_i^m = VA_i^m + \sum_n a_i^{mn} \sum_j Pr_{ij}^n (p_j^n + tm_{ij}^n) \quad (\text{B.3})$$

The coefficient tm_{ij}^n represent the monetary costs of traveling (these values come from the transportation module) and are presented in table B.7, right.

Finally we solve the linear system (B.3) to obtain the prices, values are shown in table B.9:

Table B.9: Equilibrium Prices

Sector	Zone		
	1	2	3
2	6.781	4.413	4.198
3	0.864	0.867	0.855
4	1.104	1.107	1.092

and the corresponding shadow prices are obtained subtracting the prices from ϕ . Table B.10 presents the final shadow price values, these values are centered (we subtracted the median per economical sector). Subtracting a constant from a logit does not change the probabilities.

Table B.10: Shadow Prices: absolute value (percentage of price)

Sector	Zone		
	1	2	3
2	-2.493 (-36.7%)	1.069 (24.2%)	1.424 (33.9%)
3	0.656 (75.9%)	-0.390 (-44.9%)	-0.266 (-31.1%)
4	1.251 (113.3%)	-1.014 (-91.5%)	-0.238 (-21.7%)

What we have presented here is the whole computation of equations to estimate the endogenous parameters called shadow prices in the land use and activity module.

Appendix C. Numerical aspects

Local optimisation may converge to local minima. We observed this in practice, depending on the setting of the parameter β (c.f. equation (2.4)) and on the starting point for the ϕ . An observation that seemed strange at first sight, was as follows. When estimating the ϕ for one sector, after convergence, the residuals of the cost function were all non-zero (besides for two zones for which observed production was zero). Further, all these residuals, besides for one zone, were exactly equal to one another and the residuals summed up to exactly zero. This seemingly strange behaviour has an explanation, as follows.

First, it must be noted that the sum of computed productions, does not depend on the values of the ϕ :

$$\sum_j X_j^n = \sum_j \sum_i D_i^n Pr_{ij}^n = \sum_i D_i^n \underbrace{\sum_j Pr_{ij}^n}_1 = \sum_i D_i^n$$

If the data are coherent, then the sum of computed productions must equal that of observed ones:

$$\sum_j X_j^n = \sum_j X_{0j}^n$$

Hence, the sum of residuals must be equal to zero, as was observed in practice.

The other issue concerned the fact that all non-zero residuals but one, were exactly equal to one another. This can be explained as follows. For one zone j , the value of ϕ_j was sufficiently large at some stage of the estimation, so that the computed probabilities Pr_{ij} effectively became equal to zero, for all i : the absolute value of the argument of the exponential $\exp(-\beta(\lambda\phi_j + t_{ij}))$ became so large that the exponential effectively got evaluated to zero. This in turn means that the computed production for that zone, also was computed as zero since

$$\sum_j X_j^n = \sum_j \sum_i D_i^n \underbrace{Pr_{ij}^n}_0 = 0$$

Hence, the residual for zone j is non-zero, and actually equal to the (negative of the) observed production X_{0j} . Since the sum of residuals over all zones must equal zero, as shown above, we must have:

$$\sum_{k \neq j} (X_k - X_{0k}) = -X_{0j}$$

Remember that the cost function to be minimised is the sum of squared residuals; as for the zones other than j , this means:

$$\min \sum_{k \neq j} (X_k - X_{0k})^2$$

It can be shown that given the constraint that the sum of residuals must equal a known value, the cost function is a minimum if that known value is equally apportioned to the residuals, i.e. if all the residuals are equal to that value, divided by the number of residuals:

$$\forall k \neq j : X_k - X_{0k} = -\frac{X_{0j}}{\sum_{k \neq j} 1}$$

This explains the observation made in practice, described above.

In practice, this problem can be avoided by choosing a different starting point for the optimisation algorithm.