Contributions to the calibration of integrated land use and transportation models

Thomas Capelle, Peter Sturm, Arthur Vidard, and Brian Morton

Abstract

The need for land use and transport **integrated** modelling (LUTI modelling) as a decision aid tool in urban planning, has become apparent. Instantiating such models on cities, requires a substantial data collection, model structuring and parameter estimation effort. This work is a partial effort towards the integrated calibration of LUTI models. It considers one of the most widely used LUTI models and softwares, Tranus. The usual calibration approach for Tranus is briefly reviewed, then the calibration of Tranus' land use module is reformulated as an optimisation problem, proposing a clear basis for future fully integrated calibration. We analyse the case of transportable and non-transportable economic sectors. 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. Finally, results of this methodology are presented for real world scenarios.

T. Capelle (Corresponding author) • P. Sturm • A. Vidard
Inria, Univ. Grenoble Alpes, LJK, F-38000
Grenoble, France
Email: thomas.capelle@inria.fr
Email: peter.sturm@inria.fr
Email: arthur.vidard@inria.fr

B. Morton
Center for Urban and Regional Studies, Univ. of North Carolina
Chapel Hill, USA
Email: bjmorton@email.unc.edu

1 Introduction

Integrated land use and transport modelling (LUTI modelling) has attracted the attention of researchers since 1960 (Wegener 2004). It is well known that the interaction of land use and transportation models creates complex non linear systems. Calibration of large-scale LUTI models is a challenging task. It is usually partitioned into a set of smaller, partial parameter estimation problems of individual components of a model, and an integrated calibration of the composite model, taking into account the mutual interactions between these components, is most often lacking.

In this work, we consider one of the most widely used LUTI models worldwide, Tranus, which comes together with an open-source software implementation[1]. Tranus is an equilibrium type model based on macro-economic principles and the interplay between offer and demand (of goods, workforce, housing, transportation, etc.) and prices. Instantiating the model on an urban area requires a calibration phase, which includes the estimation of several types of parameters. The goal is to reproduce as closely as possible, observations gathered on the studied area (socio-economic data, transport surveys, etc.). The usual calibration approach is semi-automatic. In an "outer loop", an expert user manually adjusts socio-economic parameters (for instance parameters expressing price-elasticity of the different modelled population groups to different goods such as housing). Given the current set of user-adjusted parameters, an "inner loop" then automatically computes adjustment coefficients (so-called shadow prices in Tranus) that essentially correct for un-modelled effects, by correcting model outputs such as to fit observed data better.

Our work addresses several shortcomings of this classical approach. First, concerning the inner loop, the estimation of shadow prices is performed by a greedy approach, which we replace by a proper formulation as an optimisation problem with an actual cost function. We then propose several efficient methods for solving this optimisation problem and further show how to decouple the entire estimation problem into an equivalent set of smaller problems. This is achieved by carefully investigating the dynamic system characteristics of Tranus and in particular by introducing auxiliary variables in the optimisation problem that allow to replace iterative computations by closed-form ones.

Second, we challenge the classical viewpoint adopted by Tranus and various other LUTI models, that calibration should lead to model parameters for which the model output perfectly fits observed data. This may indeed cause the risk of producing overfitting (as for Tranus, by using too many shadow price parameters), which will in turn undermine the models' predictive capabilities. We thus propose a model selection scheme that aims at achieving a good compromise between the complexity of the model (in our case, the number of shadow prices) and the goodness of fit of model outputs to observations. Our experiments show that at least two thirds of shadow prices may be dropped from the model while still giving a near perfect fit to observations.

[1] `http://www.tranus.com/tranus-english`

Third, our eventual goal is to couple the two main loops of the calibration process inside a single estimation process, with a coherent objective function, whereas now the two loops use different objectives. This should also reduce the manual part of the calibration process as much as possible. We have made first steps towards this goal by developing an integrated estimation scheme for the shadow prices and a subset of the outer loop's parameters. The scheme is shown to outperform calibration quality achieved by the classical approach, even when carried out by experts.

Finally, a recurring problem in the calibration of LUTI models is validation of estimated parameters. For most model parameters, it is impossible to obtain ground truth values, so the quality of parameters is most often assessed via the goodness of fit of model outputs to observations. This is in general suboptimal and the more so if a model is potentially over-parameterised. We thus propose a scheme akin to twin experiments in data assimilation that consists of generating synthetic observations that are as close as possible to real-world observations and to which the Tranus model can produce a perfect fit without any adjustment coefficients (shadow prices). In essence, we thus know the ground truth values (i.e., zero) of the shadow prices for these synthetic observations, and then use them to assess the accuracy of the calibration approach and its convergence properties.

The contribution outlined above are demonstrated on Tranus models and data from one metropolitan area in the USA.

2 Tranus land use module calibration

Tranus is a land use and transportation integrated model (LUTI), providing a framework for modelling land use and transportation 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 term sector is much more general than in the traditional concept. It may include the classical sectors in which the economy is divided (agriculture, manufacturing, mining, etc.), factors of production (capital, land and labor), population groups, employment, floorspace, land, energy, or any other that is relevant to the spatial system being represented. The **land use and activity module** of Tranus is responsible of simulating the complete spatial economic system, estimating the activities that locate in each zone and the interactions that they generate for a specific time period.

As for the **transportation module**, it estimates travel demand and assigns it to the transport supply, such that an equilibrium is reached. The two modules are linked together, serving both, as input and output to one another.

To attain an equilibrium status, Tranus runs both modules iteratively until convergence. The land use and transportation modules also need to reach their own 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 activity. This is done at current transportation costs and disutilities. The details about the land use equilibria are explained below. Second, the transportation module takes as input the transport demand output by the land use module and equilibrates the transportation network to satisfy the given demand. Both modules are ran iteratively until a general equilibrium status is found. This is achieved when neither land use nor transportation, evolve anymore.

2.1 The land use module

We present now the land use and activity model calibration process, the actual implementation in Tranus, and later our approach to this subject. In the activity model, there are two general types of sectors: economic sectors can be transportable or non-transportable. The main difference is that transportable sectors can be consumed in a different place from where they were produced. For instance, 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 must be consumed where it is produced. Transportable sectors generate fluxes, that induce transport demand. This translates in transportation costs, meanwhile, non-transportable sectors do not require transportation nor generate fluxes. Usually, three classes of economic sectors are considered: land and/or floorspace, households and businesses. Land/floorspace is usually split into two or three residential types (detached houses, apartments, mobile homes, etc.), and commercial floorspace into offices and stores. Households are usually classified by socio-economic level, based on income or the household composition. Businesses comprise industries (whose main output is exportation), services (schools, universities, recreational) and commerce. The consumption chain is as follows: Industry demands labor (households) and service businesses. Households also consume services, and services also need labor. 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 Lowry 1964). This process results in economic exchanges, sometimes inducing flux (transportable sectors) and sometimes place consumption (land). These exchanges induce prices for each economic sector.

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 and other technical coefficients. Let us introduce a few definitions:

- **Productions:** X_i^n expresses how many “items” of an economic sector n are present in a zone i.
- **Demands:** D_i^{mn} 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 located in zone i.

It is important to realise that the price in the case of land, correspond to the actual rent, whereas the “price” of a household is derived from the rent of the floorspace occupied by the household.

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 (Barra 1999). A graphical representation of this feedback is represented in 1. For instance, demand induces production and vice-versa. The iteration scheme is as follows. Prices of the current iteration are translated in intermediate variables (this will not be detailed here) which enables the computation of demand and consumption costs (noted as c in the scheme). This is done based on the current transportation costs and disutilities. Once demand and costs are known, the current production is computed and fed back to compute the new set of prices. This process is bottom-up, starting with land use prices and exogenous production and demand up to the production and prices of transportable sectors. All the above computations are repeated until convergence is attained in both, productions X and prices p (which guarantees convergence of all other variables).

We are only going to deal with a subset of model equations relevant to this paper. Demand is computed for all combinations of zone i, demanding (consuming) sector m and demanded sector n:

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

$$D_i^n = D_i^{*n} + \sum_m D_i^{mn} \tag{2.2}$$

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:

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

Here, 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:

$$Pr_{ij}^n = \frac{A_j^n exp\left(-\beta^n U_{ij}^n\right)}{\sum_h A_h^n exp\left(-\beta^n U_{ih}^n\right)} \quad . \tag{2.4}$$

Here, h ranges over all zones, 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.

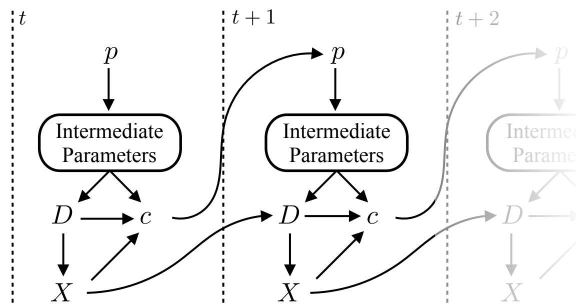

Fig. 1 Sketch of computations in the land use and activity module.

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

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

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

$$X_j^n = \sum_i X_{ij}^n = \sum_i D_i^n Pr_{ij}^n \quad . \tag{2.6}$$

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

$$\tilde{c}_i^n = \frac{\sum_j X_{ij}^n \left(p_j^n + tm_{ij}^n\right)}{D_i^n} \tag{2.7}$$

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

These finally determine the new prices:

$$p_i^m = VA_i^m + \sum_n a_i^{mn} S_i^{mn} \tilde{c}_i^n \tag{2.8}$$

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

The calibration process consists in adjusting the model parameters such as to reproduce observed base year 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.
ii. The adjustment parameters h_j^n of the utilities (2.3) (shadow prices).
iii. The remaining parameters (for example the penalisation factors and logit dispersion parameters)

After computing the external parameters [i], and giving an initial guess value to [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. Shadow prices "try" to compensate for the other parameters to reach a perfect fit – it acts as an adjustment term for the part of the utility that is not represented by the model. One would want to make the value 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 1 and the above equations):

$$h_i^{n,t+1} = (h_i^{n,t} + p_i^{n,t}) \frac{X_i^{n,t}}{X_{0,i}^n} - p_i^{n,t+1} \tag{2.9}$$

The shadow prices for the next iteration $t + 1$ increase proportionally to the ratio between computed and observed production in the previous iteration t.

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 automise 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 a 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 Reformulation as an optimisation problem

It's important to notice that a 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 base year data. To reformulate the calibration as an optimisation problem, we must compute the shadow prices that makes the productions as similar as possible to the observed productions. This can be achieved by the following optimisation problem:

$$\min_{h} \quad \|X(h) - X_0\|^2 \ . \tag{3.1}$$

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 1. The dependency of these on the shadow prices is visible from equations (2.3) to (2.6). Each evaluation of the productions $X(h)$ involves the convergence of the dynamic system exposed in figure 1. Each evaluation of the cost function involves the convergence of the dynamic system in productions and prices. This dual convergence can be avoided by including the prices in the set of optimised variables. Moreover, one can compute directly productions that are in equilibrium for a given set of shadow prices and prices. This is done by realising that the computation of demand and production involves a set of linear equations (2.1), (2.2), (2.5), and (2.6). If we re-organise these equations, knowing that only productions are needed in our cost function, we may only need to compute this. To do so, we substitute D_i^n in equation (2.5) using equations (2.1) and (2.2), giving:

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

Upon substituting this into (2.6), we obtain the following linear system in X_j^n:

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

By construction, the solution of this linear system corresponds to an equilibrium of both, productions and demands.

Computing the gradient of such a cost function (3.1) is still difficult. Each evaluation of the productions involves solving a linear system of the type (3.3), this means that the estimation of a numerical gradient by finite differencing is possible but costly. Second, although productions and demands computed as above are in equilibrium, the prices p may still evolve from the current to the next iteration. Hence, one still has to iterate model equations until prices converge.

Another observation can be made: we already have a base year production data set. If the calibration is successful, we would want to have the computed productions equal to the base year productions. Hence, we can simply impose this condition by evaluating the right hand side of(3.3) in the base year's productions. This approach enables us to compute the productions directly, without the need to solve a linear system. To address the second problem, we can add the prices to the parameters to be optimised. We use the current values of prices, and compare them against the model computed prices (2.8). The difference between the current prices and the ones computed by the model through equations (2.3) to (2.7) is added to our new optimisation function:

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

Here, $\hat{p}$ is the vector of prices computed by the model using (2.8) 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 directly. Further, it is no longer required to wait for the convergence of an 'inner loop", meaning that the cost function can be optimised by any non-linear least squares method; in our implementation we use the Levenberg-Marquardt method (Levenberg 1944).

Let us also note that other choices than the L_2 norm would of course be possible to define the cost function (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.

3.2 Land use sectors

For this particular type of sector, we can exploit the fact that land use (or floorspace) is only consumed and it doesn't consume other economical sectors – as a matter of fact, they are at the bottom of the consumption chain. Moreover, land use sectors are fixed in space (they are non-transportable), the amount of land is fixed and is consumed in place. As is standard practice for calibration, the prices p_i^n for the base year are known (they are fixed and considered as input and not the result of an equilibrium process). This translates in simplifications in the system of equations exposed above. First, the non-transportability makes the location probability (2.4) vanish, transforming equation (3.3) into:

$$X_i^n = D_i^{*n} + \sum_m \left(X_i^{*m} + X_i^m\right) a_i^{mn} S_i^{mn} \tag{3.5}$$

Being at the bottom of the consumption chain implies that the induced production of a given land use sector only depends on the other land use sectors variables. This is particularly interesting for the shadow prices: the induced production X_i^n is a function of the other land use sectors' shadow prices, but only within the same geographical zone i. These two last conditions make the estimation of the shadow prices of land use sectors much easier than for the rest of the economical sectors. Moreover, these are independent for each geographical zone i. So, the problem proposed in (3.4) is reduced to one separate optimisation problem for each geographical zone and without the price term. 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, see below.

3.3 Simultaneous estimation of shadow prices and land use substitution parameters

The sub model to compute the shadow prices of land use sectors, relies on the computation of equation (3.5). This sub model, has two parts to be estimated, first the technical coefficients a_i^{mn} and the substitution probabilities S_i^{mn}. The first, is generally estimated externally, using data from land use consumption per socio-economic category, giving normally good results. The latter, is much more complicated to estimate, because the substitution preferences refer to data that are not readily available. This substitution probability has broad uses, but usually represents the ability of households to choose between different types of land use, modelling the preferences of different households for different types of housing. For instance, rich people prefer detached housing, but could also live in apartments if these are well located. What we propose, is to calibrate the substitution model parameters at the same time as the land use sectors shadow prices. The approach is done in two steps: first we estimate a starting point with a logistic regression (Train 2009) over the substitution probability logit model, and then an optimisation layer to refine these parameters. The quality of the parameters will be quantified by the size of the corresponding shadow prices.

Tranus's substitution model for stock sectors (floor space and land): the basics. For ease of exposition, the land use sectors will be referred to as "floor space" (sector n), and the consumers of floor space will be referred to as "households" (sector m). In Tranus's substitution model, the proportion of sector m households in zone i that consume floor space sector n, S_i^{mn}, is given by the well known logit formulation (McFadden and Train 2000) with utility term $U_i^{mn} = -\omega^{mn} a_i^{mn} \tilde{c}_i^n$. The term a_i^{mn} is the average sector m household's consumption of sector n floor space; $\tilde{c}_i^n$ is the consumption cost of sector n floor space (per unit of floor space), see (2.7); ω^{mn} is the penalising factor, which is specific to both household sector m and floor space sector n; and the product $a_i^{mn}\ \tilde{c}_i^n$ may be interpreted literally as a household's expenditure on housing (say, per month).

1. **Phase 1: estimating parameters' initial values with multinomial logistic regression.** The substitution model's parameters are estimated with multinomial logistic regression (Train 2009). The data that are essential for estimation are household level observations on floor space consumption, housing expenditure, and the socio-economic 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 sep-

arately for each household sector, and they yield estimates of $-\omega^{mn}$ for each combination of floor space sector and household sector.

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 and observed productions 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:
$$f(h, \omega) = \|X(X_0, h, \omega) - X_0\|^2 \ . \tag{3.6}$$
We would like to find the values of ω that reduce the corresponding shadow prices. What we propose is to solve at first the following problem:
$$\min_{\omega \in \Omega} f(h = 0, \omega) \tag{3.7}$$
where Ω is a set of bounds over the penalising factors ω. We use 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 (3.7), then the final values for the shadow prices for the land use sectors are:
$$h^* = \arg\min_h f(h, w^*) \ .$$

4 Results

4.1 Generation of a Synthetic Scenario

The evaluation of a LUTI calibration is a difficult process, mainly due to the noise in the data and the fact that obtaining ground truth information is almost impossible. Our optimisation scheme needs as input the base years' productions and parameters $(X_0, parameters)$. Then, the calibration is done against this information. We could think of a model that does not need the shadow prices to attain a perfect fit, hence, create a synthetic scenario where the "perfect" fit is achieved with the shadow prices set to zero. To generate this "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 convergence in the prices, and compute the value of the induced production afterwards. This can be obtained with the following problem:
$$\min_p \quad \|\hat{p}(h, p) - p\|^2 \ . \tag{4.1}$$
After obtaining convergence in the prices, we compute the productions and set the synthetic base year productions to these values. This methodology produces a sce-

nario where the optimal value of the shadow prices is zero (by construction) and that reproduces the base years' productions perfectly. We could also set the shadow prices value to any other value than zero. This enables us to test our methodology and optimisation algorithms against a known ground truth.

4.2 Results of Estimation of Shadow Prices and Substitution Parameters

We applied this procedure to a real-scale LUTI model for North-Carolina, with 38 zones, 3 floorspace and 9 other economic sectors. Figure 2 shows the shadow prices for all zones and floorspace sectors, after the two phases of our process. 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 Tranus LUTI model).

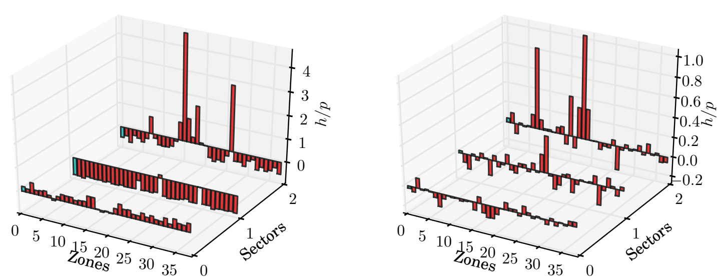

Fig. 2 Ratios of shadow prices and prices after phase 1 (left) and 2 (right). Note the different scales of the graphs.

4.3 Reducing the number of shadow prices

There are as many shadow prices as observations we are trying to fit. Having a perfect fit (near zero cost function) is the goal that most modellers try to achieve. Once this is achieved, reducing the value of the shadow prices (adjustment term) is the next goal. This is done by iteratively modifying the various parameters of the model until the desired level is reached. This may indeed cause the risk of producing overfitting (by using too many shadow price parameters), which will in turn undermine the models' predictive capabilities. We propose a different approach, consisting in sacrificing the "perfect fit", but using fewer shadow prices in the model. This is equivalent to supposing that in some specific sectors and zones, the model does not need the shadow price term to correct the utilities. So instead of having one shadow price per sector and zone (one per observed data item), we could find an "optimal" trade-off between model complexity and model fit.

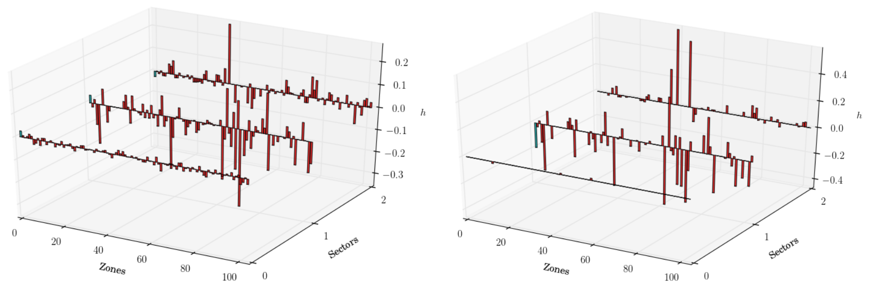

Fig. 3 100% of shadow prices of land use sectors - perfect fit (left) and 33% of shadow prices - within 3% fit (right). In the right-hand scenario, the shadow prices of floorspace sector 0 were excluded from the model in most zones.

A simple initial idea to do so, is as follows. First, we compute all shadow prices h^* based on (3.1). Then, we determine which shadow prices may be deleted from the set of model variables: this can be done by identifying the "small" shadow prices and fixing them to zero, after which we re-estimate the remaining shadow prices. We have tried this in the North-Carolina model, reducing the number of shadow prices to one third, after which the model still can be fitted to within 3% of the base year production, without a large increase in the values of the remaining shadow prices, as shown in Figure 3.

5 Conclusions and final remarks

The Tranus LUTI framework is a very powerful tool. The modelling possibilities are endless, and using LUTI evaluation should be the norm for urban and transportation planning. The complexity of these large scale models is something that must 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, and enables the use of the powerful optimisation algorithms currently available, solving 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 known and calibrated models, against our approach. This is not a straightforward conversion, because additional underlying details need to be taken into account. The proposed methodology for reducing the number of shadow prices needs additional fine tuning, but we consider the presented initial method as a first step in the right direction. We are convinced that the model "as it is" with one shadow price per observation is overfitted. Determining which shadow prices have to be removed cannot be completely automated, and the expert eye of the modeller has to have the last word. Finally, the simultaneous calibration of different parameter sets (here, shadow prices and substitution parameters), has the potential to be a powerful tool in practice. The results that we have for the North Carolina model 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).

References

Barra, T. de la (1999). *Mathematical description of TRANUS*. Tech. rep. Modelistica, Caracas, Venezuela. URL: http://www.tranus.com/tranus-english.

Levenberg, K. (1944). "A Method for the Solution of Certain Non-Linear Problems in Least Squares". In: *Quarterly of Applied Mathematics* 2, pp. 164–168.

Lowry, I.S. (1964). *A Model of Metropolis*. Tech. rep. Memorandum RM-4035-RC. Santa Monica, California: The RAND Corporation. URL: http://www.prgs.edu/content/dam/rand/pubs/research_memoranda/2006/RM4035.pdf.

McFadden, D. and K. Train (2000). "Mixed MNL Models for Discrete Response". In: *Journal of Applied Econometrics* 15, pp. 447–470.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Wegener, M. (2004). "Overview of Land-Use Transport Models". In: *Transport Geography and Spatial Systems*. Ed. by D. A. Hensher and K. Button. Handbook in Transport. Pergamon/Elsevier, pp. 127–146.