



**HAL**  
open science

## Simplified modelling of travel in low density areas

Wilfried Raballand, Pierre-Antoine Laharotte

► **To cite this version:**

Wilfried Raballand, Pierre-Antoine Laharotte. Simplified modelling of travel in low density areas. 47th European Transport Conference, Oct 2019, Dublin, Ireland. hal-02317375

**HAL Id: hal-02317375**

**<https://hal.science/hal-02317375>**

Submitted on 16 Oct 2019

**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.



## **SIMPLIFIED MODELLING OF TRAVEL IN LOW DENSITY AREAS**

Wilfried RABALLAND  
Pierre-Antoine LAHAROTTE  
Cerema Centre-Est, Bron (France)

### **1. INTRODUCTION AND OBJECTIVES**

At a time when institutions are involved with the issue of energy transition, communities with low density areas must equip themselves with tools to deal with the issue of travel needs and transport supply. In addition to environmental considerations, Schutz (2000) has identified a wide range of transportation planning needs in small communities, including safety, development, diversification and changing modes of transport.

Owing to lack of sufficient financial resources, these areas cannot collect the expensive data needed to develop a travel model according to the usual tested methods. It is therefore necessary to develop tools to propose "simplified" modelling. Such basic models have been developed since the 1970s. Ortúzar and Willumsen (2011) identify the usual methods that can be applied in such cases. More specifically, Bera and Rao (2011) have established a state of the art for models that use counting data.

As part of the study of a new road infrastructure begun in 2017 in the French town of Vesoul (35,000 inhabitants), a significant amount of data was collected, consisting of 70 automatic vehicle counting points, and 36 number plate reading (NPR) stations. This set of data enables to build an Origin-Destination (OD) matrix, which is assigned to the road network of the area. This "conventional model" based on observed flow data is used as a ground truth resource in the introduced results. Strong challenges related to the new infrastructure made this financial investment possible in terms of data and studies.

The present work uses the Vesoul case study to develop and test a simplification process to model the road trips in the area. The objective of the study lies in reducing the amount of data and assessing the performance impact. The models developed require a smaller amount of expensive data. This means that it is possible to propose a soberer, more systematic modelling approach to the local authority of the low density areas sheltering between 20000 and 50000 inhabitants. Furthermore this approach enables to spare time compared to usual studies, which is an additional saving on the cost of the assessment tool.

### **2. METHOD**

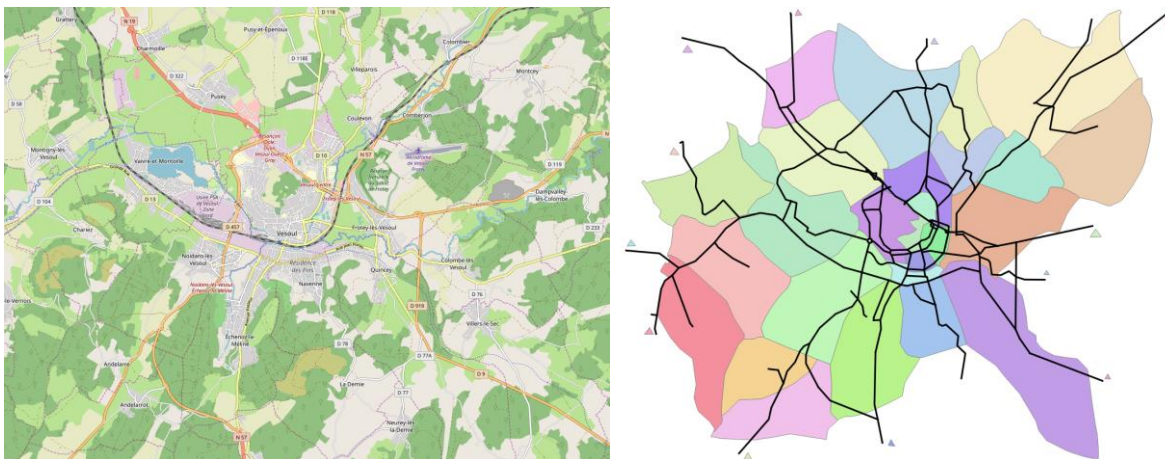
From the conventional model developed during the study begun in 2017, we keep the geographical features of the transport supply and demand:

- the road network on which the demand is assigned,
- the division of the area into zones which are either the destination or the starting point for trips.

Fig. 1 represents, on the left, a map of the urban area, and, on the right, the layout of the road network and the zoning of the model.

We propose a simplified model that uses the usual 4-step scheme: generation, distribution, modal choice, assignment. The study makes a focus on demand estimation, and on the category of light vehicles. The modal choice is considered as near-homogeneous for all OD relationships. This is because in this type of urban area, private car usage is largely predominant among all modes of transport, due to inefficient public transport, low traffic congestion and ease of parking. In addition, in order to focus on the first steps of the model, we will use the Vesoul road network which was already tuned during the path allocation stage, as well as the same commercial software to assess the simplified model.

Fig. 1: Map of the urban area of Vesoul (left). Network and zoning of the Vesoul model (right).



In order to reduce the use of expensive data, we have included in the model the free, public statistical data published by the French National Institute of Statistics and Economic Studies (INSEE), including population and employment data from the general census. These data help to explain people's demand for mobility. They make it possible to construct regression models and assumptions concerning the generation and the attractiveness of trips for each zone of the area, distinguishing between morning rush hour (MRH) and evening rush hour (ERH).

### A 3 stages process to reproduce OD matrices

The simplified model is built in three stages. Each corresponds to the treatment of one of the three types of trip that make up the demand: internal flows, exchange flows, and transit flows. At each stage, the work involves testing and validating the assumptions of the simplified generation-distribution model, using field data collected as part of the initial study. For the two stages concerning internal flows and exchange flows, we follow the usual method, which involves aggregating and standardizing the mobility behaviour of the users.

*Stage 1: reproducing the internal flows*

The first step, generation, is based on a simple expression of the emission of trips  $E_i$  as a linear combination of population  $Pop_i$  and employment  $Emp_i$  data from the internal area of origin  $i$  (see Eq. 1). This principle is validated by means of a linear regression. The second step, distribution, is based partly on a trip attractiveness function, and partly on a function of the location and/or distance of the destination zone. We begin by defining attractiveness  $A_j$  as a linear combination of population and employment data from the internal area of destination  $j$  (see Eq. 2). Next, we analyse the residual function  $f_i$  of Eq. 3 to estimate the plausible impact of the impedance  $z_{ij}$  between zones  $i$  and  $j$ , on the distributed volume of trips  $D_{ij}$ , as in the case of gravity models.  $f_i(z_{ij})$  is defined empirically, and  $K$  is the constant of proportionality of a doubly-constrained gravity model.

$$(1) \quad E_i = \gamma_{e/Pop} Pop_i + \gamma_{e/Emp} Emp_i$$

$$(2) \quad A_j = \gamma_{a/Pop} Pop_j + \gamma_{a/Emp} Emp_j$$

$$(3) \quad D_{i \rightarrow j} = K E_i A_j f_i(z_{ij})$$

#### Stage 2: reproducing the transit flows

For the third stage concerning transit flows, the process is different, because it is not possible to solve generation and attractiveness on the fictitious external zones simply and jointly. In order to replace the NPR data, it is necessary to define a model specific to transit that can be based on automatic counts only. We propose three assumptions (see Eq. 4,5 and 6), making it possible, respectively, to estimate:

- the isotropic fictive relationship between incoming exchange flows and outgoing exchange flows at rush hours,

$$(4) \quad \sum_j D_{r_k \rightarrow j}^{incoming\ exchange} = \tau_{MRH\ (or\ ERH)} \sum_i D_{i \rightarrow r_k}^{outgoing\ exchange}, \forall k \in [1, n]$$

For all  $n$  entry and exit points (as associated to the external fictive areas  $r_k, k \in [1, n]$ ), we have a relation of proportionality between the number of incoming exchange trips and the number of outgoing exchange trips, with a given coefficient  $\tau$  that depends of the rush hour.  $\tau$  can be expressed as:

$$\tau_{MRH\ (or\ ERH)} = \min_{k \in [1, n]} \left\{ \frac{\sum_j D_{r_k \rightarrow j}}{\sum_i D_{i \rightarrow r_k}} \right\}$$

where  $\sum_j D_{r_k \rightarrow j}$  is equal to the automatic count of incoming vehicles, and  $\sum_i D_{i \rightarrow r_k}$  is equal to the automatic count of outgoing vehicles at the entry and exit point  $k$ .

- the total volume of transit flows (passing through the urban area), is defined by the combination of the residual flows resulting from the difference between total flows and the fictive exchange flows defined in Eq. 4:

$$(5) \quad \sum_{r_1, r_2} D_{r_1 \rightarrow r_2}^{transit} = \sum_{k \in [1, n]} \left| \sum_i D_{i \rightarrow r_k} - \Phi \sum_j D_{r_k \rightarrow j} \right|, \Phi = \begin{cases} \tau_{MRH}, & \text{at MRH} \\ \tau_{ERH}^{-1}, & \text{at ERH} \end{cases}$$

The volume of transit flows is expressed as the anisotropic result of the difference between the total incoming flows and the incoming exchange flows.

- the distribution of transit flows between the various entry and exit points in the area.

$$(6) \quad D_{r_1 \rightarrow r_2}^{transit} = f_{transit}(\theta) \sum_i D_{i \rightarrow r_2} \sum_j D_{r_1 \rightarrow j}$$

If  $\theta$  (also noted  $\widehat{r_1 O r_2}$ ) is the angle formed by the entry point  $r_1$ , the center of the urban area O, and the exit point  $r_2$ ,  $f_{transit}(\theta)$  is a strictly increasing function from  $\theta=0$  to  $\theta=\pi$ .

#### Stage 3: reproducing the exchange flow

The exchange flow is defined in 2 steps:

- assessment of the total incoming (respectively outgoing) exchange flow for any crossing point  $r_k$ : this volume results from the difference between the counting volume and the combination of transit flows incoming (resp. outgoing) by  $r_k$ :  $\sum_j D_{r_k \rightarrow j} - \sum_i D_{r_k \rightarrow r_i}$ ;
- distribution of the total incoming / outgoing exchange flow on the OD relationship according to the emission and attractiveness of the internal areas.

At the end of each stage, we obtain a new travel demand. This new demand is then substituted within the global matrix that has been built from the field data. The resulting matrix is assigned with the same tool and the same settings used during the initial study.

The results of the simplified model in terms of traffic on the road network are evaluated on the basis of a comparison with the results of the conventional model, notably concerning:

- the reproduction of actual traffic observed on the entire road network,
- the estimation of simulated traffic shifts onto the new infrastructure project.

In addition, an inventory of the expensive data used and the corresponding budget required is drawn up in each phase. This assessment serves to quantify the savings made in technical and financial means compared to the initial data collection. It also illustrates the correlation between the amount of field data used and the quality of modelling the actual situation.

### 3. RESULTS

In the following, the results of the simplified model are compared either with data collected directly in the field, or with the demand tuned during use in the 2017 study.

#### *Emission and attractiveness assessment*

In the case of simplified modelling of internal flows and exchange flows, we apply the usual method to solve Eqs. 1 and 2. The variables explaining generation and attractiveness by zone that give the best results are as follows:

- the population for generation, and the sum (population + jobs) for attractiveness for the MRH,
- the sum (population + jobs) for generation, and also for attractiveness for the ERH.

Table 1: Validation by regression (coefficient  $R^2$ ) of the generation and attractiveness laws

	$E_i$ - Internal flows	$A_j$ - Internal flows	$E_i$ - Exchange flows	$A_j$ - Exchange flows
MRH	0.90	0.91	0.64	0.73
ERH	0.91	0.90	0.71	0.84

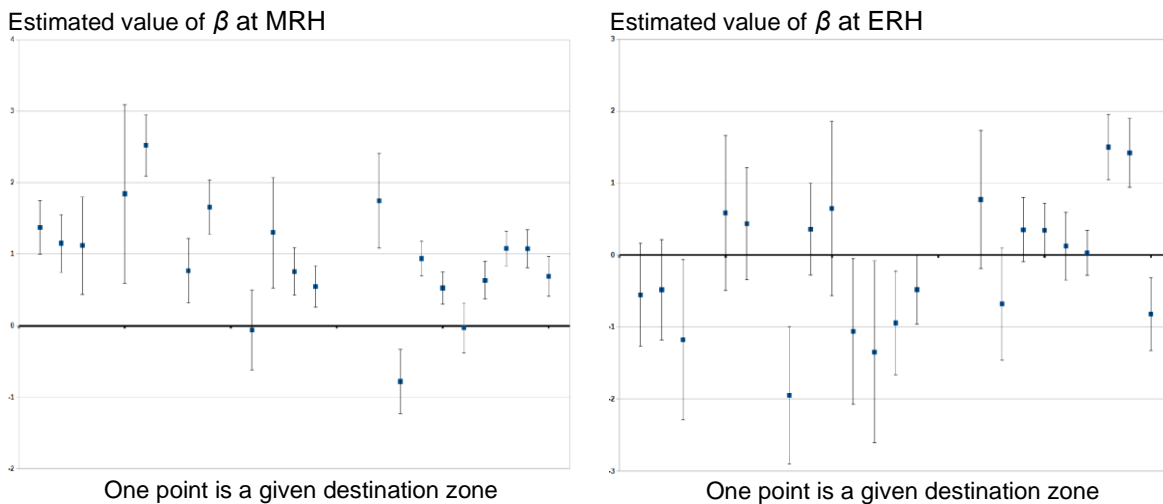
The generation and attractiveness models obtained are validated by linear regression (see Table 1). A satisfactory result overall is to be noted, with a lower quality for exchange traffic, and at the MRH.

Concerning any possible impact of the impedance between zone of origin and zone of destination on distribution, we propose testing a phenomenon of resistance to the distance travelled (see Eq. 7). Fig. 2 illustrates the resolution of coefficient  $\beta$  by regression. The results show that there is no significant resistance to distance, such that  $\beta$  is strictly negative. After an alternative analysis of the resistance to travel time, we obtain a comparable result.

$$(7) \quad f_l(z_{ij}) = d_{ij}^\beta$$

Therefore, as we are unable to resolve the dependence of distribution on remoteness, we propose to assume that  $f_l(z_{ij})=1$ .

Fig. 2: Determination of the coefficient  $\beta$  by linear regression on internal flows.





*Evaluation of the assumptions about the transit flows*

In the case of transit flows, we first examine the assumption that the ratio between incoming and outgoing exchange flows is isotropic (Eq. 4) for a given rush hour. Regression analysis gives us  $R^2=0.94$  for the MRH, and  $R^2=0.96$  for the ERH, and so validates this assumption.

Using the relation (Eq. 5), we estimate the total volume of transit flows on the basis of the anisotropic part of the traffic entering and leaving the area. Comparison between the survey data and the results of our model (Table 2) shows that the assumptions about volume estimation are acceptable.

Table 2: Validation of the estimation of the volume of transit flows

	Field survey	Simplified model	Absolute difference	Relative difference
MRH	477	526	+49	+10 %
ERH	543	549	+6	+1 %

The distribution of the transit movements between entries into and exits from the urban area is calculated using Eq. 6 and the following definition (Eq. 8) of the function  $f_{transit}(\psi)$ . A regression analysis of the real transit flows gives a satisfactory validation of this distribution function as it is defined:  $R^2=0.76$  for the MRH, and  $R^2=0.72$  for the ERH.

$$(8) \quad f_{transit}(\theta) = \begin{cases} 1 & \text{if } \theta > \pi/2 \\ \sqrt{2}/2 & \text{if } \theta = \pi/2 \\ \epsilon \ll 1 & \text{if } \theta < \pi/2 \end{cases}$$

*Global comparison*

For each development phase of the simplified model we can assess the performance of the simplified model by comparing the results obtained with the data collected in the field. We consider as comparison criterion:

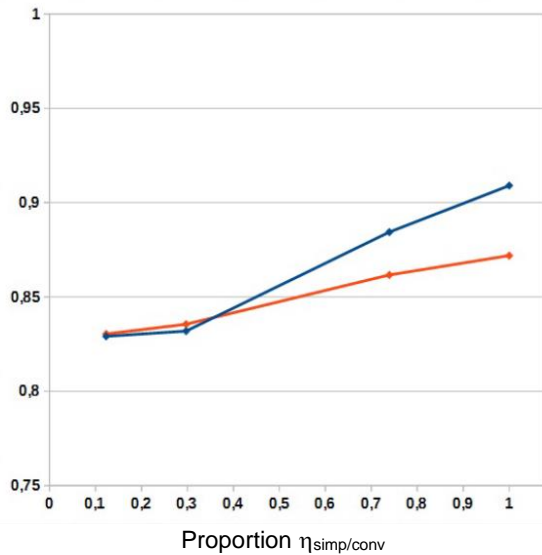
- aggregate indicators describing the correlation between observed traffic and modelled traffic:  $R^2$ , and more specifically for this type of data, the Geoffrey E. Havers (GEH) indicator (more accurately the value of the third quartile of all GEHs).
- the estimation of the volume of traffic during a working day on the new road infrastructure.

Table 3: Overall cost in expensive data of the different models

	Conventional model	Simplified internal mode	Simplified internal and exchange model	Overall simplified model
Amount	€41,400	€30,600	€12,300	€5,100
Proportion $\eta_{simp/conv}$	1	0.74	0.30	0.12

Fig. 3: Aggregate indicators of discrepancies between observed traffic and modelled traffic.

$R^2$  at MRH in blue, at ERH in red



3<sup>rd</sup> quartile of the GEHs at MRH in blue, at ERH in red

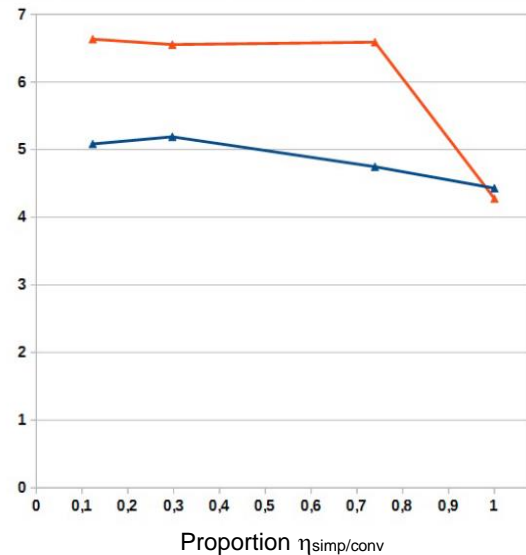
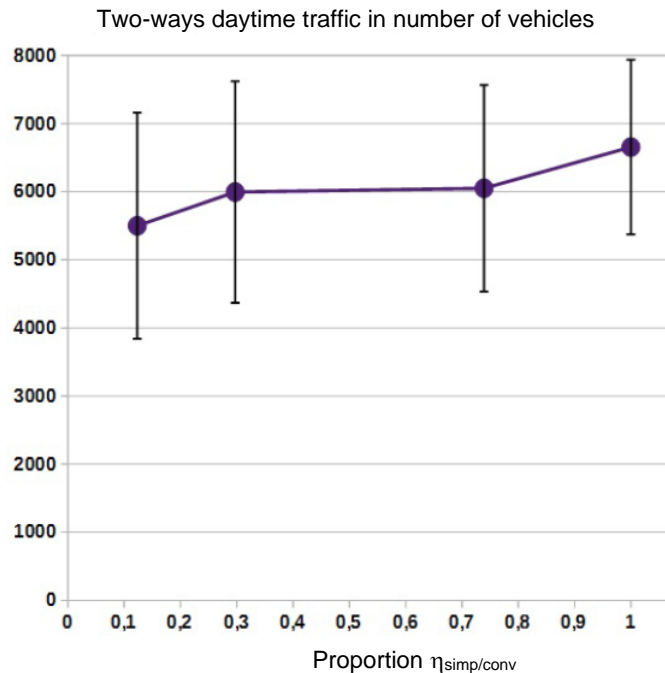


Fig. 4: Estimation of the volume of daytime traffic on the new road infrastructure. The margin of error shown opposite is calculated from the average absolute deviation between the observed data and the modelled data.



In order to more clearly situate one model in relation to the others, we plot any criteria according to the overall cost of the expensive data needed to develop it. These overall



costs are shown in Table 3, and they are the X axis of the following graphs. Fig. 3 gives the changes in the comparison criteria chosen according to the cost of the data used.

In these representations, moving from right to left, we reduce the use of expensive data and increase the use of the simplified model. We observe:

- for  $R^2$ : a moderate but homogeneous decline in data correlation,
- for the GEH: a significant, heterogeneous increase according to the rush hour.

Simplified modelling significantly affects the individual reproduction of traffic volumes as a whole. However, the divergence found is acceptable and does not presuppose by itself the capabilities of the tool developed. In particular, we can target the new infrastructure and analyse traffic estimates calculated over a working day (see Fig. 4). To within the margin of error, the estimated volume on the infrastructure does not vary significantly and retains the same order of magnitude. We can therefore conclude that the simplified model provides a comparable assessment of the project on this criterion. We can also provide a detailed inventory of the data needed for each modelling phase in terms of the quantity and types of collections used. Table 4 gives the comparison for all the models tested mentioned above. In addition to the savings achieved, the field survey is observed to be simplified and the system required less cumbersome.

Table 4: Number and types of data used according to the model examined

Type of data	Conventional model "Ground Truth Data"	Simplified internal mode	Simplified internal and exchange model	Overall simplified model
Vehicle counts within the perimeter	58	22	<10	<10
Cordon survey vehicle counts	12	12	12	12
NPR internally	22	22	-	-
Cordon survey NPR	12	12	12	-
INSEE population and employment	-	X	X	X

#### 4. CONCLUSION

The results show that modelling of travel in this type of area is not totally dependent on field data collections, and therefore on the budget available. In particular, there is a tested and validated alternative to demand estimation, which merges traffic data with public statistical data.

This work provides a development basis for modelling solutions in less urbanized areas. The Vesoul case study explored the potential and the limits of such a "simplified" modelling approach. Despite significant differences with results from a conventional model, the simplified model is a robust tool, simple to develop, and well suited for assistance with decision-making in low density areas. In particular, it can be applied in the study of the development of green and shared mobility in rural areas.



Such "turnkey" solutions can be replicated in similar areas, even though the approach is still exploratory and requires further experimentation. Ultimately, we plan to extend simplified modelling to the assignment stage by taking advantage of open source resources, particularly to avoid the cost of commercial software.

## 5. BIBLIOGRAPHY

Schutz, J.B., (2000) Transportation Planning Needs for Small and Medium-Sized Communities. In: Transportation in the New Millennium, TRB, National Research Council, Washington D.C.

Ortúzar, J.D., Willumsen, L.G., (2011) Modelling Transport. 4th Edition, Wiley, Hoboken. DOI: <http://dx.doi.org/10.1002/9781119993308>.

Bera, S., Rao, K.V.K., (2011) Estimation of origin-destination matrix from traffic counts: the state of the art. European Transport, 49, 2-23.