# A Data-Driven Routing and Scheduling Approach for Activity-Based Freight Transport Modelling

Ali Nadi[a,*], Neil Yorke-Smith[b], J.W.C Van Lint[a], Lóránt Tavasszy[a], Maaike Snelder[a,c]

[a] Department of Transport and Planning, TU Delft, Delft, The Netherlands
a.nadinajafabadi@tudelft.nl, J.W.C.vanLint@tudelft.nl, L.A.Tavasszy@tudelft.nl
[b] Electrical Engineering, Mathematics and Computer Science, TU Delft, Delft, The Netherlands
N.Yorke-Smith@tudelft.nl
[c] TNO, Den Haag, The Netherlands
M.Snelder@tudelft.nl
* Corresponding author

*Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)*
*June 19-25, 2022, Mauritius Island*

Keywords: Activity-Based Freight Modelling, Baysian Optimization, Data-Driven Vehicle Routing Problem

# 1 INTRODUCTION

Truck flow patterns can best be understood through the study of freight activities on a transportation network (Tavasszy and De Jong, 2013). These activities are often observable after the execution of tours resulting from a routing and scheduling optimization process, where the traditional Capacitated Vehicle Routing Problem (CVRP) is used to minimize multiple objectives under spatial, temporal, and vehicle constraints. Researchers have put forward in recent years that there should be distinctions between freight and passenger transport modelling due to the complexity associated with logistics (Gonzalez-Calderon and Holguín-Veras, 2019). The most important distinction is to involve the tour behaviour of carriers considering all spatial and temporal constraints (You et al., 2016). In current tour-based models, tours are constructed through incremental trip chaining in such a way that the next destination in a tour is estimated based on the conditional probability of the current stop. These types of models are based on choice modelling and provide statistical agencies with descriptive statistics and insights into freight demand. However, discrete choice methods are not subject to constraints and therefore cannot capture spatial-temporal characteristics of tours (Heinitz and Liedtke, 2010). Additionally, trip chaining decisions are made once at the tactical level and hence incremental reconstruction of the tour at the operational level is not identical to the tour planning process in reality.

To deal with this problem, micro-simulations have used a family of vehicle routing problems to model the pickup and delivery of carriers within a multi-agent microsimulation framework (Donnelly et al., 2010). Although normative models can perfectly capture space-time constraints, their outcome could deviate from observed tours due to heterogeneity in the tour planning decisions of planners. Hence, You et al. (2016) proposed an inverse optimization approach to calibrate a family of VRP using the method of successive averages. They estimate the weights of a weighted sum of multiple objectives from a set of observed tours. Parameter estimation of these methods requires fully observed truck movement patterns. However, tour data, if available, are often partially available to traffic agencies and policymakers due to privacy issues. To the best of our knowledge, the development of a method to calibrate the VRP model based on shipment flows with partially observed tour data has not yet been explored and is therefore of interest . To address this gap in the literature, we propose an efficient Bayesian optimization method to calibrate a VRP model based on partially observed tour information.

# 2 Activity-Based Freight Tour Modelling Framework

In this paper, we propose a disaggregate simulation-based optimization framework to model tour activities. In this framework, we begin with the available goods for pickup or delivery at each traffic analysis zone for each carrier. This could come from a freight generation model or available shipment flow data. Then, for all carriers, we simulate a parametrized capacitated vehicle routing problem so that all the available goods in the network are transported at the end of the planning horizon (e.g., one day). Afterwards, the planned tours are decomposed into trips and all trip characteristics are derived. The parameters of the model are updated through a surrogate-based optimization technique to minimize the difference between estimated and observed tour characteristics, such as the total tour travel times, travel costs, departure times, and arrival times. To solve this problem, we define a general approximation of the various type of VRPs with a weighted ($\beta$) sum of objectives $Z$ in Equation 1:

$$\min_{x} z_{\beta}\left(x_{ij}^{m}\right)=\min \sum_{l=1}^{L} \beta_l z_l=\beta^T Z \tag{1}$$

where $x_{ij}^{m}$ is the binary route decision variable where its elements are 1 if a vehicle visits location $j$ after $i$ in time interval $m$. Depending on the available data and the aim of the study, the modeller can define various objectives for the calibration. In this study, we only use capacity constraints for the VRP formulation. We adopted a compact form of a standard three-index formulation of the pickup and delivery problem proposed by Furtado et al. (2017). We formulate our objectives as follows:

$$z_1=\sum_{m\in M}\sum_{i\in V}\sum_{j\in V} t_{ij}^{m} x_{ij}^{m} \tag{2}$$

$$z_2=\sum_{m\in M}\sum_{i\in V}\sum_{j\in V} c_{ij} x_{ij}^{m} \tag{3}$$

$Z_1$ and $Z_2$ in Equations 2 and 3 are two cost functions that minimize the total tour travel time and distance-related costs respectively, and $Z_2$ and $Z_4$ are the disutility of a firm to visit a location and disutility of a specific commodity being transported between locations $i$ and j. All typical VRP constraints like capacity and conservation of flow and sub tour elimination constraints are used as proposed in the study of Furtado et al. (2017). For the calibration of this general approximation of CVRP, we solve the CVRP iteratively for each firm minimizing a loss function. This loss function includes the deviation between the estimated and observed tour cost, time, start time, and end time for all firms. This loss function is defined in Equation 4.

$$\min \sum_{\theta} J(\theta) = \min_{\theta} \tag{4}$$

Since the CVRP is computationally expensive and we must solve it repeatedly, a regular optimization algorithm cannot be easily used. Surrogate-based algorithms are a class of optimization methods that can find the minimum of functions that are expensive to evaluate. Such an algorithm works in the following steps:

1- Initialization: Generate $n$ initial $\Theta=\{\theta_1,\theta_2,\ldots,\theta_n\}$ parameters for the CVRP model of each carrier using Latin hypercube sampling. In our case, we started with three samples.
2- Evaluation: For each parameter setting, solve the $CVRP_{\Theta}$ and evaluate the loss function.
3- Surrogate construction: Fit a surrogate surface $f$ to the evaluated samples using a Multivariate Gaussian process with mean $\mu$ and covariance $\Sigma$.

$$f(\Theta)\sim N(\mu,\Sigma) \tag{5}$$

4- Search for minimum: Find the minimum of the surface interpolated for the sample of parameters.

$$\theta_{min}=\underset{\theta\in\Theta}{\operatorname{argmin}} f(\theta) \tag{6}$$

5- Next evaluation point: the next evaluation point $\theta^*$ in parameter space is where the expected improvement (EI) measure is maximum.

$$EI(\theta^*)=E[\max(\bar{f}(\theta_{min})-f^*(\theta),0)] \tag{7}$$

where $\bar{f}(\theta_{min})$ is the average stochastic prediction at point $\theta_{min}$

6- Update the surrogate surface: With the new parameter $\theta^*$, we update the covariance matrix and consequently the surrogate surface. The process iterates until the stopping criteria.

# 3 Results and findings from model calibration

In the Netherlands, the Central Bureau for Statistics (CBS) has digitally collected over 2.7 million records of shipments across the country for the year 2015. After data pre-processing, a set of 3,629 tour activities of the 10 largest carriers are selected for this study. All tours together visit, on average, 86 zones on a planning day. The loading/ unloading and handling time at each visiting location is proportionally distributed (based on weight and number of commodities) over visited locations and is randomly drawn from an exponential distribution. The parameters of this distribution are estimated from a sample of data. Three parameters are estimated for each feature based on the departure time of the trips, i.e., morning, noon, and afternoon. The model run time was 238 minutes for calibration and simulation of the tours. The surrogate optimizer stopped on the optimality gap after 20 iterations and 87 function evaluations. Table 1 shows the estimated parameters of the VRP model.

Table 1 –*Estimated parameters of the model for different departure time intervals*

| Param | Morning (6:00-11:00) | | Noon (11:00-15:00) | | Afternoon (15:00-20:00) | |
|---|---|---|---|---|---|---|
| | estimates | Std | estimates | std | estimates | std |
| $\theta_{\beta c}$ | 5.76 | 1.2e-3 | 4.293 | 4.38e-5 | 6.631 | 2.56e-2 |
| $\theta_{\beta t}$ | 0.098 | 1.4e-8 | 0.0819 | 3.29e-4 | 0.107 | 1.3e-2 |
| $\theta_{\beta UC}$ | -7.2 | 1.5e-6 | -7.66 | 1.3e-5 | -5.89 | 1.7e-4 |
| MAPE | 8.7% | | 4.8% | | 18.1% | |
| $R^2$ | 0.86 | | 0.91 | | 0.83 | |

The most salient findings go to the interpretation of cost ($\theta_{\beta c}$), time ($\theta_{\beta t}$), and capacity utilization of vehicles ($\theta_{\beta UC}$). Comparing tour preferences of carriers in different time intervals shows that travel time and cost are more important when carriers start a tour in the afternoon period. This means that carriers are likely to plan a shorter tour in distance with less total tour travel time. One possible reason for this is that carriers should reach the customers before the closing hours. On the contrary, the magnitude of the cost and time parameters for the noon period is relatively lower. This means that carriers are likely to make longer trips or plan tours with longer total tour travel times. Interestingly, the value of the capacity utilization parameter is also lower for the tours starting in the afternoon meaning that it is more likely to have less-than-truckload trips in the afternoon. The model estimates on average 42 minutes of (un)loading handling time using a bounded exponential distribution (minimum bound of 20 minutes). Drivers' break times are estimated together with the handling time. Our dataset does not report on the time and location of drivers' breaks but reports on the total tour duration. Therefore, the model can only capture these extra times along with the service time of each zone. The MAPE ranges between 8.7% and 18.1% with the maximum belonging to the afternoon mode and the correlation coefficient $R^2$ is 0.86, 0.91 and 0.83 for the morning, noon and afternoon respectively. These measurements confirm the overall good performance of the model.

The tour distance and number of stop distributions are two important aggregate features that should be captured well by tour models. We grouped tour distances into different clusters and compared the estimated tour distance distribution with the observations in Figure 1. The results show that the model can approximate this feature reasonably well. It is noticeable that the model generates local tours (distance <50km) more frequently as compared to the observations. Similarly, we can see that the model captures the general pattern in the frequency of tours with different numbers of stops. In general, it can be concluded that tours with a smaller number of stops are slightly overestimated.

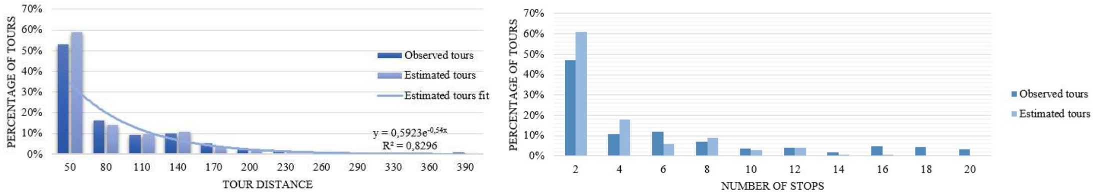


Figure 1 – *Estimated vs. observed tour distance and number of stops*

# 4 DISCUSSION

In this concluding section, we discuss the ability and limitations of the proposed model. Estimating a general tour model which can represent the collective tour activities of large carriers can become an undetermined problem especially when a limited imperfect amount of disaggregated data is available. Researchers have been dealing with these problems in transportation science for decades. The performance and accuracy of models in these cases depend largely on the amount of prior information used to help the model converge to the real behaviour of road users. Our proposed method shows that additional, although imperfect, tour information like tour duration, cost, start time, and end time can provide a robust and accurate estimation of a single tour model with meaningful parameters that can reproduce the preferences of carriers in planning tours. Developing such models with effective algorithms for robust estimation of model parameters is therefore essential due to the fact that perfect data are not always available to the public traffic sectors. This tool can also be used to understand the impact of freight activities on the traffic system. This is because the activities resulting from this model can be easily translated into a time-dependent truck OD table which can be coupled with a traffic simulation model. This could provide interesting insights into the interrelation between logistics and traffic. We also recommend a comparison study to examine the accuracy of our model among other tour models.

# Acknowledgments

We thank the TRISTAN XI reviewers. This work was supported by the Dutch Research Council (NWO), TKI Dinalog, Commit2Data, Port of Rotterdam, SmartPort, Portbase, TLN, Deltalinqs, Rijkswaterstaat, and TNO under project 'ToGRIP-Grip on Freight Trips' with grant number 628.009.001, and also partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation program under grant number 952215. The authors thank CBS and Rijkswaterstaat for providing shipment data and travel time matrices, respectively.

# References

Donnelly, R., Wigan, M. & Thompson, R. 2010. A hybrid microsimulation model of urban freight travel demand. *Transportation Research Board*.

Furtado, M. G. S., Munari, P. & Morabito, R. 2017. Pickup and delivery problem with time windows: a new compact two-index formulation. *Operations Research Letters,* 45**,** 334-341.

Gonzalez-Calderon, C. A. & Holguín-Veras, J. 2019. Entropy-based freight tour synthesis and the role of traffic count sampling. *Transportation Research Part E: Logistics and Transportation Review,* 121**,** 63-83.

Heinitz, F. M. & Liedtke, G. T. 2010. Principles of Constraint–Consistent Activity-Based Transport Modeling. *Computer-Aided Civil and Infrastructure Engineering,* 25**,** 101-115.

You, S. I., Chow, J. Y. & Ritchie, S. G. 2016. Inverse vehicle routing for activity-based urban freight forecast modeling and city logistics. *Transportmetrica A: Transport Science,* 12**,** 650-673.

TAVASSZY, L. & DE JONG, G. 2013. *Modelling freight transport*, Elsevier.