The Netherlands Research School for Transport, Infrastructure and Logistics
TRAIL

11th TRAIL Congress
November 2010

BENCHMARK FRAMEWORK FOR SENSITIVITY ANALYSIS OF DYNAMIC OD DEMAND ESTIMATION METHODS

Tamara Djukic MSc, Dr. ir. Hans van Lint, Prof. dr. ir. Serge Hoogendoorn, Prof. dr. Henk van Zuylen
Faculty of Civil Engineering and Geosciences, Department of Transport and Planning, Delft University of Technology, the Netherlands

ABSTRACT

The indicator for the quality of an OD estimation method is the sensitivity of the method to (and robustness against) random and structural perturbations of the input (data from sensors, prior OD matrices) on a few typical test networks. In this paper we propose such an assessment methodology based on the Latin Hypercube (LHC) method, which is an efficient alternative for Monte Carlo sampling and particularly suited for high-dimensional estimation problems. We demonstrate the methodology on a real urban corridor network for one well-known OD estimation method (the entropy maximization estimation method) to illustrate which results can be obtained and how these can be used to benchmark different OD estimation methods.

KEYWORDS

Benchmark methodology, Dynamic OD demand estimation, Latin hypercube sampling

INTRODUCTION

In recent years a large number of methods have been developed to estimate dynamic origin destination matrices from traffic counts and other available data. Although many methods have been proposed to solve the (dynamic) OD estimation problem, much fewer efforts have been reported to actually evaluate and cross-compare these methods under different circumstances (e.g. different network structures, different sets of data available in different qualities). Few studies have focused on evaluation of the reliability and accuracy of the estimated OD matrices in absence of ground truth OD matrix *(1)*, and on available ground truth OD matrix such as *(2, 3)*. The reason why it is difficult to cross compare and generalize results from sensitivity analysis undertaken to demonstrate different properties of the methods is that these methods differ in many aspects.

A structured and generic methodology to benchmark different OD estimation methods under different circumstances (e.g. network lay-out, data availability and quality of prior OD) would enable researchers to pinpoint the strengths and weaknesses of different OD estimation methods and their applicability and validity under different circumstances. This paper presents and demonstrates such a generic OD estimation benchmarking framework. It is generic in the sense that a wide range of OD estimation approaches can be tested under a wide range of different circumstances related to for example data availability and quality, and network lay out. One of the central components is an efficient Monte Carlo sampling method, the so called latin hypercube (LHC) method.
In the first part of the paper we will outline the overall framework and discuss its application to the OD estimation problem with various scenarios that are relevant for assessing different OD estimation methods. In the second part of this paper we will demonstrate this evaluation framework for entropy maximization estimation method on real urban corridor network under different scenarios of data quality and availability. The paper closes with a discussion of the results and conclusions.

BENCHMARKING FRAMEWORK

As reflected by equation (1), a particular OD estimation method *h* with a set of parameters and assumptions *H*, aims to infer the maximum probable OD matrix given the set of available inputs (data *D*, prior and previous OD estimates T_{prior} and $T_{t\text{-}1}$, $T_{t\text{-}2},\ldots$) and all the assumptions *H* used in the method, that is

$$T_t = h\left(T_{t-1}, T_{t-2}, \ldots, T_{prior}, D, H\right) \qquad (1)$$

Typically *h* is a highly nonlinear and dynamic function of its inputs. Since OD estimation methods differ in many aspects (objective function, optimization method, single or bi-level, etc) and since they may be assessed under many different circumstances (network type with or without route choice, different traffic conditions, varying data availability, etc) the number of simulations required for an exhaustive and statistically sound comparison between different OD estimation methods will quickly become very large.
Since the framework has be applicable to a whole range of OD estimation approaches, we consider a simulation-based approach where the OD estimator is considered as black box, providing a certain outcome (OD estimate) given certain input. The obvious choice for a simulation-based approach is to use some sort of constrained or stratified form of Monte Carlo sampling *(4)*, which does not scale with the input dimensionality. We applied the stratified sampling method (Latin hypercube (LHC) method *(5)*) that provides a computationally much more efficient alternative to random (Monte Carlo) sampling. We refer to reader to the original papers on LHC method *(5, 6)* for more details.
Given such an efficient sampling method exists, the key properties of a benchmarking framework for OD estimators are the following:

1. It needs to sample / vary over all the relevant inputs, i.e.:
    a. Network topology, considering a number of typical networks (e.g. a freeway corridor without route choice, a freeway network with a few dominant routes and a grid network), for which ground truth OD data are available
    b. Traffic conditions, considering various degrees of congestion
    c. Data availability and quality, considering many possible data scenarios
        i. availability/quality of local sensors (vehicle counts)
        ii. availability/quality of probe vehicle data / travel reports

    iii. availability / quality of a priori OD matrices

2. It must be able to assess and compare either a single or multiple OD estimators, and possibly sample different parameter settings for each method
3. It needs to be able to generate all relevant (conditional) output distributions from the results on the basis of which a wide range of performance indicators can be derived. We identify three levels of aggregation:
    a. The total OD demand (summed over all OD flows). Clearly, OD methods should at least be robust in deriving this quantity (as a function of departure time) even in cases where little data (or poor quality) is available
    b. The OD demand summed over particular destinations (or origins). Here not just the total amount, but also the distribution over the destinations may be assessed
    c. OD demand per origin destination pair, i.e. the entire OD matrix. In this case particularly the preservation of the matrix structure (i.e. the dominant OD relations versus the near zero ones) is of interest.

Requirement 1 essentially determines which input scenarios, varying in terms of network topology, traffic conditions, and data availability, need to be considered during the assessment. Requirement 2 describes which methods and variations of methods are to be assessed and compared. Requirement 3 describes the output required for assessment, cross-comparison and analysis.

EXPERIMENTAL SET UP

In this section we will describe the input data and assessment scenarios used in this paper. This case study has a purpose to demonstrate the main features of the proposed assessment framework for one particular OD estimation method. A benchmarking exercise of several OD estimators will be the subject of further studies. The entropy maximization model proposed by *(7)* is applied in this study. We use the solution algorithm based on an advanced iterative algebraic reconstruction method, referred to as the Revised Multiplicative Algebraic Reconstruction Technique (RMART), as proposed by *(8)*. For more detail explanation of this algorithm we refer to the papers *(8)*.

By implementing the LHC method described in previous section, we can simulate the fact that $\tilde{d}$ may contain errors. We consider the following three scenarios:

1. *Scaled (total) demand variation scenario*: This scenario illustrates how the variation of demand flows across the sequence of departure intervals can influence the estimation results without changes in OD matrix pattern (the all demand values within the same departure interval are scaled with a factor u_τ);
2. *Random demand variation scenario*: This scenario is based on the assumption that the prior OD matrix is the best estimate of the mean of the dynamic OD matrices. In this case $\tilde{d}$ is varied by adding uniformly random components to the ground truth OD matrix, representing the difference between the smoothed historical estimate and the particular daily realization;
3. *Systematic demand variation scenario:* This scenario addresses situations where the prior OD demand might contain other, structural errors besides the random daily fluctuations. The demand per each origin over destinations is generated from positively and negatively skewed mean values of distribution from random demand scenario.

For a comprehensive benchmark study it is important to take into account the real state and conditions (in terms of output, errors, location, etc.) of the detectors on the network. Therefore we assume the random perturbations of traffic counts measurements using the following two detector layouts:

1. *Detector layout based on maximum coverage of all links in network:* In this "best" case we want to obtain the benchmark results for comparison of OD estimates with other detector layouts.
2. *Detector layout based on the maximum link flows*: The choice of link traffic detectors is according to the maximum link flow measurements such that links with highest volume have priority. First, we identified the links with highest volume. Then detector layout is defined randomly removing the detectors from links to obtain tree scenarios with 80%, 60% and 40% coverage of total number of links on network.

The true link measurements are randomly perturbed following the same approach proposed in random demand variation scenario.

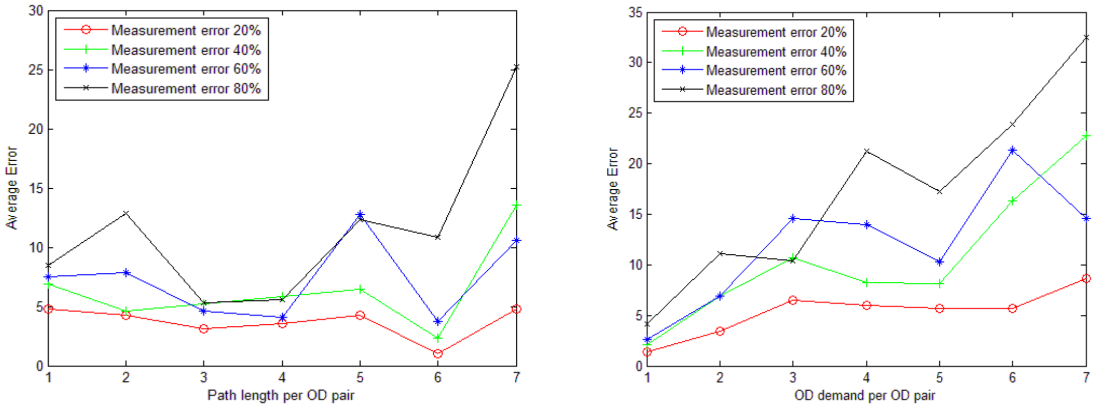
We evaluate the performance of the maximum entropy method on a real corridor network. This network was chosen because of the availability and quality of the empirical detector data on network, and because a calibrated OD matrix was available. The network consists of 214 OD pairs and 118 corresponding links. The evening peak hour reflecting the congested state at the network which was divided in 6 departure time intervals of 10 minutes is chosen for experiments.

RESULTS AND DISCUSSION

We use two criteria to check whether estimated OD matrices from the entropy method are on average equal to the true value, or show at least only a small variation around this value. The first criterion relates to the sensitivity of the total demand estimates under different scenarios of counting errors and errors in the prior OD matrix. Secondly we look at the sensitivity of certain OD pairs to the same error scenarios.

The results provided new insights into the sensitivity of EM estimation method with respect to the scale in the prior OD matrix. We found that errors in link measurements lead to the larger range of errors resulting in underestimation and overestimation of total demand and per total origin demand as well. The results finally also showed that the use of less accurate link measurements can dominate the accuracy of entropy method especially when the link flow measurements availability is limited.

As we were expected, in all tested scenarios we observed that the minimum absolute error for some OD pairs is close to zero while the maximum error is usually over 35%. It was thus interesting to identify the properties of the paths and magnitude of the demand associated with such 'problematic' OD pairs. These data are presented in Figure 1a (influence of path length) and 1b (influence of magnitude of the demand).



a) b)

Figure 1: The average errors a) path length per OD pair and b) value of OD demand

From Figure 1 we observe that the entropy method furthermore turned out to be sensitive to path lengths and the demand per OD pair. As a result, it can be concluded that it is harder to estimate OD pairs that have longer paths or higher demand accurately, since these showed larger error in the link measurements.

CONCLUSIONS

This paper proposes a new framework to assess and cross-compare the sensitivity of OD estimation methods to different types of uncertainty in the input data, and the parameters of the OD estimation methods. Central to this benchmarking framework is a well-known simulation-based stratified sampling technique that allows assessing the sensitivity without the need to perform an unfeasible number of simulations. The contribution of the proposed benchmark methodology is to enhance the understanding in how the abovementioned aspects impact dynamic OD estimation methods in terms of robustness and accuracy. Using this benchmark framework, both researchers and practitioners will gain insight in which OD estimation methods are suitable for both off-line applications, such as ex-ante simulation studies, and on-line applications, such as short-term traffic prediction, decision support systems, etc.

REFERENCES

Yang, H., Y. Iida, T. Sasaki (1991) An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts, in: *Transportation Research Part B: Methodological 25(5),* pp. 351-363.

Marzano, V., A. Papola, F. Simonelli (2009) Limits and perspectives of effective O-D matrix correction using traffic counts, in: *Transportation Research Part C: Emerging Technologies, 2009. 17(2)*, pp. 120-132.

Stathopoulos, A., T. Tsekeris (2003) Framework for analysing reliability and information degradation of demand matrices in extended transport networks, in: *Transport Reviews: A Transnational Transdisciplinary Journal, 23(1),* pp. 89 - 103.

Reuven, Y.R., (1981) *Simulation and the Monte Carlo Method*, John Wiley \& Sons, Inc. 304.

McKay, M.D., R.J. Beckman, W.J. Conover (1979) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, in:. *Technometrics, 21(2)*, pp. 239-245.

Budiman, M., B.M. Alex, (2006) A conditioned Latin hypercube method for sampling in the presence of ancillary information, in: *Comput. Geosci., 32(9)*, pp. 1378-1388.

Van Zuylen, H.J., L.G. Willumsen (1980) The most likely trip matrix estimated from traffic counts, in: *Transportation Research Part B: Methodological*, 14(3), pp. 281-293.

Wu, J. (1997) A real-time origin-destination matrix updating algorithm for on-line applications, in: *Transportation Research Part B: Methodological, 31(5)*, pp. 381-396.