DYNAMIC ORIGIN-DESTINATION ESTIMATION IN CONGESTED NETWORKS

Rodric Frederix: Katholieke Universiteit Leuven, Belgium rodric.frederix@cib.kuleuven.be
Chris M.J. Tampère: Katholieke Universiteit Leuven, Belgium chris.tampere@cib.kuleuven.be
Francesco Viti: Katholieke Universiteit Leuven, Belgium francesco.viti@cib.kuleuven.be

Abstract

In this paper three conditions for unbiased origin-destination (OD) estimation in congested networks are presented. A first condition deals with the choice of the network loading model. It is shown that an incorrect representation of queuing leads to an incorrect interpretation of information from the detectors, and thus leads to biased results. A second aspect is the relationship that is used in the minimization problem. In many OD estimation methods a linear relationship expressed by the assignment matrix is chosen. However, this relationship assumes separability of the link flows to the OD flows, which can lead to biased results when dealing with congested networks. A final source of error is the presence of multiple local minima in the objective function. An important cause of local minima is an incorrect interpretation of the information from the detectors. Since flow measurements can correspond to two distinguished traffic regimes, free flow and congestion, when the assigned initial OD matrix reproduces a different traffic regime at the detectors compared to reality, the measurements can be interpreted incorrectly. Finally a case study is presented that illustrates the bias that may be introduced if not all three conditions are met.

1. Introduction

The dynamic OD estimation problem has been tackled by many researchers in the past decades. Their studies focus on a variety of topics, dealing with the stochasticity and underdeterminedness of the problem, solution algorithms, the use of different data sources, and so forth. For a well detailed overview we refer to Lindveld (2003) and Balakrishna (2006). In this study we focus on the assignment-based dynamic OD estimation problem using traffic counts, which is a rather traditional approach to this problem. Key issue in the estimation of an OD matrix from traffic counts is the identification of the origin-destination flows that reproduce the observed traffic data. The assignment-based dynamic OD estimation problem is typically expressed as a bi-level problem:

$$\begin{aligned} x^* &= \arg\min_x \left[z_1\left(x,\hat{x}\right) + z_2\left(y(x),\hat{y}\right) \right] \\ y &= A(x)x \\ &= B(x)P(x)x \end{aligned} \tag{1}$$

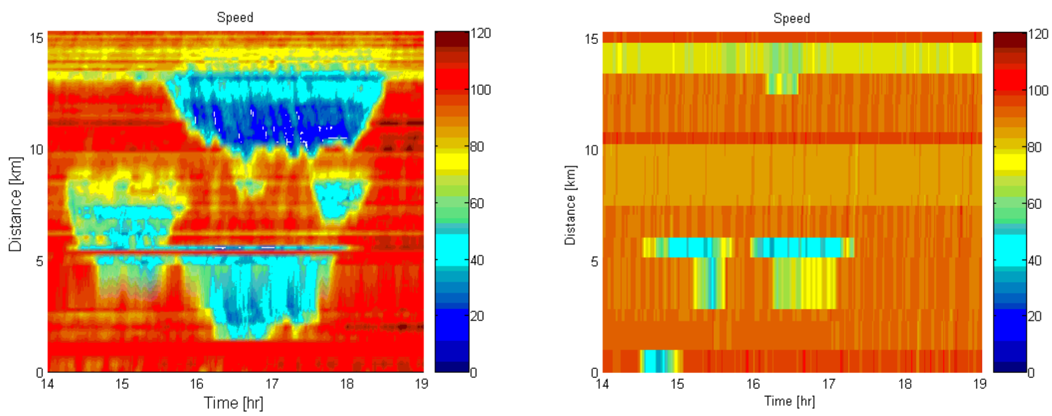
where z_1 and z_2 are distance measures , x is the unknown OD vector containing n elements, x* is the estimated OD vector and $\hat{x}$ is the target OD vector, $\hat{y}$ is a vector with traffic counts containing m elements, and y is a vector with the corresponding estimated link flows. The relationship between the OD flows and the link flows is described by the assignment matrix A, which has dimensions $m\times n$. This matrix can be further subdivided into a crossing fraction matrix B and a route fraction matrix P (see Cascetta (2001)). The elements of crossing fraction matrix B express the proportion of a route flow that passes a link, thus describing the spatio-temporal propagation of the route flows throughout the network. The elements of route fraction matrix P express the proportion an OD flow choosing a certain route. Both are obtained from the lower level, in which a dynamic OD matrix from the upper level is assigned using a Dynamic Network Loading (DNL) through a Dynamic Traffic Assignment (DTA) model. The output of this model is used to derive a relationship between the measurements and the OD flows (in general a linear relationship is used). This relationship is used in the minimization problem of the upper level to obtain a new estimate of the OD matrix. This estimate is then transferred back to the lower level to maintain consistency between the two levels. This process repeats until convergence is reached.

A number of factors affect the reliability of the estimation results. An important aspect is the effect of congestion on the estimation process. Congestion causes the relationship between the link flows and the OD flows to become highly non-linear, because of spillback and rerouting effects. This is expressed by the dependence of the assignment matrix on the OD flows. The effect of this non-linearity will be further discussed in section 2.

Many estimation methods proposed in literature do not perform sufficiently well on congested networks as they do not model appropriately the above dynamic effects. This will be shown in section 2 of this paper through some synthetic examples. Furthermore the non-linearity makes the problem non-convex, and it is possible that the estimated OD matrix is a local minimum and not a global minimum of the goal function. In many cases these local minima result in a congestion pattern that differs strongly from reality. An example of this deviation can be seen in figure 1. This figure was obtained by performing an OD estimation on the ring way around Antwerp using only flow measurements. The left picture in figure 1 is the real speed contour plot, while the right picture is the speed contour plot that is obtained by assigning the estimated OD matrix. As one can see the congestion pattern of the estimated OD matrix deviates strongly from the actual congestion pattern. Apparently the OD estimation process converged to a local minimum.

To reduce the risk of local minima many papers stress the importance of having a good initial matrix that should not deviate too much from the actual OD matrix. Very often this is referred to as the OD adjustment process. However, only in rare cases the available initial OD matrix yields to a reliable adjusted matrix. In many cases an initial OD matrix is not available at all, or it is calculated for rather different scenarios. Therefore, in practice we need good estimation models especially for cases in which we do not have a good initial OD matrix.

Congestion can thus lead to poor-quality estimates of the dynamic OD matrix. These errors are transferred onto the network flows, leading to errors in network performance measures. Since practitioners are often confronted with highly congested networks and adopt traffic models that need to be calibrated in order to be used in traffic management systems or for planning purposes, there is an urgent need for a practical OD estimation methodology that is applicable to heavily congested networks.



(a) **(b)**

Figure 1: Real traffic pattern (a) and traffic pattern with estimated OD matrix (b)

As can be deduced from the assignment relationship in equation 1 there are two sources of non-linearity due to congestion: one because of the dependence of the crossing fraction matrix B on the OD flows, the other one because of the dependence of the route fraction matrix P on the OD flows. In the present paper we only consider networks without route choice, to focus on the effect of the non-linearity of the crossing fraction matrix on OD estimation. We choose for this approach, not because we believe the effect of route choice to be less important, but because we believe it is important to first analyze and understand the different components of the problem before tackling the entire problem. The effect of route choice will be the subject of future research.

In this paper a number of directions are presented that bring insight into fundamental issues of traditional OD estimation methods, namely the inability to deal with congested networks, and the high risk of converging to a local minimum. We identify and discuss three conditions that should be met for unbiased dynamic OD estimation:

- The DNL or DTA model that is used in the estimation process should be able to represent queuing behavior in a proper way.

- The linear relationship between the OD flows and the link flows that is used in the optimization step at the upper level should account for the non-separability of link flows with respect to OD flow. For details about the issue of separability we refer to section 2.B.
- During the OD estimation process the correct bottlenecks should be activated and the simulated traffic state of the network (free flow or congestion) should match with the actual traffic state of the network.

In the next chapter we will further elaborate on each of the three conditions. A proof of concept is presented in section 3. In section 4 the theoretical properties discussed in the previous chapter are discussed, as well as the practical implications. In the last section we formulate our conclusions along with future research lines.

2. Conditions for unbiased OD estimation

A. Using a correct DNL model

In this section we examine the effect of a different dynamic queuing representation for network loading models on traffic counts based OD estimation. More specifically, we compare three different queuing approaches: point/vertical queuing (PQ), spatial/horizontal queuing (SQ) and queuing based on simplified first-order shockwave theory. For more details about these queuing approaches, as well as some more experiments that examine its effect on OD estimation, we refer to Frederix et al. (2010a). In the experiment described below, we generated synthetic traffic counts as input for the OD estimation, using the Link Transmission Model (LTM) (see Yperman et al.(2006)), a first-order shockwave model. With the LTM as a reference case, we exclude any influence other than the queuing mechanism. For the same reason the initial OD matrix used is the OD matrix that was input of LTM. In this way any deviation from this solution is due to the fact that the difference between the generated traffic counts and the estimated flows from the network loading model pushed the estimation away from the initial, perfectly correct solution. This way it is checked if all DNL models are capable of finding the correct solution if they start close to it.
The network that we use for the experiments is a simple bottleneck, and is depicted in figure 2, as well as the real OD matrix. In this network all links have a sufficiently high capacity to accommodate the input demand, while link 4 has a capacity of only 800 veh/h, and is thus acting as a bottleneck for this demand profile.

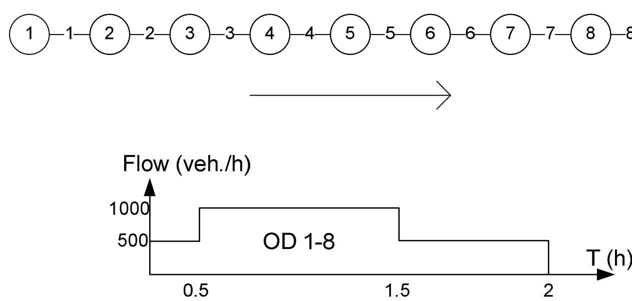


Figure 2: Bottleneck network and input demand

The results are summarized in figure 3. LTM reproduces the solution almost exactly, which is to be expected, as it starts on top of the solution and traffic counts are obtained by LTM itself. The small deviations that can be seen in figure 3 are the result of the stochastic nature of the SPSA algorithm (see Spall (1992) and Spall (1998)) that was used for optimization. The deviations of the SQ and PQ model are much larger. Due to spillback of congestion the measured traffic flow decreases in reality. However, since queuing is not properly modeled by both models, spillback does not occur, or at least not at the correct moment for the same amount of demand. So when a detector is in congestion in reality, it will be in free flow for both of these models. Because of this both the SQ and PQ model tend to misidentify the decreased flow due to spillback of congestion as a lower demand. This leads to a lower estimate of the demand when congestion is emerging and a higher estimate when congestion is dissolving. These results indicate that applying a DNL model that represents queuing properly is a necessary condition for accurate OD estimation. As trivial as this conclusion may seem, in practice there are plenty of DTA models that use the SQ principle rather than traffic flow theory.

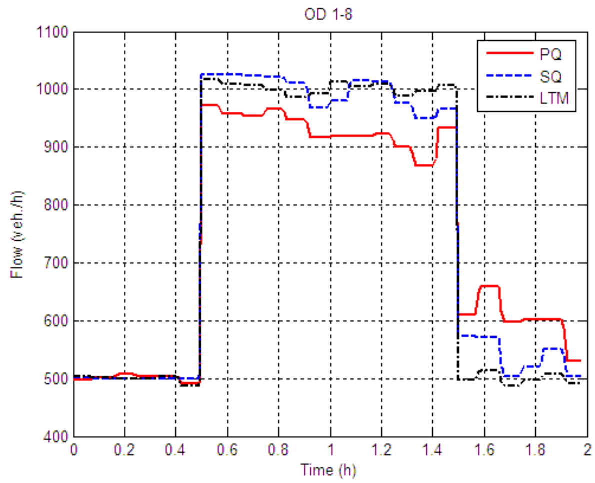


Figure 3: OD flow according to different models on link network

B. Using a correct linear model for the relationship between OD flows and link flows

The choice of a correct DNL model partly explains the discrepancy between real and estimated OD flows. Another important factor is the assumed relationship between OD and link flows that is used in the optimization step at the upper level. The most widely employed technique for formulating this optimization problem is the Generalized Least Squares (GLS) estimator (it is used for example in Cascetta & Postorino (2001), Yang et al. (2001), Bierlaire & Crittin (2004)). The GLS estimator can be formulated as follows:

$$x^* = \arg\min_x \left[(x-\hat{x})^T W^{-1} (x-\hat{x}) + (Ax-\hat{y})^T R^{-1} (Ax-\hat{y}) \right] \quad (2)$$

where W and R are error variance-covariance matrices that reflect the trust one has in the target matrix and the measurements respectively. Usually the above formulation is also constrained to non-negative values of x. This is a standard inequality constrained quadratic program, and can be solved using standard optimization techniques.

Another popular approach is the Kalman filtering technique (for example Okutani (1987), Ashok & Ben-Akiva (2002)). It uses transition equations, capturing the dynamics of the system, and measurement equations, relating the OD flows to the measurements:

$$\begin{aligned} x &= Fx + w \\ y &= Ax + r \end{aligned} \quad (3)$$

Both these approaches use a form of locally approximate traffic model that expresses the relationship between the OD flows and the link flows, namely $y = Ax$, where the assignment matrix A is now composed of scalars. The problem with this local relationship is that it neglects the fact that the assignment matrix A is itself a function of the OD flows. We will now further analyze this relationship, but first let us introduce the following notations:

y_i^m is the link flow at link i during time period m

x_j^n is the flow between OD pair j departing in time period n

$a_{i,m}^{j,n}$ is the proportion of x_j^n that passes link i during time period m

J is the set containing all OD pairs

J_i is the set of OD pairs passing link i

$r \in j$ is the set of routes between OD pair j

While it is impossible to consider this non-linear relationship in the optimization step, it is possible to have a linear model that takes the response of the assignment matrix into account. If we take a first-order Taylor approximation of the link flows at x^* this model would have the following form:

$$y_i^m = \sum_{n=1}^{m} \sum_{j=1}^{J} a_{i,m}^{j,n}\left(x^*\right) x_j^{*n} + \sum_{j=1}^{J} \sum_{n=1}^{m} \frac{dy_i^m\left(x^*\right)}{dx_j^n}\left(x_j^n - x_j^{n*}\right) \tag{4}$$

$$\begin{aligned} with \quad \frac{dy_i^m\left(x^*\right)}{dx_j^n} &= \frac{d\left(\sum_{n'=1}^{m} \sum_{j' \in J_i} a_{i,m}^{j',n'}\left(x^*\right) x_{j'}^{*n'}\right)}{dx_j^n} \\ &= a_{i,m}^{j,n}\left(x^*\right) + \sum_{j' \in J_i} \sum_{n'=1}^{m} \frac{\partial\left(\sum_{r \in j'} b_{i,m}^{r,n'}\left(x^*\right) p^{r,n'}\left(x^*\right)\right)}{\partial x_j^n} x_{j'}^{*n'} \\ &= a_{i,m}^{j,n}\left(x^*\right) + \sum_{j' \in J_i} \sum_{n'=1}^{m} \sum_{r \in j'} \frac{\partial b_{i,m}^{r,n'}\left(x^*\right)}{\partial x_j^n}\left(p^{r,n'}\left(x^*\right) x_{j'}^{*n'}\right) + \sum_{j' \in J_i} \sum_{n'=1}^{m} \sum_{r \in j'} b_{i,m}^{r,n'}\left(x^*\right)\left(\frac{\partial p^{r,n'}\left(x^*\right)}{\partial x_j^n} x_{j'}^{*n'}\right) \end{aligned} \tag{5}$$

In expression (5), the first term consists of the assignment fraction that represents the fraction of OD flow x_j^n that passes link i during interval m. The second term consists of the sensitivity of each element of the crossing fraction matrix (that was non-zero for link i and interval m) to OD flow x_j^n, multiplied with the associated route flow. The third term consists of the sensitivity of all route flows that pass link i during interval m to OD flow x_j^n, multiplied with the associated crossing fraction.

If we compare this expression with the linear relationship that is used in GLS estimators and in Kalman filtering, we notice that both approaches neglect the second and the third term of the gradient in expression (4), that represent the sensitivity of the spatio-temporal distribution and the sensitivity of the route choice fractions to the OD flows respectively. As was already mentioned in the introduction, we will focus on the second term in the rest of this paper.

This matter has already been addressed in Yang (1995), Lindveld (2003) and Tavana (2001). In Yang (1995) explicit derivatives of the link flows to the OD flows are used in the OD estimation. Since the paper deals with static OD estimation, the crossing fraction matrix simplifies to the link-incidence matrix, which is not sensitive to the OD flows. Lindveld (2003) investigates the derivative of the link flows to the OD flows, but decides not to use the last two terms because of computational complexity. In section 4 we will further discuss the complexity of calculating these terms. Tavana (2001) acknowledges the theoretical importance of including both derivatives. Although he does not find a substantial improvement compared to the standard linear relationship, the case study considers a non-congested network, and is thus not well chosen to draw any conclusions on its performance in congested networks. He argues that one of the reasons for the small amount of improvement is that the initial guess of OD flows is very close to the desired solution, because else the process may converge to another solution. This is a different aspect that will be further addressed in section 2.C.

In congested scenarios if these terms are neglected it implies that that the link flows are assumed separable in space and time with respect to the OD flows. Or in other words, it assumes that changing an OD flow of a certain time period only influences the flow on the space-time interval at which it originally passes. This assumption is incompatible with some typical phenomena in congested networks, such as spillback of congestion, time lags due to congestion effects, and interdependencies between crossing or opposing flows through intersections.

Note that there are some methods that do not make use of a linear relationship between OD and link flows to estimate a new OD matrix. Two popular approaches in OD estimation are SPSA (Balakrishna (2006), Vaze et al. (2009)) and evolutionary algorithms (Kattan & Abdulhai (2006), Tsekeris et al. (2007)). Both are gradient-free approaches and only make use of evaluation of the goal function to guide the search onto a descent direction. The use of these optimization algorithms is motivated by the fact that the ability to calculate the gradient vector and the need for a deterministic setting are often unreasonable for practical problems (Vaze et al. (2009)). In these cases SPSA and evolutionary algorithms might provide a good alternative. However, because they do not make use of a gradient, they are highly inefficient especially in large congested network problems. Therefore when it is possible to calculate the gradient and we are in a deterministic setting (which is often the case), traditional gradient-based methods should be preferred.

C. Identifying correct traffic state on links

Many researchers mention the importance of having an initial matrix that does not deviate too much from the actual OD matrix, because else the estimation process may converge to another solution (see e.g. Tavana (2001)).

The reason for deviation from the actual solution can be understood by considering the fact that we are using local approximations of the relationship between the OD flows and the link flows. This local relationship is not necessarily identical to the actual relationship, and might even suggest to adjust the OD flows in the wrong direction. This is especially true for congested networks. Spillback can cause a link flow to decrease by increasing an OD flow. For example, if we have a simple network like the one sketched in figure 2, and in reality there is spillback on a link (point (q'', k'') in figure 4), but this link is thought to be in free flow (point (q, k)) with the current estimate of the demand, the linear relationship at this point for this measurement will suggest to decrease the OD flow to get point (q'', k'). Vice versa if in reality the detector is in free flow regime (point (q, k)), but this link is in congestion regime (point (q'', k'')) with the current estimate of the demand, the linear relationship at this point for this measurement will suggest that the link flow is insensitive to any change of the OD flows, which of course only holds locally. These examples indicate that starting from an incorrect traffic regime leads to incorrect relationships between the OD flows and the link flows, which on their turn lead to convergence to local minima.

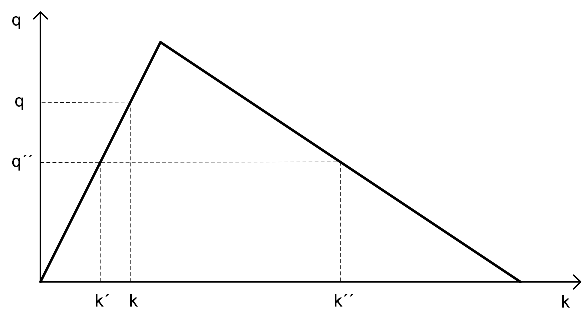


Figure 4: Incorrect traffic state induces errors

Therefore the last condition for unbiased OD estimation is that during the estimation process the correct bottlenecks are activated and the simulated traffic state of the network (free flow or congestion) should match with the actual traffic state of the network. By including this condition we hope to avoid (or at least decrease) the occurrence of local minima.

3. Case study

In this section we combine conditions 2 and 3 from the previous chapter in a case study on a very simple network. Figure 5 depicts the network and the real OD matrix. The merging of the 2 OD pairs creates a bottleneck on link 9. The flow from the right branch (OD 2-3) takes priority over the flow from the left branch (OD 1-3) and a queue forms on the left branch that spills back on links 2, 3 and 4 (indicated by the dark rectangle in figure 5). Detectors are positioned on the left and top branch of the network (indicated by the crosses in figure 5). Note that OD 2-3 is not measured directly, but can be deduced by the flow measurements on the left and top branch.

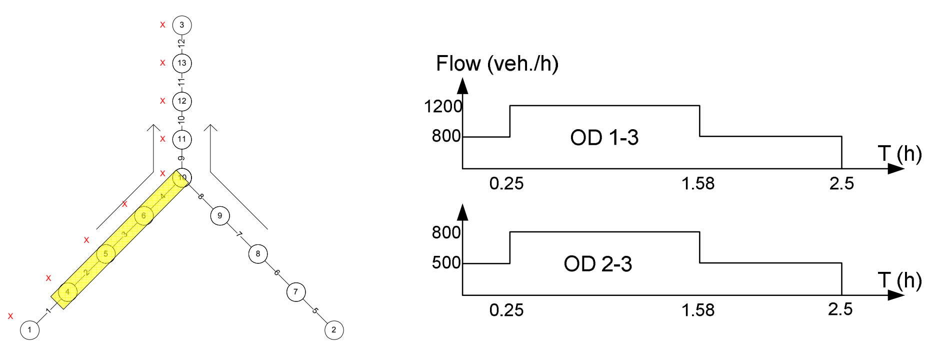


Figure 5: Merge network and input demand

The effect of starting in a different regime and the effect of a different linear model between the OD flows and the link flows is investigated. The effect of using different queuing models is not investigated in this section to limit the number of combinations, and because its effect is already described elsewhere in the paper. Four different OD estimations are performed using a gradient-based optimization method:

a) Incorrect starting regime, incorrect linear model
b) Incorrect starting regime, correct linear model
c) Correct starting regime, incorrect linear model
d) Correct starting regime, correct linear model

The initial OD matrices and the estimated OD matrices are summarized in figures 6 -10.

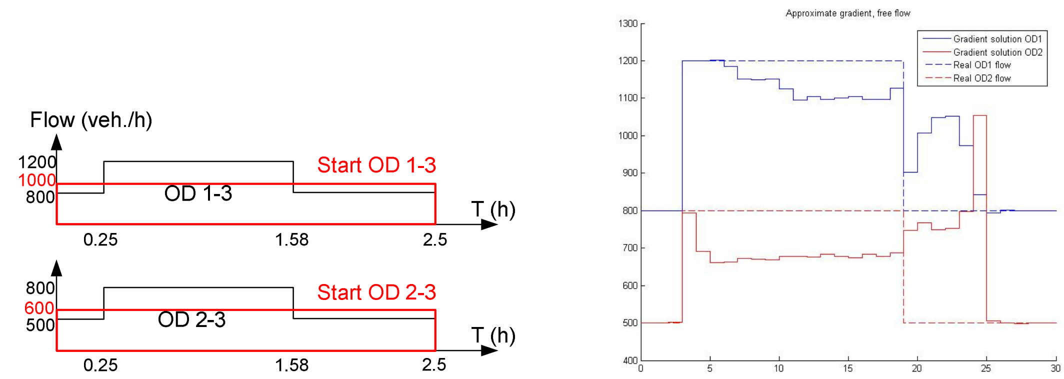


Figure 6: Incorrect regime, incorrect linear model

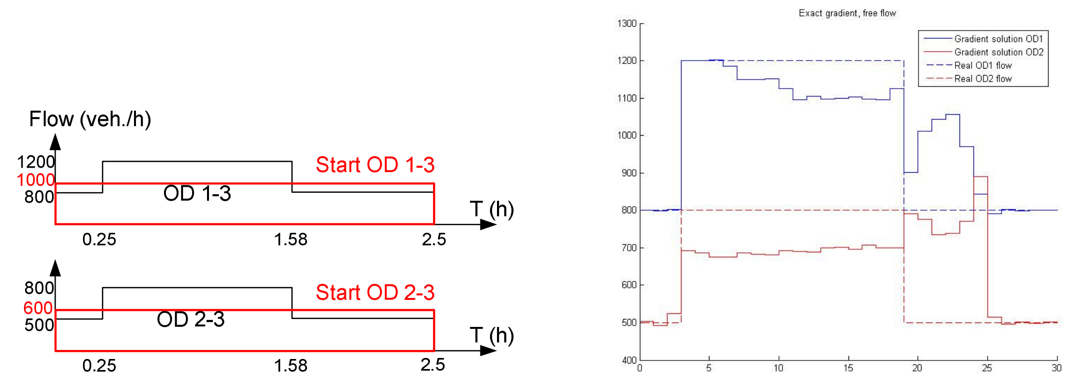


Figure 7: Incorrect regime, correct linear model

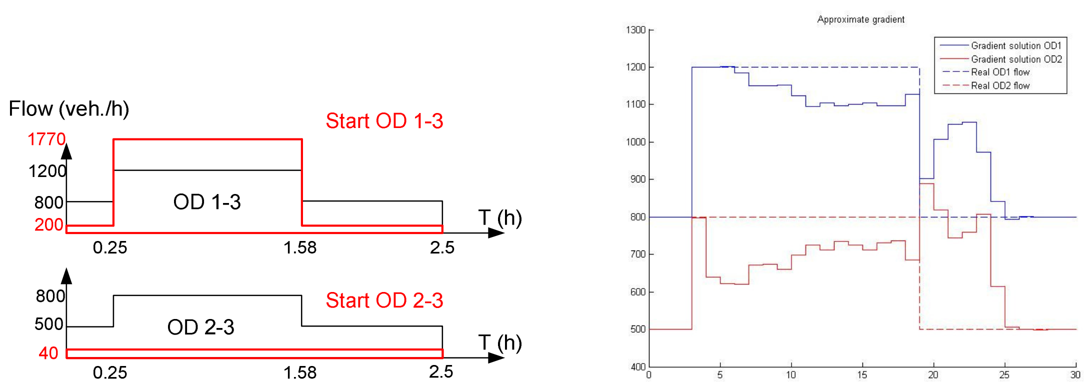


Figure 8: Correct regime, incorrect linear model

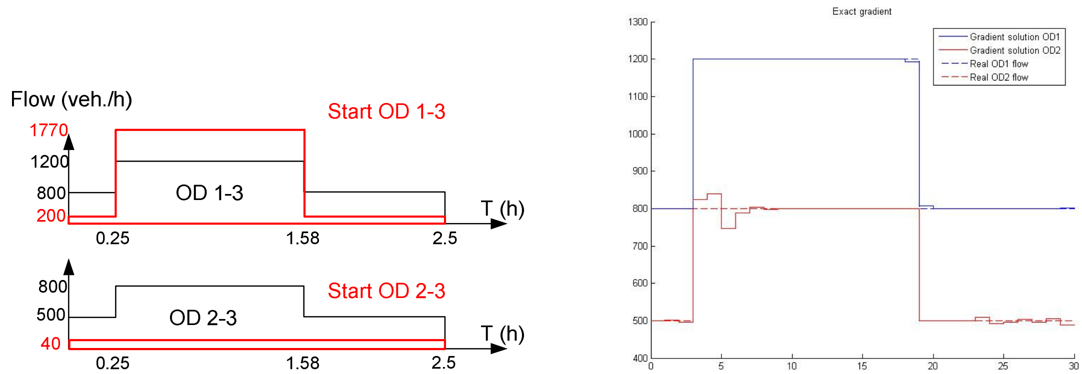


Figure 9: Correct regime, correct linear model

In all but the last case the OD estimation process converges to an incorrect solution. Note that in case d) the exact solution is not fully reached because the problem is underdetermined for OD flows from 2 to 3 in the time period between the start of the congestion spillback and the moment that this queue spills back over the first detector on the left branch. Apart from this transition effect the estimation is quite accurate.

These results suggest that both conditions need to be fulfilled to obtain an unbiased OD estimation method. The example also illustrates that the condition that the initial OD matrix should be close to the actual OD matrix can be a misleading: the average deviation from the actual OD matrix in case a) and b) is 200 veh/h for OD 1-3 and about 150 veh/h for OD 2-3, while for case c) and d) the average deviation is about 580 veh/h and 620 veh/h respectively.

4. Discussion

In this section we discuss some theoretical and practical issues dealing with the stated conditions from the previous sections.

A. Theoretical considerations

To our knowledge there are no OD estimation methods that consider all the three conditions we mentioned at once. However, this does not hinder existing methods to successfully estimate OD matrices in many cases. The reason that even without the proposed conditions a correct solution can be found can be attributed to the choice of a proper initial OD matrix and also to the specific combination of the detector configuration and the situation in reality (that determines the measurement values at these detectors). Although the OD estimation problem is non-convex in general, a specific detector configuration on a simple network can make the problem convex for a large domain in the solution space. Consider for example the network of figure 2. The goal function of this OD estimation problem is convex for the entire solution space of OD flows that do not cause congestion spillback onto the most upstream detector. Another reason for obtaining equally good estimation results could be that the real OD matrix does not result in any congestion. In this case the PQ and the SQ model would also be able to find the correct OD matrix. Also a good initial OD matrix can still belong to the attraction domain of the global minimum, even though it does not satisfy the third condition.

All these cases should be interpreted as special cases of a more general non-convex problem. The proposed conditions are therefore not necessary for each specific case, but should be interpreted as necessary in the sense that they are needed for the OD estimation to be unbiased in *any* case (regardless of specific situation). Without these conditions it cannot be guaranteed that we find a correct solution: it is very likely that we get stuck in a local minimum.

A second important point of discussion is about the sufficiency of these conditions. The question is whether we can guarantee convergence to the global minimum with these conditions. The main idea behind the condition of starting in the correct regimes is that we hope to get a monotonic relation between any link flow and any OD flow. If this is true, then each squared deviation in the goal function of equation 2 is a unimodal function[1]. If we assume there are no measurement errors, the mode of each of these functions is equal to the

[1] A function $f(x)$ is unimodal if for some value m (the mode), it is monotonically increasing for $x \leq m$ and monotonically decreasing for $x \geq m$.

solution. Then the sum of these functions is unimodal as well, which means that any local minimum is a global minimum. Note that we do not believe that the absence of measurement errors is a realistic assumption, but this argumentation shows that measurement errors are a potential source of local minima.
The above argumentation is supported by the assumption of a monotonic relation between any link flow and any OD flow that would be obtained by starting in the correct regime. However, it can be proven with an example that this assumption does not always hold.

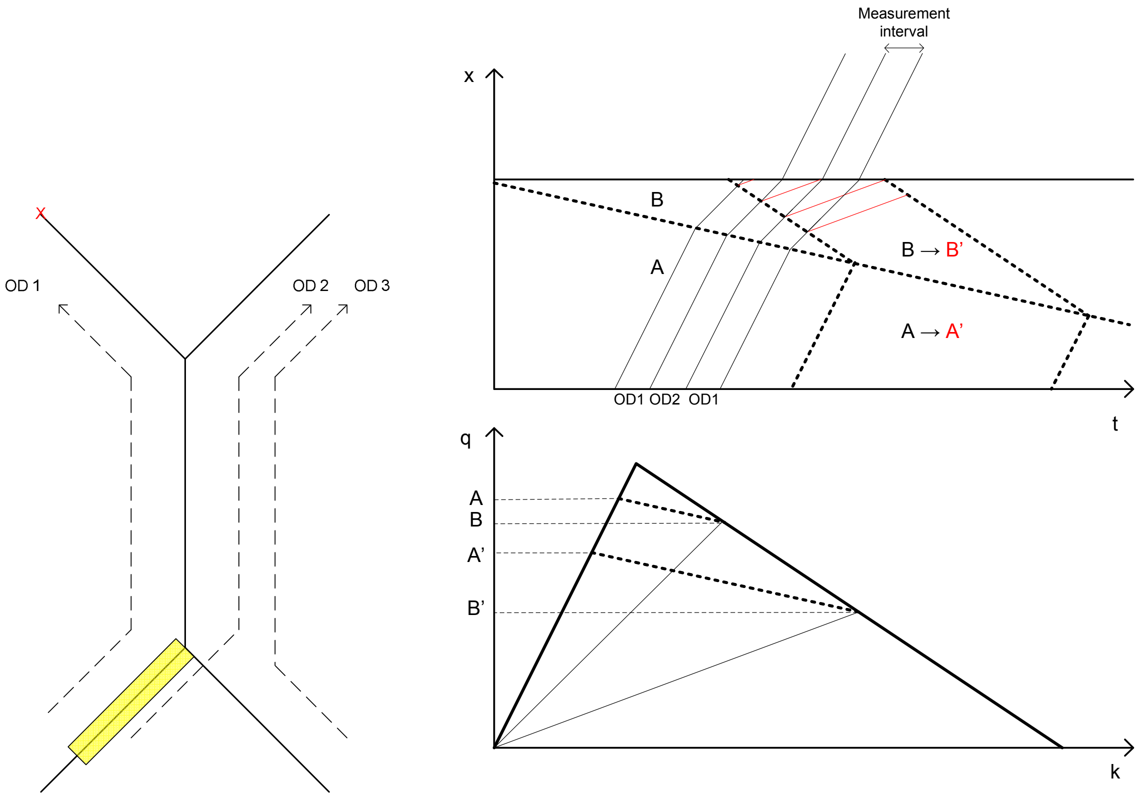
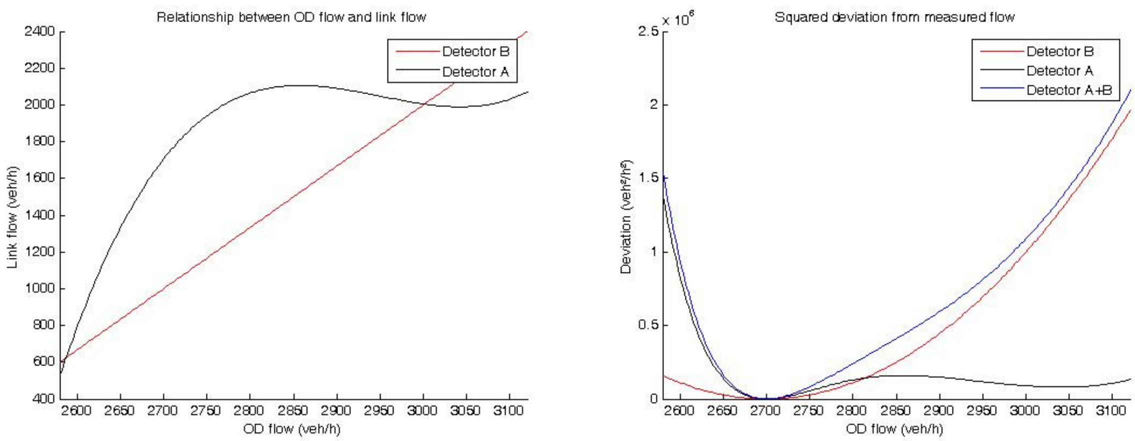


Figure 10: Example of non-monotonic relation between OD flow and link flow

Consider the network depicted in figure 10. The flow from the right branch (OD3) takes priority over the flow from the left branch (OD1 & OD2). For this reason a queue starts building on the left branch. The flow in this queue is equal to the remaining capacity (point B). The sum of OD1 and OD2 is equal to the flow A shown also on the fundamental diagram in the bottom-right picture. Now suppose we decrease OD1 and increase OD3, in such a way that the speed at which the queue spills back remains the same. The demand on the left branch then equals A', and the remaining capacity equals B'. By doing so the speed in the queue drops, which can be seen in the x-t plot (the top-right picture in figure 10) by the smaller slope of the trajectories in the congested area. Assume that the composition of A varies in time as depicted in the x-t plot: a time period with only OD1 alternates with a time period with only OD2.
Now the change of the slope alters the composition of the flow passing detector on the upper left branch during the indicated measurement interval. While originally it detected a flow equal to B (all OD1), now it does not detect any flow (because OD2 does not pass there). So by decreasing OD1 (and increasing OD3) the link flow is decreased. If now we would further decrease OD1 (and increase OD3, such that the spillback speed remains the same), the slope of the trajectories further decreases, and the detector would measure some OD1 flow passing. Thus, by decreasing OD1, the link flow decreases at first, and afterwards increases again, which is clearly a non-monotonic relationship.
Therefore the proposed conditions are not always sufficient for convergence to the global minimum. The example illustrates that delays can lead to shifts in the composition of flow, which can cause non-monotonic behaviour of the relationship between OD flows and link flows. It should be noted that a non-monotonic relationship does not necessarily lead to local minima in the goal function. Similar as the special cases, (discussed in the section dealing with the necessity of the conditions) there are cases with a specific detector configuration in which no local minima are present. Consider for instance an OD flow that has a non-

monotonic relationship with the flow on detector A described by $y_A = f_A(x)$. Suppose that this OD flow is also measured by detector B, and that the relationship between the OD flow and the flow on detector B is monotonic and described by $y_B = f_B(x)$. Then the squared deviation $(\hat{y}_A - y_A(x))^2$ in the goal function is not unimodal, but $(\hat{y}_B - y_B(x))^2$ is unimodal. Now because in the goal function we take the sum of both quadratic terms, and this is the sum of a non-unimodal and a unimodal function, the result might become unimodal. This is illustrated in figure 11.



(a) **(b)**

Figure 11: Non-monotonic and monotonic relationship between OD flow and link flow (a) and corresponding non-unimodal and unimodal relationship (b)

This is more likely when the deviation from monotonicity in the relationship between the OD flow and the flow on detector A is small. Well then, it can be expected in practice that shifts in the composition of flow caused by delays will be quite small in practice, and thus also the fluctuation of the link flow caused by the altered composition will be rather small. The fluctuation of the link flow caused by spillback on the other hand can be expected to have a larger magnitude. Therefore it seems reasonable to us to assume that local minima caused by altered travel times will originate far less frequently than those caused by congestion spillback.

B. Practical implications

In the previous sections the importance of starting in the correct regime was stressed. However, it seems that there are some feasibility problems with this condition. First of all we do not dispose of measurements that are continuous in space and time that could be used to check if we are in the correct regime. Therefore the best one can do in practice is to meet this condition for all measurement intervals (in space and time). Yet since originally it was our goal to get a monotonic relationship between the OD flows and the link flows of each time period, it seems sufficient to meet this condition for the measurement intervals.

A second practical question is how to get to the correct regime. In Tavana (2001) the use of a hybrid model is suggested that estimates OD flows based on both traffic flow and density. This approach has been tested successfully in Frederix et al. (2010b). Research is still ongoing on more efficient methods.

Also the second condition raises objections concerning the feasibility. The derivative of the link flows to the OD flows could be calculated using finite differences, but this would require running a number of simulations equal to the number of OD flows in time, which is unfeasible for medium to large-scale networks. Therefore we are doing research on an efficient method for approximately quantifying the effect of changing the demand, using similar techniques as are described in Corthout et al. (2009).

5. Conclusions & future research

Many existing OD estimation methods lack of a theoretically sound base for dealing with congested networks. The present paper analyzes a number of shortcomings in current assignment-based dynamic OD estimation models and proposes three conditions that need to be met for unbiased OD estimation in congested networks: (1) a DNL model that reproduces the correct queuing behaviour should be used, (2) in the optimization step a correct linear model that takes the non-separability of link flows with respect to OD flows into account is necessary, and (3) the traffic states on the links obtained by propagating the initial OD matrix should match with the actual traffic states. A proof of concept is presented that illustrates the necessity of all three conditions. However, we also show that these conditions cannot guarantee a successful estimation. On the other hand, we expect that in practice the conditions should suffice or at least they will highly improve the reliability of the estimation results.

The proposed conditions, although theoretical of nature, should provide practical insight for people that are dealing with OD estimation on congested networks: practitioners should have knowledge of the assumptions underlying a certain estimation method, and should realize how these assumptions can lead to errors in the estimated OD matrix. Our conditions can therefore be interpreted as a practical guideline that can be used when the outcome of an OD estimation method is not satisfactory. This paper also mentions the fact that specific detector configurations can lead to a convex goal function. This understanding might have important consequences for the sensor location problem.

Future research will concentrate on the implementation of an efficient method to calculate the derivatives of the link flows to the OD flows, as well as the implementation of an approach to get to the correct regime. These methods should then be tested on a real network using real data to study the feasibility of this approach. Also the difference between this approach and traditional approaches should be tested. Finally comparable insights and advances with respect to route choice models are required.

References

Ashok, K, & Ben-Akiva, ME 2002, *'Estimation and prediction of time-dependent origin–destination flows with a stochastic mapping to path flows and link flows'*, Transportation Science vol. 36, pp 184–198.

Balakrishna, R 2006, 'Off-line Calibration for Dynamic Traffic Assignment Models', PhD thesis, Massachusetts Institute of Technology.

Bierlaire, M & Crittin F 2004, *'An efficient algorithm for real-time estimation and prediction of dynamic OD tables',* Operations Research vol. 52(1), pp. 116-127.

Cascetta E 2001, *Transportation Systems Engineering Theory and Methods*, Kluwer Academic Publishers, The Netherlands.

Cascetta, E & Postorino, MN 2001, *'Fixed Point Approaches to the Estimation of O/D Matrices Using Traffic Counts on Congested Networks'*, Transportation Science, vol. 35, no. 1, pp. 134-147.

Corthout, R, Tampère CMJ, Immers LH 2009, *'Marginal Incident Computation: Efficient Algorithm to Determine Congestion Spillback due to Incidents'*, Transportation Research Record no. 2099, pp 22-29.

Frederix, R, Tampère, CMJ, Viti, F & Immers, LH 2010a, 'The effect of Dynamic Network Loading models on DTA-based OD estimation', *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment (Proceedings of DTA2008, Leuven), Edward Elgar, Cheltenham, UK - Northampton, MA, USA*.

Frederix, R, Viti, F & Tampère, CMJ 2010b, *'A density-based dynamic OD estimation method that reproduces within-day congestion dynamics'*, paper submitted for the IEEE-ITSC 2010 conference, September 19-21, Madeira Island, Portugal.

Kattan, L & Abdulhai, B 2006, *'Non-Iterative Approach to Dynamic Traffic Origin/Destination Estimation using Parallel Evolutionary Algorithms'*, presented at the 85th annual meeting of the Transportation Research Board, Washington, USA.

Lindveld, K 2003, 'Dynamic O–D matrix estimation: a behavioural approach', PhD thesis, Trail Thesis Series, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands.

Okutani, I 1987, *'The Kalman filtering approach in some transportation and traffic problems'*, N. H. Gartner and NHM Wilson, eds. Transportation and Traffic Theory. Elsevier, New York, pp. 397–416.

Spall, JC 1992, *'Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation'*, IEEE Transactions on Automatic Control, Vol. 37, pp. 332-341.

Spall, JC 1998, *'An Overview of the Simultaneous Perturbation Method for Efficient Optimization'*, Johns Hopkins University APL Technical Digest, Vol. 19, No. 4.

Tavana, H 2001, 'Internally-consistent estimation of dynamic network origin–destination flows from intelligent transportation systems data using bi-level optimization', PhD thesis, Department of Civil Engineering, University of Texas at Austin, Austin, Texas.

Tsekeris, T , Dimitriou, L & Stathopoulos, A 2007, *'Simultaneous Origin-Destination Matrix Estimation in Dynamic Traffic Networks with Evolutionary Computing'*, M Giacobini et al., eds. EvoWorkshops 2007, LNCS 4448, pp. 668–677.

Vaze, V, Antoniu, C, Wen, Y & Ben-Akiva, M 2009, '*Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance*', presented at the 88th annual meeting of the Transportation Research Board, Washington, USA.

Yang, H 1995, 'Heuristic algorithms for the bilevel origin–destination matrix estimation problem', *Transportation Research B,* vol. 29, pp 231–242.

Yang, H, Meng, Q, & Bell MGH 2001, *'Simultaneous estimation of the origin–destination matrices and travel-cost coefficient for congested network in a stochastic user equilibrium'*, Transportation Science vol. 35, pp 107–123.

Yperman, I, Logghe, S, Tampere, CMJ & Immers, LH 2006, *'The multi-commodity link transmission model for dynamic network loading'*, presented at the 85th annual meeting of the Transportation Research Board, Washington, USA.