A Hierarchical Approach for Dynamic Origin-Destination Matrix Estimation on Large-Scale Congested Networks

Rodric Frederix*, Francesco Viti, and Chris M. J. Tampère

Abstract—In this paper we propose a hierarchical approach for decomposing and simplifying the dynamic OD estimation procedure for large-scale congested networks. The main idea is to subdivide the network into multiple hierarchical levels. Next an OD estimation is performed on each level separately, starting with the highest level, and the output of this estimation is used as input for the OD estimation on a lower level. The main advantage of this approach is that different levels of complexity (of the DNL/DTA model and of the estimation method) can be used for different parts of the network as required. In addition, this hierarchical approach solves many practical and theoretical limitations of traditional OD estimation methods, which have been identified in previous research.

I. INTRODUCTION

The importance of dynamic traffic assignment (DTA) models needs no introduction. They form a central component in a variety of traveler information systems, e.g. Wang et al. [1] and they are essential for traffic and transportation assessment analyses. A critical component in DTA is the dynamic origin-destination (OD) matrix, which captures the spatial and temporal patterns of traffic demand, containing information about activity patterns, trip sequences and scheduling preferences in an aggregated way. The OD matrix is distributed and observed onto the network in form of link and path flows. Errors in estimating this matrix are therefore transferred onto the network flows, leading to errors in the network performance measures.

In many dynamic OD estimation methods (e.g. Cascetta & Postorino [2]; Yang et al. [3]) a bi-level problem is solved for the entire network. At the upper level the distance between estimated and measured flows is being minimized, given a certain (assumed) relationship between the link flows and the OD flows. At the lower level the estimated OD flows are propagated to generate an approximation of the relationship between link flows and OD flows. In this process, all link flow measurements are centralized in one goal function at the upper level, which is then (iteratively) optimized. At first sight this seems a theoretically sound and rigorous methodology: all information is used simultaneously to estimate the OD matrix, so the dependency between link flows is captured at once. However, there are several problems to this methodology dealing with large-scale networks. These difficulties originate from the fact that the computation time required for solving the OD estimation problem grows superlinear with respect to the size of the network. There are three components that constitute the total amount of computation effort in the conventional OD estimation procedure. One is related to the upper level of the bi-level problem, in which a new OD matrix is estimated. The number of OD pairs, and thus the size of the estimation problem grows quadratically with the size of the network. The second component is related to the lower level, in which the estimated OD matrix from the upper level is assigned to the network, i.e. the DTA component. A third component relates to the combination of both levels, more specifically the number of iterations that is required for convergence of the bi-level problem.

There are many methods described in literature that reduce the required computation time. The application area of most of these methods is real-time OD estimation. Cascetta [4] proposes the use of sequential estimators. Hereby the optimization problem is decomposed in time into smaller, more manageable ones. Bierlaire & Crittin [5] present a solution algorithm that exploits the sparsity of the assignment matrix, allowing to consider larger problems. In both approaches there is a trade-off between accuracy and a feasible computation time: they both introduce some approximations that introduce a (small) bias in the estimation.

The need for fast and efficient OD estimation methods is self-evident in the context of real-time OD estimation. But also for offline OD dynamic estimation the balance between accuracy and computation time remains an issue. In recent work (Frederix et al. [6]) the authors of this paper have shown some theoretical limitations of traditional OD estimation methods in congested networks. It was shown that the occurrence of congestion spillback requires the calculation of the sensitivity of the link flows to the OD flows: otherwise the information of the measurements can be misinterpreted, leading to biased results. However, deriving this sensitivity is cumbersome in terms of computation time. In Frederix et al. [7] an efficient method for calculating this sensitivity is presented and successfully applied on a primary road network. However, the results also suggest that in practice it is not possible to calculate
Another issue when dealing with large-scale networks relates to the Dynamic Network Loading (DNL) model that is used. In all assignment-based OD estimation methods this DNL model determines the relationship that is assumed between the OD flows and the link flows, and thus it determines how the measurements are interpreted (see [8]). There are multiple types of DNL models that each fit different applications. For example, a DNL model based on kinematic wave theory is best suited for freeway traffic modeling, while a DNL model based on travel time functions or queuing theory might be better suited for urban traffic modeling. However, each of these models will face some difficulty when applied on a mixed network, leading to errors in the estimation. Also for the sake of computational efficiency it is often desirable to use faster, less complex DNL models on parts of the network that are not that much of interest.

The above-stated problems in large-scale networks suggest the use of a ‘divide and conquer’ technique. That is to say, to subdivide the larger problem into a set of smaller problems that are feasible to solve in terms of calculation time. The benefits of such methodology are as follows. First of all, the size of the estimation problem does no longer grow quadratically with the size of the network, and thus becomes scalable. Secondly, decomposing the network creates opportunities so that different degrees of accuracy and even different estimation procedures may be adopted, in order to make it computationally tractable and efficient. More specifically, the network can be decomposed into areas where congestion dynamics largely affect the accuracy and reliability of the estimated OD flows, and others where fine-grained estimation may not be necessary. In the former case the sensitivity-based OD estimation (SBODE) methodology of Frederix et al. [6] can be applied, whereas in the latter case a traditional OD estimation procedure may suffice. It also allows to use a DNL model that is best suited for the subnetwork that is considered at the current stage.

Furthermore the subdivision of the larger network should be well-chosen. We propose to subdivide the network into different hierarchical levels that more or less coincide with the hierarchic structure of a network. The main idea behind this is that the calibration of the demand on a lower level (e.g. secondary road level) highly depends on what happens on a higher level (e.g. primary road level). Therefore no effort should be spent on calibrating the demand on a lower level before the higher level is well calibrated.

In the next section a hierarchical approach for dynamic OD estimation is presented that builds upon the philosophy of subdividing the problem into a set of smaller problems. In section 3 we present a proof of concept that illustrates the methodology and its advantages. Finally we discuss some remaining challenges.
of the computation time, i.e. the computation time for the estimation.

In both approaches the link flows resulting from the OD matrix estimate on a subnetwork are used as measurements in the OD estimation on a lower level OD estimation. However, both approaches assume that detailed knowledge on the actual travel times and route choice is available. However, in many cases where only flow measurements are available on a limited part of the network, this assumption is too restrictive. In those cases assignment-based methods are required, where a DNL/DTA model is used to model the propagation. In the context of assignment-based OD estimation, which is the focus of this paper, the idea of adding the link flows resulting from the OD matrix estimate on a subnetwork as measurements in the OD estimation on a lower level OD estimation, does not suffice for various reasons. First of all it does not allow to limit the use of the SBODE methodology to only one part of the network. The reason is that if we only add link flows from this part of the network, there is no guarantee that the congestion patterns that where calibrated using the SBODE methodology remain the same: these congestion patterns are the result of the OD flows and the propagation map. The propagation map is itself a function of the OD flows. Since in the lower level there is no restriction on the OD flows passing in the considered part of the network, there is also no restriction on the propagation map here, and thus no guarantee that the calibrated congestion patterns remain the same. A second problem related to this approach concerns calculation time. Estimating the OD matrix on a subpart of the network from the measured flows requires multiple DTA runs. When we only add flow measurements from this subpart to the lower level OD estimation, the calculations on the subpart have to be repeated.

Therefore we suggest the following approach: apart from using the link flows resulting from the OD matrix estimate on a subnetwork as input to the OD estimation on a lower level OD estimation, also the propagation map on the subnetwork needs to be added as an input. With propagation map we mean the mapping between the inflow and outflow the subnetwork. Furthermore, the propagation map on that part of the network should be kept fixed in the OD estimation process. This will introduce an inconsistency between the OD flows on the subnetwork and its propagation map during the estimation process. However, if at the end of the estimation process the inflows and outflows of the subnetwork match with the link flows that were added, there is no longer any inconsistency. Apart from transferring this fixed propagation map a similar approach of subdividing the OD estimation geographically as in [1] is advocated. The suggested hierarchical approach can therefore be seen as an extension of the distributed approach of [1] to cases where no accurate travel times or assignment matrix are available.

The algorithmic procedure for the hierarchical approach can be summarized as follows.

- In a first stage only the higher level network is considered (see Figure 2). The connection with the underlying network is modeled by creating origins and destinations, and an OD matrix for this network (with the created origins and destinations) is estimated. If there is congestion on this higher level network, the computationally expensive SBODE methodology is used.

- In a second stage a new OD estimation is performed on the combination of the higher and lower level network. Important here is that in this new OD estimation the propagation of the flow over the links from the higher level network is fixed and equal to the propagation related to the estimated OD matrix from the first stage. As we suggested above, this also has an impact on the total computation effort, since no complex calculations on the propagation of the flows on the higher level are required anymore. The simulated link flows at the boundaries of the network obtained from the first stage are used in the second stage, either as a constraint, or they can be added to the goal function with a large weight. This is done to avoid inconsistency between the link flows and the fixed propagation: if the final OD matrix produces different link flows on the links of the first stage, the assumed propagation does not necessarily hold anymore.

![Figure 2: Cutting out the higher level network.](image)

To summarize the hierarchical approach solves the following problems:

- It is now possible to use a DNL/DTA model that is best suited for the network that is considered at the current stage.
- No alterations are made to the OD flows that affect the deviation on lower level link, before the propagation map on the primary road network is of decent quality.
- The calculation effort can be expected to reduce because the propagation map of higher levels is kept fixed, and thus no longer requires using a complex DNL/DTA
• In the higher level estimation it is also possible to do consider multiple independent networks for which enough detector information is available to determine the propagation map. This would allow to perform parallel computations, allowing much faster and more objective function evaluations.

• The hierarchical approach allows for calculating the sensitivity of the link flows to the OD flows only for specific levels. Since congestion spillback is a phenomenon that mainly occurs at the higher level(s), we can suffice with calculating the sensitivity only for the higher level(s), thereby reducing the amount of calculation effort.

III. PROOF OF CONCEPT

In this section the methodology described in section III is illustrated on the synthetic network of figure 3, and compared with a traditional, single-level approach. The network consists of a highway network and an urban road network. There are two origins, two destinations and three OD pairs (O1-D1, O2-D1 and O2-D2). Note that this network does not in any way represent the order of magnitude of networks that we aim to tackle with the presented methodology. It was conceived as the smallest, most simple network to illustrate the hierarchic approach. Both networks are linked by OD 2-1, which has a route via the highway and a route via the urban network. On the highway network the flow is propagated following first-order kinematic wave theory, whereas on the urban network the travel times (and thus the propagation) are determined by a Bureau of Public Roads (BPR) function. The correct OD flows are chosen such that they cause congestion spillback on the highway network at the location of the onramp, as is depicted in figure 3 by the yellow rectangle on the left branch. Furthermore it is assumed that the route choice for OD 2-1 corresponds to a Stochastic User Equilibrium (SUE). The corresponding simulated link flows at the detector points in figure 3 are used as synthetic measurements for the OD estimation.

Figure 3: Network highway and urban roads

In the first stage the highway network is cut out, and an OD estimation is performed for this network (see figure 4). Since there is congestion spillback on this network the SBODE methodology is used, i.e. we start with an initial OD matrix that results in the actual congestion pattern, and the sensitivity of the link flows to the OD flows is calculated and used to determine the new OD matrix. This OD matrix estimate results in certain link flows and a certain propagation map. Both are used in the second stage: the simulated link flows at the boundaries of the network (i.e. at the on and off ramps) are used as virtual detectors, and the propagation map on the highway network is fixed in the second stage.

Figure 4: First stage network

In the second stage the entire network is again considered. The detectors on the highway network are replaced by virtual detectors at the on and off ramps, as can be seen in figure 5. The measurements on these virtual detectors are added to the goal function and are given a high weight. Since there is no congestion spillback on this network (the propagation map on the highway part is fixed) there is no need to use the SBODE methodology; a traditional OD estimator will suffice.
We compare this approach with a traditional, single-level approach. The measures of performance that are used are the Root Mean Square Error (RMSE) and the Root Mean Square Normalized (RMSN), defined as follows:

\[
RMSE = \sqrt{\frac{\sum(y - \hat{y}(x))^2}{n}} \tag{1}
\]

\[
RMSN = \sqrt{\frac{n \sum(y - \hat{y}(x))^2}{\sum y}} \tag{2}
\]

Both the actual OD flows and the OD flows that were estimated by the two different approaches are depicted in figure 6. In table 1 the performance measures for link flows and OD flows are presented. Both the figure and the table indicate that the hierarchic approach performs better than the traditional approach. We would like to point out that the main reason for the superior outcome is the use of the SBODE methodology on the highway part of the network. For the single-level approach this methodology was not used, since the idea was that it is not feasible to apply this on large-scale networks.

### Table 1: Comparison of results of traditional and hierarchical approach by Root Mean Square Error (RMSE) and Root Mean Square Normalized (RMSN)

<table>
<thead>
<tr>
<th></th>
<th>Traditional approach</th>
<th>Hierarchical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE link flows (veh/h)</td>
<td>80.9</td>
<td>24</td>
</tr>
<tr>
<td>RMSN link flows (%)</td>
<td>6.3</td>
<td>1.9</td>
</tr>
<tr>
<td>RMSE OD flows (veh/h)</td>
<td>140.7</td>
<td>35.3</td>
</tr>
<tr>
<td>RMSN OD flows (%)</td>
<td>14.2</td>
<td>3.6</td>
</tr>
</tbody>
</table>

IV. DISCUSSION

The hierarchical approach solves many practical and theoretical limitations of traditional OD estimation methods, and is therefore an important step towards the
applicability of dynamic OD estimation in large-scale networks. One of the challenges solved by this approach is how to connect the information from an OD estimation of a subarea of a network with the OD estimation for a different subarea: what type of information should be transferred, and how should it be handled?

Another remaining challenge is the following. By only using the subset of detectors that is present on the considered subarea it is not guaranteed that the propagation map is uniquely determined (while this might have been the case if the entire set of detectors would have been used). This means that in that case the hierarchic approach will introduce some bias compared to the traditional approach. At first sight this case seems more theoretical of nature: in practice it does not occur often that the propagation on a certain part of the network can only be deduced by information on another part of the network. Normally either enough detectors are available on a subarea to determine its propagation map, or either the amount of detectors on the entire is too low to determine the propagation map of the subarea. Note also that the hierarchic approach would only fail when the propagation map is not uniquely determined, not when the OD flows are not uniquely determined. The latter one requires much more information than the former.

REFERENCES