1 INTRODUCTION

This paper describes research on the quasi dynamic assignment model STAQ (acronym for Static Traffic Assignment with Queuing; first described in Brederode et al, 2010) applied on the province wide traffic model of Noord-Brabant, the second largest urban region in the Netherlands with some 2 million inhabitants, a large concentration of high-tech companies and the highest patent density of all European regions.

1.1 Research motivation and objective

In 2014 the province of Noord-Brabant commissioned an audit on its strategic traffic modelling methodology called the ‘Brabant Brede Model Aanpak’ (BBMA). Regarding the assignment model used within the BBMA, the audit (Gorris et al, 2014) defined two topics for improvement: (1) the traditional static traffic assignment (STA) model should be replaced by an assignment model that captures the physical effects of congestion, mainly to improve modelled travel times and broaden opportunities for analysis; and (2) convergence of the assignment model towards the user equilibrium should be guaranteed (now a fixed number of iterations is used as stop criterion). Further recommendations included that junction modelling should remain an integral part of the assignment model to realistically model urban areas.

This research is conducted to explore how and to what extent the desired improvements can be realized, considering both the required effort (in terms of development time needed for embedment and effects on model complexity, calculation time and scalability) as well as the degree of improvement (in terms of realism and convergence). Besides a report on the research findings, a (prototypical) implementation of the method within the BBMA is also part of research output, because the intention was to implement a successful method in the second generation of the regional models using the BBMA (on which development in fact started late 2015). This also means that focus within this research is to embed and optimize an existing method within the BBMA, rather than development of a new method from scratch.

1.2 Context and assumptions

The BBMA is used in all regional transport models within the province of Noord Brabant, as well as a coarser ‘base’ model of the whole province that is being used to maintain consistency between the different regional models. This
approach is described in Heynickx et al, 2016. In Figure 1 some facts and figures about these models are given.

<table>
<thead>
<tr>
<th>model</th>
<th>#centroids</th>
<th>#inhabitants</th>
<th>#jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>5.047</td>
<td>240.292</td>
<td>104.167</td>
</tr>
<tr>
<td>Midden</td>
<td>5.508</td>
<td>389.892</td>
<td>191.318</td>
</tr>
<tr>
<td>Breda</td>
<td>6.043</td>
<td>445.596</td>
<td>223.449</td>
</tr>
<tr>
<td>Noord-Oost</td>
<td>5.880</td>
<td>221.755</td>
<td>109.582</td>
</tr>
<tr>
<td>Den Bosch</td>
<td>5.979</td>
<td>417.957</td>
<td>213.063</td>
</tr>
<tr>
<td>SRE</td>
<td>7.200</td>
<td>738.855</td>
<td>319.338</td>
</tr>
<tr>
<td>BBMB ('base' model)</td>
<td>3.321</td>
<td>2.454.347</td>
<td>1.1603.917</td>
</tr>
</tbody>
</table>

Figure 1: facts and figures about the different regional models in Noord Brabant

In its base, the first generation BBMA models are aggregated four step models that feed travel times back from the assignment (route choice) model into the demand (mode and destination choice) model and use a pivot point method to project forecasted effects on a calibrated base matrix. The static assignment model tries to reach user equilibrium and distinguishes multiple user classes that have their own route choice parameters (varying among purposes and vehicle types) and networks (varying among vehicle types) and as such takes up most of the calculation time of a BBMA model. A complete description of the model system can be found in Kiel and van Grol, 2012 and Koopal and van der Werken, 2014.

Besides replacement of the assignment model that is topic of this research, the second generation BBMA models also adds a departure time choice model as a fifth step in between mode/destination choice and assignment (Zantema et al, 2016). This increases the number of feedback loops and iterations between the assignment and demand model, and thus increases the need for an assignment model with low computational requirements.

This research project did not allow for recalibration of the demand matrices using the new assignment model, whereas in practice this would be needed given the different sensitivity of the new assignment model, due to the
incorporation of the physical effects of congestion. Therefore, in this research only comparisons between the STA and the new model are made, both using uncalibrated demand matrices. Based upon these comparisons, we can draw conclusions on the performance (realism, convergence, speed, and scalability) of the new assignment model in a realistic context (overall the uncalibrated matrices do represent a realistic level of demand), but no conclusions can be drawn based on the absolute assignment outcomes throughout this paper.

2 WHY STAQ?

In this section, the choice for STAQ in this study is motivated, by comparing the assignment model to STA models (paragraph 2.1) and Dynamic traffic assignment (DTA) models (paragraph 2.2) in a strategic modelling context.

2.1 Adding physical effects of congestion

The audit referred to in section 1 revealed a major disadvantage of traditional STA models: they cannot describe the physical effects of congestion (flow metering and queue formation). This is caused by their link cost function (the relation between link flow and link cost (e.g. travel time)), which does not correspond to empirically supported traffic flow theory describing the relation between flow, speed and density as is explained below using Figure 2.

![Figure 2: measurements and fitted relation between speed and flow on a typical Dutch highway (left), and relation used within STA model (right – dashed line) and STAQ (right – blue line)](image)

In the left part of Figure 2, the relation between speed and flow as could be measured on a loop detector on a typical Dutch Highway is displayed. Each black dot represents an observation of speed and flow and the colored lines represent curve-fitted relationships. The right part of Figure 2 displays the relationship that is assumed for the same type of road section in a STA model (dashed red line) and a macroscopic (quasi) dynamic assignment model (blue line). Considering the observed relationship (left part) it becomes clear that in reality there exists a free flow branch (with high speeds) and a congested branch (with low speeds due to congestion caused downstream) that meet each other around a flow value that is considered to be the road capacity.

The static relationship does not have a congested branch. Instead, although speed does drop, flow may continue to increase well beyond capacity as demand increases. As a result, in static assignment models, congestion effects are limited to changes in route choice due to lower speeds, but can never lead to flow metering or spillback effects in the network. The (quasi) dynamic relationship does have a congested branch causing flow to decrease when...
demand exceeds capacity. In order to maintain consistent, it also describes the relationship with a third variable: the density. This is needed to be able to ‘store’ vehicles in congestion (in the form of higher densities), thereby ensuring conservation of vehicles on a section of road. This relation between flow and density is called the fundamental diagram and, together with the conservation of vehicles law, forms the basis of traffic flow theory described in Lighthill and Whitham, 1955 and Richards, 1956. Most (but not all!) macroscopic dynamic assignment models as well as the quasi dynamic assignment method STAQ are based on this theory and use it to describe flow metering and spillback.

2.2 Avoiding issues with DTA models in a strategic context

As stated in section 2.1, most DTA models can describe flow metering and spillback. DTA models have been topic of research since the early 1980’s. Since then, they have proved that they can be useful tools for both planners and engineers under certain circumstances (TRB, 2011). However, we argue that their application is limited to operational and tactical models for three reasons.

Firstly, DTA models exhibit high computational cost and memory usage and therefore limited scalability. Especially their high computational cost prohibit application on large models (in terms of number of OD-pairs and routes).

Secondly DTA models lack convergence properties needed for applications within the strategic context in which the model is used to compare long term effects of different scenarios (TRB, 2011, Peeta and Ziliaskopoulos, 2001). Researchers and practitioners state that in this context and when using static traffic assignment models, a duality gap (the metric most used to measure the level of disequilibrium) of 1E-04 or lower is needed (Boyce, Ralevic and Bar-Gera, 2004; Caliper, 2010). To the best of the authors knowledge- no DTA algorithms exist that can converge to this level. Various tests and studies by the author confirm that the required duality gap value of 1E-04 also makes sense when using STAQ.

Thirdly, the existence of a time dimension within DTA models means that much more input data such as time dependent OD-matrices (or demand models that can deliver these), traffic counts and route choice parameters are needed. These are often not available, especially for longer term (i.e. scenarios 10-20 years into the future).

3 METHODOLOGY

In this section, the assignment model STAQ is briefly described (paragraph 3.1) along with its input requirements (paragraph 3.2) and a description of modifications to the network (paragraph 3.3) and demand matrices (paragraph 3.4) from a STA model that are typically needed to be able to run STAQ.

3.1 Assignment model: STAQ

STAQ is a quasi-dynamic traffic assignment model that has been developed to overcome the problems of DTA and STA models described in section 2. STAQ
explicitly captures flow metering and queue formation due to congestion (just like DTA models do), but assumes stationary demand during a single time period (e.g. a whole peak hour, just like STA models do) and is therefore much more scalable and mathematically tractable. Furthermore, STAQ has proven to converge to the required level and does not need any more input data than a STA model does. STAQ is implemented as a propagation model within the StreamLine framework in OmniTRANS transport planning software where it is intended to be used for large scale urban transport models containing both freeways as well as urban road sections.

It uses a concave, two regime fundamental diagram for the relation between speed, flow and density on the link level and an explicit node model to describe interaction of flows on nodes. It can be used with different route choice models, but in this study the multinomial logit model is used with scale parameters set to one over 14% of the max route cost of the considered OD pair. It can be run with an additional junction modelling component, taking into account capacity and delay effects on the level of turning movement as a result of traffic rules, geometry and/or signal schemes on the junction. Furthermore, it can be run as a multi user class assignment, where each class has its own route choice parameters, free flow speed and set of network restrictions.

The combination of assumptions behind STAQ has consequences for its usage and the interpretation of its outcomes. Its hard capacity constraints and explicit node model can lead, contrary to static assignment models, to residual traffic: traffic that cannot reach its destination within the studied period as it got stuck in queues. Its lack of a time dimension means that all results should be interpreted as averages over all travelers departing in the study period. Furthermore, it forces the modeler to make assumptions on the network state before and after the study period, as there are no warm up or cool down periods to take care of this. Instead, just like static models implicitly assume, STAQ assumes an empty network before and after the study period. However, all travel time (and contributions to density and flow) of traffic that departed within the study period is accounted for in the average outputs, also when part of a trip takes place after the end of the study period (the latter cannot occur in static traffic assignment models).

Note that the STAQ algorithm is not described here; the interested reader is referred to section 4 of Bliemer et al (2013).

### 3.2 Input requirements for STAQ

STAQ needs relatively little extra input compared to static traffic assignment models. As such, it can be relatively quickly implemented in study areas for which an existing static traffic assignment model is available. Therefore, in this section, we first describe model input required for static traffic assignment models, and then describe the additional input required for STAQ.

In static assignment models, model supply is described by a digitized network (graph) of the study area consisting of nodes, directed links and centroids (the latter aggregately represent the locations of origins and destinations of trips). Required link attributes are the free flow speed and the theoretical link capacity,
whereas centroids do not require any attributes. Only for nodes where junction modelling is desired attributes are required. For these nodes, attributes describing the properties of the junction (junction type, approach and exit lane configuration and dimensions and optionally the traffic light schema) need to be defined. Model demand in static traffic assignment models is described by an OD matrix containing stationary travel demand between all origins-destination pairs in the network during the study period.

STAQ needs the same network input, but additionally needs the jam density (which is considered a single, network wide parameter) and optionally the critical speed for all links (only when a non-triangular fundamental diagram is assumed). Note that these extra data requirements can be easily measured or derived: the critical speed can be derived from free flow speeds and jam density can be derived or assumed based on the average car length and headway in a non-moving queue. Further note that STAQ does not need a link typing, since all link characteristics are derived from the fundamental diagram. On the demand side, STAQ uses a single demand matrix per class, just like STA models.

### 3.3 Modifications to STA model network

On the supply side, although STAQ makes direct use of the network from an existing STA model, its hard capacity constraints cause the model to be less forgiving for erroneous or coarse network data that usually stays unnoticed in the network of the STA model. Typically, there are three types of network errors that need to be solved.

Firstly, STAQ (and DTA models) may produce unrealistic congestion or even gridlocks due to too low link capacity values. These errors often go unnoticed on short links in networks of STA models, since the travel time penalty given for such links is marginal by definition. Typically this happens on buffer spaces in front of large intersections that are often simply not coded in the network of the STA model. In the BBMA network, the capacity needed to be increased on 29 links (red links in figure Figure 3).

Secondly, when using junction modelling, the hard capacity constraints make it necessary to model junctions integrally using a single node, not as an ‘expanded node’ (a combination of several links and nodes representing one junction). In static assignment models, this is sometimes being done at large signalized junctions and roundabouts to maintain network consistency with e.g. environmental models. Although not correct, the error introduced in the static context is relatively small, because only the turn delays calculated by junction modelling are used. Since delays are additive and travel times are only to be used for route choice within the model, the induced error is sometimes traded off for network consistency. However, because STAQ also uses the turn capacities calculated by junction modelling as hard capacity constraints this trade off can no longer be made. Because capacity is not additive each path using the junction will only be affected by the first turn on the path that forms an active constraint. If this is a turn on a node originating from an ‘in junction’ link a queue will form on the ‘in junction’ link, whereas in reality this is would be
prohibited (on signalized junctions), impossible (on junctions without mid verges) and/or would only occur when a queue formed downstream of the junction spills back onto it. In the first two situations, a queue that in reality would form on the upstream links of the junction is modelled on the junction itself, potentially blocking other turns on the junction. Because the ‘in junction’ links are relatively short, spillback on these links occurs rapidly causing almost instant gridlock on the junction, whereas this would not happen in reality. Note that this sensitivity is also present in DTA models. In the BBMA network, only one expanded intersection has been adjusted to a single node (blue node in Figure 3).

Thirdly, capacity refinements are needed when in reality the effective capacity is structurally lower than the theoretical link capacity. This mostly occurs on merges, weaving sections, highway intersections and slip lanes. These effects are often omitted in networks used by the STA model, because the model is not that sensitive to capacity reductions in the order of magnitude of these cases (around 20%) anyway. In the BBMA network, on 131 links the capacity was decreased (green links in Figure 3).

Besides these errors and adjustments, an automated check needed to be done for links far outside the study area that where coded with ‘infinite’ capacity (in STA models often coded as e.g.: 99999), which cause a critical density much larger than the network wide jam density (which was set to 170 veh/km/lane) and thus an inconsistent fundamental diagram. For these links, the capacity per lane was set to a default value and the number of lanes was increased instead to attain ‘infinite’ capacity.
3.4 Modifications to STA demand matrices

On the demand side, just like the STA model, STAQ requires OD matrices that describe some stationary (average, peak, nth percentile) demand within the study period. However, the hard capacity constraints make it necessary to define the stationary demand matrix more explicitly: it should contain all the traffic that chooses to *depart* in the study period, no matter whether it reaches its destination within that study period. This means that when using traffic counts to calibrate the OD matrix, flow metering and spillback effects of congestion should be somehow taken into account. This is something that is usually *not* accounted for in matrix estimation procedures for static traffic assignment models, since these models cannot reproduce these effects anyway. For DTA models it is common to use matrix estimation methods based on the simultaneous perturbation stochastic approximation (or SPSA) by Spall, 1998). While these methods have extensively been used for DTA models (see e.g. Tympakianaki et al, 2015 and Antoniou et al, 2015 for recent examples), to the best of the authors’ knowledge, they have not been applied in the strategic context yet. As described in section 1.2, in this research we circumvent this problem by using the uncalibrated demand matrices and only compare results to the STA model results using the same matrices.

In order to decrease calculation times, bucket rounding was applied to the OD matrices to decrease the number of OD pairs with demand, and thus the number of routes to be evaluated by STAQ. Bucket rounding is a method that merges cells that contain less trips than some threshold value by redistributing them over the destinations. Because multiple interpretations of bucket rounding exist, pseudo code from the version that was used for this research is displayed in Figure 4. The threshold value was set to 1, motivated by the fact that the level of uncertainty of the destination choice model that determines the original OD matrix (a gravity model) is much larger than that. Bucket rounding yields a much sparser matrix, but has negligible effect on the link flows, because the row totals and the grand total are guaranteed to deviate less than the threshold value from the original matrix. Analysis of the resulting trip length distribution and link flows confirmed that only marginal differences occur.

```
Initialize: Threshold = 1.0 ; inBucket = 0
For each cell (OD-relation)
    If 0 < cellvalue < threshold
        inBucket = inBucket + cellvalue
    If inBucket > threshold
        newcellvalue = threshold
        inBucket = inBucket - threshold
    else
        newcellvalue = 0
    End if
End if
```

*Figure 4: pseudo code of the bucket rounding procedure used in this research*
4 RESULTS

In this section, some of the model components that can be used within STAQ are described (paragraph 4.1) and tested on convergence and realism (paragraph 4.2), yielding an ‘optimal’ methodology that is used in the remainder of this paper. In paragraph 4.3 realism of this optimal methodology is demonstrated using a case study, whereas scalability of the method is examined in paragraph 4.4.

4.1 Tested model components

In this research, the following three model components that influence the realism, convergence, computational cost and/or scalability of the assignment model where tested.

Averaging scheme

To reach user equilibrium, STAQ uses an iterative process that distributes traffic over different routes, according to differences in route cost (travel times). This process yields a distribution over routes built up from a weighted average of the distributions of all previous iterations. The method that determines the iteration specific weights $\alpha_i$ is referred to as the averaging scheme. The simplest averaging scheme tested here is the well-known method of successive averages (MSA) which uses $\alpha_i = 1/i$, implying that all iterations are of equal importance. The other method tested is the self-regulating average (SRA, Liu et al, 2009) that uses $\alpha_i = 1/\beta_i$, where $\beta_i = \beta_{i-1} + \Gamma$ or $\beta_i = \beta_{i-1} + \gamma$. In each iteration, one of the formulae for $\beta_i$ is chosen based upon a comparison of the level of disequilibrium (in terms of ‘excess’ vehicle hours) of the latest iteration with the previous iteration. Whenever an iteration proves to be converging, it gets assigned a large weight by increasing beta with the very small increment of gamma, in this research set to 0.01. Whenever an iteration proves to be diverging, it gets assigned a small weight by increasing beta with the relatively large increment of kappa, in this research set to 1.9. SRA is a more intelligent averaging scheme because it uses information of previous iteration(s) and is expected to perform better than MSA.

Spillback

Queues are primarily caused by bottlenecks in the network. In the model, a bottleneck is defined as a link or turning movement that has a capacity that is lower than the demand for this link or turning movement. When this is the case, a queue will form with its head on the upstream node of the link or the node being traversed by the turning movement. Such a queue could spillback onto upstream links causing secondary bottlenecks. Calculation of these spillback effects can be turned off within the STAQ algorithm, in which case only vertical queues on the bottleneck nodes are modelled. Running STAQ without spillback will improve convergence, at the expense of realism around heavy bottlenecks where spillback occurs in reality. Note that it is also possible to run STAQ without spillback effects until equilibrium is reached, and then perform a post processing spillback run to translate the vertical queues into horizontal queues. In this case, the route choice is based on delays assuming vertical queues whereas the effect of horizontal queues are included in the final link flows, speeds and densities.
Junction modelling

The effect of traffic rules, geometry and/or signal schemes on turn delays and turn capacities over junctions can be taken into account. Optionally, only turn delays can be taken into account, or junction modelling can be turned off altogether.

4.2 Convergence and computational cost

In order to evaluate the effects of the tested components, 100 iterations where run on the BBMB model (3321 centroids, 142,000 links, 107,000 nodes, 16,000 junctions) for all twelve possible combinations of the tested components. The results are summarized in Figure 5 which shows the convergence (duality gap value corrected for stochasticity in the multinomial logit route choice model) in relation to the calculation time (in hours) of these runs (using a machine with Intel Core i7-4770 CPU and 32GB ram). Each curve represents a run, its color and shape indicate the combination of model components tested as displayed in the legend. The reds represent runs using the MSA averaging scheme, the greens represent runs using the SRA averaging scheme. Dashed curves represent runs where spillback is enabled, and continuous curves represent runs without spillback. The three different shades of both reds and greens represent the three different options for junction modelling (see legend in which ‘NoJM’ means no junction modelling; ‘Delays’ means that only turn delays are taken into account and ‘JM’ means that both turn delays and turn capacities are taken into account). Besides showing the trade-off between computational time and convergence, the total computational time needed to do 100 iterations can also be derived from Figure 5 by looking on the vertical axis at the point where the curve stops.

Figure 5: convergence (relative gap corrected for stochastic route choice model) vs calculation time
Recall from section 2.2 that the duality gap should be lower than 1E-04 for the assignment mode to produce outcomes that can be used in the strategic context. Then, from Figure 5 we conclude that all runs with spillback do not converge sufficiently: the duality gap keeps oscillating and never drops below 1E-04. Spillback effects are thus the most important cause for non-convergence, which makes sense when realizing that spillback is likely to cause the route cost of routes that use link(s) affected by this spillback to become diagonally non-dominant (i.e.: the demand for such a route itself is no longer the main contributor to its cost; instead demand on other routes are), whereas the route choice model and averaging scheme do not anticipate for this.

Figure 5 also shows that all runs without spillback converge sufficiently. However, there are large differences in the required computational time, mainly due to the averaging scheme. Considering the runs with full junction modelling, MSA converges to 1E-04 in 80 iterations (2:14 hours), whereas SRA reaches the same level of convergence (sustained) in only 32 iterations (0:54 hours). When removing junction modelling, SRA converges sufficiently in 12 iterations (0:19 hours) and proves to be able to converge even towards machine precision. Although this high precision result is quite impressive, for practitioners it is not important, and removing junction modelling is not an option, given the level of urbanization in Noord Brabant.

In Figure 6, the average calculation time per iteration for all twelve runs is displayed. Comparing the two averaging schemes, it becomes clear that SRA is only marginally more computationally expensive, whereas Figure 5 showed that convergence is significantly better. Adding spillback does add 15-20% calculation time, whereas adding turn delays and full junction modelling adds 4-10%. Note that these increases are primarily caused by the extra calculations needed to determine spillback effects and junction delays and capacities. On top of that, adding junction capacities will also result in slightly more internal STAQ iterations within each route choice iteration, because there are more (potential) bottlenecks in the network for which interaction effects need to be made consistent (see algorithm in section 4 of Bliemer et al, 2013).

<table>
<thead>
<tr>
<th></th>
<th>MSA</th>
<th>SRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>no spillback, no JM</td>
<td>1:32</td>
<td>1:34</td>
</tr>
<tr>
<td>no spillback, turn delays</td>
<td>1:38 (107%)</td>
<td>1:38 (104%)</td>
</tr>
<tr>
<td>no spillback, JM</td>
<td>1:41 (110%)</td>
<td>1:42 (108%)</td>
</tr>
<tr>
<td>spillback, no JM</td>
<td>1:50 (120%)</td>
<td>1:50 (116%)</td>
</tr>
<tr>
<td>spillback, turn delays</td>
<td>1:59 (130%)</td>
<td>1:58 (125%)</td>
</tr>
<tr>
<td>spillback, JM</td>
<td>1:53 (123%)</td>
<td>2:01 (128%)</td>
</tr>
</tbody>
</table>

Figure 6: calculation time per iteration (average) for all runs

Based on the results described in this section and analysis of the realism around intersections (not described here), assignment using SRA, full junction modelling and spillback as post processing was chosen as the optimum methodology and as such is used in the remainder of this paper.
4.3 Realism

As mentioned in section 1.2, OD matrices where not be calibrated within this research. Therefore, the effects of the new assignment model with respect to realism of the assignment results will be demonstrated using a case study in which we compare results of a network variant using uncalibrated demand matrices. The case study focusses on a bottleneck location in the AM peak period: the A59 freeway from Den Bosch towards Oss around the off-ramp Rosmalen (indicated with the black circle in the left part of Figure 7). In the reference situation, the STAQ results (right side of Figure 7) show a vertical queue between the off and onramp and a second, much smaller, vertical queue at the end of the off-ramp, together causing a queue spilling back all the way to highway intersection Empel (the upper left of the network cut out area displayed in the figures), whereas the static results only exhibit minor speed drops directly on the bottleneck links.

In the network variant, the capacity of the intersection at the end of the southern off-ramp is increased and an extra lane between the southern off- and onramp is added, yielding the assignment results displayed in Figure 8.

Based upon the assignment results the following conclusions are drawn with respect to the effect of the network variant:

1. The two bottlenecks around the off-ramp are effectively removed in both the static assignment and STAQ;
2. As a result, the onramp itself and all arterial roads towards it are used more (higher flows), both in the static assignment and STAQ;
3. Another result is that southbound traffic crossing the A59 returns from alternative routes to the arterial that uses the intersection with the considered off-ramp, both in the static assignment and STAQ.
4. On the A59, the queue spilling back from the considered bottleneck towards the northwest is much shorter. Furthermore, due to increased flow from this direction, a new bottleneck is activated at the merge of the highway intersection Hintham (just west of the original bottleneck location). These effects are only visible in the STAQ assignment results.

5. On the A59, downstream from the removed bottleneck, the existing bottlenecks intensify, and a new small bottleneck activates at the next off-ramp. These effects are only visible in the STAQ assignment results.

Comparing the static assignment with the STAQ results we conclude that only the direct effects and some of the route choice effects of the network variant are captured by the static traffic assignment, whereas STAQ also captures the indirect effects both up- and downstream from the removed bottleneck. This clearly demonstrates that the addition of flow metering and spillback effects strongly improves the realism of the qualitative assignment results. Quantitative comparisons of STAQ output with observed congestion patterns, speeds and travel times can be found in Wismans et al, 2012, Brederode et al, 2014 and Possel, 2015.

4.4 Scalability

To assess the scalability of the model, calculation times and memory usage of several runs on the regional model of Breda (6,000 centroids, 150,000 links, 108,000 nodes, 16,000 junctions) were monitored. STAQ assignments using SRA, full junction modelling and post processing spillback where run until convergence (duality gap of 1E-04) for three different study periods, each with their own routeset and for both the single class (person car equivalents) and the double class (cars and freight) cases. The results are displayed in Figure 9.

![Figure 9: calculation time and peak memory usages of several runs on the regional model of Breda](image)

The table reveals that convergence during the AM peak is more difficult to reach (23 or 24 iterations needed) compared to the PM peak (12 or 13 iterations) and the off peak period (9 iterations). This is caused by the number and severity of bottlenecks. Note that the (set) of normative bottlenecks determines convergence of the whole model. This means that the AM peak is not necessarily more congested network wide, but it contains certain bottlenecks that cause the need for more iterations. This could be explained by the fact that peak spreading in the AM peak occurs much less compared to the PM peak.

© AET 2016 and contributors
Furthermore, it can be derived from this table that memory usage is proportional to the number of routes, which is on its turn is roughly proportional to the number of classes (at least in this case, where route sets are not shared among modes). On average, the peak memory usage per route is around 3 KiloBytes, which translates to around 3GigaBytes needed for every million routes. Calculation times are proportional to the number of iterations times the number of routes. On average, the calculation time per route per iteration is around 0.2 milliseconds, which translates to about 3 minutes per iteration for every million routes.

Note that the calculation time needed per iteration per route found in this study is very similar to earlier tests (i.e. Bliemer et al, 2013, where 4 different models where examined and Brederode, 2014, where 7 different models where examined) and thus is relatively model-independent. The peak memory usage however proves to be much more variable over the different models, which can be explained by varying network size and level of detail causing the average number of links within each route to vary.

5 CONCLUSIONS AND RECOMMENDATIONS

This paper described how the BBMA models can be improved using the quasi dynamic assignment model STAQ which adds physical effects of congestion (flow metering and spillback) to the traffic assignment model, while avoiding limited scalability, poor convergence properties and heavy input data requirements of DTA models.

The model can be run using the network and demand matrices from an existing STA model, and does not require additional data gathering. However, capacity values of most STA networks must be refined, and any matrix calibration method should account for physical effects of congestion in order to get realistic results. In this research, demand matrices where bucket rounded to decrease the required computation time. The effect of bucket rounding proved marginal and falls well within the uncertainty level of the demand model.

Using STAQ combined with the self-regulating averaging scheme, post processing spillback and full junction modelling, convergence to a duality gap value of 1E-04 was attained on the BBMB model within an hour of calculation time; which is well acceptable for application within the demand model. Using a case study clearly demonstrated the improvement to the realism of the assignment results.

Using STAQ in the same configuration on the regional model of Breda yielded an average of 3 Gigabyte of memory and around 3 minutes of calculation time per iteration was needed for every million routes. Calculation time is rather model-independent, whereas peak memory usage may vary (between 1.25 and 3 Gigabytes) across different models.

Furthermore, the following conclusions were drawn based on general experiences during the research project as a whole. Firstly, the assignment method has proven to perform very well. It has successfully added physical
effects of congestion to two large strategic model systems, and the assignment method itself has even proved to be faster than the original STA model. Secondly, inclusion of physical effects of congestion provides more convincing model outputs: more realistic plots showing correct congestion locations and spillback effects and more (realistic) output variables such as route travel times, queue lengths and percentages of blocked demand. Thirdly, the level of convergence that has been achieved proves that it can truly be used in a strategic context. Based on these outcomes, the province of Noord Brabant decided to fully implement STAQ including the junction modelling component within the second generation of the BBMA model system, which is currently under development.

Some recommendations from this research will be given below. Note that some of these recommendations are already being brought into practice in version 2 of the BBMA model system. Firstly, to the best of the authors’ knowledge, no matrix estimation methods that account for the physical effects of congestion have been proven to be applicable on large strategic networks such as the regional models listed in Figure 1. Such a method needs to be developed, until then, heuristic workarounds (not discussed here) can be applied. Secondly, convergence of the assignment method cannot be guaranteed in very heavily congested networks. In these cases, gridlocks may occur that may result in unrealistically high travel times. To solve this problem, the model system should provide all the alternatives that would be available in reality; i.e. when gridlock prevents a traveller to reach a destination in a reasonable time, the traveller will choose another destination, mode, departure time or may choose to not travel at all. This is one of the reasons why a departure time choice model is being added in version 2 of the BBMA model system. Furthermore, improving the quality of the route set and the route choice model could also improve convergence under these circumstances. Thirdly, within the assignment model, the route choice and network loading component is now faster than the route generation component. Focus for further development of STAQ should thus be on the latter.
BIBLIOGRAPHY


Gorris, T, Snelder, M, Minderhout, M (2014) *Audit BrabantBrede ModelAanpak (BBMA) – rapportage van bevindingen* (TNO-rapport R11318), TNO, Delft


Zantema, K. Heynickx, M., Brederode, L.J.N., Koopal, R.J. Time of day and demand equilibrium in large scale models, *to be presented at the European Transport Conference 2016*