The relation between data quality and traffic management investigated for a simple route choice model

G. A. Klunder *

Transport and Planning,

Technical University of Delft, The Netherlands

Henk Taale

Transport and Planning,

Technical University of Delft, The Netherlands

***** Email: mailto:g.a.klunder-1@tudelft.nl

1 Introduction

Dynamic traffic management and information is used by road operators in order to improve network utilisation, safety or the environment. Examples are influencing the traffic flow by influencing speeds, lane use, route choice or merging operations by employing variable message signs (VMS), Dynamic Route Information Panels (DRIPs), ramp metering etc. In order to operate the measures, to generate traffic information and to choose the best suitable measure, traffic data are required. Accurate, reliable, high quality traffic data is a prerequisite for effective traffic management and information services.

Different data sources are available, such as loop detector data, floating car data (FCD) from GPS or GSM, blue tooth etc. Each data type has its own characteristics and quality. The required quality for a dynamic traffic management (DTM) measure or traffic information service differs, depending on the type of measure or information needed. Some measures are more time critical than others, while also the required accuracy requirements may differ. However, good research to establish requirements for the quality of traffic data in relation to the intended traffic management goals is lacking, while more and more new traffic data is coming available and the demand for reliable traffic information is increasing. Therefore more research on this subject is needed.

In this paper, the relation between the accuracy of traffic counts and the effect on route choice and resulting delays is studied. A queuing model is used to calculate queue lengths and delays on two route alternatives, with and without errors on the traffic counts. On a DRIP, the delays over both alternatives are shown, such that a certain fraction of the traffic will change its route choice. This study will show how sensitive this route choice and the effect on the traffic performance is for errors on the input data.

2 Background

An important development concerning collecting and distribution of traffic data in the Netherlands is the National Data Warehouse for Traffic Information (NDW). The NDW is a partnership of 15 Dutch authorities that joined their strengths to provide complete, reliable and up-to-the-minute information on the status of the basic Dutch road network. NDW will become the databank that will collect, process, store and distribute all relevant traffic data. Quality requirements have been defined and imposed to traffic data suppliers. These quality requirements are not differentiated for different road types or traffic management applications. Currently, there are discussions about redefining the quality requirements, especially to differentiate them, because the current quality requirements cannot always be met and will lead to high costs. In [1], a preliminary study was performed on the relation between data quality and dynamic traffic management, however, this research studied only the effect on the resulting information or traffic management measure, not the impact on the traffic system, and they concluded that more thorough research is needed on this. At European level it has been identified that there is a lack of common quality criteria for traffic data and services.The QUANTIS project [2] aimed to provide preliminary insights into the issue. Also in the U.S. it is recognized that the matter of data quality has become more urgent in recent years by the increase of ITS applications and various travel information systems [3].

The process from traffic data to the end user consists of a chain of several steps. It starts with the processing of the data. This includes some basic improvements to the raw data such as complementing missing data and removal of outliers, but also advanced processing techniques for combining different data sources (e.g. Kalman filtering), prediction of the future traffic state with prediction models or simulation models.

After the processing of the data, the next step is to use the data for service provision or traffic management, i.e. showing travel times on DRIPs. In this step, usually algorithms and/or models are used to translate the traffic data into the required output of the service. Finally, the information or traffic management measure will be presented to the end user, which might change its behaviour and therefore have an impact on the resulting traffic system. The aim of traffic management is generally to improve the traffic performance, but when the traffic management measure is controlled by erronuous traffic data, the effect might be adverse. Quality of data is furthermore highly related to costs for data acquisition and processing. Higher quality can be obtained with using more or more accurate monitoring instruments. It is not clear beforehand how higher costs for data acquisition relate to higher service quality at the end of the chain. So in the end, it is a matter of finding the best combination of costs and benefits.

The whole process van be summarized in a flow scheme or in a control scheme, as shown in Figures 1 and 2.

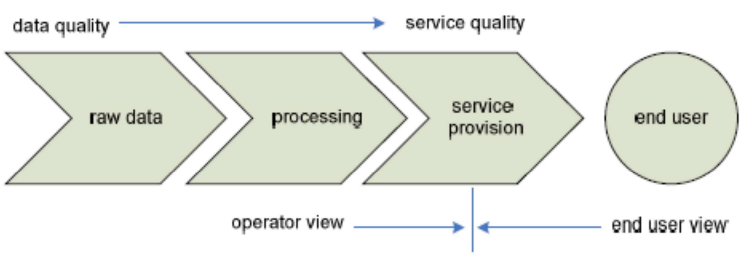

Figure 1: flow scheme from raw data to the end user [2].

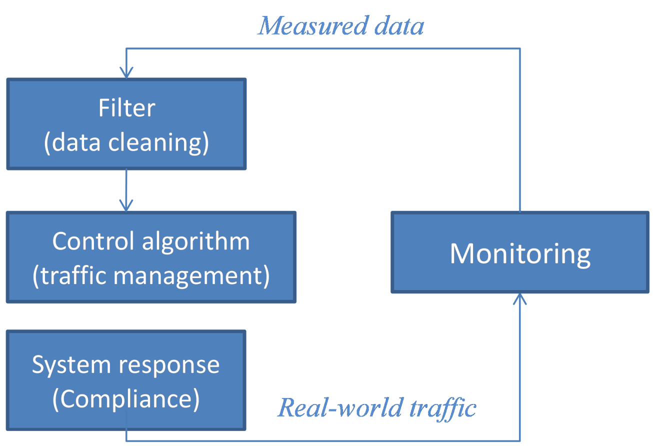

Figure 2: control scheme; the system from measured data to real-world traffic from a control point of view

Finally, next to the quantitative aspects, there are also organizational aspects concerned, because many different parties need to cooperate in order to get access to the different data sources and to implement them. These include private parties who collect traffic data, such as navigation system providers and traffic light manufacturers, and public parties like road operators and traffic management centers. It seems that while data fusion techniques have been developed since the seventies of the previous century [4], still few of them have been implemented in practice. Probably the cause of this is both lacking of good data and organizational problems.

3 Case study: route choice with non-perfect traffic data

This paper will not go into detail on the relation between traffic data quality on traffic management, but will merely illustrate the impact of non-perfect traffic data. For this the simple model proposed in (Hoogendoorn, 1997) is used. In this example, similar to the Pleijroute in Arnhem in The Netherlands, Dynamic Route Information Panels (DRIPs) provide information to the drivers about delays in the network. The network consists of two route-alternatives. For both routes, bottlenecks exist. The bottleneck of route 2 has a small capacity (2300 veh/hr) compared to the bottleneck of route 1 (4300 veh/hr). Additionally, the unconstrained travel time of route 1 is shorter (16 min) compared to the free travel time of route 2 (18 min). Therefore, drivers generally prefer route 1. See Figure 3.

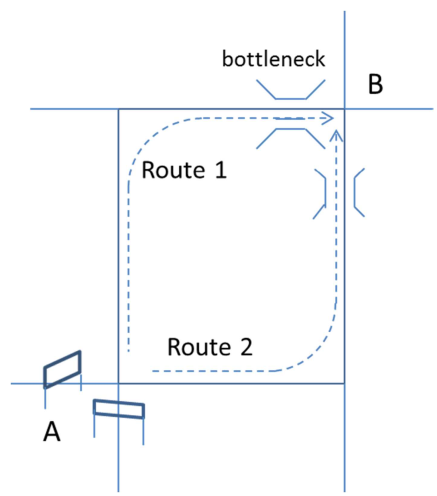

Figure 3: case study route network

A peak period is modelled where the peak flow on route 1 exceeds the capacity, such that congestion will occur. Only part of the traffic on route 1, called *i*, is able to change its route choice, namely those with the same origin and destination of both route alternatives. The rest of the traffic is the autonomous traffic *p*. The initial traffic demand is shown in Figure 4.

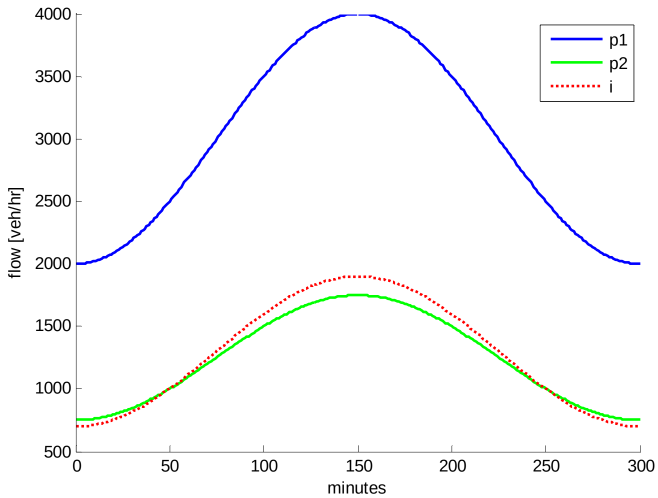

Figure 4: initial traffic conditions

Route choice

It is assumed that part of traffic *i* will change its route choice from route 1 to route 2, based on the difference in delay times of both routes, as originally proposed in [5]:

$$\alpha(t) = \max(0, \min(1, (1 - \beta\Delta T))) \quad (1)$$

$$i_1(t) = i(t) \cdot \alpha(t) \quad (2)$$

$$i_2(t) = i(t) \cdot (1 - \alpha(t)) \quad (3)$$

In which α represents the split fraction for traffic for route 1, β is a sensitivity parameter for the difference in delay between both routes and *i* is the part of the traffic that may consider to change its route.

Figure 5 shows how the route choice depends on the difference in delay times with this assumption.

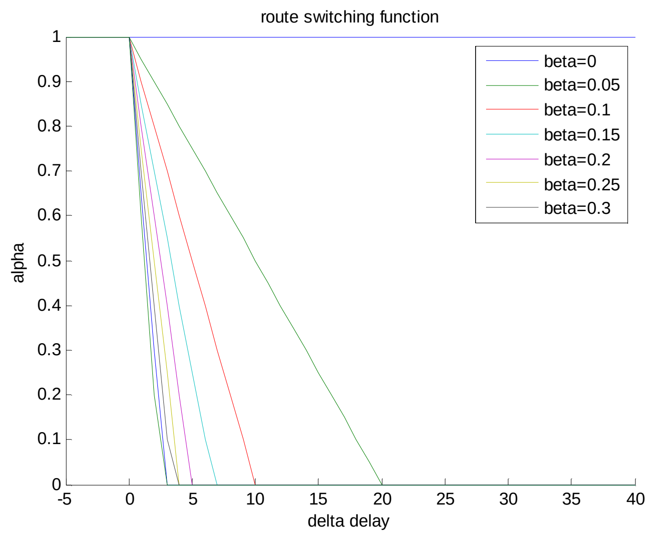

Figure 5: relation between difference in delay times and choice for route 1 as a lineair function

There are few research results on the compliance of drivers concerning route choice information on DRIPs. In a recent study in the Netherlands in which driver behaviour was investigated for delay information on DRIPs based on a questionnaire [6], about 66% of the repsondents indicated that they are considering to change their route sometimes or regularly, based on the information on a DRIP. However, based on an additional group discussion they said that they only consider to do so when the difference is large enough, around 15-20 minutes. Furthermore, only half of them chooses the indicated alternative route. About 72% finds a deviation of the delay up to 5 minutes acceptable. Translating these results into a decision function for route 1 which seems more realistic than the one proposed in [5], using a smooth s-curve and parameter beta = 0.378 in formula 4, this gives the curve as in Figure 6 where 66% switches to route 2 when the delay becomes larger than 15 minutes:

$$\alpha(t) = \frac{1}{1+e^{(\beta\Delta-6)}} \qquad (4)$$

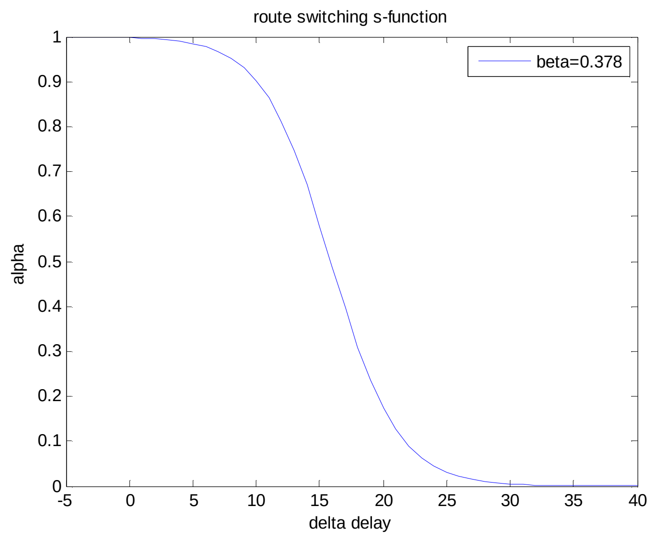

Figure 6: relation between difference in delay times and choice for route 1, using an s-function.

Queuing model

The queuing model to calculate traffic delays is the following:

$$d_i(t) = i_i(t - \tau_i) + p_i(t) \tag{5}$$

$$y_i = \begin{cases} C_i, & d_i(t) > C_i \mid r_i(t) > 0 \\ d_i(t), & \text{else} \end{cases} \tag{6}$$

$$r_i(t+1) = r_i(t) + (d_i(t) - y_i(t)) \cdot dt \tag{7}$$

$$T_i = 60 \cdot \frac{r_i(t)}{C_i} \tag{8}$$

$$F_i = 0.005 \cdot r_i(t) \tag{9}$$

With

d_i: demand of route *i* (veh)

p_i: autonomous traffic of route *i (veh/h)*

τ_i: travel time of route *i* (min)

C_i: capacity of route *i (veh/h)*

y_i: traffic after the bottleneck at route *i*

r_i: number of vehicles in the queue before the bottleneck (#veh)

T_i: delay of route *i* (min)

F_i: queue length on route *i* (m)

In the formula 5 a travel time delay is introduced, because traffic will make the route choice based on experienced travel times as they are shown on the DRIPs.

Measurement error

Subsequently, the traffic counts have been altered with several (random) errors in order to investigate the effect. The traffic input consists of the traffic counts before the bottleneck, for example measured by a loop detector. The measured traffic flow is determined by:

$$\widetilde{d_i} = (1 + e) \cdot d_i, \qquad (10)$$

where *e* is the error, according to some distribution. For example, if *e* equals 0.1, this means that the traffic counts are 10% overestimated. In practice, it is usually unkown how much error there is on loop detectors, since a perfect monitoring system would be required. Loop detectors are normally only tested in a laboratory setting. Investigation at the Centre of Transport and Navigation of the Dutch Ministry of Transport learned that the error on traffic counts is usually in the range of 5%. Every timestep, both the ‘real’ delays based on d_i and the estimated delays based on $\widetilde{d_i}$ are calculated. The drivers base their route decision on the delays based on $\widetilde{d_i}$, while the traffic performance is estimated as the total delays based on *d*:

$$T_{tot} = \sum_t (T_1(d) + T_2(d)) \qquad (11)$$

4 Case study results

Figure 7 and 8 show the results of the total delay where a fixed (not random) error is modelled on the traffic counts of either route 1 or route 2, with several different beta values and the first (linear) route choice model.

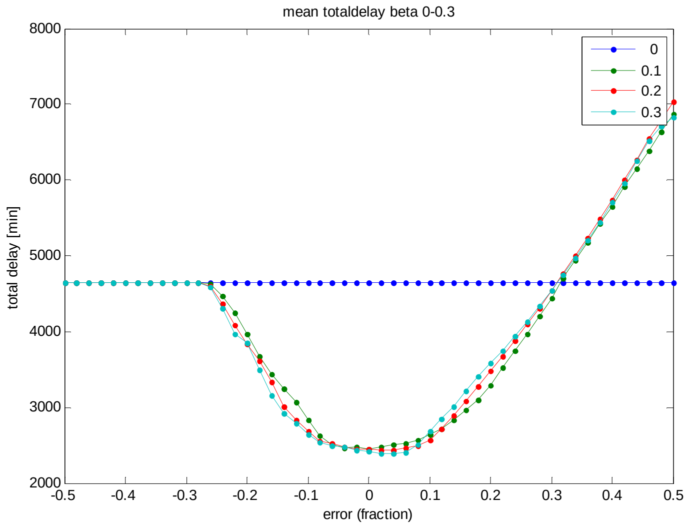

Figure 7: total delay for a range of errors on route 1 for beta = 0, 0.1, 0.2 and 0.3

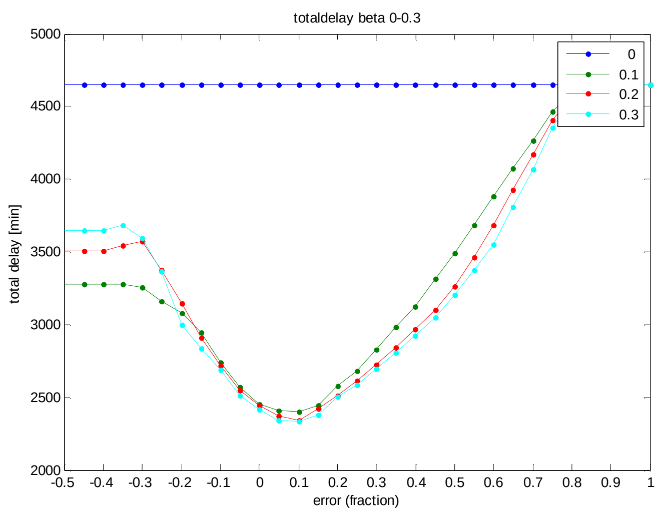

Figure 8: total delay for a range of errors on route 2 for beta = 0, 0.1, 0.2 and 0.3

The results show that when there is no information such that drivers will not change their route, the total delay of all vehicles on both routes is 4650 minutes. In the optimal situation with no error on the traffic counts, the total delay is 2350 minutes, an improvement of 49%. When an error is added to the measured traffic counts, the displayed delays are either underestimated or overestimated. As can be seen in Figure 9 below, for an error of -10% on route 1, the measured delay is only about 1/3 of the real delay.

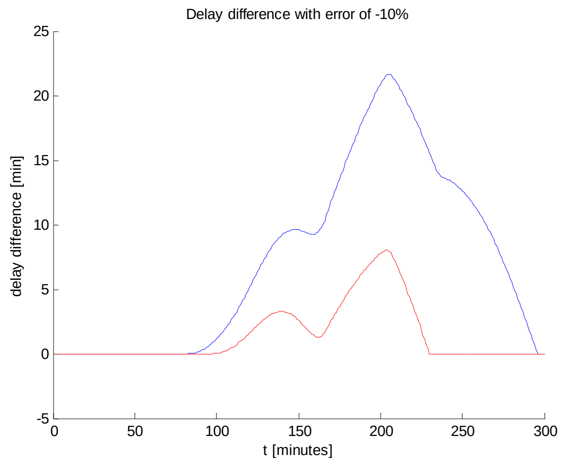

Figure 9: real delay (blue) and estimated delay (red) when the traffic counts on route 1 are underestimated with 10%

For errors smaller than -30% on route 1, no delay is measured, such that drivers will not change their initial route choice. Therefore the situation will be the same as for the initial situation with beta = 0. On the other hand, when the traffic on route 1 is overestimated, more congestion is detected on route 1 than in reality. More drivers than necessary will divert to route 2. If the error on route 1 becomes larger than 30%, the total delay becomes worse than without changing, due to congestion on route 2. However, as long as the error stays within 25%, the effect is still positive with regard to no information. This is probably due to the fact that within these limits, still part of the congestion on route 1 is detected.

The impact of an error on the traffic counts of route 2 is smaller than for route 1, as can be seen in Figure 10. This is due to the fact that there is no initial congestion on route 2, and there will only be an adverse effect when drivers will change from route 2 to route 1. For this, the traffic counts on route 2 should be highly overestimated. In case the traffic counts on route 2 are underestimated, drivers will change from route 1 to route 2 which has a positive effect.

It is remarkable that the parameter determining how many drivers will change their route choice (beta) has no big influence. This can probably be explained by the fact that the part of the traffic flow that is able to change its route choice (*i*) is rather small in comparison to the total flow.

The same calculations have been done for each combination of errors within -0.5 and 0.5 on route 1 and on route 2. This gives the result as shown in Figure 10:

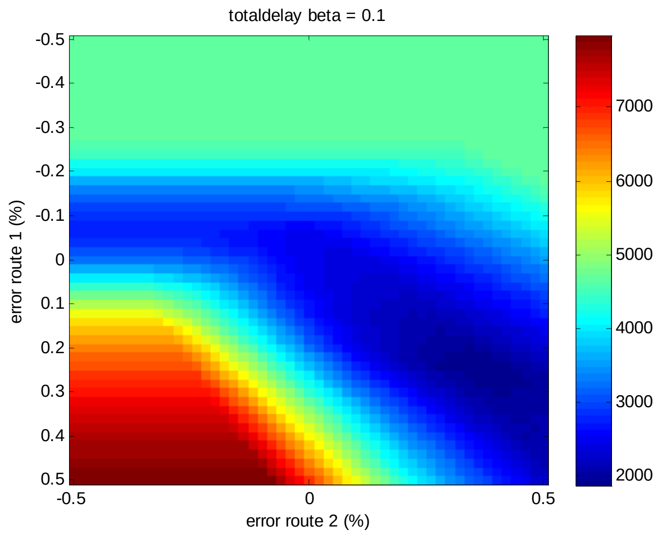

Figure 10: total delay for each combination of errors on the traffic counts of route 1 or 2

This result shows that when there is an equal large error on both routes, the effect is positive (small total delay), which is logical, since when the congestion on both routes is equally overestimated, the delay difference will show correctly more delay on route 1. On the other hand, when the traffic counts are highly underestimated on both routes, the DRIP will not show any delay, such that the situation will remain the same as without information.

Looking at the impact for errors within 10%, the total delay increases with 55% with regard to the optimum, for errors within 4%, the total delay increases with 10%.

If the error *e* is modelled according to a random distribution, the average results are the same, though for a single realization there can be peaks with much higher delays. An example if this is shown in Figure 11, where the error is normally distributed. It can therefore be good to perform an analyses on the distribution of the outcomes relative to the distribution of the error on the input.

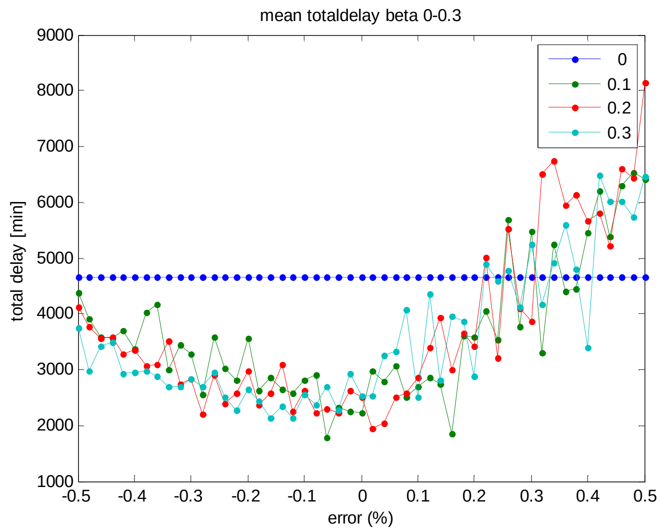

Figure 11: total delay for normally distributed errors on route 1 for beta = 0, 0.1, 0.2 and 0.3

In order to investigate the effect for other situations with different traffic demands, capacities, compliance rates and betas, calculations have been done for several variants as shown in Table 1.

The base case is as described above, with a capacity of route 1 of 4300 vehicles and of route 2 of 2300 vehicles, a peak demand on route 1 of 5900 and on route 2 of 1750 and a full compliance rate, meaning that all drivers of traffic flow *i* might change their route, according to the given decision rule with $\beta = 0.25$.

The next scenario with compliance rate 0.66 means that only 66% of the drivers of traffic flow *i* is considering to change route, in accordance with [6].

The next scenario shows the sensitivity of parameter β for the case when a smaller part of the drivers will change its route choice for the same difference in travel times.

The following scenario with an s-curve decision function uses the function of Figure 6, tuned more towards real behaviour as found in [6] compared to the simple linear function as proposed in [5].

The final two scenarios illustrate the effect of less spare capacity on the alternative route, by respectively lowering the capacity or increasing the demand.

In this table, the improvement is calculated as the relative improvement of the minimum total delay of the scenario with some error on the information with regard to the scenario without information. For example, 0.82 improvement means that the total delay is 82% lower than when none of the drivers changes route. As above, the minimum and maximum allowed errors are calculated as the errors for which the total delay will not increase with regard to the no information case.

Table 1: effects of other scenarios compared to the base case

	Base case	Compliance rate 0.66	Beta= 0.05	s-curve decision function	Capacity route 2 = 1500	Demand route 2 = 2000
improvement	0.82	0.62	0.65	0.57	0.016	0.37
minerror1	<-1	<-1	<-1	<-1	-0.18	-0.24
maxerror1	0,82	0.82	0.82	>1	0.18	0.56
minerror2	-0,26	-0.26	-0.26	<-1	<-1	-0.26
maxerror2	0,3	>1	0.36	0.34	-0.16	0.16

Looking at these results, we see the following:

- a lower compliance rate reduces the positive effect, but for 66% compliance the effect is still substantial (62%). A lower beta has a comparable effect.
- Adapting the decision function to a more realistic s-curve with 66% switching to route 2 when the delay difference is larger than 15 minutes [6], the improvement of the total delay is less, though the allowed errors are larger.
- Decreasing the capacity of route 2 from 2300 to 1500 reduces the positive effect of the DRIP a lot. This is due to the fact that less rest capacity is available on the alternative route. Also much smaller errors are allowed.
- Increasing the peak demand on route 2 from 1750 to 2000 also reduces the positive effect and allows for smaller errors.

5 Conclusions and recommendations

While more and more traffic data are coming available, not much is known about the needed data quality in order to reach the desired goals of traffic management.

In order to show how the accuracy of traffic counts can influence traffic performance when traffic management is applied, a theoretical case study is performed for a case where DRIPs show delays to redirect the traffic over two route alternatives, using a queuing model.

Conclusions

The results show that relatively large errors are allowed without deteriorating the situation with regard to the no information case, especially when the error is comparable on both routes.

Of course the results are highly dependable on the chosen parameters such as the capacity and autonomous traffic demand on both routes. With the given model, the effects for other than the given parameters can be easily investigated. When there is not much spare capacity available on the alternative route, the effect of the DRIP is smaller and smaller errors on the input data are allowed. However, without deteriorating the base case, still errors of 16% are allowed.

Recommendations

The presented case study is a simplification from reality, since it assumes that delays can be estimated with a queuing model based on a fixed bottleneck capacity. Also it assumes that the delays as shown on DRIPs can be estimated with this queuing model. In practice in the Netherlands, DRIPs show instantaneous travel times, which cannot be estimated with this model since it doesn't contain speeds.

Therefore it is recommended to do a comparable study based on more realistic traffic flows and traffic data, such as can be obtained from loop detectors or a micro simulation model.

Finally, since the relation between quality of traffic data and traffic management depends on the type of traffic management measure, a similar study should be done for other types of measures as well.

Acknowledgement

This research is based on an idea of prof. Serge Hoogendoorn and jointly funded by TNO, Netherlands Organisation for Applied Scientific Research, and TrafficQuest, a joint collaboration between TNO, Delft University of Technology, and Rijkswaterstaat, highway agency of the Dutch Ministry of Infrastructure and Environment.

References

[1] Tampère, C., Corman, F., Himpe, W., “Invloed van datakwaliteit op Dynamisch Verkeersmanagement”, Katholieke Universiteit Leuven, 2011

[2] Öörni, R., Innamaa, S., Kulmala, R., Kellermann, A., Ebner, R. and Newton, D., “Evaluation report for optimal data quality in selected European service cases”.

Deliverable 6 of QUANTIS - Quality Assessment and Assurance methodology for Traffic data and Information Services. Espoo, Finland, 2010.

[3] Ahn, K., Rakha, H., Hill, D. "Data Quality White Paper". Federal Highway Administration, 2008.

[4] Liu, H., Van Lint, H., Joksimovic, D., Zijderhand, F. and Van Nifterick, W., “Traffic Data Fusion – Traffic State Estimation for Heterogeneous Networks with Missing Data”, ATMO report. Deliverable D3d-3 (29-01-2007)

[5] Hoogendoorn, S.P., “Optimal control of dynamic route information panels”. IFAC Transportation Systems, Greece, 1997

[6] Heinen, H., Sijprsma, R. Reistijdinformatie op Dynamische Route Informatie Panelen (DRIPs) - verslag van een evaluatieonderzoek onder weggebruikers. Report by Ergo Research Intelligence. Commissioned by Rijkswaterstaat (the Dutch Ministry of Infrastructure and the Environment), 2011.