

This paper is also published in a journal. Please cite instead:
Knoop V.L., J.W.C. van Lint, J. Vries, L. Kester, I. Passchier (2013),
Relationship between application scale and maximum time latency in Intelligent
Transport Solutions,
Transportation Research Records No. 2380, p 1-9

Relationship between application scale and maximum time latency in Intelligent Transport Solutions

V. L. Knoop PhD

Delft University of Technology, Transport & Planning
Stevinweg 1, Delft, The Netherlands
+31 15 278 8413
v.l.knoop@tudelft.nl

J.W.C. Van Lint PhD

Delft University of Technology, Transport & Planning
Stevinweg 1, Delft, The Netherlands
j.w.c.vanlint@tudelft.nl

J. Vries MSc

Delft University of Technology, Transport & Planning
Stevinweg 1, Delft, The Netherlands
mail@jaspervries.nl

L. Kester PhD

TNO – Netherlands Organization for Applied Scientific Research
Oude Waalsdorperweg 63; 2597 AK Den Haag; The Netherlands
leon.kester@tno.nl

I. Passchier PhD

TNO – Netherlands Organization for Applied Scientific Research
Oude Waalsdorperweg 63; 2597 AK Den Haag; The Netherlands
igor.passchier@tno.nl

October 22, 2012

Word count:

nr of words in abstract	207
nr of words (including abstract)	5967
nr of figures& tables	$6 * 250 = 1500$
total	7474

Submitted to the 92th Annual Meeting of the Transportation Research Board

ABSTRACT

Congestion is a major problem in large urbanised areas. Intelligent Transport Solutions aim to reduce this problem. Generally, traffic is monitored using sensors, this data is processed, a traffic state is estimated and a control measure is computed and implemented. The availability and quality of the data and the processing time of the algorithms are possible limiting factors in this pattern. This paper investigates what the data requirements for different traffic control measures are. In particular, we focus on the relationship between the spatial scale of the control measure and the time delay between the measurement and the control action, the latency. A set of 17 applications shows that the larger the spatial scale of the control measure, the larger the latency can be. This relationship can be used to determine the latency quality criteria for a certain application. On the other hand, it can be analysed which type of applications are possible when data with a specific quality can be collected. The latter analysis is carried out for a province in the Netherlands. This analysis furthermore shows the difference between the availability in terms of information and communication technology (ICT) and useful data for applications. In particular, if no vehicles pass, the ICT systems are available, but no realistic measurements are produced.

1 INTRODUCTION

Many large cities face severe congestion. With Intelligent Transport Solutions (ITS), road authorities try to alleviate congestion. They are constantly trying to improve their systems to increase the road capacity without adding new infrastructure. These systems usually are feedback systems: the situation at the road is monitored, the data is processed and based on the processed data a certain traffic control scheme is implemented.

In this interaction between data availability and control schemes, the road authorities can approach the problem from either the data side or from the control side. That would lead to the following questions. First, which data is needed for a particular traffic management application. Second, given a certain availability of data, which traffic management applications are possible. The first question is relevant in the light of rapidly changing information and communication technology (ICT). Road authorities question what ICT hardware need to be installed in order to have enough data.

This paper sheds light on both these questions by relating the amount, type and quality of the data to the applications. We choose to classify the applications based on the spatial scale of their control influence. It turns out that there is clear relationship between the maximum time latency of data and the spatial scale of the application. The remainder of the paper, in particular section 3, shows several applications and their spatial and temporal scale. The authors also searched for papers explicitly relating these two for various applications, but could not find literature on it. This paper aims to fill this gap. At the end (section 5), we will argue that with larger spatial scales it suffices to have data of lower quality also for other dimensions of quality.

In this paper, first, section 2 gives the definitions for scale, latency time and quality of data which we will use later in the paper. Then, we will approach this relationship between latency and scale from two sides. In section 3 we investigate a set of ITS applications and find the corresponding time scale. In section 4 we take the an example case of the available data in the province of Delft, the Netherlands. We critically analyse the quality of the data, including availability, and check, based on the empirical relationship of section 3 which applications are possible with the data.

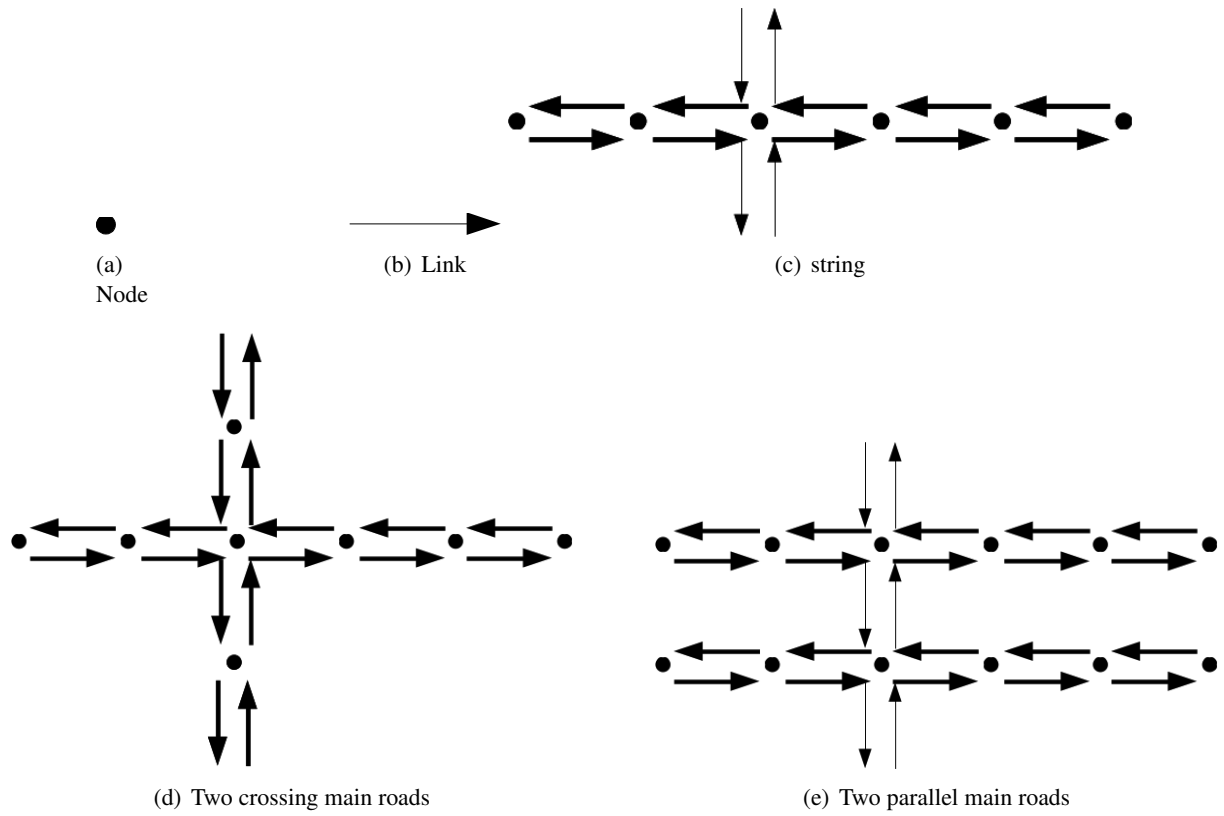
The most straightforward implication of the found relationship is to define quality levels for data (if the application is used as starting point) or to have a a minimum scale at which traffic control can be applied (if data availability is used as fixed starting point). A further implication is that it supports the idea of hierarchical control systems, as for instance proposed by Landman et al. (*1*). In such a control system, the higher level controller governs a larger geographical area. The proposed relationship shows it is feasible to have control at larger areas. Usual constraints of data transmission, processing and storage form less of an issue since at this high level this high-quality information is not needed. This discussion is found in section 5.

2 DEFINITIONS AND MEASURES OF PERFORMANCE

This section defines the levels of space and time, as well as the quality criteria of the data.

2.1 Geographic levels

We distinguish four different spatial levels, which are discussed in this section. The first is the vehicle level. For the higher spatial levels it is useful to code the network into a (functional) directed graph, with road segments represented by links and intersections by nodes (i.e. dimensionless points in space). In the ensuing we further limit the scope of traffic control to control on (major) urban arterials with a speed limit of 50 km/h and above. This limitation is arbitrary but useful nonetheless. It means that on all included roads,

**FIGURE 1 Spatial scales**

control is possible and these are not for instance (private) driveways. It also means if there are more roads included, traffic control can be coordinated over several controlled roads.

We now find the following levels: vehicle, local, string, and (sub-)networks, as defined below.

2.1.1 Vehicle

One traffic control action is directed at one vehicle, and for each vehicle a different control action is determined.

2.1.2 Local

On the local level (being a node – figure 1a – or segment – figure 1b – in the directed graph) one traffic control action is directed at one geographically localized segment (or node). Examples of traffic control on a segment are dynamic speed limitation, lane control, traffic lights or ramp metering installations. For traffic management at this level, it is required that it operates based on sensing information in the direct vicinity of the node (or on segment) and its control objectives pertain to the local traffic conditions (on the node or segment) only.

2.1.3 String

We define a string as a minimum of 2 consecutive elements (nodes/links), see figure 1c. The sensing information may come from any location within the string, and the control actions are coordinated such that a control objective (or a mix thereof) is served over the string. As an upper bound for the spatial scope of a

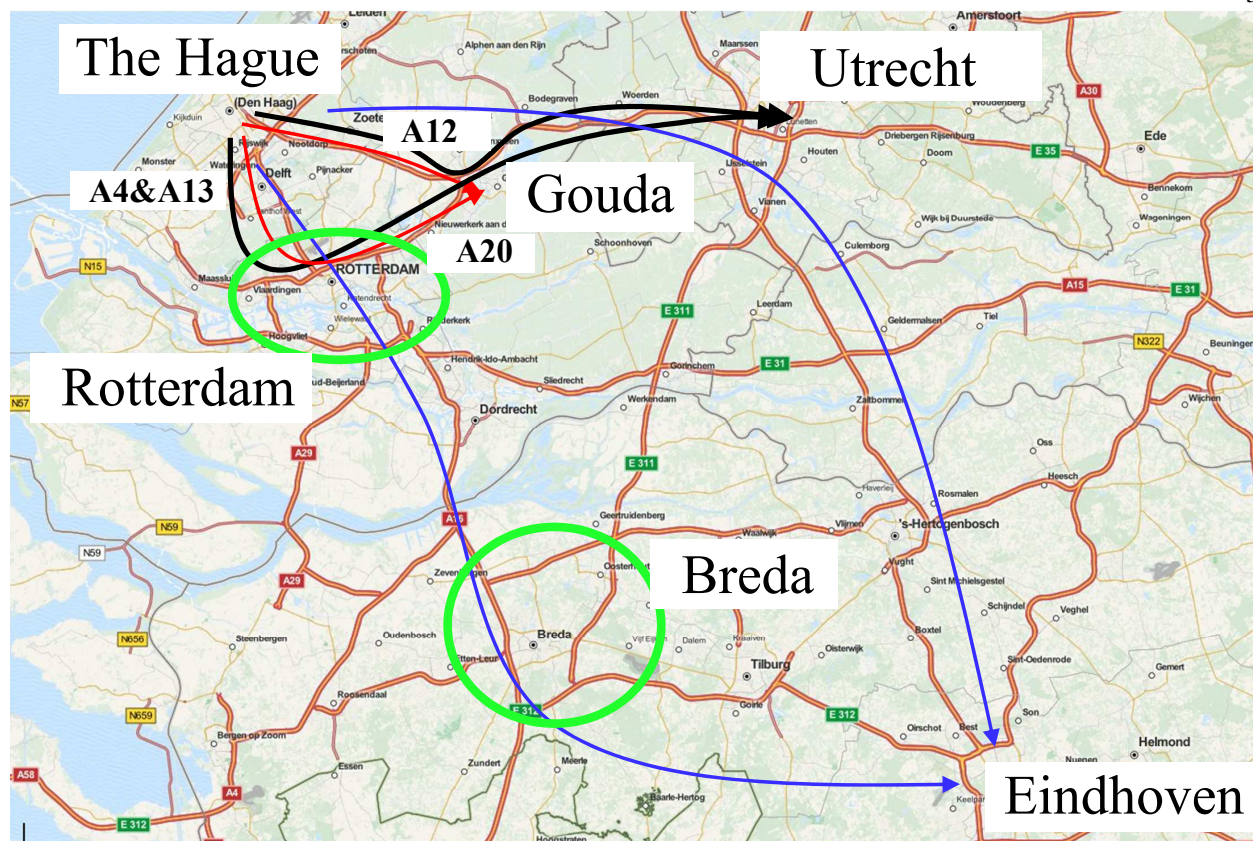


FIGURE 2 An example of overlapping subnetworks at the same infrastructure for the trip from The Hague to Gouda (red arrows) and Utrecht (black arrows). The figure also shows the network scale, for a routing alternative from The Hague to Eindhoven (blue arrows), and subnetworks as part thereof near Rotterdam and Breda (green ellipses). The distance from The Hague to Eindhoven is approximately 130 km.

string we consider a string constituted of one main stream of traffic (indicated with bold links in the figure), and probably several junctions with minor roads. Examples of traffic control on a string are green waves (2), coordinated ramp metering (3), or for example dynamic speed limit applications such as *SPECIALIST* (4).

2.1.4 (Sub-)networks

(Sub)networks are defined essentially as a cluster of a minimum of two strings, see figure 1d and 1e. Subnetworks only have meaning in relation to the network. Subnetworks are the same as networks in their structure, the only difference is that networks are (arbitrarily) considered larger. In a subnetwork, a tradeoff has to be made between at least two traffic streams (e.g., intersection control), and/or routing of traffic is possible.

A couple of clarifying remarks are in place. First, a (sub)network is a functional spatial unit, which is defined dynamically or “on the fly” within a certain control or management application. For example, in the freeway network, two routes from The Hague to Gouda (A4-A13-A20 versus A12, see figure 2) constitute a subnetwork. At the same time, a different (but overlapping) subnetwork of routes from The Hague to Utrecht can be defined. Although these subnetworks overlap and physically refer to the same infrastructure, it is meaningful to consider them as separate spatial entities in different control applications.

Secondly, and related to this first point, in a hierarchical controlled network, one can arbitrarily differentiate between large chunks of the network, and smaller ones and between a very fine-grained representation or a very coarse one. The graph representation allows for combining a cluster of parallel strings (i.e. a subnetwork) into one string or even a segment. A high-level (network-level) operator might use an aggregated description of the network, and control the main streams only. An example of this is the traffic management center deciding on routing traffic from The Hague to Eindhoven, which is possible via Utrecht or via Rotterdam and Breda (see figure 2). Within this network, there can be other, localised optimizations. For instance, there can be detailed traffic management on the advice how to pass Rotterdam or Breda, which then are subnetworks.

An approximate way to test the scale of a control measure is to what would happen if the control measure is performed one scale wrong. For example, if a control action for a vehicle (collision avoidance) is sent to the next vehicle, problems occur. If a control action on routing is sent to the next vehicle, no problems arise. The first control action thus is a control action on vehicle level (or smaller), whereas the second one has a scale larger than vehicle level.

2.2 Time scales

Two time scales can be considered, being the latency and the duration of the control scheme. In the remainder of the paper, we will focus on the latency of the data; nevertheless, it is useful to have a clear distinction between the two time scales.

2.2.1 Latency

A control scheme needs to be computed with input data. Before the data results in a control action, some of the following processes are required: collection, aggregation (hence also awaiting the end of the aggregation time), transmission to a central server, central processing, retransmission, filtering, control action calculation. All these processes take time. The time between the occurrence of an event (e.g., a car passing a detector) and the moment this event is processed in a control action scheme, is called the latency. Note that this time cannot be negative. It is possible to find the best control scheme based on a traffic prediction, i.e. a predicted traffic state for future conditions. The basis for this prediction, however, is measured data. The latency is the time between the measurements of the data and the time an algorithm has completed the computation of a control scheme. A shorter latency is always better, but it is not always required. In case of aggregated data, the latency varies over the aggregation interval. The measurements that are taken at the beginning of the aggregation interval have a longer latency than those taken at the end, since the resulting data will be available at the same time in the form of one aggregated number.

2.2.2 Duration of control action

Another time scale is the duration of a control action scheme. Once the action scheme is computed, for how long will that control scheme be active. This is not the same as the latency, or even the update time of the data. For instance, a controller might use recent data and produce a traffic control action in one minute (latency is one minute). This control scheme can last for a longer time, for instance 30 minutes in order to gradually over this time period improve traffic conditions. This shows that latency is not necessarily the same as the duration of a control action.

2.3 Quality level

Several characteristics define the quality level of data: availability, accuracy and timeliness. These characteristics were identified (5) for the Dutch National Data Warehouse (NDW) (6) and are used to determine

the quality of collected road traffic data. This section discusses the three quality levels.

2.3.1 *Availability*

The availability is defined as the fraction of time in which data is delivered to the end user for an ITS application. This is likely to be equal to the fraction of time that all measuring and communication devices in the sequence from measurement device to end user are working, or the “uptime” of the system. One has to distinguish this availability from an ICT perspective from availability from a traffic and transport application perspective. From the ICT perspective, availability generally will be equal to the uptime of the data collection systems. This is the approach chosen for the Dutch NDW. On the other hand, from the application perspective, only usable measurements will be counted towards the availability. A practical example of this difference is a system which gives average travel times over a section using number plate recognition. Note the system only gives the average travel time per aggregation period –separate inflow and outflow values are not given. Suppose halfway the section there is signal controlled intersection. The signal cycle will cause short periods of time in which vehicles that have passed the entry checkpoint are stopped at the signal, resulting in a situation where there simply are no vehicles passing the exit checkpoint which also have passed the entry checkpoint. During this period, no travel time are measured, so no travel times are produced by the system. The measurement equipment is however still technically functional and can be considered available from the ICT perspective. It will however not produce any usable traffic data and can therefore be considered unavailable from the traffic and transport perspective. This makes the traffic and transport availability a subset of the ICT availability.

2.3.2 *Accuracy*

Data accuracy largely depends on the measurement error of the equipment used and the data aggregation method chosen. The latter is related to the aggregation period. No aggregation at all will provide the most detailed description of traffic, but this level of detail is not necessary for all applications. In order to reduce the amount of data, aggregation of data can be considered. As aggregation introduces inaccuracy, the aggregation method and period needs to be chosen with care. In case of many outliers, aggregation can result in many unusable data points. To illustrate this, consider a measurement section providing the average travel time every minute. If there is a vehicle that has stopped alongside the travel time section for a short period, this will result in an unrealistically high average travel time and as a result an unusable measurement point. On the other hand, aggregation can also improve accuracy, because there is less statistical spread in aggregated data.

2.3.3 *Timeliness*

The timeliness of data is the time between a vehicle passing a measurement site and having the data of this measurement available for use. The latency as described in section 2.2.1 is for a large part influenced by the timeliness of data, which includes the aggregation period of the data. Namely, only after the aggregation interval has ended, measured data become available. Therefore, increasing the aggregation period will automatically increase the latency. Timeliness is furthermore influenced by the length of a measurement section; the longer the measurement section the more outdated the data will be once it becomes available. Speed or flow data from e.g. induction loops are instantly available and result in a negligible delay; the data describes the current situation. Measurements of realized travel time however will be less timely with increasing measurement section length, as they describe an event that took place starting the measured travel time ago.

3 APPLICATIONS AND THEIR SCALE

In case of readily available data, the geographical level, time scale and quality level of the data will determine what ITS applications are possible. The other way round, a particular ITS application will dictate the minimum geographic, time and quality levels requirements for data to be used for the application. This section discusses the implications of these requirements from the perspective of several example ITS applications and from the perspective of an existing road traffic data system. The applications are grouped in three categories, based on the environment they are applied in: highways, urban areas and miscellaneous applications. The applications and their respective time and geographical scales are summarized in table 1.

3.1 Highway

For the highway, several applications have been selected on different time and geographic scales. Several of these can also be applied to freeways. Active Cruise Controls (ACC) keep a preset distance between the follower and the leader and adapt the gap accordingly. Different brands will allow different settings (fixed time or space headway), but whatever setting is chosen, the time to react on the leader's deceleration needs to be in the order of 0.1 seconds (7). On the geographic level this application acts on the vehicle level, as information about specifically the leading vehicle is needed. The quality of this information needs to be very high, as slight deviations may have severe consequences. This implies a very high data availability and accuracy.

A merging assist will help a vehicle to merge onto the freeway. It is needed that it has information on the speeds and position of the surrounding vehicles. This information needs to be exchanged at a frequency where the error between the position and the estimated position based on the last received speed is an order of magnitude less than the size of the required gap; furthermore, this time should be shorter than the time required to perform a complete lane change: 0.1-1 seconds, (7). Data quality requirements are identical to ACC.

In the Netherlands, a system is implemented that provides guaranteed green at traffic signals for approaching trucks (called *Tovergroen*, translated *Magic Green*). Trucks (or other long vehicles) are detected 300 meters prior to an intersection. If the signal controller can prolong the green phase, the truck driver is informed of this so that it can proceed without having to brake needlessly. Evaluation of this system in the Netherlands has shown that not just traffic flow but also traffic safety improved at several test locations (8). The geographic level of this application is the local level, as one intersection is controlled. The time scale is in the order of a second. Decisions need to be made instantly to inform the truck driver in time and to not prolong the green phase needlessly causing unnecessary delays for other directions. High quality data is needed to make a distinction between trucks and other vehicles and to be able to respond in time.

A local measure to avoid traffic breakdowns is to homogenize speeds (9). For this goal, it is needed that vehicles within a 100-1000 meter range have more or less the same speeds. A typical time for this 1-10 seconds, since otherwise vehicles move too far into that area before adapting their speed.

Two similar applications of live traffic data are incident detection and event detection. Both apply to the detection of unusual traffic circumstances. Incident detection is a safety approach aimed at timely warning of approaching traffic or emergency services, including warnings of recurring congestion. Geographically, it operates at a local level, and spatially it operates at time scales of 1-60 seconds, (5); after this, warnings for road users will be too late.

Event detection is a traffic management approach aimed at informing road traffic managers of unusual traffic conditions (10). Time requirements for event detection are in the order of minutes, as even minutes after the occurrence of an event traffic management measures still can be useful. Note that events can also be at the demand side: for example excessive demand from a sports game which has finished. The geographical scale of the management measure taken is a string or a network, whereas the time scale can

be in the order of 1-10 minutes. Aggregated data of lower quality still can produce usable results in case of event detection.

A recently developed algorithm to dissolve stop-and-go waves on grade-separated roads after they started is *SPECIALIST* (4). The duration of the solution scheme is several tens of minutes, but the accuracy should be in the order of 1-10 minutes. The application operates on the string geographical level, where the location and length of the string is assigned dynamically depending on the location and severity of the congestion.

Another application is rerouting traffic over the network (11). This is also useful once unusual conditions are detected, as follow-up application for traffic management. Traffic data can be used to determine if an alternative has sufficient capacity left and in case of multiple alternatives which one is the most suitable given the circumstances. A delay of several minutes and low quality data can still produce usable results. The applicable geographic scale is the (sub)network.

3.2 Urban

For urban applications, we start at a small scale with a protection system to prevent collision with vulnerable road users (12). This system, present in some modern vehicles, will automatically stop the car if it is about to hit an object. Like ACC, the time scale for this type of collision avoidance system is less than a second, working on the vehicle geographic level.

Another application in urban areas is the indication of the time a traffic light will switch to green, or to red. Whereas sub-second accuracy or data updates are not necessary, it is required to have this indicator on a time scale of one second, which is the practical time scale at which timing schemes are adapted. As information from the traffic signal controller is used for this application, the same local geographic scale applies as would apply for traffic signals. This can be implemented for car-drivers (13) and/or for pedestrians (14).

If traffic lights are coupled over different intersections, and green waves (2) are introduced, coordination is required. The application works at the string geographic level and the applicable time scale depends on the distance between intersections, but in any case should be no longer than a minute.

Urban traffic routing is similar to the previously discussed routing on major roads, but is generally only aimed at recreational traffic and considers parking space availability more than congestion; drivers can be informed about the available parking spaces or a route (15). The geographic level is the (sub)network and the time scale is around 10 minutes. Smaller times are not necessary, as one cannot know beforehand what vehicles want to park where, so conditions may have changed by the time a vehicle arrives at the parking location.

3.3 Miscellaneous

Automatic parking is an application that allows vehicles to park themselves, an application working on the vehicle geographic scale (16). Drivers stop their vehicle close to the desired parking spot, confirm the exact location on the onboard computer and hand over control to the vehicle. The vehicle then has to check its path and nearby obstacles many times per second to adjust the exact path and check for obstacles moving into the path, giving a time scale of around 0.1 second. Like ACC, collected data needs to be highly accurate.

An application related to urban traffic routing in the previous section is that of guiding vehicles towards free parking spaces in a car park (17). For this it is assumed that individual parking spaces are monitored to see if there is a vehicle present or not. The geographic level of this application may easily be mistaken to be the vehicle level, but as information about available parking spaces is not aimed at individual vehicles but at all visitors of the car park in general, the string geographic level is more applicable. A time scale of 10 seconds may still produce usable results.

Then there are two applications related to emergency vehicles at controlled intersections. On the local scale level at an individual signal controlled intersection, emergency vehicles need to be detected within a few seconds so that green can be given to the direction where it is coming from (18). This can be expanded to the string geographic level to clear intersections in advance and provide the emergency vehicle with a green wave (19). In this case larger time scales up to around a minute can be permitted.

An application in another domain than individual cars, but still related to traffic conditions, in the area of public transportation is the prediction of arrival times at bus stops in order to for instance inform waiting passengers (20). To provide reliable information, it is necessary to have information about the current traffic conditions along the bus route. This implies the string geographical level. The time scale is in the order of a couple of minutes; with larger delays one might as well use the current GPS position of the bus and the empirical average speed. The reliability of the arrival time estimate depends on the data quality, so high quality data is required to obtain accurate estimates.

3.4 Comparison

A graphical representation of the applications and their respective time and geographical scales is given in figure 3. This figure clearly shows a relationship between geographical level and time scale. The larger the geographical scale, the longer the data processing times are allowed to be. For new applications this makes it possible to quickly get an estimate of the applicable time scale once the geographic scale is determined using the method described in section 2.1. The found time scale gives an indication of the required timeliness of the data needed for the application.

4 APPLICATION POSSIBILITIES WITH DATA: PRACTICAL ANALYSIS

Section 3 discussed the maximum (time) latency required for several applications. This section approaches the relationship from the other side, and checks which application given a certain type and quality of data available. Several Dutch road authorities of national, regional and local level collaborate in the National Data Warehouse (NDW) in order to collect and distribute road traffic data from a central location in a unified format. The available data mainly consists of speed and flow measurements collected using induction loops and travel time measurements collected using automatic number plate recognition. The geographical spacing of measurement sites differs between a few hundred meters on freeways to one or more kilometers on regional highways. Latency times include a data aggregation period of one minute and allowed data processing times of up to 75 seconds before distribution to users. In practice it is proven hard to fulfil this timeliness requirement; at the end of 2011 over half of the data transmissions did not comply.

The minimum data quality requirements for availability is 96.9%. The 2011 average availability varies by geographical region from 94.5% to 97.6%. A gradual improvement throughout the year is however visible, with the lowest scoring region at 97.1% in the last quarter of 2011 (23). These numbers apply to the availability from the ICT perspective.

From the traffic and transport view the availability will be much lower, especially in case of travel time measurements on routes with signal controlled intersections. An example of this is given in figure 4, showing availability values for all travel time sections on regional highways in the Dutch province of South-Holland for the period of February 1st to February 14th, 2012. The technical availability here is $98.5 \pm 0.4\%$, but actual travel time measurements are only available $46 \pm 23\%$ of the time. This is due to the fact that there is no traffic passing the detectors.

The quality requirements for accuracy are set at 95% for speed and flow measurements or 90% for realized travel time measurements. Conformance to this is tested by means of plausibility studies, for 2011 resulting in averages of 99.0% for speed, 96.7% for flow and 98.4% for travel time (23). The value of such a plausibility study is however questionable, as not the accuracy is tested, but instead the probability

TABLE 1 Overview of several assistance systems and their scales

Nr	Name	Time	Geographic	Comment
Highway				
1	Collision avoidance	80ms	Vehicle	Adaptive cruise control (7)
2	Merging assist	0.1-1 s	Local	Information of the surrounding vehicles is needed (21)
3	Magic Green	1 s	Local	Guaranteed green for trucks (8)
4	Homogenize traffic	1-10 s	Local	Information of surrounding vehicles and downstream vehicles is needed (9)
5	Incident detection	1-60 s	Local	Warning of upstream traffic for traffic congestion (22)
6	Event detection	1-10 min	String	Traffic management in case of traffic conditions deviating from daily patterns (10)
7	Stop-and-go wave removal	1-10 min	String	Speed needs to be reduced over several kilometers, over several minutes (4)
8	Routing	10 min	(Sub)network	Rerouting over a network (11)
Urban				
9	Collision avoidance (vulnerable road user)	0.1s	vehicle	Protection against crashing into pedestrians(12)
10	Time to green	1s	local	Indication of remaining red time for a traffic light (13, 14)
11	Green wave	1 min	String	Coordination of different traffic lights (2)
12	Urban routing	10 min	(Sub)network	Routing in the city center to carparks (15)
Miscellaneous				
13	Automated parking	0.1s	Vehicle	Vehicles that parallel park themselves (16)
14	Routing in P-garage	1-10s	String	Guidance towards free spaces (17)
15	Bus arrival prediction	1 min	String	Using traffic information to accurately estimate bus arrival times depending on local traffic (20)
16	Emergency vehicle priority	1-10s	Local	Green for the direction where a emergency vehicle is coming (18)
17	Emergence vehicle green wave	10-100s	String	Clear intersections and have green light for emergency vehicles (19)

that the accuracy requirements are met. The data values provided via the NDW are given as the mean of all measurements in a particular minute, including standard deviation where applicable. The one minute averaging will often result in measurement values with high standard deviations. From a mathematical and ICT viewpoint, these averages will generally be correct and therefore accurate. However, from a traffic and transport perspective, a high standard deviation indicates that the particular average does not resemble an average traffic state. A possible solution to this problem is to not just provide a regular mean of the aggregated measurements, but additionally also provide either the median or a Winsorized mean, i.e., a mean of the values excluding the extremes. These alternatives will generally be more true to the actual traffic state and will produce more accurate values for the traffic and transport view, while keeping mathematical and ICT accuracy the same. It falls outside the scope of this paper to further deepen algorithms that possibly

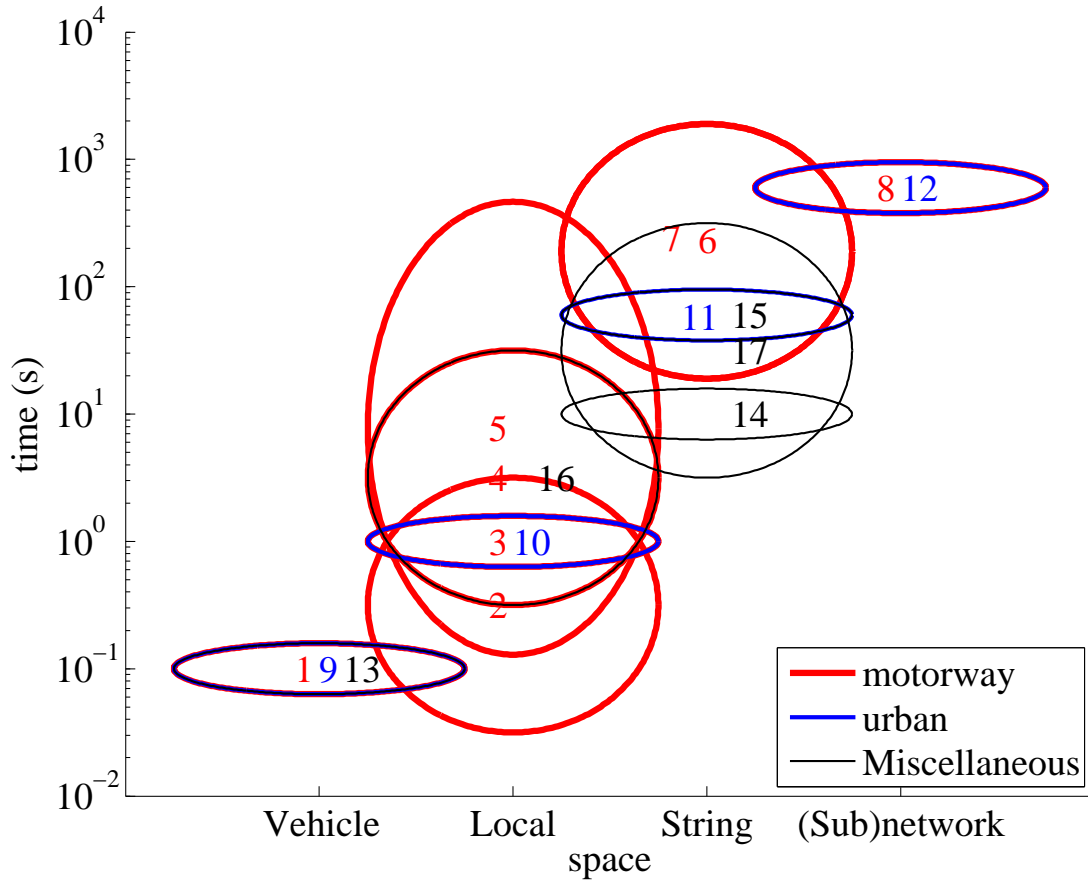


FIGURE 3 Different applications and their spatial and temporal scale shown in a graph. Note the logarithmic vertical scale.

overcome this problem.

Given these data characteristics, it can be concluded that applications working on the vehicle geographic level are impossible with traffic data from the NDW. The same applies to applications that are required to operate on a very small time scale or applications that require very accurate data. Looking back to figure 3, one can draw a line at 2 minutes and at the local vehicle level. All applications above and to the right are possible with NDW traffic data. This principle can be applied to any data source to find what types of applications are possible with it.

5 DISCUSSION AND CONCLUSIONS

In this paper, we studied ITS applications. In particular, we focussed on spatial application scale and the maximum latency in the control cycle, i.e. the time between the moment of measuring and the time of applying the control measure. We studied 17 different common cases of traffic control. The cases were collected from different application environments: 8 from a highway environment, 4 from the urban environment and 5 other examples.

From these examples, we found a remarkable correlation between the spatial scale of the application and the allowed maximum latency. This relationship on its own is a useful addition for practitioners who

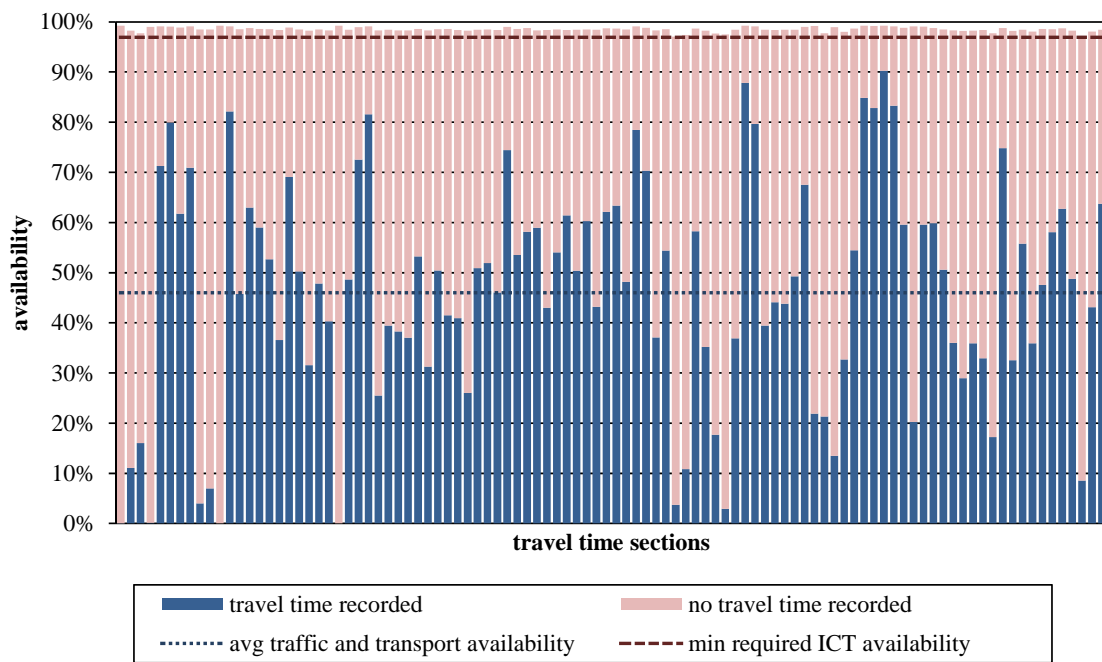


FIGURE 4 The availability of data for each of the travel time sections. Some data is not available due to malfunctioning, and sometimes there is no data because there is no traffic.

have a certain aim of traffic control. On the other hand, it can be used if a road authority knows which data (quality) is available, and wants to know what ITS solution can be implemented.

The explanation for the observed relationship is twofold. First of all, there is the speed of vehicles. For instance, a collision avoidance system (adaptive cruise control) must work at a small latency since the vehicle is driving at a certain speed, i.e. covering distance with a certain rate. The device should hence react quicker on changes in the environment of the vehicle (typically: 10 meter away) than the time the vehicle needs to get there. This is a speed relationship.

Secondly, latency can be relaxed due to the inertia of the system. Consider for instance the case of motorway routing. Even if the control would follow directly on the measurements, it takes time until a road fills. The rate at which speed decreases due to an increased demand, is based on the size of the network, and the possibility to support additional vehicles. A similar reasoning holds for the situation where the inflow in an area is reduced. If the inflow reduces, it takes time until enough vehicles left the area such that the speed is back to free flow conditions. These are inertial effects.

In many places, the network data collection and transmission protocols are being revisited. In fact, modern techniques allow much more detailed collection and faster transmission, but it all comes at a cost. This paper can give a guideline whether it is useful to spend this money.

The relationship shows that if one wants to control a large area, timeliness of the data is not very restrictive. This is useful, since especially in large areas there is much data which needs to be collected, aggregated, transmitted (several times), and processed. This research shows that typically there is time to perform all these actions. Thus, this research provides an excellent motivation to continue hierarchical traffic control. For high level control the data needs to be less timely than for low level control. An information and communication architecture can be developed which collects data on the lower level and possibly controls

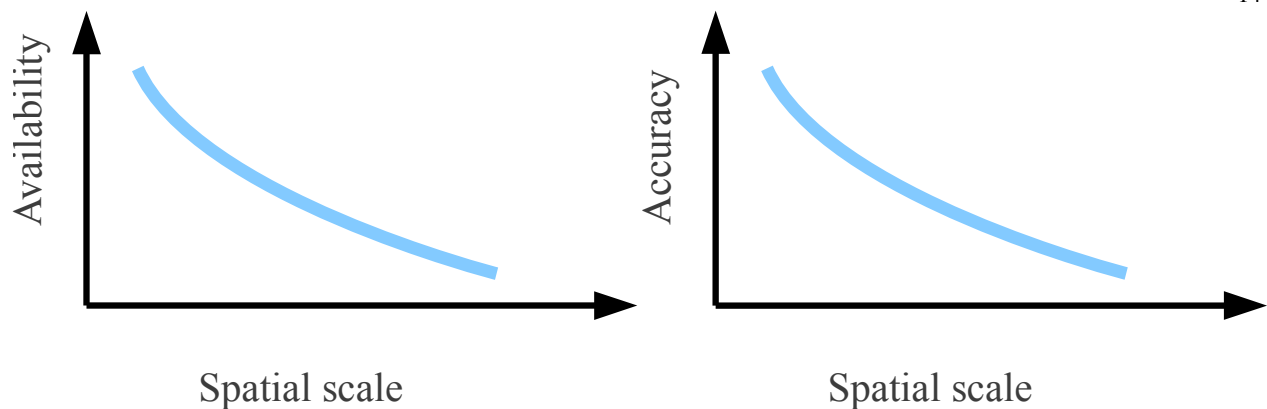


FIGURE 5 The expected relationships between the spatial scale and the other quality criteria

at this lower level. For higher level control, there is more time to aggregate the data, transmit the data and find an optimal solution. Further research will investigate the requirements for the data in terms of accuracy (how much error can there be in the available data) and reliability (what fraction of the data is required) of the data for different spatial scales.

It can be argued that for other quality criteria, availability and accuracy, similar relationships can be found, as is indicated in figure 5. We expect that the required availability decreases with a larger scale, mainly because the applications are not directly safety-related. For instance in the extreme case of active headway adjustment of individual vehicles, all data must be present at all times to avoid a collision, whereas this extreme requirement is not needed for instance for large scale routing. As is argued in the paper, for routing applications it is often sufficient to repeat the data of the previous aggregation time in case of lacking data. Also for accuracy we expect that the requirements are less strict for larger spatial scales. As soon as the traffic control requires an immediate action based on the information – as is the case with the low time scales –, it is most likely also required that this information needs to be correct. This, however, is a reasoning which needs to be investigated in more detail in the future, preferably with examples of traffic management systems, and the data they require. Both these expectations are also in line with work on data fusion, e.g. (24), stating that if more data sources are available, which is most likely the case for larger areas, the quality of the data can be lower. Averaging plays an important role in this conclusion, as well as knowledge on which (type of) data might be biased and which (type of) data is accurate for a particular aspect.

Acknowledgement This research was sponsored by a IP-CC subsidy from ICTregie/NWO in the project *SI4MS*, Sensor Intelligence for Mobility Systems, and by the foundation Next Generation Infrastructures.

REFERENCES

- [1] Landman, R., S. Hoogendoorn, M. Westerman, J. Van Kooten, and S. Hoogendoorn-Lanser, Design and Implementation of Integrated Network Management in the Netherlands. In *Proceedings of the 89th TRB Annual meeting*, 2010.
- [2] Hu, H., X. Wu, and H. Liu, A Simple Forward-Backward Procedure for Real-Time Signal Timing Adjustment on Oversaturated Arterial Networks. In *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems*, 2011.
- [3] Papamichail, I., M. Papageorgiou, V. Vong, and J. Gaffney, Heuristic Ramp-Metering Coordination

- Strategy Implemented at Monash Freeway, Australia. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2178, 2010, pp. 10–20.
- [4] Hegyi, A. and S. Hoogendoorn, Dynamic speed limit control to resolve shock waves on freeways Field test results of the SPECIALIST algorithm. In *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC 2010)*, Madeira, Portugal, 2010, pp. 519–524.
 - [5] Viti, F., S. Hoogendoorn, L. Immers, C. Tampère, and S. Hoogendoorn-Lanser, National Data Warehouse: How The Netherlands Is Creating a Reliable, Widespread and Accessible Data Bank for Traffic Information, Monitoring and Control of Road Networks. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2049, 2008, pp. 176–185.
 - [6] <http://www.ndw.nu/>, 2012, accessed July 2012.
 - [7] Ploeg, J., B. Scheepers, E. van Nunen, N. van de Wouw, and H. Nijmeijer, Design and experimental evaluation of cooperative adaptive cruise control. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, 2011, pp. 260–265.
 - [8] De Mol, J., W. Vandenberghe, S. Vlassenroot, and K. De Baets, *ITS-technieken om verkeersveiligheid te verhogen op kruispunten met verkeerslichten (VRIs): Onderzoek naar de mogelijkheden van dynamisch snelheidsadvies op VRIs (in Dutch, English summary)*. Ghent University, 2009.
 - [9] Kesting, A., M. Treiber, M. Schönhof, and D. Helbing, Adaptive cruise control design for active congestion avoidance. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 6, 2008, pp. 668–683.
 - [10] Zheng, P., M. McDonald, and D. Jeffery, Event detection based on loop and journey time data. *Intelligent Transportation Systems*, Vol. 2, 2008, pp. 113–119.
 - [11] Schlaich, J., Analyses of Route Choice Behavior using Mobile Phone Trajectories. In *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, 2010.
 - [12] Nedeveschi, S., S. Bota, and C. Tomiuc, Stereo-Based Pedestrian Detection for Collision-Avoidance Applications. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 3, 2009, pp. 380–391.
 - [13] Ibrahim, M. R., M. R. Karim, and F. A. Kidwai, The Effect of Digital Count-Down Display on Signalized Junction Performance. *American Journal of Applied Sciences*, Vol. 5, 2008, pp. 479–482.
 - [14] Eccles, K. A., R. Tao, and M. B. C., Evaluation of Pedestrian Countdown Signals in Montgomery County, Maryland. *Transportation Research Records: Journal of the Transportation Research Board*, Vol. 1878, 2004, pp. 36–41.
 - [15] Paolo, F. and Deflorio, Evaluation of a reactive dynamic route guidance strategy. *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 5, 2003, pp. 375–388, world Congress on Intelligent Transport Systems.
 - [16] Xu, J., G. Chen, and M. Xie, Vision-guided automatic parking for smart car. In *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, 2000, pp. 725–730.
 - [17] Chinrungrueng, J., U. Sunantachaikul, and S. Triamlumlerd, Smart Parking: An Application of Optical Wireless Sensor Network. In *International Symposium on Applications and the Internet Workshops, 2007. SAINT Workshops 2007*, 2007, p. 66.
 - [18] Giglio, D. and N. Sacco, Urban traffic control structure based on hybrid Petri nets. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, 2004, pp. 224–237.
 - [19] Kwon, E., S. Kim, and R. Betts, Route-based dynamic preemption of traffic signals for emergency vehicle operations. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, 2003.

- [20] Yu, B., W. H. Lam, and M. L. Tam, Bus arrival time prediction at bus stop with multiple routes. *Transportation Research Part C: Emerging Technologies*, Vol. 19, 2011, pp. 1157–1170.
- [21] Pueboobpaphan, R. and B. van Arem, Understanding the relation between driver/vehicle characteristics and platoon/traffic flow stability for the design and assessment of Cooperative Cruise Control. In *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, 2010.
- [22] Qiang, H., M. Hadi, Y. Xiao, and C. Zhan, Estimating Incident Detection Time based on Incident Management and Traffic Detector Data. In *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, 2010.
- [23] National Data Warehouse, *NDW Rapportage Kwaliteit, 4e kwartaal 2011 (in Dutch)*, 2012.
- [24] Ou, Q., *Fusing Heterogeneous Traffic Data*. TRAIL thesis series 2011/5, Delft University of Technology, 2011.