Measuring real-time traffic data quality based on Floating Car Data

Jeroen Brouwer
TomTom
De Ruyterkade 154, 1011 AC Amsterdam, The Netherlands
+31 20 75 75441, jeroen.brouwer@tomtom.com

Abstract

Congestion is a growing and global problem, impacting the majority of people traveling and goods moving. Governments are challenged to keep the traffic flowing, to reduce pollution and decrease the economic damage caused by congestion. Highly accurate traffic information helps governments to reach their goals to become smart and sustainable. Due to the increasing volume of connected navigation devices in the market (PNDs, smartphone applications, in-dash devices) Floating Car Data (FCD) becomes widely available. When this data is matched with a map and processed in a traffic fusion process real-time traffic solutions including traffic incidents and traffic flow are created. However, FCD comes in many different qualities. To guarantee the quality, different techniques need to be incorporated into the fusion process, including the abilities to measure the quality of delay data, the quality of transit time data and the quality of journalistic incident data.

Keywords: Congestion, Traffic Information Quality, Floating Car Data.

Traffic Information Services

TomTom's real time traffic information is generated by combining measurements of existing infrastructure with signals from anonymous connected GPS navigation devices and anonymous GPS equipped Smartphones. Historical traffic information is used as a background source to complement and add confidence to the real-time observed information.

TomTom combines the different data sources together in a traffic fusion process and two different traffic services result from this fusion process; TomTom Traffic and TomTom Traffic Flow. Both services provide highly accurate real time speeds and travel times on individual road stretches. TomTom Traffic is mainly used by consumer products like satellite navigation devices. Using TomTom Traffic, delays and journalistic information are delivered on all affected roads in the area requested (example in image 1.1). TomTom Traffic Flow is mainly used for governmental monitoring purposes. By using TomTom Traffic Flow, real time speeds and travel times are delivered for the full road network for all road segments (example in image 1.2). These traffic solutions are available immediately and require no infrastructural changes on the road network. As a result there is no time required to develop or install the solution.

Image 1.1 - Traffic Incidents

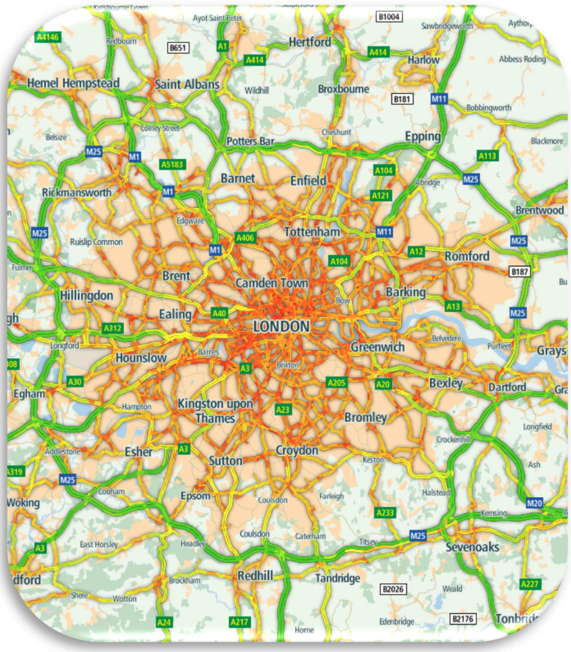


Image 1.2 - Traffic Flow

Sources

To collect real time data TomTom uses revolutionary collection techniques. As described, information from multiple sources are collected and automatically fused to generate real-time traffic information. In Europe the following sources are used to create the traffic services:

1. GPS Data from TomTom LIVE navigation devices (PNDs)
2. GPS Data from TomTom Smartphone navigation applications
3. GPS Data from Automotive in-dash and on-dash devices
4. GPS Data from TomTom Fleet Management systems (Business Solutions)
5. GPS Data from 3rd party Smartphone navigation applications
6. Public Traffic Information

By collecting GPS data from so many different sources TomTom has a fleet of connected devices driving around the world that is unsurpassed by any other floating car data (FCD) fleet. As 'connected navigation' is the popular choice for navigation, the combined fleet of GPS devices is growing day by day with impressive speed and consequently the quality and coverage of the traffic information increases. Hundreds of millions of connected devices globally are able to contribute highly accurate and anonymous GPS probe data with the consent of the user. Examples of users of TomTom real-time traffic services in the automotive market include in-dash and on-dash connected navigation devices for Renault, Mercedes-Benz, BMW, Toyota, Mazda, Fiat, Alfa Romeo, Lancia, Lexus, Opel and Ford. In the near future Peugeot and Citroen and others will join this group. TomTom also license traffic information to a number of key mobile phone handset manufacturers such as Blackberry.

These source devices all monitor road traffic conditions continuously and when each of them contacts the TomTom servers for traffic information they exchange intelligence on the congestion they have experienced in the past few minutes on their journey. This congestion information is fused into the TomTom traffic services anonymously, so that this intelligence can be passed to following customers to improve their journey.

Image 1.3; different sources of accurate GPS data

Thanks to the continuous sampling of very precise GPS measurements, all groups of connected devices contribute to the accuracy of the travel time information. In order to make sure only information from inside the car is used a special filter is implemented so bicycle, public transportation and pedestrian data is filtered out. With a new GPS breadcrumb generated every few seconds on the device it is easy to accurately understand the road conditions for the vehicles and with just a few devices reporting anonymously for a specific road stretch it is possible to generate high quality traffic information with very detailed speed and travel time information with confidence.

In addition to real time GPS traffic data, historical GPS speed data is used as a background source. 'Speed Profiles' details time-dependent speed characteristics along a road element (unique value for time-of-the-day and day-of-the-week). Speed Profiles are derived from aggregating hundreds of billions of anonymous GPS measurements from millions of devices that reflect world-wide consumer driving patterns. Travel data is analysed and normalized to ensure the highest degree of accuracy. Speed values are assigned to each routable road element in the map. The data is updated quarterly for changes in map data as well as to enhance freshness and coverage of the historical data. Image 1.4 shows an example of a road element Speed Profile. Speed Profile data contains data for a 5 minute interval for each day of the week. As you can see in the image the speed values change over time, showing speeds around 40 km/h around midnight and clear drops in speed during the weekday rush hours.

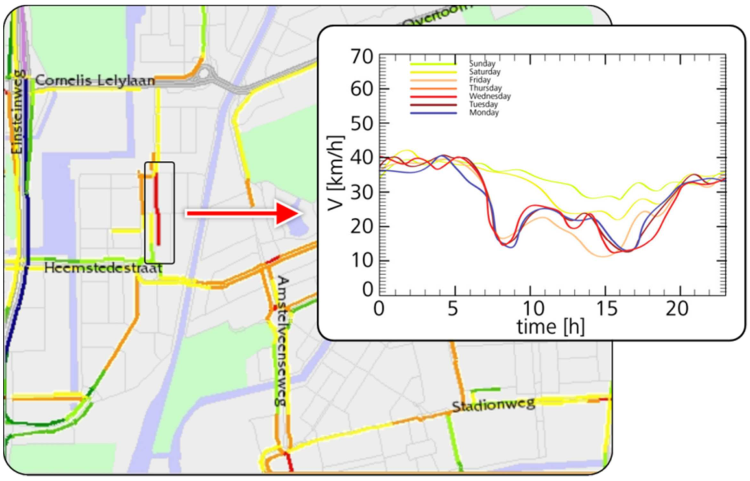


Image 1.4; Speed Profile example in Amsterdam.

Heat maps can visualize the volume of probes and the distribution over the road network. These heat-maps are essentially black sheets with all measurements plotted in top of this. The brighter the colour the more measurements were received. Please note that these are only TomTom devices (not 3rd party devices). For traffic data creation also other GPS measurements are added. In the images below you find a heatmap for Europe and North America[1].

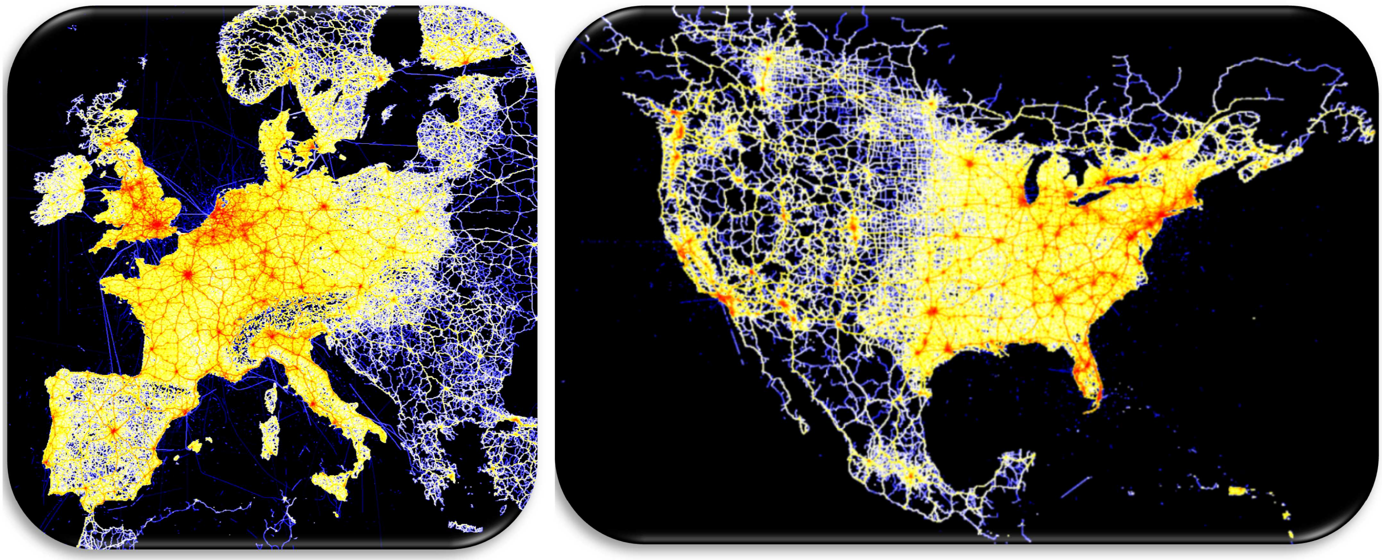


Image 1.5; TomTom GPS measurements for Q1 2013.

Using Floating Car Data from various sources result in a number of benefits compared to other ways of measuring travel time and speed information. As cars drive on all roads that are open for traffic it is possible to measure with floating car data travel times and speeds on all roads, even when construction works take place, forcing traffic for example to go on a temporary lane or drive on the other side of the dual carriageway. Also when new roads are opened congestion can easily and quick be reported. Additionally, the system does not have to wait until a vehicle has reached the end of a road section before the travel time is calculated as data can be received regularly along the route. This is especially an advantage in slow traffic conditions when it takes a long time before a vehicle has travelled to the end of a road section. Floating Car Data enables TomTom to accurately measure the speeds, also in case of low speeds or standstill conditions.

[1] Please note that TomTom is not limited to these countries but can collect data for any location on the world.

Data Fusion

TomTom has developed unique proprietary software to automatically combine all data sources to generate speed and travel time information. The software processes compares the different sources and automatically decides which source delivers the most accurate data. By comparing real time data with acquired historic data and feedback from users the system is being improved on a continually basis. The software was first developed and used for the launch of TomTom Traffic in The Netherlands in 2007. Since then the software has been continuously refined in order to improve the quality, handle more and more probe data and to have the system working in different countries with their own specific tuning characteristics. Using a number of different quality measurements, TomTom can measure if new software improves the quality as intended. Image 1.6 provides an overview of the process used to combine the data sources into traffic information to the end user.

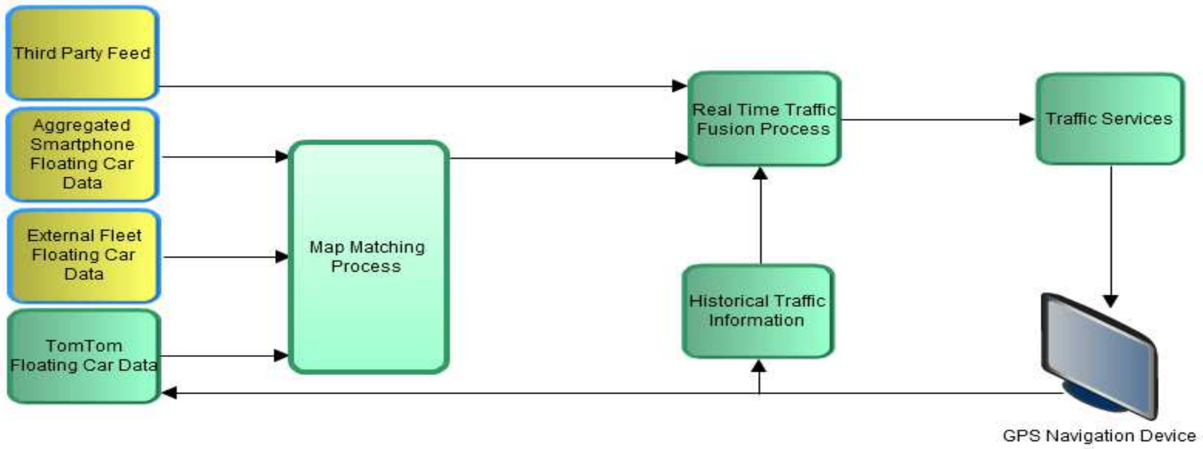


Image 1.6; data fusion architecture.

The accuracy of GPS Floating Car Data depends on a number of factors and can range from 1 millimetre up to 30- metres depending on the number of satellites in-view of the device, the GPS-chip that is used, the local terrain and weather situation. The accuracy of GPS chips used in TomTom navigation systems is better than 10 meter with a 95% confidence. The data is stored in WGS84 format with latitude and longitude degree values to 5 decimal places, which means a maximum location accuracy of 1.1 metre can be saved. TomTom also uses a refined map matching algorithm to match the GPS positions precisely to a geographical database. Therefore locations that are slightly (10 meters) of the roadway will be snapped back to the centre line of the road segment as defined in the database. The map-matching process is important as individual location measurements are combined to determine the route and travel time of a probe vehicle in the network. After map-matching the GPS data the data is sent to the fusion process. This fusion process considers a number of factors for each trace presented (for example the age of measurement, the deviation from other samples etc.) to determine the aggregated speed for every segment, hence detecting traffic jams. The confidence level is determined as part of the fusion process which is an indicator of the likely quality. With this aggregated information, traffic data is created in the different formats (e.g. flow or incidents).

As described in the introduction of this paper the output of the fusion process is traffic incident information and traffic flow information. Due to the fact that congestion can evolve very quickly, this data is updated every 30 seconds. Time-Distance images visualize the importance of fast update intervals. Image 1.7 displays a time-distance image for a highway in Belgium. Green and yellow represents freeflow and blue / purple show low speeds. As you can see the congestion is quickly moving backwards in this example, proving the urgency of fast data fusion and traffic updates to provide the freshest and most precise information to drivers.

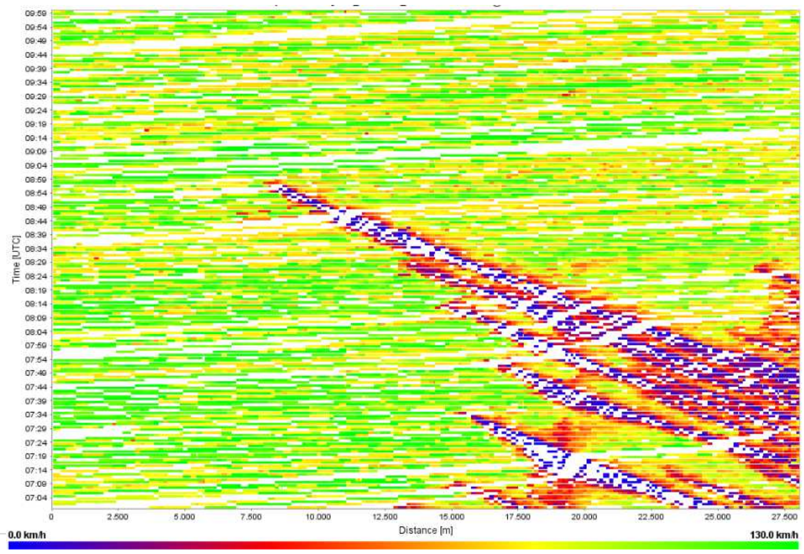


Image 1.7; time distance graph for the E19 near Antwerp, Belgium. Severe congestion is moving backwards along the road with a high speed.

Incident Data Quality

By using Floating Car Data it is possible to automatically detect and correct road closures, even when they are not reported by official sources. It is possible to measure a lack of real-time GPS probe data in the road network, indicating a closure if the amount of real-time floating car data does not match the expected amount of GPS probe data. In this case automatic closure reports are created. In image 1.8 you find an example of probe data where every blue dot represents a GPS probe data trace. When the fusion server detects a lack of traces, like on the left cloverleaf, it will trigger the system to automatically close the road until new probe data is measured on the closed road stretch.

Image 1.8; example of Probe Data Reports.

Besides automatically detecting road closures, service providers like TomTom also receive road closures from the governmental data feeds. This data is also used in the fusion process. Often, however, the closed stretch does not represent reality. By using GPS probe data it is possible to verify and correct a closure message to the actual closed stretch by removing closed stretches where GPS data is measured but also to extend a closure if needed. This means that a message can be completely removed or that the closed stretch is shortened or enlarged.

Figure 1.9; example of a trimmed roadwork report in Germany based on actual speed measurements.

With Floating Car Data it is also possible to detect the accuracy of the roadwork stretches. Governments and service providers can report very long roadwork stretches while the actual stretch of constriction is in reality much shorter. TomTom is adjusting this information by only showing roadwork reports where GPS speed data shows travel time delays. In image 1.9 you see one example of this technology on the German Autobahn. The green lines represent the reported roadwork location as reported to TomTom. The red line represents the aggregated measured speed over the stretch, indicating a clear drop in speed between km 8 and km 12. The location of the roadwork is adjusted to only the area where a drop in speed is measured.

Traffic Delay Quality

Traffic data can be split into two different services that are supplied; traffic incidents and traffic flow. Traffic incidents only provide information about traffic jams (congestion) while traffic flow provides information about all road stretches, congested or not. Both products have different use cases and configuration and hence the need for different quality benchmark methods.

In order to measure the quality of delay information it is important to have solid ground truth data which is independent of the data used to create the real time (delay) traffic information. The TomTom navigation devices can be split into two different categories: connected TomTom devices (enabled with GSM connection for real-time data transmission) and non-connected TomTom devices (not able to share data over the air, but data is uploaded via TomTom desktop software). Data from connected devices is used to generate the real-time traffic information. Non-connected probe data is a completely unique dataset that has not been used to generate traffic data. This data is what is called the Ground Truth data and this database is independent from the real-time data source. With this data it is possible to analyse the speed and congestion situations and compare these to the delay information published in the real-time traffic feeds at the time.

TomTom adopted the QKZ method to measure the quality of traffic delay information. QKZ is the most commonly recognized and followed methodology to measure the quality of congestion information. When looking at quality figures it is key to understand what the definition of congestion is. TomTom has its own definition of congestion. The definition is that the congestion needs to be at least 1 kilometre of observed speed driving less than 56% of the normal freeflow speed and with a minimum of 90 seconds delay. With traffic delay quality two components are relevant, the coverage and the error rate. Coverage is the percentage of traffic jams that were reported in the real-time traffic information and what was found to be true in the ground truth database. So when 80 delays were reported but 100 congestion locations were found in the ground truth database in the same minute, equals 80% coverage. Error rate is the percentage of delays that were inaccurately reported when comparing to the ground truth database. For example if 100 delay messages were reported but only 80 were found in the ground truth, this is represented as 20% error rate. Congestion can be summarized in a congestion cloud in a time-distance graph as displayed in image 1.10. As clearly visualised there is a hit area, miss area and ghost area. This together is summarized into coverage and error.

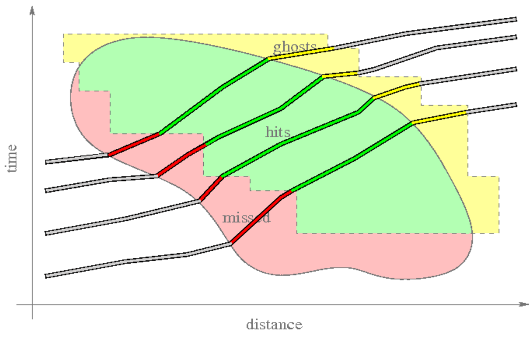
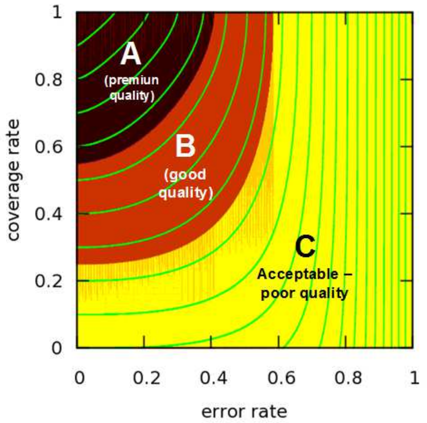


Image 1.10; example of how congestion is measured and how ghosts, hits and missed congestion are detected.



Coverage and error can be summarized in a graph displaying the quality levels of traffic data. In this graph a combination is made of both coverage and error. As explained, coverage rate goes from 0 (very bad) to 1 (very good). For error it is the other way around as value 0 is very good and value 1 is very bad. Combining both values means that with an coverage value of 1 and error rate of 0 would mean perfect quality with no mistakes made. A coverage value of 0 and an error rate of 1 would be completely incorrect traffic data. Obviously traffic quality without mistakes is difficult as congestion is changing so fast it will never be possible to cover and predict all delays very accurately. Within QKZ there are three different quality levels distinguished; premium, good and acceptable. The target quality of TomTom is premium quality (level A). By comparing the real-time data with the separate database with real driven trips it is possible to measure quality per area (e.g. country, province), per data selection (e.g. week, month) and per roadclass (Functional Road Class)

Image 1.11; Different QKZ quality area. Premium is the target area.

Image 1.12 provides an overview of the QKZ values for a range of countries, measured over March 2013. The values enable quality benchmarks, but also feed back into the development team indicating where improvements are necessary.

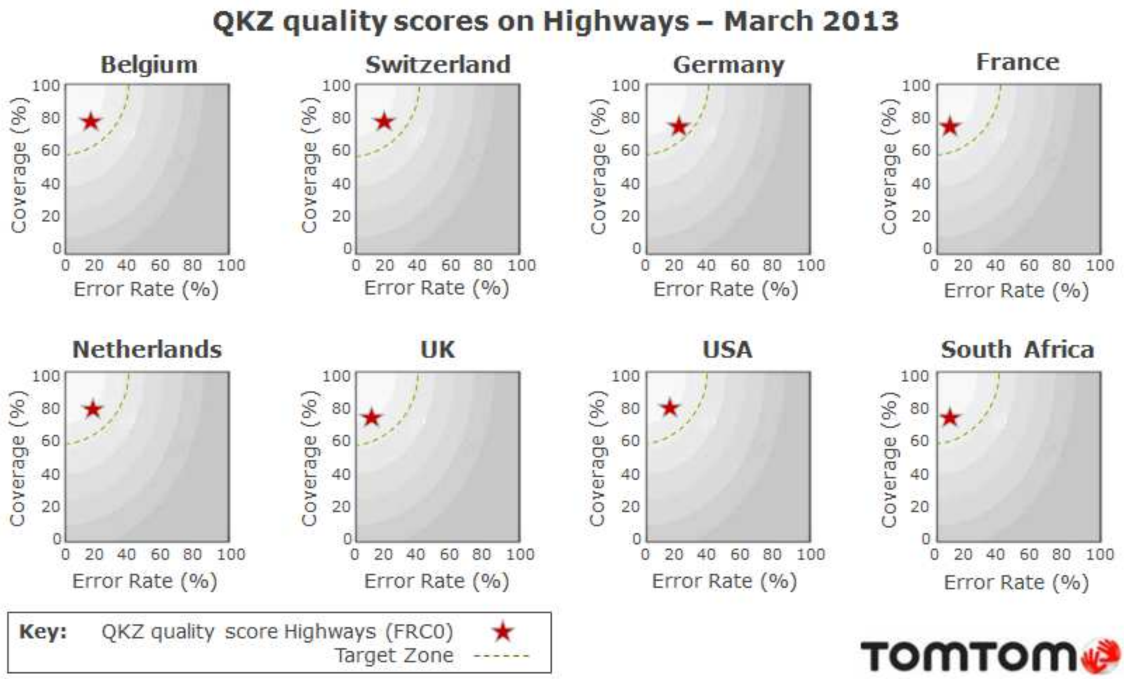


Image 1.12; QKZ values for 8 example countries.

Traffic Flow Quality

Testing flow data requires different techniques compared to testing delay data. Reason is that flow data is about the speed per road segment and the travel time attached to this. The focus in quality measurements is travel time accuracy.

QBENCH is the most commonly recognised and followed methodology for measuring travel time accuracy. This method requires three different inputs; the freeflow travel time, the ground truth travel time and the reported travel time. These inputs mean for TomTom that the historical Speed Profiles with the overnight freeflow are used as the freeflow travel time. The ground truth travel time is detected by using the database with non-connected TomTom speed measurements (see QKZ for more details) and for the reported travel time the Traffic Flow data is used. The delay for each road link with an observation in the ground truth database is calculated and compared to the delay report in the data used for the real-time flow data. Results are expressed as a value between 0 and 1 where 1 is the perfect quality score.

Relevant in QBENCH is that there needs to be a definition of the congestion threshold. Due to the fact that this threshold is open to configure it is possible that different analysis based on the same data show different results. The congestion threshold of TomTom is 50% of the overnight freeflow. Next to the threshold there is also an error tolerance and standstill cut-off as you cannot measure travel time over data that is not moving. Besides the definitions there is an option to assign different weight to penalties for over-reported delays versus under reported delays.

With QBENCH it is possible to visualize the traffic data quality. An example is displayed in image 1.13 whereby the green area represents data within the tolerance, yellow within increased tolerance (positive or negative) and red displays all traveltimes in excess of higher tolerances. Image 1.14 provides an example of QBENCH results for California, USA for Q3 2011. The yellow and red areas contain a combination of too positive and too negative traveltimes.

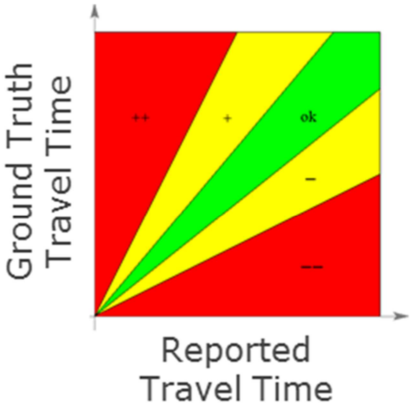


Image 1.13 (above); QBENCH principals.

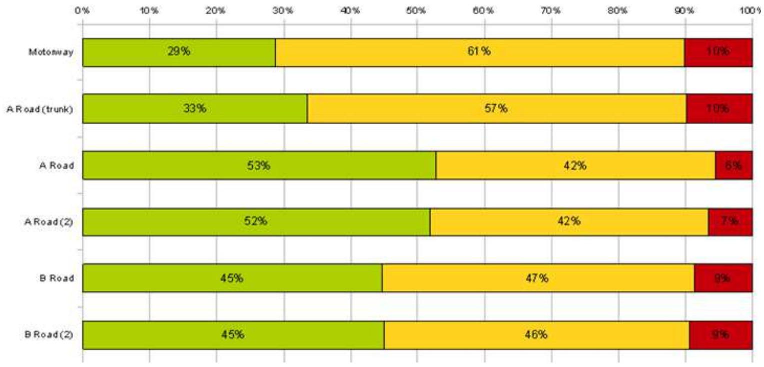


Image 1.14 (left); QBENCH results for Q3 2011 for California.

On the road

Several different techniques have been described in this paper for statistically measuring traffic information quality, but it is important not to lose the link to reality. Drivers use the data to navigate to where they want to be as quickly as possible and it is important to appreciate the user experience in the traffic information process. Hence it will always be essential to execute occasional drive tests to support the statistical quality test results. Despite the low sample size that test drives provide (it is not possible to drive past all jams) it does provide a view of how data is used. It will show the impact that the accuracy of the information displayed and estimated time of arrival (ETA) have on the user experience. Image 1.14 provides an example of a trip in the congested Ruhr area in Germany and how ETA calculations with TomTom traffic are more stable over time then a competitive traffic service.

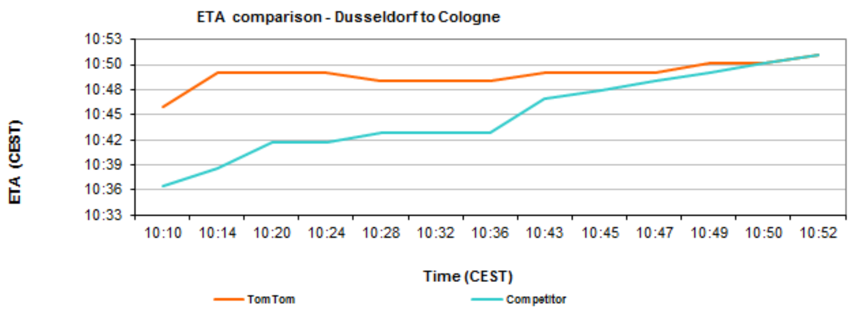


Image 1.15 (above): travel time comparisons. TomTom ETA is more stable over time.

These tests are conducted by both TomTom and independent organizations validating quality or comparing different service providers. Image 1.16 shows how a typical test drive is carried out with many different navigation devices placed in a car driving in traffic.

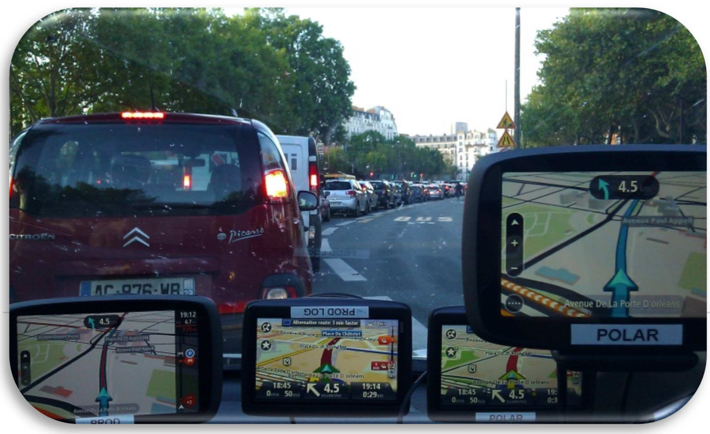


Image 1.16 (right): test drive set-up, testing traffic data in Paris City center.

Conclusions

Traffic data based on Floating Car Data enables service providers to make high quality traffic data on all roads. FCD can help make journalistic data more accurate by correcting information sources that conflict with flow data in the fusion process – and even self-generating messages where sustained abnormal flow is observed.
Quality benchmark methods such as QKZ and QBENCH are easily integrated with FCD systems – particularly if there is a significant independent ground truth source.

References

- UMTRI; Stuck in Traffic: analysing real time traffic capabilities of personal navigation devices and traffic phone applications. October 2012
- TUV Süd; NavCert: Study on TomTom High Definition Traffic. March 2010
- Deutsches Zentrum für Luft- und Raumfahrt: Project drive test – comparison of five Personal Navigation Devices (PND)