Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

Research Program TomTom-DLR (2011)

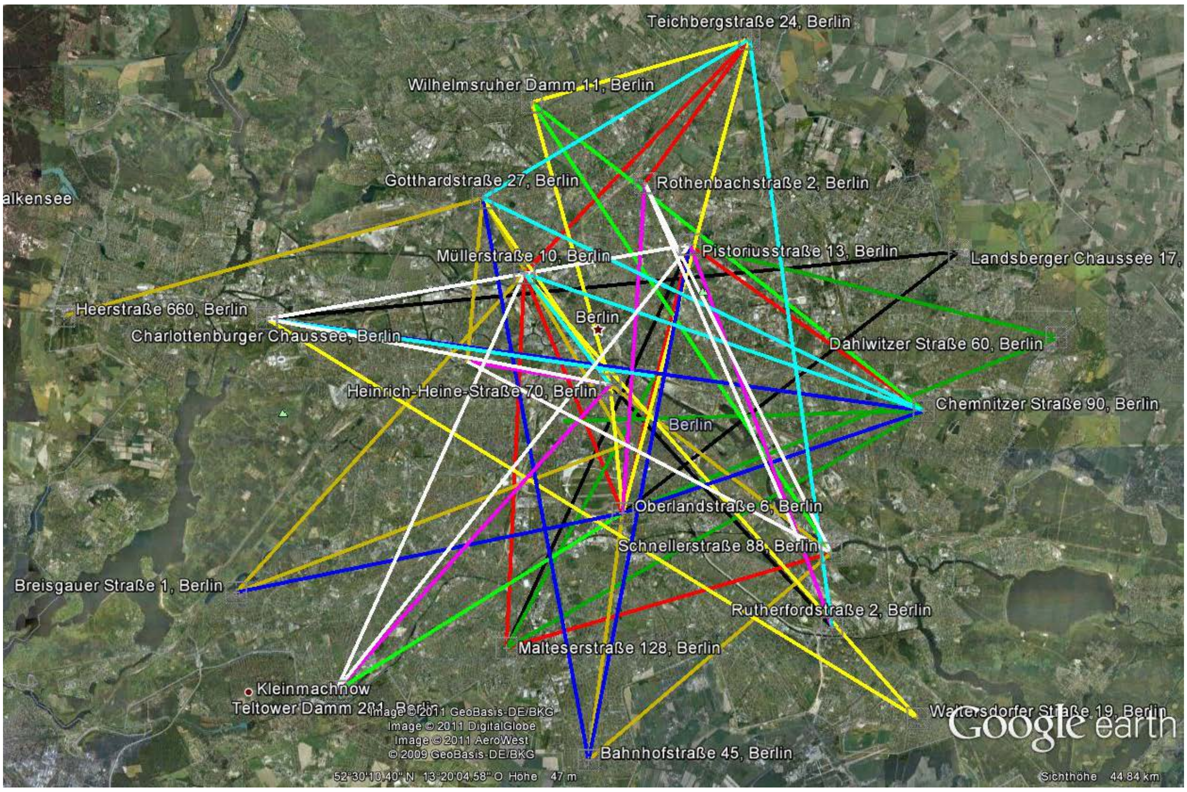
Project drive test – comparison of five Personal Navigation Devices (PND)

Project Report

Alexander Sohr and Peter Wagner

Institute of Transport Systems, German Aerospace Centre

19. December 2011



Executive Summary

This document describes the results of a measurement campaign, where five different PND's had been compared. The comparison looked for the ability of the PND's to compute and guide a driver on the fastest path to his or her destination. The basic performance metric which had been used to compare the five PND's is the relative travel time gain: for each valid trip performed, there was one PND which performed worst (the vehicle guided by this PND had the longest travel time). The difference between the travel time of the worst and the PND currently looked at is the (raw) travel time gain of this PND. Of course, a more robust indicator is in this case the median value of the relative travel time gains. Since the PND's had a similar performance, and to compare these gains, again the gain of the weakest device had been subtracted from the raw gains to lead to the following medians of relative travel time gains:

	Bosch	Garmin	Google	TomTom HD	TomTom XL
relative travel time gain (median)	0.00%	9.55%	3.19%	12.63%	9.23%

By a closer examination it turns out, that the gains can be attributed at least partly to a smaller stop time share: the TomTom PND's generate routes that lead to an about 8% reduced stop time.

Obviously, under certain circumstances, larger gains could be realized, with a typical value between 20 and 30% compared to the worst route for a certain trip.

Another measure of the quality of a PND might be its ETA, the estimated time of arrival. Here, it seems that all the devices still have deficits at least in this urban area with its very complicated structure of changing traffic patterns. In the beginning of the trip, the PND's typically underestimate (they believe the trip takes shorter travel time than it really does) the time of arrival between 10 and 30% of the final trip time taken. While the TomTom HD is among the best (together with the Google App) in this task, the TomTom XL is in the midfield of this league. Of course, it could be speculated that this is an effect of the better traffic flow information available to this device. On the other hand, nothing is known about how the computation of the ETA that is displayed by any of the devices is actually done, therefore other reasons are imaginable as well.

Finally, a small part of the campaign was devoted to the driver's subjective assessment. They were asked to fill a questionnaire to give their rating of the quality of the PND's ranging from questions regarding the routes to questions about the men/machine interface on a five-point scale from excellent to poor. Again, by taking the average of the different ratings it turned out that the devices were similar in their performance, with again the TomTom's having a small edge over the other PND's.

Contents

Figures

Tables

1 Set-up of the driving experiments

The task of this project was to find and quantify differences in the performance of PND's. While there are many means to do such a testing, this project focuses on the mere driving or routing performance: given a couple of possible trips between an origin and a destination, which of the routes generated by the PND's turns out to be the best? The final test, which one is the best route will be evaluated by simply driving the different routes generated by the PND's and log all the trips by GPS-loggers and manually by entering data during the trips into an excel sheet.

Altogether five different PND's had been compared (the names in bold will be used subsequently in all the plots and tables as short-hand notation, the sequence is alphabetical):

- The "**Bosch** Navigation" App Ver.1.2.21, running on an Apple iPhone 4 – the software had access to real-time information provided by INRIX called "Inrix Traffic".
- A **Garmin** nüLink! 2340 device as a genuine PND; this device had real-time traffic information provided by Navteq called "3D Traffic".
- The "**Google** Navigation" App running on a HTC Desire device with Android 2.3.3; this device, too, had real-time traffic information, albeit it is not clear to what degree it is being used in the computation of the routes.
- **TomTom HD** 1000, a genuine PND with online access to TomTom's HD traffic service.
- **TomTom XL**[2], which is also a genuine PND, but without access to online traffic flow information. Albeit the device has in principle at least TMC information available, the TMC radio reception was disabled. It uses, however, TomTom's IQ-Routes technology, which computes fastest routes on the basis of predetermined travel time profiles, which yield in principle different routes for different starting times.

The devices had been bought by DLR directly off the shelf and it had been made sure, that (at this time) the most recent software and map updates had been installed on the devices (where applicable).

All the devices were set to compute the fastest route between origin and destination. A PND was assigned to a vehicle, and in each vehicle there were a team of two students, one driving and the other writing down all the necessary data that could not be locked by the GPS-loggers. The teams and the assigned PND's have been rotated, so that each team was using each PND roughly for the same number of trips. This should somehow balance the differences between the drivers, albeit it is not perfect in avoiding all the biases. The drivers, however, have not been rotated through the cars, since the rental car company requires that each car is driven by the same driver. The cars itself have been typical compact cars like a Volkswagen Golf or similar.

The origin and destination points had been selected from a set of parking lots of super-markets in the German capital Berlin, because this eases the co-ordination between the teams and makes sure, that the cars leaves at roughly the same time. This list of parking lots had been selected in advance. However, the teams usually decided by themselves to which destination they should drive next and picked the destination from the list of parking lots. They tried to travel through areas with strong traffic, and they tried to find routes which took them between 30 and 40 minutes. The destinations were chosen to proceed in a star-like manner through the city, see Figure 1.

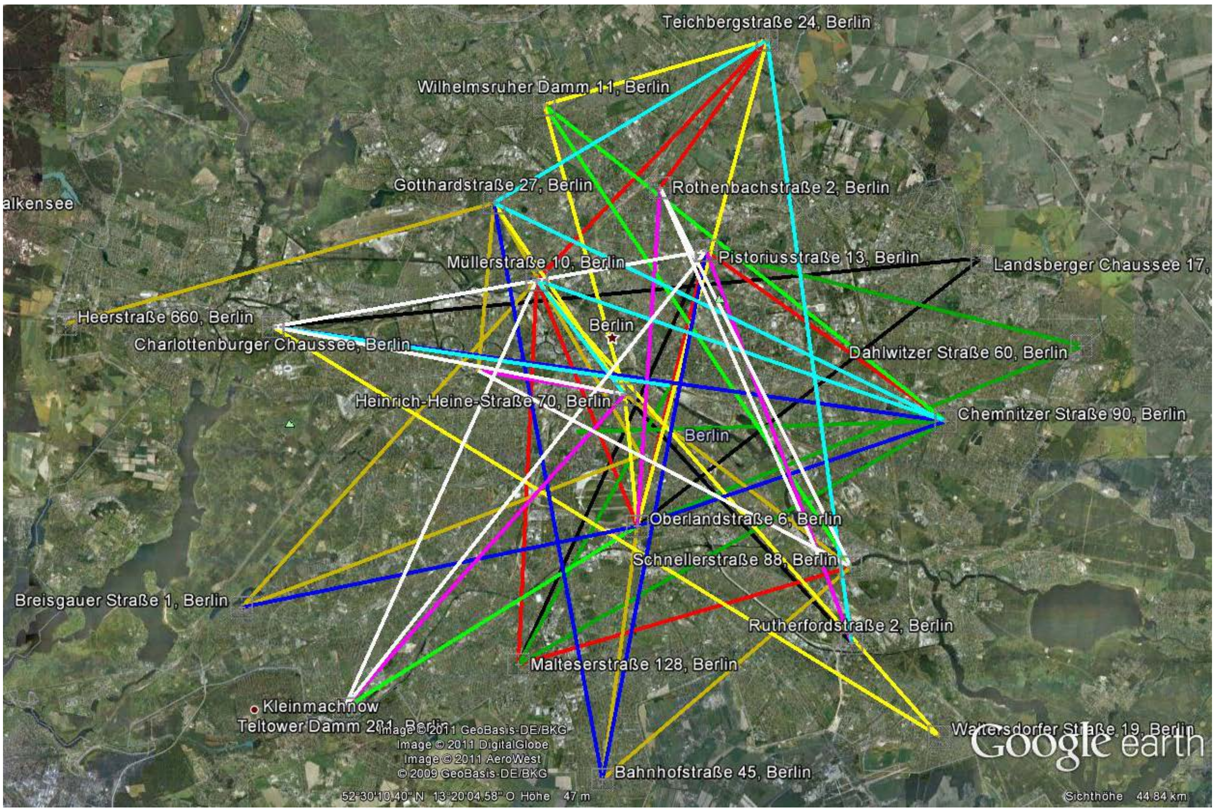


Figure 1: The origin and destination points of the trips driven during the two full weeks. Each color stands for a different day. The corresponding kmz file will be in the data package transferred to TomTom.

All the experiments have been conducted in the city of Berlin in three waves. The first wave was a one day pre-test designed to find weaknesses in the general set-up which took place on Thursday, 8. Sept. 2011. The second wave lasts from Monday 26. September to Friday 30. September 2011, the last wave was performed in the week from Monday 17. October to Friday 21. October 2011. The drives took place during the morning rush hour between 6 and 9 o'clock, and during the afternoon rush hour from 14 to 19 o'clock.

The pre-test has not found strong flaws in the design of the campaign. Apart from a number of smaller points, the only significant change was a shortening of the routes to meet the above mentioned criterion for a typical commuter trip to last between 30 and 40 minutes. Therefore, the seven trips sampled on this day had been included into the total analysis as well to improve the statistics. They have not been used, however, in the ETA and overlap analyses of sections 3.5 and 3.6.

During each trip, each team filled in an excel-sheet which was installed on a netbook, that each of the teams had with them. In Figure 2, a screen-shot of such a sheet is displayed.

STRG + h	**New Route (Header)**							
STRG + t	**Time**							
STRG + f	**final (Footer)**							
1. ETA all 3min loggen								
2. notice any events		Changes in the route, why, where, what changes						
		Cahnges in ETA						
		Traffic jams (real or PND), one entry in log-book at start and at the end of a jam						
		device o.k.? ETA constant?						
from:	Oberlandstr. 7				**GPS_start:**	06:28:02		
to:	Dahlwitzer Str. 60 / Griebenweg 2				**GPS_end:**	07:07:42	**GPS_TT:**	00:39:40
via:	Hermannstr., Skalitzer Str., Warschauer Str., B1/B5							
CAR / DRIVER:	Fzg 5 / Turgay							
Screenshot?	OK							
Traffic ON?	OK							
GPS-Logger ON?	OK							
	Time	**ETA**	**Congestion?**	**Congestion I**	**Details (changes in route- or ETA, congestion,...)**			
	06:27:36	07:04		Warning	Oberlandstr. Baustelle, Fahrbahnverengung, Anfang unc			
	06:30:07	07:05			Baustelle, Fahrbahnverengung Hermannstr.			
	06:31:59	07:05			Ende			
	06:35:03	07:05						
	06:38:03	07:05						
	06:42:35	07:07			stockender Verkehr, B96A			
	06:45:42	07:08						
	06:47:31	07:08			Fahbahnverengungen auf Frankfurter Allee, Meldung im			
	06:50:29	07:09			starker Verkehr auf Frankfurter Allee, auch stadtauswär			
	06:53:39	07:08						
	06:56:14	07:08						
	07:01:05	07:08						
	07:04:14	07:08						
	07:07:29	07:07			Ziel erreicht			
END	**Positive:**	waren Erste, kein Stau						
	Negative:	x						

Figure 2: Screenshot of the excel sheet used. Originally, the sheet is in German, some text has been translated to English for this report.

The team had the task to report the origin and the destination points, the car and the driver, in addition a screenshot was produced from the starting situation, and the team was asked to make sure that the GPS-logger is up and running, and that the devices that have an online connection to a traffic state server have a running connection ("Traffic ON" in the screen-shot). During the trip, the driver report roughly any three minutes the current time, and the ETA (estimated time of arrival) that the devices display. Unfortunately, we had to resort to this error-prone manual sampling, since it was not possible to read the data directly from the devices. In addition, the teams were asked to report anything unusual (column with header "Details"), and to report whether the PND flags a congestion of they themselves observed congestion. No clear definition of congestion had been given to them, so there is an element of subjectiveness in this assessment. At the end of the trip, the teams were asked to give a one-line assessment of their experience with the PND in this trip (labeled "Positive" and "Negative" in the last two rows.)
At the end of each of the days, each team had to fill a questionnaire (this is not visible in Figure 2) which tried to assess a more balanced and more detailed view on the experience of the drivers with the PND (more on this in section 4).
To do the final data analysis in section 3, the excel sheets had been edited for obvious errors which showed up during the analysis. Most apparent have been strange numbers in the ETA (e.g. 19:18 instead of 9:18) or some of the remarks that had been normalized (e.g. spelling errors in the texts) to yield a cleaner statistics.

2 Aggregated Statistics

The data have been analyzed mostly with the help of scripts written in python. A python library named xlrd allowed read access to the excel sheets, which could be used to perform complex and reproducible analyses of the data. A number of simple statistical numbers have been derived first from the data, to give a general idea about the data-set. All the times in the following table are in the format hour [h]:minutes [min]:seconds [s].

# trips	# invalid trip data	# of identical trips	min trip time	max trip time	average trip time	standard deviation
85	3	16	0:12:10	1:28:40	0:39:41	0:13:58

Altogether 85 trips had been driven altogether, three of them had not been used in the final analysis since a GPS logger failed or the team approached at the wrong destination. A couple of trips had not been performed at all since all the PND's displayed the same or at least roughly the same routes – the teams were asked to check for this before they go on the trip and if it happens, to choose a new destination which would give a different set of routes. This has been reported to happen 16 times, however it could not be ruled out that the teams have not reported each case, or that the analysis had not caught each case. This seems to happen especially near 9 o'clock, which is understandable: this is about the end of the rush-hour. Then, the likelihood increases, that the routes become similar, since the traffic load vanishes leading to roughly free-flow travel times. As could be seen from the table, the shortest trip was 12 min 10 s, while the average trip time was 39 min 41 s, with a standard deviation of roughly 14 min.

3 Detailed Statistics

How to compare the results of the five PND's? There is no general solution to this, therefore different methods will be explained and executed in the following.

3.1 Ranking

The arguably simplest approach is to compare the travel times and rank them. The PND with the shortest travel time gets 5 points, while the one with the longest travel time gets 1 point. Other weightings are possible and give different results. Applying this simple approach to the 82 sampled routes yields the ranking statistics displayed in Table 1. The best PND is the TomTom HD with 288 points, while the weakest PND is the Bosch App. However, the difference between the PND's is not dramatically as can be seen especially when looking at the median and the 25%- and 75%-quantile's of the ranks. How strong the PND's differ will be shown later more thoroughly by a statistical analysis of the differences of the travel times measured, were detailed significance levels can be added to the differences between the different PND's. As can be observed from the mean values and the computed 95% confidence intervals (C.I.) of the mean values, there is a certain overlap between the results of the ranking of the PND's. The confidence intervals have been computed by two different methods: the classical method which uses the formula $\mu \pm SE$ where μ is the mean rank and SE is the standard error of the sample, which is defined as standard deviation σ divided by the square root of the number of observations, $SE = \sigma/\sqrt{n}$. In addition, a so called bootstrapped approach has been used to compute these confidence intervals directly – this method is much more robust against distorted (non-normal) statistics as the classical approach. In this case, however, the differences are fairly small showing that the statistical approach used is a valid one.

Table 1: Ranking statistics with confidence limits

PND	Bosch	Garmin	Google	TomTom HD	TomTom XL
points	204	248	222	288	263
75% quantile	3	4	4	4	5
Median m	2	3	3	4	3
25% quantile	1	2	2	3	2
Mean μ	2,49	3,02	2,71	3,51	3,21
Standard deviation σ	1,38	1,40	1,33	1,26	1,48
Standard error (SE)	0,15	0,15	0,15	0,14	0,16
95% C.I. (upper, bootstrap from R)	2,80	3,38	3,01	3,79	3,56
95% C.I. (lower, bootstrap from R)	2,20	2,78	2,43	3,27	2,90
upper 95 % C.I. from μ + 1.96*SE	2,79	3,33	2,99	3,78	3,53
Lower 95 % C.I. from μ - 1.96*SE	2,19	2,72	2,42	3,24	2,89

3.2 Travel time gains

A more numerically oriented approach is to use the travel times T_i directly. However, since they differ from trip to trip (with roughly a factor of 6 between the longest and the shortest trip) comparing mean values is not an option. Therefore, for each trip the travel times have been compared to the longest travel time of this trip and each device has been assigned a "travel time gain", which is the difference between the travel time of this device and the maximum travel time (for this trip). These numbers will be worked with in the following: first, they can be averaged to yield an average gain in travel time for each device, or, by normalizing them to the maximum of the travel times, a relative gain in travel time can be defined:

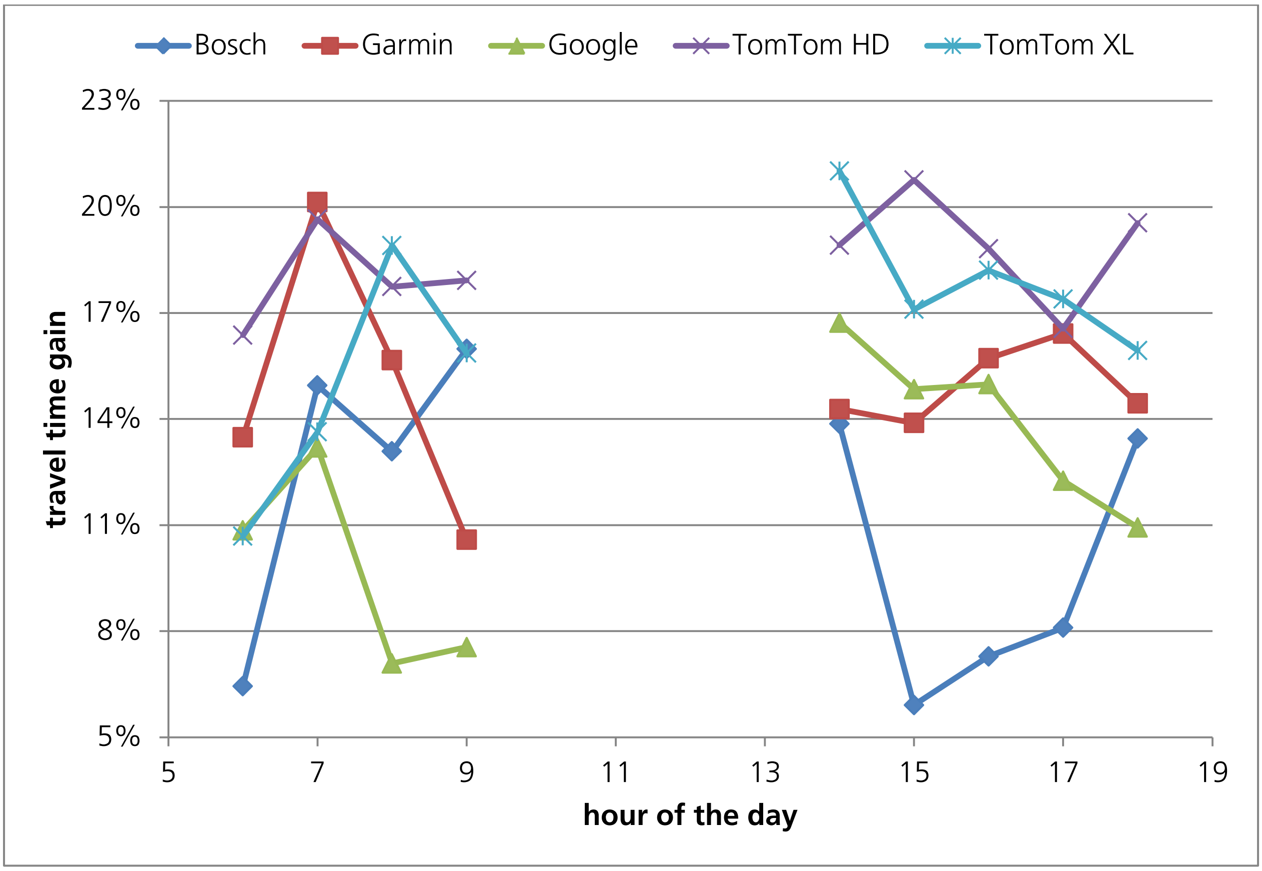
$$g_i = \frac{\max_{k=\{1,\ldots,5\}}\{T_k\} - T_i}{\max_{k=\{1,\ldots,5\}}\{T_k\}} \qquad (1)$$

These gains have been analyzed and yield the results displayed in Table 2.

Table 2: Averaged performance indicator median and average travel time gain

P.I.	Bosch	Garmin	Google	TomTom HD	TomTom XL
scaled median gain	0.0%	9.6%	3.2%	12.6%	9.2%
scaled mean gain	0.0%	4.6%	1.1%	7.9%	5.3%

More details can be obtained by computing the distributions of travel time gains, they are displayed in the Figure 3. However, this Figure has to be interpreted with care, since there are at most 16 events in each of the bins – which statistically a small number. However, it seems that at least two different patterns can be seen in Figure 3: The Bosch PND has a maximum in this distribution at zero gain, while all the other devices peak at a gain between 10 and 20%. The further details of thes distributions mirror of course the result found so far.


Bosch
Garmin
Google
TomTom HD
TomTom XL
23%
20%
17%
14%
11%
8%
5%
travel time gain
5
7
9
11
13
15
17
19
hour of the day

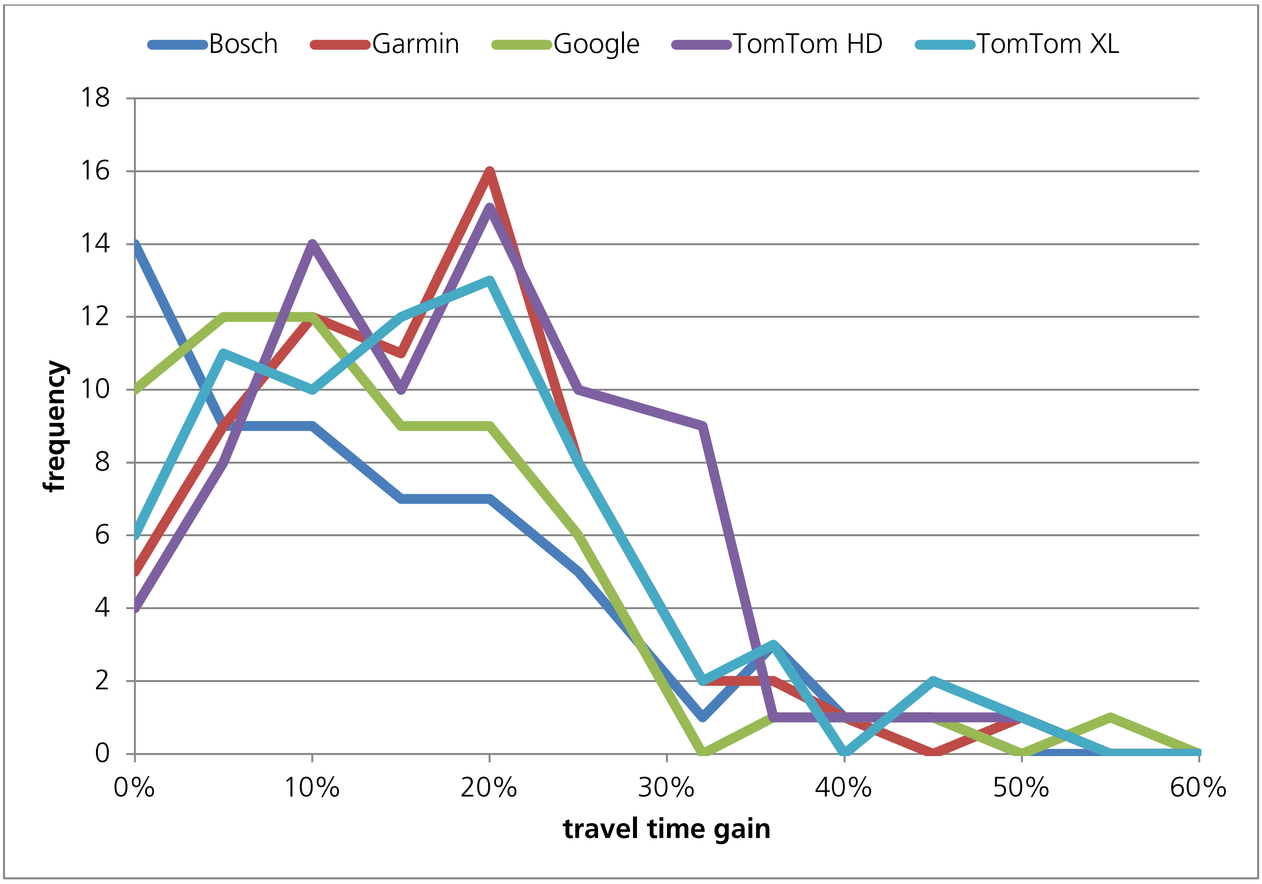


Figure 3: Histograms of the travel time gains for the five PND's.

The final result of this section is the confidence analysis in Table 3: by comparing the travel time data-sets, the significance matrix can be computed. Each number in the array gives the probability that the two devices have displayed the observed different average travel time by chance. For example: the TomTom XL and the Garmin PND share a probability of 44% that the difference between them (which is in fact fairly small) have been observed randomly – in other words, they have to be regarded as statistically not different. On the contrary, the difference between the TomTom HD and the Google PND is highly significant, i.e. the chance that this difference has occurred by chance is about 1:100000, which is highly unlikely. In accordance with the statistical parlance, the cells in the matrix have been colored in green (highly significant), orange (weakly significant) and red (not significant).

Table 3: Significance matrix

	Bosch	Garmin	Google	TomTom HD	TomTom XL
Bosch		4,76%	83,16%	0,00%	0,49%
Garmin	4,76%		2,87%	1,87%	43,92%
Google	83,16%	2,87%		0,00%	0,86%
TomTom HD	0,00%	1,87%	0,00%		10,03%
TomTom XL	0,49%	43,92%	0,86%	10,03%	

3.3 Analysis of outliers

To characterize the results obtained so far, a so called analysis of outliers has been performed. Before doing so, a working definition of outlier is needed. Within this project, an outlier has been defined as follows:

1. For each trip, compute the mean value, the minimum and the maximum value of the five travel times achieved by the five PND's.
2. From the set of mean values, compute the total mean value (this is of course the mean value of all the trips taken together)
3. Compute the mean-value of the minimum and maximum values. By normalizing them to the total mean value of all the trips, an interval is obtained which should contain most of the travel time values. For the data at hand, the interval is [0.88,1.18], and it turns out, that about 75% of all the travel times can be found in this interval – the rest will be defined as outliers.

With this definition at hand, the following Table 4 can be compiled:

Table 4: Analysis of outliers

Feature	Bosch	Garmin	Google	TomTom HD	TomTom XL
extremely fast	7	10	4	16	15
normal	63	66	64	63	62
extremely slow	11	5	13	2	4

The interpretation is straightforward: the main difference between the different PND's is just this distribution of outliers. The two TomTom devices, e.g., manage to find about 25% of routes to be very fast, while the rest of the routes are just in the bunch of normal routes – this can be due to the fact, that the system simply does not have better routes, our it could be due to the fact that the traffic information used had not been entirely correct.

3.4 Analysis by time of the day

The data have also been analyzed as a function of the hour of the day. The results are displayed in Figure 4, but they do not show a clear tendency (e.g. different devices performing better/worse for different hours of the day) apart from the ones already described. (E.g., the ranking between the PND's is roughly conserved over the hours of the day; this may have come out differently, if times with calmer traffic had been used for the data-sampling.

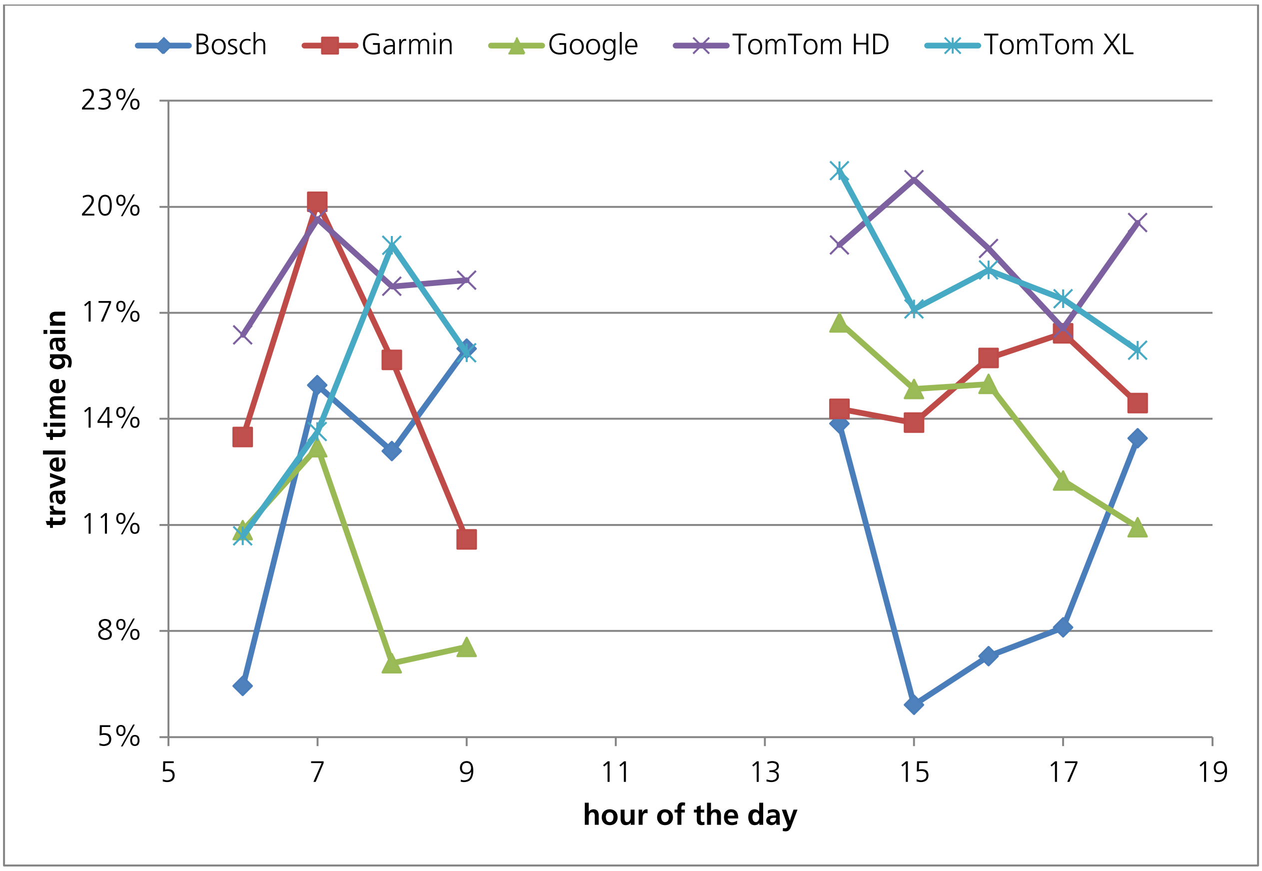


Figure 4: Travel time gain versus hour of the day. The data for the 13th hour had been excluded from the diagram, since they contain only one trip.

3.5 ETA-Analysis

Also of interest had been the quality of the ETA computations of the devices, since drivers certainly rely on them. To arrive at such an assessment, the following steps had been taken.

1. The team have recorded at roughly 3 minutes intervals the current time and the ETA as displayed by the PND. These data have been edited, since they seem to contain errors (remind, that they have not been read-off the devices in an electronic manner, but by humans). Some examples that had been identified are the wrong hour (9:18 instead of 8:18, or even 19:18 instead of 9:18) and have been corrected: this can usually be done because the ETA data before and after this faulty data contain completely different values.
2. Since the routes have different lengths, and so do the number of ETA data per route, a simple normalization scheme had been used. First, a new time-axis had been introduced ranging from 0 to 100, then the ETA-data in each of the intervals $[t_k, t_{k+1}]$ had been assigned to the bins $[100 \cdot t_k/T, 100 \cdot t_{k+1}/T]$ where T is the total travel time of this trip and the t_k are the times relative to the starting time when the ETA-data had been recorded.
3. In each of the 101 bins, the different ETA-values, subtracted by the finally achieved arrival time, and divided by the total trip time of the associated trip had been sampled and their mean-value computed.

This yields the relative ETA-error, together with the standard deviation (in each of the bins) as function of the percentage of the trip computed and is displayed in Figure 5. The error-bar is displayed only for two PND's (Google and TomTom XL) to avoid overcrowding of the figure.

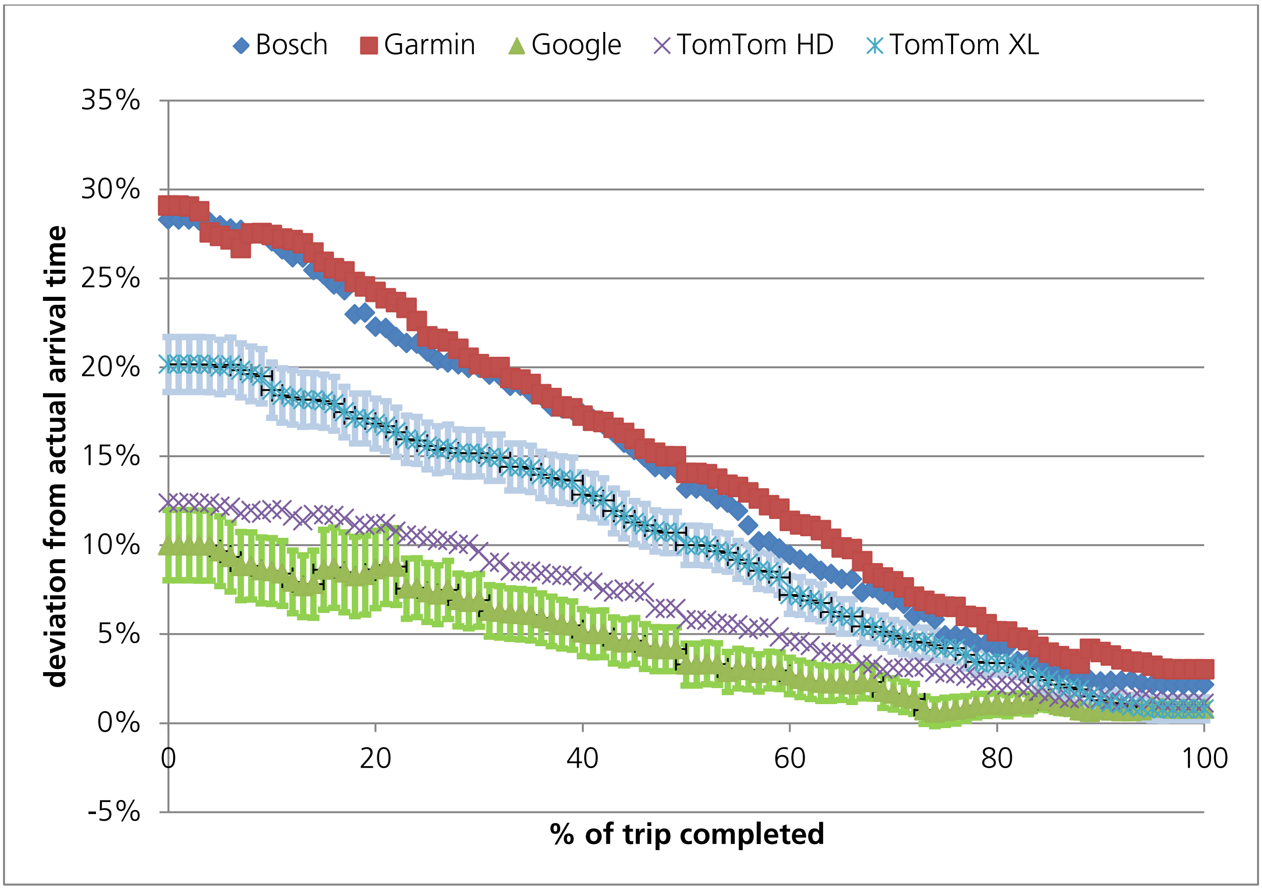


Figure 5: Analysis of the ETA error.

They are, however, similar for the other three PND's not displayed.

3.6 Route length and speeds

To analyze the route lengths and the speeds of the vehicles needs the recorded GPS-tracks. The speeds have been used directly as recorded in the GPS tracks, therefore they are measured data (and not computed e.g., from the difference between two subsequent way-points) and carry the (small) uncertainty of the GPS signal. All the data have been aggregated into distributions.
The Figure 6 displays the result for the speed distribution. Note, that the difference between the different PND's is not large – all the vehicles spend a considerable amount of time by waiting at intersections and the like. A closer look at the speed zero bin reveals however, that again the TomTom devices gain an edge: TomTom users seem to have to wait a little less, with the difference (about 9%) again similar to the overall difference (about 8%) already characterized in section 3.2.

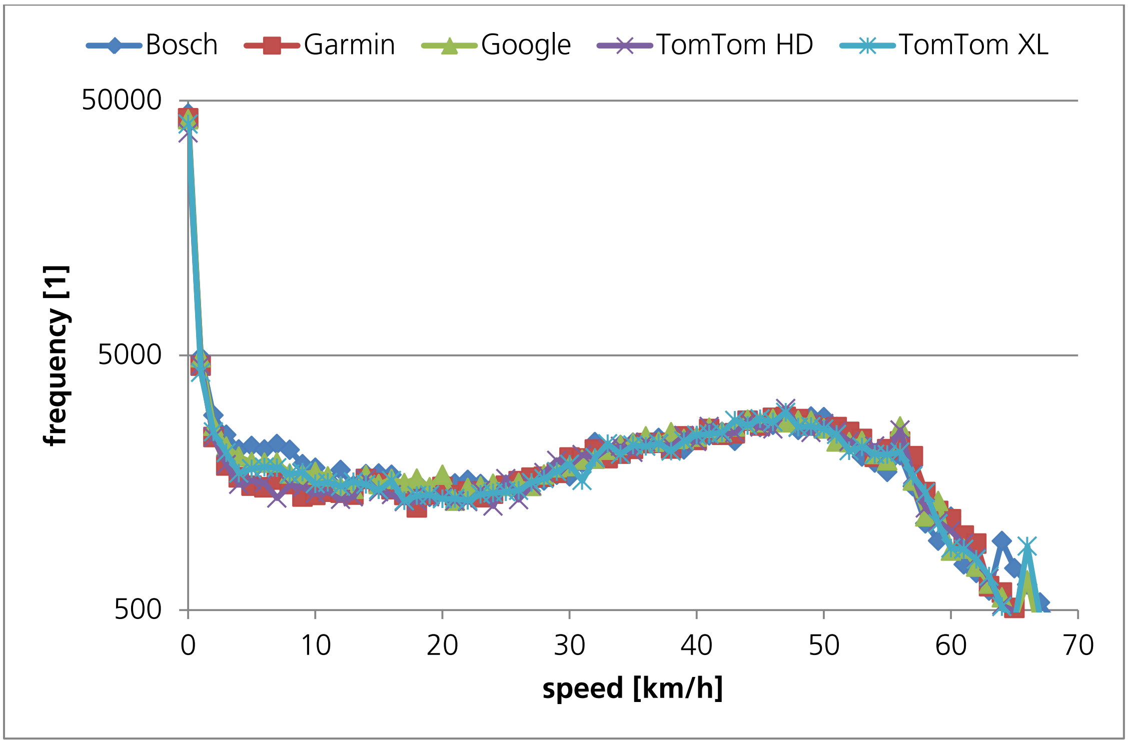


Figure 6: Speed distribution of the different PND's. Note the logarithmic scale of the y-axis.

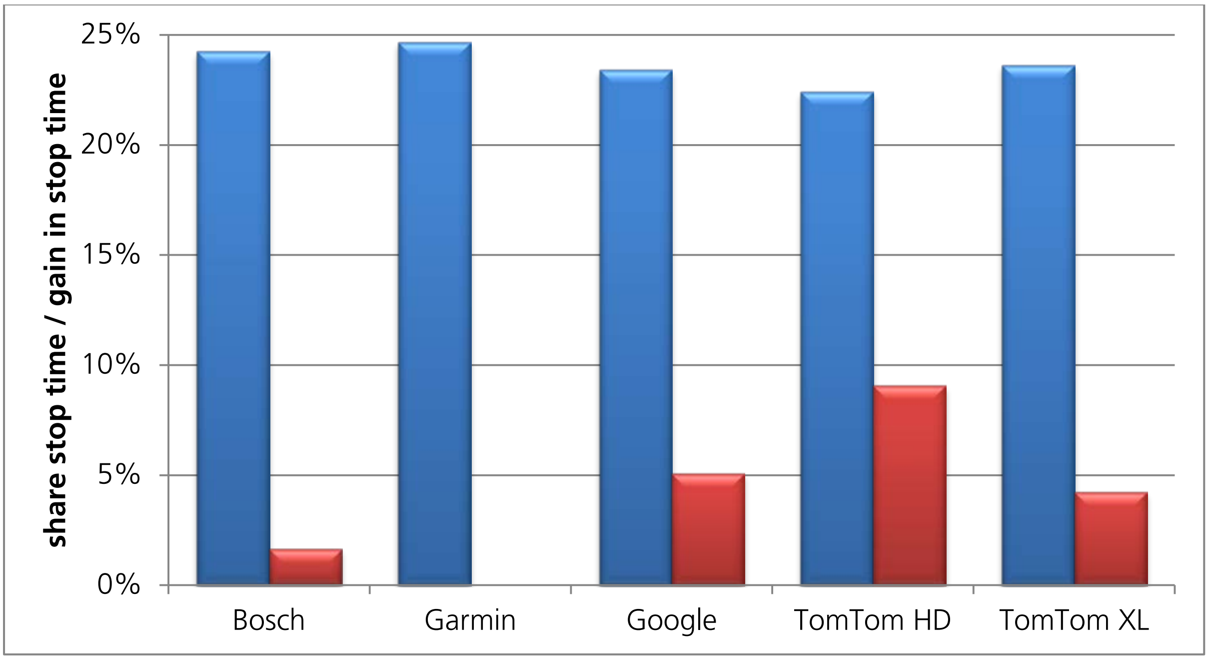


Figure 7: Share of the stop time (compared with the total time -- in blue) and the gain in stop time (red) for the five PND's.

3.7 Overlap between the routes and FRC distribution

Finally, the average overlap between the routes of the different PND's has been computed. In addition, the FRC distribution (FRC = functional road class), i.e. the share of the trip length that each trip spends in a certain FRC, has been computed, too. To do so, the tracks had been assigned to a

NavTeq digital map by the map matching algorithm in use within DLR's taxi FCD system. While the overlap matrix between the PND routes in Figure 8 is just for control purpose (to see that the routes were in fact different on average), the FRC distribution in Figure 9 contains additional information.

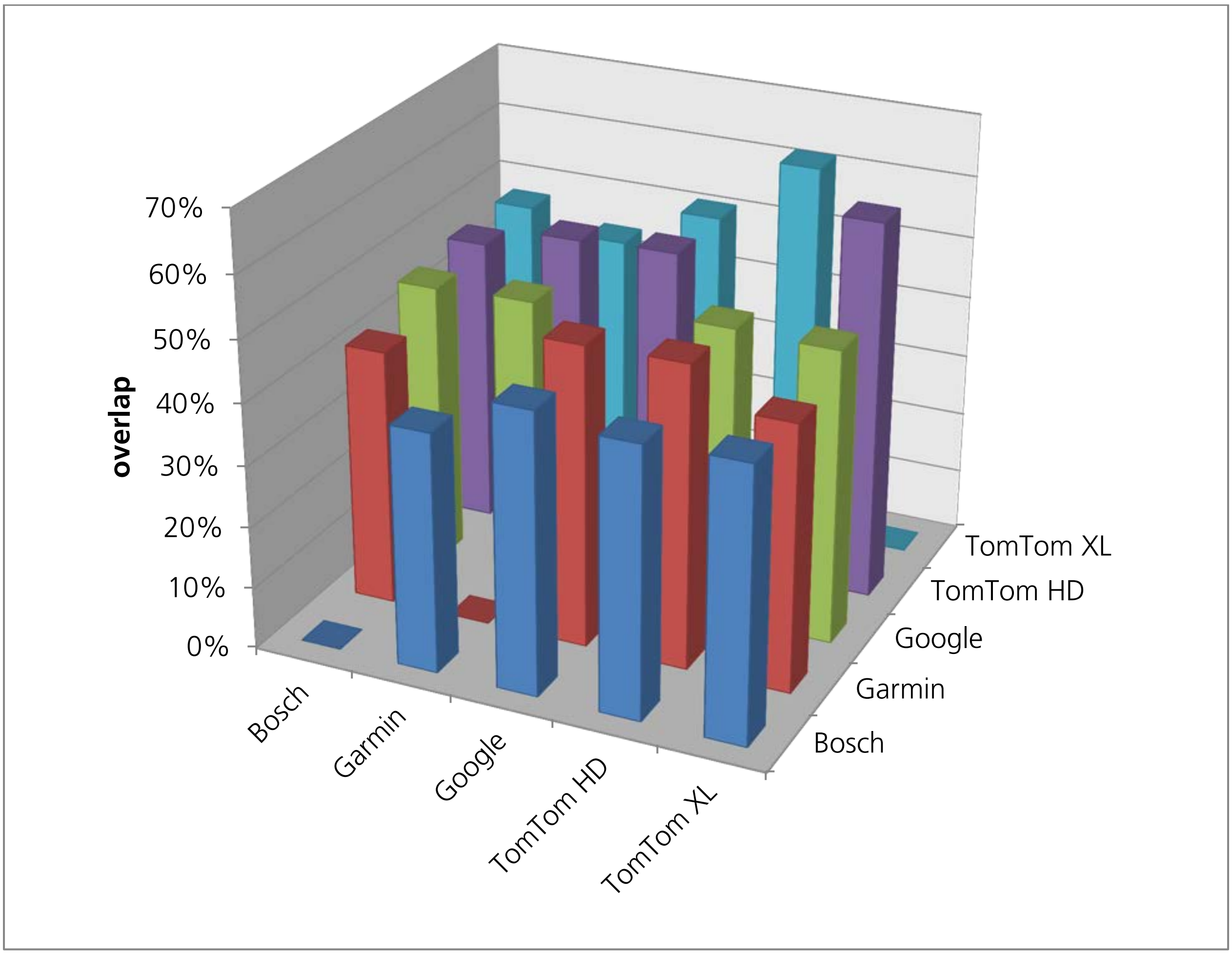


Figure 8: Average overlap between the routes of the different PND's. Overlaps are between 50% and 60%.

First, there are roughly two types of distributions visible in Figure 9: the Bosch and Google PND's have a higher share of FRC classes 0 and 1, while the remaining PND's have a maximum use on FRC class 2. In addition, the TomTom HD has the largest share of FRC class 4, which means that the dynamic traffic information in fact leads to an increased use of the smaller roads. However, the difference is still not dramatic, when looking at the numbers it could be seen that 8% of the total trip length is on FRC 4, while the other devices range between 2.5% and 5%.

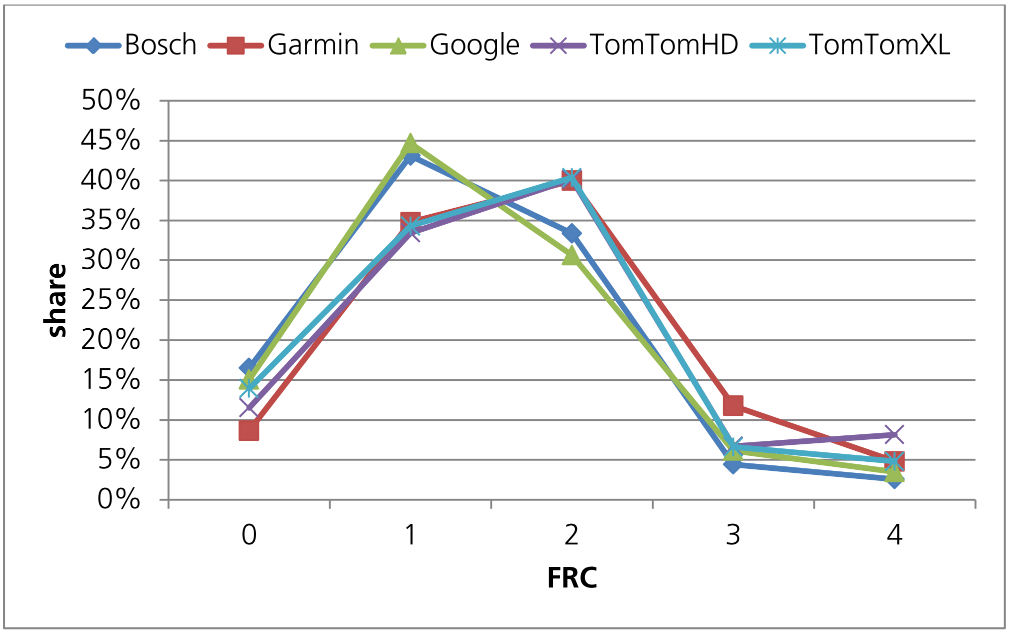


Figure 9: FRC distribution of the five PND's.

4 Remarks of the teams

During the drives, the teams were requested to note anything unusual (like re-routings, jams, falsely reported jams) as a comment in the excel sheet. Altogether four different types of remarks and comments had been collected:

1. remarks while driving, this typically occur when noting the ETA from the devices,
2. along with this, they also noted the current traffic state and the state reported by the PND,
3. after each trip the drivers had been requested to give a short account ("positive" and "negative") about this trip,
4. and at the end of the day the filled in a short questionnaire.

Only the on-trip assessment of the traffic state and the post-day questionnaire had been analyzed to a certain detail, for the other two entries the results turned out to be too disperse to be analyzed with a reasonable amount of resources: e.g., altogether 554 post-trip short accounts and 2,850 on-trip remarks have been collected. The data have been extracted from the excel sheets and summarized into separate sheets, which will be provided (together with the original data) to TomTom.

4.1 On-trip remarks

During each trip, the teams should record anything noteworthy. The first entries look as in Table 5 (in German, no translation):

Table 5: First six lines of on-trip remarks file.

File	row	time	remark
Protokoll_Bosch_2011-10-17.xls	20	6:25:47	stockender Verkehr auf B1/B5, Ampelrückstau
Protokoll_Bosch_2011-10-17.xls	21	6:28:10	stockender Verkehr auf B1/B5, Ampelrückstau

Protokoll_Bosch_2011-10-17.xls	22	6:31:10	stockender Verkehr auf B1/B5, Ampelrückstau
Protokoll_Bosch_2011-10-17.xls	23	6:34:05	aufgelöst, aber starker Verkehr
Protokoll_Bosch_2011-10-17.xls	25	6:40:44	Fahrbahnverengung, B1/B5
Protokoll_Bosch_2011-10-17.xls	26	6:41:24	Ende

Although these remarks may contain additional information about the experiences of the teams with the PND's, no systematic evaluation of the remarks had been undertaken. The relevant data are in the excel logbooks, and they are summarized in a separate file which is part of the electronic supplement to this report.

4.2 On-trip assessment on traffic state

During driving, the teams also reported the traffic state, most notably whether there was a jam or not. In addition, they reported the information displayed by the PND about jams. By simply counting how often the teams have reported observed jams and / or the PND have reported jams, a small statistical evaluation can be done that may be used to assess in principle the quality of this information. However, the results have to be taken with care, for the following reasons.
First, in order to do the evaluation, a bunch of entries had been edited by automatically substituting spelling errors, mis-guiding comments and the like. Second, despite the fact that the TomTom XL has no information on traffic state, even for this device the PND traffic state was reported, albeit in a limited number. Thirdly, the sheer number of reports differs strongly between the different PND's for reasons that are so far not understandable: they should have been of a comparable number, if the teams had followed the instructions correctly, because they were strongly coupled to the ETA recordings.

Anyway, Table 6 displays the absolute numbers extracted from the log-files, and Table 7 displays the relative (normalized) result:

Table 6: Number of jam / no jam on-trip observations

PND	jam / pnd jam	jam / no pnd jam	No jam / pnd jam	No jam / no pnd jam	sum	All remarks
Bosch	1	31	2	2	36	368
Google	91	12	115	337	555	1003
Garmin	1	16	0	111	128	806
TomTom XL	0	5	0	34	39	253
TomTom HD	36	32	3	725	796	1023

Table 7: Share of jam / no jam on-trip observations.

PND	jam / pnd jam	jam / no pnd jam	No jam / pnd jam	No jam / no pnd jam	Share of traffic-state remarks
Bosch	2,8%	86,1%	5,6%	5,6%	9,8%
Google	16,4%	2,2%	20,7%	60,7%	55,3%
Garmin	0,8%	12,5%	0,0%	86,7%	15,9%
TomTom XL	0,0%	12,8%	0,0%	87,2%	15,4%

TomTom HD	4,5%	4,0%	0,4%	91,1%	77,8%

In Table 7, the share of the jam / no jam answers is relative to the total number of jam / no jam remarks (therefore, the four columns for each device add up to 100%), while the last column is the share of traffic-state related remarks compared to the total number of all recorded remarks (i.e. all the ETA recordings) – of course, also empty remark lines are counted here, since the teams often reported the ETA only.

4.3 Post-trip short account

AT the end of each trip, the teams had to write one sentence about positive things they would like to add, and one sentence about negative things. The first entries look like in Table 8 (in German, no translation is undertaken):

Table 8: First six lines in post-trip short account file.

File	PND	Pos/ Neg	Remark
Protokoll_Bosch_2011-10-17.xls	Bosch	Pos:	"Route war ok, oder?"
Protokoll_Bosch_2011-10-17.xls	Bosch	Pos:	.(xxx und trotzdem 3. Platz!
Protokoll_Bosch_2011-10-17.xls	Bosch	Pos:	Abbiegeansage gute und rechtzeitige Ankündigung
Protokoll_Bosch_2011-10-17.xls	Bosch	Pos:	bei hoher Geschwindigkeitsüberschreitung, Tempolimiterinnerung
Protokoll_Bosch_2011-10-17.xls	Bosch	Pos:	ETA nur leicht korrigiert, gute Umleitung aufgrund der Stausituation, trotzdem nicht schnellstes Auto
Protokoll_Bosch_2011-10-17.xls	Bosch	Pos:	Fahrer sagt: "Entspannt zu fahren"

Although these remarks may contain additional information about the experiences of the teams with the PND's, no systematic evaluation of the remarks had been undertaken. The relevant data are in the excel logbooks, and they are summarized in a separate file which is part of the electronic supplement to this report.

4.4 Post-day questionnaire

Finally, the teams filled-in a short questionnaire at the end of each day. The answers have been counted and added together, to ease interpretation also an average grade have been computed both for each question as well as for all the answers. The questions had been grouped in four groups (in bold), each of the four groups have individual accountings in their headline (in bold). The results for the different PND's are highlighted by different colors. While the original questions had been in German, this table has been translated into English for this report.

Table 9: Results of the post-day questionnaire for the different PND's

Bosch	excellent	good	average	bad	poor	don't know	mean
Assessment traffic state information	**0**	**4**	**15**	**11**	**0**	**0**	**3,233**
Quantity of the traffic state information	0	1	8	1	0	0	3,000
Information richness (beginning and end of congestion, length recognizable?)	0	2	3	5	0	0	3,300
Correctness (could the jams displayed be seen in real-life?)	0	1	4	5	0	0	3,400
Route assessment	**0**	**8**	**8**	**15**	**7**	**2**	**3,553**
Was the PND successful in circumventing congestion?	0	2	2	5	1	0	3,500
Your feeling about re-routings?	0	3	3	2	1	1	3,111
Has the system chosen the fastest route?	0	2	2	2	3	1	3,667
ETA - assessment: how well was ETA upon departure?	0	1	1	6	2	0	3,900
Route guidance (map)	**1**	**17**	**9**	**2**	**1**	**0**	**2,500**
Was the route always clearly visible?	0	6	4	0	0	0	2,400
If maneuvers had to take place: had they been clearly recognizable?	0	8	1	1	0	0	2,300
The re-routings: had they been recognizable?	1	3	4	1	1	0	2,800
Route guidance (speech)	**1**	**14**	**2**	**3**	**0**	**0**	**2,350**
Were the announcements understandable?	0	8	1	1	0	0	2,300
Were the announcements accurately timed?	1	6	1	2	0	0	2,400
Number	**2**	**43**	**34**	**31**	**8**	**2**	**120**
Total score							**3,000**

Garmin	excellent	good	average	bad	poor	don't know	mean
Assessment traffic state information	**0**	**8**	**11**	**11**	**0**	**0**	**3,100**
Quantity of the traffic state information	0	3	6	1	0	0	2,800
Information richness (beginning and end of congestion, length recognizable?)	0	1	2	7	0	0	3,600
Correctness (could the jams displayed be seen in real-life?)	0	4	3	3	0	0	2,900

Route assessment	**1**	**11**	**16**	**9**	**0**	**3**	**2,892**
Was the PND successful in circumventing congestion?	1	0	6	3	0	0	3,100
Your feeling about re-routings?	0	6	1	0	0	3	2,143
Has the system chosen the fastest route?	0	4	5	1	0	0	2,700
ETA - assessment: how well was ETA upon departure?	0	1	4	5	0	0	3,400
Route guidance (map)	**0**	**11**	**12**	**3**	**1**	**3**	**2,778**
Was the route always clearly visible?	0	4	4	1	1	0	2,900
If maneuvers had to take place: had they been clearly recognizable?	0	5	4	1	0	0	2,600
The re-routings: had they been recognizable?	0	2	4	1	0	3	2,857
Route guidance (speech)	**3**	**8**	**5**	**3**	**1**	**0**	**2,550**
Were the announcements understandable?	1	8	1	0	0	0	2,000
Were the announcements accurately timed?	2	0	4	3	1	0	3,100
Number	**4**	**38**	**44**	**26**	**2**	**6**	**120**
Total score							**2,860**

Google	**excellent**	**good**	**average**	**bad**	**poor**	**don't know**	**mean**
Assessment traffic state information	**0**	**11**	**20**	**14**	**0**	**0**	**3,067**
Quantity of the traffic state information	0	0	9	6	0	0	3,400
Information richness (beginning and end of congestion, length recognizable?)	0	6	7	2	0	0	2,733
Correctness (could the jams displayed be seen in real-life?)	0	5	4	6	0	0	3,067
Route assessment	**10**	**12**	**19**	**8**	**6**	**3**	**2,782**
Was the PND successful in circumventing congestion?	0	0	1	6	6	2	4,385
Your feeling about re-routings?	4	3	6	1	0	1	2,286
Has the system chosen the fastest route?	5	4	5	1	0	0	2,133
ETA - assessment: how well was ETA upon departure?	1	5	7	0	0	0	2,462
Route guidance (map)	**5**	**23**	**14**	**3**	**0**	**0**	**2,333**
Was the route always clearly visible?	2	9	3	1	0	0	2,200

If maneuvers had to take place: had they been clearly recognizable?	1	4	9	1	0	0	2,667
The re-routings: had they been recognizable?	2	10	2	1	0	0	2,133
Route guidance (speech)	**6**	**9**	**6**	**7**	**2**	**0**	**2,667**
Were the announcements understandable?	3	6	3	3	0	0	2,400
Were the announcements accurately timed?	3	3	3	4	2	0	2,933
Number	**21**	**55**	**59**	**32**	**8**	**3**	**178**
Total score							**2,720**

TomTom HD	exce llent	good	aver age	bad	poor	don't know	mean
Assessment traffic state information	**7**	**13**	**8**	**2**	**0**	**0**	**2,167**
Quantity of the traffic state information	3	6	1	0	0	0	1,800
Information richness (beginning and end of congestion, length recognizable?)	3	3	2	2	0	0	2,300
Correctness (could the jams displayed be seen in real-life?)	1	4	5	0	0	0	2,400
Route assessment	**1**	**17**	**12**	**2**	**5**	**3**	**2,811**
Was the PND successful in circumventing congestion?	0	3	4	1	2	0	3,200
Your feeling about re-routings?	0	5	1	1	3	0	3,200
Has the system chosen the fastest route?	1	5	1	0	0	3	2,000
ETA - assessment: how well was ETA upon departure?	0	4	6	0	0	0	2,600
Route guidance (map)	**1**	**16**	**5**	**5**	**2**	**1**	**2,690**
Was the route always clearly visible?	1	6	1	2	0	0	2,400
If maneuvers had to take place: had they been clearly recognizable?	0	7	2	1	0	0	2,400
The re-routings: had they been recognizable?	0	3	2	2	2	1	3,333
Route guidance (speech)	**6**	**8**	**2**	**4**	**0**	**0**	**2,200**
Were the announcements understandable?	5	3	0	2	0	0	1,900
Were the announcements accurately timed?	1	5	2	2	0	0	2,500
Number	**15**	**54**	**27**	**13**	**7**	**4**	**120**
Total score							**2,509**

TomTom XL	exce llent	good	aver age	bad	poor	don't know	mean
Assessment traffic state information	**4**	**0**	**0**	**0**	**12**	**3**	**4,000**
Quantity of the traffic state information	2	0	0	0	4	1	3,667
Information richness (beginning and end of congestion, length recognizable?)	1	0	0	0	4	1	4,200
Correctness (could the jams displayed be seen in real-life?)	1	0	0	0	4	1	4,200
Route assessment	**8**	**9**	**9**	**7**	**6**	**0**	**2,846**
Was the PND successful in circumventing congestion?	4	0	0	2	3	0	3,000
Your feeling about re-routings?	3	1	2	0	3	0	2,889
Has the system chosen the fastest route?	0	6	2	1	0	0	2,444
ETA - assessment: how well was ETA upon departure?	1	2	5	4	0	0	3,000
Route guidance (map)	**8**	**4**	**10**	**4**	**0**	**1**	**2,385**
Was the route always clearly visible?	3	1	4	1	0	0	2,333
If maneuvers had to take place: had they been clearly recognizable?	2	1	5	1	0	0	2,556
The re-routings: had they been recognizable?	3	2	1	2	0	1	2,250
Route guidance (speech)	**5**	**9**	**2**	**2**	**0**	**0**	**2,056**
Were the announcements understandable?	2	6	0	1	0	0	2,000
Were the announcements accurately timed?	3	3	2	1	0	0	2,111
Number	**25**	**22**	**21**	**13**	**18**	**4**	**103**
Total score							**2,768**

5 Summary and Conclusions

This report demonstrates clearly the differences as well as the similarities between the five PND's that has been tested with respect to their ability to do routings. The focus was especially on routing with respect to real-time traffic information, and how the devices make use of them. Three of the devices had been genuine PND's, while the remaining two had been used as so called Apps on a smart-phone. Arguably, the smart-phone-based PND's have a small disadvantage in terms of their hard-ware, since one may assume that the GPS receiver of the smart-phone may not deliver the same quality as the GPS-receiver of the genuine PND's. And interestingly, the total ranking is just like this: the three genuine PND's are on the top, while the two Apps are at the end of the table. Note, however, that this is not a proof that a genuine PND is superior to an App, there are little doubts that the performance of the TomTom App running on an iPhone would have given very similar, if not identical results as for the TomTom HD. (It would have been easy to test that just by putting an iOS or Android device with the HD-App in the same car and try to monitor the differences between the two – this is for future reference.)

The results of this campaign show, that there is a roughly 12% difference in travel time between the best and the worst PND. This difference is not very large but statistically highly significant. The closer analysis in section 3 show, that the mechanism responsible for this difference is the fact that the TomTom devices sometimes found routes that were considerably better than what the other devices found. This seem to happen in about 25% of all the cases, while in the remaining 75% the difference in the travel times is fairly small. In addition, those different routes also display a shorter waiting time, again with the gain in waiting time being in the same order of magnitude as the overall gain.

When looking at the FRC distribution, it turns out that the routed vehicles spend most of their time in the FRC classes 0, 1, and 2. The share of the FRC 4 classes and higher is well below 10%, in accordance with a separate assessment done by DLR in a recent project. It seems, that the HD routes nevertheless utilize these FRC 4 classes and higher to a larger extend than, e.g., the TomTom XL: a value of 9% stands against a value of 5%.

For the estimated time of arrival (ETA) it has to be stated, that it is not as well as it could be. However, it has to be added, that the ETA measurement in this campaign was not as precise as, e.g., the speed distributions that had been computed from the traces of the GPS loggers. Nevertheless, after a thorough cleaning of the manually recorded ETA values it can be concluded, that the devices typically and systematically underestimate the time of arrival – by 10% for the best PND's, and up to 30% for the worst ones. Fortunately, the error in the ETA gets smaller with an increasing part of the trip complete.

In general, doing the same campaign in a different city and with different drivers may lead to different results. The same is true if it is to be done one year in the future, when the coverage with real-time information has changed.

Finally, a few remarks about potential for improvement will be made. Within this campaign, there is little that can be changed to improve the results. Of course it would have been nice to compare against normal drivers that travel to the same destination, but without a PND. But this definitely needs a completely different approach. Also true is, that the 80+ routes in this sample are still a fairly small number from a statistical point of view, however as has been demonstrated in section 3.2, this number is enough to draw firm statistical conclusions.

From the viewpoint of TomTom, it might be an interesting endeavor to monitor the difference between the XL and the HD devices. This can be done completely within TomTom's system, and it can be done with any desired degree of completeness and precision. The most direct approach is of course to compare directly two trips connecting the same origin and destination, however it is highly unlikely that this happens very often even in the large data-bases that TomTom has at their disposal. Therefore, looking at the average speed, or better, the speed distribution might be more fruitful in order to compare the two approaches systematically.