This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2020.2986820, IEEE Consumer Electronics Magazine

Precision of Wearable GPS in Marathon Races

Javier Lluch, Miguel Rebollo, Ángeles Calduch-Losa and Ramón Mollá
Universitat Politècnica de València

***Abstract*—This study determines the precision of the most relevant GPS models used in marathon races. It uses public data of the participants in the leading marathons. We have retrieved 73,865 records from 85 GPS models. There are differences in the precision obtained by road GPS models compared with models designed for trail races and mobile phones. The precision also depends on the finish time: the longer the race takes, the higher the error is. No evidence of differences among the studied brands appear. The results can be helpful for manufacturers to get field information about the behavior of the devices in real conditions. And it can be beneficial for end-users also since the data help the buy decision. On the other hand, with this information, athletes could have available more accurate measures about their pace and other data during a marathon.**

I. Overview

The evolution of technology and the rise of sportive practice have favored an active industry for the development of technological devices that improves the performance of athletes providing data regarding the execution of their training sessions. The use of wearables is increasingly widespread. Systems that can detect posture and the heart rate are essential for the monitoring of athletes[1]. They store the data on websites for later analysis[2]. The data extracted from both daily and sports activities help monitor the health care of the users[3].

Among the most common devices, GPS watches provide relevant information for a race, such as pace, distance, or height. GPS is not safe from making mistakes when calculating the mileage traveled. It is usual, especially in novice runners, to raise questions arguing that the race was wrongly measured. Many factors affect the accuracy of a GPS: width of the streets, altitude of buildings, unevenness, the existence of high voltage cables, trees, cloudy skies, and any other that hinders a good reception from satellites (Fig. 1).

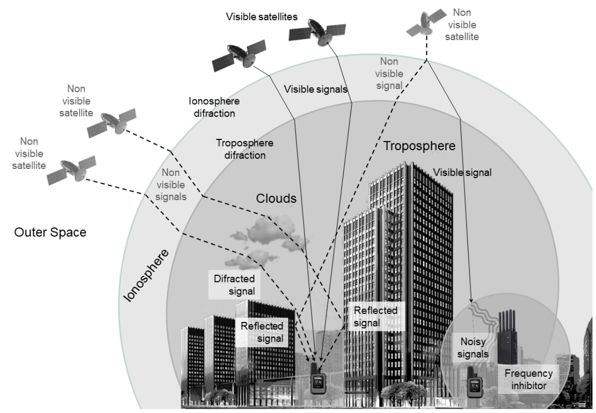

Fig. 1: Multiple sources affect the GPS precision

The aim of this work is (i) to evaluate the precision of the different models of GPS with more presence in the most relevant marathons; and (ii) validate empirically the results of other studies that affirm that the finish time sways to the total distance measured by the GPS.

II. Background

To determine the precision of the GPS devices is relevant since runners take seriously the values thrown by their devices. This fact has health implications for having correct lectures to avoid unnecessary injuries[4]. Given the difficulty of obtaining significant samples, there are few references in the scientific literature that analyze exhaustive sample sets. Bauer[5] has studied the precision of the nine apps for smartphones, making a 500-meter round trip, measuring it with a single phone model. Leong Lee et al.[6] carried out a more exhaustive study on the accuracy and precision of six smartphone models. They used 810 measurements, defining three different protocols. Pobichurin et al.[7] carry out a study similar to ours in which they investigate the precision of the measures taken by smartphones and GPS watches with a sample of 262 runners for the Trollinger-Marathon with half and marathon

2162-2248 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: UNIVERSIDAD POLITECNICA DE VALENCIA. Downloaded on July 20,2020 at 09:25:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2020.2986820, IEEE Consumer Electronics Magazine

Fig. 2: All these laps were run over the same line. The deviations in the measure are noticeable

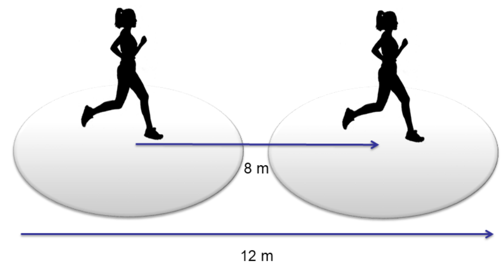

Fig. 3: A popular athlete can run 1 km in 4'10" This is 8 m in 2 seconds. With an error 2 m, it can give a measure of 12 m, which means run at a 2'45" pace, faster than the marathon WR.

distances. This race has the characteristic that it is disputed in an open zone with good satellite reception. The average distances obtained were 21.154 km for the half and 42.385 km for the marathon.

Besides the precision of GPS, the time needed to finish the race is another factor that affects the final measure. Haney and Mercer[8] found a relation between the variability of the pace and the performance on the marathon using data from GPS readings. However, they had to remove many records due to lousy precision. Hubble and Zhao[9] analyze the difference in the performance of men and women in the Houston Marathon using the data provided by the organizers at different kilometric points. The relevant conclusion of both works is the existence of a correlation between pace and finish time. Schipperijn et al.[10] compared the precision of one GPS model with walking, cycling, and vehicle lane. They combined several high-precision methods and calculated the differences in alternative scenarios from open spaces to narrow urban 'canyons.' The results confirm that higher speeds increase the precision.

With a smartphone, the signal of GNSS combined with NRTK (Network Real-Time Kinematics) positioning reduces distance-dependent error[11], but it is not available for watches. Besides, there are new chips like the BCM47755, that achieves an accuracy of 30 cms, but no sports device uses it[12].

III. Methods

The data for this observational research has been obtained from the public information for tracking sportive activities. The analyzed races have been the marathons of Berlin, Boston, Chicago, London, New York, and Valencia. For each one of them, we have obtained the two last editions among the three available at the moment of the analysis (years 2016, 2017, or 2018). The subjects of the study are popular athletics that have participated in the mentioned marathons and have published their results voluntarily. The retrieved information has been: race, runner id (anonymized), covered distance, finish time, and GPS model.

TABLE I: Participants and distance (mean± SD)

city	year	#finishers	#retrieved	distance (km)
Berlin	2016	36 054	3 769	42.74 ± 0.26
Berlin	2017	39 101	6 403	42.71 ± 0.27
Boston	2017	26 400	5 278	**42.52** ± **0.11**
Boston	2018	25 831	5 904	42.53 ± 0.12
Chicago	2016	40 608	3 714	44.22 ± 0.97
Chicago	2017	44 508	7 023	**44.65** ± **1.27**
London	2017	39 281	9 929	42.93 ± 0.52
London	2018	40 255	12 185	43.02 ± 0.56
New York	2016	51 388	5 505	42.89 ± 0.58
New York	2017	50 766	8 528	42.89 ± 0.62
Valencia	2016	15 858	2 113	42.70 ± 0.22
Valencia	2017	16 169	3 547	42.74 ± 0.27
Total		426 219	73 898	

The results have been filtered to discount the effect of anomalous data. This deviation can be due to human causes, such as delay in starting the reading, involuntarily pauses, forgetfulness in the provisioning of after passing the finish line, or other reasons not attributable to the GPS. We have been filtered the data using the interquartile range, keeping the lectures in $Q_3 \pm 1.5(Q_3 - Q_1)$.

2162-2248 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: UNIVERSIDAD POLITECNICA DE VALENCIA. Downloaded on July 20,2020 at 09:25:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2020.2986820, IEEE Consumer Electronics Magazine

The final data comprises 73,898 records after filtering outliers. Chicago gives the longest distances, followed by New York and London. Berlin, Boston, and Valencia through a similar result (see Table I). In total, 85 different GPS models have been identified. To select the GPS models to analyze, we have followed the 80-20 Pareto's law. We have chosen those devices whose apparition in the races sum the 80% of the total records. With these models, we have more than 63,000 readings. Similar models, such as Garmin Forerunner 220 and 225, appear reunited as 22x.

IV. Results

A. Device Classification

The devices have been classified according to their usage. Initially, we have considered four categories: road, triathlon, trail, and mobile apps. However, after a Tukey–Kramer's test over the ANOVA with level of significance $\alpha = 0.05$, there is no evidence to separate road and triathlon-specific models ($p = 0.94$) (see Table II). Therefore, we maintain three categories: road models, trail running, and cell phones. Despite there is no evidence for significative differences between trail and phone-based ones ($p = 0.15$), we keep the classification because (i) they are different devices, and (ii) when it has been analyzed in an individual race, small differences become significant.

B. Precision of the Devices

The second analysis calculates the average distances, aggregated by model. We pose a hypothesis test with a level of confidence of 95% ($\alpha = 0.05$). The null hypothesis is H_0: *there are no significant differences in the averages*, whereas the alternative is H_a: *at least one of the averages is significantly different from the other*. Figure 4 (top) shows the average distances in Valencia 2017 Marathon. The error bars represent the standard error of the data. Mobile apps have the highest distances, followed by trail devices (Garmin Fenix models).

TABLE II: Result of the confidence intervals.

	diff	conf. int.	p–value
road-phone	-0.55	[-0.60, -0.49]	0.00
trail-phone	-0.05	[-0.11, 0.01]	**0.15**
tri-phone	-0.55	[-0.61, -0.49]	0.000
trail-road	0.50	[0.48, 0.52]	0.00
tri-road	0.00	[-0.02, 0.01]	**0.94**
tri-trail	-0.50	[-0.53, -0.48]	0.00

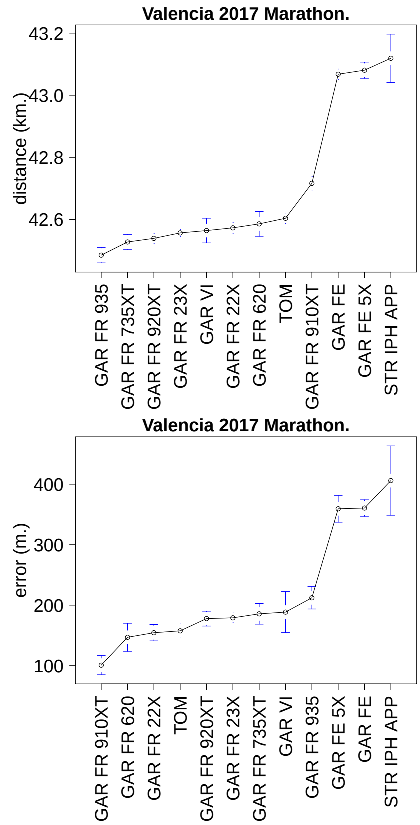

Fig. 4: Measured distance and relative error before (top) and after (bottom) the correction

An ANOVA and Tukey–Kramer's test indicates that the differences observed in the Figure 4 are significant, with a difference of -0.35 between the last road model (Garmin FR 910XT) and the first trail model (Garmin FE), being $p = 1e-5$. Results show similar behavior in different races. When the data are aggregated by type, the three groups: mobile, trail, and road models, reject the hypothesis of equality of means (see Table II), with differences of -0.55 between road and mobile, 0.50 between trail and road (both with $p = 1e-7$). Still, the difference between trail and mobile -0.05 is not significant ($p = 0.15$). We can conclude that the GPS model is a factor that affects the measured distance, and there are substantial differences between road and trail models.

2162-2248 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: UNIVERSIDAD POLITECNICA DE VALENCIA. Downloaded on July 20,2020 at 09:25:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2020.2986820, IEEE Consumer Electronics Magazine

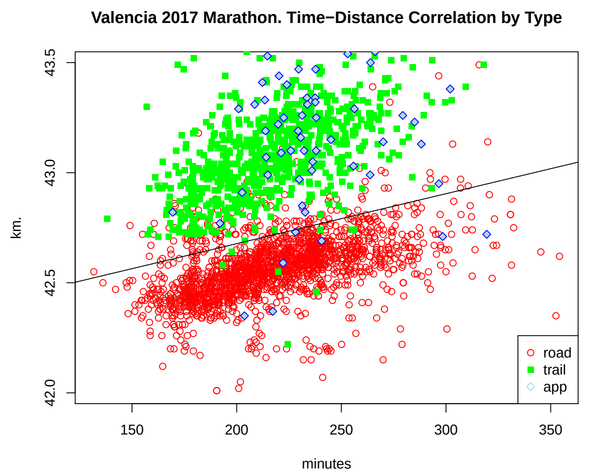

Fig. 5: Correlation among the time involved in the marathon and the distance measured by the GPS

C. Dependence from the Time/Pace

A third test consists of determining if the time needed to finish the marathon affects the measure. In this case, we make a correlation study to see if the more prolonged the runner lasts, the longer is the measure of the GPS.

Figure 5 shows the correlation between the time involved and the total distance in Marathon of Valencia. All the cases through similar results. In the case of Valencia, the correlation index is low ($r = 0.17, R^2 = 2.89\%$), but the low p-value $p = 1.2e - 15$, suggests that it is a small but significant effect, despite it do not explain the complete variability of the data. The same effect appears in the rest of the analyzed races. Therefore, we have to discount the effect of the time in the error obtained by the GPS. The three populations identified correspond to each type of watch.

There is evidence that GPS devices introduce an excess systematically in the measures. Our results indicate that there is a significant effect that depends on the time runners have needed to finish the race and the total distance given by their devices, being a positive correlation. That implies that slower runners obtain, in general, longer distances than the fastest ones. This effect has been observed in all races, with a coefficient of R^2 between 0.6% and 4%. It is a low correlation, but the small p-value $p = 1e - 15$ in all cases indicates that the effect is significant. Therefore, we can conclude that there are other factors, besides time, that affect the distances, so dedicated time does not explain the variability of the data altogether.

D. Precision after Distance Correction

To correct the effect on the distance covered in the race, the effect of the involved time has been discounted as follows. For each runner, the total official distance (42.195 km) is increased using the regression model obtained for the current marathon taking into account the time. Therefore, each participant has a different length, which depends on the time he or she has invested to finish the marathon. Then, the relative error is the difference between the measure from the GPS and this estimated distance.

Figure 4 shows the difference between the first measure and the results adjusted for each participant. The average error obtained changes in the order of the GPS. Nevertheless, the difference between road models and the rest maintains. After correcting the deviation, the hypothesis still fulfills. There is a change in the order of the GPS models, but the differences obtained in the confidence intervals among the three groups (road, trail, and phone) are significant (see Table III).

V. Conclusions

We have analyzed the data available from two editions of six marathons: Berlin, Boston, Chicago, London, New York, and Valencia.

An ANOVA over the type of device separates them into three categories: road, trail, and mobile devices. The differences obtained by the road models are not significant, but we observe a clear separation with trail and mobile models. Probably it is due to changes in the design to include other elements, such as a barometer, or use process time in the calculation of health and performance data (e.g., VO_2 max), with a smaller size and weight[13]. There are algorithms such as that use inertial sensors to increase accuracy significantly, but with a computational cost that is not acceptable for this type of device. Newer devices do not increase the precision, and the brand is not relevant. Mobile phones are the devices with the highest deviation.

In the second place, a positive correlation between the time involved in the race and the distance measured by the GPS has been observed. Despite

2162-2248 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: UNIVERSIDAD POLITECNICA DE VALENCIA. Downloaded on July 20,2020 at 09:25:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2020.2986820, IEEE Consumer Electronics Magazine

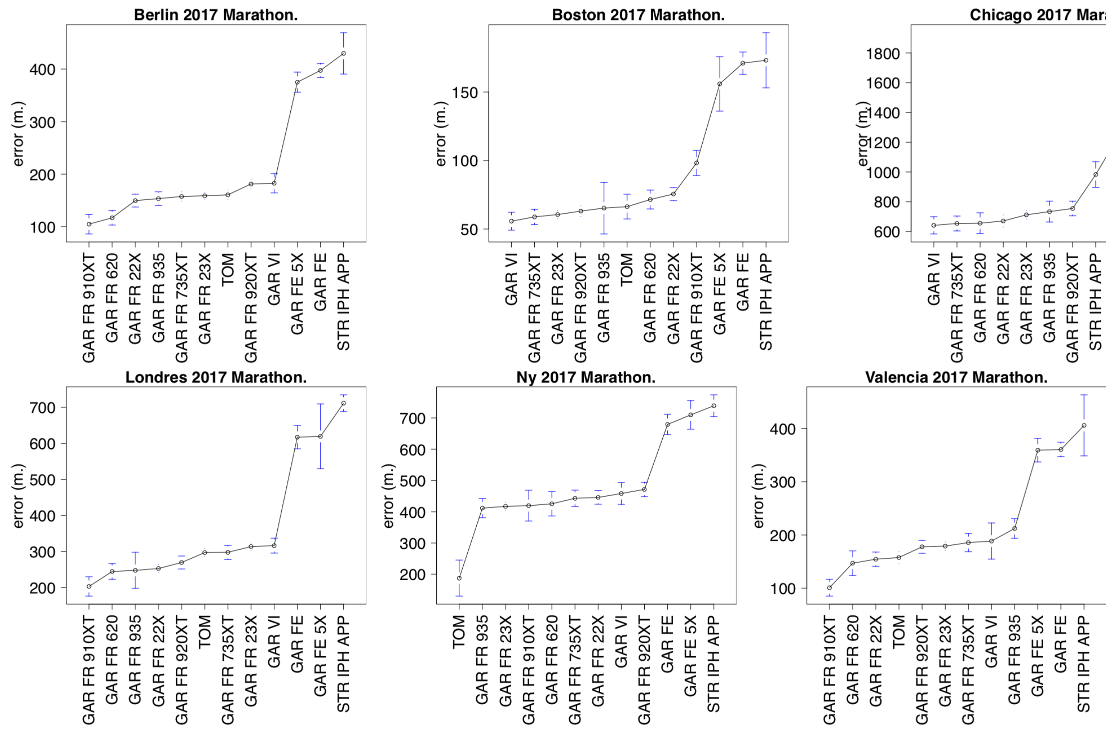

Fig. 6: Error obtained by the 12 most used devices in once the effect of the time has been discounted. In general, There is a significative difference between road models and the rest (trail and phone apps)

a low coefficient, the effect can be considered as significant. This effect has been taken into account to calculate the average error of each GPS model.

The knowledge of the devices has a direct impact on the performance of popular athletics. It allows correcting the results obtained in training, giving a more accurate view of their performance. As the measures overestimate the distance, the real pace is always faster. Furthermore, we can create applications that take into account the effect to provide the fittest feedback to the athletes while they are training. There are many factors to consider when purchasing a running watch: the kind of activity (indoor, outdoor, multisport, trail, ultra), battery life, map display, and tracking, monitoring daily activity, size, or design. Following the study data for a runner, the best choice would be a FR 235, to multisport a FR 920XT or FR 735. For trail races, it is difficult to choose any since all of them have a significant error. Anyway, Garmin Fenix 5 is one of the most popular models and it throws the best results in its group. Nevertheless, it is severely affected by the height of the buildings, which explains the bad results in cities such as NY and Chicago.

References

[1] E. Prawiro et al. A wearable system that detects posture and heart rate: Designing an integrated device with multiparameter measurements for better health care. *IEEE CEM*, 8(2):78–83, 2019.

[2] T. Coughlin. Digital storage in smartphones and wearables [the art of storage]. *IEEE CEM*, 7(2):108–120, 2018.

[3] P. Sundaravadivel et al. Everything you wanted to know about smart health care: Evaluating the different technologies and components of the internet of things for better health. *IEEE CEM*, 7(1):18–28, 2018.

[4] B. Zhang, J. Guo, and D. Xu. Wearable pace speed monitoring and reminding system. In *ICITBS Int Conf*, pages 553–556, 2018.

[5] C. Bauer. On the (in-)accuracy of gps measures of smartphones: A study of running

2162-2248 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: UNIVERSIDAD POLITECNICA DE VALENCIA. Downloaded on July 20,2020 at 09:25:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2020.2986820, IEEE Consumer Electronics Magazine

tracking applications. In *11th MoMM Int Conf*, pages 335–340, 2013.

[6] L. Lee et al. Comparison of accuracy and precision of gps-enabled mobile devices. In *IEEE CIT Int Conf*, pages 73–82, 2016.

[7] M. Pobiruchin et al. Accuracy and adoption of wearable technology used by active citizens: A marathon event field study. *JMIR mHealth uHealth*, 5(2):e24, 2017.

[8] TA. Haney and JA. Mercer. A description of variability of pacing in marathon distance running. *Int J Exerc Sci*, 4(2):133–140, 2011.

[9] C. Hubble and J. Zhao. Gender differences in marathon pacing and performance prediction. *J Sports Anal*, 2:19–36, 2016.

[10] S. Jasper et al. Dynamic accuracy of gps receivers for use in health research: A novel method to assess gps accuracy in real-world settings. *Front Pub Health*, 2:e21, 2014.

[11] P. Dabove and V. Di Pietra. Towards high accuracy gnss real-time positioning with smartphones. *Adv Space Res*, 63(1):94 – 102, 2019.

[12] SK. Moore. Super-accurate gps coming to smartphones in 2018 [news]. *IEEE Spectrum*, 54(11):10–11, 2017.

[13] Y. Chen, S. Zhao, and J.A. Farrell. Computationally efficient carrier integer ambiguity resolution in multiepoch gps/ins: A common-position-shift approach. *IEEE TCST*, 24(5):1541–1556, 2016.

Javier Lluch is Associate Professor at Universitat Politècnica de València (UPV), Valencia, SPAIN. He is a board member of the Institute of Industrial Control Systems and Computing. Contact him at jlluch@dsic.upv.es.

Miguel Rebollo is Associate Professor at UPV and a member of the Valencian Research Institute on Artificial Intelligence. Contact him at mrebollo@vrain.upv.es.

Ángeles Calduch-Losa is an Assistant Professor in the Department of Applied Statistics and Operational Research, and Quality at UPV. Contact her at mcalduch@eio.upv.es.

Ramón Mollá is an Associate Professor with the Department of Computer Systems and Computation, UPV. Contact him at rmolla@dsic.upv.es.

TABLE III: Mean error obtained by the GPS models in each one of the editions.

model	type	total		Berlin'16			Berlin'17			Boston'17			Boston'18			Chicago'16			Chicago'17	
		n	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.
GAR FR 22X	road	7 497	347	156	±12	422	150	±13	779	76	±5	528	82	±6	620	479	±32	738	670	±41
GAR FR 23X	road	13 989	337	166	±13	1 085	159	±7	1 031	60	±3	1 507	71	±3	464	583	±45	1 678	711	±28
GAR FR 620	road	3 277	266	131	±15	252	117	±14	377	72	±7	200	78	±10	274	521	±52	281	655	±69
GAR FR 735XT	road	4 085	157	157	±14	495	157	±8	280	59	±6	418	67	±5	127	571	±85	491	653	±50
GAR FR 910XT	road	1 974	165	104	±16	131	105	±19	232	98	±9	121	94	±13	149	905	±77	92	1400	±140
GAR FR 920XT	road	6 226	426	178	±10	543	181	±9	669	63	±4	574	69	±5	368	622	±50	653	754	±49
GAR FR 935	road	1 953	-	-	-	247	153	±13	16	65	±19	457	66	±5	-	-	-	314	733	±70
GAR VI	road	2 712	78	158	±21	150	183	±19	191	56	±6	222	74	±9	160	581	±81	350	641	±58
TOM	road	3 554	239	184	±12	394	161	±9	123	66	±9	105	74	±15	114	1430	±70	124	1720	±60
GAR FE	trail	6 804	431	417	±18	688	397	±13	374	171	±8	321	173	±8	293	1070	±83	598	1240	±70
GAR FE 5X	trail	2 509	-	-	-	348	375	±19	42	156	±20	309	161	±9	-	-	-	215	1810	±120
STR IPH APP	phone	5 791	131	421	±38	138	430	±40	96	173	±20	49	214	±31	354	676	±61	337	983	±87

model	type	total		London'17			London'18			NY'16			NY'17			Valencia'16			Valencia'17	
		%	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.	n	mean	C.I.amp.
GAR FR 22X	road	9.4%	1 132	253	±11	855	263	±15	764	421	±20	740	446	±22	184	135	±13	230	154	±13
GAR FR 23X	road	17.5%	1 609	314	±11	2 714	327	±8	756	435	±20	1 795	417	±13	168	157	±15	584	179	±8
GAR FR 620	road	4.1%	455	244	±22	336	216	±21	338	422	±33	277	425	±39	59	130	±31	70	147	±23
GAR FR 735XT	road	5.1%	380	298	±20	673	308	±15	279	439	±36	533	443	±26	54	139	±19	129	186	±17
GAR FR 910XT	road	2.5%	265	203	±27	147	381	5±1	182	441	±49	159	419	±49	148	94	±13	138	101	±16
GAR FR 920XT	road	7.8%	600	269	±18	505	321	±17	561	483	±26	715	471	±22	173	158	±12	286	178	±13
GAR FR 935	road	2.4%	32	248	±50	340	349	±24	-	-	-	398	412	±31	-	-	-	98	212	±18
GAR VI	road	3.4%	484	316	±20	433	329	±24	177	508	4±8	298	458	±35	46	155	±32	60	189	±34
TOM	road	4.5%	925	297	±11	867	395	±18	218	215	±24	22	188	±58	153	131	±18	216	158	±12
GAR FE	trail	8.5%	566	617	±32	658	617	±27	495	726	±40	687	679	±32	303	308	±18	529	361	±14
GAR FE 5X	trail	3.1%	70	619	±90	551	615	±29	-	-	-	403	710	±46	-	-	-	180	359	±22
STR IPH APP	phone	7.3%	928	711	±23	1 080	743	±23	526	555	±34	614	739	±35	37	351	±53	54	406	±57

2162-2248 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: UNIVERSIDAD POLITECNICA DE VALENCIA. Downloaded on July 20,2020 at 09:25:49 UTC from IEEE Xplore. Restrictions apply.