PREPRINT · RESEARCH ARTICLE · PUBLIC TRANSPORT

Accuracy of open-data versus operator real-time bus arrival predictions

A paired evaluation of the Bide+ method in Biscay, Spain

Yeray Garrido Independent researcher, Bilbao · ORCID 0009-0008-0459-0294

Cite as: Garrido, Y. (2026). Accuracy of open-data versus operator real-time bus arrival predictions: a paired evaluation of the Bide+ method in Biscay, Spain. Preprint v1.0, Zenodo. doi:10.5281/zenodo.24286027.

Key results

16%	39%	64%
Published timetable	Official real time	Bide+ (open data only)

share of predictions within two minutes of the actual passage · 555,744 paired predictions

- Same bus, same stop, same minute: 59 stops of the Bizkaibus network, six days in November 2026.
- Equally accurate below three minutes (90%); from three minutes on, Bide+ is ahead at 57 of 59 stops, every hour and every day.
- The gap comes from the travel times Bide+ learns between consecutive stops.

Abstract This paper compares an operator's real-time bus predictions with an independent predictor for the same buses at the same moments, for Bizkaibus, the interurban network of Biscay (Spain), which publishes vehicle positions and timetable as open data and shows its own waiting times in an app. For one week in November 2026, the official waiting times of 60 stops were recorded every minute with the position feed. Each official announcement was matched to the bus it referred to, confirmed by the distance the service reports, and compared with the published timetable and with the Bide+ method, which predicts arrivals from the open data alone and was frozen before the week. All three were checked against the actual passage, reconstructed from the feed and from GPS positions. Across 555,744 paired predictions on six days, the timetable was within two minutes of the actual passage in 16% of cases, the official real time in 39% and Bide+ in 64%. Below three minutes the two real-time sources were equally often within two minutes, although the official one had the smaller mean error below six minutes; from three minutes on, Bide+ was more often within two minutes, by 29 to 35 percentage points at 10–20 minutes (95% interval by stop), at 57 of 59 stops and at every hour. An ablation attributes its advantage to learned segment travel times. The results apply to the 54% of first-bus announcements whose bus was tracked and confirmed, and to one week; data and code are open.

Keywords bus arrival prediction · real-time passenger information · automatic vehicle location · paired forecast evaluation · open data · GTFS · SIRI

1 Introduction

Real-time passenger information (RTPI) shortens the time passengers wait at stops and makes that wait less uncertain (Watkins et al. 2011; Dziekan and Kottenhoff 2007; Brakewood and Watkins 2019), but only when it is accurate: wrong predictions erode trust and can make passengers miss

the bus they were told to expect (Gooze et al. 2013; Rahman et al. 2013). Many passengers receive these predictions from the operator, through its app and stop displays, or from journey planners that relay the operator's figures. Whether a third party with only the operator's *open* data can predict better than the operator is rarely measured, and no study was found that does so on the same buses and at the same moments.

This paper makes that measurement for Bizkaibus, the interurban bus network of the province of Biscay (Spain). Bizkaibus shows its own real-time waiting times in an official app and on some stop displays; in this paper they are called the *official real time*. It also publishes, as open data, the position of every vehicle in service, obtained by the Global Positioning System (GPS) and published as a SIRI Vehicle Monitoring feed (Service Interface for Real-time Information, CEN 2022), and its timetable in the General Transit Feed Specification (GTFS, MobilityData 2026b), but not its predictions. The Bide+ method predicts arrivals from those two open sources alone. It is documented in a technical report (Garrido 2026a) and specified for reimplementation in a companion paper (Garrido 2026b), both of which evaluated it against the published timetable only.

The research question is:

> *For the same bus, at the same stop and at the same minute, how close to the actual passage are the published timetable, the real-time waiting time announced by Bizkaibus and the arrival predicted by the Bide+ method?*

For one week in November 2026 the official waiting times of 60 stops were recorded every minute, together with the live feed every 20 seconds. Each official announcement was paired with the bus it referred to, confirmed by the distance that the official service gives, with that bus's published time at the stop, with the Bide+ prediction for that bus at that minute, and with the moment the bus actually passed, reconstructed in two independent ways.

The contributions are:

1. A paired protocol to compare a third-party predictor with an operator's real-time predictions when the operator publishes positions but not predictions: what to capture, how to match each announcement to a vehicle and confirm it, how to reconstruct the actual passage and check it against an independent GPS-based passage, and how to report coverage.
2. The comparison of the three sources over 555,744 paired predictions on six normal days, by actual and announced time to arrival, with inference that accounts for dependence within trips and stops, a passenger decision model, an ablation of the learned components of Bide+, and a set of sensitivity analyses.
3. The capture program, the raw data, the timetable and learned tables used, and the analysis, openly available, so that the comparison can be repeated on other days or networks.

Section 2 reviews related work, Section 3 describes the three sources, Section 4 the data, Section 5 the comparison design, Section 6 the results, Section 7 their interpretation and limitations, and Section 8 concludes. Supporting tables and figures are in Appendix A.

2 Related work

Arrival prediction from automatic vehicle location. Bus arrival prediction from automatic vehicle location (AVL) data has a long history. Early methods combined the last observed position with historical travel times between stops or timepoints (Lin and Zeng 1999; Cathey and Dailey 2003). Later work used Kalman filters (Cathey and Dailey 2003; Shalaby and Farhan 2004; Vanajakshi et al. 2009), artificial neural networks (Chien et al. 2002; Jeong and Rilett 2005) and, for stops served by several routes, models that share information across routes (Yu et al. 2011). Mazloumi et al. (2010) documented how travel time varies by time of day and segment, the variation that segment-based methods learn. Reviews (Moreira-Matias et al. 2015; Singh and Kumar 2022; Kumar et al. 2025)

compare methods with each other or with the timetable, and deep-learning predictors such as convolutional recurrent networks (Petersen et al. 2019) are now common. This paper does not compare Bide+ with that state of the art; it compares it with what passengers are actually shown by the operator. The method evaluated here belongs to the family of segment-based historical methods (Lin and Zeng 1999; Cathey and Dailey 2003): it learns the travel time of each segment between consecutive stops by time band, with shrinkage towards the all-day value, and adds the position of the vehicle.

Accuracy of operator real-time information. Evaluations of the predictions that operators actually show passengers are rare. Cats and Loutos (2016) evaluated the bus arrival information of Stockholm against AVL records and found that errors grow with the prediction horizon and depend on the stop and the time of day. A search of the literature found no study that compares an operator's live predictions with an independent predictor on the same vehicles and moments.

Value of accurate information for passengers. Real-time information reduces actual and perceived waiting times (Watkins et al. 2011; Dziekan and Kottenhoff 2007; Mishalani et al. 2006; Liu and Miller 2020), is valued by passengers (Ferris et al. 2010; Rahman et al. 2013) and can increase ridership (Tang and Thakuriah 2012; Brakewood et al. 2015); Brakewood and Watkins (2019) review these benefits. They depend on accuracy: wrong predictions reduce trust and lead to missed buses (Gooze et al. 2013).

Third-party predictors and open data. Predictors built only on public data are feasible: García-Mauriño et al. (2024) improve arrival predictions using only data available through public transport APIs, and Wessel et al. (2017) reconstruct actual passage times from a public vehicle location feed, the kind of ground truth used here.

Comparing predictors. Comparing two predictors on the same targets is a classic forecasting problem. The loss differential should be tested with errors that account for dependence (Diebold and Mariano 1995), the scoring rule should match the decision it serves (Gneiting 2011), and evaluation must be out of sample (Tashman 2000), with no information from the test period leaking into the predictor (Kaufman et al. 2012).

3 The three sources

3.1 The published timetable

Bizkaibus publishes its timetable as GTFS. Its intermediate times are interpolated at constant speed between the terminals, not observed, and on this network they drift from the actual passage as the trip advances: after the first hour of a trip most buses pass a stop more than two minutes before the published time (Garrido 2026b). The timetable is what a passenger sees in printed and online timetables and in any app without real-time data. Its prediction for a bus is the published time of its trip at the stop.

3.2 The official Bizkaibus real-time waiting times

The official app and some stop displays show, for each stop, the waiting time of the next buses of each line. The app obtains them from a public web service (`GetPasoParadaMobile_JSON`) that, for a stop, returns the next two buses of every line that serves it, each with its waiting time in whole minutes, its distance to the stop in metres and a type field (empty in every response captured). How these times are computed is not published.

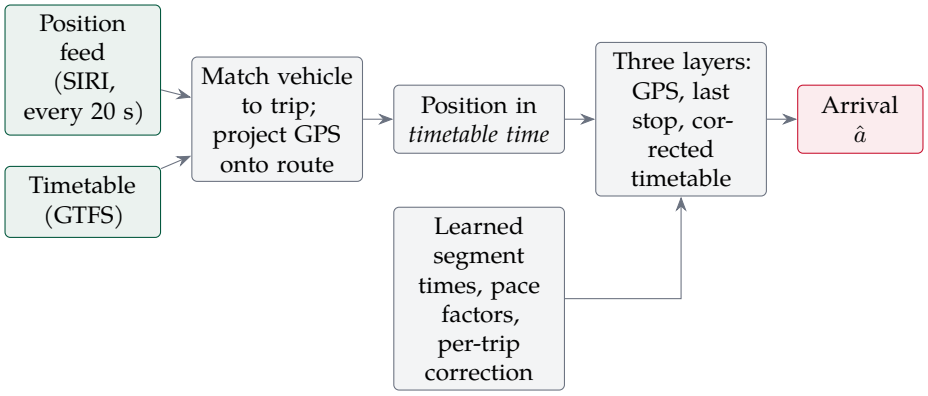


Figure 1. Outline of the Bide+ method (Garrido 2026a; Garrido 2026b). Only the operator's open data enter it: the position feed and the timetable.

3.3 The Bide+ method

The Bide+ method (Garrido 2026a; Garrido 2026b) predicts the arrival of each bus at each stop from the public position feed and the timetable (Figure 1). Its main elements are:

- **Timetable time.** Each vehicle is matched to its scheduled trip, and its GPS position is projected onto the route and converted into the scheduled time at that point, τ. The arrival at a later stop with published time a_s is then built from the timetable difference $a_s - \tau$, which remains informative even when the absolute intermediate times are not.
- **Learned tables.** The actual travel time of each segment between consecutive stops is learned from passages captured on previous days, by time band, as a median shrunk towards the all-day median of the segment; segments without data use a ratio by segment length. Two further tables hold pace factors by line and band, used where a segment has no learned time, and a per-trip correction.
- **Three layers.** The prediction uses, in order, the projected GPS position, the last stop reported by the feed, and the timetable corrected trip by trip. Each layer abstains under explicit rules, and the state of the prediction records which one was used (`live` for the first; `estimated` or `scheduled` for the two fallback layers).
- **Daily recalibration.** In production the tables are recalibrated every night, and a change is adopted only if it reduces the error on held-out data. In this paper they are frozen (Section 5).

The report (Garrido 2026a) gives every constant and the measurements behind it. The production engine is not public; an open reference implementation in Python agrees with it on all 407 test vectors produced by that engine, and it is the implementation used here. Against the published timetable, over 9.1 million predictions in 23 days, the method was within ± 2 minutes in 92–96% of cases with the bus less than three minutes away and in 67% at 10 to 20 minutes, against 23–25% for the timetable (Garrido 2026b).

4 Data

4.1 What was captured

A program running on GitHub Actions captured, during service hours:

- **The live feed**, every 20 seconds: the public SIRI Vehicle Monitoring file of Bizkaibus, with the GPS position, the trip and the next stop of every vehicle in service. The file changes every 60 seconds.
- **The official waiting times**, about once a minute for each of 60 stops: the full response of the official web service, queried stop after stop (about one request per second in total), with the

Table 1. Days of the capture (Madrid time). Only queries sent between 06:00 and 23:00 are analysed.

Date	Day	Coverage
8 Nov 2026	Sunday	From 06:46
9 Nov 2026	Monday	From 12:42 (morning lost: the capture overran its time limit); feed gap 18:17–18:38
10 Nov 2026	Tuesday	Complete; feed gap of 203 s; an early block before 06:00 is not used
11 Nov 2026	Wednesday	Complete, *analysed separately*: general strike, minimum services (about 30%)
12 Nov 2026	Thursday	Feed complete; official times until 17:12 (the service stopped answering over HTTP)
13 Nov 2026	Friday	Complete (official times over HTTPS)
14 Nov 2026	Saturday	From 09:41 (the scheduled start of the capture did not run)

time t at which each request was sent and the response latency (median 0.97 s).

The timetable is the Bizkaibus GTFS published on 25 September 2026 and valid until 24 November 2026, which covers the week; the database built from it is published with the data.

4.2 Stops

The 60 stops, 28 of them *priority* stops, were chosen before the capture. The 32 other stops are intermediate stops with many buses and lines. Each is reached at least 15 minutes after the start of the trip (median over its scheduled trips), and its name in the official service matches the timetable. The 28 priority stops are stops that passengers asked about (in San Juan, Durango and Elorrio) and the Bilbao stops where, in earlier captures, the actual passage deviated most from the published time. Because this second criterion could favour a method that corrects the timetable, results are also given without the priority stops. Each stop and the reason for its choice are listed in the repository.

4.3 Days

The week runs from Sunday 8 to Saturday 14 November 2026 (Table 1). Only queries sent between 06:00 and 23:00 are used. Wednesday 11 was a general strike with minimum service; it is analysed separately. Monday 9 lacks its morning (the capture overran the time limit of its job), and Thursday 12 lacks official times after 17:12: the service stopped answering over HTTP, and the capture has used HTTPS since, at the same request rate and with only occasional timeouts. Saturday 14 starts at 09:41, because the scheduled start of its capture did not run. The feed has two gaps longer than two minutes, of 21 minutes on Monday 9 (18:17–18:38) and of 203 s on Tuesday 10; no passage is reconstructed across them (Section 5.2).

5 Comparison design

5.1 The three predictions

For a query sent at time t to stop s, and the bus matched to it (Section 5.3):

- **Timetable:** $\hat{a}_{\mathrm{tt}}$, the published time of the bus's trip at s.
- **Official:** $\hat{a}_{\mathrm{off}} = t + 60\,m$, where m is the announced waiting time in minutes. Whether the service rounds or truncates m is unknown; truncation is tested in the sensitivity analysis.
- **Bide+:** the prediction of the reference implementation in the last copy of the feed downloaded at or before t, at most 60 seconds earlier (median 10 s). The positions in a copy are themselves older than its download: the newest position in each copy was a median of 77 s older than the download (10th and 90th percentiles 45 and 112 s), so Bide+ works with positions about one and a half minutes old. It uses the learned tables of Saturday 7

November, calibrated only with data from before the week and not updated during it. It is rounded to whole minutes from t as the app shows it, $\hat{a}_{\mathrm{B}+} = t + 60\max\big(0, \mathrm{round}((\hat{a}-t)/60)\big)$, with halves rounded away from zero.

5.2 The actual passage

From the feed. The actual passage a of a bus at a stop is the midpoint between the last copy of the feed in which the next stop of the bus is s or an earlier one and the first copy in which it is a later one, without the bus changing trip. The bus may be missing from the feed for up to five minutes in between, but the two bracketing copies of the feed must be at most 90 s apart, so no passage is interpolated across a gap. Trips are identified by their full feed reference (line, trip number and departure time), because Bizkaibus reuses trip numbers within a day. Since the feed changes every 60 seconds, a is known to within about ± 30 s.

From GPS. This definition uses the same feed field that Bide+ uses. An independent passage a_{GPS} is therefore computed from the positions alone: the moment of closest approach of the bus to the stop, interpolated between the two positions that bracket the stop, if one of them is within 500 m. It is available for 99.7% of the eligible pairs. The two agree closely: $a_{\mathrm{GPS}} - a$ has a median of -4 s, its 10th and 90th percentiles are -36 and $+34$ s, and 89% of the differences are within one minute. The main analysis uses a; a_{GPS} is used as an alternative ground truth.

Failed reconstructions. A pair is discarded if a and a_{GPS} differ by more than five minutes, or if reaching the stop by a would require travelling at more than 30 m/s from the smaller of two distances at t: the straight-line distance of the bus in the feed (a lower bound on the road distance) and the distance reported by the official service, allowing 60 s for the feed's resolution. This removes 1.9% of the eligible pairs. One stop, Atxuri (3394), where all three sources were off by the same 17 to 30 minutes, which points to a failure of the reconstruction rather than of the predictors, is excluded entirely. This exclusion was decided after inspecting the errors; the sensitivity analysis that keeps it changes no conclusion (Table A.9).

5.3 Matching announcements to buses

The official service does not identify the vehicle it announces. For a query at t to stop s, its first announcement for line ℓ (called e_1) is matched to the first bus of line ℓ, among those tracked in the feed, that actually passes s after t, and its second announcement (e_2) to the second; passages up to 60 minutes after t are considered. The match is accepted only if the distance that the official service gives for the announced bus is consistent with the straight-line distance of the matched bus from the stop in the feed: within 500 m, or between 0.6 and 2 times it. The official distance appears to be measured along the road: for buses more than 1 km away, its ratio to the straight-line distance has a median of 1.24 (10th and 90th percentiles 0.78 and 1.87); ratios below one occur because the feed position is older than the official answer. The check rejects a median of 11% of the pairs per stop (at most 40%). A stricter version, within 300 m or between 0.9 and 1.6 times, is reported as a sensitivity analysis. Lines listed twice in the same response would be left out; there were none. No stop is visited twice by the same trip, but 12 of the 60 stops are the last stop of some routes; a bus ending its trip there has no reconstructed passage, so an announcement for it would be matched to the next bus. Such a mismatch makes the official error negative, the opposite of the bias observed (Section 6.6), and most are rejected by the distance check.

When the matched bus is not in the feed at t, Bide+ can only give a fallback prediction and the bus's distance is unknown, so the identity check cannot be applied. These pairs are excluded from the main analysis and included in a sensitivity analysis.

5.4 Metrics and inference

The error of each prediction is $\hat{a} - a$: positive when the bus arrives *earlier* than predicted. The main scoring rule is the share of predictions within ± 2 minutes, chosen in advance as in the companion paper. It is a threshold rule of the kind that suits a decision with a tolerance (Gneiting 2011). The median and mean absolute errors, the shares within ± 1 and ± 3 minutes, the median signed error and the share of buses that arrive more than two minutes early are also reported. The main analysis uses pairs with $0 \leq a - t < 30$ min, by actual time to arrival; results by the announced time m use all announcements of up to 29 minutes whose bus passed within the next hour. These rules (the 90-second gap rule, the identity check, the failed-reconstruction rule and the Atxuri exclusion) were set after the data had been collected and inspected; each is varied in the sensitivity analysis.

Predictions of the same trip, and of the same stop, are not independent: one passage contributes a prediction every minute. Confidence intervals for the difference in the share within ± 2 minutes are percentile intervals from 2,000 bootstrap resamples of whole trips and, separately, of whole stops (Efron and Tibshirani 1993). The mean loss differential $d = |e_{\text{off}}| - |e_{\text{B+}}|$ is tested with standard errors clustered by stop and by stop and day, in the manner of Diebold and Mariano (1995). With six days, day-level variance cannot be estimated reliably.

Passenger model. A passenger is assumed to reach the stop k minutes before the predicted arrival ($k = 1, 2, 3$). If the bus has already passed, the passenger waits for the next bus of the same line. The model reports the share of missed buses and the expected wait including that cost, for announcements of 3 to 29 minutes.

6 Results

6.1 Pairs and coverage

Coverage is measured on the official announcements of the first bus of each line with an announced wait under 30 minutes, on the six normal days: 838,087. Of these, 606,565 (72.4%) could be paired with a bus whose actual passage was observed, and 450,630 (53.8%) entered the main analysis; the rest were lost because the bus was not in the feed at t, failed the identity check, or was at the excluded stop or had a failed reconstruction. The share entering the main analysis is lowest for announcements of 0–2 minutes (42%) and of 20–29 minutes (48%) and highest at 3–9 minutes (61–63%; Table A.2 in Appendix A). The main analysis, which also includes announcements of the second bus and of 30 minutes or more whose bus passed within 30 minutes, has 555,744 pairs: 29,296 passages of 10,472 trips, made by 320 vehicles on 73 lines, at 59 stops, with between 176 and 290 different vehicles in each hour from 06:00 to 22:00. For the announcements that could be paired but did not enter the main analysis, the official real time was as accurate as in the main analysis (41% within two minutes when the bus was not in the feed, 40% when the identity was not confirmed, against 39%), and so was Bide+ (63% and 67%, against 64%; Table A.3). Announcements that could not be paired at all cannot be evaluated. Table A.1 in Appendix A follows the pairs by day.

6.2 Accuracy by time to arrival

The published timetable is within two minutes of the actual passage in 14–20% of cases at every horizon, with a median absolute error of six to eight minutes (Table 2, Figure 2). With the bus less than three minutes away, the official real time and Bide+ are equally often within two minutes (90%; difference -0.5 to $+0.5$ points, 95% interval by trip). From three minutes on, Bide+ is more often within two minutes: 80% against 59% at 3–6 minutes, 74% against 42% at 6–10, 62% against 30% at 10–20 and 47% against 23% at 20–30 minutes. Over all horizons the shares are 16% for the timetable, 39% for the official real time and 64% for Bide+, and the ordering is the same within one and within three minutes (Table A.5 in Appendix A).

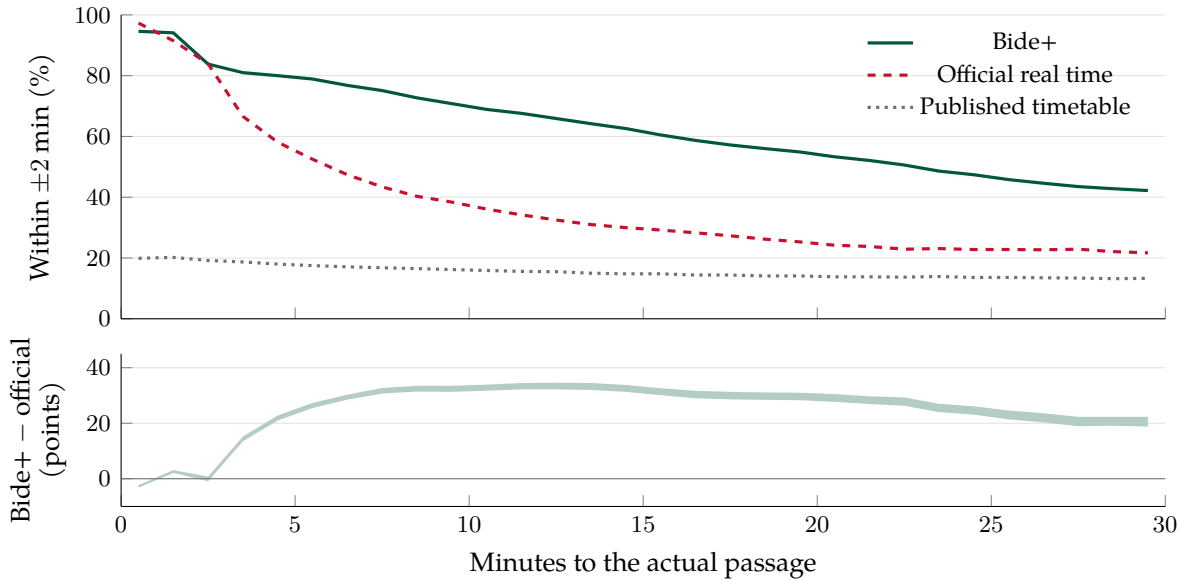


Figure 2. Top: share of predictions within two minutes of the actual passage, by minutes to that passage (one-minute bins). Bottom: 95% bootstrap interval (2,000 resamples of trips per bin) of the difference, Bide+ minus official, in percentage points. Main analysis, 555,744 pairs.

Table 2. Errors of the three sources by time to the actual passage. *Med.*: median absolute error (s). ±2: share within two minutes (%). *Early*: share of buses arriving more than two minutes before the predicted time (%; error $\hat{a} - a > 120$ s).

		Timetable			Official real time			Bide+		
Minutes	Pairs	Med.	±2	Early	Med.	±2	Early	Med.	±2	Early
0–3	49,826	370	20	55	39	90	9	40	90	6
3–6	63,557	406	18	59	98	59	25	56	80	8
6–10	88,110	434	17	61	144	42	37	66	74	10
10–20	198,662	464	15	64	214	30	53	90	62	13
20–30	155,589	500	14	67	283	23	58	128	47	13
All	555,744	452	16	63	170	39	45	82	64	11

6.3 Inference

From three minutes on, the difference in the share within two minutes is positive with either clustering; the lowest lower bound, by stop at 3–6 minutes, is 14 points (Table 3). Resampling whole lines (73) gives similar intervals, from +16 to +36 points at 3 to 20 minutes and $[-3.6, +3.4]$ below three minutes. The mean absolute error gives a more nuanced picture. Below six minutes the official real time has the smaller mean absolute error, because Bide+ has more very large errors (Section 6.6); between 6 and 10 minutes the two do not differ; from 10 minutes on Bide+ has the smaller mean absolute error, by 84 to 147 s. Bide+ is therefore more often close to the actual passage from three minutes on, while its mean error is smaller only from ten.

6.4 Accuracy by announced waiting time

A passenger sees the announced time, not the actual one. Table 4 bins all announcements of up to 29 minutes by the minutes announced. The ranking is the same: when the official service announced 6 to 9 minutes, its announcement was within two minutes of the actual passage in 54% of cases and Bide+ in 74%; at 10 to 19 minutes, 36% and 64%; at 20 to 29 minutes, 23% and 51%. Over all these announcements the shares are 17%, 42% and 65%.

Table 3. Inference. Columns 2–3: 95% bootstrap intervals (2,000 resamples) of the difference Bide+ minus official in the share within two minutes (percentage points), resampling trips or stops. Columns 4–6: mean loss differential $d = |e_{\text{off}}| - |e_{\text{B+}}|$ (s; positive favours Bide+) and $z = d/\text{SE}$ with standard errors clustered by stop or by stop and day.

Minutes	By trip	By stop	d (s)	z (stop)	z (stop-day)
0–3	$[-0.5, +0.5]$	$[-4.4, +4.8]$	-107	-7.0	-9.0
3–6	$[+20.3, +21.7]$	$[+14.2, +27.5]$	-49	-3.5	-4.8
6–10	$[+30.6, +32.4]$	$[+27.1, +35.7]$	-3	-0.3	-0.3
10–20	$[+30.7, +32.7]$	$[+28.9, +34.5]$	84	6.5	8.1
20–30	$[+23.1, +25.9]$	$[+21.5, +27.3]$	147	8.2	12.5
All	$[+24.7, +26.4]$	$[+22.7, +28.5]$	55	4.9	6.1

Table 4. Errors of the three sources by the waiting time announced by the official service (minutes, $m \leq 29$; buses passing within the next hour). Columns as in Table 2. At 0–2 minutes the official *Early* share is zero by construction, since $\hat{a}_{\text{off}} \leq t + 120$ s $\leq a + 120$ s.

		Timetable			Official real time			Bide+		
Announced	Pairs	Med.	±2	Early	Med.	±2	Early	Med.	±2	Early
0–2	53,426	357	18	45	48	79	0	46	82	4
3–5	54,591	356	20	47	85	65	5	60	76	5
6–9	68,773	378	20	53	108	54	20	65	74	7
10–19	170,264	407	17	60	170	36	41	85	64	9
20–29	159,924	462	15	65	274	23	55	116	51	12
All	506,978	408	17	58	151	42	35	81	65	9

6.5 Robustness to the ground truth

With the GPS-based passage as ground truth (546,485 pairs, re-binned by time to that passage), the shares within two minutes are 16%, 40% and 67%, and below three minutes Bide+ is slightly ahead (93% against 92%; Table A.8 in Appendix A). Shifting the feed-based passage (Figure A.1 in Appendix A) shows that Bide+ fits best when the passage is taken 30 s earlier and the official real time when it is taken as is or 30 s later. Bide+ stays ahead unless the true passage were more than two minutes later than reconstructed, an offset that the agreement with the GPS-based passage (90th percentile $+34$ s) makes unlikely. The ranking is also unchanged in the other 11 sensitivity analyses (Table A.9 in Appendix A): official minutes truncated rather than rounded, Bide+ in seconds or truncated, a stricter identity check or none, with the fallback pairs, without Zabalburu, with the excluded stop and failed passages, one random pair per passage, without the priority stops, and with live positions only.

6.6 Direction and size of the errors

The timetable errs in one direction: buses run ahead of it (median signed error $+358$ s at 10–20 minutes), and 64% of them pass more than two minutes before the published time. The official real time keeps that direction with less than half the size ($+140$ s): in more than half of the cases at 10 to 30 minutes, the bus passes more than two minutes before the announced time. Bide+ is nearly centred (-34 s at 10–20 minutes), and the bus is more than two minutes early in 13% of cases at those horizons (Figure 3; Table A.6 in Appendix A).

Bide+ has, however, more very large errors at short horizons: with the bus less than six minutes away, 1.5–1.6% of its predictions are more than ten minutes off, against 0.1–1.2% for the official real time (Table A.7 in Appendix A). Of these 1,795 errors, 626 occur at one Bilbao stop (Zabalburu, 305), and 71% come from its fallback layers: the bus was in the feed, but the GPS layer abstained and Bide+ used the last reported stop or the corrected timetable. Without that stop the shares within two minutes are 39% and 65% (Table A.9).

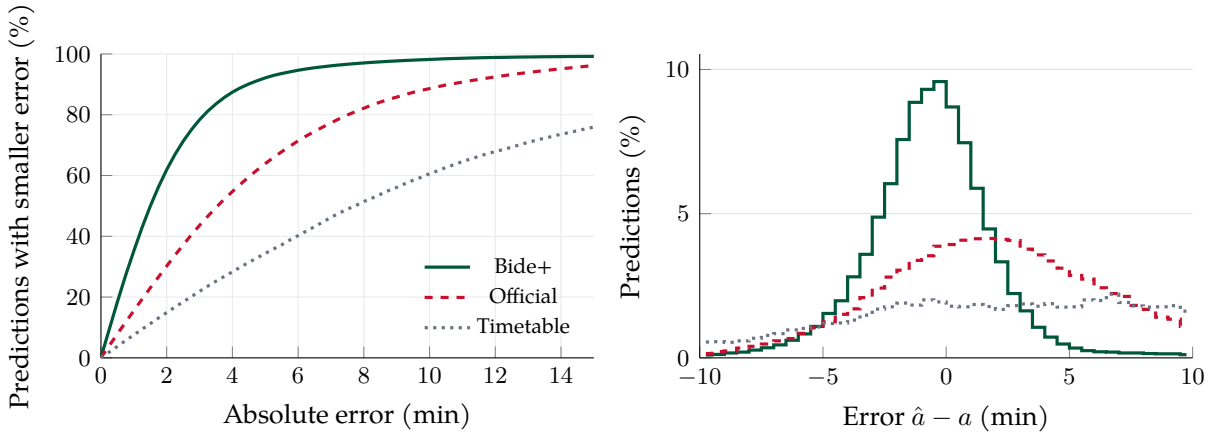


Figure 3. Errors with the bus 10 to 20 minutes away (legend in the left panel). Left: cumulative distribution of the absolute error. Right: distribution of the signed error in 30-second bins; positive means the bus arrives earlier than predicted. Beyond $+10$ minutes lie 35.0% of timetable, 11.0% of official and 1.3% of Bide+ predictions; below -10 minutes, 4.4%, 0.5% and 0.6%.

Table 5. Passenger model, announcements of 3 to 29 minutes: a passenger reaches the stop k minutes before the predicted arrival and, if the bus has gone, waits for the next one of the same line. *Miss*: share of missed buses (%). *Wait*: expected wait including misses (min).

	Timetable		Official		Bide+	
k	Miss	Wait	Miss	Wait	Miss	Wait
1 min	63	21.2	48	21.8	18	9.0
2 min	59	20.7	38	18.7	9	5.6
3 min	55	20.2	29	16.1	5	4.7

6.7 Consequences for a passenger

A passenger who arrives two minutes before the predicted time misses the bus in 59% of cases when following the timetable, 38% with the official real time and 9% with Bide+. Counting the wait for the next bus when the first is missed, the expected wait is 20.7, 18.7 and 5.6 minutes (Table 5). The ranking of missed buses holds for every margin; Bide+ gives the shortest expected wait for every margin, but with a one-minute margin the official real time gives a slightly longer one than the timetable. The model uses 395,949 of the 453,552 announcements of 3 to 29 minutes; the rest, mostly the last buses of the day, have no next bus of the line within two hours. The model is deliberately simple: it ignores that a passenger may see an updated prediction on the way to the stop.

6.8 San Juan (Amorebieta)

San Juan, in Amorebieta, is one of the stops that passengers asked about, served in both directions by lines 3513 to 3515 (Figure 4). Over 9,804 pairs with the bus 3 to 30 minutes away, the timetable was within two minutes in 18.3% of cases, the official real time in 32.6% and Bide+ in 64.8%, close to the network-wide figures for the same horizons (15.2%, 33.7% and 61.8%). The official real time was much less accurate in one direction (19% at stop 3388) than in the other (46% at 3391), while Bide+ was more similar in both (61% and 69%). The two buses drawn at random illustrate the aggregate pattern: the timetable is offset by a roughly constant amount, the official real time is closer but announces the bus later than it comes, and Bide+ follows the actual waiting time most closely. Four buses drawn at random from the whole network are shown in Figure A.2 in Appendix A.

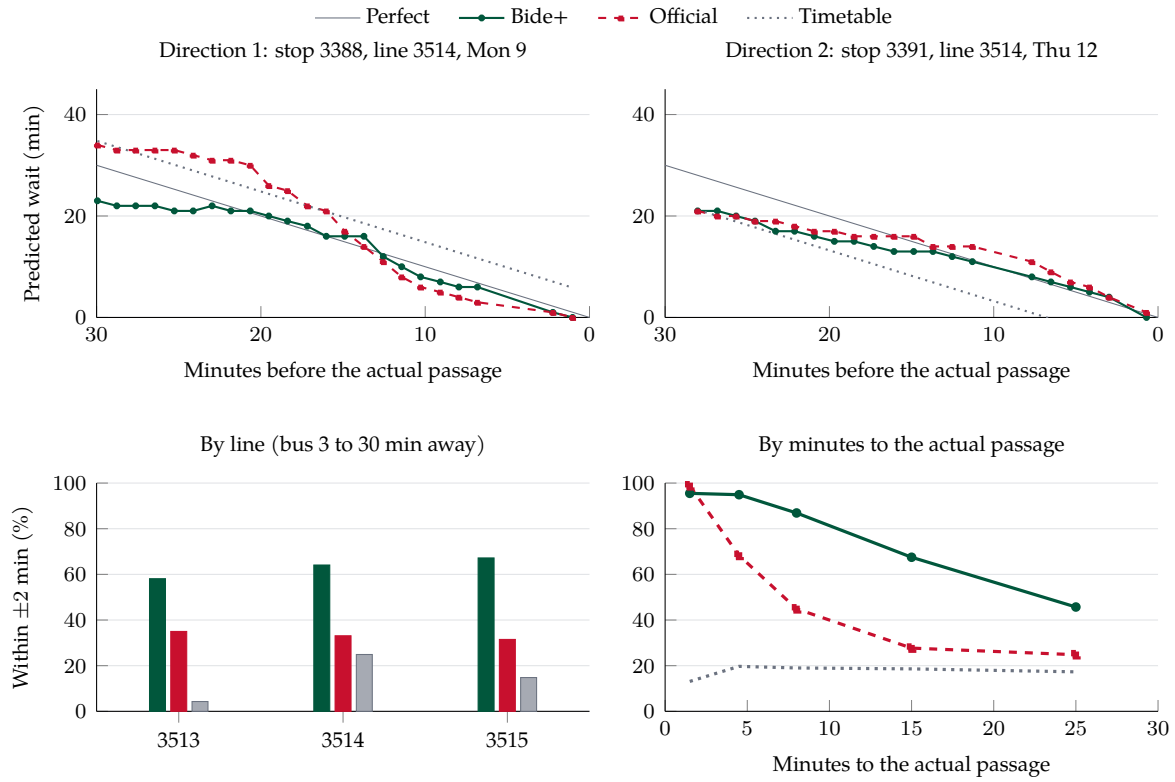


Figure 4. San Juan (Amorebieta), lines 3513 to 3515 in both directions. Top: predicted against actual waiting time for one bus drawn at random in each direction (fixed seed) among the passages with queries from at least 25 minutes before; the thin line is a perfect prediction. Bottom: share within two minutes by line (bars: Bide+, official, timetable) and by minutes to the actual passage (points at the middle of the bins of Table 2). Main analysis.

6.9 Stops, time of day and days

The ranking is not confined to a few stops, hours or days. Bide+ is more often within two minutes than the official real time at 57 of the 59 stops (Figure 5). Of the other two, at Zabalburu (305) the official real time is slightly ahead (35.9% against 34.1%), and at Labakua (2856) the timetable is the most accurate source (74%, against 67% official and 61% Bide+). The ranking timetable, official, Bide+ holds at every hour from 06:00 to 22:00 and on each of the six normal days (Figure A.3 in Appendix A), at the priority stops and at the others, on weekdays, Saturday and Sunday, and for the second announced bus (Table A.10).

6.10 Ablation: what makes Bide+ more accurate

Table 6 recomputes Bide+ on the same pairs without its learned tables. Without any of them, the arrival is built from timetable differences only, and Bide+ is within two minutes in 29% of cases, below the official real time (39%). Removing the learned segment times alone lowers it from 64% to 34% (difference +29.5 to +31.1 points, 95% interval by trip), 4 to 5 points below the official real time (interval −5.3 to −4.2 by trip, −6.3 to −3.4 by stop). Removing the per-trip correction costs one point. Removing the pace factors changes nothing, because they act only where a segment has no learned time. On these pairs and by this metric, the learned segment travel times are what puts Bide+ ahead of the operator; locating the bus in timetable time is not enough on its own.

6.11 The strike day

On Wednesday 11, with minimum service, the ranking was the same as on normal days (Table A.11 in Appendix A). Bide+ was more often early than on normal days (23% against 8–18%), which is

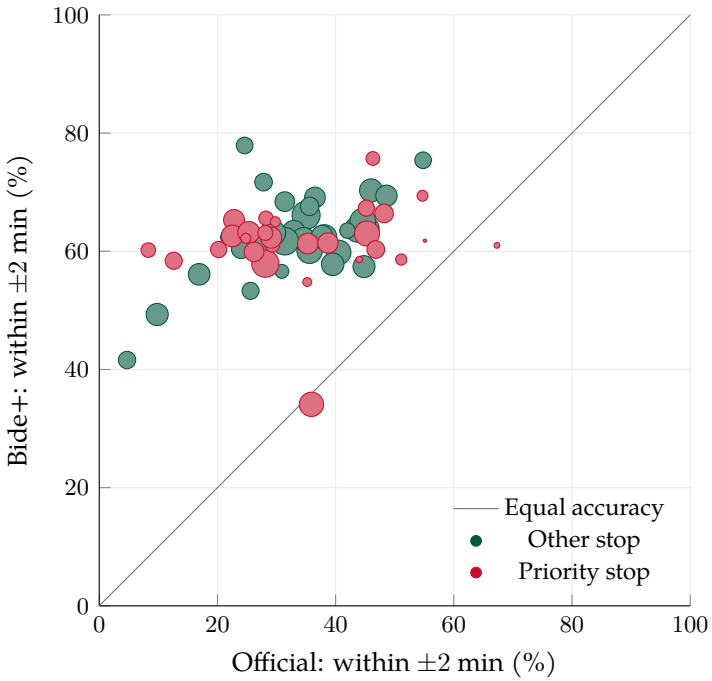


Figure 5. Each point is one stop (bus 3 to 30 minutes away; marker radius proportional to the square root of the number of pairs, on the same scale for both groups). Bide+ is more often within two minutes at 57 of the 59 stops; at 56 the unadjusted 95% interval of the difference (2,000 resamples of trips) excludes zero, and 56 remain significant after a Holm correction for 59 tests.

Table 6. Ablation: share within two minutes (%) on the same pairs, with Bide+ with all its learned tables, without any of them, and without each one in turn.

			Bide+ without				
Minutes	Pairs	Official	all tables	segment times	pace factors	trip correction	Bide+
0–3	49,826	90	77	79	90	90	90
3–6	63,557	59	41	44	80	80	80
6–10	88,110	42	32	36	74	73	74
10–20	198,662	30	22	28	62	61	62
20–30	155,589	23	16	22	47	46	47
All	555,744	39	29	34	64	63	64

consistent with buses running faster than in the normal traffic its tables were learned from; this was not tested directly.

7 Discussion

7.1 What the comparison shows

On the buses compared, the published timetable is a poor guide to when the bus will come, the operator's real time removes part of its error, and Bide+, built only on the operator's open data, is within two minutes of the actual passage more often than the operator's real time from three minutes before arrival onwards. The gap is largest, about 32 percentage points, with the bus 6 to 20 minutes away, the range in which a passenger decides when to leave for the stop. Below three minutes the two are equally often within two minutes, and below six minutes Bide+ has a slightly larger share of very large errors than the operator (1.5–1.6% against 0.1–1.2%), about a third of them at one stop.

7.2 Possible explanations

The ablation identifies the learned segment travel times as the component that puts Bide+ ahead; without them it falls below the official real time. This is consistent with the timetable on this network running behind the buses, and with the official real time keeping part of that bias (Figure 3). How the official service computes its waiting times is not published, so this paper cannot say why its errors take the form they do.

7.3 Implications

Operators that publish vehicle positions make possible independent predictions that, on this network and week, and for the announcements that could be matched and confirmed (54% of first-bus announcements under 30 minutes), were more often within two minutes than their own. The most direct implication is for evaluation: operators that also publish their predictions, for example as GTFS-Realtime trip updates (MobilityData 2026a) or SIRI estimated timetables, would allow their accuracy to be audited independently and continuously, as done here by recording a web service. The timetable itself could also be improved: with most buses passing more than two minutes before the published time at these stops, revising the intermediate times, or holding buses at timing points, would help both printed timetables and any real-time system anchored to them. Journey planners that show the operator's figures unchanged inherit their accuracy; none was measured here.

7.4 Limitations

- **Coverage.** 54% of the first-bus announcements under 30 minutes entered the main analysis, and 72% could be paired at all. Pairs whose bus was not in the feed at t (15% of the paired ones) are excluded from the main analysis; including them, without the identity check, gives 39% and 64% (Table A.9), and the official real time was as accurate on the excluded pairs as on the main ones. The 28% that could not be paired, for example buses that the official service announces but the feed does not track, cannot be evaluated, and the direction of any bias from them is unknown.
- **Ground truth.** The main ground truth comes from the feed field that Bide+ also uses, and its learned segment times were fitted on passages defined in the same way, so Bide+ is calibrated to this definition and is more sensitive than the official real time to shifting it (Figure A.1). The GPS-based passage, which does not use that field, agrees with it to within a minute in 89% of cases and gives the same ranking, although it shares the feed's vehicle-to-trip assignment and clock. Field observation was not possible.
- **Matching.** The official service does not identify the vehicle it announces. The distance check compares a road distance with a straight-line one; stricter or no checks do not change the ranking, Bide+ is ahead in every decile of the distance ratio (Table A.4), and across stops the gap is unrelated to the share of matches rejected (correlation -0.05). Some mismatches may remain, and they would count against the official source.
- **Selection of stops.** The stops were chosen by the author; 28 of them partly because the timetable deviates most there. Results without them are similar (39% and 65%, against 39% and 64%), but the stops are not a random sample of the network.
- **One week, one network.** Six normal days in November, one Saturday and one Sunday among them. Day-level variance cannot be estimated, and other seasons, networks or operators may differ.
- **Replay, not the app; source of the official times.** Bide+ was recomputed with its reference implementation and frozen tables; the app was not measured. That the web service returns what the official app and the displays show was not verified independently.

- **Analytical choices.** The scoring rule, horizons and matching follow the companion paper, and the stops were fixed before the capture, but the study was not preregistered, and the rules for reconstructing passages and confirming matches were set after the data had been inspected. A confirmatory replication on another week, ideally by a third party, would strengthen the conclusions.

8 Conclusions

For one week, at 59 stops of the Bizkaibus network, the official real-time waiting times were recorded minute by minute and compared, bus by bus, with the published timetable and with the arrivals predicted by the Bide+ method from the operator's open position feed and timetable. Over all horizons up to 30 minutes, the timetable was within two minutes of the actual passage in 16% of the paired cases, the official real time in 39% and Bide+ in 64%, with the same ranking against an independent GPS-based ground truth. Below three minutes the two real-time sources were equally often within two minutes; from three minutes on Bide+ was more often within two minutes at almost every stop, hour and day, although its mean error was smaller only from ten minutes. An ablation attributes its advantage to its learned segment travel times. These results apply to the 54% of first-bus announcements whose bus was tracked, observed and confirmed, and they call for a replication on other periods and networks.

Declaration of competing interest

The author designed the Bide+ method and develops the independent Bide+ passenger information app that uses it, and therefore has an interest in its results. The app is free to use; it accepts voluntary donations, which are used to pay for the server that runs it, and the author receives no other income from it. To limit the effect of this interest on the evaluation, the method was frozen before the test week, its predictions were recomputed with the public reference implementation, the analysis rules are stated in Section 5, and all data and code are published so that the comparison can be repeated independently. The author has no relationship with Bizkaibus, the Diputación Foral de Bizkaia or the operators of the service.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

CRediT authorship contribution statement

Yeray Garrido: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing.

Data availability

The capture program, the list of stops with the reason for each, the days and stops used and excluded, the comparison, the tables and the figure data are in the folder `articulo-comparacion` of the repository `https://github.com/Garridoparrayeray/bide-eta` (code under the MIT licence, texts and derived data under CC BY 4.0; see below for the raw official responses). The timetable database used is in the repository (`datos/horario`). An archived copy of the code, the derived data and this preprint is deposited on Zenodo (doi:10.5281/zenodo.24286027). The raw captures (feed copies, official responses and the timetable of each day) are not part of that record: they are assets of the release `comparacion-2026-11` of the GitHub repository, under the terms given below. The position feed and the timetable are open data published by the Diputación Foral de

Bizkaia and listed in the Spanish National Access Point (Ministerio de Transportes y Movilidad Sostenible 2026).

Data collection and terms of use

The official waiting times were obtained from the public web service used by the official Bizkaibus app, which requires no authentication. The capture queried it sequentially, about one request per second in total, with a user agent that identifies the project, and only during service hours. No permission was sought from the operator. The capture first used the plain-HTTP address of the service; on 12 November at 17:12 that address stopped answering (requests timed out), and from 13 November the capture used the HTTPS address of the same service. Over the week, 0.5% of the requests failed (most of them on 12 November after 17:12), all by timeout or a closed connection; no error status, rate-limit or blocking response was received. The host serves no `robots.txt`, and no terms of use for the web service or for redistributing its responses were found. The data contain no personal information: only stop, line and vehicle identifiers. The raw official responses are not covered by the CC BY licence of the derived data: they are redistributed only for research, with attribution, and will be removed if their publisher asks.

Appendix A. Supporting tables and figures

Table A.1. From official announcements to the pairs analysed, by day (first and second bus, queries 06:00–23:00). *Paired*: announcements matched to a bus whose passage was observed within 60 minutes. *Beyond 30* and *Within 30*: that passage more or less than 30 minutes after the query. *Fallback*: the bus was not in the feed at t (Bide+ gives its fallback prediction). *Excl.*: excluded stop or failed reconstruction. *Not conf.*: distance check failed. *Main*: pairs in the main analysis. * General strike, not part of the main results.

Day	Queries	Paired	Beyond 30	Within 30	Fallback	Excl.	Not conf.	Main
Sun 8	58,320	181,691	84,604	97,087	10,331	571	10,018	76,167
Mon 9	34,524	198,395	83,246	115,149	18,340	2,164	12,593	82,052
Tue 10	54,814	329,836	139,860	189,976	30,341	3,419	21,114	135,102
Wed 11*	58,481	150,024	67,533	82,491	13,500	1,944	9,353	57,694
Thu 12	35,547	224,149	98,735	125,414	20,134	2,783	13,377	89,120
Fri 13	46,227	267,145	114,219	152,926	23,010	4,152	16,593	109,171
Sat 14	40,713	164,292	74,356	89,936	13,613	2,398	9,793	64,132

Table A.2. Coverage of the official announcements of the first bus with an announced wait under 30 minutes (six normal days, 06:00–23:00), by minutes announced: share paired with a bus whose passage was observed within 30 minutes, and share in the main analysis (%).

Announced	Announcements	Paired (<30 min)	Main
0–2	124,851	71	42
3–5	88,414	78	61
6–9	106,468	77	63
10–19	270,452	76	58
20–29	247,902	66	48
All	838,087	72	54

Table A.3. Share within two minutes (%) in the main analysis and in the paired announcements excluded from it (all horizons up to 30 minutes, six normal days, excluded stop and failed reconstructions removed). For pairs whose bus was not in the feed, Bide+ gives its fallback prediction.

Pairs	Number	Timetable	Official	Bide+
Main analysis	555,744	16	39	64
Bus not in the feed at t	110,829	34	41	63
Identity not confirmed	83,488	16	40	67

Table A.4. Share within two minutes (%) by decile of the ratio between the official distance and the straight-line distance of the matched bus (main analysis, bus more than 500 m away). Bide+ is ahead in every decile, so the result does not depend on the pairs whose identity is least certain.

Decile	Ratio	Pairs	Official	Bide+
1	0.01–0.80	52,706	69	79
2	0.80–0.94	52,706	53	73
3	0.94–1.03	52,706	44	69
4	1.03–1.10	52,706	35	65
5	1.10–1.16	52,706	29	62
6	1.16–1.22	52,706	29	60
7	1.22–1.28	52,706	26	56
8	1.28–1.36	52,706	26	54
9	1.36–1.52	52,706	26	57
10	1.52–2.00	52,707	24	56

Table A.5. Further metrics, all horizons up to 30 minutes (main analysis).

Metric	Timetable	Official	Bide+
Within ±1 min (%)	8	22	39
Within ±2 min (%)	16	39	64
Within ±3 min (%)	23	52	79
Median absolute error (s)	452	170	82
Mean absolute error (s)	689	257	202
More than 10 min off (%)	38.7	10.3	2.0

Table A.6. Share within one minute (%) and median signed error (s) by time to the actual passage. A positive signed error means that the bus arrives earlier than predicted.

	Timetable		Official		Bide+	
Minutes	±1	Signed	±1	Signed	±1	Signed
0–3	10	+198	68	−9	68	−10
3–6	9	+266	32	+2	53	−10
6–10	8	+314	22	+48	46	−15
10–20	8	+358	15	+140	35	−34
20–30	7	+399	12	+196	25	−76
All	8	+339	22	+81	39	−30

Table A.7. Mean absolute error (s) and share of predictions more than ten minutes off (%), by time to the actual passage.

	Timetable		Official		Bide+	
Minutes	MAE	>10 min	MAE	>10 min	MAE	>10 min
0–3	630	32.2	58	0.1	165	1.6
3–6	657	35.1	128	1.2	177	1.5
6–10	683	37.1	193	4.3	196	1.7
10–20	700	39.4	289	11.4	205	1.8
20–30	711	42.1	370	19.4	224	2.7

Table A.8. With the GPS-based passage as ground truth: share within two minutes (%) by time to that passage, and 95% bootstrap interval (2,000 resamples of trips) of the difference Bide+ minus official (percentage points).

Minutes	Pairs	Timetable	Official	Bide+	Difference
0–3	46,052	20	92	93	[+0.5, +1.4]
3–6	63,206	18	62	84	[+21.2, +22.7]
6–10	88,168	17	44	77	[+31.3, +33.1]
10–20	197,393	15	31	64	[+32.0, +34.2]
20–30	151,666	14	23	49	[+24.2, +27.0]
All	546,485	16	40	67	–

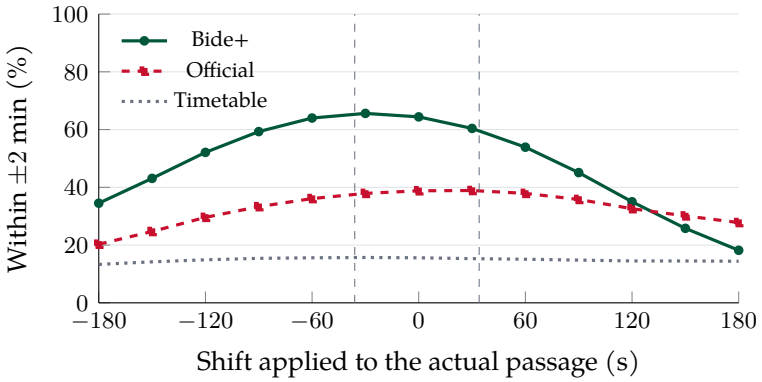


Figure A.1. Share within two minutes, all horizons, when the feed-based actual passage is shifted by the amount shown (positive: later). The dashed lines mark the 10th and 90th percentiles of the difference between the GPS-based and the feed-based passage.

Table A.9. Sensitivity analyses: share within two minutes (%) over all horizons and with the bus 10 to 20 minutes away.

		All horizons			10–20 min	
Variant	Pairs	Timet.	Official	Bide+	Official	Bide+
Main analysis	555,744	16	39	64	30	62
Actual passage −60 s	555,744	16	36	64	27	63
Actual passage +60 s	555,744	15	38	54	32	51
Actual passage from GPS	546,485	16	40	67	31	64
Official +30 s (minutes truncated)	555,744	16	38	64	29	62
Bide+ in seconds	555,744	16	39	65	30	62
Bide+ minutes truncated	555,744	16	39	61	30	58
Stricter identity check	444,787	16	36	63	31	62
No identity check	639,232	16	39	65	29	62
No identity check, with fallback	750,061	18	39	64	31	62
Without Zabalburu (305)	540,457	15	39	65	30	62
With stop 3394 and failed passages	566,016	16	38	63	30	61
One pair per passage (random)	29,296	18	43	66	31	61
Without priority stops	357,824	16	39	65	31	62
Bide+ with live position only	510,488	15	39	66	31	63

Table A.10. Errors by group, all horizons up to 30 minutes. Columns as in Table 2.

		Timetable			Official real time			Bide+		
Group	Pairs	Med.	±2	Early	Med.	±2	Early	Med.	±2	Early
Weekdays	415,445	426	16	60	164	40	42	84	63	10
Saturday	64,132	499	15	68	168	39	45	79	66	10
Sunday	76,167	596	14	73	216	33	56	75	69	18
Priority stops	197,920	525	15	66	178	38	46	86	63	12
Other stops	357,824	420	16	62	166	39	44	80	65	11
First bus (e_1)	512,218	452	16	62	166	40	43	80	65	11
Second bus (e_2)	43,526	452	13	73	218	29	58	114	52	14

Table A.11. Errors by day, all horizons up to 30 minutes. Columns as in Table 2. * General strike, shown for reference and not part of the main results.

		Timetable			Official real time			Bide+		
Day	Pairs	Med.	±2	Early	Med.	±2	Early	Med.	±2	Early
Sun 8	76,167	596	14	73	216	33	56	75	69	18
Mon 9	82,052	480	15	64	169	39	45	76	68	13
Tue 10	135,102	418	17	59	160	41	41	86	62	9
Wed 11*	57,694	539	16	71	194	36	52	82	65	23
Thu 12	89,120	404	17	59	156	41	39	88	62	8
Fri 13	109,171	418	15	60	170	38	44	86	63	10
Sat 14	64,132	499	15	68	168	39	45	79	66	10

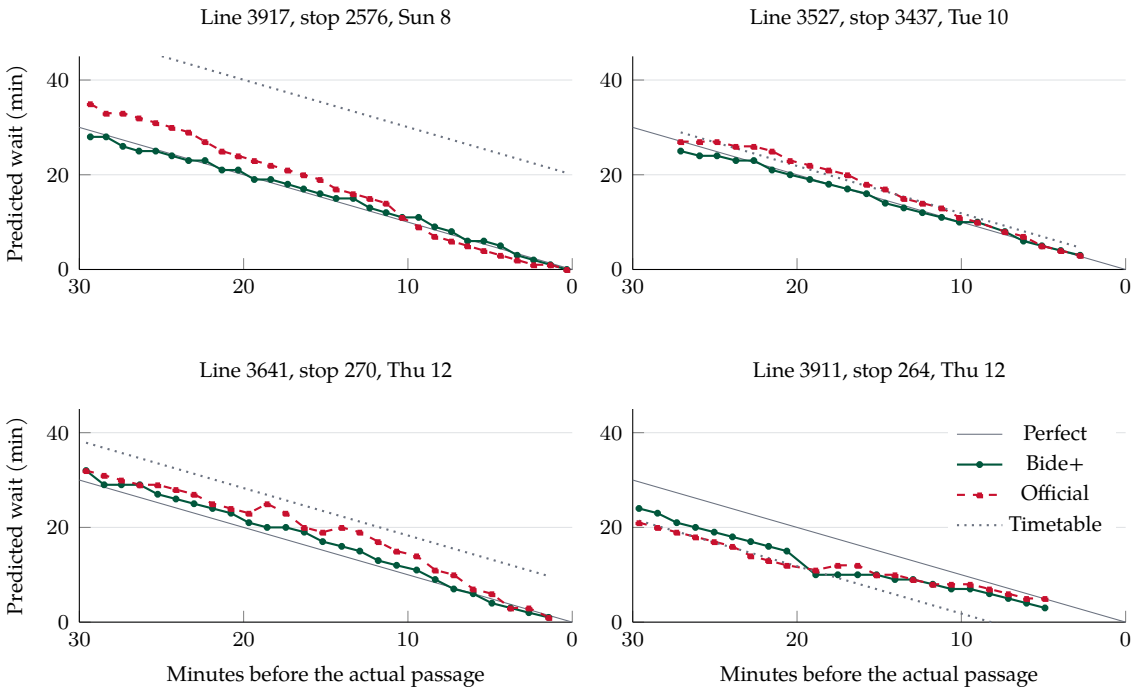


Figure A.2. Predicted against actual waiting time for four buses drawn at random from the whole network (fixed seed), among the passages with queries from at least 25 minutes before. The thin line is a perfect prediction.

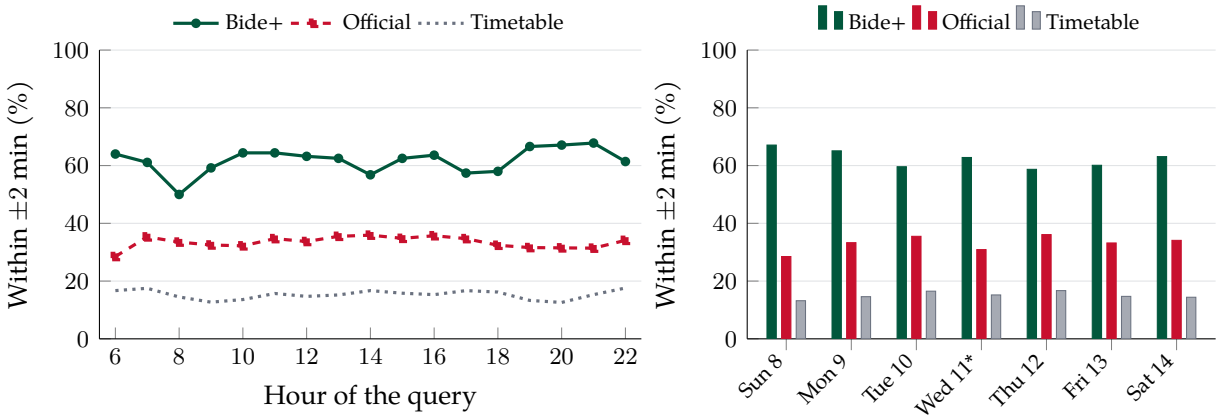


Figure A.3. Share within two minutes with the bus 3 to 30 minutes away. Left: by hour of the query (six normal days, 13,000 to 42,000 pairs per hour). Right: by day (* general strike, not part of the main results).

References

Brakewood, C., G. S. Macfarlane, and K. Watkins (2015). "The Impact of Real-Time Information on Bus Ridership in New York City". In: *Transportation Research Part C: Emerging Technologies* 53, pp. 59–75. DOI: 10.1016/j.trc.2015.01.021.

Brakewood, C. and K. Watkins (2019). "A Literature Review of the Passenger Benefits of Real-Time Transit Information". In: *Transport Reviews* 39.3, pp. 327–356. DOI: 10.1080/01441647.2018.1472147.

Cathey, F. W. and D. J. Dailey (2003). "A Prescription for Transit Arrival/Departure Prediction Using Automatic Vehicle Location Data". In: *Transportation Research Part C: Emerging Technologies* 11.3–4, pp. 241–264. DOI: 10.1016/S0968-090X(03)00023-8.

Cats, O. and G. Loutos (2016). "Real-Time Bus Arrival Information System: An Empirical Evaluation". In: *Journal of Intelligent Transportation Systems* 20.2, pp. 138–151. DOI: 10.1080/15472450.2015.1011638.

Chien, S. I.-J., Y. Ding, and C. Wei (2002). "Dynamic Bus Arrival Time Prediction with Artificial Neural Networks". In: *Journal of Transportation Engineering* 128.5, pp. 429–438. DOI: 10.1061/(ASCE)0733-947X(2002)128:5(429).

Diebold, F. X. and R. S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.3, pp. 253–263. DOI: 10.1080/07350015.1995.10524599.

Dziekan, K. and K. Kottenhoff (2007). "Dynamic At-Stop Real-Time Information Displays for Public Transport: Effects on Customers". In: *Transportation Research Part A: Policy and Practice* 41.6, pp. 489–501. DOI: 10.1016/j.tra.2006.11.006.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall. ISBN: 978-0-412-04231-7.

European Committee for Standardization (CEN) (2022). *Public Transport. Service Interface for Real-Time Information Relating to Public Transport Operations. Part 3: Functional Service Interfaces*. European Standard EN 15531-3:2022.

Ferris, B., K. Watkins, and A. Borning (2010). "OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1807–1816. DOI: 10.1145/1753326.1753597.

García-Mauriño, C., P. J. Zufiria, and A. Jarabo-Peñas (2024). "Improving Bus Arrival Time Predictors Using Only Public Transport API Data". In: *Transportation Letters* 16.8, pp. 804–813. DOI: 10.1080/19427867.2023.2245994.

Garrido, Y. (2026a). *Predicción de llegadas de autobuses con posiciones en tiempo real y horarios interpolados*. Tech. rep. Version 2.1. Zenodo. DOI: 10.5281/zenodo.24041572. In Spanish.

— (2026b). *Predicting Bus Arrivals When the Feed Reports Positions but Not Delays: A Reproducible Three-Layer Method with an Open Reference Implementation*. Preprint. Version 1.0. Zenodo. DOI: 10.5281/zenodo.24285541.

Gneiting, T. (2011). "Making and Evaluating Point Forecasts". In: *Journal of the American Statistical Association* 106.494, pp. 746–762. DOI: 10.1198/jasa.2011.r10138.

Gooze, A., K. E. Watkins, and A. Borning (2013). "Benefits of Real-Time Transit Information and Impacts of Data Accuracy on Rider Experience". In: *Transportation Research Record: Journal of the Transportation Research Board* 2351, pp. 95–103. DOI: 10.3141/2351-11.

Jeong, R. and L. R. Rilett (2005). "Prediction Model of Bus Arrival Time for Real-Time Applications". In: *Transportation Research Record: Journal of the Transportation Research Board* 1927, pp. 195–204. DOI: 10.1177/0361198105192700123.

Kaufman, S., S. Rosset, C. Perlich, and O. Stitelman (2012). "Leakage in Data Mining: Formulation, Detection, and Avoidance". In: *ACM Transactions on Knowledge Discovery from Data* 6.4, 15. DOI: 10.1145/2382577.2382579.

Kumar, B. A., R. Singh, H. E. Shaji, and L. Vanajakshi (2025). "Bus Arrival Time Prediction: A Comprehensive Review". In: *IEEE Transactions on Intelligent Transportation Systems* 26.6, pp. 7362–7379. DOI: 10.1109/TITS.2025.3545695.

Lin, W.-H. and J. Zeng (1999). "Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data". In: *Transportation Research Record: Journal of the Transportation Research Board* 1666, pp. 101–109. DOI: 10.3141/1666-12.

Liu, L. and H. J. Miller (2020). "Does Real-Time Transit Information Reduce Waiting Time? An Empirical Analysis". In: *Transportation Research Part A: Policy and Practice* 141, pp. 167–179. DOI: 10.1016/j.tra.2020.09.014.

Mazloumi, E., G. Currie, and G. Rose (2010). "Using GPS Data to Gain Insight into Public Transport Travel Time Variability". In: *Journal of Transportation Engineering* 136.7, pp. 623–631. DOI: 10.1061/(ASCE)TE.1943-5436.0000126.

Ministerio de Transportes y Movilidad Sostenible (2026). *Punto de Acceso Nacional de transporte multimodal*. URL: https://nap.transportes.gob.es (visited on 2026-10-31).

Mishalani, R. G., M. M. McCord, and J. Wirtz (2006). "Passenger Wait Time Perceptions at Bus Stops: Empirical Results and Impact on Evaluating Real-Time Bus Arrival Information". In: *Journal of Public Transportation* 9.2, pp. 89–106. DOI: 10.5038/2375-0901.9.2.5.

MobilityData (2026a). *GTFS Realtime Reference*. URL: https://gtfs.org/documentation/realtime/reference/ (visited on 2026-10-31).

— (2026b). *GTFS Schedule Reference*. URL: https://gtfs.org/documentation/schedule/reference/ (visited on 2026-10-31).

Moreira-Matias, L., J. Mendes-Moreira, J. F. de Sousa, and J. Gama (2015). "Improving Mass Transit Operations by Using AVL-Based Systems: A Survey". In: *IEEE Transactions on Intelligent Transportation Systems* 16.4, pp. 1636–1653. DOI: 10.1109/TITS.2014.2376772.

Petersen, N. C., F. Rodrigues, and F. C. Pereira (2019). "Multi-Output Bus Travel Time Prediction with Convolutional LSTM Neural Network". In: *Expert Systems with Applications* 120, pp. 426–435. DOI: 10.1016/j.eswa.2018.11.028.

Rahman, M. M., S. C. Wirasinghe, and L. Kattan (2013). "Users' Views on Current and Future Real-Time Bus Information Systems". In: *Journal of Advanced Transportation* 47.3, pp. 336–354. DOI: 10.1002/atr.1206.

Shalaby, A. and A. Farhan (2004). "Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data". In: *Journal of Public Transportation* 7.1, pp. 41–61. DOI: 10.5038/2375-0901.7.1.3.

Singh, N. and K. Kumar (2022). "A Review of Bus Arrival Time Prediction Using Artificial Intelligence". In: *WIREs Data Mining and Knowledge Discovery* 12.4, e1457. DOI: 10.1002/widm.1457.

Tang, L. and P. V. Thakuriah (2012). "Ridership Effects of Real-Time Bus Information System: A Case Study in the City of Chicago". In: *Transportation Research Part C: Emerging Technologies* 22, pp. 146–161. DOI: 10.1016/j.trc.2012.01.001.

Tashman, L. J. (2000). "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review". In: *International Journal of Forecasting* 16.4, pp. 437–450. DOI: 10.1016/S0169-2070(00)00065-0.

Vanajakshi, L., S. C. Subramanian, and R. Sivanandan (2009). "Travel Time Prediction under Heterogeneous Traffic Conditions Using Global Positioning System Data from Buses". In: *IET Intelligent Transport Systems* 3.1, pp. 1–9. DOI: 10.1049/iet-its:20080013.

Watkins, K. E., B. Ferris, A. Borning, G. S. Rutherford, and D. Layton (2011). "Where Is My Bus? Impact of Mobile Real-Time Information on the Perceived and Actual Wait Time of Transit Riders". In: *Transportation Research Part A: Policy and Practice* 45.8, pp. 839–848. DOI: 10.1016/j.tra.2011.06.010.

Wessel, N., J. Allen, and S. Farber (2017). "Constructing a Routable Retrospective Transit Timetable from a Real-Time Vehicle Location Feed and GTFS". In: *Journal of Transport Geography* 62, pp. 92–97. DOI: 10.1016/j.jtrangeo.2017.04.012.

Yu, B., W. H. K. Lam, and M. L. Tam (2011). "Bus Arrival Time Prediction at Bus Stop with Multiple Routes". In: *Transportation Research Part C: Emerging Technologies* 19.6, pp. 1157–1170. DOI: 10.1016/j.trc.2011.01.003.