https://pubsonline.informs.org/journal/trsc

TRANSPORTATION SCIENCE
Vol. 58, No. 1, January–February 2024, pp. 8–11
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

2021 Amazon Last Mile Routing Research Challenge: Data Set

Daniel Merchán,[a] Jatin Arora,[a] Julian Pachon,[a] Karthik Konduri,[a] Matthias Winkenbach,[b,*] Steven Parks,[b] Joseph Noszek[b]

[a] Amazon.com, Inc., Bellevue, Washington 98004; [b] Massachusetts Institute of Technology, Center for Transportation & Logistics, Cambridge, Massachusetts 02139
*Corresponding author

Contact: damercha@amazon.com, https://orcid.org/0000-0003-1822-083X (DM); arorjati@amazon.com (JA); pachonj@amazon.com (JP); kkondur@amazon.com (KK); mwinkenb@mit.edu, https://orcid.org/0000-0002-8237-625X (MW); stevenp@mit.edu, https://orcid.org/0000-0002-8018-9565 (SP); jnoszek@mit.edu, https://orcid.org/0000-0002-6214-5585 (JN)

Received: July 30, 2022
Revised: August 23, 2022
Accepted: August 25, 2022
Published Online in Articles in Advance: September 14, 2022

https://doi.org/10.1287/trsc.2022.1173

Abstract. The 2021 Amazon Last Mile Routing Research Challenge, hosted by Amazon.com's Last Mile Research team, and scientifically supported by the Massachusetts Institute of Technology's Center for Transportation and Logistics, prompted participants to leverage real operational data to find new and better ways to solve a real-world routing problem. In this article, we describe the data set released for the research challenge, which includes route-, stop-, and package-level features for 9,184 historical routes performed by Amazon drivers in 2018 in five metropolitan areas in the United States. This real-world data set excludes any personally identifiable information: all route and package identifiers have been randomly regenerated and related location data have been obfuscated to ensure anonymity. Although multiple synthetic benchmark data sets are available in the literature, the data set of the 2021 Amazon Last Mile Routing Research Challenge is the first large and publicly available data set to include instances based on real-world operational routing data.

History: This paper has been accepted for the *Transportation Science* Special Issue on Machine Learning Methods and Applications in Large-Scale Route Planning Problems.

Keywords: vehicle routing • machine learning • optimization • sequencing • last-mile delivery • data set

1. Introduction

On March 15, 2021, Amazon.com launched the *2021 Amazon Last Mile Routing Research Challenge* (referred to as the *research challenge* in the following) with scientific support from a team of researchers at the Massachusetts Institute of Technology's Center for Transportation and Logistics. The research challenge encouraged participants worldwide to develop innovative approaches leveraging data science and optimization methods to produce delivery routes that reflect the tacit knowledge of seasoned drivers gleaned through years of experience. Specifically, participants were asked to build a model that would use data on historically observed routes and their stop sequences to propose stop sequences for new sets of orders in previously unseen routes for which the historical stop sequences are unknown to the model. The proposed route sequences should be as similar as possible to high-quality stop sequences operated by experienced delivery drivers.

To this end, participants were provided with data for model training from 6,112 historical routes traversed by Amazon delivery drivers across five major metropolitan areas in the United States. Each route is characterized by myriad features including a cost matrix of historically realized transit times, the driver-operated sequence of delivery stops, the weights and dimensions of the packages delivered on each route, and delivery time windows specified by customers, among others. Algorithms developed by participants were evaluated against another set of 3,072 historical routes. In total, 45 final submissions to the challenge were received and evaluated; winners were announced on July 30, 2021.

From an academic perspective, the research challenge aimed to engage students and academics across the world with an underrepresented research stream on solving routing problems. Data-driven, learning-based solution approaches to the vehicle routing problem (VRP) and its variants are still relatively absent in the academic literature compared with conventional, optimization-driven methods. Furthermore, most routing problems discussed in the literature aim to minimize a clear, explicit objective function such as total route time or the sum of vehicle operational costs. The routing problem defined in the research challenge pushed participants to instead propose techniques that capture the complex decision making and tacit knowledge of seasoned delivery drivers and generate route sequences that a typical driver would consider to be of good quality, implicitly factoring in quality criteria such as convenience, efficiency, and safety.

The primary objective of this paper is to introduce the research challenge data set. Benchmark data sets have been extensively used for over four decades to directly

compare newly proposed solution methods to solve the VRP and its variants, both in terms of computational efficiency and solution quality. Classic benchmark data sets such as the ones presented by Christofides, Mingozzi, and Toth (1979), Solomon (1987), and Golden et al. (1998) are still in common use. Many other established data sets in the literature see, e.g., (Taillard, Laporte, and Gendreau 1996) are modifications of these preceding data sets. We refer the reader to the recent article by Gunawan et al. (2021) for a comprehensive taxonomy of commonly used benchmark data sets. To the best of our knowledge, all benchmark data sets available in the extant literature have been generated based on synthetic data. In this paper, we introduce a novel, publicly available, real-world data set that researchers may use as a test bed to develop and evaluate new ideas on how to solve realistically sized instances of real-world routing problems. This unique data set establishes a large collection of benchmark instances based on real operational data that researchers can use to advance the state-of-the-art in vehicle routing research and to expand its applications to industry settings.

2. Data Set Description

The routing challenge data set is available through the Registry of Open Data on Amazon Web Services (AWS) at https://registry.opendata.aws/amazon-last-mile-challenges. It has been split into training and evaluation sets.

The training data set contains 6,112 routes historically performed by Amazon drivers in 2018 in the metropolitan areas of Seattle, Los Angeles, Austin, Chicago, and Boston. Each route is characterized by a variety of route-level, stop-level, and package-level features, summarized in Table 1 and explained in Sections 2.1 through 2.3. Notable summary statistics of select route features in the training data set are provided in Table 2. Each route in this data set is labeled according to its perceived route quality (i.e., *low*, *medium*, and *high*) so that it can serve as a basis to train machine learning models or tune the parameters of traditional optimization models. These labels were created by analyzing route quality attributes and delivery defects along with various other dimensions of interest. These dimensions of interest can broadly be categorized as related to driver experience, customer satisfaction, and productivity.

The evaluation data set contains the same features for an additional 3,072 historically observed routes. However, routes in this evaluation set are of *high* perceived quality exclusively. For the research challenge, this evaluation data set was fed to models fitted using the training set to predict high-quality stop sequences and to assess model performance and solution quality. The summary statistics of select route features in the evaluation data set are provided in Table 3.

No personally identifiable information has been included in the data. All package and route identifiers were randomly regenerated, and related location data were obfuscated and perturbed to ensure anonymity and protect the privacy of drivers and delivery recipients.

2.1. Route-Level Features

The following features describe the individual routes.

- Route ID: A randomly generated alphanumeric string that uniquely identifies each route. No personally

Table 1. High-Level Description of Data Fields Provided in the Research Challenge Data Set

Data field	Description	Unit/format
	Route information	
Route ID	Unique and anonymized identifier of each route	—
Station code	Unique identifier for a delivery station where routes begin	—
Date	Date of route execution	YYYY-MM-DD
Departure time	Time when vehicle leaves the station	—
Executor capacity	Volumetric capacity of vehicle	cm^3
Stops	Each stop on route	—
Observed sequence	Sequence in which stops were visited	—
Route score	Quality of the observed sequence	Categorical
	Stop information	
Stop ID	Unique identifier of each stop on a route	—
Latitude/longitude	Obfuscated coordinates of each stop	—
Type	Type of stop	Categorical
Zone ID	Geographical planning area	—
Packages	Packages delivered at each stop	—
Transit time	Estimated transit time to every other stop on route	Seconds
	Package information	
Package ID	Unique and anonymized identifier of each package	—
Status	Delivery status of package	Categorical
Time window	Start and end of time window, when applicable	—
Planned service time	Time that serving the package is expected to require	Seconds
Dimensions	Length, width, and height of package	cm

Table 2. Summary Statistics of the 6,112 Training Routes Included in the Data Set

	No. of stops	No. of packages	No. of packages with time windows	Total route time (h)	Transit time (h)	Service time (h)
P_{10}%	105.0	197.0	2.0	7.1	2.6	3.4
Mean	148.0	238.5	18.6	8.1	3.6	4.5
P_{90}%	186.0	278.0	40.0	9.1	4.6	5.6
Standard deviation	31.0	31.0	15.2	0.8	0.7	0.9

identifiable information of delivery drivers has been included.

- Station Code: An alphanumeric string that uniquely identifies the delivery station (or depot) at which the route began.
- Date: The date the delivery vehicle departed from the station.
- Departure Time: The time when the vehicle departed the delivery station, specified in Coordinated Universal Time (UTC).
- Executor Capacity: The volumetric capacity of the vehicle, specified in cm^3.
- Stops: A list of each stop that was served in the route.
- Observed sequence: The order in which the stops were visited.
- Route Score: The quality of the observed sequence. This is in the form of a categorical variable that can take the values of *high*, *medium*, or *low*.

2.2. Stop-Level Features

The following features describe each stop along each route.

- Stop ID: A unique identifier of each stop within a route. Each stop ID consists of two letters. All stop IDs within a route are unique. However, a stop on one route can share an ID with an unrelated stop on another route.
- Latitude/Longitude: The coordinates of each stop, specified via the World Geodetic System (WGS) 84 projection system. Coordinates have been anonymized and perturbed to protect the privacy of delivery recipients.
- Type: Categorical variable denoting the type of stop, either *station* or *dropoff*. The delivery vehicle acquires all packages at the station and delivers them at subsequent drop-off locations.
- Zone ID: A unique identifier denoting the geographical planning area into which the stop falls. The numeral before the dash denotes a high-level planning zone. The text after the dash denotes the subzone within the high-level zone. All zones are devised internally by Amazon.
- Packages: A list of packages assigned to be delivered at the stop.
- Transit times: For a given pair of stops, the average of historically realized transit times between all combinations of package delivery locations between those stops, specified in seconds.

2.3. Package-Level Features

The following features describe each package delivered at each stop.

- Package ID: A randomly generated alphanumeric string that uniquely identifies each package within a route. Package identifiers are not shared among routes. No personally identifiable information of delivery recipients has been included.
- Scan Status: A categorical variable denoting the delivery state of a package, either *DELIVERED*, *DELIVERY_ATTEMPTED*, or *REJECTED*. In some cases, a package may experience a failed delivery attempt and a successful delivery in the same route. Where this occurs, the package will still have a single ID, but the associated stop will have two stop IDs, one corresponding to the failed attempt and another to the successful attempt.
- Time Window: The interval of time in which package delivery is acceptable, defined by *start_time_utc* and *end_time_utc*, both specified in UTC. If a package's *start_time_utc* and *end_time_utc* fields are *NaN*, no time window was specified.

Table 3. Summary Statistics of the 3,072 Evaluation Routes Included in the Data Set

	No. of stops	No. of packages	No. of packages with time windows	Total route time (h)	Transit time (h)	Service time (h)
P_{10}%	96.0	193.0	1.0	6.9	2.3	3.4
Mean	141.9	237.5	17.7	8.0	3.4	4.6
P_{90}%	183.0	279.0	41.0	8.9	4.4	5.8
Standard deviation	33.2	32.7	15.8	0.8	0.8	0.9

• Planned Service Time: Time in seconds that serving the package is expected to require.

• Dimensions: Length, width, and height of the package, each specified in centimeters.

3. Conclusion

This paper introduces the first large, real-world, publicly available routing data set, which was initially released for the 2021 Amazon Last Mile Routing Research Challenge. This novel and unique data set, which was anonymized and obfuscated to remove any personally identifiable information, includes a training set of 6,112 historical routes and an evaluation set of 3,072 historical routes traversed by Amazon delivery drivers across five major metropolitan areas in the United States. For each observed route instance, package-, stop-, and route-level features are included. We expect this data set will help narrow the existing gap between theoretical route planning and real-life route execution by incorporating into methodological approaches mechanisms to learn from tacit knowledge of experienced delivery drivers about the complex operational environments in which they operate.

4. Downloading the Data

The AWS command line interface (CLI) (Amazon Web Services 2022) is the recommended tool to access the research challenge data set. After installing the AWS CLI, the following command can be executed to download the data to a local directory specified by the user: `aws s3 sync --no-sign-request s3://amazon-last-mile-challenges/almrrc2021/{local directory}`

For further instructions and documentation, we refer the reader to the data set page in the Registry of Open Data on AWS available at https://registry.opendata.aws/amazon-last-mile-challenges.

Acknowledgments

The authors thank the editorial team at *Transportation Science*, specifically Karen Smilowitz, Jean-François Cordeau, and Barrett Thomas, for valuable insights and helpful suggestions. The authors also thank André Snoeck and Mehmet Candas at Amazon.com and Connor Makowski and Tim Russell at Massachusetts Institute of Technology for supporting the research challenge.

References

Amazon Web Services (2022) AWS command line interface. Accessed July 15, 2022, https://aws.amazon.com/cli/.

Christofides N, Mingozzi A, Toth P (1979) The vehicle routing problem. Christofides N, Mingozzi A, Toth P, Sandi C, eds. *Combinatorial Optimization* (Wiley, Chichester, UK), 315–338.

Golden BL, Wasil EA, Kelly JP, Chao IM (1998) The impact of metaheuristics on solving the vehicle routing problem: Algorithms, problem sets, and computational results. *Fleet Management and Logistics* (Springer, Boston, MA), 33–56.

Gunawan A, Kendall G, McCollum B, Seow HV, Lee LS (2021) Vehicle routing: Review of benchmark datasets. *J. Oper. Res. Soc.* 72(8):1–14.

Solomon MM (1987) Algorithms for the vehicle routing and scheduling problems with time window constraints. *Oper. Res.* 35(2): 254–265.

Taillard ÉD, Laporte G, Gendreau M (1996) Vehicle routing with multiple use of vehicles. *J. Oper. Res. Soc.* 47(8):1065–1070.