TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Benghe, C., Graure, V., Shreedhar, T., & Mohan, N. (2026). Horizon: Understanding and Predicting Global Starlink Performance. *Performance Evaluation Review*, *54*(1), 171-173. https://doi.org/10.1145/3820171.3806896

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

*This work is downloaded from Delft University of Technology.*

# Horizon: Understanding and Predicting Global Starlink Performance

Cristian Benghe*
C.Benghe@student.tudelft.nl
TU Delft
The Netherlands

Vlad Graure*†
vladgraure@yahoo.ca
TU Delft
The Netherlands

Tanya Shreedhar
t.shreedhar@tudelft.nl
TU Delft
The Netherlands

Nitinder Mohan
N.Mohan@tudelft.nl
TU Delft
The Netherlands

## Abstract

Starlink has deployed over 7,800 satellites serving millions of subscribers, yet predicting its performance remains an open challenge. Rapid orbital dynamics, frequent handovers, and weather-induced signal attenuation create variability that existing models, built on a handful of instrumented terminals in limited regions, cannot capture at global scale. We present `Horizon`, the first global-scale machine learning system for predicting LEO satellite Internet performance. Our key insight is that crowdsourced measurement platforms, while noisier than controlled experiments, provide the geographic diversity necessary to build globally generalizable models. `Horizon` integrates 11 months of measurements from M-Lab and Cloudflare spanning 90+ countries with meteorological data and satellite orbital propagation features. On a fully held-out one-week temporal window, `Horizon` achieves mean absolute errors of 17.76 ms for latency and 25.63 Mbps for throughput; on a standard 80/20 split it outperforms all baselines, including adaptations of state-of-the-art architectures. Feature importance analysis reveals that geographic position dominates prediction, with latitude alone contributing 42–46%, while weather features account for 14–15%, quantifying the impact of atmospheric conditions on Ku/Ka-band links. Leave-one-location-out experiments confirm that `Horizon` generalizes to regions absent from training, enabling performance estimation where measurement infrastructure does not yet exist. Our dataset and pipeline are publicly available, providing a foundation for global LEO network performance visibility.

## CCS Concepts

• **Networks** → **Network performance modeling**; **Network performance analysis**.

## Keywords

LEO Internet, Crowdsourced Telemetry, Performance Prediction, Starlink, Machine Learning

*Both authors contributed equally to this research.
†Now employed at Google.



This work is licensed under a Creative Commons Attribution 4.0 International License.
*SIGMETRICS Abstracts '26, Ann Arbor, MI, USA*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2570-8/2026/06
https://doi.org/10.1145/3801489.3806896

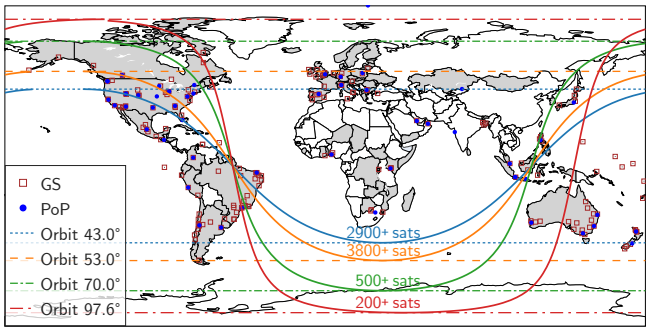



**Figure 1: Starlink infrastructure: ground stations, PoPs, and four orbital shells with satellite counts, as of January 2026. Countries with availability shown in grey.**

**ACM Reference Format:**
Cristian Benghe, Vlad Graure, Tanya Shreedhar, and Nitinder Mohan. 2026. Horizon: Understanding and Predicting Global Starlink Performance. In *Abstracts of the 2026 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS Abstracts '26), June 08–12, 2026, Ann Arbor, MI, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3801489.3806896

## 1 Introduction

The global demand for reliable Internet connectivity continues to grow, yet large populations in rural, remote, and underserved regions remain poorly connected because deploying terrestrial infrastructure is often economically infeasible [12, 17]. Low-Earth-Orbit (LEO) satellite constellations have emerged as a promising alternative, and Starlink [20] operates the largest such constellation, spanning four orbital shells with over 7,800 active satellites serving millions of subscribers across more than 140 countries [19], as illustrated in fig. 1. User terminals communicate with overhead satellites via high-frequency Ku-band and Ka-band links in a bent-pipe architecture, where traffic is relayed through one or more satellites connected via inter-satellite links to ground stations [16].

The core difficulty in predicting LEO satellite performance stems from the unique and complex characteristics of these networks. Unlike terrestrial networks where infrastructure is relatively static,

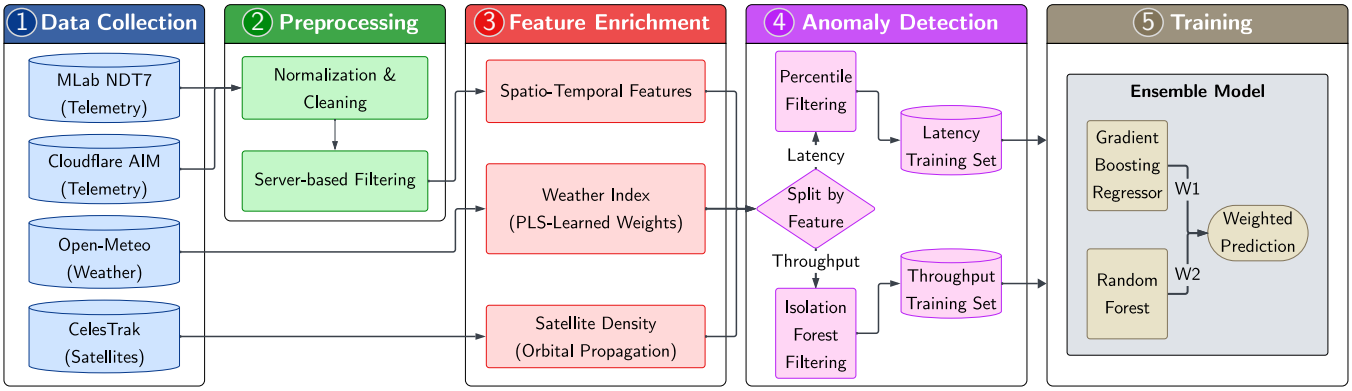



**Figure 2: Overview of the Horizon pipeline. Phase 1 collects crowdsourced measurements, weather data, and satellite positions. Phase 2 preprocesses and validates cross-dataset compatibility. Phase 3 engineers features including a novel Weather Index and satellite density. Phase 4 applies location-aware anomaly filtering. Phase 5 trains the ensemble prediction model.**

LEO satellites move at approximately 7.6 km/s, forcing user terminals to execute frequent handovers every ≈15 seconds as satellites enter and exit view [7]. Simultaneously, ground infrastructure evolves continuously as new ground stations and points of presence alter routing paths [4]. Weather compounds the variability further: these high-frequency bands are inherently susceptible to atmospheric attenuation, with studies documenting up to 50% throughput reduction during heavy precipitation [10, 13]. Together, these three factors create enormous performance variability that changes dramatically across geographic regions, times of day, and environmental conditions [16].

Prior research has attempted to characterize and predict Starlink performance. Multiple studies have documented the impact of weather on connectivity [8, 10, 13], and predictive models have been explored for download latency and throughput [9, 11, 21]. However, all existing work relies on professionally instrumented terminals deployed in limited geographic regions, predominantly concentrated in Europe and North America where infrastructure is mature and measurement costs are manageable [18]. This geographic bias severely constrains the generalizability of existing models. The vast majority of Earth's population lives in Africa, South America, and Asia, regions where Starlink could have the highest societal impact but where existing measurement infrastructure is sparse. This creates a critical gap where the regions that could benefit most from reliable LEO connectivity are precisely those where we understand the network least well.

We present Horizon, the first global-scale machine learning system for understanding and predicting LEO satellite network performance. Our key insight is that crowdsourced measurement platforms [6], provide the geographic diversity and longitudinal coverage necessary to build prediction models that actually generalize across different regions of the world. Please see [1] for full details.

## 2 The Horizon Pipeline

Figure 2 shows the end-to-end Horizon architecture. 1 *Data Collection* includes two major crowdsourced measurement platforms. M-Lab NDT7 [14, 15] provides speed tests contributed voluntarily by internet users worldwide, operating over 500 measurement servers across 60 metropolitan areas and generating ≈15.6M speed test results from Starlink subscribers during our 11-month study period (January–November 2025). Each test measures application-level download and upload performance by transferring data over a secure WebSocket connection using the BBR congestion control algorithm. Cloudflare AIM [22] provides quality scores for different Internet use cases like gaming, streaming, and video conferencing, generating ≈157K measurements including download and upload throughput, latency metrics, jitter measurements, and packet loss statistics evaluated using multiple test methodologies including HTTPS transfers and UDP-based real-time communication. Cross-dataset compatibility is validated via Jensen-Shannon Divergence analysis, confirming the distribution similarity between the two platforms. Combined, these data sources provide measurements spanning 90+ countries across six continents. We complement these with high-resolution meteorological data from the OpenMeteo API [25], capturing temperature, precipitation, wind speed, cloud cover at hourly resolution, and satellite density estimates computed via TLE/SGP4 orbital propagation [24] for each geographic location. We validated weather data reliability by comparing OpenMeteo against ground-truth observations from national meteorological institutes across 20 weather stations globally, finding strong temporal agreement: temperature closely tracks observations with clear diurnal cycles, precipitation differences remain below 1 millimeter per hour, wind speed follows consistent trends, and cloud cover matches in relative variation.

2 *Preprocessing* addresses server selection bias, a critical issue where users can be connected to distant servers under network congestion, infrastructure failures, or suboptimal routing policies, inflating latency values. We apply server filtering to identify the best performing servers for each geographic location on a monthly basis by selecting servers that consistently deliver the lowest latency for that region, ensuring stability across months. We retain only measurements from best-performing servers (servers hat achieved a 1st percentile latency per location per month), removing 9.15% of Cloudflare and 5.82% of NDT7 data, with substantially higher rates in African countries with sparse infrastructure.

3 *Feature Enrichment* produces spatio-temporal features (coordinates, client-server distance, hour of day, day of week, month),

a novel Weather Index combining precipitation, cloud cover, temperature and wind speed via Partial Least Squares regression and satellite density estimates via orbital propagation [24]. ❹ *Anomaly Detection* applies percentile-based filtering (75th percentile threshold) for latency and Isolation Forest with 0.25 contamination for throughput to remove outliers while preserving natural performance variability. ❺ *Model Training* employs an ensemble of Random Forest (100 trees) and Gradient Boosting (100 estimators, learning rate 0.1, max depth 5) regressors [5]. Random Forest provides robustness through bagging and captures non-linear interactions, while Gradient Boosting iteratively corrects predictions. We compare against Linear Regression, KNN, Decision Trees, XGBoost, and deep learning architectures including LSTM (two-layer, 64 hidden units, 20% dropout following T3P [21]) and GRU Seq2Seq with attention mechanisms (inspired by StarNet [11]), all trained on identical feature sets and evaluated using Mean Absolute Error (MAE), Root Mean Squarred Error (RMSE), and Coefficient of Determination ($R^2$) metrics.

## 3 Results

On a global held-out test set spanning 24–30 November 2025, `Horizon` achieves MAE of 17.76 ms for latency and 25.63 Mbps for throughput, with RMSE of 26.21 ms and 38.41 Mbps and $R^2$ values of 0.574 and 0.369 respectively. On an 80/20 random train-test split, `Horizon` substantially outperforms all baselines, including XGBoost, and deep learning methods: LSTM [21] achieved 26.54 ms and 41.75 Mbps; GRU Seq2Seq with attention [11] achieved 25.12 ms and 41.81 Mbps, confirming that tree-based ensemble methods better capture these complex nonlinear relationships on tabular data [5]. Feature importance analysis reveals geographic position as the dominant predictor: latitude alone contributes 42–46% of total prediction accuracy, reflecting fundamental physical constraints on satellite visibility and ground infrastructure proximity. Weather features account for 14.0–15.3% overall, with precipitation and cloud cover as the dominant factors, validating atmospheric Ku/Ka-band signal attenuation [10, 23]. Leave-one-location-out experiments across three continents confirm that `Horizon` generalizes to regions absent from training, enabling performance estimation where measurement infrastructure does not yet exist. `Horizon` operates as a continuously deployed system generating hourly forecasts for over 300 H3 hexagons worldwide, incorporating real-time weather forecasts and weekly model retraining. Full experimental details, regional analysis, and limitations are presented in the full paper [1]. Our dataset [2] and code [3] are publicly available.

## Acknowledgements.

This research was supported by the National Growth Fund through the Dutch 6G flagship project "Future Network Services" and the Internet Society Pulse Research Fellowship.

## References

[1] Cristian Benghe, Vlad Graure, Tanya Shreedhar, and Nitinder Mohan. Horizon: Understanding and Predicting Global Starlink Performance. *Proc. ACM Meas. Anal. Comput. Syst.*, 10(2), jun 2026.

[2] Cristian Benghe, Vlad Graure, Tanya Shreedhar, and Nitinder Mohan. Horizon: Understanding and Predicting Global Starlink Performance – Dataset. https://doi.org/10.4121/0bf59468-e5cb-433f-aeb2-e04cf694b65c, 2026. Accessed: 2026-03-27.

[3] Cristian Benghe, Vlad Graure, Tanya Shreedhar, and Nitinder Mohan. Horizon: Understanding and Predicting Global Starlink Performance – Source Code. https://github.com/spear-lab/Horizon-Predicting-Starlink-Performance, 2026. Accessed: 2026-03-27.

[4] Rohan Bose, Jinwei Zhao, Tanya Shreedhar, Jianping Pan, and Nitinder Mohan. Investigating Web Content Delivery Performance over Starlink. In *Proceedings of the ACM Web Conference*, 2026.

[5] Leo Grinsztajn, Edouard Oyallon, and Gael Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 507–520. Curran Associates, Inc., 2022.

[6] Matthias Hirth, Tobias Hoßfeld, Marco Mellia, Christian Schwartz, and Frank Lehrieder. Crowdsourced Network Measurements: Benefits and Best Practices. *Computer Networks*, 90:85–98, 2015.

[7] Geoff Huston. The ISP Column: Using LEOs and GEOs. https://www.potaroo.net/ispcol/2022-04/leogeo.html, 2022.

[8] Mohamed M. Kassem, Aravindh Raman, Diego Perino, and Nishanth Sastry. A Browser-side View of Starlink Connectivity. In *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22)*, pages 151–158. Association for Computing Machinery, 2022.

[9] Eric Lanfer, Dominic Laniewski, Daniel Otten, and Nils Aschenbruck. Weather-Based Link Prediction for LEO-Satellite Networks using the WetLinks Dataset. In *Proceedings of the 23rd IFIP/IEEE Networking Conference (IFIP Networking 2024)*, pages 586–588. IEEE, 2024.

[10] Dominic Laniewski, Eric Lanfer, Bernd Meijerink, Roland van Rijswijk-Deij, and Nils Aschenbruck. WetLinks: A Large-Scale Longitudinal Starlink Dataset with Contiguous Weather Data. In *Proceedings of the 8th Network Traffic Measurement and Analysis Conference (TMA 2024)*, pages 1–9. IEEE, 2024.

[11] Zikun Liu, Fan-Xue Gabriella Reidys, Sarah Tanveer, and Deepak Vasisht. Vivisecting Starlink Throughput: Measurement and Prediction. *Proc. ACM Netw.*, 3(CoNEXT4), November 2025.

[12] Lyntia. Telecommunications Technologies Helping Close the Digital Divide. https://www.lyntia.com/en/news/telecommunications-technologies-to-close-the-digital-divide/, 2024. Accessed: 2025-12-28.

[13] Sami Ma, Yi Ching Chou, Haoyuan Zhao, Long Chen, Xiaoqiang Ma, and Jiangchuan Liu. Network Characteristics of LEO Satellite Constellations: A Starlink-Based Measurement from End Users. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM 2023)*, pages 1–10. IEEE, 2023.

[14] Kyle MacMillan, Tarun Mangla, James Saxon, Nicole P. Marwell, and Nick Feamster. A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7). *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 7(1):1–26, 2023.

[15] Measurement Lab. NDT (Network Diagnostic Tool). https://www.measurementlab.net/tests/ndt/, 2025. Accessed: 2025-12-28.

[16] Nitinder Mohan, Andrew E. Ferguson, Hendrik Cech, Rohan Bose, Prakita Rayyan Renatin, Mahesh K. Marina, and Jörg Ott. A Multifaceted Look at Starlink Performance. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*, pages 2723–2734. Association for Computing Machinery, 2024.

[17] Ilana Newman. For Rural Communities, Broadband Expansion Is No Single Thing. https://www.governing.com/infrastructure/for-rural-communities-broadband-expansion-is-no-single-thing, 2023. Accessed: 2025-12-28.

[18] Jianping Pan, Jinwei Zhao, and Lin Cai. Measuring a Low-Earth-Orbit Satellite Network. *arXiv preprint arXiv:2307.06863*, 2023.

[19] SpaceX. Starlink Network Update. https://starlink.com/updates/network-update, 2025. Accessed: 2026-01-10.

[20] SpaceX. Starlink: Reliable High-Speed Internet from Space. https://www.starlink.com, 2025. Accessed: 2025-12-28.

[21] Shubham Tiwari, Saksham Bhushan, Aryan Taneja, Mohamed Kassem, Cheng Luo, Cong Zhou, Zhiyuan He, Aravindh Raman, Nishanth Sastry, Lili Qiu, and Debopam Bhattacherjee. T3P: Demystifying Low-Earth Orbit Satellite Broadband, 2023.

[22] David Tuber. Measuring network quality to better understand the end-user experience. https://blog.cloudflare.com/aim-database-for-internet-quality/, 2023. Accessed: 2025-12-28.

[23] Muhammad Asad Ullah, Antti Heikkinen, Mikko Uitto, Antti Anttonen, and Konstantin Mikhaylov. Impact of Weather on Satellite Communication: Evaluating Starlink Resilience. *arXiv preprint arXiv:2505.04772*, 2025.

[24] David A. Vallado, Paul Crawford, Richard Hujsak, and T. S. Kelso. Revisiting Spacetrack Report #3. *AIAA/AAS Astrodynamics Specialist Conference*, 2006.

[25] Patrick Zippenfenig. Open-Meteo.com Weather API. https://open-meteo.com/, 2023. CC-BY-4.0; Creative Commons Attribution 4.0 International; Accessed: 2025-12-28.