Assessing the Impact of Spatial Data Quality on Deep Learning Models for Travel Time Estimation

Diêgo de A. Correia[1], Rafael Luciano L. Silva[1], Vinicius Santiago[1], Jose Matheus S. Alves[1], Ruan T. Melo[1], J. Rui R. S. Fernandes[1], Fábio J. Coutinho[1]

[1]Instituto de Computação – Universidade Federal de Alagoas (UFAL)

{dac, rlls, vgss, jmsa, rtm, jrrsf, fabio}@ic.ufal.br

Abstract. *Data quality is fundamental to the reliability of predictive models. In the context of spatial data, understanding how quality dimensions influence robustness is critical for intelligent transportation systems. This study investigates the impact of spatial data quality on Deep Learning models for Travel Time Estimation (TTE). The methodology encompasses the steps of data pollution, model training, and testing under two complementary scenarios: polluted training and test data. We evaluate three models using the Porto Taxi dataset while considering the impact of the Positional Accuracy and Completeness quality dimensions defined in ISO 19157-1:2023. The results indicate that the impact of data quality degradation varies according to both model architecture and application phase.*

1. Introduction

In recent years, due to the growing recognition of data as a key factor in improving the effectiveness of Artificial Intelligence (AI)-based systems, researchers have advocated a shift from a model-centric perspective to a data-centric approach. Within this context, the concept of Data Readiness for AI emphasizes the importance of preparing data and ensuring its quality, accessibility, and suitability for AI applications [Hiniduma et al. 2025].

In Travel Time Estimation (TTE) problems, the data used by Deep Learning (DL) models must capture and represent complex temporal and spatial dependencies. TTE plays a fundamental role in Intelligent Transportation Systems (ITS), supporting modern location-based services, such as navigation, ride-hailing, and route planning [Zheng et al. 2023, Zuo et al. 2026]. To support the assessment of spatial data quality, ISO 19157-1:2023 defines key quality dimensions, including completeness, logical consistency, positional accuracy, temporal, and thematic quality [Silva et al. 2025].

Some effort has been devoted to assessing the impact of data quality on machine learning performance for specific data types. [Mohammed et al. 2025] analyzed selected quality dimensions in tabular data to understand how Machine Learning (ML) algorithms behave under varying levels of data degradation by introducing controlled data pollution. However, to the best of our knowledge, no previous study has evaluated the impact of spatial data quality on DL models for TTE problems. This work presents an experimental evaluation of the influence of spatial data quality on TTE problems by analyzing three different DL models under progressively polluted training and testing datasets.

The remainder of this paper is divided as follows: Section 2 reviews related works on the effects of spatial data quality for AI. Section 3 presents the setup for the experiment evaluated in this study, delving into the data pollution and model training process. Section

4 presents the results for the experiments applied on the different proposed scenarios. Finally, Section 5 concludes the paper and outlines directions for future research.

2. Related Work

Preprocessing methods and data quality assurance play a decisive role in the performance and analyses generated by intelligent systems [Chen 2025]. Several studies in the literature evaluate the impact of data quality on the use of ML models, which are briefly discussed below.

The study conducted by [Si et al. 2023] highlights the critical relevance of data quality for the success of models in transportation applications, such as the TTE. However, the paper does not specifically address the spatial dimensions defined by ISO 19157-1 (2023). Furthermore, while the work conducts experiments demonstrating data quality improvements in the utilized databases, it neither implements nor evaluates the corresponding impact on prediction models.

For common tabular data, [Mohammed et al. 2025] presents a comprehensive empirical analysis of how data quality affects the performance of ML models and distinguishes three experimental scenarios based on the stage of the pipeline where data pollution occurs: in the training set, in the test set, or in both. The study concludes that the dimensions of completeness and feature accuracy have the greatest impact across all scenarios and addresses the effect on ML performance in each case.

Although these works reinforce the importance of evaluating the robustness of DL models under polluted data conditions, research specifically addressing the impact of spatial data quality remains limited. To bridge this gap, this work investigates how spatial data quality attributes affect DL models for the TTE problem. Our analysis leverages real-world experiments and open-source models to ensure practical relevance.

3. Experiment

This section describes the experimental framework developed to evaluate the impact of spatial data quality degradation on deep learning models for TTE. The experimental analysis is grounded in the Porto Taxi Dataset[1]. This dataset contains 1.71 million trajectories with GPS traces logged every 15 seconds, recorded over a year from 442 taxis operating in the city of Porto, Portugal. Each entry in the dataset represents a single taxi ride.

The methodology adopted in this experiment encompasses the steps of data pollution, model training, and testing under two complementary scenarios, and the evaluation of performance metrics. Figure 1 illustrates this methodology. The following sections discuss each of these steps in detail.

3.1. Data Pollution

Assessing model robustness against data noise requires systematic pollution strategies. [Vitagliano et al. 2023] formalized this approach for tabular data, generating controlled degraded versions to analyze system robustness. We bring this paradigm into the spatial domain, applying controlled pollution over real-world GPS trajectory data to evaluate the resilience of DL models for TTE under quality degradation scenarios.

[1]https://www.kaggle.com/datasets/crailtap/taxi-trajectory

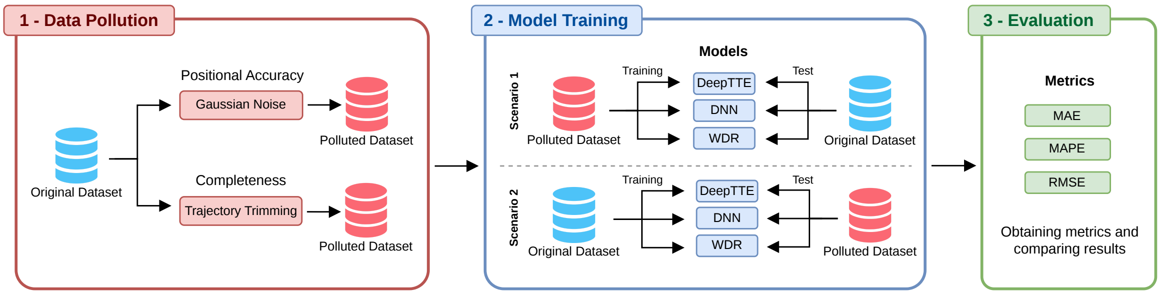

Figure 1. Methodology for evaluating data quality impact in TTE models.

Following ISO 19157-1:2023, we select Positional Accuracy and Completeness as our pollution targets. These dimensions consistently rank among the most investigated in extrinsic spatial quality evaluations [Minghini and Frassinelli 2019]. Both are further classified as key metrics for evaluating spatial data through machine learning [Tavra et al. 2026], providing a principled foundation for stress-testing TTE models against positional distorting and sequential data omission.

3.1.1. Positional Accuracy

To pollute the positional accuracy of the trajectory dataset, we first modeled the noise consistent with real-world Global Navigation Satellite System operating conditions. The positioning error experienced by user equipment is influenced by local environmental factors, such as multi-path interference and signal attenuation. The *2020 GPS Standard Positioning Service Performance Standard*, published by the U.S. government[2], establishes that, for the *Global Average Position Domain*, horizontal position errors (i.e., latitude and longitude) shall not exceed 15 meters for the *Worst-Site Position Domain*. Our methodology adopts a noise model initially calibrated around this value. Thereafter, to simulate stronger disturbances along the trajectory, we stretched these values to 30 and 50 meters.

We implemented Gaussian Noise as the positional distortion method, where the parameters are chosen such that the resulting positional perturbations reflect the degraded conditions consistent with the aforementioned thresholds, thereby simulating dense urban areas. The noise injection is applied to the complete trajectory dataset, following the augmentation methodology discussed by [Li et al. 2018] and [Deng et al. 2022] for trajectory representation learning. Specifically, for each trajectory in the dataset, the fraction of GPS coordinate pairs subjected to a perturbation of 15, 30 or 50 meters is governed by a distortion rate parameter, randomly sampled from the discrete set $\{0.0, 0.2, 0.4, 0.6, 0.8\}$. This selection aims to reflect heterogeneous signal conditions found in urban environments.

3.1.2. Completeness

To investigate the impact of Completeness pollution on model performance, we focused on the data omission aspect due to its higher prevalence in the literature [Li et al. 2018, Deng et al. 2022]. To avoid random approaches (e.g., Random Down-sampling) as well

[2]GPS.gov

as more complex methods that require map matching (e.g., Road-aware Masking), we implement the Trajectory Trimming strategy. This approach simulates contiguous signal loss, such as abrupt signal drops, by discarding a continuous sequence of GPS points from the extremities of a trip. In addition to this scenario, other approaches will be investigated in future extensions of this work.

Following [Jiang et al. 2023], our methodology applies this operation to either the origin or the destination of the trajectory with equal probability to maintain the travel semantics. This completeness pollution is injected uniformly across the entire dataset. For each trip, the exact length of the omitted trajectory segment is determined by a trimming ratio selected from the discrete set $\{0.3, 0.5, 0.7, 0.9\}$, adapting the masking ratios of [Zhou et al. 2024] to a degradation analysis context.

3.2. Model Training and Evaluation

The objective was to assess model robustness under baseline and pollution scenarios:

- *Baseline:* Standard training and testing using the original data.
- *Scenario 1:* Training on polluted trajectories and testing on the original data.
- *Scenario 2:* Training on the original data and testing on polluted data.

To evaluate the impact of spatial data pollution on TTE, this study investigates three different DNN architectures with distinct inductive biases regarding trajectory geometry. The DNN relies exclusively on static origin-destination and temporal features, serving as a geometry-free baseline. The WDR [Wang et al. 2018b] separates static categorical features from LSTM-based sequential processing, where the Wide and Deep branches encode global origin-destination information independently of perturbations in the intermediate GPS sequence. DeepTTE [Wang et al. 2018a], by jointly modeling path geometry and trip dynamics as its primary representation, is the architecture most exposed to errors in the sequential GPS input.

The DNN and WDR models were trained using a random split of the dataset into training (70%), validation (10%), and test (20%) sets. For DeepTTE, a temporal partitioning strategy was adopted instead — consistent with the approach in the original work [Wang et al. 2018a] — in which, over the one-year dataset, the final two months were reserved for testing, the preceding month for validation, and the remaining data for training. During training, the DNN and WDR models were optimized using Adam for up to 100 epochs, employing an early stopping mechanism with a patience of 15 epochs on the validation loss to prevent overfitting. The DeepTTE model was trained following the default configuration from the original implementation [Wang et al. 2018a]. The performance of these models was evaluated using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE).

4. Results

This section discusses the impact of data pollution on the three evaluated models. Tables 1 and 2 provide the results, with errors denoting the 95% confidence interval half-width.

4.1. Completeness

Scenario 1. The DNN behaves as an unstable model in this scenario, with MAE increasing from 280.74s to 2053.83s and RMSE reaching 2324.49s at 90% pollution. In contrast, DeepTTE exhibits remarkable resilience to completeness degradation. Its MAE increases only from 133.78s in the baseline to 216.08s at 90% pollution, remaining consistently below both DNN and WDR across all degradation levels. This behavior suggests that the model is able to preserve most of its predictive capability even when trained on heavily truncated trajectories. The WDR degrades more sharply: MAE remains near the baseline up to 50% pollution, but increases steeply to 740.12s at 70% and 3609.96s at 90%, suggesting that the Bi-LSTM branch becomes increasingly sensitive to trajectory trimming as sequence length decreases. Since the Wide and Deep branches use only origin-destination features, they cannot compensate for the loss of sequential information.

Scenario 2. The relative stability of DNN is better explained by architectural insensitivity to trajectory structure than by active robustness to completeness degradation. In contrast, WDR relies more on trajectory continuity, and severe trimming removes enough sequential information to degrade the Bi-LSTM predictions sharply. In this scenario, DeepTTE exhibits a more pronounced degradation relative to its own baseline, with MAE rising

Table 1. Impact of completeness degradation on model performance.

Scenario	Metric	Model	Baseline	30%	50%	70%	90%
Scenario 1	MAE (s)	DNN	280.74 ± 1.37	293.71 ± 1.62	405.61 ± 1.62	715.62 ± 2.01	2053.83 ± 3.61
		WDR	243.22 ± 2.02	287.18 ± 2.62	369.90 ± 1.92	740.12 ± 2.31	3609.96 ± 2.95
		DeepTTE	133.78 ± 1.18	197.97 ± 1.71	201.34 ± 1.67	206.24 ± 1.62	216.08 ± 1.73
	RMSE (s)	DNN	561.23 ± 12.12	559.38 ± 11.46	637.38 ± 11.74	927.58 ± 6.75	2324.49 ± 4.94
		WDR	658.05 ± 37.39	808.81 ± 135.49	681.93 ± 26.34	1024.94 ± 31.96	3716.64 ± 4.34
		DeepTTE	356.94 ± 25.83	512.24 ± 25.84	511.16 ± 25.61	504.13 ± 25.68	521.82 ± 27.27
	MAPE (%)	DNN	51.86 ± 0.26	51.19 ± 0.23	72.79 ± 0.26	125.73 ± 0.37	354.37 ± 0.73
		WDR	32.68 ± 0.52	44.15 ± 0.99	70.00 ± 0.50	153.67 ± 0.73	684.59 ± 1.14
		DeepTTE	16.95 ± 0.21	24.55 ± 0.12	26.61 ± 0.13	28.60 ± 0.13	31.82 ± 0.14
Scenario 2	MAE (s)	DNN	280.74 ± 1.37	275.96 ± 1.62	281.18 ± 1.59	303.39 ± 1.73	325.87 ± 1.81
		WDR	243.22 ± 2.02	232.63 ± 1.85	288.67 ± 1.81	456.28 ± 1.73	656.22 ± 1.99
		DeepTTE	133.78 ± 1.18	220.27 ± 1.40	332.43 ± 1.60	439.62 ± 1.95	526.17 ± 2.31
	RMSE (s)	DNN	561.23 ± 12.12	556.38 ± 11.82	563.82 ± 11.19	592.62 ± 11.40	631.87 ± 11.39
		WDR	658.05 ± 37.39	610.67 ± 33.01	610.33 ± 25.82	706.37 ± 26.25	894.05 ± 17.76
		DeepTTE	356.94 ± 25.83	450.89 ± 19.65	566.61 ± 15.71	684.94 ± 15.69	808.81 ± 18.67
	MAPE (%)	DNN	51.86 ± 0.26	49.69 ± 0.33	48.11 ± 0.33	48.77 ± 0.33	49.75 ± 0.32
		WDR	32.68 ± 0.52	28.35 ± 0.57	39.22 ± 0.54	65.63 ± 0.54	90.07 ± 0.55
		DeepTTE	16.95 ± 0.21	26.25 ± 0.21	39.88 ± 0.21	53.00 ± 0.21	63.46 ± 0.24

Table 2. Impact of positional accuracy degradation on model performance.

Scenario	Metric	Model	Baseline	15m	30m	50m
Scenario 1	MAE (s)	DNN	280.74 ± 1.37	260.68 ± 1.62	254.29 ± 1.61	280.39 ± 1.65
		WDR	243.22 ± 2.02	257.81 ± 1.94	244.44 ± 2.06	244.82 ± 1.98
		DeepTTE	133.78 ± 1.18	165.36 ± 1.38	176.62 ± 1.13	153.96 ± 1.30
	RMSE (s)	DNN	561.23 ± 12.12	535.87 ± 11.73	537.53 ± 11.38	561.08 ± 12.36
		WDR	658.05 ± 37.39	633.81 ± 30.04	654.11 ± 32.42	645.33 ± 31.83
		DeepTTE	356.94 ± 25.83	409.89 ± 22.65	379.83 ± 21.87	402.04 ± 26.19
	MAPE (%)	DNN	51.86 ± 0.26	47.90 ± 0.35	45.89 ± 0.35	51.81 ± 0.26
		WDR	32.68 ± 0.52	39.35 ± 0.45	32.81 ± 0.79	33.41 ± 0.72
		DeepTTE	16.95 ± 0.21	19.24 ± 0.09	21.37 ± 0.20	18.89 ± 0.13
Scenario 2	MAE (s)	DNN	280.74 ± 1.37	282.64 ± 1.71	282.05 ± 1.72	282.06 ± 1.66
		WDR	243.22 ± 2.02	242.85 ± 2.04	243.08 ± 1.95	242.85 ± 2.00
		DeepTTE	133.78 ± 1.18	128.41 ± 1.08	168.84 ± 1.56	299.94 ± 2.48
	RMSE (s)	DNN	561.23 ± 12.12	569.22 ± 11.99	569.22 ± 12.67	569.21 ± 12.31
		WDR	658.05 ± 37.39	648.76 ± 35.43	650.34 ± 33.20	648.31 ± 35.48
		DeepTTE	356.94 ± 25.83	340.56 ± 24.61	450.30 ± 20.23	738.64 ± 16.25
	MAPE (%)	DNN	51.86 ± 0.26	52.13 ± 0.33	52.14 ± 0.32	52.14 ± 0.33
		WDR	32.68 ± 0.52	32.67 ± 0.56	32.69 ± 0.62	32.72 ± 0.57
		DeepTTE	16.95 ± 0.21	16.91 ± 0.19	21.62 ± 0.20	36.84 ± 0.23

from 220.27s at 30% to 526.17s at 90% pollution, as the trimmed inputs represent out-of-distribution samples relative to the full trajectories seen during training.

4.2. Positional Accuracy

Scenario 1. In this scenario, across MAE, RMSE, and MAPE, the DNN remains close to its baseline, with only minor variation at the tested noise levels. This is consistent with its dependence on origin-destination distance, where 15 to 50 meters perturbations are small relative to the trip scale. The WDR is also stable; all three metrics remain nearly flat. This is consistent with its hybrid design, in which the Wide and Deep origin-destination branches are weakly affected by positional perturbations, like DNN, while the Gaussian noise in the Bi-LSTM branch does not accumulate enough distortion across the sequence to strongly impact the results. DeepTTE is also relatively stable under training-time positional noise, but it operates at a much lower error level than the other models, remaining the best-performing architecture in the tested range (Table 2).

Scenario 2. The WDR and DNN remain close to their respective baselines across all pollution levels, suggesting that both architectures are relatively robust to positional corruption during training and testing. DeepTTE is more sensitive in this scenario; the three metrics increase as noise reaches 50m, showing that local geometric distortions on testing data affect the trajectory representation learned by the model. Since DeepTTE relies directly on fine-grained path geometry through its GeoConv layers, larger positional perturbations distort local movement patterns and reduce the model's ability to correctly interpret the trajectory during inference. Despite this degradation, DeepTTE still achieves lower MAPE than DNN even at 50m noise. At 15m and 30m, it also maintains lower MAE, RMSE, and MAPE than both DNN and WDR, indicating that its sensitivity to positional corruption emerges only under more severe geometric distortion (Table 2).

5. Conclusions and Future Work

This work analyzed the impact of spatial data quality on DL models for the TTE problem, considering different levels of data pollution on Positional Accuracy and Completeness. To this end, we employ MAE, MAPE, and RMSE metrics under polluted training and test data. Despite its sensitivity to data pollution, DeepTTE outperforms DNN and WDR in terms of prediction accuracy, remaining the preferred model across all evaluated scenarios. In real scenarios, due to its higher accuracy when compared to the other two models, DeepTTE emerges as a reliable choice even in unstable environments, such as in autonomous vehicle systems, where signal blockage caused by tall buildings can affect its precision. The pollution algorithms used in this study are available in a public repository on GitHub[3] and can be used to reproduce the experiments.

Future work includes extending the pollution analysis to embrace all data quality dimensions defined in ISO 19157-1:2023. We also plan to investigate the robustness of models based on other algorithms across additional datasets, broader data pollution scenarios, and predictive contexts beyond TTE. Future experiments will also include multiple runs under identical parameter settings.

The work is supported by the project "IntegraSanca: Conectando Dados, Transformando São Carlos" (Proc. FAPESP 2025/07160-0) and the PIBIC program at UFAL.

[3]https://github.com/Spatial-Data-Quality-for-TTE/spatial-data-quality-TTE

References

Chen (2025). Exploration of intelligent geographic information system development in big data era. *Journal of Computer, Signal, and System Research*, 2(3):24–30.

Deng et al. (2022). Efficient trajectory similarity computation with contrastive learning. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22)*, page 365–374. Association for Computing Machinery.

Hiniduma et al. (2025). Data readiness for AI: A 360-Degree Survey. *ACM Computing Surveys*, 57(9).

Jiang et al. (2023). Self-supervised trajectory representation learning with temporal regularities and travel semantics. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 843–855.

Li et al. (2018). Deep representation learning for trajectory similarity computation. ICDE 2018, pages 617–628.

Minghini et al. (2019). Openstreetmap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1):1–17.

Mohammed et al. (2025). The effects of data quality on machine learning performance on tabular data. *Information Systems*, 132:102549.

Si et al. (2023). Data quality analysis and improvement: A case study of a bus transportation system. *Applied Sciences*, 13(19).

Silva et al. (2025). QualiBus: implementando métricas de qualidade para dados de transporte coletivo. SBBD 2025, pages 589–601. SBC.

Tavra et al. (2026). Determining the data quality-influencing characteristics of participants in volunteered geographic information campaigns – the case of maritime disasters. *Geomatics, Natural Hazards and Risk*, 17(1):2659907.

Vitagliano et al. (2023). Pollock: A data loading benchmark. *VLDB*, 16(8):1870–1882.

Wang et al. (2018a). When will you arrive? estimating travel time based on deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Wang et al. (2018b). Learning to estimate the travel time. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*, page 858–866. Association for Computing Machinery.

Zheng et al. (2023). Data-driven methods for travel time estimation: A survey. ITSC 2023, pages 1292–1299. ISSN: 2153-0017.

Zhou et al. (2024). Red: Effective trajectory representation learning with comprehensive information. *Proc. VLDB Endow.*, 18(2):80–92.

Zuo et al. (2026). TriDGNet: Triple feature encoder-based dual granularity graph learning network for enhanced travel time estimation. *IEEE Transactions on Intelligent Transportation Systems*, 27(1):1606–1620.