Quality-aware wireless sensor networks using deep-learning for real-time urban planning

Carrow Morris-Wiltshire*[1], Stuart Barr[2], Phil James[3] and Tom Komar[4]

[1]Newcastle University

January 14, 2025

Summary

The emergence of smart city initiatives has led to the widespread deployment of Internet of Things (IoT) sensors for monitoring urban dynamics. While these sensors generate vast amounts of data, the development of real-time data quality assessment frameworks remains a significant challenge. This research presents an initial prototype library for developing quality-aware pedestrian data stream pipelines, tested on data from the Newcastle Urban Observatory. The library currently implements machine learning capabilities for data quality assessment and prediction, with planned extensions for data interpretation and automated decision-making functionality. This work provides a foundation for understanding the relationship between data quality and predictive performance in urban sensor networks.
KEYWORDS: Data Quality, Internet of Things, Urban Computing, Machine Learning, Time Series Analysis

1 Background and Motivation

Urban observatories worldwide are collecting unprecedented volumes of mobility data through networks of sensors (Smith and Turner 2019; Rusli et al. 2023). Data quality dimensions provide a framework for assessing the fitness-for-use of data across multiple attributes (R. Y. Wang and Strong 1996). These dimensions include accuracy (how well the data represents reality), completeness (the presence of all necessary values), timeliness (how current the data is), and consistency (the degree to which data follows defined rules) (Karkouch et al. 2016; Mansouri et al. 2023). For pedestrian monitoring systems, these dimensions are particularly critical - accuracy affects the reliability of crowd measurements, completeness impacts the ability to detect patterns, timeliness determines the feasibility of real-time responses, and consistency ensures reliable trend analysis (Chen et al. 2021). Current cities typically have limited decision-making functionality due to slow, incomplete, and largely disconnected information transmission systems (Barr et al. 2020). While improving spatial resolution with massive low-cost sensor networks is possible, it comes with increased data quality challenges as cheaper sensors are more prone to errors and inconsistencies (Karkouch et al. 2016).

*c.morris-wiltshire@ncl.ac.uk

Understanding and monitoring these quality dimensions becomes essential for developing reliable urban monitoring systems (Teh, Kempa-Liehr, and K. I.-K. Wang 2020).

2 Methodology

Our approach integrates statistical analysis with deep learning techniques to assess and monitor data quality in real-time sensor networks. The methodology consists of three interconnected components that form a pipeline for processing and analysing sensor data streams. First, a data quality assessment framework evaluates multiple dimensions including completeness, sequence length, temporal consistency, and signal periodicity - moving beyond simple metrics to understand how different aspects of data quality affect model performance. Second, a feature engineering pipeline processes the raw sensor data, using techniques like Lomb-Scargle Periodograms (Press and Rybicki 1989; VanderPlas 2018) to handle the irregular temporal nature of IoT data streams and extract meaningful patterns. Finally, a deep learning architecture uses LSTM (Hochreiter and Schmidhuber 2025; Gers, Schmidhuber, and Cummins 2000) units to capture both short-term fluctuations and long-term dependencies in pedestrian movement patterns, with the goal of making reliable predictions even with imperfect data quality and more perhaps more importantly, understanding the limits on data quality for effective prediction. This integrated approach enables both the assessment of data quality impacts on predictive performance and the development of models that can adapt to varying levels of data quality across different sensor types and generations.

2.1 Data Quality Assessment and Preprocessing

The data preprocessing stage focuses on data cleaning and quality assessment before any feature engineering takes place. While deep learning models can handle missing data, these models tend to be more complex and tuning them can be challenging. Another option is using complex imputation methods to fill in missing data, but these are often computationally expensive and can affect the integrity of the data. For these reasons, the data has been left in its raw form, and a consecutive sequence detection and labelling algorithm has been developed.

The sequence detection algorithm works by scanning through a pandas DataFrame row by row, building sequences where the time difference between consecutive rows equals the minimum time delta (normally 15-minutes). When there's a gap in the timestamps or when starting fresh, it checks if the current sequence meets a minimum length requirement. Valid sequences are stored and numbered sequentially.

2.2 Feature Engineering

The feature engineering stage encapsulates functions that create additional features used to train the model. Frequency features are calculated using a Lomb-Scargle Periodogram, which is similar to a Fourier Transform but applicable to data with missing records. This approach is particularly important given the irregular nature of our sensor data streams.

The features are then normalised using the Standard Scaler equation:

$$X_i' = \frac{X_i - \mu}{\sigma} \tag{1}$$

where X_i is the original value, μ is the mean of the feature, and σ is the standard deviation of the feature. The standard scalar normalisation technique is chosen to ensure that any future data points falling outside of the existing range can be scaled appropriately. Calculating the mean and standard deviation in the scaling process also serves as a check for future data drift.

2.3 Model Architecture and Training

For recurrent neural networks, such as LSTMs, we employ a fixed-size window approach for data loading. Each window contains multiple sequences of data: the primary sensor readings and additional engineered features. Each window is associated with a label which represents a future data point that the model aims to predict. Two key hyper-parameters define the window structure: the window size, which determines how many historical time steps are included in each window, and the horizon, which specifies how far into the future the model predicts.

The model architecture incorporates LSTM units to capture long-term dependencies in the sequential data. These architectures have proven highly effective at handling sequential data processing tasks. A range of hyperparameters are tuned including window sizes (4-48 timesteps), batch sizes (32-128), hidden units (32-256), number of layers (1-3), learning rate (1e-5 to 1e-1), and dropout rate (0-0.2). Early stopping with a patience of 5 epochs is implemented to prevent overfitting. Model performance is evaluated using three standard metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). Full documentation for the project's python library can be found here.

3 Results

Our analysis of pedestrian sensor data from urban observatories revealed several key insights into the relationship between data quality and model performance. In examining data completeness, we used the 12 sensors with the most data, with completeness values ranging from 0.75 to 0.85, and corresponding RMSE values between 0.35 and 1.0 (Figure 1). Counter-intuitively, sensors with lower completeness values demonstrated better model performance, a finding that led to deeper investigation of temporal data patterns.

The impact of sequence length emerged as a critical factor in model performance. Our analysis revealed that optimal performance was achieved with average sequence lengths of at least 35 hours, allowing models to capture full daily cycles of pedestrian activity. Notably, shorter sequences of approximately 20 hours showed higher RMSE values despite having higher completeness metrics, highlighting the importance of continuous data streams in capturing urban mobility patterns (Figure 2).

Comparing model architectures in Table 3, we found that multivariate (MV) models demonstrated increasingly superior performance over univariate (UV) models as prediction horizons extended.

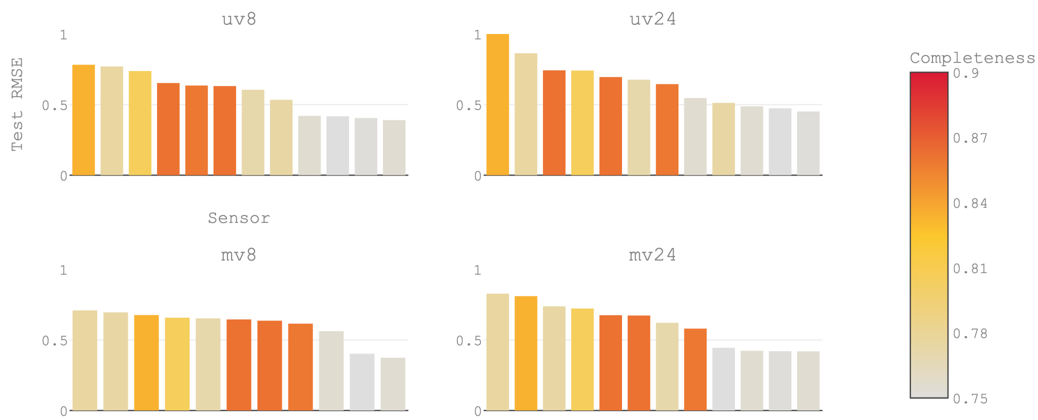

Figure 1: RSME vs Data Completeness

Table 1 show mostly positive values in the diff column highlighting the superior performance of the smaller univariate models, whereas Table 2 show mostly negative values suggesting longer predictions benefit from larger models which is unsurprising. The optimal configuration emerged as an LSTM architecture with 64 hidden units and 2 layers. Through extensive hyperparameter tuning, we determined that batch sizes of 32 and window sizes of 24 time steps consistently yielded the best results across different sensor locations and time periods. This configuration balanced computational efficiency with predictive accuracy, making it suitable for real-time applications.

4 Discussion

The research reveals critical insights about the relationship between data quality and prediction reliability in urban monitoring systems. Our findings demonstrate that for longer-term predictions (6-hours), multivariate models that combine harmonic analysis with deep learning marginally outperform simpler approaches, but this superior performance is contingent on specific data quality requirements. Specifically, we found that effective long-term prediction requires both high completeness values (>0.80) and substantial average sequence lengths (>35 hours). Initially counterintuitive, our finding that sensors with lower overall completeness yielded better model performance highlights how single-dimension quality metrics can be misleading. Newer sensors in the network, designed with a feature that halts data recording when no pedestrians are detected, demonstrated this limitation. While this design choice marginally reduced database related expenses, it inadvertently created an ambiguity in the data stream - making it impossible to differentiate between sensor failures and genuine zero-pedestrian states, and reducing the average data sequence length

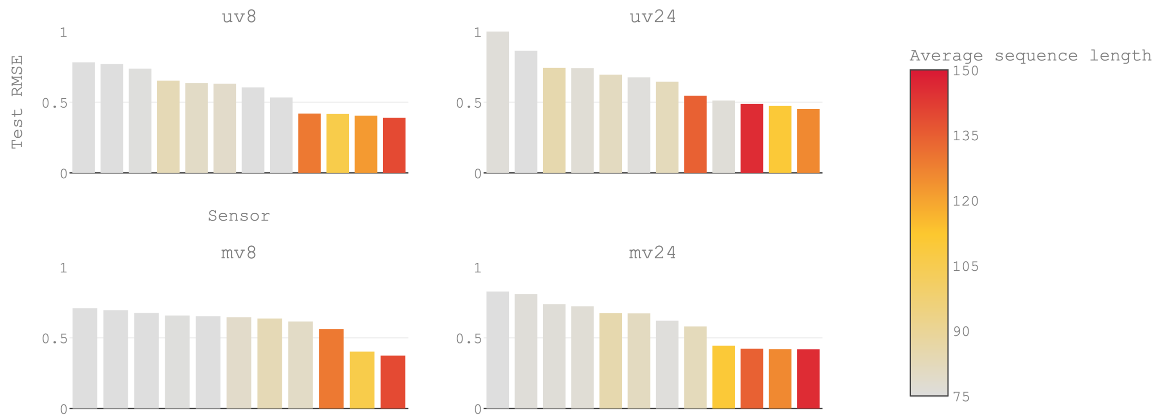

Figure 2: RSME vs Average Sequence Length

Table 1: Comparison of test metrics for UV/MV models (8 timesteps (2-hour) horizon)

Test MAPE	**diff**	**Test RMSE**	**diff**	**Test MAE**	**diff**
2.148/3.308	1.160	0.623/0.656	0.033	0.484/0.515	0.031
1.632/1.922	0.290	0.752/0.683	-0.069	0.604/0.542	-0.063
1.755/2.047	0.292	0.701/0.680	-0.021	0.528/0.515	-0.013
1.650/2.317	0.667	0.632/0.683	0.051	0.485/0.549	0.064
1.332/1.509	0.177	0.639/0.639	0.000	0.478/0.474	-0.003
16.16/13.558	-2.607	0.384/0.387	0.003	0.282/0.283	0.002
1.123/1.174	0.051	0.380/0.373	-0.007	0.270/0.275	0.005
2.221/2.504	0.283	0.405/0.464	0.059	0.304/0.366	0.062
0.865/0.755	-0.110	0.414/0.397	-0.017	0.297/0.310	0.012
1.466/1.869	0.403	0.603/0.614	0.011	0.458/0.479	0.021
1.446/2.289	0.843	0.528/0.605	0.077	0.407/0.476	0.068
1.986/2.212	0.226	0.748/0.681	-0.067	0.600/0.531	-0.069

Table 2: Comparison of test metrics for UV/MV models (24 timesteps (6-hour) horizon)

Test MAPE	**diff**	**Test RMSE**	**diff**	**Test MAE**	**diff**
1.623/1.945	0.322	0.696/0.671	-0.025	0.546/0.535	-0.011
2.790/3.648	0.858	0.928/0.830	-0.098	0.760/0.680	-0.081
2.463/2.687	0.224	0.680/0.675	-0.005	0.509/0.520	0.011
1.431/1.442	0.011	0.665/0.610	-0.055	0.506/0.465	-0.040
15.748/15.547	-0.201	0.629/0.602	-0.027	0.470/0.448	-0.021
1.653/1.270	-0.383	0.492/0.442	-0.050	0.369/0.371	0.001
1.324/1.180	-0.144	0.450/0.356	-0.094	0.315/0.271	-0.044
1.321/0.980	-0.341	0.533/0.431	-0.102	0.400/0.359	-0.041
0.827/0.930	0.103	0.437/0.356	-0.081	0.313/0.277	-0.035
17.913/22.985	5.072	0.593/0.594	0.001	0.465/0.479	0.015
1.646/2.990	1.344	0.513/0.667	0.154	0.390/0.547	0.156
3.008/4.392	1.384	0.861/0.848	-0.013	0.681/0.668	-0.013

Table 3: Comparison of hyperparameters across experiments

Parameter	**UV-24**	**MV-24**	**UV-8**	**MV-8**
window_size	24	24	24	24
horizon	24	24	8	8
stride	1	1	1	1
batch_size	32	32	32	32
model_type	lstm	lstm	lstm	lstm
hidden_dim	256	64	64	64
num_layers	2	2	1	2
dropout	0.064	0.053	0.072	0.173
learning_rate	0.00011	0.00024	0.00382	0.00012
scheduler_step_size	7	4	4	3
scheduler_gamma	0.518	0.564	0.440	0.802

to under 21 hours. The low average sequence length significantly impacted model performance, and the ambiguity between true periods of inactivity and sensor failures made interpolation of missing data challenging.

In contrast, older sensors, despite experiencing more extended periods of downtime due to maintenance requirements and connectivity issues, provided more consistent blocks of complete data when operational. These uninterrupted sequences, though less frequent, better captured the natural periodicities in pedestrian movement patterns, particularly around the crucial 24-hour cycle identified in our temporal analysis. This finding underscores an important lesson in urban sensing: higher completeness metrics alone do not guarantee better data utility. The temporal structure of data gaps and the ability to interpret them meaningfully are equally, if not more, important. Our research demonstrates that only through measuring multiple dimensions of data quality - including completeness, sequence length, temporal consistency, and signal periodicity - can we fully understand the suitability of data for predictive modelling and achieve reliable long-term predictions.

5 Conclusions and Future Work

The development of quality-aware networks that can adapt to and account for different types of data gaps, while maintaining interpretability, will be essential for building reliable smart city systems. This research offers a framework for implementing such systems and highlights the importance of considering data quality as a multidimensional characteristic. This research provides a foundation for developing quality-aware urban data stream processing systems. The findings contribute to the broader field of urban computing by providing practical tools and methodologies for handling the quality challenges inherent in large-scale sensor networks. Most significantly, this work establishes the critical relationship between data quality dimensions and predictive performance in urban sensor networks, demonstrating that reliable long-term predictions require both sophisticated modelling approaches and high-quality data streams with specific characteristics. The library provides a pathway toward more reliable and scalable urban monitoring systems, essential for developing effective smart city management solutions.

6 Acknowledgements

The authors would like to thank the Newcastle Urban Observatory for providing access to their pedestrian monitoring data infrastructure. Special thanks to the Centre for Doctoral Training (CDT) in Geospatial Systems at Newcastle University for their academic support and guidance. We also acknowledge the valuable feedback received from colleagues in the School of Engineering and fellow CDT researchers.

References

Barr, S. L. et al. (Sept. 2020). "FLOOD-PREPARED: A NOWCASTING SYSTEM FOR REAL-TIME IMPACT ADAPTION TO SURFACE WATER FLOODING IN CITIES". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* VI-4-W2-

2020, pp. 9–15. ISSN: 2194-9042. DOI: 10.5194/isprs-annals-VI-4-W2-2020-9-2020. (Visited on 01/07/2025).

Chen, L. et al. (July 2021). "Estimating Vehicle and Pedestrian Activity from Town and City Traffic Cameras". In: *Sensors (Basel)* 21.13, p. 4564. ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). DOI: 10.3390/s21134564.

Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins (Oct. 2000). "Learning to Forget: Continual Prediction with LSTM". In: *Neural Computation* 12.10, pp. 2451–2471. ISSN: 0899-7667. DOI: 10.1162/089976600300015015. (Visited on 01/08/2025).

Hochreiter, S and Jürgen Schmidhuber (2025). *Long Short-Term Memory | MIT Press Journals & Magazine | IEEE Xplore*. https://ieeexplore.ieee.org/abstract/document/6795963. (Visited on 01/08/2025).

Karkouch, Aimad et al. (2016). "Data Quality in Internet of Things: A State-of-the-Art Survey". In: *Journal of Network and Computer Applications* 73, pp. 57–81. ISSN: 1084-8045.

Mansouri, Taha et al. (2023). "IoT Data Quality Issues and Potential Solutions: A Literature Review". In: *The Computer Journal* 66.3, pp. 615–625. ISSN: 0010-4620.

Press, W. H and G. B. Rybicki (1989). "Fast Algorithm for Spectral Analysis of Unevenly Sampled Data". In: *Fast algorithm for spectral analysis of unevenly sampled data* 338.1, pp. 277–280. ISSN: 0004-637X.

Rusli, Noradila et al. (2023). "A Review on Worldwide Urban Observatory Systems' Data Analytics Themes: Lessons Learned for Malaysia Urban Observatory (MUO)". In: *Journal of Urban Management*. ISSN: 2226-5856.

Smith, Luke and Mark Turner (2019). "Building the Urban Observatory: Engineering the Largest Set of Publicly Available Real-Time Environmental Urban Data in the UK". In: *Geophysical Research Abstracts*. Vol. 21. ISBN: 1029-7006.

Teh, Hui Yie, Andreas W. Kempa-Liehr, and Kevin I-Kai Wang (Feb. 2020). "Sensor Data Quality: A Systematic Review". In: *Journal of Big Data* 7.1, p. 11. ISSN: 2196-1115. DOI: 10.1186/s40537-020-0285-1. (Visited on 06/17/2024).

VanderPlas, Jacob T. (2018). "Understanding the Lomb–Scargle Periodogram". In: *The Astrophysical Journal Supplement Series* 236.1, p. 16. ISSN: 0067-0049.

Wang, Richard Y. and Diane M. Strong (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers". In: *Journal of management information systems* 12.4, pp. 5–33. ISSN: 0742-1222.

Biographies

Carrow Morris-Wiltshire is a PhD student at the Geospatial Systems CDT in the School of Engineering at Newcastle University. His research interests include the application of machine learning and deep learning to sensor data for complex urban systems.

Stuart Barr is professor of Geospatial Systems Engineering in the School of Engineering Newcastle University UK.

Phil James is Professor of Urban Data at Newcastle University and leads the Urban Observatory Programme.

Tom Komar is a Machine Learning Operations Engineer at Urban Observatory Newcastle University.