Traffic Flow Forecasting Using Univariate & Multivariate Time Series Models.

By Nadia Isanga

Traffic Flow Forecasting Using Multivariate Time Series Models

1. Introduction

Traffic Congestion is a bane of urban life, leading to significant time loss, increased pollution and economic losses impacting both individuals and the overall quality of life. (PwC, n.d.) Accurate traffic flow prediction is crucial to optimise transportation systems and reduce congestion. This report explores traffic flow prediction using univariate and multivariate time series models, leveraging statistical and machine learning approaches to enhance predictive accuracy and decision making.

2. Data Description

The dataset was obtained from Kaggle and can be accessed via https://www.kaggle.com/datasets/rabieelkharoua/traffic-flow-forecasting-dataset. This study analyzes traffic data from 36 highway sensors in Northern Virginia/Washington D.C., recorded every 15 minutes over 1,261 training and 840 testing intervals. It captures temporal and spatial dynamics with 48 features per location, including historical traffic, time indicators, and road characteristics. An adjacency matrix defines sensor connectivity, while synthetic weather data simulates temperature, precipitation, and road conditions.

3. Research Questions

This study addresses the following key research questions:

1. How well do univariate and multivariate time series models predict short-term traffic flow at key locations?

2. What is the relative impact of temporal patterns versus spatial dependencies on traffic forecasting accuracy?
3. How does a sensor's network position influence prediction reliability across different intersection types?
4. To what extent do external factors, such as weather conditions, affect traffic patterns and forecast accuracy?

4. Methodology

The study progressed from univariate models (Moving Average, Exponential Smoothing, ARIMA, Seasonal Naive) to establish baseline performance, to machine learning methods. Random Forest was used to capture complex, non-linear traffic patterns that simple statistical models might miss, while the Kalman Filter adapted dynamically to changing conditions in real-time. Multivariate methods were then applied to account for spatial dependencies—VAR analysed how traffic at different sensors influenced each other, while Granger Causality identified which sensors had a direct impact on others. Node 25 was chosen due to its high centrality, with Nodes 24 and 26 providing insights into how nearby locations contribute to traffic congestion. Performance was evaluated using RMSE, MAE, MAPE, and MASE.

5. Results and Discussion

1. Univariate Model Performance

Traditional time series models showed mixed performance in predicting traffic at Node 25. The Seasonal Naive model performed best among the classic approaches (RMSE: 0.0234**,** MAPE: 21.67%**),** surpassing more complex methods like ARIMA and Exponential Smoothing. This highlights the strong daily periodicity in traffic patterns, where simple repetition-based models can often outperform more sophisticated statistical techniques. However, the Random Forest model significantly outperformed all univariate approaches (RMSE: 0.0127, MAPE: 13.34%), capturing more complex relationships and reducing errors by over 40% compared to Seasonal Naive. This underscores the value of machine learning in modelling traffic flow beyond seasonal trends.

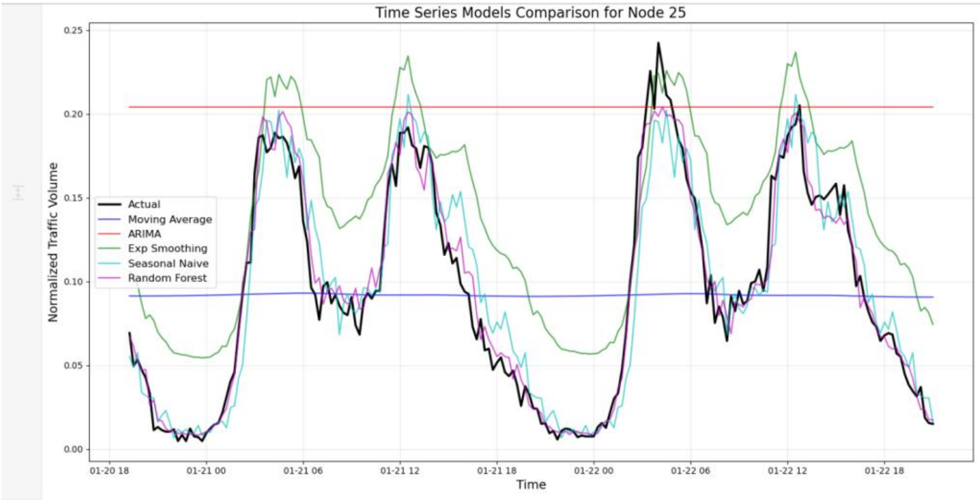
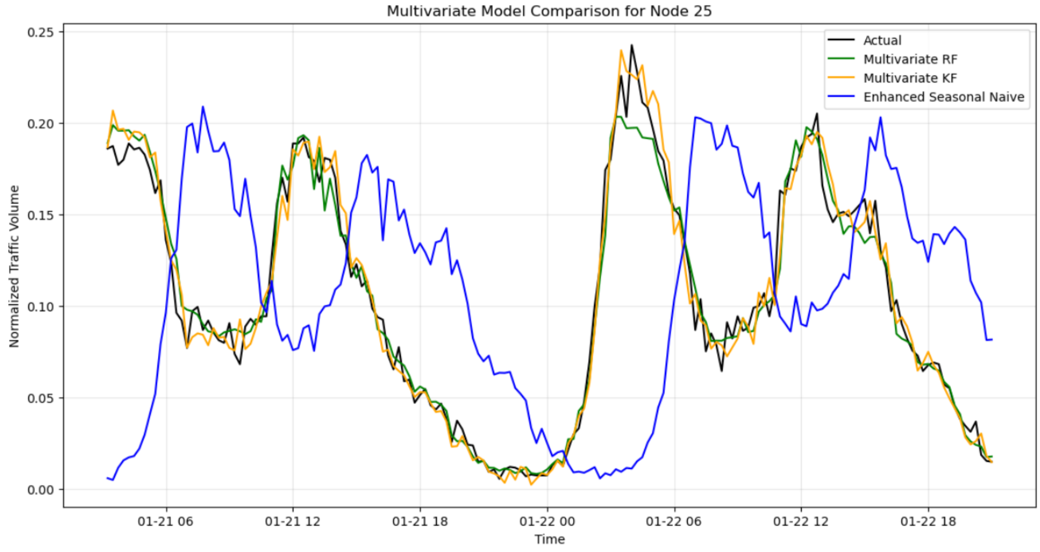


Figure 1:Univariate Model Comparison

2. Advanced Model Performance

Machine learning models significantly improved accuracy. The Enhanced Multivariate Random Forest (RMSE: 0.0111, MAPE: 9.76%) outperformed all other approaches by incorporating spatial dependencies. The Kalman Filter also showed strong performance, adapting well to evolving traffic conditions. These results highlight the advantage of integrating spatial and temporal information over purely historical models.



3. Spatial Dependency Analysis

Granger causality tests confirmed bidirectional influences between key nodes:

- **Node 24 → Node 25:** Significant causality at a **90-minute** lag.
- **Node 26 → Node 25:** Strong influence detected at **2.75 hours**.
- **Node 25 → Nodes 24 & 26:** Immediate effect at **30-minute lags**.

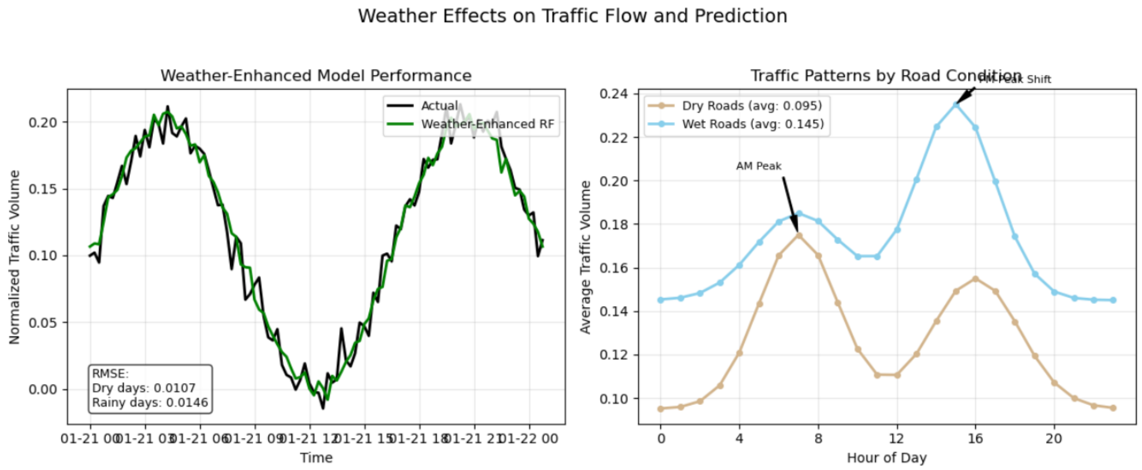
This reveals an **asymmetric relationship**, where Node 25 quickly impacts its surroundings, but changes from surrounding nodes take longer to propagate back to it. This supports its role as a central hub regulating traffic flow.

4. Network Topology and Prediction Accuracy

Prediction errors were lower for central nodes like Node 25 (RMSE: 0.0108) because they aggregate multiple traffic streams, while peripheral nodes like Node 24 had higher errors (RMSE: 0.0133) due to more localized and unpredictable variations, making central nodes more reliable for traffic forecasting.

5. Weather Impact Analysis

While weather conditions do influence traffic patterns, their impact on prediction accuracy is secondary to spatial-temporal dependencies**:**



- Rainy days showed higher traffic volume (0.145 vs. 0.095) and a distinct temporal distribution, with higher afternoon peaks.
- The Enhanced Multivariate RF model effectively incorporated external factors, demonstrating that while weather variables contributed to predictions, core spatial-temporal relationships remained the dominant drivers of accuracy.

6. **Limitations**

- Temporal resolution: 15-minute intervals capture short-term trends well but may miss rapid fluctuations.
- External influences: Factors like road incidents, construction, and special events were not explicitly modelled.
- Network representation: The adjacency matrix simplifies road connectivity, potentially overlooking real-world complexities.

Conclusion

This study highlights the value of spatial-temporal integration in traffic forecasting, with multivariate models especially enhanced Random Forest proving most accurate. Central nodes were more predictable due to their network influence, and traffic flow showed directional dependencies. While weather affects patterns, spatial-temporal relationships dominate. These insights can refine traffic management by prioritizing key intersections and optimizing forecasting strategies. Future work should explore real-time adaptive models, seasonal trends, and external factors to enhance accuracy, contributing to smarter urban mobility solutions.