QUBAHAN JOURNAL OF SOCIAL SCIENCE
VOL. 3, NO. 1, March 2026
https://doi.org/10.48161/qjss.v3n1a13

Applications of Stochastic Processes in Modeling Uncertainty and Random Variation: Evidence from Hourly Traffic Flow Dynamics

Chalang Suleiman Hussein[1*], Chyavan Mohammed Abdulrahman[2]

[1] Accounting Techniques Department, College of Administrative Techniques, Duhok Polytechnic University, Duhok, Iraq;
[2] Sport Science, Physical Education and Sport Science, University of Duhok, Duhok, Iraq;

* **Corresponding author:** chalangsleman@gmail.com.

ABSTRACT: Stochastic processes provide a rigorous probability-based foundation for analyzing systems that evolve over time under uncertainty and random variation. This study develops and evaluates an empirical stochastic-process framework for modeling hourly traffic-flow dynamics using a real multi-junction traffic dataset. The framework integrates Markov-chain state transitions, high-traffic event modeling, random-variation analysis, and Monte Carlo simulation to examine how traffic conditions transition among low, moderate, high, and congested states. The dataset contains 48,120 hourly observations from four traffic junctions, with no missing values or duplicate records. We defined traffic states using junction-specific calibration quartiles, and we reserved the final month of observations for out-of-sample evaluation. The results show strong temporal dependence in traffic dynamics. Lag-one autocorrelation ranged from 0.630 to 0.963, while weekly autocorrelation remained high for Junctions 1 and 2, indicating persistent operational cycles. Markov transition matrices revealed substantial state persistence, particularly in low and congested states, while upward and downward transitions captured short-run changes in traffic intensity. One-step stochastic forecasts achieved an average state-classification accuracy of 68.72%, with an average 95% prediction-interval coverage of 89.10%. High-traffic event counts exhibited overdispersion across all junctions, supporting the use of negative-binomial robustness alongside Poisson event-rate estimation. Monte Carlo simulation generated interpretable weekly traffic scenarios and uncertainty intervals for mean, peak, and total vehicle counts. The study contributes a transparent empirical framework that connects classical stochastic-process theory with operational traffic monitoring and uncertainty-aware decision support.

Keywords: stochastic processes; uncertainty modeling; random variation; Markov chains; traffic flow dynamics.

I. INTRODUCTION

Modern transportation systems operate under continuous uncertainty. Traffic volume changes by hour, day, location, and operational state, and these changes are rarely deterministic. Even when daily and weekly patterns repeat, short-run changes in vehicle counts are still affected by random arrivals, local congestion, driver behavior, and temporal dependence. For this reason, methods that treat observed traffic as a dynamic stochastic process help explain how current conditions evolve into future states.

Stochastic processes offer a coherent mathematical language for studying systems that change over time according to probabilistic rules. Rather than treating uncertainty as residual error around a deterministic model, stochastic-process methods place uncertainty at the center of the analytical structure. Markov chains represent transitions between discrete states, Poisson-type processes model random event occurrences,

random-walk logic describes cumulative variation, and simulation procedures translate transition and event models into probabilistic future scenarios.

The present study applies these ideas to hourly traffic-flow data collected from four junctions. The empirical setting is appropriate because vehicle counts are sequential, non-negative, temporally dependent, and subject to sudden changes. A purely average-based analysis would summarize traffic levels but would not explain the probability of movement from low to high congestion, the frequency of high-traffic events, or the uncertainty surrounding near-future traffic scenarios. A stochastic-process framework can address these issues while remaining interpretable for operational decision-making.

This study aims to develop and evaluate a structured stochastic-process framework for modeling uncertainty and random variation in hourly traffic-flow dynamics. The paper estimates traffic-state transition matrices, examines counts of high-traffic events, evaluates random hourly variation, and generates Monte Carlo scenarios for future traffic behavior. The study is an empirical, probability-based analysis rather than a purely theoretical review. This makes the contribution relevant for transportation analytics, operations research, and risk-aware monitoring systems.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature on stochastic processes, traffic-flow uncertainty, Markov modeling, count-event modeling, and probabilistic forecasting. Section 3 identifies the research gap and study contribution. Section 4 presents the proposed stochastic-process framework. Section 5 explains the dataset, model design, equations, and evaluation procedure. Section 6 reports the results. Section 7 discusses the findings and integrates the theoretical contributions, practical implications, limitations, and future research directions. Section 8 concludes the study.

II. LITERATURE REVIEW

Stochastic-process theory is widely treated as the mathematical foundation for analyzing systems that evolve through random changes over time. Classical formulations explain how indexed random variables, Markov chains, birth–death processes, and diffusion models represent dynamic uncertainty, which matters for hourly traffic analysis because vehicle counts occur sequentially rather than as isolated observations [1]. Large-scale traffic research has applied Markov modeling to simulate movement across road networks, showing that probabilistic state transitions can reproduce complex traffic behavior, though this work focuses more on network-wide simulation than on junction-specific hourly state evaluation [2].

Queueing theory provides another important foundation by examining systems characterized by random arrivals, limited service capacity, waiting times, and congestion. This perspective directly relates to transportation because vehicle arrivals can be interpreted as stochastic demand on road infrastructure with constrained capacity [3]. Markovian event modeling in other fields also shows how sequential random events can estimate the distribution of possible system outcomes, supporting the broader use of transition-based probability methods even though the original application differs from traffic-flow analysis [4].

Count-data modeling connects stochastic-process theory with the empirical analysis of events occurring within fixed time intervals. Poisson models are appropriate when the event-count variance is close to the mean, whereas negative-binomial models are preferable when counts are overdispersed, which is particularly relevant to clustered high-traffic events [5]. Bayesian traffic-flow inversion provides a more advanced approach by combining observed traffic information with prior distributions and computational sampling, although its complexity may reduce operational transparency when only basic hourly vehicle-count data are available [6].

General stochastic-process theory distinguishes among Markov chains, Poisson processes, renewal processes, martingales, and other dynamic probability models. This distinction supports using separate analytical components for traffic-state transitions, event occurrence, random variation, and simulation, rather than applying one universal model to all forms of uncertainty [7]. Stochastic network link-transmission modeling extends this reasoning to connected road systems by representing uncertain vehicle movements, traffic propagation, and queue distributions, although it requires more detailed network information than junction-level hourly counts [8].

Probabilistic forecasting emphasizes assessing uncertainty through predictive distributions rather than point forecasts alone. Calibration and sharpness matter because a useful traffic forecast should indicate both the likely outcome and the plausible range of future vehicle counts [9]. Negative-binomial modeling

complements this forecasting perspective by allowing event-count variance to exceed the mean, making it suitable for high-traffic periods that occur in clusters rather than at a stable homogeneous rate [10].

Applied time-series forecasting literature stresses the importance of preserving temporal ordering, using holdout periods, examining seasonality, and reporting transparent error measures. These principles support using a chronologically later test period because random data splitting would disrupt the hourly, daily, and weekly dependence in traffic observations [11]. Stochastic traffic-flow theory further shows that traffic models can incorporate uncertainty directly rather than treating it only as independent residual noise, providing a more realistic representation of evolving traffic density and flow [12].

Gaussian approximations simplify the estimation of complex stochastic traffic-flow models while retaining probabilistic information about system variation. This approach improves analytical tractability but is more technically demanding than direct Markov transition estimation from observed traffic states [13]. Probabilistic stationary speed–density modeling similarly shows that traffic relationships should not always be represented by one deterministic curve because driver behavior and operational conditions can produce a distribution of possible traffic outcomes [14].

Brownian motion and stochastic calculus provide a theoretical basis for continuous random variation, cumulative uncertainty, and diffusion over time. Although hourly vehicle counts are discrete, these concepts remain useful for interpreting how small short-run changes can accumulate into uncertain traffic trajectories [15]. Collective traffic forecasting shows that observations across multiple traffic locations and congestion states may share predictive patterns, though it primarily emphasizes forecast performance over interpretable state-transition probabilities [16].

Diffusion-based traffic-network modeling applies continuous stochastic processes to the design and control of road systems under uncertainty. This work is important for understanding network-level variation, but it requires richer traffic and infrastructure information than the basic count variables used in the present analysis [17]. Stochastic fundamental-diagram modeling incorporates uncertainty into relationships among traffic flow, density, and vehicle classes, confirming that traffic heterogeneity should be treated as a structural characteristic rather than as simple measurement noise [18].

Probabilistic forecast communication research emphasizes that decision-makers need understandable information about the range and likelihood of future outcomes. This principle supports the reporting of prediction intervals and simulated weekly traffic scenarios rather than presenting only single deterministic estimates [19]. Applied probability modeling also shows how Markov chains, Poisson processes, queueing models, renewal processes, and simulation can be combined based on a system's empirical characteristics, supporting an integrated stochastic framework [20].

Gaussian-process approximation has been used to evaluate stochastic traffic-flow models efficiently while preserving estimates of predictive uncertainty. This method offers considerable flexibility but may be less transparent to operational users than transition matrices and state-specific simulation outputs [21]. The stochastic cell-transmission model incorporates uncertainty directly into traffic-state surveillance and assignment by representing probabilistic traffic movement across connected road cells, although it requires more detailed road-segment data than hourly junction counts [22].

General stochastic-modeling theory emphasizes that model selection should correspond to the structure of the observed process. Discrete states, event counts, cumulative changes, and continuous fluctuations represent different forms of uncertainty and therefore require distinct but complementary analytical tools [23]. Research on stochastic Poisson intensities further shows that event rates may change over time rather than remain constant, which is relevant to traffic systems because high-volume events are often concentrated within particular hours or operating periods [24].

Physics-regularized Gaussian-process modeling combines empirical flexibility with established traffic-flow principles. This approach can produce uncertainty-aware predictions while remaining consistent with transportation theory, though it generally requires richer data and more complex estimation than a transparent junction-level monitoring framework [25]. Stochastic Lagrangian traffic-state estimation models vehicle movement under driver heterogeneity and uncertain traffic dynamics, but its reliance on detailed trajectory and state information limits its direct applicability to datasets containing only timestamps, junction identifiers, and vehicle counts [26].

Stochastic Lagrangian dynamics have also been used to examine stop-and-go oscillations and the propagation of traffic disturbances. These findings show that small random changes can develop into

persistent traffic waves, supporting analyses of state persistence and unusually large hourly traffic variations [27]. Probabilistic load-forecasting research shows how predictive distributions and regularized estimation apply to highly seasonal demand systems, offering transferable methodological support because electricity demand and traffic flow both exhibit hourly dependence, repeated cycles, and uncertain peaks [28].

Overall, the literature provides strong foundations for stochastic-process theory, traffic-flow modeling, count-event analysis, probabilistic forecasting, and simulation. However, most previous studies concentrate on one component of uncertainty, such as state transitions, network diffusion, event counts, traffic assignment, or predictive distributions. This study addresses this gap by integrating Markov traffic-state transitions, overdispersed high-traffic event counts, random hourly variation, out-of-sample prediction intervals, and Monte Carlo scenarios into a single, transparent framework based on multi-junction hourly traffic data.

Table 1. Literature-review comparison table.

Ref.	Study	Focus	Main contribution	Relevance/gap for this study
1	Allen (2010)	Stochastic-process theory	Explained core stochastic-process models and biological applications.	Provides general theoretical foundation but not traffic-specific empirical testing.
2	Besenczi et al. (2021)	Markov traffic simulation	Used Markov modeling for large-scale road-network traffic simulation.	Focuses on network simulation rather than hourly count-state evaluation.
3	Bhat (2015)	Queueing theory	Presented queueing models for arrivals, service, and system delay.	Useful for arrival logic but not designed for multi-junction traffic counts.
4	Black and Ross (2015)	Stochastic epidemic modeling	Demonstrated computation of final-size distributions in Markovian models.	Shows stochastic event logic outside transportation.
5	Cameron and Trivedi (2013)	Count-data modeling	Explained Poisson and negative-binomial approaches for event counts.	Does not address traffic-state transition dynamics.
6	Coullon and Pokern (2022)	Bayesian traffic-flow inversion	Applied MCMC to motorway traffic-flow parameter uncertainty.	Methodologically advanced but less transparent for simple operational monitoring.
7	Durrett (2016)	Stochastic-process foundations	Covered Markov chains, Poisson processes, renewal processes, and martingales.	Theoretical foundation without empirical traffic application.
8	Flötteröd and Osorio (2017)	Network link transmission	Modeled vehicular networks through stochastic link transmission and queue distributions.	Network-level focus; not centered on observed hourly state transitions.
9	Gneiting and Katzfuss (2014)	Probabilistic forecasting	Clarified calibration and predictive distributions.	Forecasting-focused rather than traffic-process-specific.
10	Hilbe (2011)	Negative-binomial modeling	Explained overdispersion and negative-binomial regression for count data.	Supports robustness but not a complete stochastic-process framework.
11	Hyndman and Athanasopoulos (2018)	Forecast evaluation	Presented practical time-series forecasting and evaluation principles.	Focuses on forecasting methods rather than stochastic state processes.
12	Jabari and Liu (2012)	Stochastic traffic-flow theory	Developed theoretical foundations for stochastic traffic-flow modeling.	Theoretical emphasis; limited direct empirical state-transition framework.
13	Jabari and Liu (2013)	Gaussian approximation	Derived Gaussian approximation and estimation for stochastic traffic-flow models.	More complex than transparent applied monitoring designs.
14	Jabari et al. (2014)	Speed-density relation	Proposed a probabilistic stationary speed-density relation.	Focuses on the traffic-flow relation, not event-rate and simulation integration.

Ref.	Study	Focus	Main contribution	Relevance/gap for this study
15	Karatzas and Shreve (2012)	Brownian motion	Provided formal background for Brownian motion and stochastic calculus.	Important for diffusion theory but not ideal for count traffic data alone.
16	Lippi et al. (2010)	Traffic forecasting	Studied collective traffic forecasting across congestion states.	Forecasting-focused with less emphasis on interpretable stochastic processes.
17	Mandjes and Storm (2020)	Diffusion traffic model	Formulated stochastic road-network design and control with diffusion logic.	Focuses on network design rather than empirical multi-model comparison.
18	Ngoduy (2011)	Stochastic fundamental diagrams	Used stochastic fundamental diagrams in multiclass traffic modeling.	Does not combine Markov, count-event, and Monte Carlo components.
19	Raftery (2016)	Probabilistic forecast communication	Explained the decision value of calibrated probabilistic forecasts.	General forecast communication rather than traffic-state modeling.
20	Ross (2014)	Probability models	Presented applied probability models, including Markov chains and Poisson processes.	General applied foundation without empirical traffic validation.
21	Storm et al. (2022)	Gaussian-process approximation	Evaluated stochastic traffic-flow models using Gaussian-process approximation.	Advanced approximation approach, less focused on simple traffic-state interpretability.
22	Sumalee et al. (2011)	Stochastic cell transmission	Introduced a stochastic dynamic traffic model for surveillance and assignment.	Freeway-oriented and model-theoretic rather than simple hourly traffic-count application.
23	Taylor and Karlin (2014)	Stochastic modeling	Presented core stochastic-modeling concepts for applied systems.	General reference without traffic-specific empirical testing.
24	Wang et al. (2020)	Stochastic Poisson intensities	Estimated stochastic Poisson intensities for high-intensity traffic/service traces.	Relevant to event rates but not traffic-state transitions.
25	Yuan et al. (2021)	Gaussian-process traffic modeling	Integrated physics regularization with Gaussian processes for traffic modeling.	High predictive focus but limited operational transparency.
26	Zheng et al. (2018)	Stochastic Lagrangian dynamics	Modeled traffic-state estimation under driver heterogeneity.	Requires richer traffic-state information than the present count dataset.
27	Zheng et al. (2019)	Traffic oscillation	Modeled stop-and-go traffic using stochastic Lagrangian dynamics.	Targets oscillation mitigation rather than multi-junction traffic-count uncertainty.
28	Ziel and Liu (2016)	Probabilistic forecasting	Used probabilistic forecasting for uncertain electricity demand.	Different domain but useful for interval-based evaluation.

The Table summarizes the main theoretical and empirical studies used to position the present framework. The review shows that prior studies provide strong stochastic foundations, but fewer studies combine Markov transition modeling, high-event count analysis, random-variation summaries, and simulation in a simple hourly traffic-count framework.

III. RESEARCH GAP

The reviewed literature confirms the importance of stochastic-process methods for dynamic systems, traffic-flow modeling, probabilistic forecasting, and count-event analysis. However, three gaps remain. First, many stochastic traffic-flow studies are theoretically sophisticated but require dense traffic-state variables, detailed road-network structure, or specialized calibration settings. Second, applied count-based traffic studies often estimate arrival or event rates without linking them to state-transition behavior. Third, few empirical studies present a compact, interpretable framework that integrates state transitions, event counts, random variation, and simulation using a simple multi-junction hourly dataset.

The present study addresses these gaps by developing a framework implementable with ordinary traffic-count observations. It does not require detailed vehicle trajectories, road geometry, or weather covariates. Instead, it converts hourly vehicle counts into operational states, estimates Markov transition probabilities, evaluates high-traffic event counts, summarizes random variation, and generates Monte Carlo scenarios.

This makes the framework suitable for contexts where basic traffic monitoring data are available but advanced sensor infrastructure is limited.

IV. STUDY CONTRIBUTION

The study's contribution is fourfold. First, it provides an empirical stochastic-process framework for modeling uncertainty and random variation in hourly traffic-flow dynamics. Second, it estimates traffic-state transition matrices that describe how junctions move between low, moderate, high, and congested conditions. Third, it evaluates high-traffic event counts and demonstrates why overdispersion checks are necessary before relying on a simple Poisson assumption. Fourth, it uses Monte Carlo simulation to produce practical uncertainty intervals for weekly mean, peak, and total traffic scenarios.

V. PROPOSED STOCHASTIC-PROCESS FRAMEWORK

The proposed framework treats traffic volume as a discrete-time stochastic process observed hourly. Let X_t represent the vehicle count at hour t for a given junction. Because the raw count is non-negative, temporally ordered, and highly variable, the framework analyzes the process through four connected layers: state transition, event occurrence, random variation, and simulation.

In the first layer, traffic counts are transformed into four junction-specific states: low, moderate, high, and congested. This transformation improves comparability because each junction has a different traffic-volume scale. In the second layer, we estimate a Markov transition matrix from the calibration period. In the third layer, we define high-traffic events as those above the 90th percentile of the calibration distribution and count them by day. In the fourth layer, random hourly changes and simulation scenarios are used to evaluate uncertainty in future traffic behavior.

X_t = vehicle count observed at hour t for a given junction (1)

$S_t \in \{\text{Low, Moderate, High, Congested}\}$ (2)

$p_{ij} = \Pr(S_{t+1} = j \mid S_t = i)$ (3)

$$N_d = \sum_{t \in d} I\,(X_t \geq q_{0.90}) \quad (4)$$

$$\Delta X_t = X_t - X_{t-1} \quad (5)$$

$$\hat{y}_{t+1} = \sum_{j} p_{ij}\,\mu_j, \quad \mu_j = \textit{calibration mean of state } j \quad (6)$$

VI. METHODOLOGY

1. RESEARCH DESIGN

The study uses an empirical quantitative design based on sequential traffic-count observations. The analysis does not estimate causal effects; instead, it models stochastic dependence, event frequency, and uncertainty in dynamic traffic behavior. The design follows a calibration-test structure: observations before June 2017 are used for model calibration, and observations from June 2017 are used for out-of-sample evaluation.

The empirical unit is the junction-hour. Each record represents the number of vehicles observed at a specific junction during one hour. The analysis is conducted separately by junction because traffic scale and volatility differ across locations. Aggregated summaries are used only when discussing average framework performance.

2. DATASET DESCRIPTION AND PREPROCESSING

The dataset contains hourly vehicle counts from four traffic junctions. It includes four original variables: DateTime, Junction, Vehicles, and ID. We converted DateTime into a time index and derived hour, day of week, date, and month for descriptive analysis. The dataset contains 48,120 observations, no missing values, and no duplicate rows. Junctions 1–3 cover 1 November 2015 to 30 June 2017, while Junction 4 covers 1 January 2017 to 30 June 2017.

Because the four junctions have different traffic-volume levels, we estimated state thresholds separately for each junction using the calibration period. The first, second, and third quartiles define the boundaries between low, moderate, high, and congested states. This avoids classifying a normal value at a low-volume junction as low simply because it is smaller than values at a high-volume junction.

Table 2. Dataset audit and descriptive statistics by junction.

Junction	Start	End	Observations	Mean	SD	Minimum	Median	Maximum
1	2015-11-01	2017-06-30	14592	45.05	23.01	5	40.00	156
2	2015-11-01	2017-06-30	14592	14.25	7.40	1	13.00	48
3	2015-11-01	2017-06-30	14592	13.69	10.44	1	11.00	180
4	2017-01-01	2017-06-30	4344	7.25	3.52	1	7.00	36

Table 2 shows that the dataset is complete and suitable for stochastic-process analysis. Junction 1 has the largest average traffic volume, while Junction 4 has shorter time coverage and lower mean vehicle counts.

Table 3. Junction-specific traffic-state thresholds.

Junction	Q1	Q2	Q3
1	27.00	39.00	57.00
2	9.00	12.00	17.00
3	7.00	11.00	17.00
4	5.00	7.00	9.00

Table 3. reports the calibration-period quartile thresholds used to classify hourly vehicle counts into traffic states. The thresholds differ by junction, which preserves within-junction comparability.

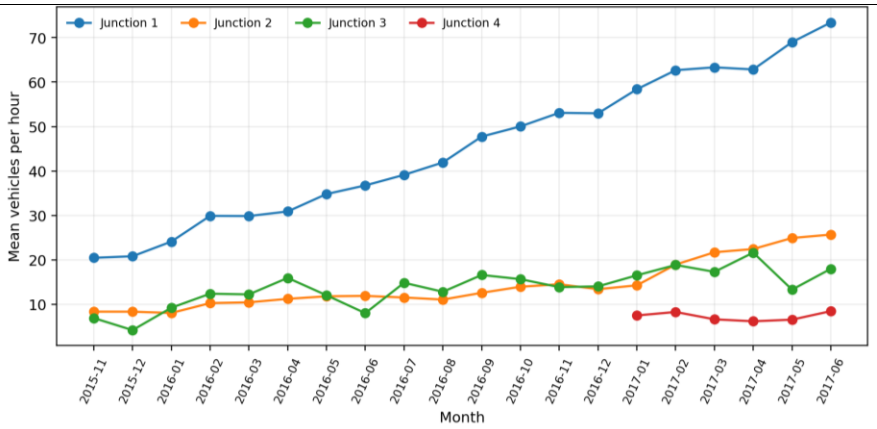


FIGURE 1. Monthly mean traffic volume by junction.

FIGURE 2 shows the monthly variation in mean traffic volume. Junction 1 has the highest average traffic level and a visible upward movement over time, while Junctions 2–4 operate at lower average counts.

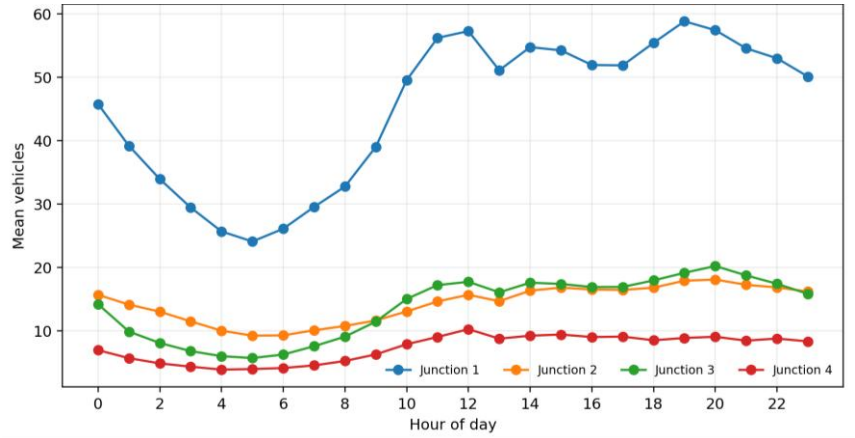


FIGURE 2. Average hourly traffic pattern by junction.

FIGURE 3. shows the intra-day traffic pattern. Hourly seasonality is evident, supporting the treatment of traffic as a temporally dependent stochastic process rather than independent observations.

3. ANALYTICAL PROCEDURE

- Step 1: Audit the dataset for missing values, duplicate records, date ranges, and junction-specific coverage.
- Step 2: Split the data into calibration and test periods using June 2017 as the out-of-sample evaluation period.
- Step 3: Define four traffic states using junction-specific calibration quartiles.
- Step 4: Estimate Markov transition matrices for each junction using sequential state changes in the calibration period.
- Step 5: Define high-traffic events using the junction-specific 90th percentile and evaluate daily event-count dispersion.
- Step 6: Summarize random hourly variation using first differences, absolute changes, and autocorrelation at 1-hour, 24-hour, and 168-hour lags.
- Step 7: Generate one-step stochastic forecasts and Monte Carlo weekly scenarios using transition probabilities and empirical state-specific emissions.

4. EVALUATION MEASURES

The framework is evaluated using both state and count-based indicators. State accuracy measures whether the most probable next state matches the observed next state. MAE, RMSE, and MAPE evaluate one-step vehicle-count forecasts generated from the transition matrix and state-specific means. Prediction-interval coverage evaluates whether observed vehicle counts fall within the simulated 95% interval. The dispersion index evaluates whether daily high-traffic events follow a Poisson-like variance structure or require an overdispersion-robust interpretation.

$$\text{MAE} = n^{-1}\sum_{t=1}^{n}|y_t - \hat{y}_t| \qquad (7)$$

$$\text{RMSE} = \left[n^{-1}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2\right]^{1/2} \qquad (8)$$

$$\text{Coverage}_{0.95} = n^{-1}\sum_{t=1}^{n} I\,(L_t \le y_t \le U_t) \qquad (9)$$

$$\text{Dispersion index} = \frac{\text{Var}(N_d)}{E(N_d)} \qquad (10)$$

5. VALIDITY, RELIABILITY, AND ETHICS

Temporal splitting, junction-specific state construction, and separate evaluation of state accuracy, point error, interval coverage, event dispersion, and simulation uncertainty support validity. Reliability is supported by reproducible rules for state classification and transition estimation. The dataset is publicly available and contains aggregated traffic counts, not personally identifiable information. Therefore, the analysis involves no human-subject risk, privacy disclosure, or individual-level tracking.

VII. RESULTS

1. MARKOV TRANSITION STRUCTURE

The transition matrices show that traffic states are persistent but not static. Junction 1 has strong persistence in the low and congested states, with probabilities of 0.865 and 0.875, respectively. This indicates that once Junction 1 enters a low or congested condition, it is likely to remain in the same state during the next hour. Junctions 2 and 3 also show meaningful persistence, while Junction 4 exhibits more frequent state movement because its lower traffic scale and shorter observation period produce less stable state patterns.

Table 4. Markov transition probability matrix for Junction 1.

State	Low	Moderate	High	Congested
Low	0.87	0.13	0.00	0.00
Moderate	0.15	0.69	0.16	0.00
High	0.00	0.16	0.72	0.12
Congested	0.00	0.00	0.12	0.88

Table 4. illustrates the transition structure for Junction 1. The diagonal probabilities are high, showing strong state persistence, while adjacent-state transitions capture gradual movement between traffic conditions.

Table 5. State persistence and transition direction by junction.

Junction	Persistence (%)	Upward (%)	Downward (%)
1	78.90	10.50	10.60
2	63.30	18.37	18.33
3	64.85	17.57	17.58
4	44.44	27.27	28.29

Table 5. summarizes the share of same-state, upward, and downward transitions. Junction 1 has the most persistent state behavior, while Junction 4 has the highest share of downward transitions.

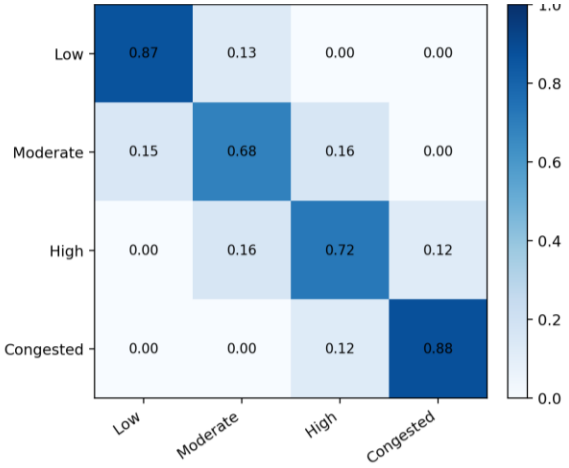


FIGURE 3. Markov transition probability matrix for Junction 1.

FIGURE 4. visualizes the transition matrix for Junction 1. The darker diagonal cells show that low and congested traffic states tend to persist across consecutive hours.

2. ONE-STEP STOCHASTIC FORECAST PERFORMANCE

The one-step evaluation uses June 2017 as the test period. Across the four junctions, the framework achieves an average MAE of 8.03 vehicles and an average RMSE of 11.22 vehicles. The average state-classification accuracy is 68.72%, indicating that a simple Markov-chain structure captures much of the short-run state dependence. Average 95% prediction-interval coverage is 89.10%. Coverage is below the nominal level for Junctions 1 and 2 but high for Junctions 3 and 4, suggesting that interval calibration can be improved for higher-volume junctions.

Table 6. Out-of-sample one-step forecast evaluation.

Junction	MAE	RMSE	MAPE (%)	State acc. (%)	PI95 cov. (%)	PI95 width
1	17.26	22.17	21.02	82.22	81.53	56.55
2	7.91	10.15	27.32	76.81	83.33	22.80
3	4.35	8.76	24.73	68.47	97.78	30.10
4	2.61	3.78	37.89	47.36	93.75	11.32

Table 6 reports forecast accuracy and interval performance for June 2017. The framework performs best in absolute-error terms for Junctions 3 and 4 and produces the widest intervals for Junction 1 because it has the largest traffic volume and volatility.

3. HIGH-TRAFFIC EVENT-COUNT MODELING

High-traffic events were defined using the 90th percentile of each junction's calibration distribution. The daily high-event counts are overdispersed for all junctions because the variance exceeds the mean. This means a simple homogeneous Poisson process can provide an interpretable baseline event rate, but negative-binomial or quasi-Poisson models are more appropriate for modeling uncertainty in event frequency.

Table 7. High-traffic event-rate and overdispersion results.

Junction	Threshold	Mean/day	Variance	Dispersion	Max/day	Event hours (%)
1	76.00	2.93	26.69	9.12	17.00	12.19
2	22.00	3.22	36.97	11.48	21.00	13.41
3	24.00	2.90	19.76	6.82	20.00	12.07

Junction	Threshold	Mean/day	Variance	Dispersion	Max/day	Event hours (%)
4	11.00	3.52	13.34	3.79	16.00	14.66

Table 7. shows that daily high-traffic events are not purely Poisson-like. Dispersion indices greater than one indicate clustered high-traffic periods and justify an overdispersion-aware interpretation.

4. RANDOM HOURLY VARIATION AND AUTOCORRELATION

Random variation was evaluated using hourly first differences. The mean hourly change is close to zero at all junctions because increases and decreases offset each other over the full period. However, the standard deviation and upper-tail absolute changes reveal substantial short-run volatility, particularly at Junctions 1 and 3. Junction 3 has a maximum absolute hourly change of 137 vehicles, indicating rare but substantial shifts in observed volume.

Autocorrelation results confirm strong temporal dependence. Junction 1 has a lag-one autocorrelation of 0.963, a 24-hour autocorrelation of 0.855, and a 168-hour autocorrelation of 0.949. Junction 2 also has strong daily and weekly dependence. Junctions 3 and 4 show weaker but still meaningful dependence, especially at the 1-hour lag. These patterns support Markov- and simulation-based modeling rather than independent-count assumptions.

Table 8. Random hourly traffic variation by junction.

Junction	Mean Δ	SD Δ	Mean |Δ|	P75 |Δ|	P95 |Δ|	Max |Δ|
1	0.00	6.26	4.60	6.00	12.00	63.00
2	0.00	2.97	2.31	3.00	6.00	19.00
3	0.00	5.84	3.32	4.00	9.00	137.00
4	0.00	3.03	2.30	3.00	6.00	27.00

Table 8 summarizes first-difference behavior. Although average change is near zero, the standard deviation and upper-tail absolute changes show substantial random variation.

Table 9. Autocorrelation of hourly traffic counts.

Junction	Lag (h)	Autocorrelation
1	1	0.96
1	24	0.85
1	168	0.95
2	1	0.92
2	24	0.81
2	168	0.89
3	1	0.84
3	24	0.58
3	168	0.45
4	1	0.63
4	24	0.47
4	168	0.45

Table 9 shows that traffic observations are strongly autocorrelated, especially at short and weekly lags. This temporal dependence supports stochastic-process modeling.

5. MONTE CARLO SIMULATION SCENARIOS

Monte Carlo simulation used the final observed traffic state of each junction and generated 3,000 simulated 168-hour paths. For each junction, the simulation produced distributions of weekly mean, peak, and total traffic. Junction 1 has the highest simulated weekly mean and peak, with a weekly mean interval from 32.12 to 60.35 vehicles per hour and a peak interval from 90.00 to 143.00 vehicles. Junction 4 has the narrowest weekly mean interval because of its lower traffic scale.

Table 10. Monte Carlo 168-hour simulation scenarios.

Junction	Metric	Mean	2.5%	97.5%
1	Weekly_mean	45.59	32.12	60.35
1	Weekly_peak	111.35	90.00	143.00
1	Weekly_total	7659.03	5395.90	10138.53
2	Weekly_mean	13.97	11.19	17.17
2	Weekly_peak	38.22	31.00	44.00
2	Weekly_total	2347.23	1879.98	2884.02
3	Weekly_mean	13.84	9.96	18.30
3	Weekly_peak	75.86	36.98	134.98
3	Weekly_total	2324.55	1672.98	3074.00
4	Weekly_mean	7.05	6.23	7.92
4	Weekly_peak	20.25	14.00	36.00
4	Weekly_total	1184.34	1047.00	1331.00

Table 10. reports simulated weekly mean, peak, and total traffic scenarios. These intervals translate stochastic transition behavior into practical uncertainty summaries for decision support.

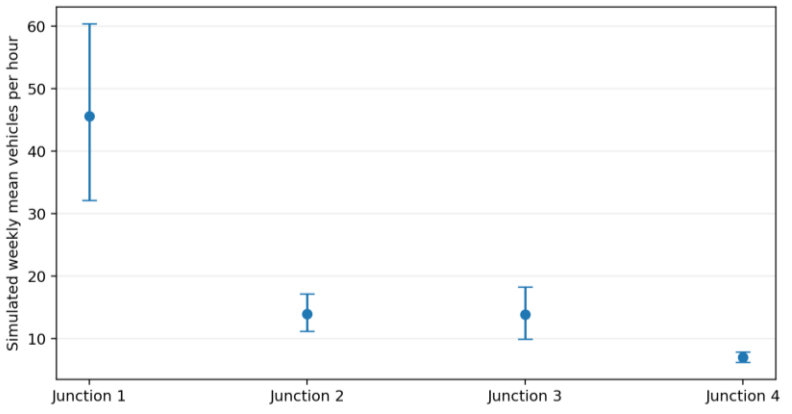


FIGURE 4. Monte Carlo 95% intervals for weekly mean traffic.

FIGURE 5. compares simulated weekly mean traffic across junctions. Junction 1 has the highest expected traffic and the widest uncertainty interval, while Junction 4 has the lowest and narrowest simulated range.

VIII. DISCUSSION

The findings show that stochastic processes provide a useful, interpretable foundation for modeling uncertainty and random variation in traffic-flow systems. The Markov-chain results show that traffic conditions are state-dependent: the next-hour state is strongly influenced by the current state. This is especially evident at Junction 1, where low and congested states persist. This persistence matters operationally because congestion, once formed, may persist without intervention, while low-volume periods also tend to remain stable over short horizons.

The event-count analysis adds a second layer of interpretation. High-traffic events are not evenly scattered in time; they cluster and are overdispersed. This means the Poisson process is a useful baseline model, but a strict Poisson assumption is too restrictive for the observed data. The negative-binomial interpretation is more realistic because it allows the event-count variance to exceed the mean. For empirical traffic studies, this distinction is essential because underestimating variance can lead to overconfident risk assessments and overly narrow operational thresholds.

The random-variation results show that average hourly change alone is not informative. Because increases and decreases cancel out over time, the mean difference is close to zero, but the standard deviation and high-percentile absolute changes reveal important volatility. For example, Junction 3 has a rare but very large hourly jump. This supports the study's use of random-walk logic and simulation: dynamic systems may appear stable in the mean while still producing large short-run deviations.

The autocorrelation analysis confirms that hourly traffic counts are not independent. Strong lag-one dependence indicates short-run continuity, while daily and weekly autocorrelation reveal repeated cycles. This temporal dependence justifies a stochastic-process framework and also explains why simple independent Poisson assumptions cannot fully describe the data. Traffic flow is both patterned and random: it follows recurring cycles, but within those cycles it remains subject to uncertainty.

The out-of-sample forecast results are encouraging for a simple, transparent model. Average state accuracy of 68.72% indicates that the Markov-chain structure captures meaningful traffic-state dynamics. The prediction intervals provide useful uncertainty ranges, although the average coverage of 89.10% is below the nominal 95% target. This suggests that future work could improve interval calibration by using hour-specific transition matrices, day-type-specific models, or overdispersion-adjusted emission distributions.

The Monte Carlo simulation results show the practical value of stochastic processes for decision support. Rather than producing a single deterministic forecast, the framework generates distributions for weekly mean, peak, and total traffic. This helps traffic managers because planning often requires understanding best-case, typical, and high-pressure scenarios. The simulation intervals are especially informative for Junction 1, where traffic volume and uncertainty are both high.

The results show how classical stochastic-process models can be combined in a unified empirical framework. Markov chains describe traffic-state transitions, Poisson-type event logic provides a baseline for high-traffic occurrences, random-walk analysis captures hourly variation, and Monte Carlo simulation translates estimated stochastic behavior into future uncertainty scenarios. This integrated design shows that stochastic-process theory can be operationalized with ordinary traffic-count data while remaining interpretable.

The discussion also advances traffic-flow modeling by emphasizing state-level uncertainty rather than point forecasting alone. Instead of treating vehicle counts as isolated observations, the analysis models traffic as a sequential process with persistent states, clustered events, and temporally dependent variation. This helps connect applied traffic analytics with probability theory and operations research.

From a practical perspective, the framework provides traffic-monitoring agencies with a transparent method for early warning and congestion monitoring. Transition matrices can identify whether a junction is likely to remain congested, move toward congestion, or return to a lower state. High-event counts can support threshold-based monitoring and resource allocation. Random-variation summaries can identify junctions with unstable hourly changes. Simulation intervals can support planning for peak-load scenarios, staffing, signal timing, and risk-aware operational decisions.

The framework is also useful in data-limited settings. Because it requires only timestamp, junction, and vehicle count variables, it can be implemented where weather data, road geometry, vehicle trajectories, or sensor-fusion systems are unavailable. This makes the approach suitable for municipalities and transport organizations with basic monitoring infrastructure.

Interpreting these findings should also account for several limitations. First, the dataset includes traffic counts but does not include weather, road conditions, accidents, holidays, signal timing, or lane capacity. These missing covariates may explain part of the unexplained variation. Second, Junction 4 has a shorter time period than the other junctions, which may reduce transition stability. Third, the Markov model assumes that the next state depends primarily on the current state; future studies could test higher-order or time-varying Markov models. Fourth, we use a Poisson process as a baseline event model, but observed overdispersion suggests considering negative-binomial and self-exciting event models in future work.

Future research can improve the framework by estimating hour-specific and weekday-weekend transition matrices, adding weather and incident covariates, comparing Markov models with hidden Markov models, calibrating prediction intervals to achieve nominal coverage, and applying the framework to other dynamic systems such as energy demand, hospital arrivals, equipment-failure events, and network traffic.

IX. CONCLUSION

This study developed an empirical stochastic-process framework for modeling uncertainty and random variation in hourly traffic-flow dynamics. Using 48,120 observations from four traffic junctions, the paper showed that traffic counts can be represented through state transitions, high-traffic event counts, random hourly variation, and simulation-based future scenarios. The Markov transition results revealed strong state

persistence, particularly for low and congested traffic conditions. The event-count results showed that high-traffic events are overdispersed, so a negative binomial model is preferable to a strict Poisson assumption. The random-variation and autocorrelation results confirmed that traffic flow is temporally dependent and subject to both recurring cycles and sudden short-run changes. Out-of-sample evaluation produced meaningful state prediction accuracy and practical prediction intervals, while Monte Carlo simulation translated stochastic dynamics into interpretable weekly mean, peak, and total-traffic scenarios. Overall, the findings support the value of stochastic processes as transparent tools for modeling dynamic uncertainty. The framework is especially useful when only basic traffic-count data are available, because it does not depend on complex sensor fusion or detailed road-network geometry. The study therefore provides a practical probability-based contribution for traffic monitoring, operations research, and uncertainty-aware decision support. Future studies can extend the framework by incorporating weather, incidents, time-varying transition matrices, and interval-calibration procedures.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1]. Allen, L. J. S. (2010). An introduction to stochastic processes with applications to biology (2nd ed.). Chapman and Hall/CRC.
[2]. Besenczi, R., Bátfai, N., Jeszenszky, P., Major, R., Monori, F., & Ispány, M. (2021). Large-scale simulation of traffic flow using a Markov model. PLOS ONE, 16(2), e0246062.
[3]. Bhat, U. N. (2015). An introduction to queueing theory: Modeling and analysis in applications. Birkhäuser.
[4]. Black, A. J., & Ross, J. V. (2015). Computation of epidemic final size distributions. Journal of Theoretical Biology, 367, 159–165.
[5]. Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (2nd ed.). Cambridge University Press.
[6]. Coullon, J., & Pokern, Y. (2022). Markov chain Monte Carlo for a hyperbolic Bayesian inverse problem in traffic flow modeling. Data-Centric Engineering, 3, e4.
[7]. Durrett, R. (2016). Essentials of stochastic processes (3rd ed.). Springer.
[8]. Flötteröd, G., & Osorio, C. (2017). Stochastic network link transmission model. Transportation Research Part B: Methodological, 102, 180–209.
[9]. Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. Annual Review of Statistics and Its Application, 1, 125–151.
[10]. Hilbe, J. M. (2011). Negative binomial regression (2nd ed.). Cambridge University Press.
[11]. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts.
[12]. Jabari, S. E., & Liu, H. X. (2012). A stochastic model of traffic flow: Theoretical foundations. Transportation Research Part B: Methodological, 46(1), 156–174.
[13]. Jabari, S. E., & Liu, H. X. (2013). A stochastic model of traffic flow: Gaussian approximation and estimation. Transportation Research Part B: Methodological, 47, 15–41.
[14]. Jabari, S. E., Zheng, J., & Liu, H. X. (2014). A probabilistic stationary speed-density relation based on Newell's simplified car-following model. Transportation Research Part B: Methodological, 68, 205–223.
[15]. Karatzas, I., & Shreve, S. E. (2012). Brownian motion and stochastic calculus (2nd ed.). Springer.
[16]. Lippi, M., Bertini, M., & Frasconi, P. (2010). Collective traffic forecasting. In Knowledge Discovery in Databases (pp. 259–273). Springer.

[17]. Mandjes, M. R. H., & Storm, P. J. (2020). Road traffic network design and control using a multi-class diffusion model. INFORMS Transportation Science and Logistics Conference Proceedings.
[18]. Ngoduy, D. (2011). Multiclass first-order traffic model using stochastic fundamental diagrams. Transportmetrica, 7(2), 111–125.
[19]. Raftery, A. E. (2016). Use and communication of probabilistic forecasts. Statistical Analysis and Data Mining, 9(6), 397–410.
[20]. Ross, S. M. (2014). Introduction to probability models (11th ed.). Academic Press.
[21]. Storm, P. J., Mandjes, M. R. H., & van Arem, B. (2022). Efficient evaluation of stochastic traffic flow models using Gaussian process approximation. Transportation Research Part B: Methodological, 164, 126–144.
[22]. Sumalee, A., Zhong, R. X., Pan, T. L., & Szeto, W. Y. (2011). Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. Transportation Research Part B: Methodological, 45(3), 507–533.
[23]. Taylor, H. M., & Karlin, S. (2014). An introduction to stochastic modeling (4th ed.). Academic Press.
[24]. Wang, R., Jaiwal, P., & Honnappa, H. (2020). Estimating stochastic Poisson intensities using deep latent models. arXiv preprint arXiv:2007.06037.
[25]. Yuan, Y., Yang, X. T., Zhang, Z., & Zhe, S. (2021). Macroscopic traffic flow modeling with physics-regularized Gaussian process: A new insight into data-driven applications in transportation. Transportation Research Part B: Methodological, 146, 88–110.
[26]. Zheng, F., Jabari, S. E., Liu, H. X., & Lin, D. (2018). Traffic state estimation using stochastic Lagrangian dynamics. Transportation Research Part B: Methodological, 115, 143–165.
[27]. Zheng, F., Lu, L., Li, R., Liu, X., & Tang, Y. (2019). Traffic oscillation using stochastic Lagrangian dynamics: Simulation and mitigation via control of autonomous vehicles. Transportation Research Record, 2673(4), 529–539.
[28]. Ziel, F., & Liu, B. (2016). Lasso estimation for GEFCom2014 probabilistic electric load forecasting. International Journal of Forecasting, 32(3), 1029–1037.

Appendix A. Additional transition matrices

Table A1. Markov transition probability matrix for junction 2.

State	Low	Moderate	High	Congested
Low	0.73	0.23	0.04	0.00
Moderate	0.33	0.41	0.25	0.01
High	0.04	0.20	0.57	0.18
Congested	0.00	0.01	0.21	0.77

Table A1 provides the junction-specific transition matrix used for state prediction and simulation.

Table A2. Markov transition probability matrix for junction 3.

State	Low	Moderate	High	Congested
Low	0.76	0.21	0.03	0.00
Moderate	0.28	0.46	0.24	0.02
High	0.04	0.22	0.55	0.19
Congested	0.00	0.02	0.19	0.79

Table A2 provides the junction-specific transition matrix used for state prediction and simulation.

Table A3. Markov transition probability matrix for junction 4.

State	Low	Moderate	High	Congested
Low	0.59	0.28	0.09	0.05
Moderate	0.37	0.31	0.19	0.13
High	0.19	0.30	0.28	0.24
Congested	0.06	0.15	0.26	0.53

Table A3 provides the junction-specific transition matrix used for state prediction and simulation.