Where are those trucks going? Using a Bayesian framework to synthesize freight origin-destination data in the Canadian Prairie Region

Nasim Deljouyi, M.Sc.
PhD Student, Civil Engineering Department, University of Manitoba. Obtained her M.Sc. in Electrical Engineering at Sharif University.

Babak Mehran, PhD, P.Eng.
Associate Professor, Civil Engineering Department, University of Manitoba. Obtained his PhD at Nagoya University.

Jonathan D Regehr, PhD, P.Eng.
Associate Professor, Civil Engineering Department, University of Manitoba. Obtained his PhD at University of Manitoba.

Abstract
Planning and designing road networks suitable for truck operations require a comprehensive knowledge of their spatial and temporal patterns between freight origins and destinations (ODs). However, efficient road transportation planning and logistics are constrained by the limited availability of truck OD data. This issue stems from the competitive nature and privacy concerns of the trucking industry, making the process of collecting and labeling truck trips, a time-consuming and resource-intensive task. Given these challenges, this study focuses on generating synthetic truck OD trip data through Bayesian analysis, with the goal of enhancing predictive accuracy in transportation planning of heavy vehicles. The reliability of the synthesized data is evaluated by comparing with simulation-based data generated from TFlowFuzzy method in PTV Visum. Utilizing the Canadian Freight Analysis Framework (CFAF) database, which provides estimates of Canadian freight flows, we conduct a comparative analysis of the results obtained from these two approaches.

Keywords: Freight Transport, Road Network Resilience, Freight Origin-Destination, Synthetic Data Generation, Bayesian Inference, TFlowFuzzy.

1. Introduction

For the road freight transport network to remain safe, reliable, and resilient, it is critical to analyze the spatial and temporal patterns of truck activities. Specifically, understanding the exchanges of different commodities between freight origins and destinations is crucial for identifying major trucking routes, and consequently, for analyzing the network's vulnerability and resilience. Origin-destination (OD) data are also used for informing highway upgrade investments to accommodate specific economic opportunities, allocating on-road enforcement resources, and conducting safety studies to enhance traffic conditions. Traditionally, expensive in-the-field surveys have been used to construct OD flow matrices. Given the competitive nature of the industry and privacy concerns, the direct collection of extensive real-world truck OD data is impractical. Therefore, there is a growing concern toward more practical and economically viable methods to estimate freight OD matrices.

The problem with OD estimation can become more complex at the network level, primarily due to the vast coverage involved. AI has demonstrated substantial capabilities in managing these large-scale challenges and extensive datasets. The proficiency of AI positions it as a feasible alternative for OD estimation, providing a robust solution where traditional methods may struggle. However, one inherent challenge with employing AI approaches is their need for large datasets for effective training, and existing OD data samples are very abstract and limited, failing to serve as extended training data needs for AI models. Thus, we fill this gap by reproducing the synthetic data based on limited available OD data.

Given the challenges mentioned, this research explores the use of Bayesian analysis to generate synthetic truck OD trip data. Bayesian techniques are particularly adept at synthesizing data while explicitly incorporating uncertainty and leveraging prior knowledge, thus providing a robust dataset. However, theoretical models alone, such as Bayesian analysis, may not fully capture the complexities of real-world traffic patterns and variations without practical validation. To ensure the real-world relevance of the synthetic data, it will be benchmarked against simulated Origin-Destination data integrated with a layer of temporal variability, generated using professional transport modelling software, PTV Visum. This comparison will offer a rigorous assessment by contrasting the results of a theoretical approach with those of real-world transportation planning. Hence, the primary contributions of this study are:

- Augmenting data by leveraging posterior predictive distributions to generate new freight trip counts, incorporating observed data characteristics and inherent uncertainty.
- Developing a framework to address the challenge of limited sample sizes, enabling the expansion of the existing dataset and contributing to the broader goal of training an AI-based OD estimation model.
- Providing a framework that bridges the gap between theoretical data generation methods and real-world transportation modelling, supporting more informed decision-making and policy development.

In this paper, we investigate the process of synthesizing freight origin-destination trips using Bayesian analysis and explore its implications for enhancing predictive accuracy in transportation planning for heavy vehicles. We organize our discussion into the following sections. The next section provides a literature review. In section 3, we outline our Bayesian approach for modelling OD trips and simulating future observations based on current data. We then introduce a TFlowFuzzy-based method for generating synthetic data. In section 4, we elaborate on our case study, the highway network of the Canadian Prairie, as well as the dataset utilized, the Canadian Freight Analysis Framework (CFAF) database. In section 5, we present a comparative analysis of the results obtained from the proposed approaches in section 3, highlighting their respective strengths and limitations. Finally, in the concluding section, we discuss the implications of our findings, acknowledge the limitations of our study, and outline potential avenues for future research.

2. Literature Review

The advancement of artificial intelligence (AI) in transportation and logistics can be limited by the difficulty in accessing comprehensive datasets required for model training and testing. This constraint is prevalent across various data types, including freight origin-destination (OD) trips (Basso et al., 2022). This issue stems from the industry's competitive nature and privacy concerns regarding the use of real-world data (Yao et al., 2024). Also, the process of collecting and labeling data requires considerable time and resources. Given these challenges, the idea of creating synthetic data stands out as a viable option, allowing for data sharing and utilization in ways that real-world data cannot facilitate (Jordon et al., 2022).

Just as they have provided solutions to numerous problems, machine learning approaches also demonstrate substantial potential in the data generation process. Generative AI models, also referred to as deep generative or distribution learning models, can learn the distribution of data from existing datasets and produce new data objects. Techniques such as Generative Adversarial Networks (GANs) (Goodfellow et al., 2014; Goodfellow et al., 2020), Variational Autoencoders (VAEs) (Kingma and Welling, 2013), and other deep learning-based models (Ghosh et al., 2017; Ho et al., 2020) have become popular for their ability to generate high-quality synthetic datasets. However, these machine learning methods often require large datasets for effective training, and their performance deteriorates with smaller datasets due to overfitting and instability in generalization.

Traditional statistical approaches often fall under the frequentist paradigm, generating synthetic data by resampling from established statistical models. In these models, parameters are assumed to be fixed and known, with probability defined as the long-term frequency of occurrences. Even though frequentist methods can indeed be used for probabilistic forecasting, disregarding uncertainty in parameter estimation can yield overly confident predictions that misrepresent the true level of variability in the data, especially, when available data is sparse or incomplete (Wagenmakers et al., 2008).

Bayesian methods, on the other hand, are well-regarded for their ability to incorporate uncertainty directly into the model, which can be particularly beneficial for synthetic data generation when dealing with limited reference data (Wharrie et al., 2022; Falconer et al., 2023). They can also offer interpretable models that provide insights into the relationships between variables and the data generation process through Bayesian networks (Gogoshin et al., 2021; Chen et al., 2019). Bayesian models have found extensive applications in transportation research. Davis (2000) introduces a Bayesian method using Gibbs sampling to estimate traffic accident rates while accounting for uncertainties in traffic volume estimates. The study explains how traditional methods underestimate error by ignoring uncertainty in traffic volume, leading to inaccurate assessments of hazardous locations. Hewett at al. (2024) use a Bayesian framework with a two-stage Markov Chain Monte Carlo (MCMC) procedure to model spatio-temporal collision rates, incorporating temporal dependence and spatial correlations across zones. Parry and Hazelton (2013) develop a Bayesian methodology using adapted MCMC methods to sample latent route flows and estimate traveler behavior parameters. Bayesian analysis often requires the calculation of complex integrals, particularly due to the intricate posterior distributions and the lack of closed-form solutions. Markov Chain Monte Carlo and Monte Carlo sampling methods facilitate the approximation of these distributions and play a crucial role in synthetic data generation, parameter inference, and predictive modelling within Bayesian frameworks.

Although the study of freight OD matrices has lagged behind passenger OD matrices due to challenges in obtaining comprehensive data (Holguin and Patil, 2007), several efforts have been made in the literature to estimate freight OD matrices using different data sources and different models, ranging from genetic algorithms (Al and Kaysi, 2005) to fuzzy theory (Shan and Li, 2008). However, the application of machine learning and deep learning models remains underexplored in this domain. Notably, the use of AI techniques in OD traffic flow estimation (Lorenzo and Matteo, 2013; Chu et al., 2019; Afandizadeh et al., 2021) has demonstrated promising results in terms of accuracy and scalability, suggesting that these methods have significant potential for improving freight flow predictions as well. Recognizing the potential of AI in estimating origin-destination trips, but acknowledging the scarcity of comprehensive freight OD data necessary for developing such models, this study focuses on synthesizing freight OD trip data through Bayesian analysis. This method aims to address existing data limitations and enhance the predictive accuracy of supervised deep-learning models dedicated to the estimation of Canadian freight OD flow. This method is then compared with TFlowFuzzy (Yousefikia et al., 2013; Crișan and Filip, 2016), a simulation-based technique specifically developed for demand matrix correction within the transport planning software PTV Visum (PTV Group, 2023).

3. Methodology

This study aims to generate synthetic freight origin-destination (OD) trips, where each trip represents the movement of goods from the start to the end point of a logistics activity, using two distinct approaches. In a network comprising N nodes, the OD trip matrix is represented as an $N \times N$ matrix, where each element corresponds to the number of trips between a specific origin and destination pair. Modelling these trip counts is essential for synthesizing realistic trip data. In this section, we review the two approaches: the proposed Bayesian model in Subsection 3.1 and the TFlowFuzzy technique in Subsection 3.2.

3.1 Bayesian Modeling

Given that the random variables of interest (number of OD trips) are considered counts, the Poisson distribution might be a reasonable fit. The statistical framework proposed by (Vardi, 1996) involves this assumption that the origin-destination matrix is derived from a series of independent Poisson distributions for each OD pair. However, the Poisson distribution has strict requirements, including the assumption that the mean equals the variance and remains constant over the basic analysis unit. These conditions may not always hold in the case of real-world OD trips, especially because of overdispersion or when the data represents tonnage rather than discrete counts. For counts where the means are sufficiently large, the normal distribution offers a more precise approximation (Maher, 1983; Lam et al., 2008). This is because, as the mean increases in the case of Poisson distributions, the distribution tends to become more symmetric, gradually approximating the characteristics of a normal distribution.

Therefore, we assume that each OD pair in the origin-destination trip matrix, denoted as x in the equations, follows an independent normal distribution. Next, we will detail our approach to modelling x and describe the process of simulating potential future observations y given the current data x. Consider a random sample $X_1, \dots, X_n$ with the likelihood function $f(x|\theta)$. Assuming θ has the prior distribution $\pi(\theta)$, the posterior distribution of the parameter θ based on the observed data x is given by (Bolstad and Curran, 2016),

$$\pi(\theta|x) = \frac{f(x|\theta) \times \pi(\theta)}{\int f(x|\theta) \times \pi(\theta)\, d\theta} \tag{1}$$

Here, the denominator serves as a normalizing constant, ensuring that the posterior distribution is a proper probability distribution. Due to the presence of this constant, the posterior distribution can be proportionally expressed as follows

$$\pi(\theta|x) \propto f(x|\theta) \times \pi(\theta) \tag{2}$$

Given that x is assumed to follow a normal distribution with μ representing the mean and σ^2 the variance, the likelihood function is expressed as

$$f(x|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \tag{3}$$

Incorporating this into the Bayesian framework in Equation (2), the joint posterior distribution for $\theta = (\mu, \sigma^2)$ is updated to

$$\pi(\mu, \sigma^2|x) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \times \pi(\mu, \sigma^2) \tag{4}$$

Once we gain better insight into the uncertainty in θ from the posterior $\pi(\theta|x)$, we can predict future observations y by generating samples from the predictive posterior distribution $f(y|x)$, which can be formulated as

$$f(y|x) = \int f(y|\theta) \times \pi(\theta|x)\, d\theta \tag{5}$$

The integration in Equation (5) can be difficult in practice, but we can use Monte Carlo methods to sample from it. Thus far, we have modeled x and new observations y. What follows will detail the modelling of the data's parameters μ and σ^2, along with a comprehensive overview of the data generation process.

There are several options for selecting priors for parametric models. When specific prior information about θ is available, it should be incorporated into the prior distribution. Informative priors, such as conjugate priors, can capture this information while aligning with the model structure and simplifying computations. Noninformative priors, like Jeffreys prior, are particularly valuable when specific prior information about θ is absent. These priors convey vague or broad knowledge about a parameter, adhering to the principle of allowing "the data to speak for themselves" (Gelman et al., 2013). Each of these approaches has distinct advantages, depending on the availability of prior information and the specific requirements.

Previously, we noted that each freight origin-destination pair is modeled as coming from a normal distribution with μ representing the mean and σ^2 the variance. When the available data is limited, the sample mean and variance may not reliably represent the true population parameters. To account for this uncertainty and due to the lack of prior information, both the mean and variance are considered unknown. A Jeffreys prior is applied to these parameters to incorporate a high degree of uncertainty into the model. The Jeffreys prior, known for its invariance under reparameterization, is specifically designed to facilitate an automated approach for deriving a noninformative prior suitable for parametric models in such scenarios.

This scenario presents a situation with maximum uncertainty about parameters, allowing us to explore how Bayesian analysis can proceed from a state of minimal prior knowledge. Adopting the Jeffreys prior of $\frac{1}{\sigma^3}$ for the parameter vector (μ, σ^2), which is derived from the Fisher Information matrix, Equation (4) is revised to

$$\pi(\mu, \sigma^2|x) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \times \frac{1}{\sigma^3} \tag{6}$$

To simulate samples for μ and σ^2 from the bi-variate posterior distribution $\pi(\mu, \sigma^2|x)$, Gibbs sampling is utilized (Brooks and Roberts, 1998). This technique, which is one of the Markov Chain Monte Carlo algorithms, requires the identification of full conditional distributions. Once we have samples for μ and σ^2 from their full conditional distributions, we can predict new observations using Equation (5) by applying the Monte Carlo technique.

3.2 TFlowFuzzy

The TFlowFuzzy (TFF) approach is a method designed to refine origin-destination matrices in transportation modelling by accounting for data uncertainty and variability. The process begins with an initial demand matrix, which is then adjusted using specified traffic counts and predefined tolerance limits. This approach, as implemented in PTV Visum, builds on the work of Rosinowski (1994), who applied the Fuzzy Sets Theory to model traffic count values, effectively managing data uncertainties and making the model both flexible and robust.

In this study, we utilized a multi-step approach to create a set of synthetic truck origin-destination matrices incorporating a layer of temporal variability by employing the TFlowFuzzy method within PTV Visum. This approach combines Annual Average Daily Truck Traffic (AADTT) with the variations across different months and days of the week in traffic patterns. By integrating these temporal changes, we were able to develop an OD dataset that is both realistic and reflective of actual truck traffic conditions. The following paragraphs provide a detailed explanation of each step involved in this process.

Initial OD matrix assignment:
The process starts by assigning a base OD matrix, derived from a database containing the annual average daily trips between freight OD pairs, to the highway network model. This assignment is carried out using PTV Visum, which simulates traffic flows based on the OD matrix and results in a set of link volumes across the network.

Temporal disaggregation using factors:
To account for temporal variability, the AADTT link volumes generated in the first step are adjusted using predefined factors for both months and days of the week. Specifically, 12 monthly factors (M) and 7 day-of-the-week factors (D) are applied. Dividing AADTT by the monthly factor yields the Monthly Average Daily Truck Traffic (MADTT), and similarly, dividing AADTT by the day-of-the-week factor results in the Annual Average Day-of-Week Truck Traffic (AADWTT). By combining these two, the Average Daily Truck Traffic (ADTT) for each day of the week in each month is calculated as $\frac{\text{AADTT}}{\text{M}\times\text{D}}$.

This process results in the creation of $12 \times 7 = 84$ distinct ADTT scenarios, as each day of the week within each month corresponds to a unique combination of monthly and day-of-the-week adjustments. The monthly and day-of-the-week factors used in this study are based on the most recent findings of research projects discussed in (Reimer and Regehr, 2013; Regehr and Reimer, 2013; Grande et al., 2022).

Refinement with TFlowFuzzy:
With the 84 ADTT scenarios established, the next step involves refining the original OD matrix to align with these disaggregated traffic volumes. This is achieved using TFlowFuzzy, an iterative process within PTV Visum designed for matrix adjustment. This process is repeated for all 84 scenarios, producing a corresponding set of 84 updated OD matrices from the initial base matrix.

Iterations for multiple base OD matrices:
This entire procedure is systematically repeated for each base OD matrices from the database.

To simplify and automate the entire process, a custom Python script was developed. PTV Visum supports automation through its COM-API, which allows external control of the software via Python. By leveraging this capability, it was possible to programmatically execute the OD matrix refinement process across all scenarios, significantly reducing the time and effort required compared to manual execution.

4. Case Study

The database used for our analysis is sourced from the Canadian Freight Analysis Framework (CFAF), provided by Statistics Canada (Statistics Canada, 2018). A key component of CFAF is the Trucking Commodity Origin and Destination Survey (TCOD), conducted by Statistics Canada to track commodity movements in the Canadian trucking industry (Statistics Canada, 2019). By integrating data from TCOD and other sources, CFAF offers a comprehensive view of freight flows across the country. The data are disaggregated by geography, commodity, and mode of transport. The publicly available dataset covers annual freight flow data from 2011 to 2017 and includes key metrics such as tonnage, value, tonne-kilometers, and number of shipments, providing a foundation for economic and logistical studies. Regarding geographic detail, the CFAF primarily includes major cities across Canada's provinces. Since our focus is on the highway network within the Prairies, we have specifically used part of the data related to major cities in Manitoba, Saskatchewan, and Alberta, as well as neighboring provinces, Eastern Canada, and the United States.

We utilized the metric of shipments and normalized the number of shipments for each year by dividing by 365 to obtain the average daily shipments for a given calendar year. Given that the dataset spans 7 years, we generated 7 OD matrices, one for each year of available data. Since the number of shipments does not directly correspond to the number of truck trips, we first refined these matrices using TFlowFuzzy. By integrating actual link counts from the highway network in 2019 (Grande et al., 2022), we converted the shipment-based OD matrices into truck-based OD matrices, offering a more realistic representation of truck movements. The converted OD matrices are then used as the base OD matrices in the Bayesian and TFlowFuzzy approaches. This preprocessing step is not required if the data are initially provided in terms of trips or the number of trucks.

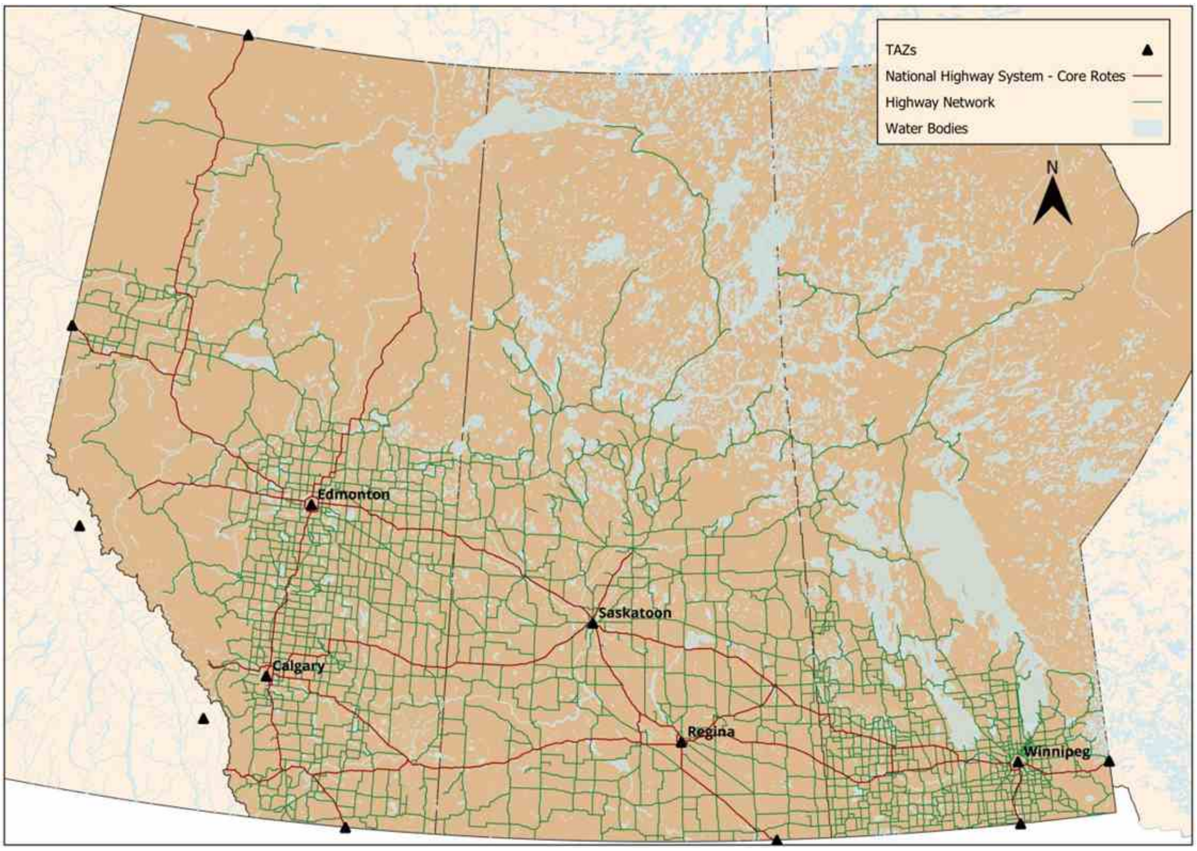


Figure 1. Core routes of the NHS in the Prairie and selected TAZs

Since our study concentrates on major cities within the Prairie provinces, we selected the core routes of the National Highway System (NHS) of Canada as the primary network for analysis. These routes are critical to Canada's transportation infrastructure, serving as the backbone for interprovincial and international trade and travel. The NHS core routes connect major population and commercial centers across Canada, including Manitoba, Saskatchewan, and Alberta. These highways are essential for facilitating the movement of goods and providing reliable connections to major border crossings and transport hubs. Therefore, our analysis focuses on the Traffic Analysis Zones (TAZs) corresponding to the major cities in the Prairies and the junctions where the NHS core routes intersect with bordering regions, including Eastern Canada, British Columbia, the Northern Territories, and the United States, providing an insightful examination of freight flows across this transportation network.

Figure 1 illustrates the highway network within the Prairie provinces, highlighting the core routes of the National Highway System (NHS) that this study focuses on, along with the centroid Traffic Analysis Zones (TAZs) selected for analysis. The map was created using QGIS Version 3.28 and integrates three separate GIS files to create a cohesive network. Although the GIS files used to create the map were pre-existing (Grande et al., 2022), significant improvements were made to create a cohesive network.

5. Results

In this section, we present and analyze the results of the Bayesian and TFlowFuzzy approaches applied to the generation of trip counts. Figure 2 presents the trace plots of the generated samples for the mean (μ) and standard deviation (σ) of trip counts for three selected origin-destination (OD) pairs, Regina-Winnipeg, Saskatoon-Edmonton, Calgary-Regina.

These samples were obtained through Gibbs sampling from the posterior distribution of μ and σ. A total of 10,000 iterations were performed, with the first 5,000 iterations considered as burn-in and discarded. The trace plots demonstrate good mixing and stationarity, indicating that the Gibbs sampler effectively converged for both parameters.

Figure 3 presents the annual trip counts generated from the posterior predictive distribution using the remaining Gibbs samples after burn-in. The histogram and scatter plots illustrate the distribution and variability of the generated counts for the same OD pairs, demonstrating the uncertainty captured through the Bayesian approach. In contrast, Figure 4 presents the histogram and scatter plots of the trip counts generated using the TFlowFuzzy method. A total of 588 OD matrices were produced, derived from 7 base matrices with 84 updated matrices for each. While the Bayesian approach generated 5,000 samples for all OD pairs in just 57 seconds, the TFlowFuzzy approach took approximately 11 hours to generate the 588 samples, highlighting the superior time efficiency of the Bayesian method.

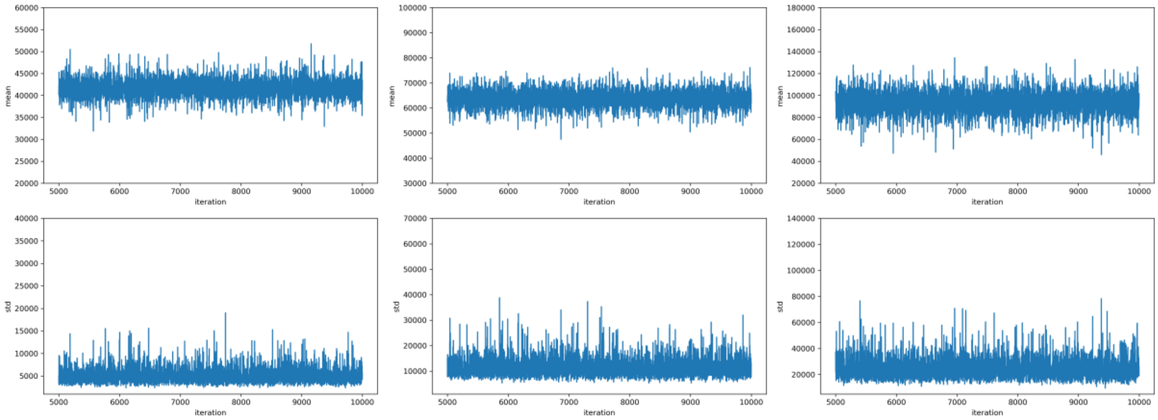
Table 1 provides statistics of the generated samples for seven additional OD pairs from both the Bayesian and TFlowFuzzy approaches, alongside the base OD matrices used. The results show that the Bayesian approach generates a wider range of trip counts compared to the TFlowFuzzy method, with a broader minimum-to-maximum spread. This wider distribution offers greater variability in the generated data, making it more suitable for training AI models. By capturing a broader spectrum of potential scenarios, the Bayesian method provides a more comprehensive dataset, improving the AI model's ability to generalize and perform well across unseen data.

6. Conclusion

In this study, we discussed the Bayesian framework and TFlowFuzzy approach adopted for generating synthetic truck OD trip data. The OD data is essential for supporting various truck-related analytical applications, including network resilience analysis, enforcement strategies, investment planning, and asset management. The results highlight the Bayesian approach's superior computational efficiency and ability to generate a broader range of trip counts, offering greater variability in the dataset compared to the TFlowFuzzy method. This variability is critical for training AI models, enhancing their ability to generalize and perform well on unseen data.

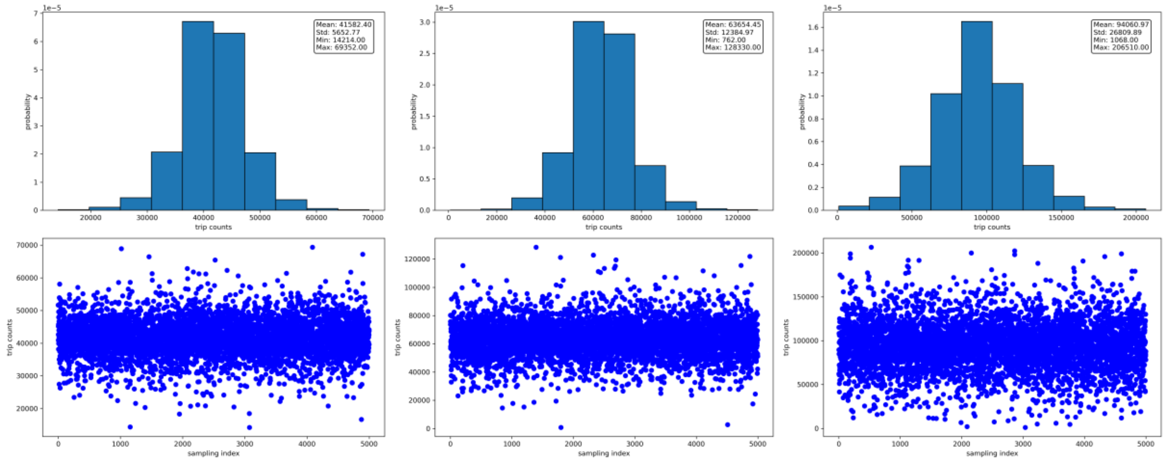
While the Bayesian approach is entirely data-driven and does not incorporate information about the highway network, the TFlowFuzzy approach, by contrast, has the advantage of utilizing network data. This feature becomes particularly valuable as the complexity of the highway network increases, with more routes between each OD pair. However, the significant computational time required by the TFlowFuzzy method makes it impractical for generating the large sample sizes necessary for training AI models. Despite this limitation, its network-aware nature presents opportunities for alternative uses.

A potential direction for future work lies in integrating these two approaches. TFlowFuzzy could be employed to provide prior information about the network, which would then inform the Bayesian model, allowing it to generate a higher volume of samples. This hybrid framework would combine the strengths of both methods, i.e., leveraging the network insights from TFlowFuzzy while maintaining the computational scalability of the Bayesian approach, ultimately improving the generation of synthetic data for AI training in OD estimation tasks.



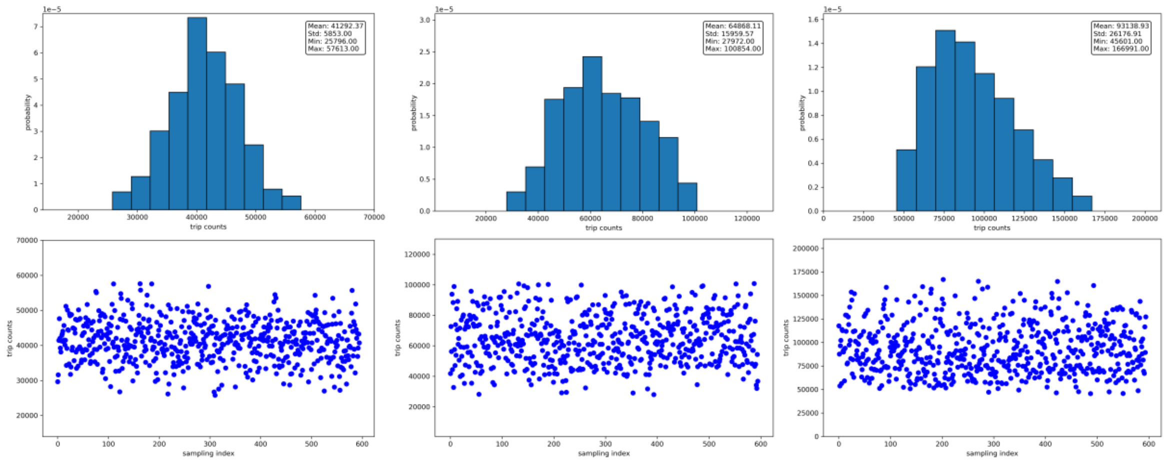
OD: Regina-Winnipeg OD: Saskatoon-Edmonton OD: Calgary-Regina

Figure 2. Trace plots of drawn μ and σ for three selected OD pairs



OD: Regina-Winnipeg OD: Saskatoon-Edmonton OD: Calgary-Regina

Figure 3. Histogram and scatter plots of drawn trip counts using the Bayesian approach for three selected OD pairs



OD: Regina-Winnipeg OD: Saskatoon-Edmonton OD: Calgary-Regina

Figure 4. Histogram and scatter plots of drawn trip counts using the TFF approach for three selected OD pairs

Table 1. Statistics of original freight trip counts vs. those generated from Bayesian and TFlowFuzzy approaches

O-D	Method	Mean	Std	Min	Max
Winnipeg - Regina	Base OD	44953.66	13024.62	31545	67055
	TFlowFuzzy	46803.82	18197.29	10025	95860
	Bayesian	45363.39	15920.28	283	131295
Regina - Calgary	Base OD	108662.32	28749.11	75747	147089
	TFlowFuzzy	110138.91	38024.97	46892	208157
	Bayesian	109318.27	34424.81	1453	269334
Saskatoon - Regina	Base OD	120504.59	20321	95160	147179
	TFlowFuzzy	121016.29	60738.84	18694	265128
	Bayesian	120926	34556.41	1008	274862
Calgary - Edmonton	Base OD	767219.04	155787.44	539079	954599
	TFlowFuzzy	763761.53	260765.7	316760	1296537
	Bayesian	772504.19	195401.23	67141	1890515
Edmonton - Winnipeg	Base OD	28217.1	6691.82	12916	33639
	TFlowFuzzy	29428.16	8708.49	7409	46961
	Bayesian	28474.03	8132.08	455	71375

7. References

- Afandizadeh Zargari, S., Memarnejad, A., & Mirzahossein, H. (2021). Hourly origin-destination matrix estimation using intelligent transportation systems data and deep learning. Sensors, 21(21), 7080.
- Al-Battaineh, O., & Kaysi, I. A. (2005). Commodity-based truck origin-destination matrix estimation using input-output data and genetic algorithms. Transportation Research Record, 1923(1), 37-45.
- Basso, F., Pezoa, R., Tapia, N., & Varas, M. (2022). Estimation of the origin-destination matrix for trucks that use highways: a case study in Chile. Sustainability, 14(5), 2645.
- Bolstad, W. M., & Curran, J. M. (2016). Introduction to Bayesian Statistics. John Wiley & Sons.
- Brooks, S. P., & Roberts, G. O. (1998). Assessing convergence of Markov chain Monte Carlo algorithms. Statistics and Computing, 8(4), 319-335.
- Chen, J., Zhang, R., Dong, X., Lin, L., Zhu, Y., He, J., Christiani, D. C., Wei, Y., & Chen, F. (2019). Shinybn: An online application for interactive Bayesian network inference and visualization. BMC Bioinformatics, 20, 1-5.
- Chu, K.-F., Lam, A. Y., & Li, V. O. (2019). Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions. IEEE Transactions on Intelligent Transportation Systems, 21(8), 3219-3232.
- Crișan, H. G., & Filip, N. (2016). Traffic modeling aspects using Visum software and effects on the traffic optimization. In Proceedings of the European Automotive Congress EAEC-ESFA 2015 (pp. 495-506). Springer.
- Davis, G. A. (2000). Estimating traffic accident rates while accounting for traffic-volume estimation error: A Gibbs sampling approach. Transportation Research Record, 1717(1), 94-101.
- Falconer, T., Kazempour, J., & Pinson, P. (2023). Bayesian regression markets. arXiv preprint arXiv:2310.14992.

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC.
- Ghosh, S., Chollet, M., Laksana, E., Morency, L.-P., & Scherer, S. (2017). Affect-lm: A neural language model for customizable affective text generation. arXiv preprint arXiv:1704.06851.
- Gogoshin, G., Branciamore, S., & Rodin, A. S. (2021). Synthetic data generation with probabilistic bayesian networks. Mathematical Biosciences and Engineering: MBE, 18(6), 8603.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.
- Grande, G., Regehr, J. D., Eng, P., & Mehran, B. (2022). Applying a hierarchical approach to develop a regional truck traffic volume model.
- Hewett, N., Golightly, A., Fawcett, L., & Thorpe, N. (2024). Bayesian inference for a spatio-temporal model of road traffic collision data. Journal of Computational Science, 80, 102326.
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.
- Holguin-Veras, J., & Patil, G. R. (2007). Integrated origin-destination synthesis model for freight with commodity-based and empty trip models. Transportation Research Record, 2008(1), 60-66.
- Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). Synthetic data—what, why and how? arXiv preprint arXiv:2205.03257.
- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
- Lam, W. H., Shao, H., & Sumalee, A. (2008). Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. Transportation Research Part B: Methodological, 42(10), 890-910.
- Lorenzo, M., & Matteo, M. (2013). OD matrices network estimation from link counts by neural networks. Journal of Transportation Systems Engineering and Information Technology, 13(4), 84-92.
- Maher, M. J. (1983). Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. Transportation Research Part B: Methodological, 17(6), 435-447.
- Parry, K., & Hazelton, M. L. (2013). Bayesian inference for day-to-day dynamic traffic models. Transportation Research Part B: Methodological, 50, 104-115.
- PTV Group. (2024). PTV Visum Traffic Software. Retrieved from https://www.ptvgroup.com/en/solutions/products/ptv-visum/ (Accessed: Aug. 19, 2024).
- Regehr, J. D., & Reimer, M. J. (2013). Methodology to estimate the distance traveled by trucks on rural highway systems. Journal of Transportation Engineering, 139(4), 390-398.
- Reimer, M., & Regehr, J. D. (2013). Hybrid approach for clustering vehicle classification data to support regional implementation of the mechanistic-empirical pavement design guide. Transportation Research Record, 2339(1), 112-119.
- Rosinowski, J. (1994). Entwicklung und Implementierung eines ÖPNV-Matrixkorrekturverfahrens mit Hilfe von Methoden der Theorie Unscharfer Mengen (Master’s thesis). University of Karlsruhe.

- Shan, J., & Li, X.-h. (2008). Estimating the highway freight origin-destination matrix from multi-source data based on fuzzy programming theory. In 2008 International Workshop on Modelling, Simulation and Optimization (pp. 204-209). IEEE.
- Statistics Canada. (2018). Canadian freight analysis framework (CFAF). Retrieved from https://www150.statcan.gc.ca/n1/pub/50-503-x/50-503-x2018001-eng.htm (Accessed: Nov. 16, 2024).
- Statistics Canada. (2019). Trucking commodity origin and destination survey. Retrieved from https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurveyId=1243542 (Accessed: Nov. 16, 2024).
- Vardi, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. Journal of the American Statistical Association, 91(433), 365-377.
- Wagenmakers, E.-J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. Bayesian Evaluation of Informative Hypotheses, 181-207.
- Wharrie, S., Yang, Z., Raj, V., Monti, R., Gupta, R., Wang, Y., Martin, A., O’Connor, L. J., Kaski, S., Marttinen, P., et al. (2022). Hapnest: an efficient tool for generating large-scale genetics datasets from limited training data. In NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research.
- Yao, Z., Tan, L., Yi, J., Fu, L., Zhang, Z., Tan, X., Xie, J., She, K., Yang, P., Wu, W., et al. (2024). Sensitive data privacy protection of carrier in intelligent logistics system. Symmetry, 16(1), 68.
- Yousefikia, M., Mamdoohi, A. R., Moridpour, S., Noruzoliaee, M. H., & Mahpour, A. (2013). A study on the generalized tflowfuzzy OD estimation. In Proceedings of Australasian Transport Research Forum 2013, Brisbane, Australia.