SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Analysis of Gradient-Based Minimization with Finite Difference Method and Automatic Differentiation on Double-Lane Roundabout

Cheah Yuat Hoong[1*], Yeak Su Hoe[2]

[1] Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Malaysia. E-mail: y_hoong0617@yahoo.com
[2] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia.
E-mail: s.h.yeak@utm.my
*Corresponding Author: Cheah Yuat Hoong.

Abstract
To ensure that the double-lane four-arm roundabout model can represent the real traffic performance, parameter estimation is carried out using two optimization approaches: partial exact gradient minimization via the Finite Difference Method and full exact gradient minimization via Automatic Differentiation, a built-in tool in MATLAB for machine learning applications. In the partial gradient approach, linear interpolation is applied to generate new data for the Finite Difference Method, due to the mismatch between data set sizes caused by varying time steps determined by the Courant-Friedrichs-Lewy (CFL) condition. To better reflect real traffic conditions, the model considers the entry and exit rates for both the inner and outer lanes. As a result, a total of sixteen parameters are estimated. Six pseudo-experiments are initially conducted for result analysis, acknowledging the potential non-uniqueness of the estimated parameters. Simulation results are evaluated based on two key criteria: computational time and accuracy. The partial exact gradient minimization shows strong accuracy but is time-consuming due to slower convergence. In contrast, the full exact gradient minimization demonstrates both faster convergence and higher robustness, making it more suitable for the fitting process when estimating a large number of parameters. Finally, the fitting process is successfully implemented, enabling the roundabout model to study the real traffic flow performance when actual Total Travel Time and Total Waiting Time data are available.

Keywords: Double-lane roundabout, Parameter estimation, CFL condition, Finite Difference Method and Automatic Differentiation.

Introduction

In recent years, roundabouts have gained significant popularity worldwide and are widely implemented in both urban and rural areas (Pulvirenti, Distefano et al. 2021). They represent a form of intersection control that can operate with or without traffic signals, using a circular geometric layout to create a self-regulated traffic system (Giuffrè, Granà and Tumminello 2016).

To date, a double-lane roundabout model has been developed (Cheah and Yeak 2026). To ensure that the proposed roundabout model can accurately analyze the real performance of traffic flow at roundabouts, a fitting procedure involving a minimization process must first be conducted. Therefore, two minimization methods are proposed in this research: partial exact gradient minimization and full exact gradient minimization. However, it should be noted that this model does not yield a unique solution.

Gradient-based minimization has been widely applied in fitting processes. This method has been shown to achieve faster convergence and more accurate parameter estimation in roundabout models compared to non-gradient approaches (Cheah and Yeak 2023). Another study demonstrated the using gradients in parameter estimation from qualitative data greatly enhances optimization efficiency and reliability, as it achieves faster convergence, better scalability and more accurate model fits than gradient-free methods (Schmiester, Weindl and Hasenauer 2021). Meanwhile, research on stochastic simulations has shown that gradient estimation plays a crucial role in optimization and sensitivity analysis. Techniques such as finite differences, perturbation analysis, likelihood ratios, and weak derivatives have enabled the use of gradient-based minimization for faster and more reliable parameter estimation (Fu 2014). Furthermore, in groundwater modelling, a hybrid approach combining global search with gradient-based refinement has been shown to improve parameter estimation accuracy (Agyei and Hatfield 2006).

In the partial exact gradient minimization approach, the Finite Difference Method requires two datasets of the same size. Since the datasets used in this study have different numbers of data points, linear interpolation is applied to generate additional points between the existing values. This produces datasets with matching sizes before the gradient is computed. Linear interpolation was selected because it is straightforward to implement and provides sufficiently accurate estimates for intermediate values. Previous studies have also reported that it performs better than the mean method when handling missing values in environmental datasets (Noor, Al Bakri Abdullah et al. 2015). Similarly, in statistical surveys, it provided an excellent fit for missing data compared to quadratic and cubic interpolation techniques (Noor, Yahaya et al. 2014).

Beyond producing well-fitted data, linear interpolation is also more efficient than many other interpolation methods in terms of algorithm implementation and computational speed (Jung and Chong 2017). More recently, it has served as the foundation for the Adaptive Multiple Linear Interpolation (AMLI) method, which addresses challenges such as insufficient prediction sample sizes or large errors between observed samples and actual distributions in machine learning applications (Du, Jin et al. 2023).

To overcome the issue of differing dataset sizes in the Finite Difference Method, the full exact gradient minimization approach is introduced. In this method, the Finite Difference Method is replaced with Automatic Differentiation (AD), a technique rooted in deep learning within machine learning. Today, AD has become a vital tool in scientific computing and machine learning because it enables accurate derivative computation without manual derivation. In differential equation modelling, AD facilitates parameter estimation, neural differential equations, and model diagnostics. Forward-mode AD is efficient for small systems, while continuous adjoint methods scale better for larger problems. Additionally, static AD outperforms tape-based approaches, particularly for nonlinear PDEs (Ma, Dixit et al. 2021).

In PDE shape sensitivity analysis, AD eliminates the need to code separate sensitivity solvers. Efficiency can be further improved by reformulating the PDE to remove mesh sensitivities, thus reducing both memory and computational costs (Margossian 2019). Moreover, in computational statistics, AD simplifies differentiation in Bayesian inference, Hamiltonian Monte Carlo, and neural networks. Modern implementations achieve high efficiency through operator overloading, memory optimization, and hybrid semi-analytical techniques (Baydin, Pearlmutter et al. 2018). In machine learning and broader scientific domains, AD generalizes backpropagation and underpins differentiable programming, providing gradients and Hessians at a cost comparable to evaluating the original function, thereby outperforming symbolic and numerical differentiation methods in both accuracy and efficiency (Borggaard and Verma 2000).

In this research, both gradient-based minimization methods are applied on the double-lane roundabout model as presented in the paper "Modeling of Double-Lane Roundabouts Using the Waiting System" (Cheah and Yeak 2026).The model operates as a macroscopic model.

Research Methodology

(a) Determination of Parameter Estimation and Pseudo Experiment Data

Referring to the paper "The fitting of roundabout model with gradient-based minimization" (Cheah and Yeak 2023), the exiting the roundabout rate, β, is treated as a single parameter estimation. In contrast, this study considers both the exiting the roundabout rate, β and the crossing the roundabout rate, P, as parameters to be estimated, with the Total Travel Time (TTT) on the road network and the Total Waiting Time (TWT) at the entrance of the secondary road serving as input variables for the minimization process. This approach is adopted because travel time and driver behaviour are inherently interconnected.

Logically, when drivers travel at higher speeds, both TTT and TWT decrease, and vice versa. This relationship extends to the exiting and entering rates: as driving speed increases, these rates also rise. Driving speed is influenced by behavioural factors, reflecting personality traits and temperament. Consequently, this relationship captures the realistic nature of traffic flow within roundabouts, where travel time and driver behaviour are mutually dependent.

(b) Mathematical Derivation

Based on the above determination, TTT and TWT are treated as the input variables for the pseudo experimental data, while the exiting the roundabout rate, β and the crossing the roundabout arm junction rate, P are regarded as the parameters to be estimated. Accordingly, following the approach described in (Cheah and Yeak 2023), parameter fitting for the double-lane roundabout model (Cheah and Yeak 2026) within the minimization process is performed by minimizing the derived objective function as follows:

$$\mathrm{E}(\beta_{\mathcal{P}}, \mathrm{P}_{\mathcal{P}}) = \sum_{\mathrm{i}=1}^{\mathrm{N}_{\mathrm{F_{in}}}} \left(\mathrm{TTT_k}(\mathrm{F_{in}}, \beta_{\mathcal{A}}, \mathrm{P}_{\mathcal{A}}) - \mathrm{TTT_k}(\mathrm{F_{in}}, \beta_{\mathcal{P}}, \mathrm{P}_{\mathcal{P}})\right)^2 + \left(\mathrm{TWT_k}(\mathrm{F_{in}}, \beta_{\mathcal{A}}, \mathrm{P}_{\mathcal{A}}) - \mathrm{TWT_k}(\mathrm{F_{in}}, \beta_{\mathcal{P}}, \mathrm{P}_{\mathcal{P}})\right)^2, \mathrm{k} = 1,2 \qquad (1)$$

where $\mathcal{A}$ and $\mathcal{P}$ represents the actual data and predicted data respectively, in which

$$TTT_k = \sum_k \sum_{n=1}^{N} \int_0^T \int_{I_n,k}^{N} \rho(x,t)\, dx\, dt + \sum_k \sum_{n=1}^{N} \int_0^T l_{n,k}(t)\, dt + T \cdot \sum_k \sum_{n=1}^{N} \int_{I_n,k}^{N} \rho(x,T)\, dx + T \cdot \sum_k \sum_{n=1}^{N} l_{n,k}(T) \qquad (2)$$

and

$$TWT_k = \sum_k \sum_{n=1}^{N} \int_0^T l_{n,k}(t)\, dt + T \cdot \sum_k \sum_{n=1}^{N} l_{n,k}(T) \quad (3)$$

where N is the number of arms of the roundabout, I is the interval between the arm junctions, Fin is the flux entering the secondary lane, l is the queue length, $N_{F_{in}}$ is the number of fluxes, $k = 1$ represents the inner lane, $k = 2$ represents the outer lane. Note that, there is no unit for TTT_k and TWT_k for Equations (2) and (3).

The next derivation is for the gradient calculation which means differentiating the Equation (1) with respect to β and P. Taking the partial derivative of Equation (1) with respect to β yields

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} E(\beta_{\mathcal{P}}, P_{\mathcal{P}}) = \sum_{i=1}^{N_{F_{in}}} \frac{\partial}{\partial \beta_{\mathcal{P}}} \left(TTT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right)^2 + \frac{\partial}{\partial \beta_{\mathcal{P}}} \left(TWT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right)^2$$
$$= \sum_{i=1}^{N_{F_{in}}} -2\left(TTT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right) \frac{\partial}{\partial \beta_{\mathcal{P}}} TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) + \sum_{i=1}^{N_{F_{in}}} -2\left(TWT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right) \frac{\partial}{\partial \beta_{\mathcal{P}}} TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) \quad (4)$$

where

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) = \sum_k \sum_{n=1}^{N} \int_0^T \int_{I_n,k}^{N} \frac{\partial}{\partial \beta_{\mathcal{P}}} \rho(x, t; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dx\, dt + \sum_k \sum_{n=1}^{N} \int_0^T \frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(t; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dt + T \cdot \sum_k \sum_{n=1}^{N} \int_{I_n,k}^{N} \frac{\partial}{\partial \beta_{\mathcal{P}}} \rho(x, T; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dx + T \cdot \sum_k \sum_{n=1}^{N} \frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(T; \beta_{\mathcal{P}}, P_{\mathcal{P}}) \quad (5)$$

and

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) = \sum_k \sum_{n=1}^{N} \int_0^T \frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(t; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dt + T \cdot \sum_k \sum_{n=1}^{N} \frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(T; \beta_{\mathcal{P}}, P_{\mathcal{P}}) \quad (6)$$

Next, taking the partial derivative of Equation (1) with respect to P yields

$$\frac{\partial}{\partial P_{\mathcal{P}}} E(\beta_{\mathcal{P}}, P_{\mathcal{P}}) = \sum_{i=1}^{N_{F_{in}}} \frac{\partial}{\partial P_{\mathcal{P}}} \left(TTT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right)^2 + \frac{\partial}{\partial P_{\mathcal{P}}} \left(TWT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right)^2$$
$$= \sum_{i=1}^{N_{F_{in}}} -2\left(TTT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right) \frac{\partial}{\partial P_{\mathcal{P}}} TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) + \sum_{i=1}^{N_{F_{in}}} -2\left(TWT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}}) - TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}})\right) \frac{\partial}{\partial P_{\mathcal{P}}} TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) \quad (7)$$

where

$$\frac{\partial}{\partial P_{\mathcal{P}}} TTT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) = \sum_k \sum_{n=1}^{N} \int_0^T \int_{I_n,k}^{N} \frac{\partial}{\partial P_{\mathcal{P}}} \rho(x, t; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dx\, dt + \sum_k \sum_{n=1}^{N} \int_0^T \frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(t; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dt + T \cdot \sum_k \sum_{n=1}^{N} \int_{I_n,k}^{N,K} \frac{\partial}{\partial P_{\mathcal{P}}} \rho(x, T; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dx + T \cdot \sum_k \sum_{n=1}^{N} \frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(T; \beta_{\mathcal{P}}, P_{\mathcal{P}}) \quad (8)$$

and

$$\frac{\partial}{\partial P_{\mathcal{P}}} TWT_k(F_{\text{in}}, \beta_{\mathcal{P}}, P_{\mathcal{P}}) = \sum_k \sum_{n=1}^{N} \int_0^T \frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(t; \beta_{\mathcal{P}}, P_{\mathcal{P}})\, dt + T \cdot \sum_k \sum_{n=1}^{N} \frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(T; \beta_{\mathcal{P}}, P_{\mathcal{P}}) \quad (9)$$

Since the double-lane roundabout model does not yield a unique solution, the model's accuracy is evaluated by defining the errors ε_k^{TTT} and ε_k^{TWT} as the differences between the computed Total Travel Time $TTT_k^{\mathcal{P}}$ and Total Waiting Time $TWT_k^{\mathcal{P}}$ from the parameter estimation and the actual Total Travel Time $TTT_k^{\mathcal{A}}$ and Total Waiting Time $TWT_k^{\mathcal{A}}$.

$$\varepsilon_k^{TTT} = TTT_k^{\mathcal{P}} - TTT_k^{\mathcal{A}} \quad (10)$$

$$\varepsilon_k^{TWT} = TWT_k^{\mathcal{P}} - TWT_k^{\mathcal{A}} \quad (11)$$

(c) Framework Topology

In general, the unknown parameters of dynamic models are estimated from the pseudo experimental data. In practice, it is essential to verify the accuracy of the obtained parameter estimation. The overall research procedure is illustrated as Figure 1.

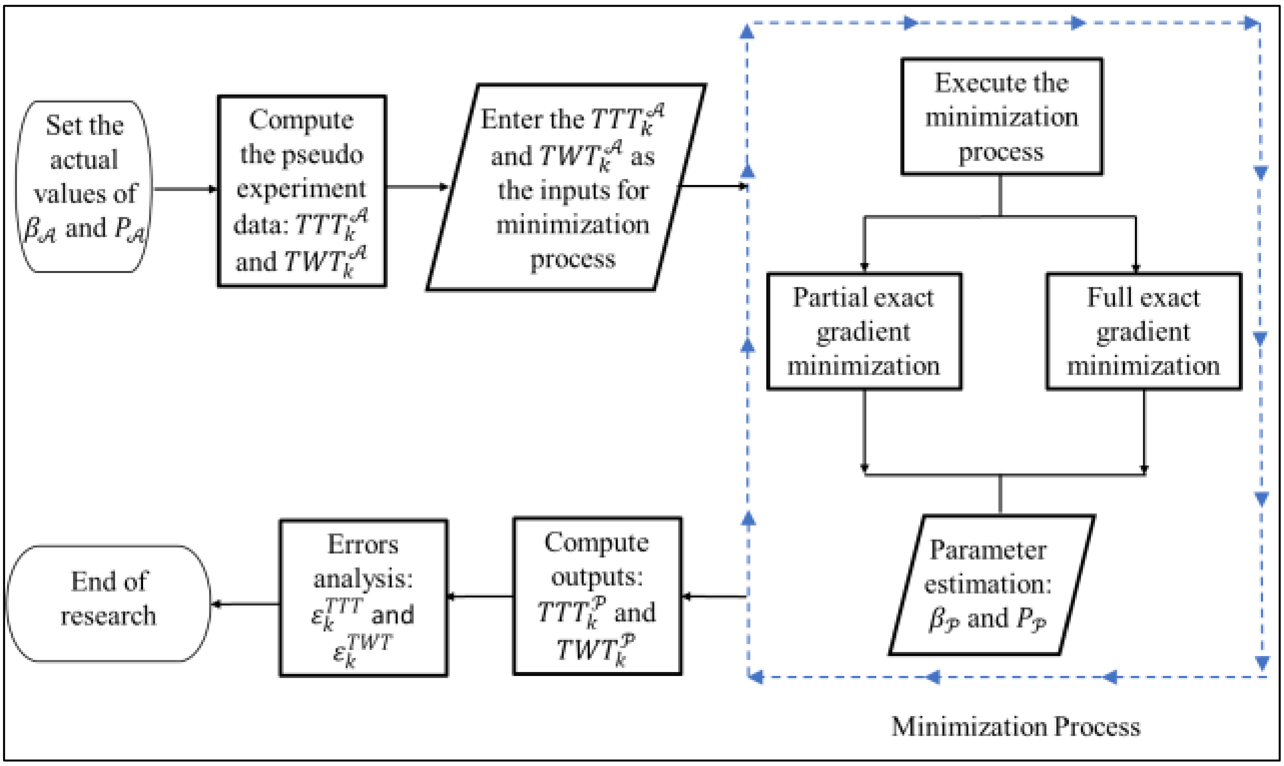

Figure 1 Research Framework

(d) Partial Exact Gradient Minimization

In this method, the Finite Difference Method is applied to the right-hand side of the Equations (5), (6), (8) and (9).

The central difference method is used when $\beta_{\mathcal{P}} + \Delta\beta$, $\beta_{\mathcal{P}} - \Delta\beta$, $P_{\mathcal{P}} + \Delta P$ and $P_{\mathcal{P}} - \Delta P$ are in the closed interval $[0, 1]$, yielding

$$\frac{\partial}{\partial \beta_{\mathcal{P}}}\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,t;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-\rho(x,t;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{2\Delta\beta} \quad (12)$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}}\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,T;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-\rho(x,T;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{2\Delta\beta} \quad (13)$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}}l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(t;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-l_{n,k}(t;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{2\Delta\beta} \quad (14)$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}}l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(T;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-l_{n,k}(T;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{2\Delta\beta} \quad (15)$$

$$\frac{\partial}{\partial P_{\mathcal{P}}}\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{2\Delta P} \quad (16)$$

$$\frac{\partial}{\partial P_{\mathcal{P}}}\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{2\Delta P} \quad (17)$$

$$\frac{\partial}{\partial P_{\mathcal{P}}}l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{2\Delta P} \quad (18)$$

$$\frac{\partial}{\partial P_{\mathcal{P}}}l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{2\Delta P} \quad (19)$$

Otherwise, the forward difference method is applied for cases where $\beta_{\mathcal{P}} - \Delta\beta < 0$ and $P_{\mathcal{P}} - \Delta P < 0$,

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} \rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,t;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta\beta} \tag{20}$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} \rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,T;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta\beta} \tag{21}$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(t;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta\beta} \tag{22}$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(T;\beta_{\mathcal{P}}+\Delta\beta,P_{\mathcal{P}})-l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta\beta} \tag{23}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} \rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta P} \tag{24}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} \rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta P} \tag{25}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta P} \tag{26}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}+\Delta P)-l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}})}{\Delta P} \tag{27}$$

and for cases where $\beta_{\mathcal{P}} + \Delta\beta > 1$ and $P_{\mathcal{P}} + \Delta P > 1$, the backward difference method is applied,

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} \rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}})-\rho(x,t;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{\Delta\beta} \tag{28}$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} \rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}})-\rho(x,T;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{\Delta\beta} \tag{29}$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}})-l_{n,k}(t;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{\Delta\beta} \tag{30}$$

$$\frac{\partial}{\partial \beta_{\mathcal{P}}} l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}})-l_{n,k}(T;\beta_{\mathcal{P}}-\Delta\beta,P_{\mathcal{P}})}{\Delta\beta} \tag{31}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} \rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}})-\rho(x,t;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{\Delta P} \tag{32}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} \rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}})-\rho(x,T;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{\Delta P} \tag{33}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}})-l_{n,k}(t;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{\Delta P} \tag{34}$$

$$\frac{\partial}{\partial P_{\mathcal{P}}} l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}) \approx \frac{l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}})-l_{n,k}(T;\beta_{\mathcal{P}},P_{\mathcal{P}}-\Delta P)}{\Delta P} \tag{35}$$

where $\mathcal{A}$ is actual data and $\mathcal{P}$ is prediction data.

When the Finite Difference Method is applied by adding and subtracting the $\Delta\beta$ and ΔP in Equations (12) – (35), the sizes of the density and queue lane datasets may differ due to variations in the time step, which are governed by the Courant–Friedrichs–Lewy (CFL) condition. This condition requires that the time step be sufficiently small to maintain numerical stability in simulations of partial differential equations. Since the sizes of the density and queue data depend on the time step, differences in dataset lengths can occur. In such cases, the subtraction operation in the Finite Difference Method cannot be directly implemented in Equations (10) – (33). Therefore, the linear interpolation method is employed to generate new data from the longer dataset, enabling it to match the shorter dataset for subtraction. This scenario is illustrated in Figure 2.

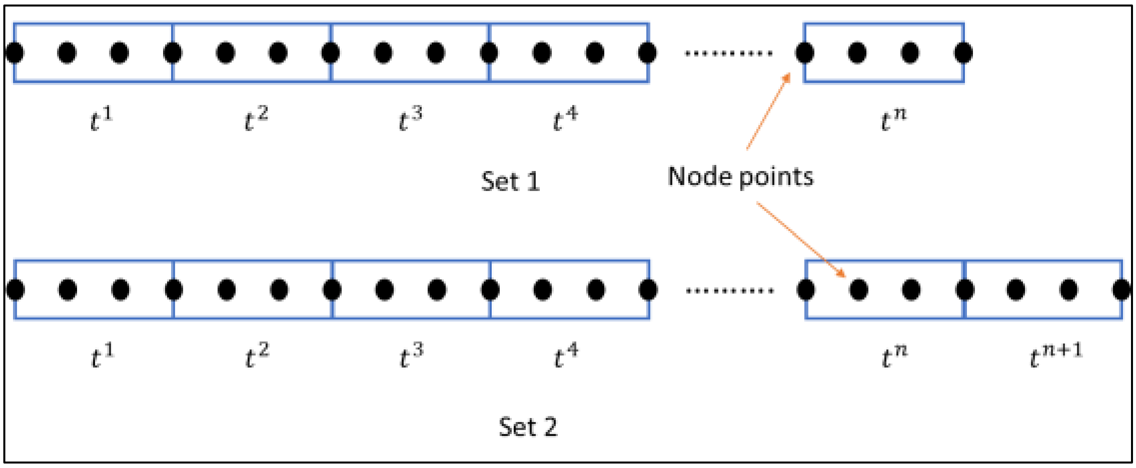

Figure 2 Scenario of Two Distinct Discretization of Time

After applying linear interpolation to the larger dataset (Set 2), the newly generated data replace the original Set 2 values, resulting in both datasets (Set 1 and Set 2) having equal sizes, as illustrated in Figure 3. Consequently, the Finite Difference Method described in Equations (12) – (35) can then be implemented.

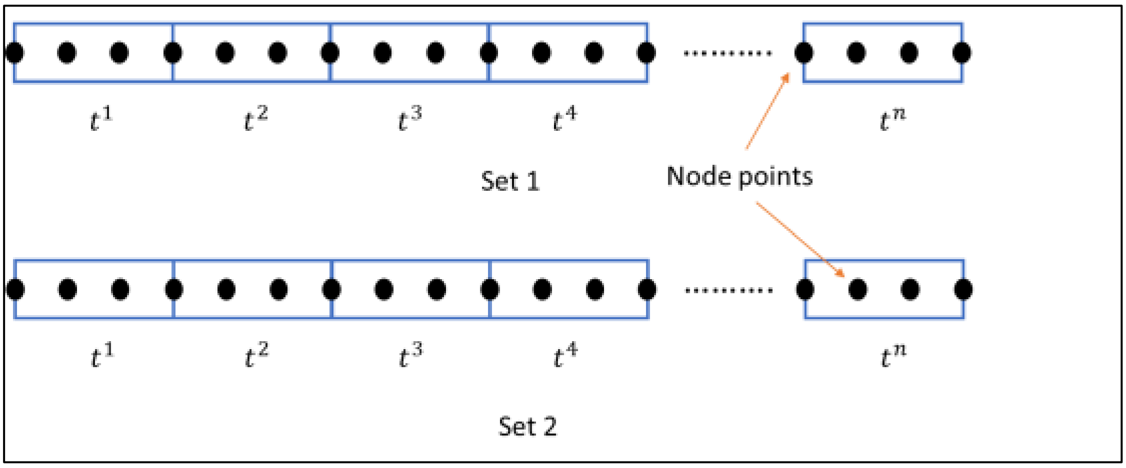

Figure 3 Scenario of Two Set of Data After the Execution of Linear Interpolation

In this minimization process, parameter estimation is carried out using constrained nonlinear optimization implemented through the fmincon solver in MATLAB.

Step 1: Computation of experiment data
Compute $TTT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}})$ and $TWT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}})$ by defining:

- Four identical sets of values of F_{in} for inner and outer lanes on each junction for four pseudo experiments.
- Eight values of exiting the roundabout rate, $\beta_{\mathcal{A}}$ for inner and outer lanes on each junction.
- Eight values of crossing the roundabout arm junction rate, $P_{\mathcal{A}}$ for inner and outer lanes on each junction.

Step 2: Setting of optimization boundaries and constraints

- Parameter values are restricted within the range $[0, 1]$.
- No additional linear or nonlinear constraints are applied beyond the bounded domain.
- A constraint function is included but set to return empty vectors, indicating unconstrained optimization.

Step 3: Performing the optimization algorithm

- MATLAB's fmincon solver is used for constrained nonlinear optimization.
- The Sequential Quadratic Programming (SQP) algorithm is applied.
- Solver options include iterative display of progress.

Step 4: Computation of objective function

- The computation of the objective function in Equation (1) is defined as the sum of squared errors (SSE) between experimental data and predicted data.

• For each pseudo experiment, the corresponding simulated values for the flux entering the secondary lane, F_{in} are computed.
• The squared differences are summed across all experimental scenarios to form the error measure.

Step 5: Computation of gradient
• Gradients of the error function with respect to each parameter are calculated explicitly based on the Equations (4) to (9).
• These gradients are accumulated across all experiments and supplied to the optimization solver.

Step 6: Parameter estimation output
• The optimization returned a parameter vector partitioned as:
○ Output $(1{:}4) \rightarrow \beta_{\mathcal{P}_{1:4,1}}$
○ Output $(5{:}8) \rightarrow \beta_{\mathcal{P}_{1:4,2}}$
○ Output $(1{:}4) \rightarrow P_{\mathcal{P}_{1:4,1}}$
○ Output $(5{:}8) \rightarrow P_{\mathcal{P}_{1:4,2}}$

• These represent the estimated parameters for exiting the roundabout rate, β and crossing the roundabout arm junction rate, P.

(E) Full Exact Gradient Minimization
The main objective of this method is to eliminate the issue arising from having two distinct datasets. Therefore, Automatic Differentiation (AD) is employed instead of the Finite Difference Method in Equations (12) – (35), using the commands dlfeval and dlgradient in the simulation. The overall procedure consists of the following steps:

Step 1: Computation of experiment data
Compute $TTT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}})$ and $TWT_k(F_{\text{in}}, \beta_{\mathcal{A}}, P_{\mathcal{A}})$ by defining:
• Four identical sets of values of F_{in} for inner and outer lanes on each junction for four pseudo experiments.
• Eight values of exiting the roundabout rate, $\beta_{\mathcal{A}}$ for inner and outer lanes on each junction.
• Eight values of crossing the roundabout arm junction rate, $P_{\mathcal{A}}$ for inner and outer lanes on each junction.

Step2: Setting of optimization boundaries and constraints
• Parameter values are restricted within the range $[0, 1]$.
• No additional linear or nonlinear constraints were applied beyond the bounded domain.
• A placeholder constraint function is included to ensure flexibility for future modifications.

Step 3: Performing the optimization algorithm
• MATLAB's fmincon solver is used for constrained nonlinear optimization.
• The Sequential Quadratic Programming (SQP) algorithm is applied.
• Solver options are configured to:
○ Display iterative progress.
○ Specify the objective gradient explicitly (enabled via automatic differentiation).
• A wrapper function is used to:
○ Convert input parameters to dlarray format.
○ Call the objective function via dlfeval.
○ Extract scalar loss and gradient for use in fmincon.

Step 4: Computation of objective function
• The computation of the objective function in Equation (1) is defined as the sum of squared errors (SSE) between experimental data and predicted data.
• For each pseudo experiment, the corresponding simulated values for the flux entering the secondary lane, F_{in} are computed.
• The squared differences are summed across all experimental scenarios to form the error measure.
• The squared differences are accumulated across all experiments to form the loss.
• The function is evaluated using dlfeval, which enables MATLAB to trace operations on dlarray objects for gradient computation.

Step 5: Computation of gradient
• Automatic differentiation (dlgradient) is employed to compute exact gradients of the loss function with respect to the parameters.

• This approach replaces the finite difference approximations, improving computational accuracy and efficiency.

Step 6: Parameter estimation output
• The optimization returns a parameter vector partitioned as:
○ Output $(1:4) \rightarrow \beta_{\mathcal{P}_{1:4,1}}$
○ Output $(5:8) \rightarrow \beta_{\mathcal{P}_{1:4,2}}$
○ Output $(1:4) \rightarrow P_{\mathcal{P}_{1:4,1}}$
○ Output $(5:8) \rightarrow P_{\mathcal{P}_{1:4,2}}$
• These represent the estimated parameters for exiting the roundabout rate, β and crossing the roundabout arm junction rate, P.

Numerical Setting
In this research, two main numerical settings are defined, differing only in the length of F_{in}. Each of these numerical settings is further divided into six pseudo experiments. The configurations of the six pseudo experiments are distributed as Table 1 and Table 2 as follows:

Table 1. Pseudo Experiments 1 to 3 with $F_{\text{in}} = 0.2, 0.4, 0.6, 0.8$

Pseudo Experiment	Pseudo Experiment 1	Pseudo Experiment 2	Pseudo Experiment 3
Method	Partial exact gradient minimization	Partial exact gradient minimization	Full exact gradient minimization
F_{in}	0.2, 0.4, 0.6, 0.8		
Time, T	2		
$\Delta\beta$	0.1	0.01	-
ΔP	0.1	0.01	-
$\beta_{n,1}$	0.2, 0.5, 0.7, 0.9 ($n = 1, 2, 3$ and 4)		
$\beta_{n,2}$	0.2, 0.5, 0.7, 0.9 ($n = 1, 2, 3$ and 4)		
$P_{n,1}$	0.1, 0.3, 0.6, 0.8 ($n = 1, 2, 3$ and 4)		
$P_{n,2}$	0.1, 0.3, 0.6, 0.5 ($n = 1, 2, 3$ and 4)		

Table 2. Pseudo Experiments 4 to 6 with $F_{\text{in}} = 0, 0.2, 0.4, 0.6, 0.8, 1$

Pseudo Experiment	Pseudo Experiment 4	Pseudo Experiment 5	Pseudo Experiment 6
Method	Partial exact gradient minimization	Partial exact gradient minimization	Full exact gradient minimization
F_{in}	0, 0.2, 0.4, 0.6, 0.8, 1		
Time, T	2		
$\Delta\beta$	0.1	0.01	-
ΔP	0.1	0.01	-
$\beta_{n,1}$	0.2, 0.5, 0.7, 0.9 ($n = 1, 2, 3$ and 4)		
$\beta_{n,2}$	0.2, 0.5, 0.7, 0.9 ($n = 1, 2, 3$ and 4)		
$P_{n,1}$	0.1, 0.3, 0.6, 0.8 ($n = 1, 2, 3$ and 4)		
$P_{n,2}$	0.1, 0.3, 0.6, 0.5 ($n = 1, 2, 3$ and 4)		

Results and Discussions

This section presents a comparative analysis of accuracy and computational performance between the partial exact gradient minimization and full exact gradient minimization approaches using six pseudo experiments. Figures 4 to 11 present a comparative evaluation of the partial exact gradient minimization (FDM) and full exact gradient minimization (AD) approaches in terms of both accuracy and computational efficiency. The computed Total Travel Time, $TTT_k^{\mathcal{P}}$ and Total Waiting Time, $TWT_k^{\mathcal{P}}$ obtained from the parameter estimation are compared with the actual Total Travel Time, $TTT_k^{\mathcal{A}}$ and Total Waiting Time, $TWT_k^{\mathcal{A}}$ under different flux entering the secondary lane rate, F_{in}.

In both Figures 4 and 5, the computed $TTT_k^{\mathcal{P}}$ shows close agreement with the actual $TTT_k^{\mathcal{A}}$ across all rates of flux entering the secondary lane rate, F_{in}. Meanwhile, in Figures 6 and 7, both methods also reproduce the expected results. However, in terms of computation time, the partial exact gradient minimization required approximately nine days to complete the simulation, while the full gradient minimization completed the same tasks in about five days, demonstrating that the full exact gradient minimization achieves superior accuracy with considerably lower computational cost.

When extending the flux entering the secondary lane rate, F_{in}, both methods maintain high accuracy for moderate rates as depicted in Figures 8 to 11. Both methods show slightly reduced accuracy of $TTT_k^{\mathcal{P}}$ and $TWT_k^{\mathcal{P}}$. Particularly, in Figure 11, the partial exact gradient minimization produces noticeably larger discrepancies, with its estimated $TWT_k^{\mathcal{P}}$ deviating significantly from the actual $TWT_k^{\mathcal{A}}$. For these simulations, partial exact gradient minimization took approximately six days, while the full exact gradient simulation completed the same runs in only three days. This again shows that the full exact gradient approach is efficient.

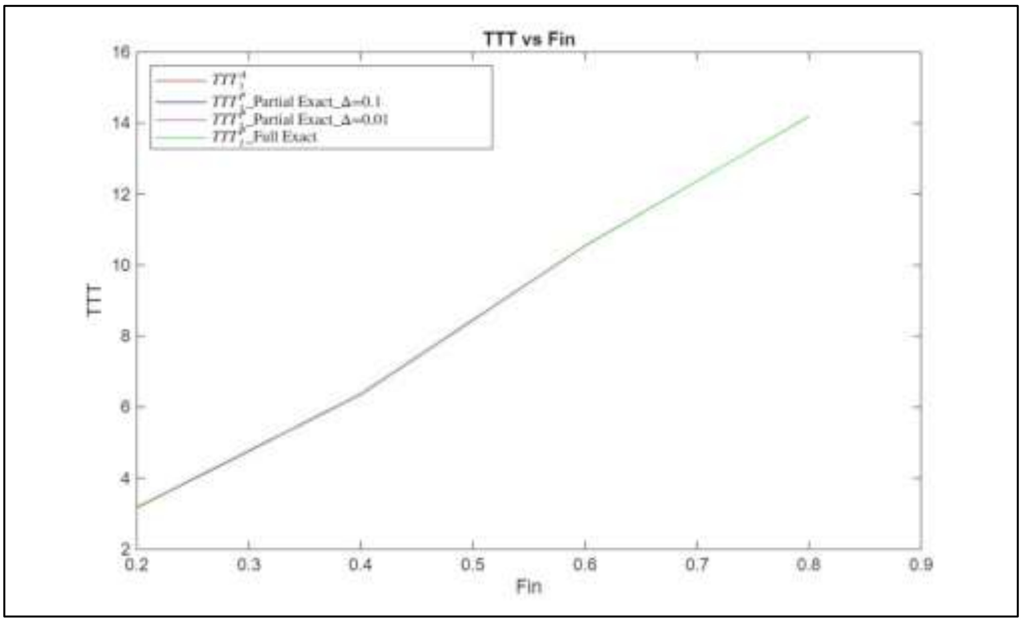

Figure 4 Comparison between $TTT_1^{\mathcal{A}}$ **and** $TTT_1^{\mathcal{P}}$ vs $F_{\text{in}} = 0.2$, 0.4, 0.6, 0.8

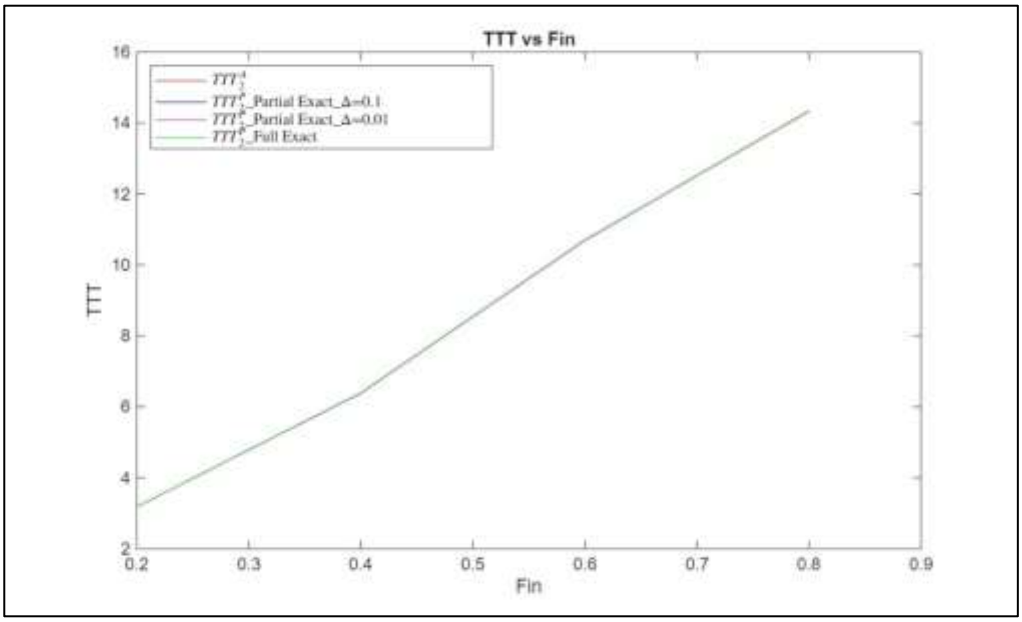

Figure 5 Comparison between $TTT_2^{\mathcal{A}}$ **and** $TTT_2^{\mathcal{P}}$ vs $F_{\text{in}} = 0.2$, 0.4, 0.6, 0.8

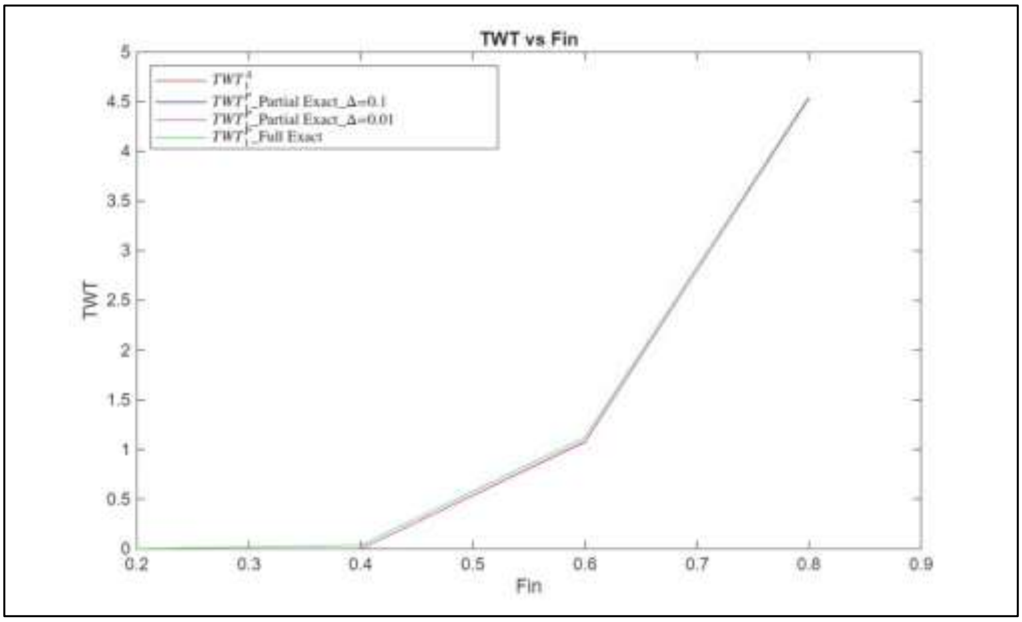

Figure 6 Comparison between $TWT_1^{\mathcal{A}}$ **and** $TWT_1^{\mathcal{P}}$ vs $F_{\text{in}} = 0.2\ , 0.4, 0.6, 0.8$

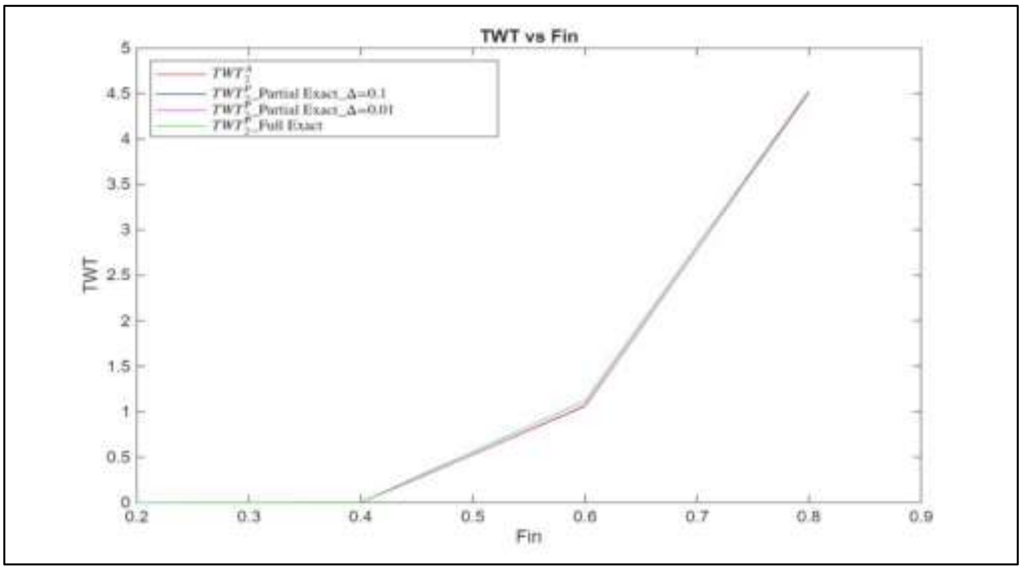

Figure 7 Comparison between $TWT_2^{\mathcal{A}}$ **and** $TWT_2^{\mathcal{P}}$ vs $F_{\text{in}} = 0.2\ , 0.4, 0.6, 0.8$

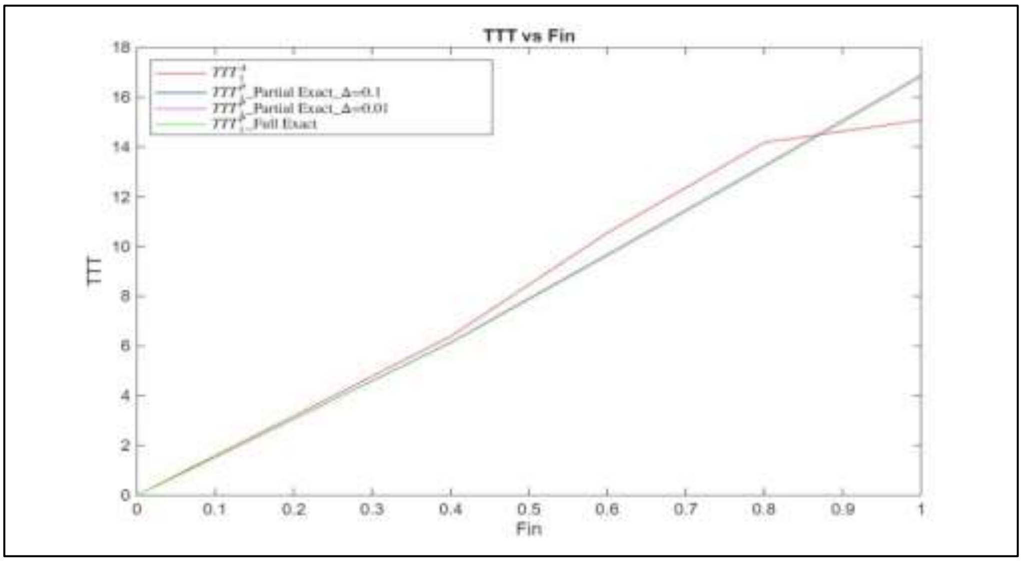

Figure 8 Comparison between $TTT_1^{\mathcal{A}}$ **and** $TTT_1^{\mathcal{P}}$ vs $F_{\text{in}} = 0, 0.2\ , 0.4, 0.6, 0.8, 1$

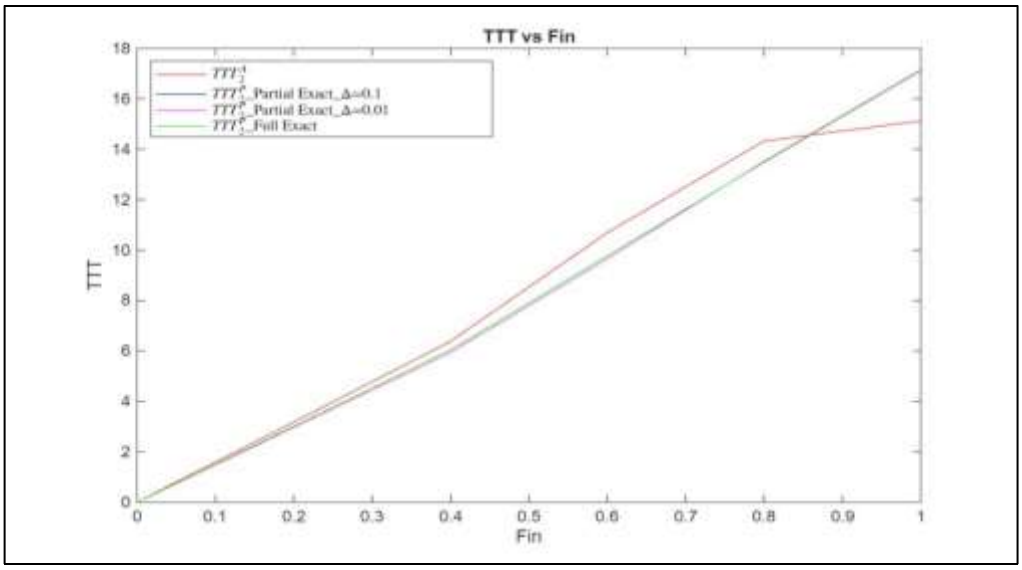

Figure 9 Comparison between $TTT_2^{\mathcal{A}}$ **and** $TTT_2^{\mathcal{P}}$ vs $F_{\text{in}} = 0, 0.2\ , 0.4, 0.6, 0.8, 1$

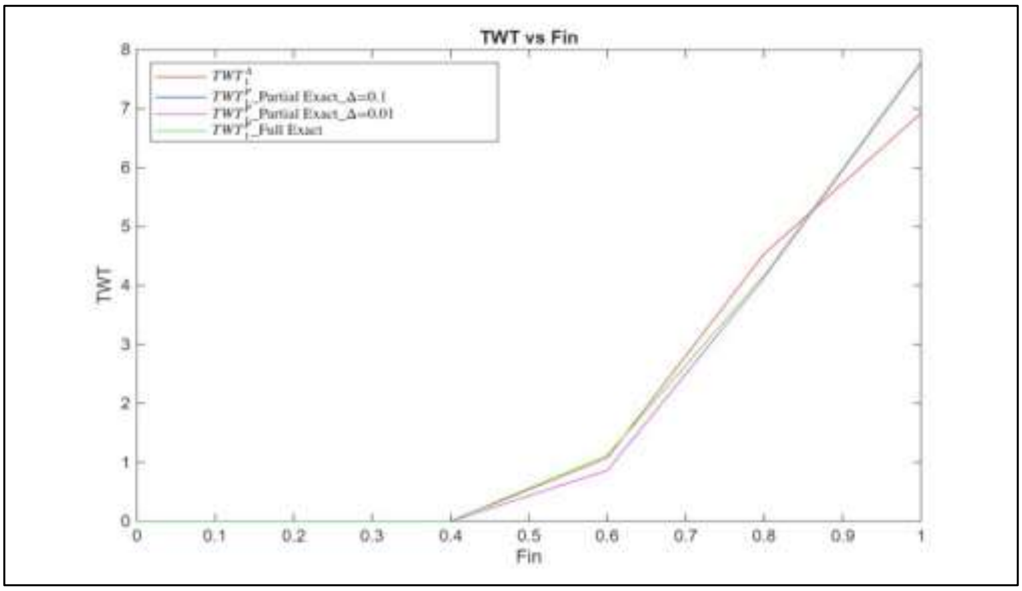

Figure 10 Comparison between $TWT_1^{\mathcal{A}}$ **and** $TWT_1^{\mathcal{P}}$ vs $F_{\text{in}} = 0, 0.2\ , 0.4, 0.6, 0.8, 1$

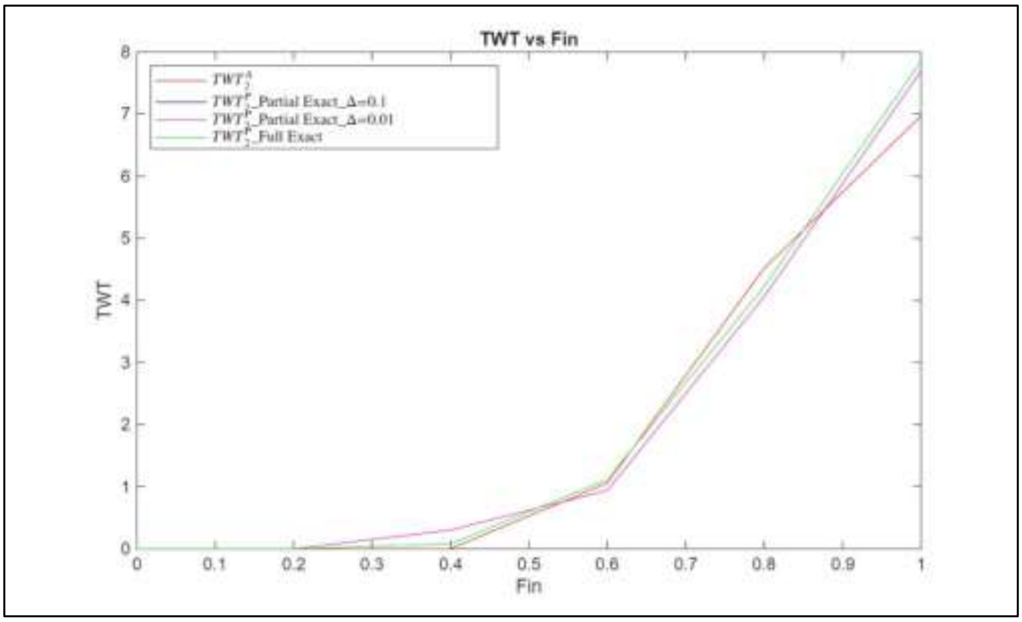

Figure 11 Comparison between $TWT_2^{\mathcal{A}}$ **and** $TWT_2^{\mathcal{P}}$ vs $F_{\text{in}} = 0, 0.2\ , 0.4, 0.6, 0.8, 1$

Furthermore, the computed errors, ε_k^{TTT} and ε_k^{TWT}, were computed based on Equations (10) and (11) as tabulated in Tables 3 and 4, to assess the accuracy and simulation time of both methods.

In pseudo experiments 1, 2 and 3, both methods achieve high accuracy across the flux entering the secondary lane rate, where the ε_k^{TTT} and ε_k^{TWT} remain within ± 0.01. However, as mentioned earlier, the partial exact gradient minimization required approximately nine days to complete, compared to only five days for the full minimization.

When the F_{in} range is extended in pseudo experiments 4 and 5, the partial exact gradient method exhibited larger error magnitudes with $|\varepsilon_k^{TTT}|$ exceeding ± 2.0 and $|\varepsilon_k^{TWT}|$ approaching ± 1.0, particularly near F_{in} boundaries. The simulation time was shortened to around six days. Compared to the full exact gradient method in pseudo experiment 6, $|\varepsilon_k^{TTT}|$ remains within ± 2.05 and $|\varepsilon_k^{TWT}|$ approaching ± 0.9.

Based on the results, the full exact gradient minimization results show significant advantages in both accuracy and computational efficiency. This is because in pseudo experiment 3, the simulation completed within five days, achieving small errors below 0.05. Moreover, in pseudo experiment 6, the simulation time was reduced further to only three days.

Table 3. ε_k^{TTT} and ε_k^{TWT} of Pseudo Experiments 1 to 3

Errors	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}
F_{in}	0.2		0.4		0.6		0.8	
Pseudo Experiment	Pseudo Experiment 1							
$k = 1$	-0.0073	0	0.0019	0	-0.0035	0.0045	0.0047	-0.0037
$k = 2$	-0.0031	0	0.0013	0	-0.0073	0.0063	0.0078	-0.0066
Pseudo Experiment	Pseudo Experiment 2							
$k = 1$	-0.0073	0	0.0019	0	-0.0035	0.0045	0.0047	-0.0037
$k = 2$	-0.0031	0	0.0013	0	-0.0073	0.0063	0.0078	-0.0066
Pseudo Experiment	Pseudo Experiment 3							
$k = 1$	-0.0241	0	-0.04	0.0347	-0.0164	0.0433	0.008	0.0156
$k = 2$	0.0091	0	0.021	0.0019	0.016	0.0573	0.0224	0.0132

Table 4. ε_k^{TTT} and ε_k^{TWT} of Pseudo Experiments 4 to 6

Errors	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}	ε_k^{TTT}	ε_k^{TWT}
F_{in}	0		0.2		0.4		0.6		0.8		1	
Pseudo Experiment	Pseudo Experiment 4											
k = 1	0	0	-0.1119	0	-0.2339	0	-0.8685	-0.2131	-0.9252	-0.4011	1.8059	0.8395
k = 2	0	0	-0.2083	0	-0.4343	0.301	-1.0265	-0.1194	-0.8114	-0.4555	2.02	0.7249
Pseudo Experiment	Pseudo Experiment 5											
k = 1	0	0	-0.1119	0	-0.2339	0	-0.8685	-0.2131	-0.9252	-0.4011	1.8059	0.8395
k = 2	0	0	-0.2083	0	-0.4343	0.301	-1.0265	-0.1194	-0.8114	-0.4555	2.02	0.7249
Pseudo Experiment	Pseudo Experiment 6											
k = 1	0	0	-0.1258	0	-0.2582	0	0	0.0433	-0.988	-0.3663	1.7431	0.8742
k = 2	0	0	-0.1576	0	-0.333	0.077	0	0.0573	-0.8429	-0.2954	1.9886	0.8849

Conclusion

In conclusion, both methods have achieved high accuracy, in particular, the full exact gradient minimization method demonstrates superior performance, offering a better balance between accuracy and computational efficiency. Therefore, it is a more practical and dependable option for conducting fitting when estimating a vast number of parameters. As a result, the fitting process has been effectively carried out, allowing the roundabout model to study real traffic performance when actual Total Travel Time and Total Waiting Time data are collected from field measurements.

Acknowledgments
This work was supported by Xiamen University Malaysia (XMUM) under the XMUM Research Fund (XMUMRF) (Grant number: XMUMRF/2024-C14/ISFS/0002).

References

1. Agyei, E. and K. Hatfield (2006). "Enhancing gradient-based parameter estimation with an evolutionary approach." Journal of Hydrology **316**(1): 266-280.
2. Baydin, A. G., et al. (2018). "Automatic differentiation in machine learning: a survey." Journal of machine learning research **18**(153): 1-43.
3. Borggaard, J. and A. Verma (2000). "On efficient solutions to the continuous sensitivity equation using automatic differentiation." SIAM Journal on Scientific Computing **22**(1): 39-62.
4. Cheah, Y. H. and S. H. Yeak (2023). The fitting of roundabout model with gradient-based minimization. Journal of Physics: Conference Series, IOP Publishing.
5. Cheah, Y. H. and S. H. Yeak (2026). "Modeling of Double-Lane Roundabouts Using the Waiting System." MATEMATIKA **42**(1): 49-71.
6. Du, Y., et al. (2023). "An adaptive multipath linear interpolation method for sample optimization." Mathematics **11**(3): 768.
7. Fu, M. C. (2014). "Stochastic gradient estimation." Handbook of simulation optimization: 105-147.
8. Giuffrè, O., et al. (2016). "Gap-accepteance parameters for roundabouts: a systematic review." European Transport Research Review **8**(1): 2.
9. Jung, I. and K. Chong (2017). "Interpolation and spatial matching method of various public data for building an integrated database." WIT Trans Built Environ **176**: 307-318.
10. Ma, Y., et al. (2021). A comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions. 2021 IEEE High Performance Extreme Computing Conference (HPEC), IEEE.
11. Margossian, C. C. (2019). "A review of automatic differentiation and its efficient implementation." WIREs Data Mining and Knowledge Discovery **9**(4): e1305.
12. Noor, M. N., et al. (2014). "Filling Missing Data Using Interpolation Methods: Study on the Effect of Fitting Distribution." Key Engineering Materials **594-595**: 889-895.
13. Noor, N. M., et al. (2015). "Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set." Materials Science Forum **803**: 278-281.
14. Pulvirenti, G., et al. (2021). "Are double-lane roundabouts safe enough? A chaid analysis of unsafe driving behaviors." Safety **7**(1): 20.
15. Schmiester, L., et al. (2021). "Efficient gradient-based parameter estimation for dynamic models using qualitative data." Bioinformatics **37**(23): 4493-4500.