Context-aware Bayesian choice models

Miroslawa Łukawska*,†, Anders Fjendbo Jensen†, and Filipe Rodrigues†

*Corresponding author (mirlu@dtu.dk)
†Technical University of Denmark, Department of Technology, Management and Economics, Bygningstorvet 116b, 2800 Kgs. Lyngby

Declarations of interest: none

Highlights

• We propose an approach to model context-dependent intra-respondent heterogeneity.
• Contextual information is mapped to additive shifts on preference parameters.
• The model allows non-linear interactions between continuous and discrete variables.
• Scenario analyses for different context attributes are possible in almost no time.
• We estimate a C-MMNL bicycle route choice model on 110,083 trips in 80 minutes.

Abstract

The mixed multinomial logit (MMNL) model assumes constant preference parameters of a decision-maker throughout different choice situations, which may be considered too strong for certain choice modelling applications. This paper proposes an effective approach to model context-dependent intra-respondent heterogeneity and introduces the idea of Context-aware Bayesian Mixed Multinomial Logit (C-MMNL) Model, where a neural network maps contextual information to shifts in the preference parameters of each individual in each choice occasion. The proposed model offers several key advantages. First, it supports for both continuous and discrete variables, as well as complex non-linear interactions between both types of variables. Secondly, each specification of the context is considered jointly as a whole by the neural network rather than each variable being considered independently. Finally, since the parameters of the neural network are shared across all decision-makers, it can leverage information from other decision-makers and use it to infer the effect of a particular context. Even though the C-MMNL model allows for flexible interactions between attributes, there is hardly an increase in the complexity of the model and the computation time, compared to the MMNL model. We present two real-world case studies from travel behaviour domain - a travel mode choice model and a bicycle route choice model. The bicycle route choice model is based on a large-scale, crowdsourced dataset of GPS trajectories including 110,083 trips made by 8,555 cyclists.

Keywords: choice context, bayesian modelling, neural networks, bicycle route choice, big data

1 Introduction

A long-standing major concern in discrete choice modelling is heterogeneity in the decision-making process. To overcome it, the mixed multinomial logit (MMNL) model (McFadden and Train 2000) adds flexibility to the original multinomial logit (MNL) model formulation (Boyd and Mellman 1980) by allowing each decision-maker \( n \) to have their own preference parameters \( \beta_n \), but constraining them to the population density \( f(\beta) \). However, this entails the assumption that the preference parameters of the decision-maker are constant throughout time and throughout different choice situations, which may be deemed too strong for certain choice modelling applications. When facing the same choice situation
multiple times, an individual might make different decisions, depending on a context. One can think of a context as an external factor that varies across individuals as well as choice situations and captures temporal (e.g. weather) or long-term (e.g. pandemics) circumstances. This often requires more complex model specification considerations.

Hess and Rose (2009) extended the MMNL model formulation by allowing for intra-respondent heterogeneity on top of the inter-respondent heterogeneity. Becker et al. (2018) further provided a Bayesian treatment of this intra-inter heterogeneity approach, and proposed a Markov-chain Monte Carlo (MCMC) procedure for performing inference. Danaf et al. (2020) extended this procedure by relaxing the constraint of normality assumptions. However, these approaches do not allow for systematic variations in preference parameters as a function of contextual variables.

With the development in computational hardware, Bayesian approaches to choice modelling have been gaining research interest. Washington et al. (2009) elaborated on the theory and specified the task to route choice modelling. Train (2001) compared the Bayesian approach to mixed multinomial logit with Maximum Likelihood Simulation (an experiment recently repeated and extended by Elshiewy et al. (2017)). When using non-informative priors, estimates in both approaches are similar, especially for large datasets (Congdon 2007; J. Huber and Train 2001). However, utilising the Bayesian approach allows the researchers to include more information within the estimation procedure, resulting in an improvement on the behavioural explanation within the sample. It also gives a possibility of obtaining full posterior distributions over the model parameters (including the individual-specific taste parameters) and taking advantage of modern approaches, e.g. utility generation (Rodrigues et al. 2020) or inference (Rodrigues 2022). Further extensions of the classical mixed logit model are possible using Variational Bayes for posterior inference, as shown in Krueger et al. (2019), where a method for including unobserved inter- and intra-individual heterogeneity in behaviour was derived.

In this work, we propose an approach to model context-dependent intra-respondent heterogeneity by introducing a neural network that maps contextual information to shifts in the preference parameters of the decision-maker in a Bayesian mixed multinomial logit framework. By making use of Stochastic Variation Inference (SVI) and GPU-hardware acceleration, we are able to scale the proposed model to handle a large-scale revealed preference (RP) dataset for bicycle route choice modelling, thereby reducing estimation times by several orders of magnitude when compared to Maximum Simulated Likelihood (MSL) estimation. Lastly, despite relying on a black-box function approximator (neural network), we show that the proposed model is highly interpretable and preserves the links to economic theories of the original MMNL.

The remainder of this paper consists of three sections. Section 2 briefly describes the standard MMNL model from the Bayesian perspective and introduces the proposed Context-aware Bayesian Mixed Multinomial Logit (C-MMNL) Model as an extension to the MMNL. Section 3 provides two extensive real-world case studies on travel behaviour: travel mode choice and bicycle route choice, both illustrating the estimation and interpretation of the C-MMNL model, and highlighting the advantages of the proposed approach. Section 4 concludes the paper.

2 Method

2.1 Mixed Multinomial Logit Model

Let us consider a standard discrete choice setup where on each choice occasion \( t \in \{1, \ldots, T_n\} \) a decision-maker \( n \in \{1, \ldots, N\} \) derives a random utility \( U_{ntj} = V(x_{ntj}, \eta_n) + \epsilon_{ntj} \) from each alternative \( j \) in the choice set \( C_{nt} \). The systematic utility term \( V(x_{ntj}, \eta_n) \) is assumed to be a function of covariates \( x_{ntj} \) and a collection of taste parameters \( \eta_n \), while \( \epsilon_{ntj} \) is a stochastic noise term. We consider the general setting under which the tastes \( \eta_n \) can be decomposed into a vector of fixed taste parameters \( \alpha \in \mathbb{R}^K \) that are shared across decision-makers, and random taste parameters \( \beta_n \in \mathbb{R}^k \), which are individual-specific.

We assume the random (individual-specific) taste parameters \( \beta_n \) to be distributed according to a multivariate normal distribution, i.e. \( \beta_n \sim \mathcal{N}(\zeta, \Omega) \). The fixed taste parameters \( \alpha \) and the mean vector \( \zeta \) are assumed to follow multivariate normal distributions: \( \alpha \sim \mathcal{N}(\lambda_\alpha, \Sigma_\alpha) \) and \( \zeta \sim \mathcal{N}(\mu_\zeta, \Sigma_\zeta) \). As for the covariance matrix \( \Omega \), as recommended in Hilbe (2009) and Barnard et al. (2000), we decompose our prior into a scale and a correlation matrix as follows: \( \Omega = \text{diag}(\tau) \times \Psi \times \text{diag}(\tau) \), where \( \Psi \) is a correlation matrix and \( \tau \) is the vector of coefficient scales. The components of the scale vector \( \tau \) are then given a vague half-Cauchy prior, e.g. \( \tau_k \sim \text{half-Cauchy}(10) \), while for the correlation matrix we...
employ a LKJ prior (Lewandowski et al. 2009), such that \( \Psi \sim \text{LKJ}(\nu) \), where the hyper-parameter \( \nu \) directly controls the amount of correlation that the prior favours.

The generative process of the MMNL model can be summarised as follows:

1. Draw fixed taste parameters \( \alpha \sim N(\lambda_0, \Xi_0) \)
2. Draw mean vector \( \zeta \sim N(\mu_0, \Sigma_0) \)
3. Draw scales vector \( \theta \sim \text{half-Cauchy}(\sigma_0) \)
4. Draw correlation matrix \( \Psi \sim \text{LKJ}(\nu) \)
5. For each decision-maker \( n \in \{1, \ldots, N\} \)
   (a) Draw random taste parameters \( \beta_n \sim N(\zeta, \Omega) \)
   (b) For each choice occasion \( t \in \{1, \ldots, T_n\} \)
      i. Draw observed choice \( y_{nt} \sim \text{MNL}(\eta_n, X_{nt}) \)

where \( \Omega = \text{diag}(\theta) \times \Psi \times \text{diag}(\theta) \) and \( \eta_n = [\alpha, \beta_n] \).

### 2.2 Context-aware Bayesian Choice Models

We present the idea of Context-aware Bayesian Mixed Multinomial Logit (C-MMNL) Model, where the context information is included in the form of an easily interpretable context-specific bias term \( \mu_t \), a non-linear function of the context information that shifts the preference parameters of each individual \( n \) in each choice occasion \( t \), i.e. \( \eta_{nt} = \eta_n + \mu_t \), where \( \eta_n = [\alpha, \beta_n] \). The adjustment term \( \mu_t \) is assumed to be Gaussian distributed, where the mean is determined by a neural network that takes as input the context information \( c_t \), i.e.: \( \mu_t \sim N(\mu_t | \text{NNet}_{\theta_{\text{net}}}(c_t), \sigma, I) \). In order to share statistical strength across individuals, we assume that all individuals shift their preference parameters in a similar way when faced with a given choice context \( c_t \), and therefore the parameters of the neural network, \( \theta_{\text{net}} \), are shared for all individuals. However, we note that, if this assumption is considered too strong for some applications, one can relax it by allowing the neural network to also take into account, for example, the socio-demographic characteristics of the decision-maker. This would allow for complex interactions between the latter and the context information \( c_t \). Figure 1 shows the graphical model representation of the proposed C-MMNL model.

The generative process of the proposed C-MMNL model can then be summarised as follows, where the main changes to the data-generating process assumed by the original MMNL model (described in Section 2.1) have been highlighted:

1. Draw fixed taste parameters \( \alpha \sim N(\lambda_0, \Xi_0) \)
2. Draw mean vector \( \zeta \sim N(\mu_0, \Sigma_0) \)
3. Draw scales vector \( \theta \sim \text{half-Cauchy}(\sigma_0) \)
4. Draw correlation matrix \( \Psi \sim \text{LKJ}(\nu) \)
5. For each choice occasion \( t \in \{1, \ldots, T_n\} \)
   (a) Draw context-specific shift term \( \mu_t \sim N(\mu_t | \text{NNet}_{\theta_{\text{net}}}(c_t), \sigma, I) \)
6. For each decision-maker \( n \in \{1, \ldots, N\} \)
   (a) Draw random taste parameters \( \beta_n \sim N(\zeta, \Omega) \)
   (b) For each choice occasion \( t \in \{1, \ldots, T_n\} \)
      i. **Compute context-adjusted taste parameters:** \( \eta_{nt} = \eta_n + \mu_t \), with \( \eta_n = [\alpha, \beta_n] \)
      ii. Draw observed choice \( y_{nt} \sim \text{MNL}(\eta_{nt}, X_{nt}) \)
Fig. 1: Graphical model representation of the proposed Context-aware Bayesian Mixed Multinomial Logit (C-MMNL) Model, where the key changes to the original MMNL model have been highlighted in red.

where \( \Omega = \text{diag}(\theta) \times \Psi \times \text{diag}(\theta) \). Kindly notice that we opt not to place a prior distribution on the neural network parameters \( \theta_{\text{nn}} \). Instead, the latter are treated as point parameters in the model to be estimated using a type-II maximum likelihood approach (also commonly referred to as “empirical Bayes”).

Letting \( z = \{ \alpha, \zeta, \theta, \mu_{1:T}, \beta_{1:N} \} \) denote the set of all latent variables in the C-MMNL model, the joint distribution \( p(z, y_{1:N} | \theta_{\text{nn}}) \) factorises as:

\[
p(z, y_{1:N} | \theta_{\text{nn}}) = p(\alpha | \lambda_0, \Xi_0) p(\zeta | \mu_0, \Sigma_0) \ p(\theta | \sigma_0) \ p(\Psi | \nu) \left( \prod_{t=1}^{T} p(\mu_t | c_t, \theta_{\text{nn}}, \sigma_c) \right) \\
\times \prod_{n=1}^{N} p(\beta_n | \zeta, \theta, \Psi) \prod_{t=1}^{T} p(y_{nt} | X_{nt}, \eta_{nt}). \tag{1}
\]

Our goal is then to use Bayesian inference to compute the posterior distribution of \( z \) given a dataset of observed choices: \( p(z | y_{1:N}, \theta_{\text{nn}}) \), while simultaneously finding a maximum likelihood estimate for the neural network parameters \( \theta_{\text{nn}} \). Unfortunately, exact Bayesian inference in this model is intractable and therefore we must resort to approximate inference methods. Concretely, in Section 2.4 we describe a Stochastic Variational Inference (SVI) algorithm for performing Bayesian inference in the proposed C-MMNL model.

2.3 Neural network

The neural network with parameters \( \theta_{\text{nn}} \) introduced in the previous section takes as input a vector \( c_t \) describing the context corresponding to the choice occasion \( t \), and outputs a vector \( o_t \in \mathbb{R}^{L+K} \), corresponding to the expected value of the context-dependent adjustments/shifts \( \mu_t \) to the fixed taste parameters \( \alpha \in \mathbb{R}^L \) and random taste parameters \( \beta_n \in \mathbb{R}^K \). We propose to approximate this mapping using a neural network architecture consisting of fully-connected layers with ReLU activations for the hidden layers and a linear activation for the output layer. The number of hidden layers required naturally depends on the complexity of the mapping. In practice, we found that a single hidden layer to be sufficient for the datasets considered in Section 3. Additionally, we found that the use of Dropout between hidden layers to generally lead to better results.

Given the flexibility of the neural network to capture complex patterns in the context data and their non-linear relation with the context-dependent shifts/adjustments of the taste parameters of a decision-maker \( n \), the proposed C-MMNL approach offers several key advantages over traditional approaches to include context information in the utility functions. First, it naturally supports for both continuous and discrete variables, as well as complex non-linear interactions between both types of variables. As a consequence, non-linear shifts in taste parameters that result from context variables (such as amount ....
of rain, see Section 3.2) can be captured. Secondly, each specification of the context is considered jointly as a whole by the neural network, rather than each variable being considered independently. The neural network then leverages observed choice data from multiple contexts to learn to extrapolate across different combinations of the context variables and their mapping to shifts in taste parameters. Therefore, even if one has never seen a particular combination of the context variables in the data, the neural network is expected to learn to non-linearly interpolate across contexts to estimate its effect on the taste parameters. Lastly, since the parameters $\theta_m$ of the neural network are shared across all decision-makers, it can leverage information from other decision-makers to make inferences about the effect of a particular context on the taste parameters of a given decision maker. In doing so, we can leverage the ability of neural networks to generalise across observations and use it to infer the effect of a particular context, even if we have never observed that decision-maker make choices in that context. Additionally, one can further supplement the context description vector $c$, with socio-demographic information about the decision-maker in order to make the generalization across decision-makers richer and more personalised.

### 2.4 Estimation

Given that exact Bayesian inference in the proposed C-MMNL model is intractable, we propose a Stochastic Variational Inference (SVI) procedure similar to one described in Rodrigues et al. (2020). Letting $z$ denote the set of all latent variables in the C-MMNL model, and $\theta_m$ denote the parameters of neural network, our goal is to compute the posterior distribution of $z$ and find point estimates for the neural network parameters $\theta_m$ given a dataset of observed choices. In Variational Inference (VI), we consider a tractable family of distributions $q(z|\phi)$ parameterised by $\phi$, which are referred to as the variational parameters. The goal of VI is then to find the values of $\phi$ that make the variational distribution $q(z|\phi)$ as close as possible to the true posterior distribution $p(z|y_{1:N}, \theta_m)$, thereby effectively reducing the inference problem into an optimization problem. Following the theory of VI (Jordan et al. 1998), this can be achieved by considering a lower bound on the model evidence, which in the case of C-MMNL model is given by

$$
\log p(y|\theta_m) \geq \mathbb{E}_q[\log p(y,z|\theta_m)] - \mathbb{E}_q[\log q(z|\phi)] = \mathcal{L}(\phi, \theta_m).
$$

Maximizing the evidence lower bound $\mathcal{L}(\phi, \theta_m)$, or “ELBO” for short, w.r.t. $\phi$ is then equivalent to minimizing the KL-divergence between $q(z|\phi)$ and $p(y,z|\theta_m)$.

Unfortunately, while $\nabla_\phi \mathbb{E}_q[\log q(z|\phi)]$ can generally be computed analytically given a tractable choice of approximate distribution (e.g. fully-factorised, or mean-field, approximation; Jordan et al. (1999)), computing $\nabla_\phi \mathbb{E}_q[\log p(y,z|\theta_m)]$ exactly is infeasible for the MMNL model described in Section 2.1, regardless of the choice of $q(z|\phi)$. To overcome this issue, we follow a general approach for non-conjugate models using Monte Carlo gradient estimation, as proposed in Rodrigues et al. (2020) for mixed multinomial logit models. In essence, this approach consists in reparameterizing the latent variables in the C-MMNL model, $z$, in terms of a known base distribution and a differentiable transformation. For example, if for a given latent variable $z$ in the model we assume $q(z|\phi) = \mathcal{N}(z|\mu, \sigma^2)$, with $\phi = \{\mu, \sigma\}$, we can reparameterise it as

$$
z \sim \mathcal{N}(z|\mu, \sigma^2) \iff z = \mu + \sigma \epsilon, \quad \epsilon \sim \mathcal{N}(0,1).
$$

We can then compute gradients of an arbitrary function of $z$, $f(z)$, such as the ELBO, w.r.t. $\phi$ by using a Monte Carlo approximation with draws from the base distribution $\mathcal{N}(0,1)$ in the example above, since

$$
\nabla_\phi \mathbb{E}_{\epsilon}[f(z)] \approx \mathbb{E}_{\mathcal{N}(\epsilon[0,1])}[\nabla_\phi f(\mu + \sigma \epsilon)].
$$

As for the neural network parameters $\theta_m$, they can be estimated using a type-II maximum likelihood approach. At convergence of VI, the bound in Eq. 2 is tight, thus making the ELBO a good proxy to the log marginal likelihood of the model $\log p(y|\theta_m)$. We can therefore find maximum likelihood estimates of the neural parameters by maximizing $\mathcal{L}(\phi, \theta_m)$ w.r.t. $\theta_m$. In practice, the optimization of the ELBO w.r.t. $\phi$ and $\theta_m$ is done jointly using stochastic gradient descent, where mini-batches of data are used to obtain an approximation to the gradients. As the ELBO is not a convex function, these noisy gradient approximations not only speed-up the optimization procedure, but they often turn out to help it escape poor local optima (Hoffman et al. 2013). Moreover, the use of mini-batches makes it possible to scale VI to large datasets that don’t fit in memory, as is the case of the route choice dataset considered in
Section 3.2. Convergence is assumed when the ELBO does not improve for a number of consecutive iterations.

The proposed C-MMNL model and the estimation procedure described above were implemented in Python and PyTorch, thereby allowing it to make use of GPU acceleration which, as shown in Rodrigues et al. (2020) and also Arteaga et al. (2022), can lead to significant improvements in terms of computational efficiency, especially when large datasets are considered, as we do in Section 3. The source code for estimating the C-MMNL model, including a few examples, is available at: [code in preparation].

3 Case studies: results and discussion

In this section we apply the proposed Context-aware Bayesian Mixed Multinomial Logit Model to analyse travel behaviour. We present two real-world case studies - a travel mode choice model and a bicycle route choice model. While the first example should mainly explain the idea of the proposed C-MMNL model, the second example shows its advantages regarding the scalability and results interpretation.

3.1 Mode choice model

Data  We consider a real-world case study based on the London Passenger Mode Choice dataset provided by Hillel et al. (2018). This case study investigates mode choice on an urban multi-modal transport network. The dataset consists of revealed preferences (RP) and the authors have developed a data fusion framework to add individual-specific mode-alternative level-of-service (LOS) variables (e.g. in-vehicle travel time, public transport fares, fuel cost, etc.) to historical trip records (Hillel et al. 2018). Based on this data, we consider the problem of mode choice between three alternatives: walking, public transport (PT), and car. Given the uncertainty regarding the choice set associated with each trip, a few simplifying assumptions are made (Rodrigues (2022) describes the data preprocessing procedure in detail). The resulting dataset consists of a total of 43,778 trips made by 18,020 individuals.

Context data  Besides the standard travel time and travel cost information, the London Passenger Mode Choice dataset contains information about the trip purpose (home-based work “HBW”, home-based education “HBE”, home-based other “HBO”, employers’ business “B”, non-home-based other “NHBO”). For the mode choice problem, we consider two context variables: trip purpose and weather (rain). We created a new binary variable “is commute” referring to commute trips - its value equals 1 if the trip purpose attribute has one of the following values: “HBW”, “HBE”, “B”, and 0 otherwise. As for the weather, we created an “is rain” variable referring to whether it was raining within the hour preceding the start of the trip or not. This information is based on attributes from OpenWeatherMap1 data.

Example  In this example, we compare two approaches: i) A conventional approach where context variables are included via interaction with mode-related attributes, ii) The proposed C-MMNL approach with mapping of context information on parameter shifts.

For the first approach, we begin with estimating a simple MMNL model, where the variables associated with travel cost (TC) and travel time (TT) are treated as random parameters and the alternative-specific constants are treated as fixed effects. We assume a simple form of utility functions as follows (the notation for random parameters is simplified for readability reasons):

\[
V_{\text{car}} = \beta_{\text{TC}} \cdot x_{\text{TC}}^{\text{car}} + \beta_{\text{TT}} \cdot x_{\text{TT}}^{\text{car}} \\
V_{\text{pt}} = \text{ASC}_{\text{pt}} + \beta_{\text{TC}} \cdot x_{\text{TC}}^{\text{pt}} + \beta_{\text{TT}} \cdot x_{\text{TT}}^{\text{pt}} \\
V_{\text{walk}} = \text{ASC}_{\text{walk}} + \beta_{\text{TC}} \cdot x_{\text{TC}}^{\text{walk}} + \beta_{\text{TT}} \cdot x_{\text{TT}}^{\text{walk}}
\]  

Furthermore, we are interested in how the preference parameters are influenced by context attributes. Therefore, we interact the respective parameters accordingly, and extend the respective basic

---

1https://www.openweathermap.org/
utility specifications by interaction terms: \( ASC_{\text{walk, rain}}, ASC_{\text{walk, com}}, ASC_{\text{pt, rain}}, ASC_{\text{pt, com}}, \beta_{\text{TC, rain}}, \beta_{\text{TC, com}}, \) and \( \beta_{\text{TT, rain}}, \beta_{\text{TT, com}}, \) and obtain the following set of utility specifications:

\[
V_{\text{car}} = \beta_{\text{TC}} \cdot x_{\text{TC}} + \beta_{\text{TT}} \cdot x_{\text{TT}} + \beta_{\text{TC, com}} \cdot x_{\text{TC}} \cdot 1_{\text{com}} + \beta_{\text{TT, com}} \cdot x_{\text{TT}} \cdot 1_{\text{com}} \\
+ \beta_{\text{TC, rain}} \cdot x_{\text{TC}} \cdot 1_{\text{rain}} + \beta_{\text{TT, rain}} \cdot x_{\text{TT}} \cdot 1_{\text{rain}}
\]

\[
V_{\text{pt}} = ASC_{\text{pt}} + ASC_{\text{pt, rain}} + ASC_{\text{pt, com}} + \beta_{\text{TC}} \cdot x_{\text{TC}} + \beta_{\text{TT}} \cdot x_{\text{TT}} + \beta_{\text{TC, com}} \cdot x_{\text{TC}} \cdot 1_{\text{com}} + \beta_{\text{TT, com}} \cdot x_{\text{TT}} \cdot 1_{\text{com}}
+ \beta_{\text{TC, rain}} \cdot x_{\text{TC}} \cdot 1_{\text{rain}} + \beta_{\text{TT, rain}} \cdot x_{\text{TT}} \cdot 1_{\text{rain}}
\]

\[
V_{\text{walk}} = ASC_{\text{walk}} + ASC_{\text{walk, rain}} + ASC_{\text{walk, com}} + \beta_{\text{TC}} \cdot x_{\text{TC}} + \beta_{\text{TT}} \cdot x_{\text{TT}} + \beta_{\text{TC, com}} \cdot x_{\text{TC}} \cdot 1_{\text{com}} + \beta_{\text{TT, com}} \cdot x_{\text{TT}} \cdot 1_{\text{com}}
+ \beta_{\text{TC, rain}} \cdot x_{\text{TC}} \cdot 1_{\text{rain}} + \beta_{\text{TT, rain}} \cdot x_{\text{TT}} \cdot 1_{\text{rain}}
\]

where \( 1_{\text{com}} \) equals 1 if a trip is a commute trip, and 0 otherwise, and \( 1_{\text{rain}} \) equals 1 if a trip started in rain, and 0 otherwise.

The results of these two models - one with and one without the interaction terms - are presented in Table 1.

| Attribute                  | MMNL          | MMNL with interactions |
|----------------------------|---------------|------------------------|
| ASC_{\text{pt}}           | 1.796***      | 1.612***               |
| ASC_{\text{pt, rain}}     | —             | +0.136*                |
| ASC_{\text{pt, com}}      | —             | +0.570***              |
| ASC_{\text{walk}}         | 2.964***      | 2.796***               |
| ASC_{\text{walk, rain}}   | —             | -0.081                 |
| ASC_{\text{walk, com}}    | —             | +0.471***              |
| \beta_{\text{TT}}        | -12.816***    | -12.689***             |
| \beta_{\text{TT, com}}   | —             | +0.206                 |
| \beta_{\text{TT, rain}}  | —             | -0.471**               |
| \beta_{\text{T}}         | -0.824***     | -0.961***              |
| \beta_{\text{T, com}}    | —             | +0.509***              |
| \beta_{\text{T, rain}}   | —             | -0.140**               |
| \beta_{\text{TT}} (sd)   | 7.417***      | 7.291***               |
| \beta_{\text{T}} (sd)    | 1.427         | 1.389                  |

No. of observations: 18,020 18,020

Tab. 1: Mode choice model estimations from simple MMNL model and MMNL model with interaction terms. The significance levels of coefficients are given by stars: * - 5%, ** - 1% and *** - 0.1%.

In the second approach, we return to the simple utility specifications from Eq. 5 – 7, and additionally input the context information according to the C-MMNL model framework described in Section 2.2, instead of extending the utilities with further interaction parameters, as in Eq. 8 – 10. We now apply the learned model structure to analyse four different scenarios resulting from all possible combinations for variables “is commute” and “is rain”. These results are presented in Table 2 in the form of additive shifts to the “zero” scenario (where the value of both context attributes equals 0).
Tab. 2: C-MMNL model results

The signs and the magnitude of the main parameters in general coincide in the two applied models (MMNL model with interaction terms and the proposed C-MMNL model). One should be aware that the interpretation of the direct results from MMNL model with interaction terms and the proposed C-MMNL model slightly differs. By means of interacting the variables in the first approach, we account for the marginal effect of the context variable. On the other hand, in the C-MMNL model, the effect of a single variable is not isolated, but rather a whole scenario must be defined where all context variable receive a value from the domain. The computation of the average effect of a particular context variable is, however, still possible by averaging over all the possible values of the remaining context variables (for discrete variables). Furthermore, in the C-MMNL approach we allow for non-linear relationship between context variables (we observe how the effect of rain is mitigated by the effect of commute). With the conventional procedure, this would have required including further interaction terms and testing multiple model specifications. For example, the parameters for the two commute scenarios barely differ whether there is rain or not, indicating that rain does not effect preferences for commuting trips. In the MMNL interaction model, only general effects of rain for each attribute can be observed.

Adding multiple interactions to the MMNL model results in an increase of model complexity, which is reflected in the estimation time (276 seconds for MMNL vs 491 seconds for the MMNL with interactions), while the structure of C-MMNL model does not contribute to the increase of the estimation time that much (317 seconds). This feature becomes more pronounced when working with large datasets. The C-MMNL model shows major advantages over the conventional MMNL model, especially if the modeller does not know which interactions are worth testing, has a very high number of possible interactions, encounters variables of another types than binary, or expects non-linear relationships between variables. These benefits are further illustrated and highlighted on the second example in Section 3.2.

3.2 Bicycle route choice model

GPS data. In this example, we utilised a large-scale, crowdsourced dataset of GPS trajectories collected from advanced bicycle head protection devices from the company Hövding\(^2\). The initial dataset covers the entire Copenhagen metropolitan area in the period from the 16\(^{th}\) September 2019 until 31\(^{st}\) May 2021 and consists of 365,813 trips from 10,049 individuals. We applied a combination of steps to process the GPS trajectories based on Lißner and S. Huber (2021) and Schuessler and Axhausen (2009), which we refer to for further details. The data was map-matched to a highly disaggregated bicycle network based on Open Street Map (OSM\(^3\)) and the choice set was generated by the “categorical” approach from Rasmussen et al. (2021). The analysis is based on a highly detailed bicycle network (420,973 directed links and 324,492 nodes) and takes into account charactersitics such as land-use, elevation, surface conditions, and bicycle infrastructure. Table 4 provides overview of all network attributes included in the model. The final dataset for route choice modelling, restricted to a maximum of 40 trips per cyclist to avoid sample imbalance, consists of 110,083 trips made by 8,555 users.

---

\(^2\)https://www.hovding.com

\(^3\)https://www.openstreetmap.org
Context data  
Ton et al. (2017) list possibly relevant contextual trip attributes for the case of route choice behaviour of cyclists, such as weather, daylight, trip purpose, or cycling season. In some of the bicycle route choice studies, the models account also for additional trip-related context variables, e.g. weather (Prato et al. 2018), trip time (Dane et al. 2019; Khatri et al. 2016; Prato et al. 2018; Shah and Cherry 2021; Ton et al. 2017) or trip purpose (Bernardi et al. 2018; Broach et al. 2012; Dane et al. 2019; Hood et al. 2011). The studies found, for example, that cyclists on commute trips and on trips during peak hours are relatively less likely to detour from the shortest paths compared to cyclists on other utilitarian trips or trips during off-peak hours, respectively. The vast majority of the models have been estimated using the standard Maximum Likelihood Estimation procedure. Recently the researchers have been applying more elaborate methods such as link-based models (Zimmermann et al. 2017), spatial models (Alattar et al. 2021), Bayesian approach (Fitch and Handy 2020), or other Machine Learning techniques (Magnana et al. 2022). By applying the C-MMNL model for bicycle route choice problem, we add to this line of research by proposing a rather simple model formulation that is able to handle complex behavioural patterns in the data.

To better understand cyclists’ behaviour in different context situations, we merged multiple external datasets and incorporated various contextual attributes inspired by the existing literature. Table 3 summarises the context variables attributed to a trip (based on the value at its start point) and accounted for by the final bicycle route choice model.

| Context variable | Source | Description                      |
|------------------|--------|-----------------------------------|
| peak hours       | timestamp | 6-9am/3-6pm, Monday to Friday |
| darkness         | package Sun in Python | based on sunrise/sunset time |
| utilitarian trips| heuristics from Lißner and S. Huber (2021) | utilitarian or leisure purpose |
| amount of rain   | OpenWeatherMap | amount of rain within the last hour before the start of a trip |

Tab. 3: Context variables

Network attributes and model specification  
To create a set of possible determinants influencing cyclists’ route choice behaviour, we processed and manipulated attributes of the bicycle network of the Copenhagen metropolitan area. Table 4 provides more details about each of the categories and their relation to the OSM network. For the infrastructure attributes, we combined three categories: road type, road size, and type of bicycle infrastructure. This resulted in 16 disjoint, clearly defined attributes which helps avoid collinearity issues and aids the interpretation of model results.

We further assume that each of the attributes derived from the network contribute to the utility function linearly, i.e. the systematic utility term is assumed to be a linear function of attributes and a collection of taste parameters. The utility expression associated with alternative \( i \) in choice situation \( t \) for individual \( n \) is defined as:

\[
U_{ntj} = \eta_n x_{ntj} + \beta_{PS} \ln PS_{ntj} + \epsilon_{ntj},
\]

where \( x_{ntj} \) is a set of network attributes, \( PS_{ntj} \) is the path-size coefficient accounting for the overlap between alternatives (Frejinger and Bierlaire 2007), and \( \epsilon_{ntj} \) is the random error (iid, EV1-distributed).

Full Route Choice example  
Table 5 presents four different scenarios, where each of the binary context attributes is present one at a time, and their effect is conditional on the value of the other attributes. Again, the results for the C-MMNL model are included in the form of additive shifts to the “zero” scenario (where the value of all context attributes equals 0). The C-MMNL model reveals, for example, that cyclists on utilitarian trips prefer the shortest routes more than cyclists on leisure (but non-loop) trips. Generally, the context “is utilitarian” results in the most differences in shifts, especially for the infrastructure category. For example, cyclists on utilitarian trips disprefer the interaction with pedestrians even more than they already do on leisure trips. Also, the preference for cycleways is higher for both utilitarian trips and trips within the peak hours. Some of the results might be attributed to high sparsity in the presence of attributes, as in e.g. “Living streets” or “No. of stair segments”. The highest variety in shifts is visible for the land-use attributes. For these attributes no clear effect and a high heterogeneity
| Attribute                              | Unit     | Description (optional)                                                                 |
|---------------------------------------|----------|---------------------------------------------------------------------------------------|
| **Trip characteristics**              |          |                                                                                       |
| Length                                | [m]      | Measure of overlap between alternatives                                               |
| ln(Path-Size)                         |          |                                                                                        |
| **Infrastructure**                    | [m]      |                                                                                        |
| Medium roads w/ protected bicycle tracks |          | Medium roads: OSM tags `highway`={'primary', 'secondary', 'tertiary', 'unclassified'} AND |
| Medium roads w/ painted bicycle lanes  |          |                                                                                        |
| Medium roads w/o bicycle infrastructure |          |                                                                                        |
| Large roads w/ protected bicycle tracks |          | Large roads: OSM tags `highway`={'primary', 'secondary', 'tertiary', 'unclassified'} AND  |
| Large roads w/ painted bicycle lanes   |          |                                                                                        |
| Large roads w/o bicycle infrastructure |          |                                                                                        |
| Residential roads w/ protected bicycle tracks | | Residential roads: OSM tag `highway`='residential'.                                    |
| Residential roads w/ painted bicycle lanes |          |                                                                                        |
| Residential roads w/o bicycle infrastructure |      |                                                                                        |
| Cycleways                             |          | OSM tag `highway`='cycleway'                                                          |
| Footways                              |          | OSM tag `highway`='footway'                                                           |
| Living streets                        |          | OSM tag `highway`='living_street'                                                     |
| Shared paths                          |          | OSM tags `highway`={'path', 'track', 'service'}                                       |
| Pedestrian zones                      |          | OSM tag `highway`='pedestrian'                                                        |
| Stairs                                |          | OSM tag `highway`='steps'                                                             |
| No. of stair segments                 | [1]      |                                                                                        |
| **Elevation gradient**                | [vertical m] | Computed between 10 meter splits of the network                                       |
| Flat                                  |          |                                                                                        |
| Steep uphill (10 – 35 m/km)          |          |                                                                                        |
| Very steep uphill (> 35 m/km)         |          |                                                                                        |
| Steep downhill (10 – 35 m/km)        |          |                                                                                        |
| Very steep downhill (> 35 m/km)       |          |                                                                                        |
| **Intersection type**                 | [1]      |                                                                                        |
| Road hierarchy downgraded             |          | Intersection where the cyclist has right of way                                       |
| Road hierarchy upgraded               |          | Intersection where the cyclist has to yield                                           |
| Roundabouts                           |          |                                                                                        |
| Traffic lights                        |          |                                                                                        |
| **Surface type**                      | [m]      | Based on OSM attributes                                                               |
| Asphalt                               |          |                                                                                        |
| Cobblestones                          |          |                                                                                        |
| Gravel                                |          |                                                                                        |
| **Cycle superhighways**               | [m]      | Based on an external GIS layer                                                        |
| Existing                              |          |                                                                                        |
| Proposed                              |          |                                                                                        |
| **Wrong way**                         | [m]      | Based on OSM attributes                                                               |
| **Land-use (right-hand side)**        | [m]      | Referes to “no land-use” attribute                                                    |
| High-rise urban areas                 |          | High-rise urban areas and the city centre                                             |
| Green areas                           |          | Green restricted areas, parks, and forests                                            |
| Areas near water                      |          |                                                                                        |
| Industrial areas                      |          |                                                                                        |
| Low-rise urban areas                  |          |                                                                                        |
| Open landscape                        |          |                                                                                        |

Tab. 4: Network attributes for route choice modelling
is expected and hence we treat them as random parameters in the model. We also observe the highest variation in context-related behavior shifts for this attribute category, even though the magnitude of the effects is not pronounced.

Applying the proposed C-MMNL model enables behavioural analysis of many scenarios, without increasing the computational burden of the estimation. Please note that the estimation times of both route choice models are similar (4,553 seconds for MMNL vs. 4,836 seconds for C-MMNL), indicating that even though the C-MMNL model allows for flexible interactions between attributes, there is hardly an increase in the complexity of the model. Once the model has learned the structure from the input data, an analysis of different scenarios for context attributes (and their combinations) is possible in almost no time.

**Amount of rain** The fourth contextual variable included in the bicycle route choice C-MMNL model is the amount of rain within the hour preceding the start of the trip, ranging from 0 to around 14 (mean=0.063, sd=0.310). The formulation of the proposed C-MMNL model enables to include context variables of the continuous type and analyse their influence on both fixed and random parameters, including extrapolation on values outside of the domain of the context variable. Figure 2 shows how rain conditions influence cyclists’ preferences for chosen attributes. We present the course of the (standardised) parameter change in the “zero” scenario and observe, for example, that cyclists’ preference for cycleways increases with heavier rain in this case.

![Graph showing relative change of the effect of the contextual variable “Amount of rain” on selected model parameters in the “zero” scenario (values normalised and shifted to values at point 0). Due to data scarcity and for illustration purposes, the plot domain was limited.](image)

**Fig. 2:** Relative change of the effect of the contextual variable “Amount of rain” on selected model parameters in the “zero” scenario (values normalised and shifted to values at point 0). Due to data scarcity and for illustration purposes, the plot domain was limited.

### 4 Conclusion

This paper presented an effective way to include contextual effects in the model while retaining the structure and properties of the classical Mixed Multinomial Logit model. The proposed Context-aware Bayesian Mixed Multinomial Logit (C-MMNL) Model uses a neural network to systematically map the context corresponding to the choice occasion to the expected value of the context-dependent shifts on the base parameters, in all combinations. While delivering the results in an easily interpretable form, the model allows for complex interactions between attributes of different datatypes. It learns the underlying structure in the input data and the analysis of different scenarios for context variables is possible without an additional computational cost. This saves the modeller the time usually spent on testing multiple parameter interactions. The source code for estimating the C-MMNL model, including a few examples, was made publicly available.
| Attribute                                      | MMNL  | C-MMNL  |
|-----------------------------------------------|-------|---------|
| **Trip characteristics**                      |       |         |
| Length                                        | -8.300*** | -8.682  | -0.984 |
| In(Path-Size)                                 | 0.540* | 0.560   | -0.047 | +0.042 |
| **Infrastructure**                            |       |         |
| Medium roads w/ protected bicycle tracks      | ref.  | -       | -       | -       |
| Medium roads w/ painted bicycle lanes         | -0.610*** | -0.694  | +0.156 |
| Large roads w/o bicycle infrastructure       | -0.926*** | -1.111  |        |        |
| Large roads w/ protected bicycle tracks      | 0.113 |         |        |        |
| Large roads w/ painted bicycle lanes          | -2.118*** | -1.969  | -0.231 |
| Large roads w/o bicycle infrastructure       | -1.894*** | -1.874  | -0.194 |
| Residential roads w/ protected bicycle tracks| -0.662*** | -0.798  | +0.059 |
| Residential roads w/ painted bicycle lanes   | 0.347*  | 0.295   | +0.034 | -0.101 |
| Residential roads w/o bicycle infrastructure | -1.650*** | -1.852  |        |        |
| Cycleways                                     | 0.328*** | 0.255   | +0.017 | +0.097 |
| Footways                                      | -3.665*** | -3.766  |        |        |
| Living streets                                | -0.481*  | -0.888  | +0.068 | +0.443 |
| Shared paths                                  | -1.563*** | -1.709  |        |        |
| Pedestrian zones                              | -2.737*** | -2.898  | -0.197 |
| Stairs                                        | -0.117 |         |        |        |
| No. of stair segments                         | -1.225*** | -1.421  | +0.147 |
| **Elevation gradient**                        |       |         |
| Flat                                          | ref.  | -       | -       | -       |
| Steep uphill (10 – 35 m/km)                  | -0.094*  | -0.868  | +0.068 | +0.443 |
| Very steep uphill (> 35 m/km)                | -0.137** | -0.183  |        | +0.010 |
| **Intersection type**                         |       |         |
| Road hierarchy downgraded                    | -0.201 |         |        |        |
| Road hierarchy upgraded                      | -0.349*  | -0.396  | -0.048 |
| Roundabouts                                  | -0.098 |         |        |        |
| Traffic lights                                | -0.064 |         |        |        |
| **Surface type**                              |       |         |
| Asphalt                                       | ref.  | -       | -       | -       |
| Cobblestones                                  | -2.249*** | -2.336  |        |        |
| Gravel                                        | -1.010*** | -1.057  | +0.061 |
| **Cycle superhighways**                      |       |         |
| No classification                            | ref.  | -       | -       | -       |
| Existing                                      | 0.272 |         |        |        |
| Proposed                                      | 0.220 |         |        |        |
| **Wrong way**                                 |       |         |
| High-rise urban areas                         | ref.  | -       | -       | -       |
| Green areas                                   | -0.119 |         |        |        |
| Areas near water                              | 0.477*** | 0.841   | -0.044 | +0.061 |
| Industrial areas                              | -0.396*** | -0.258  | -0.036 | -0.020 |
| Low-rise urban areas                          | -0.318*** | -0.190  | -0.038 | +0.071 |
| Open landscape                                | -0.555*** | -0.269  | -0.037 | +0.125 |
| High-rise urban areas (sd)                    | ref.  | -       | -       | -       |
| Green areas (sd)                              | 1.581*** | 1.638   |        |        |
| Areas near water (sd)                         | 2.683*  | 1.638   |        |        |
| Industrial areas (sd)                         | 3.419**  | 1.941   |        |        |
| Low-rise urban areas (sd)                     | 3.479*  | 1.531   |        |        |
| Open landscape (sd)                           | 4.211**  | 1.988   |        |        |
| **No. of observations:**                      |       |         |
| No. of individuals                            | 110,083 | 8,555   | 38     | 38     |
| No. of utility parameters:                    |       |         |
| Log likelihood                                | -169,582.0 | -167,187.9 | 46.34  | 47.76  |
| % of correct predictions:                    |       |         |
| Average prob. of the choice:                 | 0.34   | 0.35    |        |        |
| Estimation time [s]:                          | 4.553  | 4.836   |        |        |

Tab. 5: Estimated bicycle route choice models. The significance levels of coefficients are given by stars: * - 5%, ** - 1% and *** - 0.1%. Only parameters that are significant in MMNL model are analysed in C-MMNL model scenarios and only shifts indicating relative change of at least 5% are included. The value of variable "amount of rain" is 0 for all scenarios.
We applied the proposed C-MMNL model on two real-life datasets for mode choice and bicycle route choice problems. The case studies illustrate the major advantages of the C-MMNL model in terms of scalability and flexibility, as well as the interpretability of the results.

Future work could encompass further investigation of the Neural Network architecture or consideration of various distribution types for some of the priors for the preference parameters suggested e.g. in Prato et al. (2018). The assumption that all individuals shift their preference parameters in a similar way when faced with a given choice context could also be relaxed, allowing for complex interactions between the individual characteristics and the context information.

References

Alattar, M. A., C. Cottrill, and M. Beecroft (2021). “Modelling cyclists’ route choice using Strava and OSMnx: A case study of the City of Glasgow”. In: Transportation research interdisciplinary perspectives 9, p. 100301. DOI: 10.1016/j.trip.2021.100301.

Arteaga, C., J. Park, P. B. Beeramole, and A. Paz (2022). “xlogit: An open-source Python package for GPU-accelerated estimation of Mixed Logit models”. In: Journal of Choice Modelling 42, p. 100339. DOI: 10.1016/j.jocm.2021.100339.

Barnard, J., R. McCulloch, and X. L. Meng (2000). “Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage”. In: Statistica Sinica 10.4, pp. 1281–1311. URL: https://www.jstor.org/stable/24306780.

Becker, F., M. Danaf, X. Song, B. Atasoy, and M. Ben-Akiva (2018). “Bayesian estimator for logit mixtures with inter-and intra-consumer heterogeneity”. In: Transportation Research Part B: Methodological 117, pp. 1–17. DOI: 10.1016/j.trb.2018.06.007.

Bernardi, S., L. La Paix Puello, and K. Geurs (2018). “Modelling route choice of Dutch cyclists using smartphone data”. In: Journal of transport and land use 11.1, pp. 883–900. DOI: 10.5198/jtlu.2018.1143.

Boyd, J. H. and R. E. Mellman (1980). “The effect of fuel economy standards on the US automotive market: an hedonic demand analysis”. In: Transportation Research Part A: General 14.5-6, pp. 367–378.

Broach, J., J. Dill, and J. Gliebe (2012). “Where do cyclists ride ? A route choice model developed with revealed preference GPS data”. In: Transportation Research Part A 46.10, pp. 1730–1740. DOI: 10.1016/j.tra.2012.07.005.

Congdon, P. (2007). Bayesian statistical modelling. John Wiley & Sons.

Danaf, M., B. Atasoy, and M. Ben-Akiva (2019). Logit mixture with inter and intra-consumer heterogeneity and flexible mixing distributions”. In: Journal of Choice Modelling 35.August 2019, p. 100188. DOI: 10.1016/j.jocm.2019.100188.

Dane, G., T. Feng, F. Luub, and T. Arentze (2019). “Route choice decisions of E-bike users: Analysis of GPS tracking data in the Netherlands”. In: International Conference on Geographic Information Science. Springer, pp. 109–124. DOI: 10.1007/978-3-030-14745-7_7.

Elshiewy, O., G. Zenetti, and Y. Boztug (2017). “Differences Between Classical and Bayesian Estimates for Mixed Logit Models: A Replication Study”. In: Journal of Applied Econometrics 32.2, pp. 470–476. DOI: 10.1002/jae.2513.

Fitch, D. T. and S. L. Handy (2020). “Road environments and bicyclist route choice: The cases of Davis and San Francisco, CA”. In: Journal of Transport Geography 85.April 2019, p. 102705. DOI: 10.1016/j.jtrangeo.2020.102705.

Frejinger, E. and M. Bierlaire (2007). “Capturing correlation with subnetworks in route choice models”. In: Transportation Research Part B: Methodological 41.3, pp. 363–378. DOI: 10.1016/j.trb.2006.06.003.

Hess, S. and J. M. Rose (2009). “Allowing for intra-respondent variations in coefficients estimated on repeated choice data”. In: Transportation Research Part B: Methodological 43.6, pp. 708–719. URL: 10.1016/j.trb.2009.01.007.

Hilbe, J. M. (2009). Data Analysis Using Regression and Multilevel/Hierarchical Models. Vol. 30. Book Review 3. Cambridge university press. DOI: 10.18637/jss.v030.b03.

Hillel, T., M. Z. E. B. Elshafie, and Y. Jin (2018). “Recreating passenger mode choice-sets for transport simulation: A case study of London, UK”. In: Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction 171.1, pp. 29–42. DOI: 10.1680/jsmic.17.00018.
Hoffman, M. D., D. M. Blei, C. Wang, and J. Paisley (2013). “Stochastic variational inference”. In: Journal of Machine Learning Research.

Hood, J., E. Sall, and B. Charlton (2011). “A GPS-based bicycle route choice model for San Francisco, California”. In: Transportation letters 3.1, pp. 63–75. DOI: 10.3328/TL.2011.03.01.63–75.

Huber, J. and K. Train (2001). “On the similarity of classical and Bayesian estimates of individual mean partworths”. In: Marketing Letters 12.3, pp. 259–269. URL: 10.1023/A:101112092698.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul (1998). “An introduction to variational methods for graphical models”. In: Learning in graphical models. Springer, pp. 105–161. DOI: 10.1023/A:1007665907178.

— (1999). “Introduction to variational methods for graphical models”. In: Machine Learning 37.2, pp. 183–233. DOI: 10.1023/A:1007665907178.

Khatri, R., C. R. Cherry, S. S. Nambisan, and L. D. Han (2016). “Modeling route choice of utilitarian bikeshare users with GPS data”. In: Transportation research record 2587.1, pp. 141–149. DOI: 10.3141/2587-17.

Krueger, R., P. Bansal, M. Bierlaire, R. A. Daziano, and T. H. Rashidi (2019). “Variational Bayesian Inference for Mixed Logit Models with Unobserved Inter- and Intra-Individual Heterogeneity”. In: January. arXiv: 1905.00419.

Lewandowski, D., D. Kurowicka, and H. Joe (2009). “Generating random correlation matrices based on vines and extended onion method”. In: Journal of Multivariate Analysis 100.9, pp. 1989–2001. DOI: 10.1016/j.jmva.2009.04.008.

Lißner, S. and S. Huber (2021). “Facing the needs for clean bicycle data—a bicycle-specific approach of GPS data processing”. In: European Transport Research Review 13.1, pp. 1–14.

Magnana, L., H. Rivano, and N. Chiabaut (2022). “Implicit GPS-based bicycle route choice model using clustering methods and a LSTM network”. In: PLoS one 17.3, e0264196. DOI: 10.1371/journal.pone.0264196.

McFadden, D. and K. Train (2000). “Mixed MNL models for discrete response”. In: Journal of applied Econometrics 15.5, pp. 447–470.

Prato, C. G., K. Halldórsdóttir, and O. A. Nielsen (2018). “Evaluation of land-use and transport network effects on cyclists’ route choices in the Copenhagen Region in value-of-distance space”. In: International Journal of Sustainable Transportation 12.10, pp. 770–781. DOI: 10.1080/15568318.2018.1437236.

Rasmussen, T. K., M. Lukaw ska, and M. Paulsen (2021). An easily interpretable and efficient choice set generation method for bicycle route choices. Extended abstract submitted to the 5th Cycling Research Board Annual Meeting (CRBAM2021), Copenhagen, Denmark.

Rodrigues, F. (2022). “Scaling Bayesian inference of mixed multinomial logit models to large datasets”. In: Transportation research part B: methodological 158, pp. 1–17. DOI: 10.1016/j.trb.2022.01.005.

Rodrigues, F., N. Ortelli, M. Bier laire, and F. C. Pereira (2020). “Bayesian automatic relevance determination for utility function specification in discrete choice models”. In: IEEE Transactions on Intelligent Transportation Systems. DOI: 10.1109/TITS.2020.3031965.

Schuessler, N. and K. Axhausen (2009). “Processing raw data from global positioning systems without additional information”. In: Transportation Research Record 2105, pp. 28–36. DOI: 10.3141/2105-04.

Shah, N. R. and C. R. Cherry (2021). “Different safety awareness and route choice between frequent and infrequent bicyclists: findings from revealed preference study using bikeshare data”. In: Transportation research record 2675.11, pp. 269–279. DOI: 10.1177/03611981211017136.

Ton, D., O. Cats, D. Duives, and S. Hoogendoorn (2017). “How do people cycle in amsterdam, netherlands?: Estimating cyclists’ route choice determinants with gps data from an urban area”. In: Transportation research record 2662.1, pp. 75–82. DOI: 10.3141/2662–09.

Train, K. (2001). “A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit”. In: University of California, Berkeley, pp. 1–13. URL: https://emlab.berkeley.edu/~train/compare.pdf.

Washington, S., P. Congdon, M. G. Karlaftis, and F. L. Mannering (2009). “Bayesian multinomial logit”. In: Transportation Research Record 2136, pp. 28–36. DOI: 10.3141/2136-04.

Zimmermann, M., T. Mai, and E. Frejinger (2017). “Bike route choice modeling using GPS data without choice sets of paths”. In: Transportation research part C: emerging technologies 75, pp. 183–196. DOI: 10.1016/j.trc.2016.12.009.