ABM3 BIKE ROUTE CHOICE MODEL

January 2026

Prepared for San Diego Association of Governments

Report Title:

ABM3 Bike Route Choice Model

Report Prepared by:

RSG

Report Prepared for:

San Diego Association of Governments
1011 Union Street
Suite 400
San Diego, CA 92101

For additional information regarding this report, or for questions about permissions or use of findings contained therein, please contact:

RSG (Headquarters)
55 Railroad Row
White River Junction, VT 05001
(802) 295-4999
www.rsginc.com

ABM3 BIKE ROUTE CHOICE MODEL
January 2026

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

The San Diego Association of Governments (SANDAG) bicycle route choice model was developed to address policy questions related to bicycle infrastructure planning and evaluation. This report documents the latest implementation of the model as part of the Activity-Based Model version 3 (ABM3) framework. The work builds upon the previous ABM2 bicycle route choice model, extending its functionality and improving maintainability through a new Python-based implementation.

In route choice modeling, utility represents the overall attractiveness of an alternative, combining observable factors such as travel time, comfort, and safety with unobserved influences. In this implementation, a logit model converts these utilities into choice probabilities, with higher-utility routes being chosen more often but not exclusively. The logsum, the output of such a model, is the logarithm of the sum of alternatives' exponentiated utilities and represents the expected maximum utility across all available routes, serving as a compact measure of overall accessibility or choice quality. In bicycle routing, a stochastic route choice model represents the idea that cyclists do not always choose the single "best" path deterministically but instead choose among multiple feasible routes with probabilities.

Formally, each possible route between origin and destination has an associated utility, and the probability that a cyclist chooses a given route depends on that utility relative to alternatives. Randomness captures:

- Differences in individual preferences (risk tolerance, fitness, comfort)
- Imperfect knowledge of the network
- Unobserved factors (weather perception, traffic stress tolerance)
- Day-to-day variability in behavior

Thus, two cyclists (or the same cyclist on different days) may select different routes even under identical network conditions. A stochastic route choice model thus improves realism over all-or-nothing (AON) routing, especially for bicycle networks with many similar-quality paths, as AON routing computes a single "shortest" or "least-cost" path (e.g., minimum travel time, distance, or generalized cost), then assigns all trips between an origin-destination (OD) pair. The AON approach is deterministic and can lead to small changes in inputs to cause dramatic changes in routing, thus minimizing its usefulness for policy analysis, especially because it does not allow for explicit modeling of comfort, stress, safety, or other preferences.

The bicycle route choice model generates alternative bike paths between OD pairs using a generalized utility function applied to the regional bicycle network. For each OD pair, a set of feasible paths is identified, and a logsum is computed across these alternatives to represent the level of service for bicycle travel. This logsum serves as an input to SANDAG's ABM, influencing mode choice calculations.

Key enhancements described in this document include:

- **Re-implementation in Python** for improved transparency, maintainability, and integration with ABM3 workflows.
- **Updated model structure and coefficients**, informed by a recent literature review.
- **Refined path generation algorithms**, incorporating stochastic resampling, path size factors, and strategies to balance computational efficiency with behavioral realism.

This document provides a summary of the methodology, details of the newly implemented coefficients, and results comparing the new logsums with the previous ABM2+ implementation. Source code and configuration files for the ABM3 bike route choice model are maintained in the SANDAG ABM repository. This report serves as a technical reference for model users and stakeholders.

2.0 IMPLEMENTATION

This section focuses on the implementation details of the SANDAG ABM3 bike model and is broken down into two parts: a methodology section describing how the bike route choice code works, and a second section focusing on the implemented utility calculations deciding the bike route choice decisions.

2.1 METHODOLOGY

The SANDAG bike route choice model was ported to Python to improve maintainability, transparency, and integration with modern data science workflows. The implementation follows the ABM2+ approach based on Dijkstra's algorithm but leverages an optimized version provided by SciPy, replacing the custom Java implementation used previously. This change reduces complexity and improves runtime performance.

General Steps of the Bike Route Choice Model

1. **Network Preparation**
    - Read SANDAG bicycle network shapefiles for nodes and links.
    - Transform the network into a graph representation using NetworkX
    - Attribute links and traversals with fields required for utility calculations (e.g., length, slope, facility type).
    - Transform the graph to treat links as graph nodes and traversals as graph edges to conform to the network structure required by SciPy's Dijkstra's algorithm.
2. **Utility Specification**
    - Apply a generalized cost function based on link attributes and model coefficients. Utility terms include bike infrastructure, elevation gain, distance, roadway type, etc.
3. **Path Generation**
    - Use Dijkstra's algorithm to identify shortest paths under stochastic generalized costs.
    - Perform iterative resampling with randomized coefficient vectors to generate a diverse set of feasible paths.
    - Apply path size factors to account for route overlap.
4. **Logsum Calculation**
    - Compute logsums across all generated paths for each origin–destination pair. This is performed at both the TAZ and MGRA level.
    - Output logsums as level-of-service measures for bicycle travel, feeding into ABM mode choice and accessibility models.

Primary Python Dependencies

- **NumPy** – Numerical operations and vectorized computations.
- **SciPy** – Optimized Dijkstra's algorithm implementation for shortest path search.

- **Pandas** – Data handling and tabular processing of network attributes.
- **NetworkX** – Graph representation and traversal operations for the bicycle network.

SANDAG Bike Network

The SANDAG bike network is represented in link and node shapefiles based on a historical OpenStreetMap network. The network consists of 154,468 nodes, of which 24,321 nodes are master geographic reference area (MGRA) centroids and 4,947 are transportation analysis zone (TAZ) centroids; and 183,986 bidirectional links, of which 29,268 are centroid connectors. The bidirectional links are duplicated to create directional links of two classes: A and B links. The relevant attributes of the link and node shapefiles are reflected in Table 1 and Table 2, respectively.

TABLE 1: LINK SHAPEFILE ATTRIBUTES

FIELD	DATA TYPE	DESCRIPTION
A	Integer	Tail node ID for A links / Head node ID for B links
B	Integer	Head node ID for A links / Tail node ID for B links
Distance	Float	Length of link in feet
AB_Gain	Integer	Nonnegative elevation gain for A links, in feet
BA_Gain	Integer	Nonnegative elevation gain for B links, in feet
ABBikeClas	Integer	Bicycle classification for A links
BABikeClas	Integer	Bicycle classification for B links
AB_Lanes	Integer	Number of roadway lanes on A links[1]
BA_Lanes	Integer	Number of roadway lanes on B links[1]
Func_Class	Integer	Roadway functional classification of the link
Bike2Sep	Boolean	Indicates the presence of a cycle track facility on the link
Bike3Blvd	Boolean	Indicates the presence of a bike boulevard facility on the link
SPEED	Integer	Speed limit of the adjacent roadway, in miles per hour

[1] The data quality of the _Lanes field is known to be poor, as several instances have been identified where the value is inaccurate.

TABLE 2: NODE SHAPEFILE ATTRIBUTES

FIELD	DATA TYPE	DESCRIPTION
NodeLev_ID	Integer	Node ID
MGRA	Integer	If node is an MGRA, the MGRA ID (= node ID – 100,000,000), else 0
TAZ	Integer	If node is a TAZ, the TAZ ID (= node ID – 200,000,000), else 0
Signal	Boolean	Indicates whether the node is a signalized intersection

Network Transformation

The network transformation process converts raw node and link shapefiles into a graph representation suitable for path search and utility evaluation. This involves generating traversals through the convolution of incoming and outgoing links, then attributing the links and traversals such that they are ready for the utility calculations. The final step is converting bike links to graph nodes and the traversals to graph edges to become compatible with the Dijkstra's algorithm implementation provided by the SciPy dependency.

Step 1: Preparing the Link and Node Framework

The first stage reads the node and link shapefiles and reformats them into a consistent structure. Links are represented as directional edges, meaning each physical segment is duplicated to allow travel in both directions. This ensures that routing algorithms can evaluate asymmetric conditions such as slope or facility type. Nodes are standardized to include essential spatial coordinates and control attributes, such as signalization and centroid status, which influence traversal penalties.

Step 2: Establishing Directionality and Geometry

Edges are attributed with measures of length, slope, and facility characteristics. Each directional edge is assigned a heading value derived from its endpoints' geometric coordinates. These headings form the basis for later turn classification, allowing the model to distinguish between straight-through movements, left turns, right turns, and reversals. By embedding these geometric relationships early, the network becomes capable of supporting nuanced behavioral penalties without requiring complex geometry checks during path search.

Step 3: Generating Traversals

Traversals represent movements through an intersection from one incoming edge to an outgoing edge. They are constructed by pairing edges that share a common node, excluding U-turns. This process effectively creates a matrix of possible turning movements at each junction. Traversals are then attributed with angular relationships between the paired edges, normalized to a consistent range to facilitate classification. These angular measures, combined with the number of available outbound legs, allow the model to infer intersection type.

Step 4: Deriving Movement Characteristics

Once the traversal set is established, additional attributes are computed to reflect operational constraints and behavioral considerations. These include indicators for signalized intersections, unsignalized left turns, and crossings of major or minor arterials. The procedure aggregates information from the adjacent links, such as the functional class of connected links and the count of arterial crossings, to approximate the difficulty or risk associated with a movement. These derived attributes feed directly into the utility function, enabling penalties for movements that are less desirable from a cyclist's perspective. Computing these more complex traversal attributes occurs in software instead of the utility specification files because they do not fit well with the one-line limitations of the spec files. Details of the traversal attributes are discussed in the following section on Coefficient Transfer.

Step 5: Consolidating for Path Search

The final output consists of three integrated tables: nodes, edges, and traversals. Nodes provide spatial and control context, edges define the directional connectivity and physical characteristics of the network, and traversals capture intersection-level complexity. Together, these structures form a graph representation optimized for shortest-path algorithms while preserving the detail necessary for realistic route choice modeling.

To enable efficient path search using SciPy's implementation of Dijkstra's algorithm, the network is reoriented so that links become graph nodes and traversals become graph edges. This inversion is necessary because SciPy's solver operates on adjacency matrices where rows and columns represent vertices, and connections between them represent allowable transitions. In the context of bicycle routing, the behavioral penalties associated with intersection movements, such as left-turn disbenefits or signalized delays, apply not to the link itself but to the transition between links. Treating traversals as edges in the graph allows these movement-specific costs to be embedded directly in the adjacency matrix, ensuring that shortest-path calculations reflect both segment-level and intersection-level conditions. This design also supports efficient batch processing of origin-destination searches to help with runtime and memory constraints.

Resampling and Coefficient Fuzzing

For each origin, the model samples paths to all relevant destinations by repeatedly applying Dijkstra's algorithm under a stochastic generalized cost framework. Dijkstra's algorithm is a classic shortest-path algorithm used to find the minimum-cost route between a given origin node and all other nodes in a network. It works by iteratively expanding outward from the origin, always selecting the next unvisited node with the lowest accumulated cost, and updating the costs of its neighbors until all reachable nodes have been evaluated. Because it guarantees the optimal (lowest-cost) path under non-negative link costs, it is widely used in transportation, routing, and network analysis. (This restriction of using non-negative link costs was a key requirement of selecting bike model coefficients, as discussed in the next section.) For this bike route choice model, Dijkstra's algorithm is used because it deterministically finds lowest-cost routes based on generalized cost efficiently on a large network.

Rather than using fixed coefficients, the algorithm introduces randomness to reflect variability in cyclist preferences and to generate a diverse set of feasible routes. At the start of each sampling iteration, each coefficient is scaled by ± 10%. These randomized coefficients remain constant within an iteration but differ across iterations, ensuring that each run of Dijkstra's algorithm explores a slightly different cost landscape.

As the search expands outward from the origin, the cost of each link is computed as the dot product of its attributes and the uniformly randomized coefficients, then multiplied by a discrete per-link random link cost multiplier to introduce additional variability at the link level equal to ± 20% of the total link utility. (Note that these random adjustments of 10% and 20% are substantially reduced from the 50% and 70% that were implemented in the ABM2+ Java version and were settled on based on visualization of individual OD pairs and utility traces of the newly implemented specification.)

This double-randomization, at both coefficient and link levels, creates a richer and more behaviorally plausible set of alternative paths. It reflects the reality that cyclists do not perceive utilities uniformly and that route choices vary even under similar conditions.

Path sampling is repeated to create the set of path alternatives. A default value of 10 iterations is set based on what was implemented in the previous Java program[2].

Path Size and Logsum Calculation

The path size term is applied during the post-processing of sampled paths for each origin-destination pair, after the iterative path generation stage has produced a set of feasible alternatives. Its purpose is to adjust the utility of each path to account for overlap among routes in the alternative set. Without this adjustment, multiple paths that share long segments would artificially inflate the perceived choice set, biasing the logsum calculation.

The size of path alternative n_i in an alternative set $\mathbf{n}$ is calculated using the formula:

$$size(n_i) = \sum_{l \in n_i} \frac{edgeLength(l)}{pathLength(n_i) * pathsUsing(l, \mathbf{n})}$$

In the above expression, $pathsUsing$ is the integer number of paths in the alternative set $\mathbf{n}$ which contain link l. Its use derives from the theory of aggregate and elemental alternatives, where a link is an aggregation of all paths that use the link. If multiple routes overlap, their "size" is less than one. If two routes overlap completely, their size will be one-half. Partial overlaps yield intermediate values.

---

[2] The previous Java program had a more sophisticated approach to create a variable number of paths based on a set target for effective path size and the OD distance – OD pairs further apart would allow for more paths to be created, and shorter paths would stop after fewer iterations if the target effective path size was not reached. However, this functionality was turned off and the number of paths generated was set to 10 for all OD pairs. We assume this was due to simplify the outputs and / or reduce runtime.

The logsum represents the composite utility of all feasible paths for an OD pair and is computed as:

$$\text{logsum} = \ln\left(\sum_{i \in n} e^{U_i + \ln(\text{size}_i)}\right)$$

Here, $U_i$is the utility of path i, and $\ln(\text{size}_i)$is the path size adjustment. This integration of path size in the logsum calculation ensures paths with high overlap (low size) contribute less to the logsum, thereby preventing inflated utility from redundant routing options while still rewarding networks that offer genuinely distinct paths.

Logsums are generated for each OD pair and written to the final output file. Intra-zonal logsums were calculated as the product of half the distance to the nearest neighbor zone and the coefficient for residential streets with no bicycle facilities. The bike route choice file is run twice, once for MGRA centroids and again for TAZ centroids leading to separate MGRA and TAZ output files.

2.2 BIKE MODEL COEFFICIENTS

The original multinomial logit (MNL) model documentation as memorialized in the "ABM AT Enhancements" memo included a series of MNL coefficients for network links and turning movements, including utility penalties for overall and uphill distances, distances using various bicycle facility types, wrong-way riding, and riding on controlled-access facilities. These parameters were generally cited as being sourced from studies in Monterrey[3], Portland[4], and San Francisco[5]. Each cited paper is well over a decade old as of this writing, and in addition to the relative age of these studies, there exist several unexplained divergences from the original source document parameter values as well as values asserted without sources known to the project team. Additionally, the ABM2+ model documentation and the values used in the Java implementation were not fully consistent.

These discrepancies motivated a literature review investigating recent bicycle route choice model development, which we attach as Appendix A. After careful assessment of the literature review,

[3] Hood, J., Erhardt, G., Frazier, C., Schenk, A. (2014). "Developing a Stand-Alone Bicycle Facility Emission Reduction Benefit Estimator: Incremental Nested Logit Analysis of Bicycle Trips in California's Monterey Bay Area," presented at the 93rd annual meeting of the Transportation Research Board, Washington, D.C. Jan. 12-16, 2014. http://docs.trb.org/prp/14-0965.pdf

[4] Broach, J., Gliebe, J., Dill, J. (2011). "Bicycle route choice model developed using revealed preference GPS data," presented at the 90th annual meeting of the Transportation Research Board, Washington, D.C. Jan. 23-27, 2011. http://otrec.us/main/document.php?doc_id=858

[5] Hood, J., Sall, E., Charlton, B. (2011). "A GPS-based bicycle route choice model for San Francisco, California." Transportation Letters: The International Journal of Transportation Research, Vol. 3, pp. 63-75. http://www.sfcta.org/sites/default/files/content/IT/CycleTracks/BikeRouteChoiceModel.pdf

a more recent study in Copenhagen[6] was selected for its relative similarity to, and therefore, compatibility with, the pre-existing model structure.

The Copenhagen study upon which we base this implementation consists of utility parameters of varying degrees of compatibility with the pre-existing bike route choice model. In this subsection, we detail, for those parameters which could easily be transferred to the bike model framework, the adjustments required to incorporate those parameters as well as a discussion of the barriers which precluded other model parameters from this implementation.

Coefficient Transfer

In addition to the previously discussed path size term, the Copenhagen model consists of nine separate classes of model parameter: an overall length term, terms related to the amount of upslope present on each edge (i.e., link), a series of terms corresponding to the modeled bike link facility types, traversal (i.e., turning movement) parameters, and five other groups which were omitted (we discuss the omitted classes in the next section). Each retained class required some varying degree of manipulation to comport to the pre-existing model framework.

Link Coefficients

First, the overall distance term is presented in the Copenhagen as a value with units of utiles per kilometer. As a foundational step, we convert this to the equivalent value in utiles per mile. Next, we note that, for the medium- and high-upslope parameters, the utility term was applied by discretizing the OpenStreetMap network at ten-meter (or one-decameter) intervals and evaluating upslope on each ten-meter cell. As a result, we transfer this term by converting from a term with effective units of utiles per decameter to utiles per mile. This is required for coefficient transfer due to a lower resolution of elevation data available for the SANDAG bike network but generally ensures the application of these terms is scale-invariant with respect to network resolution. We retain the existing bounds of the medium and high upslope bins, with medium upslopes defined as a slope between 1% and 3.5% and high upslopes defined as those greater than 3.5%, respectively.

We note that the Copenhagen model segments bike facilities into three roadway classes (large, medium, and residential facilities) based on OpenStreetMap attributes[7] and four non-roadway facility types (cycleway, shared use path, pedestrian zone, and footways) according to the OpenStreetMap link type and lane count attributes. As these attributes are not defined for the SANDAG bike network, we substitute definitions based on roadway functional classification[8], bicycle functional classification, and cycle track status, as detailed below. Roadway facilities are

[6] Łukawska, M., Paulsen, M., Rasmussen, T. K., Jensen, A. F., & Nielsen, O. A. (2023). "A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset." Transportation Research, Part A: Policy and Practice, 176, 103834. https://doi.org/10.1016/j.tra.2023.103834

[7] Large roadways are defined in the Copenhagen model as primary, secondary, tertiary, or unclassified roadways with either three or more lanes in one direction or two or more lanes in both directions. All other primary, secondary, tertiary, or unclassified roadways are classified as medium roadways.

[8] A roadway functional classification is a system that categorizes roads based on the primary role they serve in the transportation network—such as mobility versus access—ranging from high-speed, long-distance facilities to local streets providing direct property access.

grouped both by size and bicycle facility availability, while non-roadway facilities are considered distinct classes of link.

- Roadway size
    - Large roadway – roadway functional classes 3 and 4
    - Medium roadway – roadway functional classes 5 and 6
    - Residential roadway – roadway functional class 7
- Roadway bicycle facility
    - Protected bike lane – is a cycle track[9]
    - Painted bike lane – is not a cycle track and has bicycle functional class 2
    - No bike facility – is not a cycle track and has bicycle functional class other than 2 (example: roadway with sharrows)
- Non-roadway facilities
    - Cycleway – link other than a centroid connector, large, medium, or residential roadway which has painted bike lanes
    - Shared use path – link other than a centroid connector, large, medium, or residential roadway which has bicycle functional class 1
    - Pedestrian zone – link other than a centroid connector, large, medium, or residential roadway which is not a cycleway or shared use path

Note that, due to lack of network distinction, the pedestrian zone and sidewalk terms were consolidated, and the pedestrian zone coefficient was selected due to its larger magnitude10. Finally, a penalty for distance traveling in the wrong direction on roadway facilities was included, defined for all links which do not have protected or painted bike lanes and which have a zero number of lanes. The implemented set of link coefficient values is presented in Table 3.

Traversal Coefficients

Only two traversal coefficients are included in the Copenhagen model specification. The first is a term which penalizes intersections at which a biker must yield. While the original model specification depended on attributes of the relevant OpenStreetMap intersection, we substitute a definition which leans on five of the original six ABM2+ traversal penalty definitions. Specifically,

---

[9] OpenStreetMap defines a cycle track as "separated from the road by curbs, parking lots, grass verges, trees, bollards, or another physical barrier, but […] running parallel and next to the road. In North America this is called a protected bike lane, separated bike lane, greenway, green lane, or [California bikeway] class IV facility." For more details, see *Key:cycleway* at the OpenStreetMap wiki. The SANDAG bike network link attribute Bike2Sep was used to determine cycle track status.

[10] OpenStreetMap defines pedestrian zones as those "reserved to pedestrians or where pedestrians have priority and can walk on the full width of the way," while a sidewalk "can be mapped as a pedestrian [zone] or as a shared [use path]." For more details, see *Guidelines for pedestrian navigation* at the OpenStreetMap wiki. For SANDAG bike network links, these are any non-roadway facility that is not a cycleway or shared-use path.

we define a traversal at an intersection in which a biker must yield as one which meets any of the following criteria:

- Any traversal of a signalized intersection which has a conflicting roadway turning movement (i.e., all except right turns or T-junction through movements)
- An unsignalized left turn from a major arterial (see large roadway definition above)
- An unsignalized left turn from a minor arterial (see medium roadway definition above)
- A crossing of, or left turn onto, a major arterial at an unsignalized intersection
- A crossing of, or left turn onto, a minor arterial at an unsignalized intersection

In the Copenhagen model, a second traversal term penalizes any intersection in which a biker has right-of-way but faces a downgrade in roadway functional classification. We implement this as any such traversal which does not meet the above criteria for the yield penalty and which transitions from one roadway facility (functional classes 3 through 7) to another with a higher roadway functional class. That is, a yield traversal (any of the bulleted situations) is a roadway upgrade while a downgrade is taken to be the opposite. Finally, while the Copenhagen model does not consider centroid connectors, we provide an extra traversal term to preclude bikers traveling from one centroid connector to another, thus preventing teleportation-esque shortcuts across zones' centroids. The coefficient values for traversals are presented in Table 4. A conversion table is provided in Appendix C.

TABLE 3: IMPLEMENTED LINK COEFFICIENTS

COEFFICIENT DESCRIPTION	VALUE
General link length (miles)	-10.93
Distance traveling in opposition to traffic on one-way roadways without bike lanes	-3.202
FACILITY GRADE TERMS	
Distance with medium upslope (1% ≤ grade ≤ 3.5%)	-9.012
Distance with high upslope (grade > 3.5%)	-18.19
LARGE ROADWAY BIKE FACILITY TERMS	
Large roadway distance with protected bike lanes	0.1748
Large roadway distance with painted bike lanes	-3.158
Large roadway distance with no bike lanes	-2.513
MEDIUM ROADWAY BIKE FACILITY TERMS	
Medium roadway distance with protected bike lanes	*-N/A-*
Medium roadway distance with painted bike lanes	-0.5464
Medium roadway distance with no bike lanes	-1.235
RESIDENTIAL ROADWAY BIKE FACILITY TERMS	
Residential roadway distance with protected bike lanes	-0.9835
Residential roadway distance with painted bike lanes	0.9288
Residential roadway distance with no bike lanes	-1.901
NON-ROADWAY BIKE FACILITY TERMS	
Cycleway distance	0.4152
Shared-use path distance	-1.705
Pedestrian zone distance	-4.021

TABLE 4: IMPLEMENTED TRAVERSAL COEFFICIENTS

COEFFICIENT DESCRIPTION	VALUE
Traversal from one centroid connector to another	-999.9

Traversal through a junction where bikers must yield	-0.301
Traversal through a junction where bikers have right-of-way and facility functional classification is downgraded	-0.164

Dropped Coefficients

Several coefficients present in the Copenhagen model were dropped from the SANDAG implementation for one of two reasons. Two facility type classification link utility terms (living streets and stairs) were dropped due to statistical insignificance. Furthermore, all link utility terms relating to land use, roadway surface, and cycle superhighway (a form of bike route) classification were dropped due to lack of data on these characteristics in the SANDAG bike model. Finally, one traversal term detailing the utility of roundabouts was omitted, as roundabouts are not modeled in the SANDAG bike network.

3.0 RESULTS

In this section, we present three separate investigations into the functionality of this new bike model implementation. First, we discuss the need to find a reasonable cutoff threshold for the bike path search space and present a search mechanism to do so, with results for both the TAZ and MGRA zone system searches. Performing searches at two different resolutions allows the model to provide detailed granularity at shorter distances and computational efficiency for larger distances. We then present a comparison of the found paths' logsums across the current and previous bike model implementation. Finally, we provide examples of origin-destination pairs and comment on the differences in the current and prior implementations' found path set.

3.1 THRESHOLDING

In the shortest path search process, an early search termination can be provided by Dijkstra's algorithm, ceasing further exploration if the encountered path cost exceeds a provided cutoff threshold. This allows a modeler to intuit a limit where no reasonable person would select bicycle as their travel mode, thus preventing the search from exploring paths which are unlikely to ever be used. The underlying SciPy implementation of Dijkstra's algorithm requires this cutoff value to be expressed in the same cost units as would be used in the path search, namely path utility. However, as the concept of utility is quite abstract, it is preferable to consider a cutoff in terms of distance. As the path utilities and distances are not directly correlated, we instead select the 99^{th} percentile of distances of paths found using a given utility threshold as its distance equivalent.

As the set of paths found within a given utility threshold is unknown a priori, finding a distance equivalent approximating some desired value requires a trial-and-error search. A script was developed that, given an initial estimate of the utility threshold equivalent to the target distance, performs an exponential (or doubling) search until a bounding interval on the utility has been found. Once this interval is established, a binary (or bisection) search is performed to find a utility threshold which has a distance equivalent within some specified margin of the target distance. By default, the script allows for a 10% margin on the input target distance.

At the conclusion of the search process, the utility threshold-distance equivalent and utility threshold-runtime relationships are charted for review. Figure 1 illustrates this graph for the TAZ network using a starting value of 20 utiles, a target distance of 20 miles, and a 10% margin. The exponential search progressively doubled the input threshold up to a value of 160 utiles before the script switched to the binary search which then led to a threshold value of 120 utiles. The graph shows a near-linear relationship between the threshold value and both the runtime and 99^{th} percentile distance, indicating the exponential-binary search strategy can be improved upon in future work using linear extrapolation as a heuristic. This is further evidenced by the MGRA search result, as shown in Figure 2, where a starting value of 25 utiles was used for a target value of 3 miles.

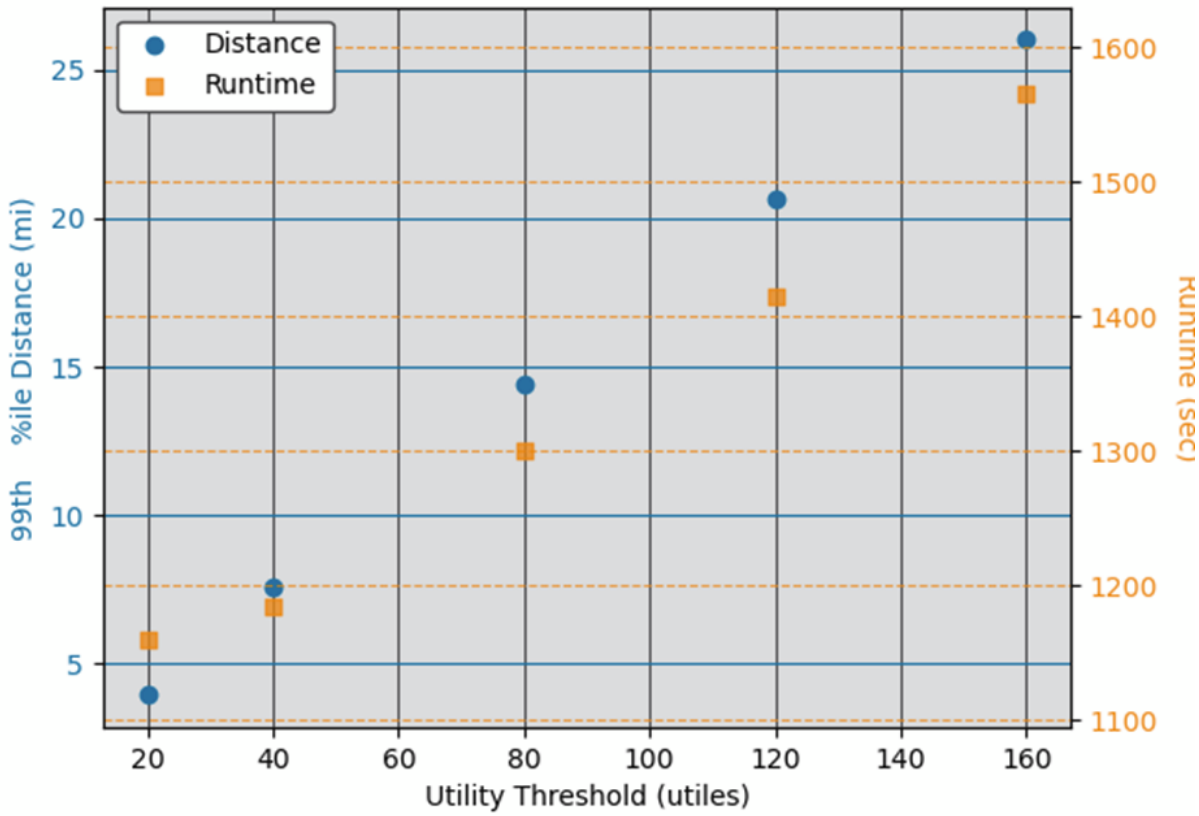

Figure 1: TAZ Threshold Search statistics graph

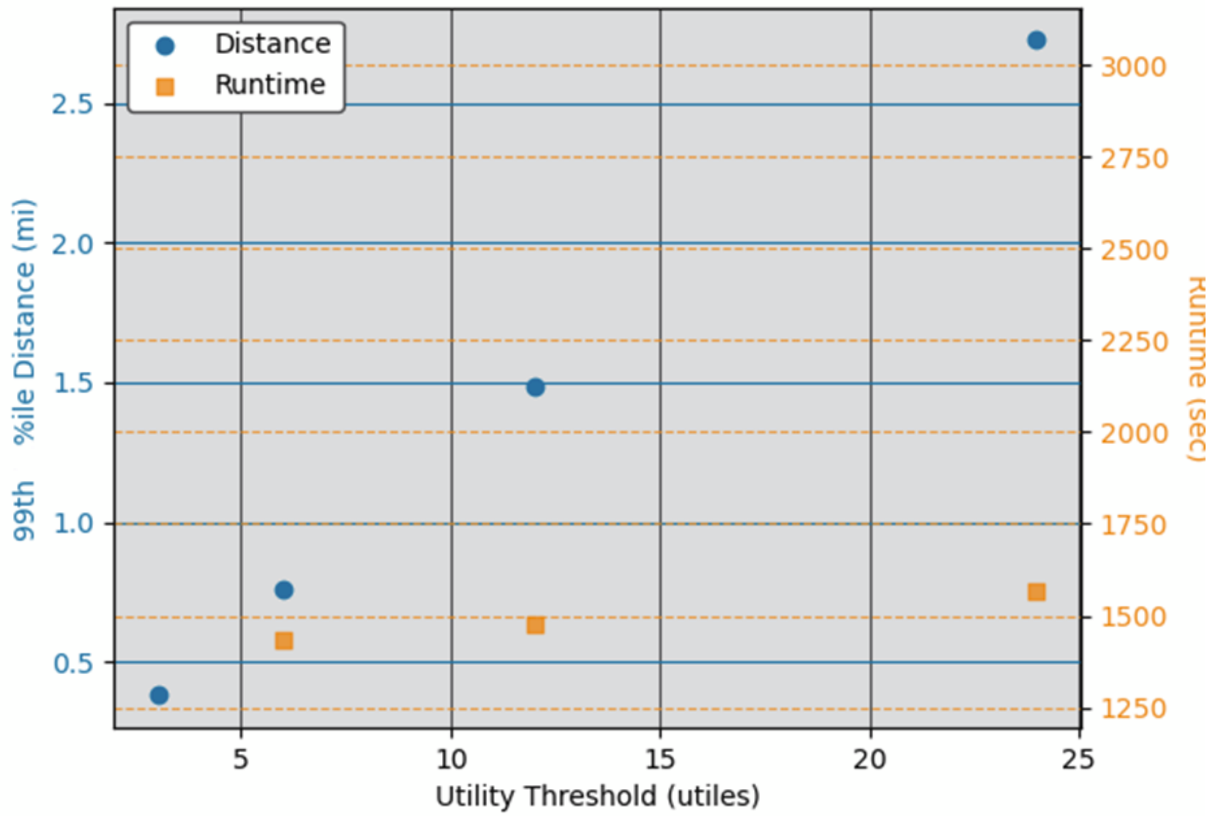

Figure 2: MGRA THRESHOLD SEARCH STATISTICS GRAPH

3.2 LOGSUM COMPARISON

To investigate the changes in origin-destination pair coverage between the old and new implementations, a comparison of the zone pairs for which the shortest path algorithm was able to find a set of paths was undertaken. The results for TAZ pairs are shown in Table 5 and the results for MGRA pairs are shown in Table 6. These tables illustrate that, for every path between TAZs found by the new implementation, a path set was found by the previous implementation as well. However, we see that there are several instances in which the previous implementation was able to find paths where the new implementation did not.

This is an expected consequence of the implemented Dijkstra's algorithm with a set utility threshold as described in the previous section. Thresholding criteria could be changed to increase the utility limit to capture more paths, but this was not done because we did not think the additional computational time was worth it to generate more long-distance paths that were very unlikely to be used.

TABLE 5: TAZ PATH INCLUSION SUMMARY

OLD \ NEW	PATH SET FOUND	NO PATH SET FOUND	*TOTAL*
PATH SET FOUND	2,819,795 (11.518%)	7,583,905 (30.977%)	*10,403,700 (42.494%)*
NO PATH SET FOUND	0 (0.000%)	14,079,004 (57.506%)	*14,079,004 (57.506%)*
TOTAL	*2,819,795 (11.518%)*	*21,662,909 (88.483%)*	*24,482,704*

TABLE 6: MGRA PATH INCLUSION SUMMARY

OLD \ NEW	PATH SET FOUND	NO PATH SET FOUND	*TOTAL*
PATH SET FOUND	3,226,528 (0.545%)	1,158,050 (0.196%)	*4,384,578 (0.741%)*
NO PATH SET FOUND	56,644 (0.010%)	587,118,462 (99.250%)	*587,175,106 (99.259%)*
TOTAL	*3,283,172 (0.555%)*	*588,276,512 (99.445%)*	*591,559,684*

Furthermore, Figure 3 and Figure 4 illustrate the correlation relationship between the old and new logsums for those zone pairs found by both algorithms. As can be seen, while there is a general positive correlation, the new logsums tend to be around an order of magnitude larger than their previous counterparts due to the much higher coefficients in the Copenhagen donor model. What's more, as is especially visible in Figure 4, this relationship appears to vary across a set of three or so groupings – a small cluster of pairs with new TAZ logsum between 0 and -2 and old logsums ranging from -1 to 10 (note: this may represent intrazonal logsums, with the positive values reflecting a known bug in the ABM2 model implementation) and two overlapping sections where the new TAZ logsums range from roughly -2 to -32 and old logsums ranging from -1 to -5.5 for one grouping and from 3.5 to -0.5 for the other. Further investigation is needed to understand this behavior.

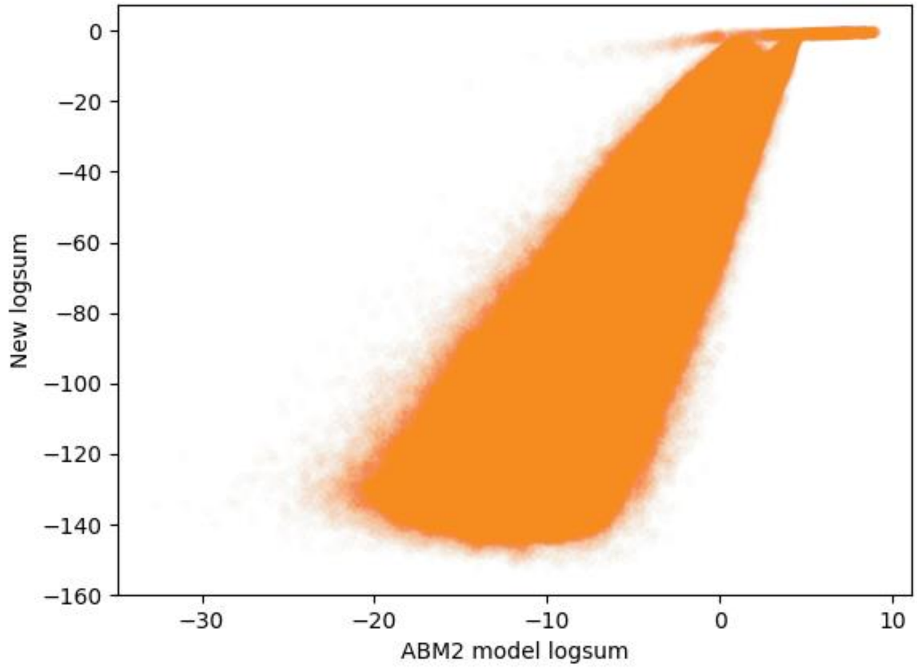

FIGURE 3: SCATTERPLOT OF TAZ LOGSUMS FROM PREVIOUS AND NEW BIKE MODELS

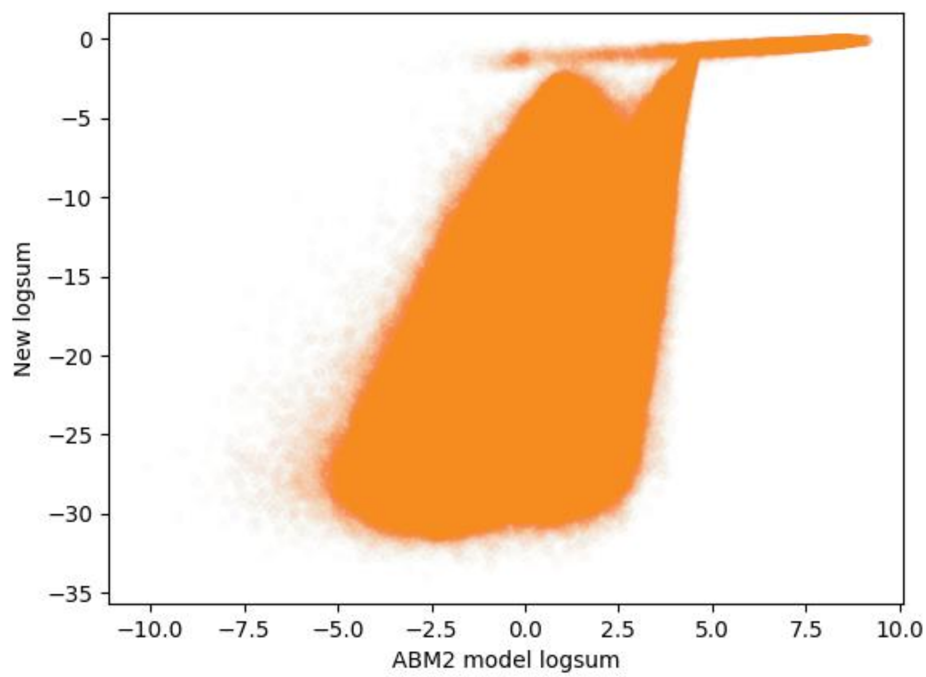

FIGURE 4: SCATTERPLOT OF MGRA LOGSUMS FROM PREVIOUS AND NEW BIKE MODELS

3.3 ORIGIN-DESTINATION CASE STUDIES

To illustrate the quality of path selection by the new implementation, a pair of case studies are mapped showing the paths selected, with the color of the path indicating its path cost relative to the other paths found between the same origin-destination pair. Figure 5 presents a TAZ pair in the Barrio Logan and shows that the path search tends to prefer a straight-line routing that also minimizes the amount of upslope along the path. Furthermore, the routing alternatives show that the randomization of costs is sufficient to generate the required multi-path set without creating chaotic or senseless routings. However, as can also be seen in Figure 6, when only a single path is realistic, the randomization is not so great as to create large diversions from this main path.

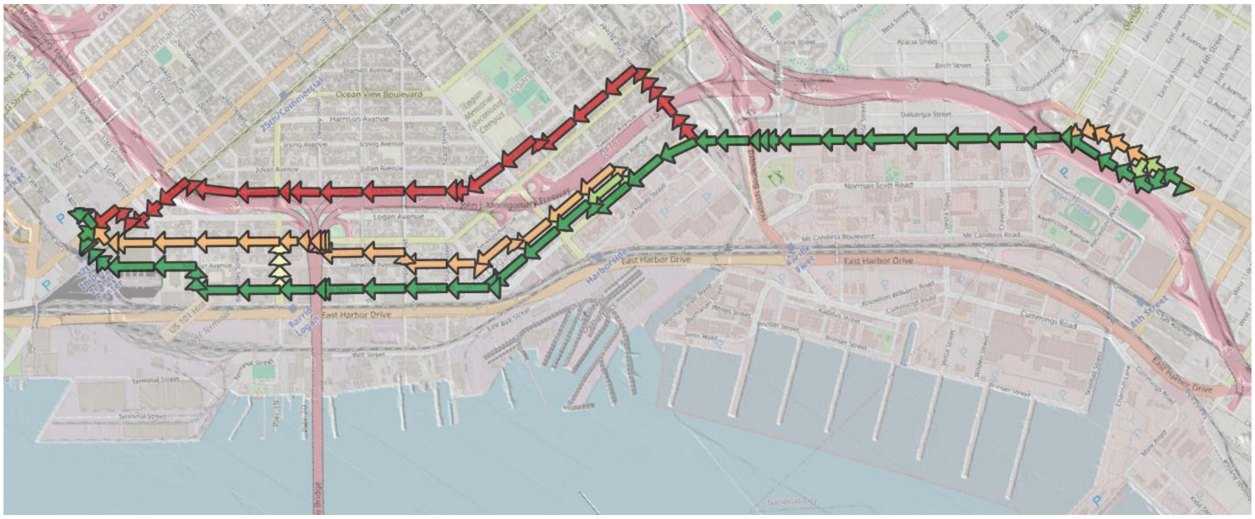

FIGURE 5: EXAMPLE PATH SET FROM TAZ 2589 TO TAZ 2030

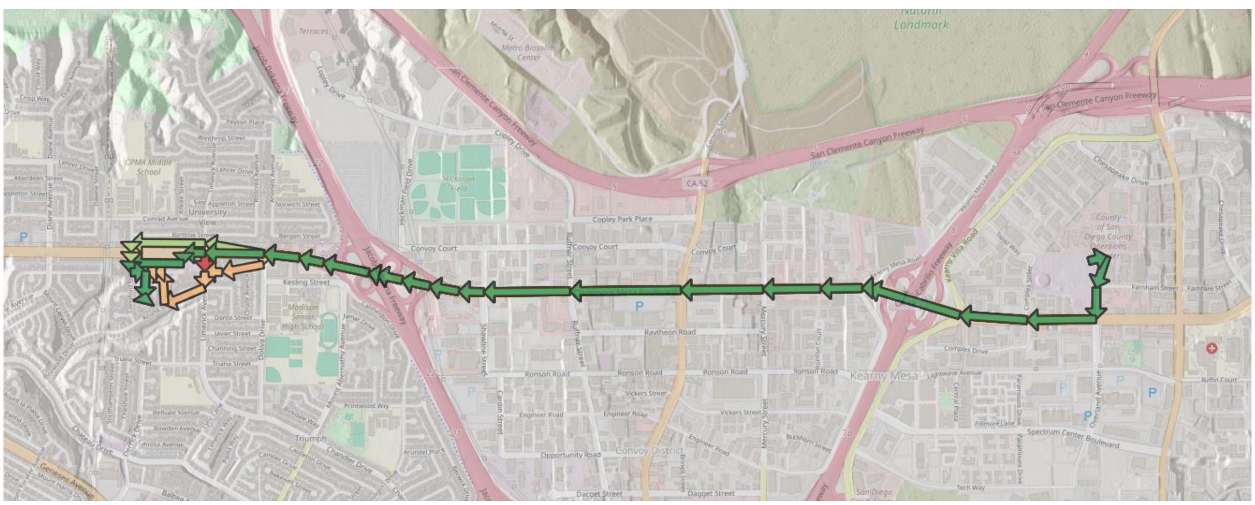

FIGURE 6: EXAMPLE PATH SET FROM TAZ 2315 TO TAZ 152

APPENDIX A. USER GUIDE

This appendix provides a practical guide for setting up, configuring, and running the Python-based active transport bike route choice model. It is intended for analysts, model integrators, and developers who need to operate or modify the model as part of the ABM3 environment.

3.4 MODEL SETUP AND RUNNING

The model requires a Python environment that includes GeoPandas and SciPy as special packages generally not included in python 3. For development of this project, we used the python environment of ActivitySim version 1.5 to ensure consistency with SANDAG's ABM3.

The model is executed through the primary driver script *bike_route_choice.py.* By default, the script reads model settings from *bike_route_choice_settings.yaml.* You can also specify the yaml file explicitly. To run the model, run the command

python bike_route_choice.py <settings_file.yaml>

This allows for parallel configurations – e.g. separate MGRA and TAZ settings, threshold experiments, or scenario testing. Model settings and their default values are listed below in Table 7.

TABLE 7: BIKE ROUTE CHOICE MODEL DEFAULT SETTINGS

SETTING	DESCRIPTION	DEFAULT VALUE
node_file	Shapefile containing bike network nodes	SANDAG_Bike_Node.shp
link_file	Shapefile containing bike network links	SANDAG_Bike_Net.shp
data_dir	Directory containing model input data	${path}\input
output_path	Directory where model outputs will be written	${path}\output\bike\taz
output_file_path	Location of final TAZ logsum output file	${path}\output\bikeTazLogsum.csv
edge_util_file	CSV defining link-level utility expressions	bike_edge_utils.csv
traversal_util_file	CSV defining traversal-level utility expressions	bike_traversal_utils.csv
bike_speed	Bicycle speed used for travel time calculations	{mode-nonmotorized-bike-speed:} taken from ABM3 run files
zone_subset	Optional list of zones for subset processing	None
zone_level	Zone type to treat as centroids	taz

max_dijkstra_utility	Utility cutoff for early path search termination	120
number_of_iterations	Maximum number of path sampling iterations	10
min_iterations	Minimum number of paths required per OD pair	0
number_of_batches	Number of batches for processing origins	16
number_of_processors	Number of processors used per batch	47
generate_shapefile	Whether to output traced paths as shapefiles	True
crs	Coordinate reference system for shapefile output	epsg:2230
random_scale_coef	Random coefficient scaling factor	0.1
random_scale_link	Random link cost multiplier	0.2
trace_bike_utilities	Whether to output traced utilities	False
trace_origins	List of origins to trace	None
trace_destinations	List of destinations to trace	None
read_cached_bike_net	Reuse cached derived network if available	False
save_bike_net	Save derived network to CSV	False

The model generates:

- **OD-level logsums** (for both MGRA and TAZ runs).
- **Sampled path sets** (optionally traced).
- **Derived network tables** (if caching enabled).
- **Diagnostic logs** supporting debugging and threshold calibration.

Paths and network tables are written to the directory specified in output_path.

Recommended workflow:

1. Create the Python environment.
2. Confirm network files and utility specs in the settings YAML.
3. Run the threshold calculator (optional) to tune early-termination settings.
4. Run: bash python bike_route_choice.py <settings.yaml>

5. Inspect outputs, including logsums and (if enabled) traced path shapefiles.
6. Adjust hyperparameters (iterations, processors, tracing) as needed.

APPENDIX B. LITERATURE REVIEW

1.0 Introduction

The purpose of this literature review is to provide an update in the current status of bicycle modeling practices found in literature. This review is focused primarily on the body of work that has developed following the last iteration of the SANDAG Activity Based Modeling enhancements that included bicycle modeling updates. The last round of updates to the bicycle modeling component of the SANDAG model occurred in 2015. Additionally, this review responds to several questions that have been posed regarding situating the current SANDAG modeling efforts within the current literature, to address desired improvements.

This review of literature took place in July 2025 and examined publications and studies that have occurred after the year 2015.

2.0 Review of Literature

Several excellent reviews currently exist discussing the wealth of studies of bicycle behavior and modeling (see for example Łukawska, 2024 and Tarkala et. al, 2024). For the sake of this literature review, the 33 papers presented in Łukawska (2024) were reviewed for potential value to the SANDAG model. These 33 studies were conducted using GPS data through data loggers, smartphone applications, fitness applications or bikeshare systems. Of these 33 studies, studies that were based on fitness application data, bicycle sharing system data due to relevance to the application to a general population model. Additionally, studies that were conducted before the last iteration of the SANDAG model were omitted based on the assumption that this body of work was known prior to the last development. Two studies based on university samples were also omitted from review for similar reasoning. Lastly, several studies were omitted from the in depth review due to the content of the analysis or the methods used, or the sample composition. The remaining studies were reviewed. In total, 10 studies were detailed and are presented for comparison to the SANDAG model. It should be noted that this review of existing literature was conducted through the lens of applicability to the current SANDAG modeling effort, and special attention was given to the similarity in methodology to the current model in terms of shortest path and estimation procedures for potential enhancements. Studies were broken into US or Canada based work and those outside of the US or Canada (which were all European studies).

Of the remaining ten studies, four were located in cities within the United States or Canada. The models were evaluated on the applicability of use for model improvements within the SANDAG framework.

Zimmerman et al.- US based study of smartphone application data collection. The analysis consisted of 648 trips from 103 users. Data consisted of GPS observations of cyclists trajectories in the city of Eugene, Oregon using the CycleLane smartphone application (which was built on code from San Franciscos CycleTracks application). The authors implemented four models, two being recursive logit models and the other two being nested recursive logit models. One of each of these types of models included Link Size attributes. Link attributes are included and used in the model formulation rather than percentages or proportions of the whole route a key difference in methodology between the SANDAG model and the models presented in this research. For this reason, this model was not considered the best comparison or resource for SANDAG improvements.

Chen et al. 2018- US smartphone application based data collection between 2009 and 2014. GPS data were collected using GPS devices in Seattle for a period of 3.5 years resulting in a sample of 538 trips. The authors used principal components analysis to describe route attributes. They then developed a mixed logit model of route choice and included distance. Included the variable of cycletrack, but this fell out of the logit model as not significant. Included distance in the model, which allows for ratio comparisons. Small sample size may lead to some weakness in the model and significant variables when comparing and looking for enhancements.

Misra and Watkins, 2018- US based study of smartphone app and survey data collected from 2012 to 2014. This analysis is based on 437 trips. Trip data from a GPS enabled smartphone app (Cycle Atlanta), linked to socio-demographic data. Data for models was narrowed down from an original sample of 15,463 trips across 1,495 users of the app. The authors developed route choice models for females only and males only to analyze differences in significant coefficients. A pooled model of all data was then developed with interaction terms for female or male specific differences in coefficients. The pooled model also included an age segmentation. This model included distance as a coefficient in the models, but this was the only coefficient overlap to the current SANDAG formulation and might not give insight to comparisons for coefficients.

Li et al. 2017- Canada based smartphone app and survey based data collection between 2014 and 2015. The analysis is conducted on 8,829 and 17,049 trips depending on the analysis. Data used in the study were a result of a data collection effort by the City of Toronto using a smartphone application. Participants recorded trips and optionally responded to a survey to collect socio-demographic information. Survey was not representative of the population. 72.8% of respondents did not respond to question of gender, but for most other questions in the survey, response rate was around 80%. Approximately half of respondents reported that they have cycled since childhood, and approximately 50% reported that they did not cycle in the winter. The authors used a combination of deterministic methods for choice set generation (path labeling and link penalties). Alternatives were derived on the basis of shortest path routes and energy saving routes. Energy saving routes were based on the ability of cyclists to overcome rolling resistance, air resistance and uphill grade. Several models were specified to compare treatment of variables and estimation techniques. Specifically, path-size and multinomial models were estimated. Key findings indicate that the path-size logit model performed better than the multinomial logit model. Additionally, a model that considered 20 alternative routes performed better than the model with 10 alternatives. Lastly, the transformation of distance into a logarithmic form improved the model fit.

Six additional studies from European cities have also been evaluated for applicability of model improvements.

Ghanayim and Bekhor, 2018- Israel based study using GPS device and survey data collected between 2013 and 2014, resulting in a sample of 545 trips. GPS data from a data logger was used from a large household survey effort in the Tel Aviv metropolitan area between Dec 2013 and June 2014. 545 trips were extracted from the survey, performed by 221 respondents. The authors compared Multinomial logit, c-logit and path-size logit models. Results indicate a superior performance of the Path-Size Logit model. Additionally, the authors compared choice set size in a second analysis focused on choice set generation methods. These methods include deterministic link elimination and link penalty generation and stochastic simulation (calculating the shortest path for each draw of link impedance from a log normal distribution). Results showed that the coefficients did not significantly differ between the models. Lastly, the authors compared 15 models with reduced choice set sizes (35%, 50%, 65% and 80%) of the original three models to examine the model robustness over varying choice set sizes. Results indicate that if the choice set is reduced by half (10 routes per observation), the parameter estimates are like the original model. The choice set size has a larger influence on the standard deviation of the coefficient estimated in comparison to the average coefficients.

Łukawska et al. 2023- Denmark based study using bicycle airbag device data collected between 2019 and 2021. Data used in this study were collected via GPS enabled and smartphone connected bicycle airbag devices. The data are recognizably skewed as the study includes only those cyclists who have purchased the 300 euro airbag device. Only those trips within the metropolitan area of Copenhagen are considered, resulting in 134,169 trips from 6,523 participants. The authors in this study employed a choice set generation approach based on the repeated shortest path searches that use different preference categories for each search. Each search was constructed based on a behavioral interpretation (negative impact on steep uphill, strongly prefers nature). A penalty coefficient was applied to specific terms in the cost function, which penalized specific links in the network. The authors used 31 categories to generate choice sets for each trip. These categories were created using attributes of bicycle infrastructure type, road type, elevation, land use, surface type, existence of a cycle superhighway and riding against traffic in one-way streets. Route choice was modeled using a joint path-size logit model. The model distinguished between parameters in value-of-distance space and preference space. All attributes that included measures of distance were expressed in the VoD space, whereas all other coefficients in other units (for instance count) were included in the preference space. Two models are jointly estimated one for long/short distance trips and one for infrequent/frequent cyclists. Although this study has limits in the representative nature of the data, this study was retained in the review due to the extensive set of attributes included as coefficients in the model and possible interest for model improvement.

Prato et al. 2018- Denmark based study using GPS device and survey data collected between 2012 and 2013. GPS device data collection was based off of the data collection effort and surveys of the Danish National Travel Survey, which collects travel diaries from Danish people between the age of 10 and 84. Those recruited for the bicycle data collection effort had participated in the National Travel Survey within 6 to 12 months, reported bicycle activity in their travel diary, lived in the Copenhagen Region and were at least 16 years old. Survey recruitment occurred in three rounds, October-December 2012, June to July 2013 and August to October 2013 and resulted in 291 participants with 3,384 trips. The authors used a doubly stochastic generation function to generate a choice set, and estimated a path-size logit choice model to model the choice process. The authors used value of distance space for parameter interpretation. This model is an extensive model with a large number of attributes included in the specification, which may be of interest to SANDAG model improvements efforts.

Huber et al. 2021- German based data collection in 2018. This study was carried out by the University of Dresden. Sample included about 200 study participants (distribution of age and gender corresponding to the population of the city), from which 3,200 bicycle trips were used for model estimation. The authors use BRouter to generate three alternative routes to the observed route in the dataset using the routing profiles of "trekking, shortest and fastbike-verylowtraffic." The authors then used a Multinomial Logit model to estimate route choice. The treatment of distance specifically varies from the SANDAG model as coefficients are included as percentages of the route for each variable (for instance percentage of route with bicycle lane, or with max speed less than 31 km/hr), not as a straight distance.

Meister et al. 2023- Swiss based study using smartphone app and survey data collected in 2021. Data used were from the greater Zurich area. A total of 4,432 trips from 100 individuals were used for the study. Data contained both regular and e-bicycles. Socio-demographic information was also collected from participants. Descriptive statistics and model results are segmented by regular and e-bikes. The authors combined BFSLE algorithm with a link penalty method for choice set generation. The authors estimated a path-size logit model in value of distance space (VoD). VoD is directly estimated in the model with the level of service attributes being represented by a value of distance parameter. Of specific interest to possible comparisons and model improvements, the authors included length, bike path, bike lane, speed limit, traffic signals, slope (2%-6%, 6%-10% and over 10%, and max traffic (k cars/day).

Dane et al. 2017- Dutch based study using data collected via smartphone app in 2014. GPS and basic user information collected from 742 respondents resulting in 17,626 bicycle/e-bike trips. The authors used K-shortest path method to develop choice sets of five alternatives. The authors then used a path-size logit to estimate regular and e-bike trips separately. Socio-demographic characteristics and trip related factors are found to influence route choice, and differences exist between e-bike and regular bicycle route choice. This study focused more on the role of sociodemographic attributes and bicycle type on route choice. This study does not include physical/built environment route attributes and includes all coefficients as interactions with distance, which may not offer any direct paths for model improvements to SANDAG.

After further investigation and discussion, a subset of the aforementioned studies was selected to directly compare coefficients between the studies presented and the current SANDAG bicycle route choice model. The Path-Size Logit model of Misra and Watkins (2018) is provided in Table B1, the Path-Size Logit model of Li et. al (2017) is presented in Table B2, the model of Ghanayim and Behkor (2018) is presented in Table B3, and the Path-Size Logit model presented in Value of Distance space found in Łukawska et. al (2023) is provided in Table B4. Additionally, the coefficients that overlap with the current iteration of the SANDAG model are compared as a ratio to the coefficient of distance in each of the respective models.

Table B1: Misra and Watkins (2018) Path-Size Logit Model with SANDAG Model Coefficients

Variables	Coefficient	ratio to dist	SANDAG Variables that overlap	coefficient	ratio to dist
ASC_alternative 2	-1.78*** (-3.4)				
ASC_alternative 3	-1.38**(-2.8)				
ASC_ alternative 4	-2.12* (-1.95)				
ASC_alternative 5	-3.13*** (-2.89)				
Distance	-2.76*** (-3.22)		Distance (tot) mi.	–0.858	
AADT	-0.62*** (-3.18)				
Speed	-0.187* (2.12)				
Percent cycle facilities	1.262*** (5.72)				
Average number of lanes	-0.400* (-2.47)				
Slope	-0.32** (-2.85)	0.1159	Cumulative gain in elevation, ignoring declines (ft.)	–0.010	0.0117
AADT x female	-0.357* (-2.52)				

Speed x female	-0.280* (-2.12)				
Log PS	1.990**(2.8)	-0.7210	Log Path Size	1	-1.1655

***p=0.001 **p=0.01 *p=0.05
Model statistics: LL(0)= -1328.34, LL(MS)= -1181.37, LL(Model)= -654.341, McFadden's p2 (base MS)= 0.45

Table B2: Li et al (2017) Model with SANDAG model coefficients

Variable	Coeff.		ratio to dist	SANDAG Variable	coefficient	ratio to dist
Distance (km)	-0.159	−20.26 (***)		Distance, total (mi.)	-0.858	
Energy consumption (1,000s kJ)	-0.196	−10.05 (***)				
Number of transit stops	0.00	7.61 (***)				
Annual average daily traffic volumes (1,000s vpd)f	-0.223	−13.09 (***)				
TP with major arterial	19.0	119.53 (***)	-119.4969	Distance on “boulevard” class III bike routes	0.43	-0.5012
TP with minor arterial	17.9	109.83 (***)				
TP with collector	17.3	94.90 (***)				
TP with local street	12.0	56.90 (***)				
TP with shared lane	0.341	2.95 (***)	-2.1447	Distance on class III bike routes	0.085	-0.0991
TP with bike lane	0.174	1.72 (***)	-1.0943	Distance on class II bike lanes	0.314	-0.3660
TP with multiuse trail	18.3	70.12 (***)	-115.0943	Distance on class I bike paths	0.61	-0.7110
TP with cycle track	2.540	11.72 (***)	-15.9748	Distance on “cycle track” class II bike lanes	0.12	-0.1399
TP without bicycle facilities	-3.040	−36.71 (***)	19.1195	Distance on arterials without bike lanes	-1.05	1.2238
Path size factor	5.190	106.9(***)	-32.6415	Log Path Size	1	-1.1655

PSL (n=20) TP= Trip proportion *p < .05; **p < .01; ***p < .001. Final log likelihood = −38,857.988; ρ2 = .616

Table B3: Ghanayim and Behkor (2018) Model with SANDAG Model Coefficients

Variable	Coefficient (Mixed PSL)	t-stat	Ratio to dist	SANDAG Variable	coefficient	Ratio to dist.
Total route length (km)	-8.04	-14.2		Distance, total (mi.)	-0.858	
Route length in category A (bike paths) (km)	2.23	7.3				
Route length in category C (urban arterials and highways) (km)	-4.59	-9.3	0.5709	Distance on arterials without bike lanes	-1.05	1.2238
Average street length (m)	0.0543	9				
Dwelling units / m	2.12	10.7				
Route length “near sea” (km)	1.28	7.5				
Route length “near park” (km)	1.72	5.1				
PS factor	1.77	9.3	-0.2201	Log Path Size	1	-1.1655
Sigma (σ)	1.14	5.6				
Hit-ratio	69%					

Initial Log-likelihood= -1420.5, Final Log-likelihood=-970.5, Likelihood Ratio= 900, Rho-bar squared= 0.316, Number of observations= 545, Number of individuals= 221

Table B4: Łukawska et al (2023) Path-Size Logit in Value of Distance Space Model With SANDAG Model Coefficients

Variable	Coefficient (p < .01)	ratio to dist	SANDAG Variable	coefficient	ratio to dist
Parameters in preference space					
Length	-6.790		Distance, total (mi.)	-0.858	
Measure of overlap across alternatives					
ln(Path-Size)	0.614				
Elevation gradient					
Flat or downhill	--				
Steep uphill (10 – 35 m/km)	-0.056	0.0082	Cumulative gain in elevation, ignoring declines (ft.)	-0.01	0.0117
Very steep uphill (> 35 m/km)	-0.113	0.0166			

Intersection type					
Road hierarchy downgraded	-0.164				
Road hierarchy upgraded	-0.301				
Roundabouts	-0.121				
Traffic lights	-0.029				
Infrastructure					
No. of stair segments	-1.010				
Parameters in VoD space					
Length	1.000				
Infrastructure					
Medium roads w/ protected bicycle tracks	--				
Medium roads w/ painted bicycle lanes	0.050	-0.0074	Distance on class II bike lanes	0.314	
Medium roads w/o bicycle infrastructure	0.113	-0.0166	Distance on class III bike routes	0.085	
Large roads w/ protected bicycle tracks	-0.016				
Large roads w/ painted bicycle lanes	0.289				
Large roads w/o bicycle infrastructure	0.230				
Residential roads w/ protected bicycle tracks	0.090				
Residential roads w/ painted bicycle lanes	-0.085				
Residential roads w/o bicycle infrastructure	0.174	-0.0256	Distance on arterials without bike lanes	-1.05	
Cycleways	-0.038	0.0056	Distance on “cycle track” class II bike lanes	0.12	
Footways	0.506				
Living streets	(insig)				
Shared paths	0.156				
Pedestrian zones	0.368				
Stairs	(insig)				
Land use (right-hand side)					
High-rise urban areas	--				
Green areas	-0.066				

Areas near water	-0.177				
Industrial areas	-0.022				
Low-rise urban areas	-0.040				
Open landscape	-0.049				
Wrong way	0.293	-0.0432			
Surface type					
Asphalt	--				
Cobblestones	0.271				
Gravel	0.130				
Cycle superhighways					
No classification	--				
Existing	-0.021				
Proposed	-0.047				

Number of observations= 134,169, Number of parameters=33, Final log likelihood=-220,808.50, Rho-square-bar for the init. Model= 0.404, Akaike Information Criterion=441,683

3.0 Specific Topics of Interest Related to SANDAG's Model Improvement

The following presents several topics of particular interest to the SANDAG bicycle route choice effort, with the goal of highlighting any alternatives approaches and situating the current SANDAG model within the existing literature.

Topic: handling of traversal coefficients in the model

The SANDAG model enumerates and adds utilities to each possible traversal through each intersection. There are a variety of manners in which different research efforts have handled turns and traversals in the route, however, the current SANDAG is more detailed in the way turns are broken into detailed movements through intersections and through different classifications of streets. Zimmerman et al. (2017) includes coefficients for no turns, no turns with crossroads, left through medium traffic at an unsignalized intersection, and left turn through heavy traffic at an unsignalized intersection. Prato et al (2018) include the turn, but no details about the intersection other than directionality in their estimation. Łukawska et al. (2023) include the intersection type in their model with a road hierarchy downgraded or upgraded, as well as roundabout and traffic lights. In this model, each of these intersection types have a negative effect. Meister et al. (2023) includes traffic signals, but no further specifics on whether there was a turn. This model also does not include unsignalized turns. Huber et al. (2021) included the number of intersections with traffic lights into the original specification of the model, but this coefficient was not significant.

Topic: Additional possible coefficients estimated for other bike route choice models that are missing from the SANDAG model

Slope is a commonly tested and significant coefficient in the literature that is not in the 2015 version of the SANDAG model. The treatment of elevation is primarily represented by slope or percent gain, or even maximum slope. Zimmerman et al. 2017 state:

> *As a result of the link-additivity assumption, link characteristics such as slope need to be carefully incorporated in the utility function. For example, it is not possible to include slope as a continuous variable, since the average slope of a path consisting of two links is not equal to the added average slopes of each link. In our case, these inherently non-link-additive attributes are slope, traffic volume and presence of bike facilities. The solution we adopt is to specify a dummy variable δ_a for each of these attributes and let the dummy variables interact with the link length attribute. On each link a, the variable δ_a takes the value 1 if the attribute is present or greater than a chosen threshold in case of continuous attributes, and 0 else. The interaction term is simply the product of the two attribute values. Not only does this specification allow us to include important characteristics in a way that respects link additivity, but the interpretation is also simple and intuitive.*

Average annual daily traffic volumes or maximum traffic volumes have also been considered in several studies (Misra and Watkins 2028, Li et al. 2017, Meister et al. 2023) as well as speed and speed limit (Chen et al. 2018, Misra and Watkins 2018, Huber et al. 2021), which are not present from the last SANDAG model update. Li et al. (2017) included a metric of energy consumption, which was a composite of rolling resistance, air resistance and uphill grade, and a cyclist’s ability to overcome these factors.

Additionally, several authors incorporate a variety of land uses such as residential areas (Prato, et al. 2018), number of dwelling units (Ghanayim and Bekhor 2018) green space such as parks and water proximity (Prato et al. 2018 and Ghanayim and Bekhor 2018).

Łukawska (2024) provides a more comprehensive review of literature on bicycle route choice between 2010 and 2023 including a table comparing studies and factors that influence route choice and is provided in Table B5.

Table B5: Coefficients found in literature (Łukawska 2024)

Table 3. Network factors associated with cyclists' route choice. "+" means a positive influence of a factor on the route choice probability, "–" means a negative influence, and "x" means that this influence was diverse, dependent on other variables, or unclear. For studies employing comparison to the shortest path as a modelling method, all attributes are marked with "x".

	Network attributes									
Reference	Length	Bicycle infrastructure	Gradient	Traffic lights	Turns	Land-use	Road size (no. of motorised traffic lanes or road type)	Traffic volume	Paths dedicated for or shared with pedestrians	Riding the wrong way
Menghini et al. (2010)	–	+	–	+						
Hood et al. (2011)	–	+	–		–				+/–	–
Broach et al. (2012)	–	+	–	–	–			–		
Casello and Usyukov (2014)	–	+	–					+/–		
Yeboah and Alvanides (2015)*	x	x					x		x	
Khatri et al. (2016)	–	+		+	–			–		
Li et al. (2017)	–	+					+/–	–		
Ton et al. (2017)	–									
Zimmermann et al. (2017)	–	+	–		–			–	+/–	
Bernardi et al. (2018)	+/–	–		+		x				
Chen et al. (2018)	–	+	–			x				
Ghanayim and Bekhor (2018)	–	+				x	–			
Lu et al. (2018)*	x	x	x		x		x	x		
Misra and Watkins (2018)	–	+	–				–	–		
Prato et al. (2018)	–	+/–	–	–	–	x	–		–	–
Skov-Petersen et al. (2018)	–	+		+	–	x	+/–		+/–	
Park and Akar (2019)*	x	x	x	x	x	x	x	x		
Sobhani et al. (2019)	–	+	–		+/–	x				–
Dane et al. (2020)	+									
Fitch and Handy (2020)	–	+	–	+/–			+/–			
Huber et al. (2021)	–	+	–							
Koch and Dugundji (2021a)	+	+		–						
Koch and Dugundji (2021b)	+/–	+		–		x				
Scott et al. (2021)	+	+/–	+/–		–		+/–		–	
Shah and Cherry (2021)	–	+		+	–			+/–		+/–
Cho and Shin (2022)	–	+	–	+						
de Jong et al. (2022)*	x	x	x			x			x	
Magnana et al. (2022)										
Fosgerau et al. (2023)	–	+	–			x	–		–	–
Łukawska et al. (2023)	–	+	–	–		x	–		–	–
Meister et al. (2023)	–	+	–	–				+/–		
Rupi et al. (2023)*	x	x		x	x	x	x			
Schirck-Matthews et al. (2023)	+/–	+		–	–	x	–	+/–	+/–	

Table 3. Continued

	Network attributes										
Reference	Car speed limit	Intersections	Unpaved surface	Bicycle infrastructure on bridges	Stairs	Cycle superhighway	Roundabouts	Motorised traffic bridge	Public transit stops	Stop signs	Light
Menghini et al. (2010)											
Hood et al. (2011)											
Broach et al. (2012)				+						–	
Casello and Usyukov (2014)	–										
Yeboah and Alvanides (2015)*											
Khatri et al. (2016)											
Li et al. (2017)									+		
Ton et al. (2017)		–									
Zimmermann et al. (2017)				+/–				–			
Bernardi et al. (2018)											
Chen et al. (2018)	–										+
Ghanayim and Bekhor (2018)		–									
Lu et al. (2018)*		x									
Misra and Watkins (2018)											
Prato et al. (2018)		–	–	+	–		+	–			
Skov-Petersen et al. (2018)											
Park and Akar (2019)*	x										
Sobhani et al. (2019)	–								–		
Dane et al. (2020)											
Fitch and Handy (2020)	–										
Huber et al. (2021)	–		–								
Koch and Dugundji (2021a)									+		
Koch and Dugundji (2021b)											
Scott et al. (2021)		–									
Shah and Cherry (2021)	+										
Cho and Shin (2022)		–									
de Jong et al. (2022)*	x					x			x		
Magnana et al. (2022)											
Fosgerau et al. (2023)			–		–	+					
Łukawska et al. (2023)		–	–		–	+	–				
Meister et al. (2023)	+/–										
Rupi et al. (2023)*	x	x									
Schirck-Matthews et al. (2023)		+									

*Employed comparison to the shortest path.

Scott et al, 2021 estimated a model using bike share user data and is provided in Table B6. Of note, this study included number of turns, but also included a route directness index (ratio of route distance to straight line distance), as well as a longest leg coefficient. Both of these coefficients show that cyclists have strong preference for routes that minimize path circuity.

Table B6: Coefficients estimated in Scott et al (2021)

Table 2
Results for global, medium G_n, and high G_n PSL models.

Attributes	Global		Medium G_n		High G_n	
	coef.	t-stat	coef.	t-stat	coef.	t-stat
Route distance	0.21	19.71*	0.72	28.86*	2.48	49.64*
RDI	−0.01	−3.06*	−0.46	−27.94*	−2.25	−48.81*
Total turns	−0.01	−4.48*	0.0003	0.08	−0.07	−10.35*
Distance btw intersections	1.37	8.43*	−0.11	−0.40	1.16	5.08*
Longest leg length	0.65	39.23*	0.87	31.82*	2.12	41.29*
Slope 0–2%	−0.21	−3.37*	−0.14	−1.59	−0.53	−3.12*
Slope 2–4%	−0.02	−0.30	0.02	0.15	0.62	3.17*
Slope 4–6%	−0.48	−3.88*	0.10	0.55	−5.96	−13.84*
Major roads without BL	0.13	5.03*	0.08	2.22*	−1.44	−16.43*
Major roads with BL	0.27	5.57*	0.34	4.76*	0.25	1.81
Minor roads with BL	−0.08	−2.85*	0.03	0.76	−2.47	−28.79*
Trails	1.41	21.84*	1.30	12.49*	1.61	12.97*
Cautionary un-signed BR-HT	−0.66	−9.19*	−0.46	−3.94*	−1.69	−4.67*
Cautionary un-signed BR-MT	0.32	7.35*	−0.01	−0.21	−0.22	−2.37*
Cautionary un-signed BR-LT	1.14	16.16*	0.94	7.14*	0.94	7.29*
Signed on-street BR-MHT	−0.06	−0.28	0.04	0.09	0.18	0.23
Signed on-street BR-LT	0.19	6.21*	0.16	3.48*	0.12	1.44
Paved multi-use trail	−1.39	−4.04*	−1.30	−2.41*	−13.98	−7.08*
ln(PS)	−1.34	−140.93*	−1.50	−104.70*	−2.68	−96.37*
Number of OD hub pairs	5561		1257		250	
Number of routes	41,369		2902		2472	
Adjusted ρ^2	0.054		0.069		0.195	

* $p \leq 0.05$.

Tarkkala et al (2024) reviewed multiple sources of literature to examine types of factors common across multiple studies in route choice. Findings are provided in Table B7 and Table B8 taken from the review.

Table B7: Emerging Factors from Literature Tarkala (2024)

Table 1
Types of emerging factors in route choice studies.

Road characteristics		Trip characteristics		User characteristics		Built environment features	
A	Gradient	L	Trip length	R	Age	CC	Scenery
B	Signalized intersection	M	Average speed	S	Gender	DD	Street green
C	Bike facility	N	Travel time	T	Cycling purpose	EE	Crowdedness
D	Traffic volume	O	Wrong-way travel	U	Cycling experience	FF	Residential unit density
E	Road quality	P	Left and right turns	V	Income	GG	Street lighting coverage
F	On-street parking	Q	Road continuity	W	Cycling frequency	HH	City features
G	Speed limit			X	Cycling with children	II	Land use mixture
H	Number and width of lanes			Y	Car ownership	JJ	Public transport service level
I	Road class			Z	Level of education		
J	Bridge facility			AA	Weather influence		
K	Tram tracks			BB	Bike operation cost		

Table B8: Significant Factors by Study Reviewed in Tarkala (2024)

Table 2
Most significant identified factors in each study.

Publication	A	B	C	D	E	G	I	J	L	N	O	P	Q	S	T	CC	DD	FF	GG	JJ	Study area	Data
Stinson and Bhat (2003)		✓	✓		✓		✓	✓		✓											US	SP
Sener et al. (2009)		✓		✓		✓				✓			✓								Texas, US	SP
Menghini et al. (2010)	✓		✓						✓												Zurich, Switzerland	RP
Hood et al. (2011)	✓		✓						✓		✓			✓	✓						San Francisco, US	RP
Winters et al. (2011)	✓		✓	✓	✓											✓					Vancouver, Canada	SP
Broach et al. (2012)	✓	✓	✓	✓					✓			✓									Portland, US	RP
González et al. (2016)			✓						✓								✓			✓	Providencia, Chile	SP, RP
Grond (2016)	✓		✓	✓					✓			✓									Toronto, Canada	SP, RP
Vedel et al. (2017)		✓	✓				✓										✓				Copenhagen, Denmark	SP
Zimmermann et al. (2017)	✓		✓	✓				✓	✓												Eugene, US	RP
Chen et al. (2018)	✓		✓			✓			✓										✓		Seattle, US	RP
Majumdar and Mitra (2018)			✓	✓															✓		Kharagpur & Asansol, India	SP
Bernardi et al. (2018)									✓					✓	✓						The Netherlands	RP
Ghanayim and Bekhor (2018)			✓				✓		✓							✓		✓			Tel Aviv, Israel	RP
Hardinghaus and Papantoniou (2020)			✓		✓					✓											Munich, Germany & Athens, Greece	SP

From this review, they created a stated preference survey using six factors:

- Bike facility- Mixed traffic, Bike lane, Adjacent cycle path and Separated cycle path
- Road class- Local street, Main street, Arterial road
- Traffic volume- Light (0–500 veh/h), Moderate (500–1000 veh/h), Substantial (1000–2000 veh/h), Heavy (>2000 veh/h)
- Signalized- No traffic signal, Few traffic signals, Many traffic signals
- Gradient- No hills (0–2.5%), Moderate hills (2.5%–5%), Steep hills (>5%)
- Trip length

Using this data, the authors estimated two route choice models (M2 includes an interaction effect between trip length and gradient while M1 excludes it). The results of these models are provided in Table B9. Their resulting model was integrated into HELMET, the four-step travel demand model of Helsinki. It should be noted that this research uses stated preference data rather than revealed preference, and for this reason it might not be the best model to use for SANDAG model improvements.

Table B9: Estimated Model Coefficients in Tarkala (2024)

Table 4
Route choice model estimation results.

Factor	Parameter	M1		M2	
		Coefficient	t Stat.	Coefficient	t Stat.
Road class × length	β_{Local}	−4.96***	−33.1	−3.99***	−29.6
	β_{Main}	−4.14***	−28.7	−3.31***	−25.3
	$\beta_{Arterial}$	−4.05***	−25.4	−3.28***	−21.8
Bike facility × length	β_{Mixed}	−5.03***	−40.1	−4.47***	−37.7
	β_{Lane}	−2.98***	−26.2	−2.56***	−22.5
	$\beta_{Adjacent}$	−2.52***	−19.8	−1.85***	−15.3
	$\beta_{Separated}$	−2.61***	−16.0	−1.70***	−11.4
Traffic volume (Ref.: Light)	$\beta_{Moderate}$	−0.008***	−7.7	−0.005***	−5.9
	$\beta_{Substantial}$	−0.008***	−16.7	−0.006***	−14.2
	β_{Heavy}	−0.009***	−24.2	−0.008***	−3.0
Signalized intersections (Ref.: No signal)	$\beta_{FewSignals}$	−0.59*	−1.7	−0.99***	−3.0
	$\beta_{ManySignals}$	−2.00***	−10.8	−1.75***	−9.9
Gradient (×length in M2) (Ref.: No hills)	$\beta_{ModerateHills}$	−2.32***	−14	−17.20***	−10.3
	$\beta_{SteepHills}$	−2.50***	−18.6	−23.40***	−18.4
Model fit					
No. of estimated parameters		14		14	
Sample size		8232		8232	
Initial likelihood		−5706.0		−5706.0	
Final likelihood		−4031.4		−4076.1	
Initial likelihood ratio test		3349.1		3259.7	
Rho squared		0.293		0.286	
AIC		8090.8		8180.2	
BIC		8189.1		8278.4	

* Significant at 10% ($p \leq 0.10$).
*** Significant at 1% ($p \leq 0.01$).

Many of the studies recognize that preferences and impacts of route attributes are linked to sociodemographics such as gender and age, as well as the type of bicycle used (ebike vs regular bicycle). However, the implementation of this knowledge is varied. Dane et al. (2017) incorporated age, gender and bicycle type directly into the model, but several categories were found not to be significant. Meister et al. (2023) use a mixed Path-Size Logit to incorporate the random taste variation into their model estimates. From this, they derived the median value of distance indicators for gender, age and bicycle types for a variety of road and bicycle facility attributes, which can be found in Table B10.

Table B10: Meister et al. 2023 Mixed Path-Size Logit Model

Table 4
Median VoD indicators. Missing values are in a ±10% range of the value for the average respondent.

	avg.	Male	Female	<30y.	30–50y.	>50y.	Bike	ebike
Bike path [km]	−0.36	−0.15	−0.42	−0.15	−0.15	−0.15	−0.23	−0.12
Bike lane [km]	−0.66	−1.00	−0.27	−1.00	−1.00	−1.00	−1.00	−1.12
Speedlimit 30 [km]	−0.16	0.09	–	–	0.07	0.07	−0.13	0.37
Traffic signals [n]	0.19	0.42	0.23	0.29	0.31	0.40	0.28	0.53
Slope 2%–6% [km]	0.55	0.31	0.46	0.24	0.77	–	0.41	0.09
Slope 6%–10% [km]	3.11	1.69	2.58	3.08	2.44	1.42	2.51	1.01
Slope > 10% [km]	4.33	2.96	4.38	7.38	–	–	–	2.78
max. traffic 1–10k [0,1]	0.07	0.16	–	–	–	–	0.08	0.27
max. traffic > 10k [0,1]	0.11	0.07	0.25	0.09	0.09	0.09	0.01	0.15

Topic: SANDAG currently uses double stochasticity in the model, how does this compare to other bicycle route choice models.

Prato et al (2018) conducted a study in Copenhagen using the Danish National Travel Survey GPS data to investigate the influence of land use and transport network effects on cyclist route choice. In this study, they evaluate methods for the generation of choice sets, including breadth first search on link elimination (BSF-LE), doubly stochastic generation function (DSGF), and branch and bound algorithm. Their review and tests show a superior performance of the DSGF method of choice set generation, which accounts for differences in travelers link cost and differences in travelers attribute preferences by drawing random costs and random parameters from probability distributions.

The doubly stochastic choice set generation approach followed by Prato et al (2018) is fully described in Bovy and Fiorenzo-Catalano (2007) in which the authors stochastically vary network attributes and attribute preferences around their measured values to generate a route choice set.

There are papers and researchers who recognize superiority in using the double stochastic approach (Prato et al 2014, Duncan et al 2020), there is not a common opinion of a strong preference of a different method over a doubly stochastic modeling approach. While many models do use double stochasticity to model route choice, there are still known limitations in the methodology and how it represents real life decision making. Additionally, Halldórsdóttir et al. recognize in a comparison of Doubly Stochastic Generation Function (DSGF), Breadth First Search on Link Elimination (BFS-LE) and Branch and Bound (B&B) algorithms, that while BFS-LE and DSGF both produce realistic routes, BFS-LE out performed in computational cost, while DSGF produced more heterogeneity in the routes produced.

Topic: Possible metrics to evaluate bike model outputs, specifically between base vs build scenarios.

Studies and tools (like CRUSE) for development of bicycle network planning (Lovelace, et al 2024) reflect a need to consider more than the distance in the successful uptake in bicycle infrastructure use. Studies have indicated that distance for leisure trips made by bicycle are longer than those that are made for non-leisure trips. Cycle Route Uptake and Scenario Estimation (CRUSE) uses metrics like quiet routes (one in which busy intersections are minimized and routes preferentially treat bicycle paths over roads and interactions between vehicles and bicycles) and green segments, which lead to longer distances for routes, but are preferable due to the lower traffic stress for the cyclist. In this work, the authors model the cycling network, simulate a variety of trip purposes and trips to supplement origin/destination data, and produce estimates of the quietness of a route for each route segment (a proxy for cyclist comfort and route preference). Through this effort, the authors generate cycling potentials on every road in the country of Ireland. Through analyzing the network for cycling potential, the authors are able to suggest prioritizing a continuous quiet network first and eventually to “cycle proof” the fastest network, making the faster network more amenable to bicycling (Lovelace, et al 2024). Observations of detouring for the sake of a safer, protective environment have been identified by (Winters, et al 2010) who found that the observed bicycle routes taken by their study participants had an average of two more traffic-calming features, 10 more stencils and seven more signs than the predicted shortest route. However, other features such as hills, air pollution, greenness and street connectivity did not lead to significant differences between the actual and shortest routes.

There has been some research in developing bike network connectivity analytic methods, which aid in infrastructure development by prioritizing accessibility to activity destinations (Iacono et al., 2010, Lowry and Loh, 2017, Furth et al., 2018, Kent and Karner, 2019). This is done by using shortest-path algorithms to route potential trips on eligible links and summarize the number of opportunities available in a specified distance. As mentioned in Gehrke however, these tools simply rely on distance, travel time and opportunities, and don’t consider other factors like bicycle stress on links or the differential value of nearby versus distant opportunities (Gherke et. al, 2020).

4.0 Citations

Ahmad, F., & Al-Fagih, L. (2024). Travel behaviour and game theory: A review of route choice modeling behaviour. In *Journal of Choice Modelling* (Vol. 50). Elsevier Ltd. https://doi.org/10.1016/j.jocm.2024.100472

Anker, O., & Giacomo, C. (2025a). General rights Efficiency of choice set generation methods for bicycle routes. In *Downloaded from orbit.dtu.dk on*.

Bernardi, S., Geurs, K., & Puello, L. L. P. (2018). Modelling route choice of dutch cyclists using smartphone data. *Journal of Transport and Land Use*, *11*(1), 883–900. https://doi.org/10.5198/jtlu.2018.1143

Bovy, P. H. L., & Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: Behavioral and probabilistic foundations. *Transportmetrica*, *3*(3), 173–189. https://doi.org/10.1080/18128600708685672

Broach, J., & Dill, J. (2016). Using predicted bicyclist and pedestrian route choice to enhance mode choice models. *Transportation Research Record*, *2564*, 52–59. https://doi.org/10.3141/2564-06

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005

Chen, P., Shen, Q., & Childress, S. (2018). A GPS data-based analysis of built environment influences on bicyclist route preferences. *International Journal of Sustainable Transportation, 12*(3), 218–231.

Dane, G., Feng, T., Luub, F., & Arentze, T. (2020). Route choice decisions of e-bike users: Analysis of GPS
tracking data in the Netherlands. Geospatial Technologies for Local and Regional Development: Proceedings of the 22nd AGILE Conference on Geographic Information Science 22 (pp. 109–124).

Duncan, L. C., Watling, D. P., Connors, R. D., Rasmussen, T. K., & Nielsen, O. A. (2020). Path Size Logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. *Transportation Research Part B: Methodological*, *135*, 1–40. https://doi.org/10.1016/j.trb.2020.02.006

Fitch, D. T., & Handy, S. L. (2020). Road environments and bicyclist route choice: The cases of Davis and San Francisco, CA. *Journal of Transport Geography*, *85*. https://doi.org/10.1016/j.jtrangeo.2020.102705

Furth, P.G., Putta, T.V.V.K., Moser, P., 2018. Measuring low-stress connectivity in terms of bike-accessible jobs and potential bike-to-work trips: A case study evaluating alternative bike route alignments in Northern Delaware. J. Transp. Land Use 11 (1), 815–831.

Gehrke, S. R., Akhavan, A., Furth, P. G., Wang, Q., & Reardon, T. G. (2020). A cycling-focused accessibility tool to support regional bike network connectivity. *Transportation Research Part D: Transport and Environment*, *85*. https://doi.org/10.1016/j.trd.2020.102388

Ghanayim, M., & Bekhor, S. (2018). Modelling bicycle route choice using data from a GPS-assisted household survey. *EJTIR Issue*, *18*(2), 158–177. http://tlo.tbm.tudelft.nl/ejtir

González, F., Melo-Riquelme, C., & de Grange, L. (2016). A combined destination and route choice model for a bicycle sharing system. *Transportation*, *43*(3), 407–423. https://doi.org/10.1007/s11116-015-9581-6

Grudgings, N., Hagen-Zanker, A., Hughes, S., Gatersleben, B., Woodall, M., & Bryans, W. (2018). Why don't more women cycle? An analysis of female and male commuter cycling mode-share in England and Wales. *Journal of Transport and Health*, *10*, 272–283. https://doi.org/10.1016/j.jth.2018.07.004

Huber, S., Lißner, S., Lindemann, P., Muthmann, K., Schnabel, A., & Friedl, J. (2021). Modelling bicycle route choice in German cities using open data, MNL and the Bikesim web-app. 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) (pp. 1–6).

Iacono, M., Krizek, K.J., El-Geneidy, A., 2010. Measuring non-motorized accessibility: Issues, alternatives, and execution. J. Transp. Geogr. 18, 133–140.

Kent, M., Karner, A., 2019. Prioritizing low-stress and equitable bicycle networks using neighborhood-based accessibility measures. Int. J. Sustain. Transp. 13 (2),100–110.

Li, S., Muresan, M., & Fu, L. (2017). Cycling in toronto, ontario, canada: Route choice behavior and implications for infrastructure planning. Transportation Research Record: Journal of the Transportation Research Board, 2662(1), 41–49. https://doi.org/10.3141/2662-05

Lovelace, R., Talbot, J., Vidal-Tortosa, E., Mahfouz, H., Brick, E., Wright, P., O'Toole, G., Brennan, D., & Meade, S. (2024). Cycle route uptake and scenario estimation (CRUSE): an approach for developing strategic cycle network planning tools. *European Transport Research Review*, *16*(1). https://doi.org/10.1186/s12544-024-00668-8

Lowry, M., Loh, T.H., 2017. Quantifying bicycle network connectivity. Prev. Med. 95, S134–S140.

Łukawska, M. (2024). Quantitative modelling of cyclists' route choice behaviour on utilitarian trips based on GPS data: associated factors and behavioural implications. *Transport Reviews*, *44*(5), 1045–1076. https://doi.org/10.1080/01441647.2024.2355468

Łukawska, M., Paulsen, M., Rasmussen, T. K., Jensen, A. F., & Nielsen, O. A. (2023). A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset. *Transportation Research Part A: Policy and Practice*, *176*. https://doi.org/10.1016/j.tra.2023.103834

Meister, A., Felder, M., Schmid, B., & Axhausen, K. W. (2023). Route choice modeling for cyclists on urban networks. *Transportation Research Part A: Policy and Practice*, *173*. https://doi.org/10.1016/j.tra.2023.103723

Meister, A., Liang, Z., Felder, M., & Axhausen, K. W. (2024). Comparative study of route choice models for cyclists. *Journal of Cycling and Micromobility Research*, *2*, 100018. https://doi.org/10.1016/j.jcmr.2024.100018

Misra, A., & Watkins, K. (2018). Modeling Cyclist Route Choice using Revealed Preference Data: An Age and Gender Perspective. *Transportation Research Record*, *2672*(3), 145–154. https://doi.org/10.1177/0361198118798968

Muhs, C. D., & Clifton, K. J. (2016). Do characteristics of walkable environments support bicycling? Toward a definition of bicycle-supported development. *Journal of Transport and Land Use*, *9*(2), 147–188. https://doi.org/10.5198/jtlu.2015.727

Nielsen, T. A. S., & Skov-Petersen, H. (2018). Bikeability – Urban structures supporting cycling. Effects of local, urban and regional scale urban form factors on cycling from home and workplace locations in Denmark. *Journal of Transport Geography*, *69*, 36–44. https://doi.org/10.1016/j.jtrangeo.2018.04.015

Park, Y., & Akar, G. (2019). Why do bicyclists take detours? A multilevel regression model using smartphone GPS data. *Journal of Transport Geography*, *74*, 191–200. https://doi.org/10.1016/j.jtrangeo.2018.11.013

Pearson, L., Berkovic, D., Reeder, S., Gabbe, B., & Beck, B. (2023). Adults' self-reported barriers and enablers to riding a bike for transport: a systematic review. *Transport Reviews*, *43*(3), 356–384. https://doi.org/10.1080/01441647.2022.2113570

Pearson, L., Gabbe, B., Reeder, S., & Beck, B. (2023). Barriers and enablers of bike riding for transport and recreational purposes in Australia. *Journal of Transport and Health*, *28*. https://doi.org/10.1016/j.jth.2022.101538

Prato, C. G., Halldórsdóttir, K., & Nielsen, O. A. (2018). Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space. *International Journal of Sustainable Transportation*, *12*(10), 770–781. https://doi.org/10.1080/15568318.2018.1437236

Rosas-Satizábal, D., Guzman, L. A., & Oviedo, D. (2020). Cycling diversity, accessibility, and equality: An analysis of cycling commuting in Bogotá. *Transportation Research Part D: Transport and Environment*, *88*. https://doi.org/10.1016/j.trd.2020.102562

Scott, D. M., Lu, W., & Brown, M. J. (2021). Route choice of bike share users: Leveraging GPS data to derive choice sets. *Journal of Transport Geography*, *90*. https://doi.org/10.1016/j.jtrangeo.2020.102903

Segadilha, A. B. P., & Sanches, S. da P. (2014). Identification of Factors that Influence Cyclistś Route Choice. *Procedia - Social and Behavioral Sciences*, *160*, 372–380. https://doi.org/10.1016/j.sbspro.2014.12.149

Tarkkala, K., Vosough, S., West, J., & Roncoli, C. (2024). Evaluating the influence of cyclists' route choices incorporation into travel demand modelling: A case study in greater Helsinki. *Transportation Research Interdisciplinary Perspectives*, *27*. https://doi.org/10.1016/j.trip.2024.101224

Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. (2017). How do people cycle in amsterdam, Netherlands?: Estimating cyclists' route choice determinants with GPS data from an Urban area. *Transportation Research Record*, *2662*(1), 75–82. https://doi.org/10.3141/2662-09

Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society*, *13*, 105–117. https://doi.org/10.1016/j.tbs.2018.07.001

Wang, H., Palm, M., Chen, C., Vogt, R., & Wang, Y. (2016). Does bicycle network level of traffic stress (LTS) explain bicycle travel behavior? Mixed results from an Oregon case study. *Journal of Transport Geography*, *57*, 8–18. https://doi.org/10.1016/j.jtrangeo.2016.08.016

Watkins, K. E., le Dantec, C., Mokhtarian, P., Dilkina, B., & Laval, J. (n.d.). *MAPPING BICYCLIST ROUTE CHOICE USING SMARTPHONE BASED CROWDSOURCED DATA*.

Winters, M., Teschke, K., Grant, M., Setton, E.M., Brauer, M., 2010. How far out of the way will we travel? Transp. Res. Rec. J. Transp. Res. Board 2190 (1), 1–10.

Xie, L., & Spinney, J. (2018). "I won't cycle on a route like this; I don't think I fully understood what isolation meant": A critical evaluation of the safety principles in Cycling Level of Service (CLoS) tools from a gender perspective. *Travel Behaviour and Society*, *13*, 197–213. https://doi.org/10.1016/j.tbs.2018.07.002

Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modeling using GPS data without choice sets of paths. *Transportation Research Part C: Emerging Technologies*, *75*, 183–196. https://doi.org/10.1016/j.trc.2016.12.009

APPENDIX C. COEFFICIENT CONVERSION TABLE

Term	Conversion					
Length			-6.790	utiles/km	-1.093E+01	utiles/mi
Med uphill (1%-3.5%)	-0.056	utiles/dam	-5.600	utiles/km	-9.012E+00	utiles/mi
Steep uphill (>3.5%)	-0.113	utiles/dam	-11.300	utiles/km	-1.819E+01	utiles/mi
Medium roads w/ protected lanes			0.000	utiles/km	0.000E+00	utiles/mi
Medium roads w/ painted lanes			-0.340	utiles/km	-5.464E-01	utiles/mi
Medium roads w/o lanes			-0.767	utiles/km	-1.235E+00	utiles/mi
Large roads w/ protected lanes			0.109	utiles/km	1.748E-01	utiles/mi
Large roads w/ painted lanes			-1.962	utiles/km	-3.158E+00	utiles/mi
Large roads w/o lanes			-1.562	utiles/km	-2.513E+00	utiles/mi
Residential roads w/ protected lanes			-0.611	utiles/km	-9.835E-01	utiles/mi
Residential roads w/ painted lanes			0.577	utiles/km	9.288E-01	utiles/mi
Residential roads w/o lanes			-1.181	utiles/km	-1.901E+00	utiles/mi
Cycleways			0.258	utiles/km	4.152E-01	utiles/mi
Shared paths			-1.059	utiles/km	-1.705E+00	utiles/mi
Footways			-3.436	utiles/km	-5.529E+00	utiles/mi
Ped zones			-2.499	utiles/km	-4.021E+00	utiles/mi
Wrong way			-1.989	utiles/km	-3.202E+00	utiles/mi

100 dam = 1 km

1.609344 km = 1 mi

the science of insight