# Calibrating Route Choice Sets for an Urban Public Transport Network using Smart Card Data


**Presenter:**

Sanmay Shelat

**Authors:**

S. Shelat, O. Cats, N. van Oort, J.W.C. van Lint
*Dept. of Transport and Planning,*
*Delft University of Technology*


TU Delft

# Introduction

## Problem + Literature

- Choice set identification is important for estimation and application of route choice models
- It is a non-trivial task due to its combinatorial nature and dependence on traveller preferences
- Choice set identification approaches in literature can be classified as:
  1. Direct identification
     a) Asking travellers what they considered
     b) Routes observed from population (smart card data)
  2. Choice set generation methodology (CSGM)
     a) Deterministic/stochastic shortest path
     b) Constrained enumeration

# Introduction
## Contributions

- We propose a constrained enumeration route CSGM that:

Avoids subjective assumptions regarding traveller preferences …

… by using (increasingly available) smart card data to calibrate parameters …

… of an intuitive and accepted behaviour model.

# Behavioural Model

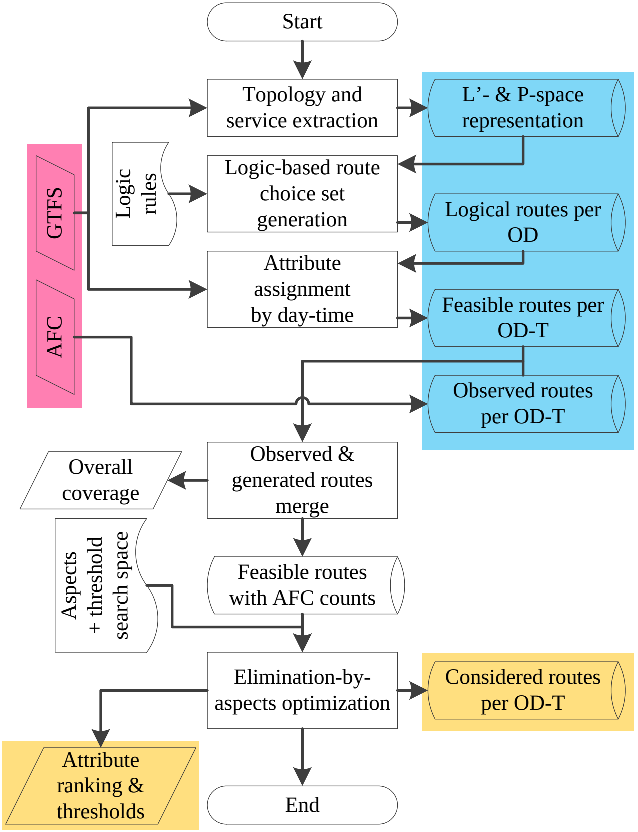
## Non-compensatory Decision Models

- Non-compensatory evaluation is typical for choice set formation from a large number of alternatives

1. Disjunctive/Conjunctive
   - Sets minimum thresholds for important attributes; either comply with at least one or require all thresholds to be met (e.g., detour thresholds)
2. Lexicographic
   - Attributes ranked by importance; alternatives selected on the basis of performance in top-ranking attribute (e.g., link-labelling approach)
3. Elimination-by-aspects (deterministic)
   - Combines attribute ranking and setting thresholds
   - Used in this study

# Choice Set Generation Methodology

## Overview



# Choice Set Generation Methodology

## Representation → Generated-feasible Routes

- Network represented as infrastructure (L-space) and service (P-space) graphs
- A breadth-first search algorithm is used to enumerate routes, constrained:
    - Depth-wise by a maximum of two transfers
    - Breadth-wise by disallowing loops and transfers between common lines
- The following attributes are assigned to the route alternatives for each hour:
    - Waiting time (frequency-based), in-vehicle time, number of transfers
- Infeasible routes (no service) and dominated alternatives are removed
- For the EBA calibration, observed routes not in generated-feasible route set are discarded

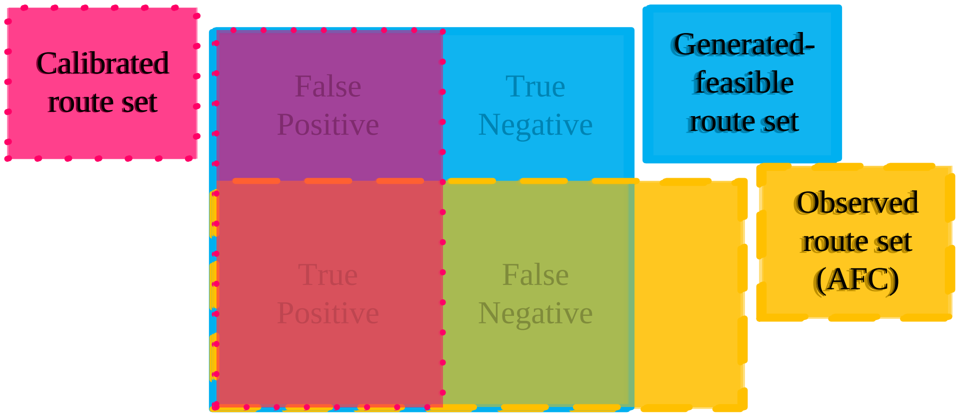
# Choice Set Generation Methodology

## EBA: Calibration Route Set

- The elimination-by-aspects (EBA) method has the following parameters:
    1. Attribute ranks
    2. Attribute thresholds
- For a given EBA parameter set:
    - The calibrated route set is obtained from the generated-feasible route set
    - By removing routes that perform worse than threshold on attributes, sequentially in order of descending attribute rank
- To obtain the optimal behavioural parameters, we need performance indicators

# Choice Set Generation Methodology

## EBA: Indicators




$$coverage = \frac{\sum_{i,j} q_{ij}^{TP}}{\sum_{i,j} q_{ij}^{TP} + q_{ij}^{FN}}$$

$$efficiency = \frac{\sum_{i,j} \left|R_{ij}^{TN}\right| q_{ij}}{\sum_{i,j} \left(\left|R_{ij}^{FP}\right| + \left|R_{ij}^{TN}\right|\right) q_{ij}}$$

$$\min\left(abs\left(coverage_a - efficiency_a\right)\right)$$

# Choice Set Generation Methodology

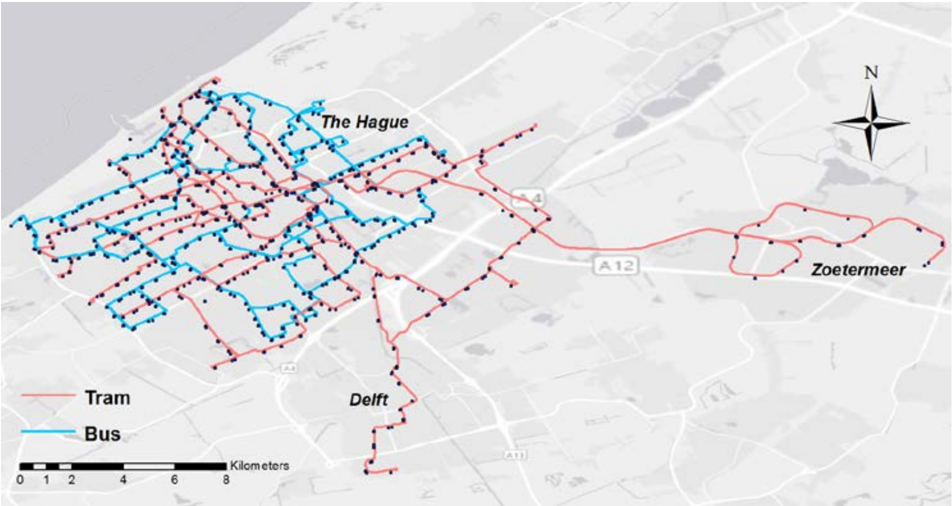
## EBA: Brute Force Optimization

- Generally, only a few (but important) aspects/attributes are available from smart card data
- Furthermore, it is reasonable to expect that:
    - Potential thresholds are close to smallest values
    - A very high precision (<0.001) in threshold values is not required (because the differences will be imperceptible to choice makers)
- Therefore, it is feasible to employ a brute force algorithm to obtain the optimal EBA parameters

# Case Study: The Hague Tram+Bus

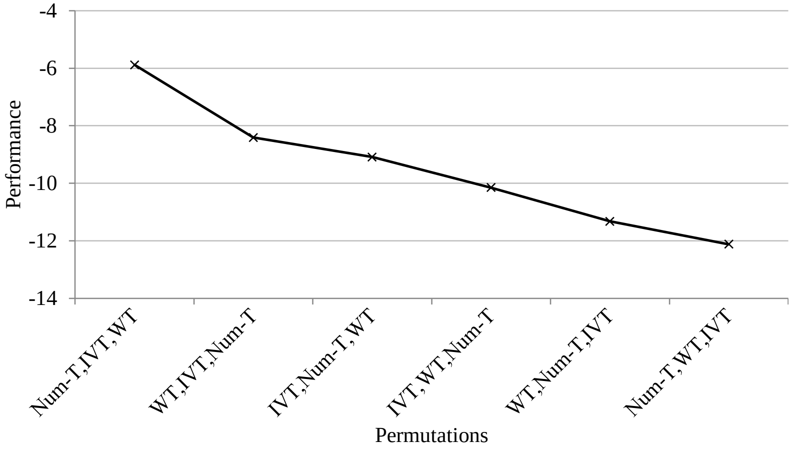
## Description

- 12 Tram + 8 Bus bidirectional lines serving 459 stations
- Smart card data from March 2015
- Weekdays, 0600h-1100h



# Case Study: The Hague Tram+Bus

## Results




Performance for different attribute rankings (lower is better)

- Num-T: number of transfers,
- WT: waiting time,
- IVT: in-vehicle time

| Rank | Attribute | Threshold | Sequential Coverage |
|---|---|---|---|
| **1** | Number of transfers | 0 | 99.3% |
| **2** | Waiting time | 1.1 | 82.0% |
| **3** | In-vehicle time | 1.1 | 78.4% |

Optimal attribute ranking and thresholds

# Case Study: The Hague Tram+Bus

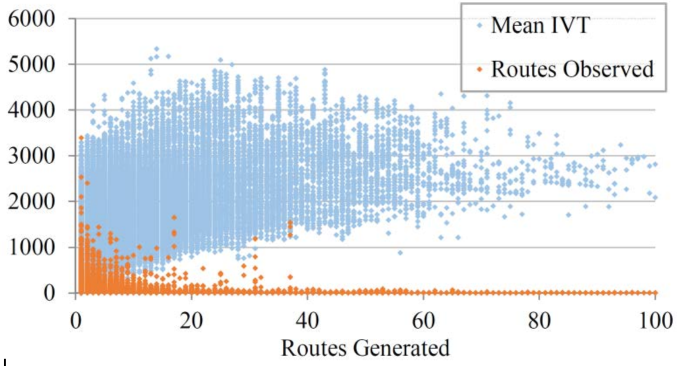
## Discussion

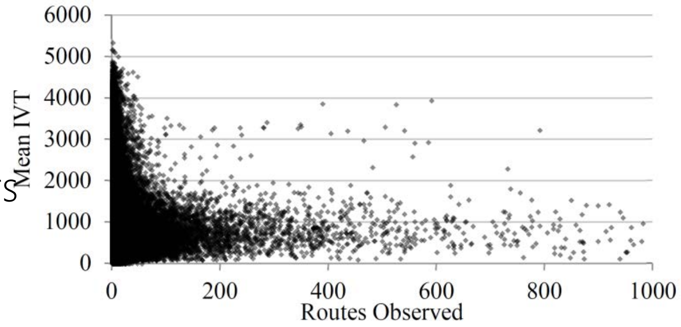
- Possible explanation for the rather restrictive threshold values:
- Some observations:

1. OD pairs with high demand are nearby
   - Routes Observed ↑- Mean IVT ↓ (black)
2. OD pairs with more route alternatives are farther
   - Routes Generated ↑- Mean IVT ↑ (blue)

- An hypothesis:

3. Travellers have stricter thresholds for OD pairs that are nearby




TUDelft

# Case Study: The Hague Tram+Bus

## Discussion

1. OD pairs with high demand are nearby

3. Travellers have stricter thresholds for OD pairs that are nearby

$$coverage = \frac{\sum_{i,j} q_{ij}^{TP}}{\sum_{i,j} q_{ij}^{TP} + q_{ij}^{FN}}$$

⇒ Coverage values are high at low thresholds

2. OD pairs with more route alternatives are farther

$$efficiency = \frac{\sum_{i,j} |R_{ij}^{TN}| q_{ij}}{\sum_{i,j} (|R_{ij}^{FP}| + |R_{ij}^{TN}|) q_{ij}}$$

⇒ Non-selected alternatives between far-away OD pairs disproportionately affect efficiency

TU Delft

# Conclusions

- A constrained enumeration CSGM is developed that employs deterministic elimination-by-aspects as the behavioural model which is calibrated using smart card data
- Results from the urban public transport network in the Hague show that num. of transfers is the most important factor, followed by waiting time and in-vehicle time and that the thresholds for these are quite restrictive thresholds
- Further research aims to overcome the assumption of a frequency-based system, account for possible reasons underlying restrictive thresholds, and compare results from different non-compensatory decision models.