Improving Transit System Performance with the Benefit of Automatic Data Collection Systems

Nigel H.M. Wilson
Transit Agencies Are at a Critical Transition in Data Collection Technology:

**Manual**
- low capital cost
- high marginal cost
- small sample sizes
- aggregate
- unreliable
- limited spatially and temporally

**Automatic**
- high capital cost
- low marginal cost
- large sample sizes
- more detailed, disaggregate
  - errors and biases can be estimated and corrected
  - ubiquitous

Examples of ADC systems:
- **AVL:** Automatic Vehicle Location Systems
- **APC:** Automated Passenger Counting Systems
- **AFC:** Automated Fare Collection Systems
Existing ADCS Applications to Public Transport Functions

- Service and Operations Planning
  - Sample application: TRITAPT*

- Service Control and Management
  - Sample application: CAD/AVL

- Customer Information
  - Sample application: Next vehicle arrival systems

- Performance Measurement and Monitoring
  - Sample application: TRITAPT

Applications can be:

- off-line
- real-time

- TRIp Time Analysis in Public Transport
  (Developed by Theo Muller at T.U. Delft)
ADCS - Potential and Reality

**Potential**

- Integrated ADCS database
- Software to support many agency decisions using ADCS database
- Providing insight into normal operations, special events, unusual weather, etc.

**Reality**

- Most ADCS systems are implemented independently
- Data collection is ancillary to primary ADC function
  - AVL - emergency notification, stop announcements
  - AFC - fare collection and revenue protection
- Many problems to overcome:
  - not easy to integrate data
  - requires substantial resources
Three Applications

• New Metrics for London Underground Journey Times\(^1\)
• Estimation of OD Matrix for London Underground\(^1\)
• Bus Passenger OD Matrix Estimation in Chicago\(^2\)

New Metrics for London Underground Journey Times

- London has about 75% market penetration with the Oyster smart card system
- Tap-in and tap-out provides total time on system for most Oyster card users
- Large sample sizes permit characterization of travel time distributions for many OD pairs
- Provide a customer perspective on service times
Traditional London Underground Travel Time Metric

JTM -- Journey Time Metric; average passenger excess time by line and for network

- compares measured and scheduled travel times
- relies on field surveys, automated data sources and models
- used as performance monitoring and management tool

Weaknesses:

- reliance on multiple data sources
- modeled components are not easy to update
- field surveys are expensive
- masks the impact of unreliability
Oyster-based Journey Time Metrics

Oyster elapsed journey time characteristics:

OD level
• Need to exclude extreme journey times that represent individual behavior instead of service performance

Line level
• Considers only same-line journeys
• OD level results are weighed by demand

Results are calculated from 100% Oyster sample
• For any time period and over any number of days, e.g. monitor day-to-day fluctuations
Oyster-based Journey Time Metrics

- Waterloo to Canary Wharf (59609)
- Bounds Green to South Kensington (1310)
- North Harrow to Farringdon (1301)
- Uxbridge to Bank & Monument (149)
- Kingsbury to Ladbroke Grove (80)
Oyster-based Journey Time Metrics
Oyster-based Journey Time Metrics

![Graph showing journey time metrics]

- Actual Journey Time
- Scheduled Journey Time

Minutes

Journey Time Percentile
Journey Time Reliability Metric (JTRM)

OD Reliability Factor = Upper Threshold – Median JT

- Upper threshold should depend on:
  - OD journey time distribution
  - Desired sensitivity of reliability results
  - Service standards of transit agency

- Sample Size Issues
  - Discard OD pairs with fewer than 20 samples
  - For OD sample size between 20 and 200,
    Upper Threshold = Max (Second maximum journey time, 95th percentile JT)
  - For OD sample size larger than 200,
    Upper Threshold = 95th percentile JT

Line Level Reliability Factor

- Considers only same-line “trunk” journeys to minimize headway and journey length variations across lines
- “Trunk” is the portion of the line served by the most frequent service
JTRM Results (1) – Period Average

- Bakerloo (16/25) • Central (20/49) • Jubilee (17/27) • Piccadilly (24/51) • Victoria (13/16)
Conclusions – Journey Time Metrics

• Limitations and Challenges
  • Difficult to disaggregate journey time components to evaluate causes of delays and unreliability
  • Assumption of identical OD journey time distributions of Oyster and magnetic stripe users
  • Consideration of only trunk journeys when over 40% of same-line journeys are non-trunk

• Suggested Improvements
  • More accurate scheduled access and egress times: station-specific access/egress factors are needed to transform the JTM scheduled times
  • Extension to non-trunk journeys for JTRM
  • Extension to network level results: accurate path choice model is needed to assign journeys onto different lines
  • Separation of in-train and out-of-train times using train operations data
Estimation of OD Matrix for London Underground

Traditional Approach:
• Rail Origin Destination Surveys (RODS) interview ≈20,000 Underground users per year to document OD patterns
• Expand to full OD matrix based on station entry and exit control totals

Oyster Approach:
• Expand Oyster OD matrix to entry and exit control totals
Qualitative Comparison between RODS and Oyster-based Methodologies

Advantages of RODS over Oyster:
- Path choice information
- Access and egress mode information
- Fair representation of all passengers in RODS

Advantages of Oyster over RODS:
- Sampling frequency
- Strong and increasing sample size
- Data Completeness
- Cost Effectiveness
AM Peak OD Results: Network

- Oyster-based matrix contains 1% more journeys than RODS (probably due to ridership increase from November 2005 to January 2007)
- The common OD pairs capture 87% of Oyster-based journeys and 98% of RODS journeys

- OD pairs captured in both RODS and Oyster-Based Matrices
  16,465
  (22.2%)

- OD pairs captured in Oyster-Based Matrix only
  25,436
  (34.3%)

- OD pairs captured in RODS Matrix only
  956
  (1.3%)

- OD pairs with no observed travel
  31,399
  (42.3%)
AM Peak OD Results: Oyster Bias

- The methodology corrected the under-representation bias of Oyster journeys at non fully gated stations.
AM Peak OD Results: 50 Largest Origins

These 50 stations account for 53% of entries in the AM Peak
AM Peak OD Results:
Destination Coverage

- On average, an origin station has 90 more destinations (150%) in Oyster than in RODS
AM Peak OD Results: Critical Link Loads

- 15 critical links: 8 Central WB, 5 Victoria NB and 2 Victoria SB
- Oyster-based and RODS matrices are assigned by computing the minimum generalized cost of travel in TransCAD (without calibration)
  - RODS link loads are systematically higher than the assigned matrices, with an average difference of 5%
  - Average difference between the RODS and Oyster-based assigned matrices is <2%
Conclusions – Time Period Level OD Matrix

• Limitations of Oyster-based methodology
  • Requires accurate station level counts which are hindered by annual manual count surveys at non fully gated stations
  • May be inaccurate for time periods shorter than one hour due to the use of RODS proportions to calculate expansion factors

• Suggested modifications of RODS for the short run
  • Simplify survey to increase response rate
  • Target at non fully gated stations where Oyster data are weak
  • Target at passengers using magnetic stripe tickets
Conclusions – Time Period Level OD Matrix

- Innovative strategies for the long run
  - More frequent manual count surveys or installing video-based automatic passenger count technology
- Cost effective ways to collect path choice information
  - Extensive Path Choice Survey – valid until the network changes
  - Automatic path choice data collection by installing Oyster readers at interchange stations
  - Internet-based surveys – target at Oyster user’s most recent transfer journey
- Reduction of Oyster bias at non fully gated stations
Bus Passenger OD Matrix Estimation in Chicago

Objective:

• Estimate bus passenger OD matrix for CTA at:
  • single route level
  • network level

CTA Network attributes:

• multi-modal rail and bus system
• entry-control-only operations
Trip Chaining: Basic Idea

Each AFC record includes:

• AFC card ID
• transaction type
• transaction time
• transaction location: rail station or bus route
Each AFC record includes:

- AFC card ID
- transaction type
- transaction time
- transaction location: rail station or bus route
Trip Chaining: Basic Idea

Each AFC record includes:
- AFC card ID
- transaction type
- transaction time
- transaction location: rail station or bus route

The destination of many trip segments is also the origin of the following trip segment.
Trip Chaining: Basic Idea

Each AFC record includes:

- AFC card ID
- transaction type
- transaction time
- transaction location: rail station or bus route

The destination of many trip segments (TS) is also the origin of the following trip segment.
Trip Chaining: Basic Idea

Each AFC record includes:

- AFC card ID
- transaction type
- transaction time
- transaction location: rail station or bus route

The destination of many trip segments (TS) is also the origin of the following trip segment.

**Note:**
1) each bus boarding requires a new AFC transaction: $TS_{bus}$ represents an unlinked bus trip
2) rail-rail transfers do not require a new AFC transaction: $TS_{rail}$ represents a path on the rail network
Trip-Chaining Method for OD Inference

Based on three assumptions:

Key Assumptions for Destination Inference to be correct:
- No intermediate private transportation mode trip segment
- Passengers will not walk a long distance
- Last trip of a day ends at the origin of the first trip of the day
Current Project: Bus OD Matrix: Description of data sources

All bus passenger trips

Bus passenger trips using farecard (~60%*)

Can identify boarding stop using AVL (~80%)

Can infer destination (~50% → seed matrix)

APC data (~15% sample)

*The percentages are based on 2005 data
Summary Information on the data used

- Overall AFC data (single weekday, all bus routes):
  - 545,000 bus passenger trips using farecard
  - From these, 436,000 with boarding stop (~80% identification rate)
  - From these, 244,000 with destination (~56% inference rate)
Overview of the approach

1. Data Cleaning and Processing (AFC, APC, AVL, Transfer card)

2. Seed Matrix from Trip-Chaining Method

3. Obtain single-route OD matrixes
   Using Iterative Proportional Fitting (IPF) or Maximum Likelihood Estimation (MLE)*

4. Joining Single Routes incorporating transfer flows

*These are two techniques to combine APC data with seed matrix to get full OD matrix
Input:

• Boarding and alighting counts for all stops (from APC and AVL systems)
• Sample of passenger OD to serve as the seed matrix (from AFC system)
• Total transfer flows (from AFC and APC data)

Key steps:

• Single route OD estimation
• Network linked OD estimation