Integrating Origin and Destination (OD) Study into GIS in Support of LIRR Services and Network Improvements

GIS In Transit Conference, Washington, DC

Pyung ho Kim
Date: September 3, 2015
Over 180 years of service - Chartered in 1834
Operating 24/7, averaging 298,429 customers each weekday on 740 daily trains
11 branches/124 passenger stations
670 track miles
Over 6,700 employees
$1.9 billion dollar annual operating budget (2015)
Background - Unique Needs of LIRR

- Survey Needs
  - Open system without turnstiles/fare gates and complex fare structure with multiple ticket types and passes
  - Limited customer travel behavior data collected through ticket sales

- Frequency
  - Undertaken approximately every 7 years (e.g., 1998, 2006, 2013)

- Methods
  - On/off passenger counts undertaken onboard all LIRR trains
  - Self-administered customer surveys distributed on westbound trains
  - Over 119,000 completed survey records
Purpose of OD Survey

- Provide a current understanding of LIRR customers’ trip patterns and behaviors
- Serve as an input into the planning process for LIRR operations
- Enhance and Validate MTA’s Regional Transit Forecasting Model
Rationale of Integrating OD survey into GIS Framework

Trip Patterns & Behaviors (Purpose, frequency, boarding and alighting station, origin and destination, socio-demographic characteristics)

Temporal Components (Temporal Constraints):
- Work schedule
- Train service schedule
- Time budget, etc.

Role of GIS:
- Convert non-geographic data into GIS data
- Relate trips within places
- Diagnose connectivity amongst trips and places
- Provide location analytics

Geographic Components (Spatial Constraints):
- Home/work
- Distance to station
- Parking lot, etc.

Long Island Rail Road
Conduct a survey and expand the overall sample by the station count (control)

Geocode survey data

Build Geodatabase (survey and subsidiary datasets)

Build Network data (base vs. scenario-specific)

Run analysis or model

Visualize the results (e.g., maps, diagrams, tables, web-based)
Integration/OD Project Management

Project Organization Structure

- Project Name
  - Geoprocess
  - GIS output
  - History
  - Original data
  - Reference data
    - Reference docs
      - email
      - gis request
      - meeting
      - memo
      - ppt
      - tsow
  - report

Implementation of Project Management

- Files and directories:
  - PID14_1_MN_Penn_GCT_Analysis
    - geoprocess
    - gis_output
    - analysis
    - mn_trips
      - mn_trips_wgs84
    - WB_Trips_SP
    - WB_Trips_WGS84
      - coastline
    - od_survey_2013.gdb
      - mn_trips
      - mn_trips_wgs84
      - WB_Trips_SP
      - WB_Trips_WGS84
      - q17_geo_all_wgs84_85273
      - od_survey_2013_1.gdb
      - od_2013_mn_destination_q17zone_60572_v1.xlsx
      - od_2013_survey_32715.xlsx
  - history
  - original_data
  - ref_data
  - ref_docs
  - report
  - PID14_1_MN_Destination_GCT_Penn_Analysis_040915.mxd
  - PID14_1_MN_Destination_GCT_Penn_Analysis_041015.mxd
Type of LIRR OD Analyses

- Potential station ridership analysis (e.g., Manhattan - Penn vs. Grand Central Terminal (GCT) analysis)

- Origin and destination paired with boarding and alighting station analysis (e.g., Port Washington (PW) station ridership to GCT/Penn)

- Origin vs. closest station analysis by travel time and distance (e.g., Ronkonkoma station ridership)
Why Manhattan?

Importance of Manhattan-bound trips

- Destination for 88% of all AM peak trips
- Major employment center with two Central Business Districts
- Currently served by one LIRR terminal (Penn Station)

In summary, both analyses (Euclidean distance & Network travel time) result that 36% of our customers will better served by GCT.
Manhattan Terminal Analysis (Penn vs. GCT) – Euclidean Distance Analysis

Step 1: Prepare Subway Network Data
Step 2: Conduct Proximity Analysis

Goal

- Identify potential LIRR customers (i.e., the number of customers), who would be better served by GCT considering travel distance

Data (Preliminary)

- Total trips: 119,014
- Manhattan-bound AM peak trips: 30,101
- Weighted Manhattan AM peak trips: 81,505

Result

<table>
<thead>
<tr>
<th>Trip Distribution Analysis/Euclidean Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCT/Penn</td>
</tr>
<tr>
<td>Trip Count (%)</td>
</tr>
<tr>
<td>Level 1/Linked (%)</td>
</tr>
</tbody>
</table>
Manhattan Terminal Analysis (Penn vs. GCT) : Network Travel Time Analysis (Spatial-temporal aspect)

Goal

Identify potential LIRR customers (i.e., the number of customers), who would be better served by GCT considering travel time with service planning schedule

Result

GCT and Penn Trip Distribution Analysis/Network-based Travel Time

<table>
<thead>
<tr>
<th>GCT/Penn</th>
<th>GCT</th>
<th>Penn</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip Count (%)</td>
<td>11,062(36.75%)</td>
<td>19,039(63.25%)</td>
<td>30,101(100%)</td>
</tr>
<tr>
<td>Level 1/Linked (%)</td>
<td>29,326(35.98%)</td>
<td>52,179(64.02%)</td>
<td>81,505(100%)</td>
</tr>
</tbody>
</table>

2013 OD Data
Manhattan Terminal Analysis (Penn vs. GCT)  
Port Washington(PW) Branch Stations

**Goal**

- Identify potential PW branch customers, who would be better served by GCT

**Result**

45% will be better served by GCT
Ronkonkoma Origin vs. Near Station Analysis – Travel Distance/Time (Network-based Analysis)

Data
Ronkonkoma Home-Station Customers (sample data)
1,355 of the total of 39,850 records (3.4%)

Result

<table>
<thead>
<tr>
<th>Shortest Network Distance Analysis</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorter Distance</td>
<td>7.6%</td>
</tr>
<tr>
<td>Same Distance</td>
<td>23.2%</td>
</tr>
<tr>
<td>More Distance</td>
<td>69.2%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shortest Network Time Analysis</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorter Time</td>
<td>0.0%</td>
</tr>
<tr>
<td>Same Time</td>
<td>23.2%</td>
</tr>
<tr>
<td>More Time</td>
<td>76.8%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Challenges and Issues

- **Geocoding Accuracy**
  - Data scrubbing/cleaning with over 119,000 survey trip records
  - Multiple address locators
  - Analysis requiring different location accuracy

- **Building Base Layers from Multiple Sources**
  - Boundary mismatch
  - Projection/coordinates differences
  - Building network database
Challenges and Issues - Geocoding Accuracy (1)

Data Scrubbing/Cleaning with over 119,000 survey trip records

Geocoding accuracy with different address locators

Solution

- Used Bing Map for the entire area except Manhattan; NYC DCP’s Geosupporter for Manhattan

Bing Map – geocode at street level

Geosupporter/NYC Planning Department (DCP) – geocode at building lot level
Challenges and Issues - Geocoding Accuracy (2)

Meeting trip sampling requirement for the analysis with different location accuracy

Solution
- Geocode the trips at different accuracy levels; groups (1 through 4)

Geocoding accuracy - Building-specific zip codes (e.g., zip code 10001 in Manhattan)
Challenges and Issues – Building base layers

Boundary mismatches

Difference in projection and coordinate system
- Original survey data in WGS 84 and LIRR’s GIS in Long Island State Plane

Building network database from various source data

Solution
- Pick data with accurate geography and add additional geography or necessary attributes from other data
- Re-project all the reference data into LI State Plane
- Choose a base network specific to individual analysis
Conclusion

- OD and other survey data are important input data for LIRR’s major services and network improvements.

- GIS can provide an effective analytical framework in revealing spatio-temporal aspects of LIRR customer travel patterns and behaviors by relating OD and other survey data with geography.

- Integrating OD study into GIS facilitates the process of identifying trip demand and service planning needs.
Comments and Questions?

Pyung ho Kim
Regional Transportation Planner
E-mail: phkim@lirr.org
Tel: 718-558-3656