Techniques for Establishing and Measuring Data Quality in an Archived Data User Service

Christopher Monsere
Kristin Tufte, Robert L. Bertini, and Soyoung Ahn
Intelligent Transportation Systems Laboratory
Maseeh College of Engineering and Computer Science
Portland State University
Data Quality

- As usage expands, so does data quality expectations
- Initial use of ATMS
  - Fixed-time ramp metering, incident identification
- Next
  - Traveler information
- Expanding Use
  - Adaptive ramp metering (SWARM), Performance measurement (travel time validation, bottleneck identification, emissions)
What’s in the PORTAL Database?

Loop Detector Data
20 s count, lane occupancy, speed from 500 detectors (1.2 mi spacing)

Incident Data
140,000 since 1999

Bus Data
1 year stop level data
140,000,000 rows

Weather Data
Every day since 2004

Days Since July 2004
About 300 GB
4.2 Million Detector Intervals

VMS Data
19 VMS since 1999

WIM Data
22 stations since 2005
30,026,606 trucks

Crash Data
All state-reported crashes since 1999 - ~580,000

001497

NATMEC 2008
Our Vision for Data Quality Improvement

PSU – Archive

ODOT - Operations

ODOT - Maintenance
PORTAL Data Quality Objectives

- Systematically identify erroneous detector data
- Prioritize detectors in need of maintenance
- Monitor and disseminate data quality information
PORTAL Data Fidelity Plot

(data fidelity 2006-07-26 highway I-5)

Interstate Bridge @ MP 308
I-84 @ MP 302
I-405 @ MP 300
US-217 @ MP 292

Time
Data Provided by Oregon DOT
Detection of Malfunctioning Detectors

- Count > 17 (≈ 3060 vph)*
- Occupancy > 95 %*
- Speed > 100 mph*
- Speed < 5 mph*
- Low maximum occupancy (thresholds vary by locations)
- Low average occupancy during peaks

* Adapted from criteria proposed by TTI
Identifying Malfunctioning Detectors

- Two primary categories of loop detector malfunction
  - Detector or controller configuration errors
  - Communication failures
## Configuration Errors Report

### Detect. | Location | Milepost | Lane | Tests Failed
---|---|---|---|---
1011 | I-5 N Rd WB to NB | 286.3 | 3 | Low Avg Occ
1244 | I- SB Swift Blvd | 307.35 | 1 | Speed < 5
1123 | I-5 N Multnomah Blvd NB | 296.6 | 3 | Occ > 95, Speed < 5
1204 | I-5 N Denver Ave NB | 306.51 | 2 | Low Max & Avg Occ
1205 | I-5 N Denver Ave NB | 306.51 | 3 | Low Avg Occ
1206 | I-5 N Denver Ave NB HOV | 306.51 | 1 | Low Avg Occ
1212 | I-5 N Delta Park NB | 306.51 | 2 | Low Max & Avg Occ
1695 | US 26 West Skyline Rd WB | 71.07 | 3 | Occ > 95
1637 | US 26 East 185th Ave SB | 64.6 | 2 | Occ > 95
1787 | I-205 43 NB-NB | 9.03 | 2 | Low Avg Occ
1481 | I-84 West 33rd WB | 2.1 | 2 | Speed < 5

All weekdays in June 2006
A Good Detector...
Observed Data Quality Issues

I-5 N Delta Park; Lane 2 (July 26, 2006)
I-5 S Swift Blvd; Lane 1 (July 27, 2006)
I-84 W 33rd; Lane 2 (July 27, 2006)
US 26 W Skyline; Lane 3 (July 20, 2006)
Site Visits

- Visited three detector stations with UW ALEDA tool for calibrating loop detectors

- Stations visited:
  - I-5 NB – Multnomah
    - High number of readings with occupancy > 95%
  - I-5 NB Terwilliger
    - Low maximum occupancy
  - I-5 NB OR 217/Kruse Way
    - Low overnight speeds
I-5 NB Terwilliger Blvd

Sept 19, 2007 - Before

Occupancy (%)

Time

0:00  4:00  8:00  12:00  16:00  20:00  0:00
I-5 NB Terwilliger Blvd

Sept 19, 2007 – Day of Visit
I-5 NB Terwilliger Blvd

Sept 21, 2007 - After
I-5 NB OR 217/Kruse Way

Sept 18, 2007; Lane 2

Sept 18, 2007; Lane 3

January 17, 2008; Lane 2

January 17, 2008; Lane 3
Communication failure is a second major cause of loop detector data errors.

Look for detectors reporting ‘communication failure’ and detectors reporting ‘no traffic’ during congested periods.

Communication failures communicated to ODOT subsequently fixed.
Impact of Adaptive Ramp Metering

I-205 NB, Pre-Timed, PM

- Division
- Powell
- Foster
- Johnson Creek
- Sunnyside
- Sunnybrook
- Lawnfield
- Clackamas Hwy
- Gladstone

I-205 NB, SWARM, PM

- Division
- Powell
- Foster
- Johnson Creek
- Sunnyside
- Sunnybrook
- Lawnfield
- Clackamas Hwy
- Gladstone

Communication Failures (%)
Conclusions

- Data quality is crucial to transportation data archives
- Detectors require regular maintenance; limited maintenance resources
- Need automated process to detect and report malfunctioning detectors
  - Requires communication between researchers and maintenance staff
  - Desire mapping between data anomalies and problems in the field
Acknowledgements

- National Science Foundation (PORTAL)
- Oregon Transportation Research Education and Consortium
- TransPORT
- Oregon DOT
- Yinhai Wang and Num Cheevarunothai
  University of Washington
Thank You!

www.its.pdx.edu
Bonus!
(2) Mainline speeds dropped prior to meter activation

(3) Note speed oscillations

(1) Metering activates at earliest possible time (1PM) under SWARM.

(4) While not confirmed by data collection, it is apparent in the archived data that metering activates at around 1:30PM under SWARM.