Implementing Comprehensive Traffic Data Quality Examination Tools: A Case Study for WisTransPortal V-SPOC Quality Assurance/Quality Control Enhancements

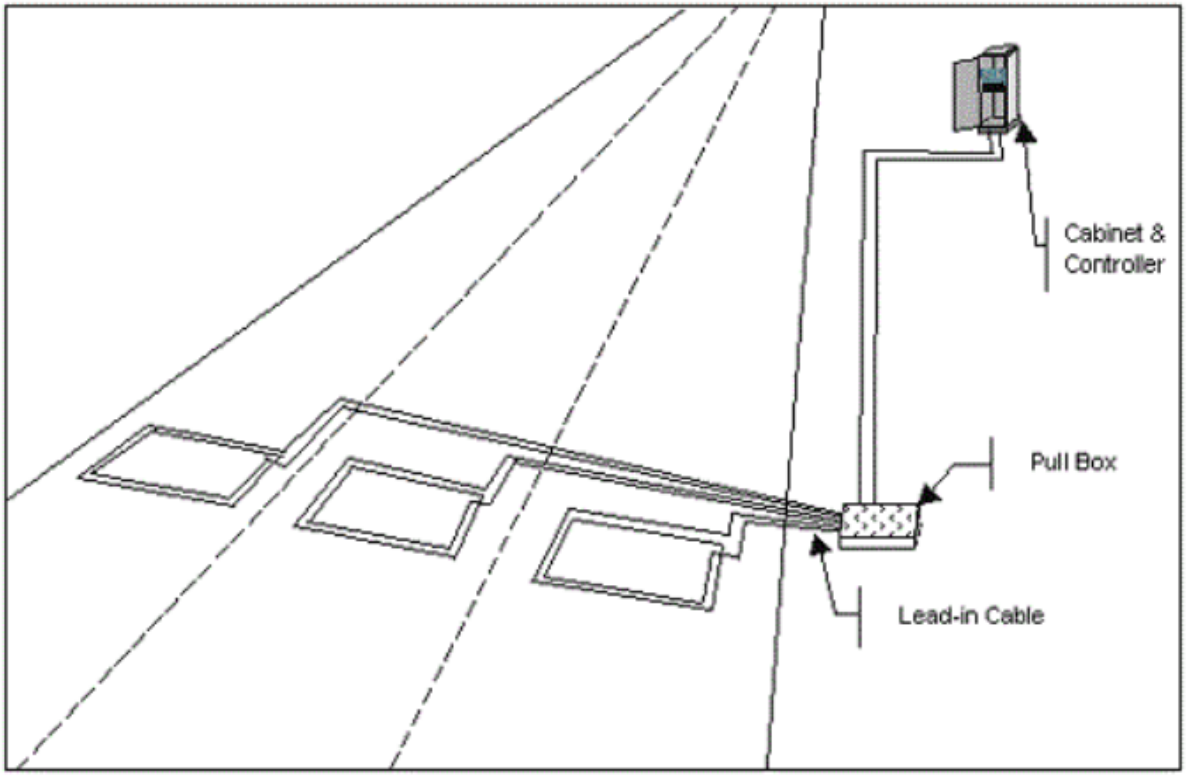
Glenn Vorhes

Transportation Operations and Safety Lab, UW Madison

VSPOC Detectors

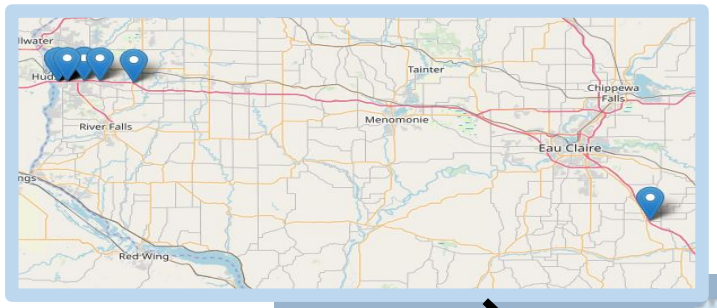
- **V** - Volume as vehicle/lane/hour
- **SP** - Speed as mile/hour
- **OC** - Occupancy as percentage

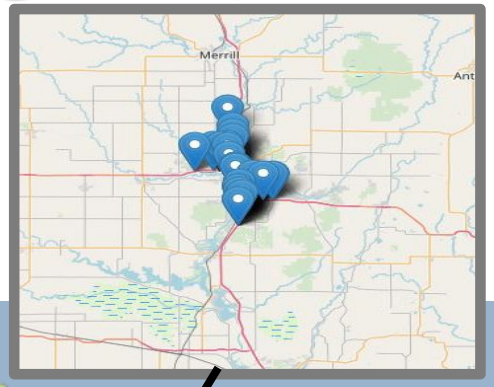
Inductive loop, microwave/radar

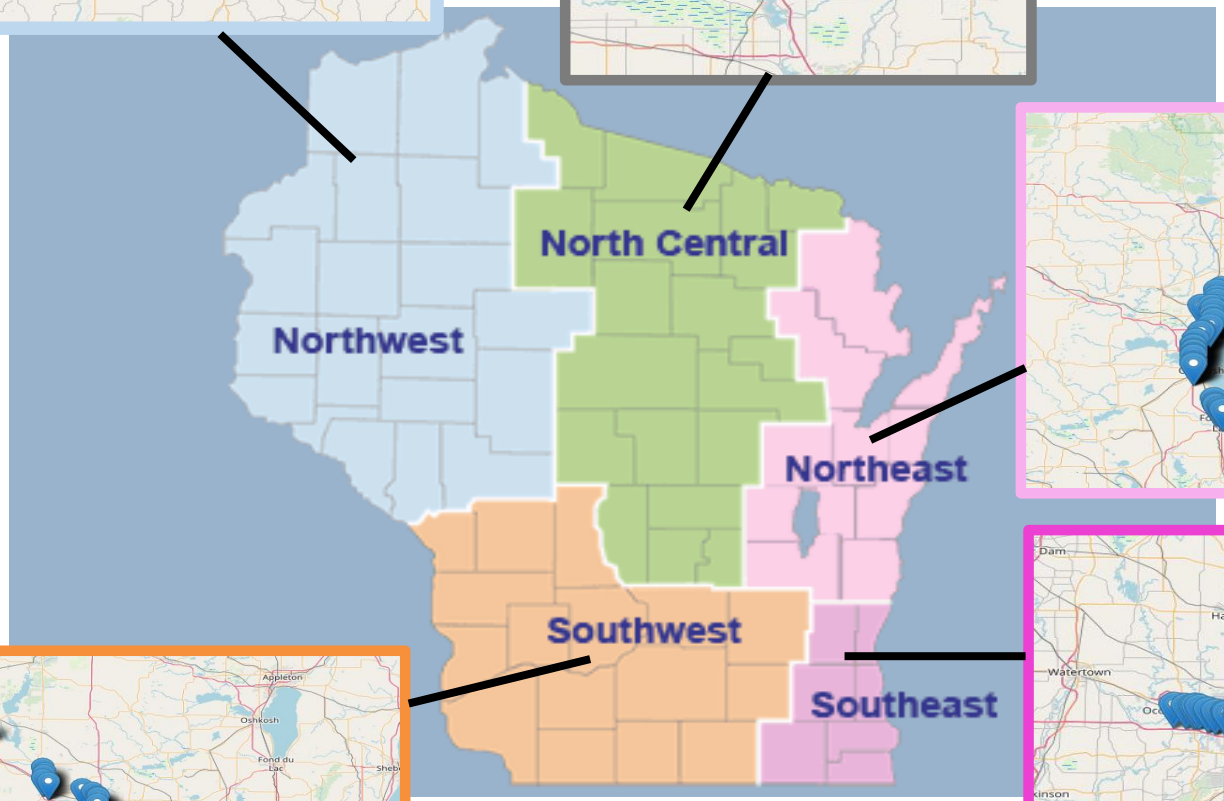


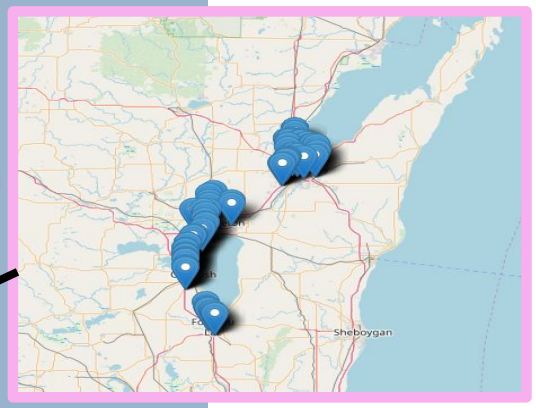
http://ops.fhwa.dot.gov/freewaymgmt/publications/frwy_mgmt_handbook/images/fig15-2.gif

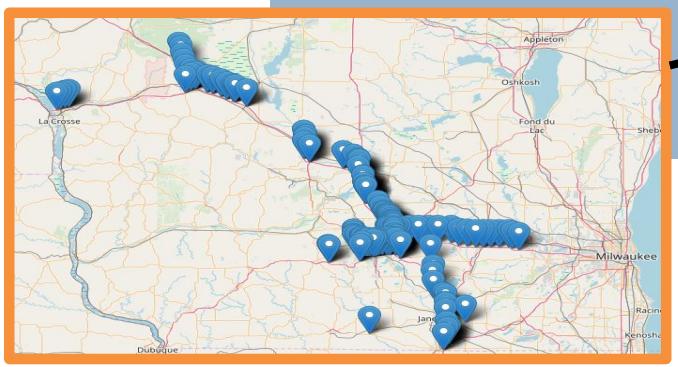
“VSPOC” Coverage – 4,000+ Detectors





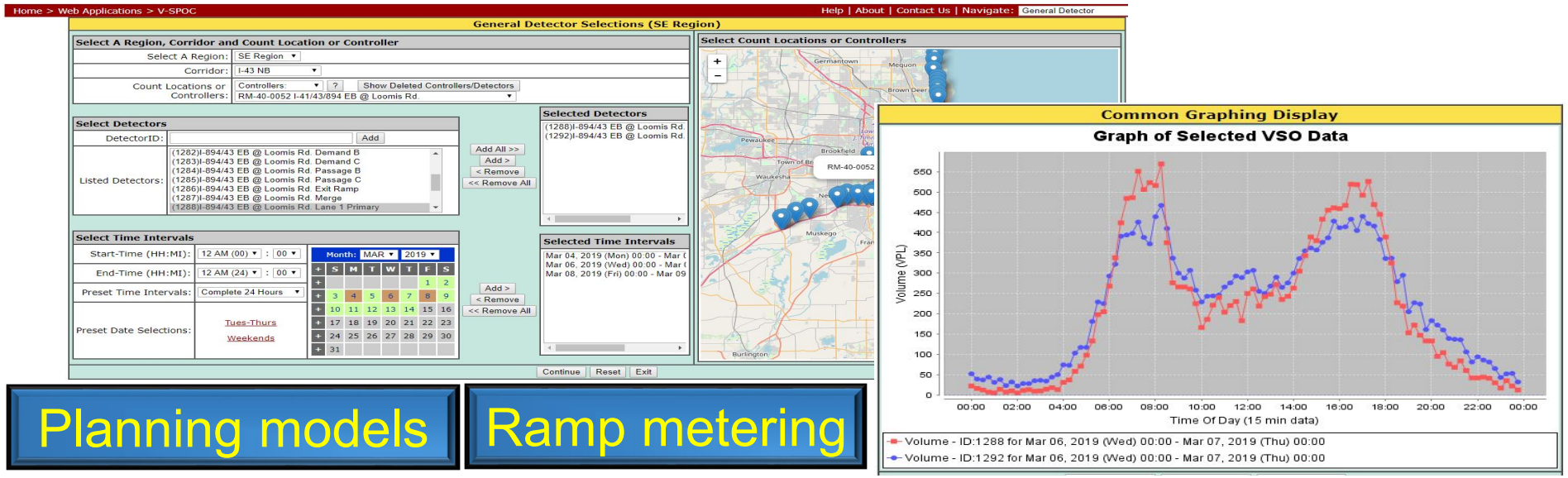






WisTransPortal "VSPOC" System

- Archived data
- Query & retrieve freeway traffic detector data from the WisDOT Traffic Management Center
  - Available as a daily CSV file that is then imported into Oracle database – VSPOC measurements for each detector
  - Signal processing for the detectors is outside the scope
- Initial System Deployed in 2006
- Contains archive of detector data from 1996



Planning models

Ramp metering

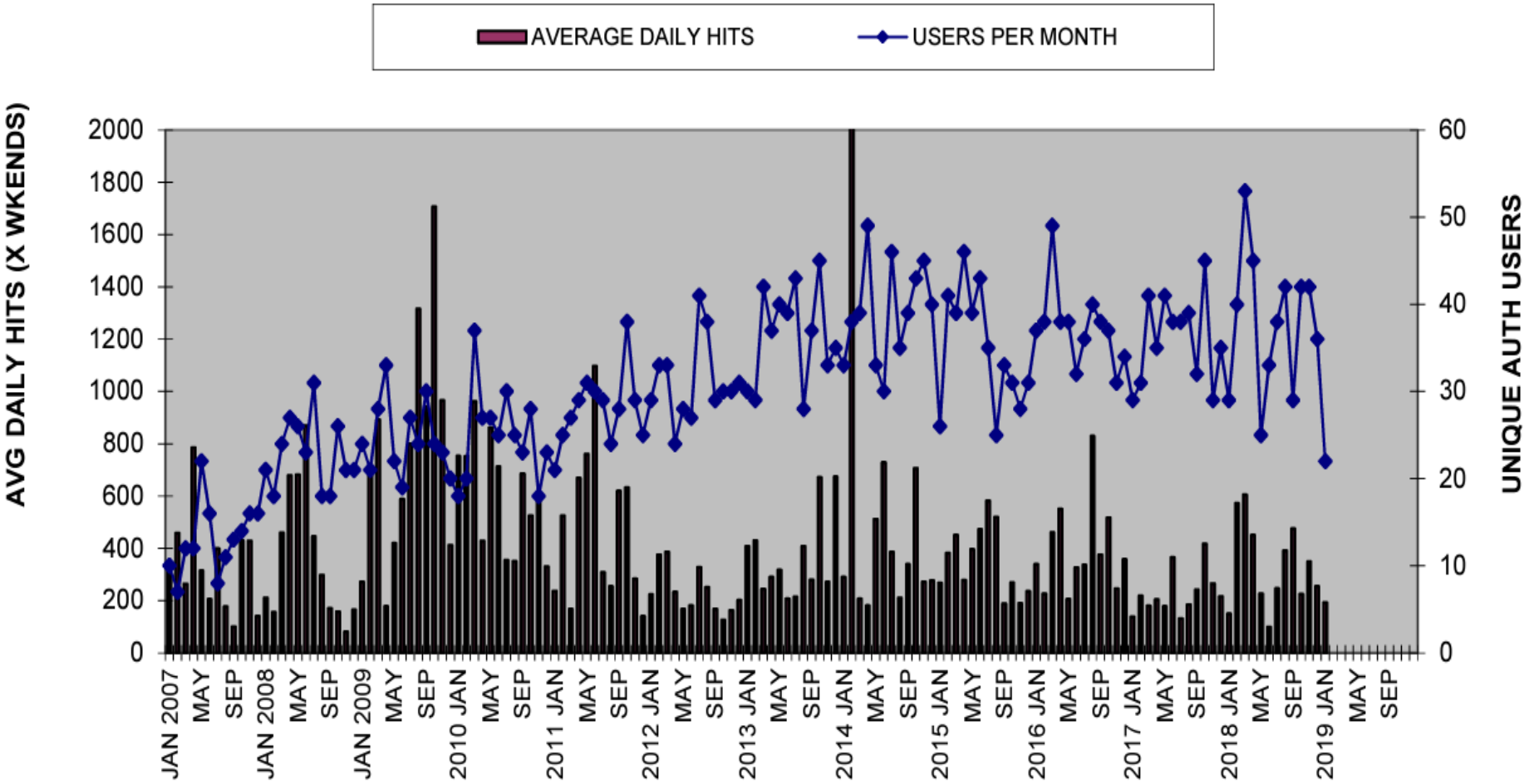
Bottleneck analysis

Capacity analysis

Work zone analysis

Incident management

WisTransPortal VSPOC Data Retrieval Facility



Motivation Data Quality Metrics

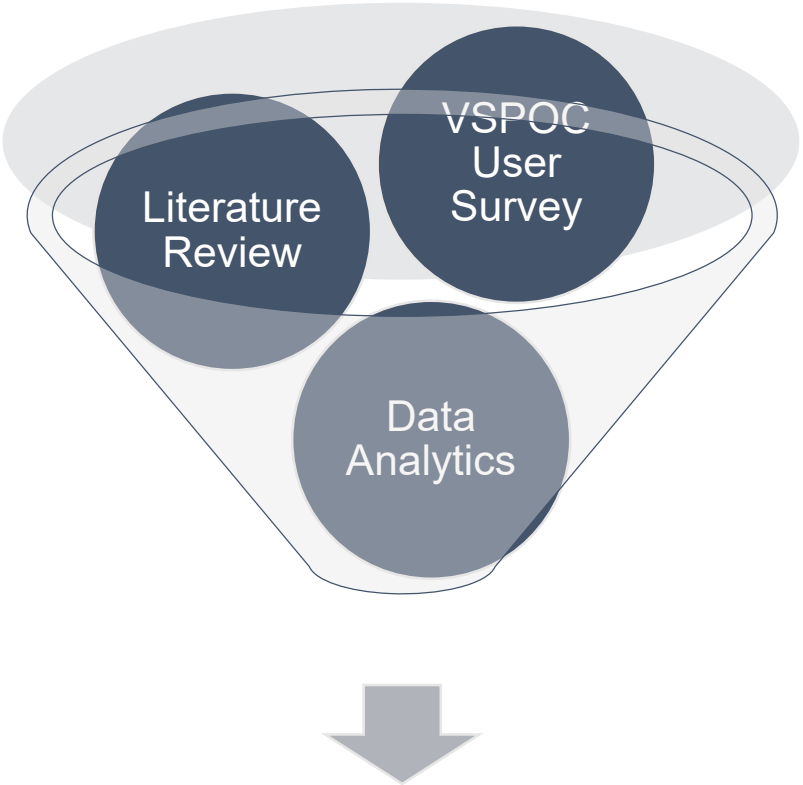
- Various data validity tests to flag obvious erroneous or potentially invalid traffic data.
  - Type 1: Univariate Tests – value within reasonable range
  - Type 2: Multivariate Tests – combination makes sense
  - Type 3: Temporal Consistency – match historical trends
- Archived Data Management Systems (ADMS) used to store historical travel data
  - Typical tests are less complex than those proposed in the literature: basic ones are widely implemented, but more advanced tests are limited.
  - Data might fail to maintain enough data quality with currently implemented validity tests.
  - This is especially if data quality issues are not easily discernable.

In Phase I

- Identify and recommend useful and practical data validity tests
- Develop a flagging procedure for potential implementation.

Archived Data Management System (ADMS) Survey

- ✓ California PeMS (Performance Measurement System),
- ✓ CATT (Center for Advanced Transportation Technology), Maryland,
- ✓ Central Florida Data Warehouse,
- ✓ Kentucky ADMS,
- ✓ Virginia ADMS,
- ✓ Phoenix RADS (Regional Archived Data Server),
- ✓ PORTAL (Portland Oregon Regional Transportation Archive Listing),
- ✓ WisTransportal V-SPOC.
- ✓ FHWA Mobility Monitoring Program



Literature Review

- Current application
- Preferred traffic variables
- Data specifications (e.g., time frame and resolution)
- Data quality checking methods

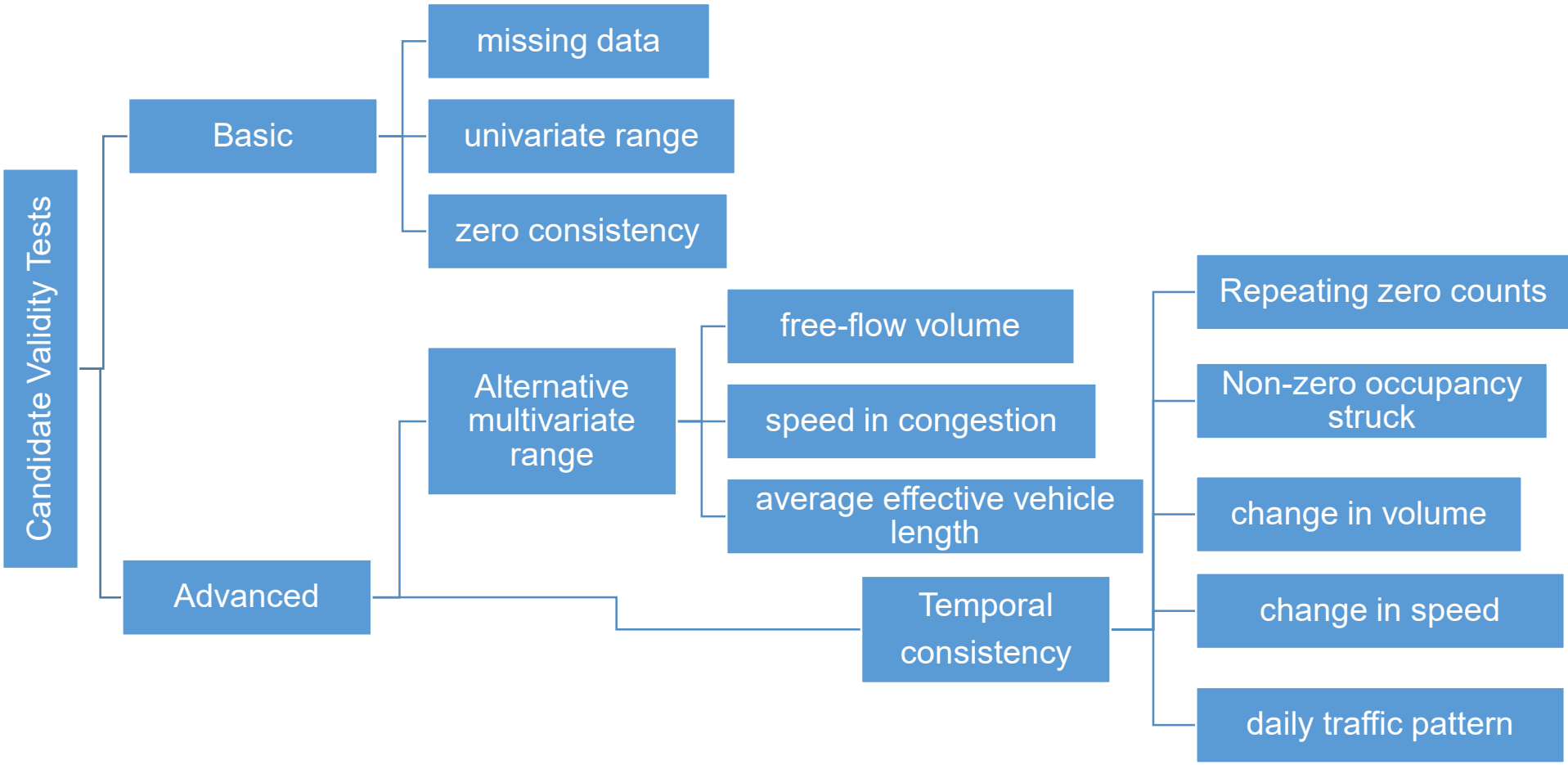
Initial Efforts

- Extensive tests
- Sensitivity analysis
- Case studies
- A combination of *rule-based* and *data-driven* approaches

What did we learn from literature review?

- Typical data QA/QC procedures include
  a) *univariate & multivariate range*: validate a single traffic variable or a combination of traffic variables against a predetermined threshold.
  b) *temporal consistency:* monitor changes within a time period or compare with historical trends.
  c) *spatial consistency:* compare with the traffic patterns from nearby detectors.
- Data validity tests in practice are closely alike, and most are less complex and probably, less effective.
- Potential to leverage new research to improve detector data quality.

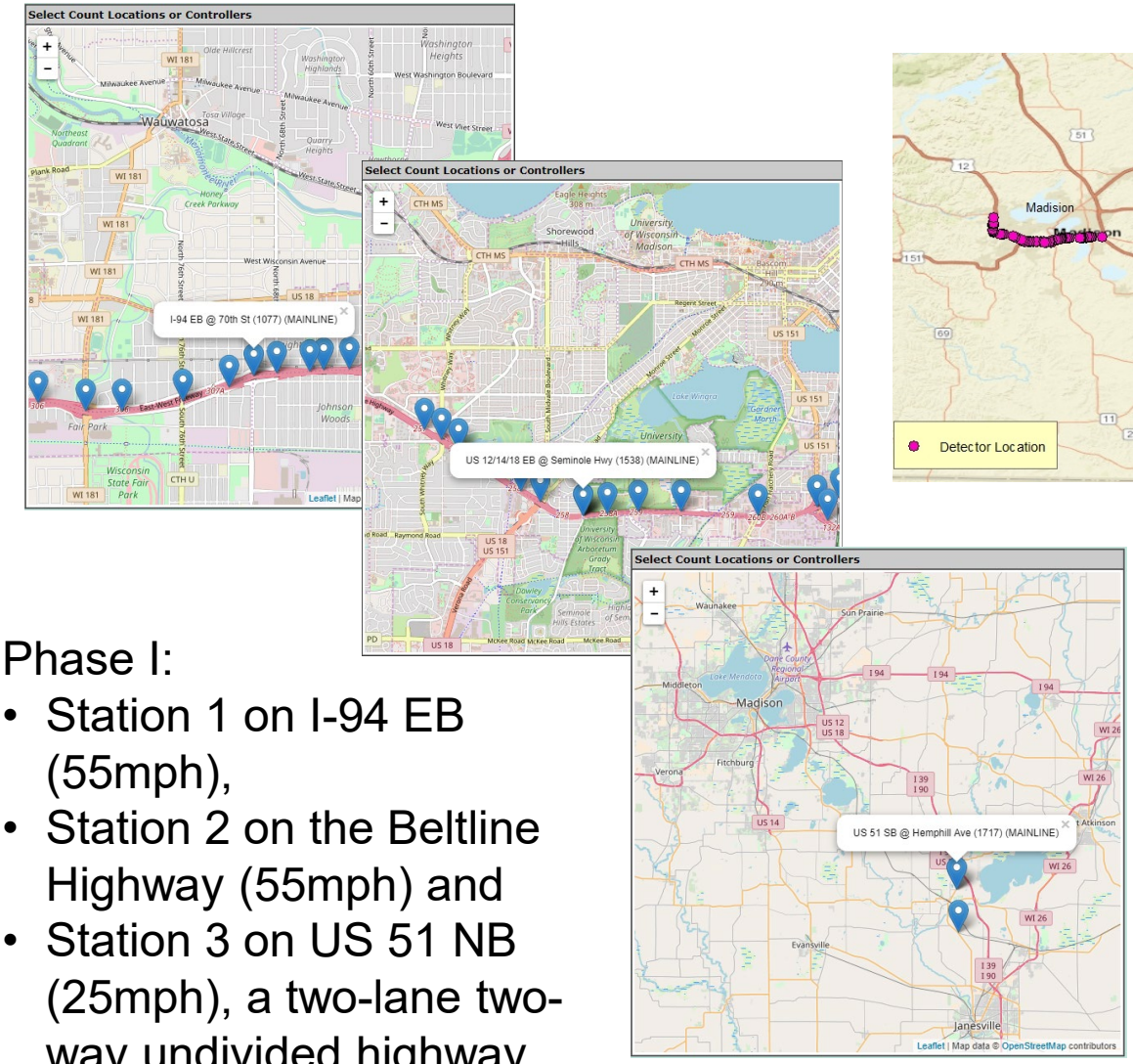
Candidate Data Validity Tests



Phase II Objectives

- On a large data set
  - 239 detectors September – December, 2017 and 2018
  - Calibrate test parameters
  - Improve the flagging procedure
- Develop a prototype
  - Use the flagging procedure
  - Integrate to the current VSPOC capabilities
- Collect user feedback regarding the prototype

Test Sites



Phase I:

- Station 1 on I-94 EB (55mph),
- Station 2 on the Beltline Highway (55mph) and
- Station 3 on US 51 NB (25mph), a two-lane two-way undivided highway

Phase II

- Milwaukee I-94, 152 detectors
- Madison Beltline highway, 87 detectors
- Sept-Dec, 2017 and 2018

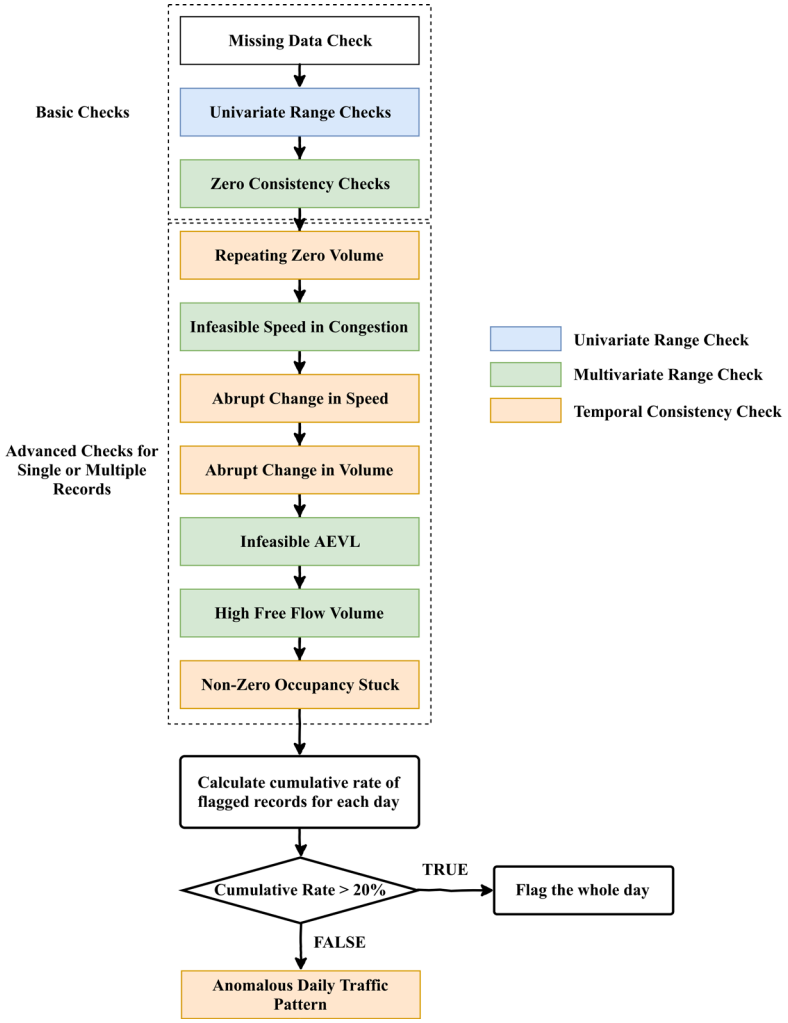
Improve the Flagging Procedure

Phase I

- Calibrate test parameters
  - Check the location-sensitivity for traffic conditions (speed limit, etc) and road types
  - Identify self-adjustment methods – data driven
- Improve the flagging procedure
  - Balancing computation cost and effectiveness

Phase II – Implement Tests

Tests applied to metrics, 5 minute aggregation window. Break at failure.



Tests

Test 1: Missing records:

If (Speed is "Null") | (Volume is "Null") | (Occupancy is "Null"), flag = 1 else 0.

"Null" = missing record/blank.

Test 2: Univariate range:

If (Volume is less than 0 or greater than 3,100) | (Speed is less than 0 or greater than 100) | Occupancy is less than 0 or greater than 100), flag = 1 else 0.

Test 3. Positive volume or occupancy with zero speed:

If Speed = 0 & (Occupancy > 0 | Volume > 0), flag = 1 else 0.

Test 4: Positive speed or occupancy with zero volume:

If Volume = 0 & (Occupancy > 0 | Speed > 0), flag = 1 else 0.

Test 5. Anomalous speed and volume with zero occupancy:

If Occupancy = 0 & (Volume > 0 | Speed > 0), flag = 1 else 0.

Test 6: Infeasible speed in congestion

$$\text{If} \quad \left(\left(Speed(t) > \frac{a_1}{\text{Occupancy}(t)} - b_1\right) \quad | \quad \left(Speed(t) \quad < \frac{a_2}{Occupancy(t)} - b_2\right) \quad \right) \&$$

$$\text{Occupancy}(t) > 10\,\% \text{ , then flag = 1 else 0.}$$

Where, a_1, b_1 are the estimable coefficient for the upper limit speed function (i.e. 97.5th percentile) and a_2, b_2 for lower limit speed function (i.e. 2.5th percentile).

Tests

Test 7: High free flow volume:

If (Volume > 1200) & (Occupancy < 5%), flag = 1 else 0.

Test 8: Average Effective Vehicle Length (AEVL):

If (AEVL < 8 feet) | (AEVL > 60 feet), flag = 1 else 0.

$$AEVL(ft/veh) = 52.80 \frac{(Speed \times Occupancy)}{Volume}$$

Test 9: Abrupt change in volume

If $|\Delta V_t| > \Delta V_{max}$, flag = 1 else 0.

ΔV_{max} is the threshold value which is determined by the distribution of $|\Delta V_t|$.

$$\Delta V_t = V_t - \frac{V_{t-1} + V_{t+1}}{2}$$

Where, ΔV_t represents the change in volume at time t and V_{t-1}, V_t, and V_{t+1} are the volumes at time $t-1$, t and $t+1$.

Test 10: Abrupt change in speed:

If $|\Delta S_t| > \Delta S_{max}$, flag = 1 else 0.

ΔS_{max} is the threshold value which is determined by the distribution of $|\Delta S_t|$.

$$\Delta S_t = S_t - \frac{S_{t-1} + S_{t+1}}{2}$$

Where, ΔS_t represents the change in speed at time t and S_{t-1}, S_t, and S_{t+1} are the speeds at time $t-1$, t and $t+1$.

Tests

Test 11: Non-zero (Occupancy) stuck:

If $(1\% < \text{Occupancy} < 100\%)$ & (

$$\sum_{i=-6, i\neq 0}^{6} I(\text{Occupancy}(t) = \text{Occupancy}(t+i)) > 3)$$, flag = 1 else 0.

Where, $I = \begin{cases} 1, & \text{Occupancy}(t) = \text{Occupancy}(t+i) \\ 0, & Otherwise \end{cases}$

Test 12: Repeating zero (volume):

If (Volume (t) = 0) & ($\sum_{i=-4, i\neq 0}^{4} I(\text{Volume}(t) = \text{Volume}(t+i)) > J$), flag = 1 else 0.

Where, $I = \begin{cases} 1, & Volume(t) = Volume(t+i) \\ 0, & Otherwise \end{cases}$

The value of J can be found in the following table. In the table, μ is the average historical 5-min volume and is calculated for two separate time periods: 6 am-10 pm and 6 am-10 pm. For example, if μ is 5 for 6 am-10 pm, J is 2. If μ is 1 for 10 pm-6 am, J is 7.

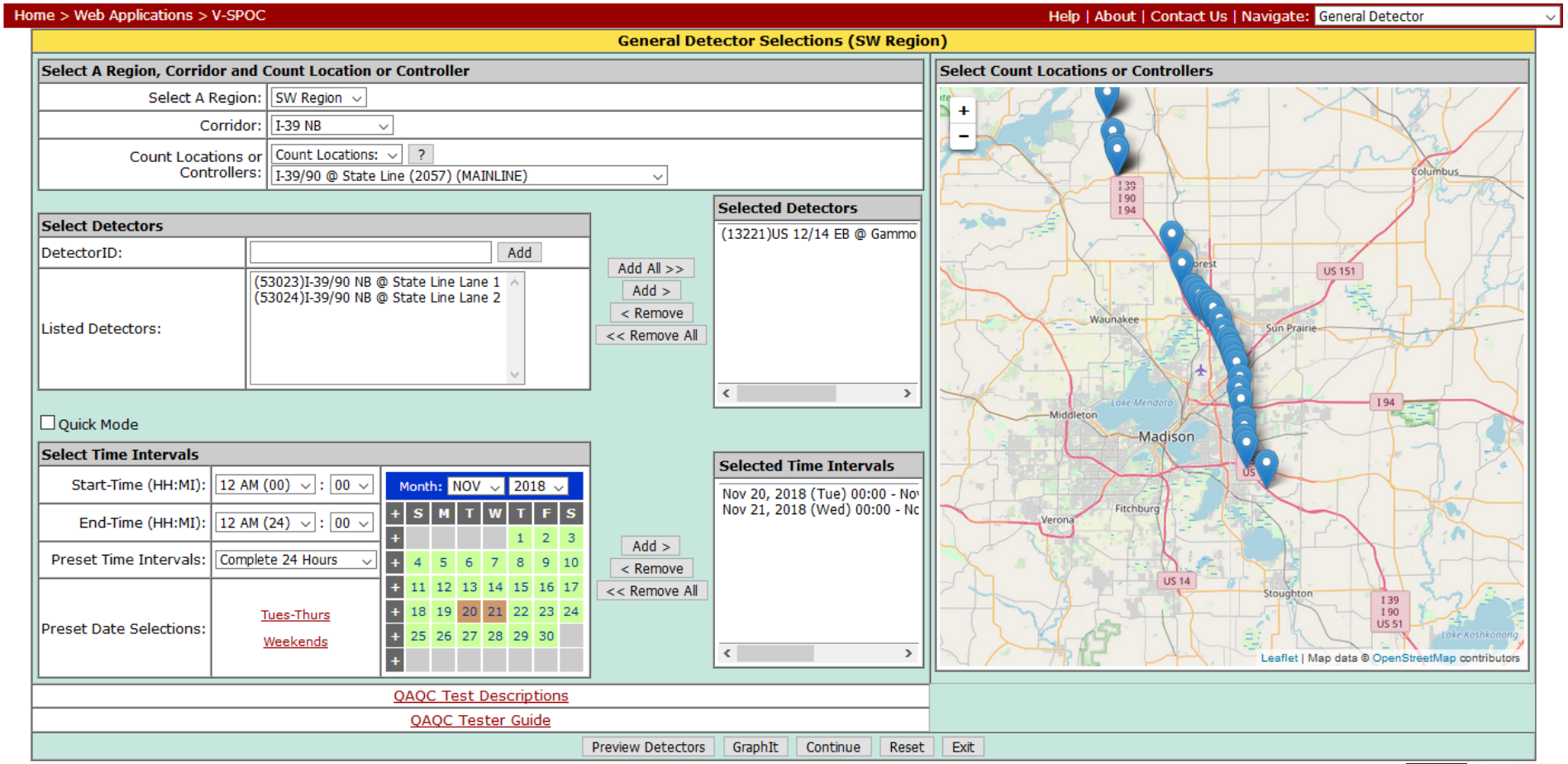
μ	$\mu=0$ and $V_t=0$	$0<\mu\leq 2$	$2<\mu\leq 3$	$3<\mu\leq 4$	$\mu \geq 4$
J	**Not Available**	7	5	3	2

Target Use Cases

- Find Bad Data
  - Traffic data managers
  - Detector managers
  - To identify data quality problems and potential detector issues
- Find Good Data
  - Data users – Nice assurance of data quality
  - Predetermined/preferred locations but flexible time ranges

Sample Interfaces (1)

- Data Retrieval Support



Last Modified 6/2/2017. Please send comments to transportal@topslab.wisc.edu.

Sample Interfaces (2)

- Data Quality Display
  - Show/Hide Various Tests
  - Download Quality Results (Aggregated/Disaggregate)
  - Quality Color-code

Preview Detectors

Preview Detectors Column Headers Legend		
☑	Test ID	Test Name
☑	TEST1	Missing Records
☑	TEST2	Univariate Range
☑	TEST3	Positive Volume or Occupancy with no Speed
☑	TEST4	Positive Speed or Occupancy with no Volume
☑	TEST5	Anomalous Speed and Volume with no Occupancy
☑	TEST6	Infeasible Speed by Occupancy Regime
☑	TEST7	High Free Flow Volume
☑	TEST8	Infeasible Average Effective Vehicle Length (AEVL)
☑	TEST9	Abrupt Change in Volume
☑	TEST10	Abrupt Change in Speed
☑	TEST11	Non-Zero Stuck
☑	TEST12	Repeating Zero

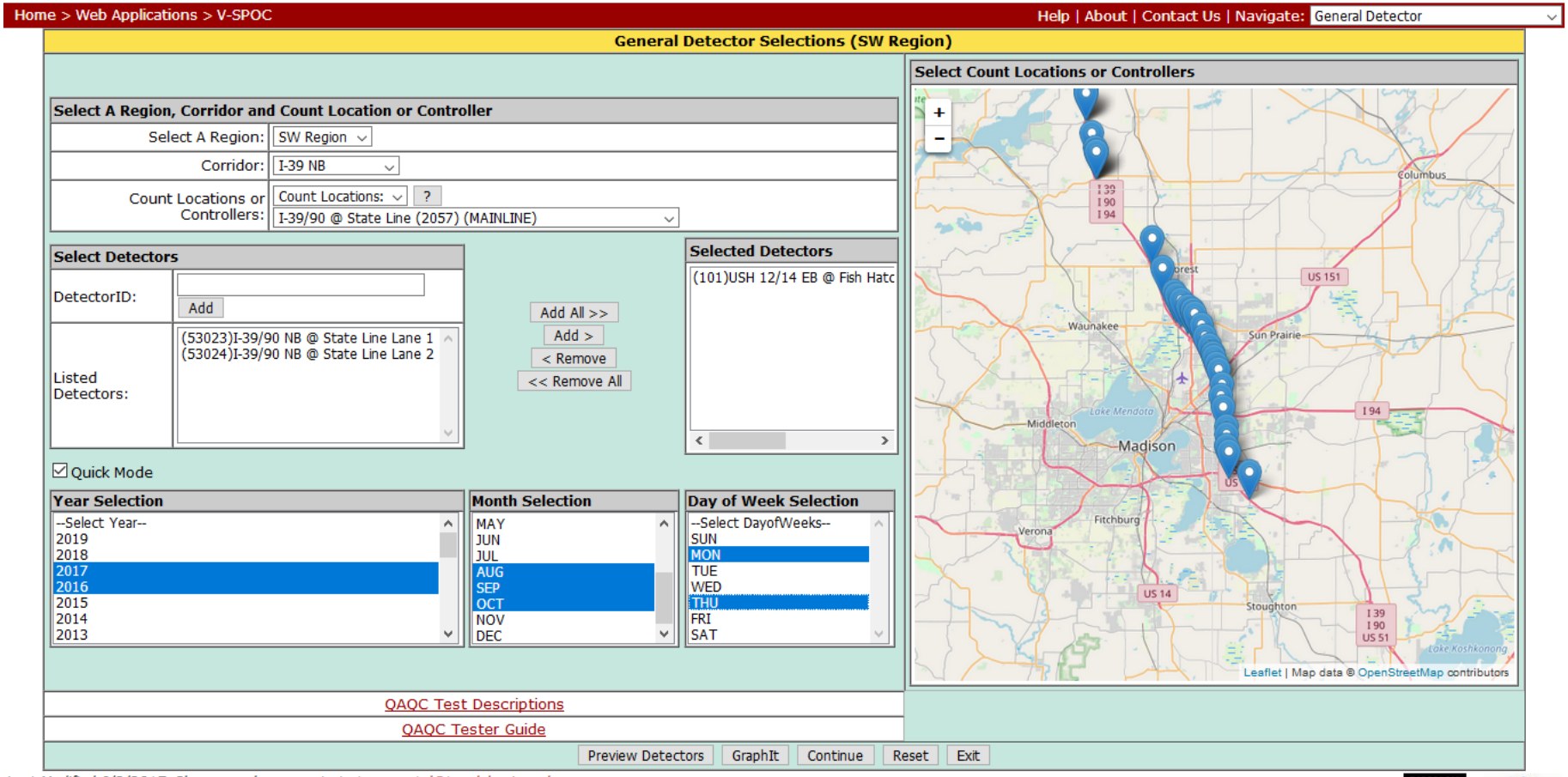
Result

☑	DID	Detector Description	Date	Time Interval	TEST1	TEST2	TEST3	TEST4	TEST5	TEST6	TEST7	TEST8	TEST9	TEST10	TEST11	TEST12
☑	101	USH 12/14 EB @ Fish Hatchery Rd Lane 3 Primary(D)	NOV 20, 2018 (TUE)	00:00 - 24:00	100.0 %	100.0 %	100.0 %	84.72 %	99.31 %	100.0 %	99.65 %	100.0 %	1.0 %	75.0 %	100.0 %	100.0 %
☑	101	USH 12/14 EB @ Fish Hatchery Rd Lane 3 Primary(D)	NOV 21, 2018 (WED)	00:00 - 24:00	100.0 %	100.0 %	100.0 %	89.24 %	96.88 %	99.31 %	99.31 %	100.0 %	1.0 %	74.31 %	100.0 %	100.0 %

Save Selected | Download Raw Data | Exit

Sample Interfaces (3)

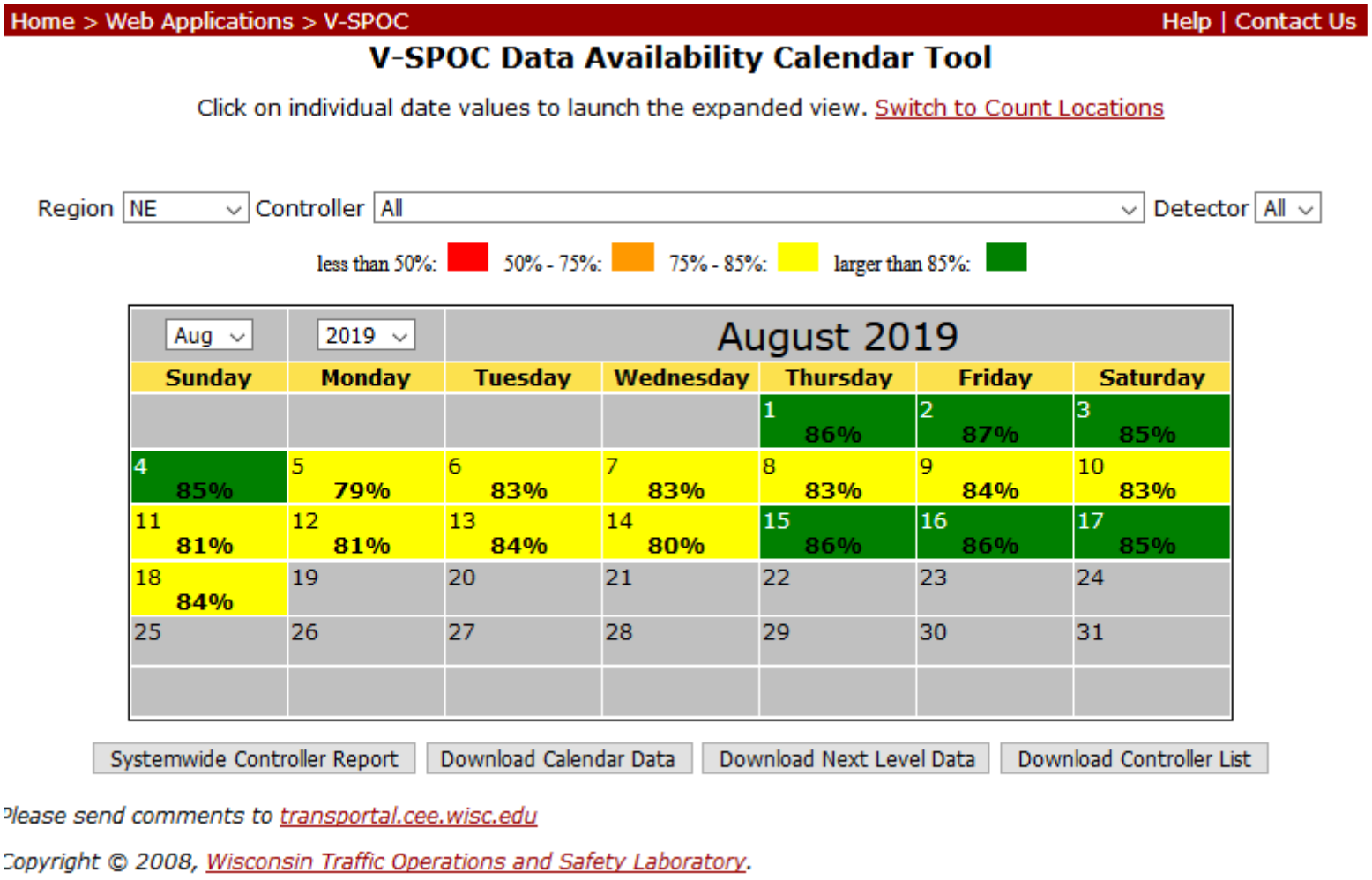
- Quick Quality Scan



Last Modified 6/2/2017. Please send comments to transportal@topslab.wisc.edu.

Sample Interfaces (4)

- The Data Quality Calendar Improvement



Project Status

- Finishing up the prototype development
- Starting the user review for the prototype very soon
- Future Work in Phase III
  - Improve the prototype based on user feedback
  - Implement the flagging procedure along with the data archiving process
  - Rollout the prototype to VSPOC production

Thank you!

Contact Information:

Elizabeth Schneider, PE,
State Freeway Operations Engineer
BTO, Wisconsin Department of Transportation
Elizabeth1.Schneider@dot.wi.gov

Yang Cheng, Ph.D.
Traffic Operations and Safety (TOPS) Laboratory
University of Wisconsin-Madison
cheng8@wisc.edu

Xiao (Shaw) Qin, Ph.D., PE
Associate Professor
Director, Safe and Smart (S^2) Traffic Laboratory
University of Wisconsin-Milwaukee
qinx@uwm.edu