PORTAL Aggregation Analysis and Documentation

Kristin A. Tufte

Portland State University

Introduction

This document describes the aggregation procedures for PORTAL . Analysis of different aggregation methodologies is included and is used to select the aggregation methodologies. Aggregations for 5-minute, 15-minute and 1-hour granularities are described as well as procedures for aggregating across lanes. Finally, a discussion of ADUS (Archived Data User Service) aggregation standards is included.

Table of Contents

1. 20-Second Data Readings

Volume:

Description: The number of vehicles that passed over the detector during the 20-second collection period. Volume is reported as an integer.

Units: none (this is a count)

Speed:

Description: The average speed of cars passing over the detector during the 20-second collection period. It is unknown how the controllers measure and/or calculate this speed reading. Volume is reported as an integer.

Units: miles per hour

Occupancy:

Description: The average occupancy of the detector during the 20-second collection period, where occupancy means the percent of time the detector is 'occupied', that is the percent of time a vehicle is physically positioned over the detector.

Units: Percent expressed as a value between 0 and 100.

Status: A data quality status value reported by the ATMS system. Details on the status values and their descriptions can be found in the SWARM documentation. Table 1 lists a summary of those values. Also see the October PORTAL report (link here) for details on other types of error values reported by the ATMS System.

Table 1 ATMS Status Flags

Status Flag	Description	Comments
0	Inhibited	May indicate communication error.
1	Disabled	
2	Ok	
3	Suspect	To the best of our knowledge, "Suspect" data is used in SWARM calculations.
4	Soft fail	
5	Hard fail	

Table 2 Database Table and View Names for Data and Aggregations

Table/View Name	Description
loopdata_YYYY_MM_DD	Daily loopdata table. Contains 20-second data for the day YYYY-MM-DD.
loopdata_5min_raw	5-minute granularity aggregations by detector
loopdata_15min_raw	15-min granularity aggregations by detector
loopdata_1hr_raw	1-hour granularity aggregations by detector
loopdata_5min_raw_bystation	5-minute granularity aggregations by station; this is a view
loopdata_15min_raw_bystation	15-minute granularity aggregations by station; this is a view
loopdata_1hour_raw_bystation	1-hour granularity aggregations by station; this is a view

2. Raw Aggregations over Time Introduction

The proposal for PORTAL Raw Aggregations is shown below. These aggregations are intended to be raw, unfiltered aggregations that display the data collected as is. Filtered and imputed aggregations will be considered in the future. This proposal first addresses the Raw Aggregations for the 5-minute, 15-minute and 1-hour aggregation tables for PORTAL. Aggregating over lanes is addressed in a separate section.

3. 5-Minute Raw Aggregations

We list the calculations first, followed by notes and assumptions.

Volume:

Calculation: Sum of 20-second volumes

Description: This calculation provides the total volume of cars travelling through the segment during the 5-minute period.

Units: none (this is a count)

Speed:

Calculation: Volume-weighted average of 20-second speeds

Description: This calculation provides the average speed of cars travelling over the detector during the 5-minute period. A volume-weighted average is used as that is the current ADUS standard as indicated by email from Rich Margiotta of Cambridge Systematics and Shawn Turner of TTI.

Units: miles per hour

Occupancy:

Calculation: Average of 20-second occupancies (not weighted)

Description: This calculation provides the average occupancy of the detector of cars travelling over the detector during the 5-minute period.

Units: Percent expressed as a value between 0 and 100.

CountReadings*:*

Calculation: Count of 20-second readings received from ODOT.

Description: Total valid readings observed for this detector during the 5- minute period. Readings indicating communication errors and disabled detectors are not included in aggregations and so are not included in this count.

Units: none – this is a count.

Vehicle Miles Travelled (VMT):

Calculation: 5-minute volume * length of influence area

Description: The estimated total miles travelled by cars in this segment during the 5-minute period.

Units: miles

Vehicle Hours Travelled (VHT):

Calculation: 5-min volume * (length of influence area/5-minute speed)

Description: The estimated total hours vehicles spent travelling in this segment during the 5-minute period.

Units: hours

Travel Time:

Calculation: (length of influence area/5-min speed)*60

Description: The estimated average time it took a vehicle to cross through this segment during the 5-minute period. Current PORTAL calculation provides travel time in hours.

Units: minutes

Delay:

Calculation: (*5*-min travel time – (length of influence area/free-flow speed))* 60

Description: The average minutes of delay experienced by a vehicle in this segment during the 5-minute period. This calculation provides average delay; total delay can be calculated by multiplying by volume. Current PORTAL calculation provides total delay in hours.

Units: minutes

Question: Is using non-weighted averages for speed and occupancy correct?

Free Flow Speed: Free flow speed is assumed to be 60 mph.

Speed = 0: VHT, Delay and Traveltime all require dividing by speed. When speed=0; VHT, Delay and Traveltime are set to NULL.

Filtering: Records indicating no communications from the detector or a disabled detector are removed from the aggregations. In technical terms, a record is not included in the aggregation if speed, volume and occupancy are all NULL. In addition, 5-minute periods with zero volume will have speed set to NULL as a volume-weighted average cannot be computed when volume is zero.

Missing Readings: Currently in PORTAL, the readings not received from ODOT are simply missing rows. The PORTAL team should modify the system so that the missing rows are replaced with rows with volume, speed, occupancy and status set to NULL. Note that rows indicating no communications have volume, speed, and occupancy set to NULL and status 0. So, with this scheme communications failures reported by ODOT (communication failure between ODOT and the detector infrastructure) can be distinguished from the situation where PSU does not receive data from ODOT. Such failure of PSU to receive data is sometimes due to communications issues between ODOT and PSU, but is typically due to ATMS outages.

Action: The PORTAL team should modify the system so that the missing rows are replaced with rows with volume, speed, occupancy and status set to NULL.

Weighted Speeds and Occupancies vs. Non-Weighted Speeds and Occupancies

For informational purposes, we include as Table 3 and Table 4 a comparison of metrics calculated with weighted and non-weighted speeds. In Table 4, the Mean Absolute Difference (MAD) and Mean Difference (MD) are reported. The mean difference is calculated as (weighted-unweighted). Note that the volume and VMT metrics rely only on volume, and do not use speed, and as such are not affected by the use of weighted or non-weighted speeds.

Table 3 Means and Standard Deviations for Metrics with Weighted and Non-Weighted Speeds

	Mean	Standard Deviation	Count Nulls
Volume	62.63	45.61	0
Speed – Non-Weighted	55.75	15.57	44,492
Speed – Volume-Weighted	56.03	14.59	67,781
VMT	67.67	84.39	172,380
VHT- Non-Weighted	1.45	8.23	274,609
VHT – Vol-Weighted Speed	1.38	2.92	276,366
Delay – Non-Weighted	0.38	9.12	274,609
Delay-Vol-Weighted Speed	0.21	2.23	276,366
Traveltime – Non-Weighted	1.43	9.66	274,609
Traveltime – Vol-Weighted Speed	1.25	2.40	276,366

Table 4 Differences between Weighted and Non-Weighted Aggregates

	MAD	MD (W-UW)
Speed (mph)	1.32	-0.15
Occupancy (pct 0-100)	5.85	-5.85
VHT (hours)	0.051 (3.1 min)	0.015
Delay (minutes)	0.115 (6.9 sec)	-0.054
Traveltime (minutes)	0.115 (6.9 sec)	-0.054

4. 15-Minute Raw Aggregations

Descriptions and units are the same as for 5-minute aggregations; calculations are provided here. Calculations are provided followed by a discussion and analysis supporting why these particular calculations were supported.

Volume: *Calculation:* Sum of 5-min volumes

Speed: *Calculation:* Volume-weighted average of 5-minute speeds

Occupancy: *Calculation:* Average of 5-minute occupancies (not weighted)

CountReadings*: Calculation:* Sum of 5-minute CountReadings.

Vehicle Miles Travelled (VMT): *Calculation:* 15-min volume * length of influence area

Vehicle Hours Travelled (VHT): *Calculation:* 15-min volume * (length of influence area/ 15-min speed)

Travel Time: *Calculation:* (length of influence area/15-minute speed)*60

Delay: *Calculation:* ((length of influence area/15min speed)-(length of influence area/free flow speed))*60

Delay is average minutes of delay; total minutes of delay can be obtained by multiplying by volume.

Free Flow Speed: Free flow speed is assumed to be 60 mph.

Speed = 0: VHT, Delay and Traveltime all require dividing by speed. When speed=0; VHT, Delay and Traveltime are set to NULL.

Aggregation Method Selection

VMT, VHT, Delay and Traveltime can be calculated in different ways. The purpose of this section is to present the different methods of calculation and to present analysis and discussion supporting the selection of calculation method.

We need to calculate 15-minute aggregates for each metric (VMT, VHT, Delay and Traveltime), there are two possible aggregation methods proposed, labeled Method 1 and Method 2. The 5-minute aggregates are calculated directly from the 20-second data. The most accurate method of calculating the 15-minute aggregates is to calculate them from the 20-second data; however, such calculation is expensive. Calculating from 15-minute speed, volume and occupancy or calculating from 5-minute aggregates is much cheaper and is preferable if the accuracy is reasonable.

For each metric, below we propose two methods of calculating that metric. The methods are labeled Method 1 and Method 2. In all cases, Method 1 uses 15-minute aggregates (calculated as specified above) and formulas similar to the 5-minute formulas. Method 2 calculates the metric using 5-minute aggregates. The methods are described below followed by analysis of the accuracy of each method.

VMT Methods:

Method 1: 15-min volume * length of influence area

Method 2: sum of 5-min VMTs

Let L = Length of influence area.

Method 1 and Method 2 are theoretically the same and are equal up to rounding errors in implementation.

VHT Methods:

Method 1: 15-min volume * (length of influence area/ 15-min speed)

Method 2: sum of 5-min VHTs

Method 1 and Method 2 are not identical.

Travel Time Methods:

Method 1: (length of influence area/15-minute speed)*60

Method 2: Volume-weighted average of 5-min traveltimes.

Method 1 and Method 2 are not identical.

Delay (in minutes):

Method 1: ((length of influence area/15min speed)-(length of influence area/free flow speed))*60

Method 2: Volume-weighted average of 5-min delays.

Method 1 and Method 2 are not identical.

Aggregate Method Accuracy Analysis

To determine the accuracy of the two aggregation methods, we compare the values calculated using each method to values calculated from 20-second data. The analysis uses data from the month of September 2009 for main line detectors.

Table 5 compares the means and standard deviations of the 20-second-based aggregations to the means and standard deviations produced from the aggregations using Method 1 and Method 2. Recall that Method 1 and Method 2 are theoretically the same for VMT, but not for VHT, Delay and Traveltime. Overall, the means and standard deviations show that Method 1 produces results that are more similar to the 20-second-based aggregations than Method 2 and that Method 1, in general, is quite accurate with respect to the 20-second-based aggregations. Table 6 shows error metrics for Method 1 and Method 2. The metrics are defined below. We observe that in general for Method 1, error is very low. For Method 1, the Mean Absolute Percent Error (MAPE) ranges from 0.03% for VMT – Method 1 to 1.16% for Delay – Method 1. We believe this range of error is acceptable.

Conclusion: Method 1 is the method of choice for computing 15-minute aggregates.

Table 5 Comparison of Means and Standard Deviations for Aggregation Methods

	Mean	Standard Deviation
Speed – From 20 sec (mph)	55.7	15.0
Speed – From 5 min (mph)	56.5	14.1
Volume – From 20 sec (count)	187.4	135.1
Volume – From 5 min (count)	187.4	135.1
Occupancy – From 20 sec (pct)	6.6	11.7
Occupancy – From 5 min (pct)	6.6	11.7
VMT – From 20 sec (miles)	202.5	251.2
VMT – Method 1 (miles)	202.4	251.1
VMT – Method 2 (miles)	202.4	251.1
VHT – From 20 sec (hours)	4.3	29.9
VHT – Method 1 (hours)	4.2	27.3
VHT – Method 2 (hours)	4.3	21.8
Delay – From 20 sec (minutes)	0.42	18.1
Delay – Method 1 (minutes)	0.33	11.7
Delay – Method 2 (minutes)	0.30	6.8
Traveltime – From 20 sec (minutes)	1.5	18.1
Traveltime – Method 1 (minutes)	1.4	11.7
Traveltime – Method 2 (minutes)	1.4	6.9

Table 6 Error Metrics for Aggregation Methods

	MAE	MAPE	ME
Speed – 20sec vs. 5min	0.93	2.69%	0.35
Volume – 20 sec vs. 5 min	0.0	0.0%	0.00
Occupancy – 20 sec vs. 5 min	0.0	0.36%	0.0
VMT1 vs. VMT 20 sec (miles)	0.0	0.0%	0.0
VMT2 vs. VMT 20 sec (miles)	0.0	0.0%	0.0
VHT1 vs. VHT 20 sec (hours)	0.16	2.0%	-0.06
VHT2 vs. VHT 20 sec (hours)	0.40	2.7%	0.00
Delay1 vs. Delay 20 sec	0.15	6.5%	-0.10

Delay2 vs. Delay 20 sec	0.27	7.5%	-0.12
Traveltime1 vs. Traveltime 20 sec	0.15	2.0%	-0.10
Traveltime2 vs. Traveltime 20 sec	0.27	2.5%	-0.12

Mean Percent Error (MAE): The average of absolute errors. Taking the absolute value prevents positives and negative values from canceling each other out and so MAE gives an indication of the average magnitude of the error.

Mean Absolute Percent Error (MAPE): The average of the absolute value of the percent errors. For example, for VHT Method 1, the percent error is calculated as: ((vht1 – vht)/vht)*100 where vht is the vht calculated from the 20 second data. The value in Table 6 is the average of absolute percent errors for all main line readings in September 2008. Readings where the (20-second-based) vht is 0 are excluded from the calculation to prevent division by zero.

Mean Percent Error (ME): The average of the errors. In this calculation, the positive and negative errors cancel out, thus this number provides a measure of the bias of the estimates.

5. 1-Hour Raw Aggregations

Computing 1-hour aggregations from 15-min aggregations is similar to computing 15-min aggregations from 5-min aggregations. Calculations are provided below.

Volume: *Calculation:* Sum of 15-min volumes

Speed: *Calculation:* Average of 15-minute speeds (not weighted)

Occupancy: *Calculation:* Average of 15-minute occupancies (not weighted)

CountReadings*: Calculation:* Sum of 15-minute CountReadings.

Vehicle Miles Travelled (VMT): *Calculation:* 1-hr volume * length of influence area

Vehicle Hours Travelled (VHT): *Calculation:* 1-hr volume * (length of influence area/ 1-hr speed)

Travel Time: *Calculation:* (length of influence area/1-hr speed)*60

Delay: *Calculation:* ((length of influence area/1-hr speed)-(length of influence area/free flow speed))*60

Delay is average minutes of delay; total minutes of delay can be obtained by multiplying by volume.

Free Flow Speed: Free flow speed is assumed to be 60 mph.

Speed = 0: VHT, Delay and Traveltime all require dividing by speed. When speed=0; VHT, Delay and Traveltime are set to NULL.

6. Aggregation across Lanes

The previous sections described how PORTAL aggregates data across time. In this section, we discuss how PORTAL aggregates data across space – that is across lanes.

Volume: *Calculation:* Sum of by-lane volumes

Speed: *Calculation:* Weighted average of by-lane speeds

Occupancy: *Calculation:* Average of by-lane occupancies (unweighted)

CountReadings*: Calculation:* Sum of by-lane CountReadings.

Vehicle Miles Travelled (VMT): *Calculation:* station volume * length of influence area

Vehicle Hours Travelled (VHT): *Calculation:* station volume * (length of influence area/station speed)

Travel Time: *Calculation:* (length of influence area/by-lane speed)*60

Delay: *Calculation:* ((length of influence area/by-lane speed)-(length of influence area/free flow speed))*60

7. ADUS Aggregation Standards

The following information is based on an email exchange with Shawn Turner of Texas Transportation Institute and Rich Margiotta of Cambridge Systematics, Inc.

The ADUS Traffic Monitoring Standard, which is being revised, addresses aggregation. The aggregation procedures specify that volume-weighted speed averages should be used, but occupancy averages should not be volume weighted (this procedure is in the standard and is what TTI does). When speeds from adjacent locations are combined, VMT should be used to weight instead of volume to account for differing "virtual" link lengths. The standard does not specify what levels the data should be stored at, but Rich Margiotta indicates that when aggregation occurs one must specify:

- any quality control results (user-specified) that occurred during the aggregation
- any imputation that occurred during the aggregation
- The number of valid (QC-passed) records that are in the set of records used to compute an aggregated metric
- The total number of records actually used in the aggregation

The above should be stored as data elements.

The distinction between the 3rd and 4th bullet points is the use of imputed or edited/altered values. Imputed or edited values are not counted in the QC-passed records used in the aggregation, but are counted in the total number of records used for the aggregation. The distinction allows the user to know how many “good” measurements went into the calculation and how much was based on imputation, etc. Rich notes that the above documentation is a minimum; systems/users may do more if they wish.

Shawn Turner comments that he sees only small differences between calculating higher-level aggregates (15-min, 1-hour) from 20-second data vs. calculating from 5-min aggregates, etc. He believes calculating from the finest granularity (mist disaggregate) data is best, but in practice, often calculates from 5-minute averages as that saves processing time. We note that the analysis we present in this report shows similar results.