Title

The Freeway Performance Measurement System (PeMS), PeMS 9.0: Final Report

Author
Varaiya, Pravin

Publication Date
2009-03-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

The Freeway Performance Measurement System (PeMS), PeMS 9.0: Final Report

Pravin Varaiya

California PATH Research Report
UCB-ITS-PRR-2009-25

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

Final Report for RTA 51A0356

March 2009

ISSN 1055-1425

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

The Freeway Performance Measurement System (PeMS), PeMS 9.0: Final Report

Pravin Varaiya

Department of Electrical Engineering and Computer Science
University of California, Berkeley CA 94720
Tel: (510) 642-5270, Fax: (510) 643-2356
varaiya@eecs.berkeley.edu
November 30, 2008

Abstract

Under development and operation since 1999, PeMS now is the *de facto* repository for all Caltrans fixed-location detector data. The PeMS 9.0 effort completed seven tasks: (1) Control Algorithm Support; (2) Integration of Census Data in PeMS; (3) Investigation of Fidelity of 3rd Party Detectors; (4) OD estimation algorithm based on the fusion of the FasTrak data and the ITS-sensors; (5) Incorporating speed measurements reported from the field; (6) Incorporation of New Photolog Format; and (7) Development of Diagnostic Routines for Rural Facilities.

Keywords: California Freeway Performance Measurement System, PeMS, Loop detector diagnostics, speed measurement, census data, OD estimation, FasTrak, data fusion

Executive Summary

Under development and operation since 1999, PeMS now is the *de facto* repository for all Caltrans fixed-location detector data. The system is used to investigate everything from the characteristics of individual loop detectors to highly aggregated performance trends across Caltrans districts. PeMS usage has grown continuously and it now supports 3000-4000 reports per day (plots, tables, etc). It distributes data to 55 registered Value Added Resellers, 20 of which actively retrieve real-time data.

PeMS is located on the U.C. Berkeley campus. During the past few months, a copy of PeMS has been installed at Caltrans. Caltrans is currently determining the steps to transition that instance of PeMS to active service. Once it goes live, Caltrans PeMS will be a 24×7 production facility. UCB PeMS will then become a platform for developing new algorithms and features.

The PeMS 9.0 effort completed seven tasks.

1. Control Algorithm Support. The first part of this task is to investigate what sort of information is needed for each control application and with what accuracy. The second part is to determine whether detector system measurements can be processed to produce this information.
2. Investigation of Census Data. This task involved investigating and understanding the Caltrans Census data and then integrating it into PeMS.
3. Investigation of Fidelity of 3rd Party Detectors. This task involved associating labels of ownership on each controller in the system. We then modified the various diagnostic reporting tools so that users can easily investigate the fidelity of the detectors from different owners.
4. Investigate FasTrak Data Fusion Methodology. This task was to implement an OD estimation algorithm based on the fusion of the FasTrak data and the ITS-sensors. The results are displayed in tabular form.
5. Investigate Using Field Speed Measurements. With this task we added the ability to use speed measurements that are reported from the field. This is done on a district-by-district basis or on a sensor-by-sensor basis.
6. Investigation of User Responses to New Photolog Format. Caltrans has upgraded their photolog system. The original system was a series of movies, each spanning 5 miles of freeway. The new system is a set of images taken every 1/100th of a mile in the front and side directions. We've incorporated these into PeMS with a new user interface based on Google Maps and inspired by Google Street Views.
7. Research Appropriate Diagnostic Routines for Rural Facilities. We've enhanced the diagnostic routines in PeMS so that we can have a different set of diagnostic thresholds for each detector in the district. This was done in response to a problem that arose in D10, which is mainly rural.

PeMS 9.0 is the latest of ten task orders devoted to research, development, and maintenance of the PeMS system (there was one mid-year task order, PeMS 6.5). PeMS collects, processes, stores, and makes available online data from nine Caltrans districts (D3-8, 10-12). The data are obtained from 25,750 loops[1], grouped into 10,073 vehicle detector stations (VDS). These loops cover 3,841 out of 30,572 directional-miles of interstate and state highways in California.

We now describe the accomplishments under the seven tasks that constitute the PeMS 9.0 project.

[1] Some of these 'loops' are microwave radar detectors that act like loops.

1.0 Task 1: Control Algorithm Support

PeMS provides an accounting framework for tracking freeway performance and a suite of diagnostic tools that pinpoint the weaknesses in freeway operations. These weaknesses are also opportunities for short term productivity gains. The Strategic Growth Plan depends on these opportunities to achieve its congestion mitigation goals through a set of operational improvements, ramp metering, incident management, traveler information and demand management, all of which are real-time control applications. The objective of this task was to explore the extent to which detector system measurements can adequately support the needs of these real-time control applications.

This task was formulated within a dynamic freeway model in which we could pose these applications as feedback control problems. We then determined the data needed to calibrate the model and the variables that had to be measured in real time in order to be used in implementation of the feedback control. We focused on ramp control.

We used the Cell Transmission Model (CTM) for freeway dynamics.[2] This first-order model can be calibrated from PeMS data. The most serious difficulty with calibration is that PeMS sometimes lacks on-ramp data and very frequently lacks off-ramp data. We strongly recommend that ramps along freeways selected in the Corridor Improvement Program be equipped with detectors.

There are different ramp metering schemes, most notably ALINEA and SWARM. ALINEA seeks to keep the density at the location where congestion is most likely to occur below its critical value. When an on-ramp merges into the mainline, this location is typically *downstream* of the merge point. Thus if one wants to use the ALINEA algorithm for ramp metering, a detector should be placed downstream of the ramp, in contrast to current Caltrans practice which places a detector between an on-ramp and the upstream off-ramp, i.e., upstream of the on-ramp. SWARM, on the other hand, relies on measuring or estimating the difference between the flow and the capacity at the next downstream bottleneck, which suggests a different arrangement for locating detectors.

Ramp metering must be accompanied by queue control, that is, care must be taken to ensure that the primary task of preventing mainline congestion does not lead to excessive queues on the ramps that may cause spill back on urban streets that feed into the ramp. In Caltrans practice, a so-called queue detector is placed close to the ramp entrance, and when its occupancy becomes sufficiently high, the queue override control takes over and increases the metering rate until the queue detector occupancy drops, causing mainline congestion. It is well-known that the interaction between the mainline metering and queue override control causes oscillatory waves of congestion on the mainline and urban street. A much better behavior can be achieved by a detection system that can measure the queue length on the ramp.

Lastly, ramp control should be designed to take into account short-term forecasts in demand, including the uncertainty in those demands. PeMS does not provide demand forecasts, but it seems plausible that PeMS data archives could be used to estimate such forecasting models, which could then be made part of a more effective feedback control design.

2.0 Task 2: Investigation of Census Data

The Traffic Census Group collects data from traffic sensors, which we'll call Census data or stations, to accurately determine traffic flows on the state highway system. The traffic flow at a point is typically captured in a single number that represents the average daily travel over the year, called AADT. The Traffic Census Group is comprised of personnel in each district as well as a coordinating team at headquarters. The traffic volumes that are collected by the Traffic Census Group are leveraged to make a number of decisions in Caltrans.

---

[2] G. Gomes, R. R. Horowitz, A.A. Kurzhanskiy, J. Kwon, and P.Varaiya. Behavior of the cell transmission model and effectiveness of ramp metering. *Transportation Research*, *C*, 16(4):485-513, August 2008.

In this task we developed a feature to store non-continuous Census count data in PeMS. This turned out to be the largest feature in PeMS 9. The decoding of the storage of the census data in TSN turned out to be quite a bit more difficult than what we expected.

We started out by extracting all of the Census hourly counts from TSN and importing them into PeMS. We did this with programs written in PERL. Since TSN is inside of the Caltrans corporate network, we can't talk to it directly from the UCB PeMS machines. Hence we leveraged the "Green Box". The Green Box is a machine that we've stuck inside of Caltrans to assist with the deployment of PeMS inside of Caltrans. The Green Box is on the Caltrans corporate network and has network connectivity to both TSN and the UCB instance of PeMS. The program for extracting the Census data from TSN runs on the Green Box. It exports the count data from TSN to flat files and then transfers the data to UCB PeMS. We then load the data in by hand. The program to grab and transfer the data initially ran to grab everything. In the future we plan to run this program once per month to grab the new Census data and load it in to PeMS.

We were able to grab the raw hourly count data from every mainline and ramp Census station in TSN. This data goes back to 1992 for some stations. In addition, we were able to grab the AADT values as computed and published by Caltrans.

We made the decision that we wouldn't impute for the Census data. Hence all of the data that we show is the actual data in TSN. They only place where this issue comes up is when we aggregate data to show values per month or per quarter. In those cases we overload the concept in PeMS of the "percentage of samples observed" to mean the percentage of hourly samples that existed. If this is every below 100% then the data in the aggregation shouldn't be used.

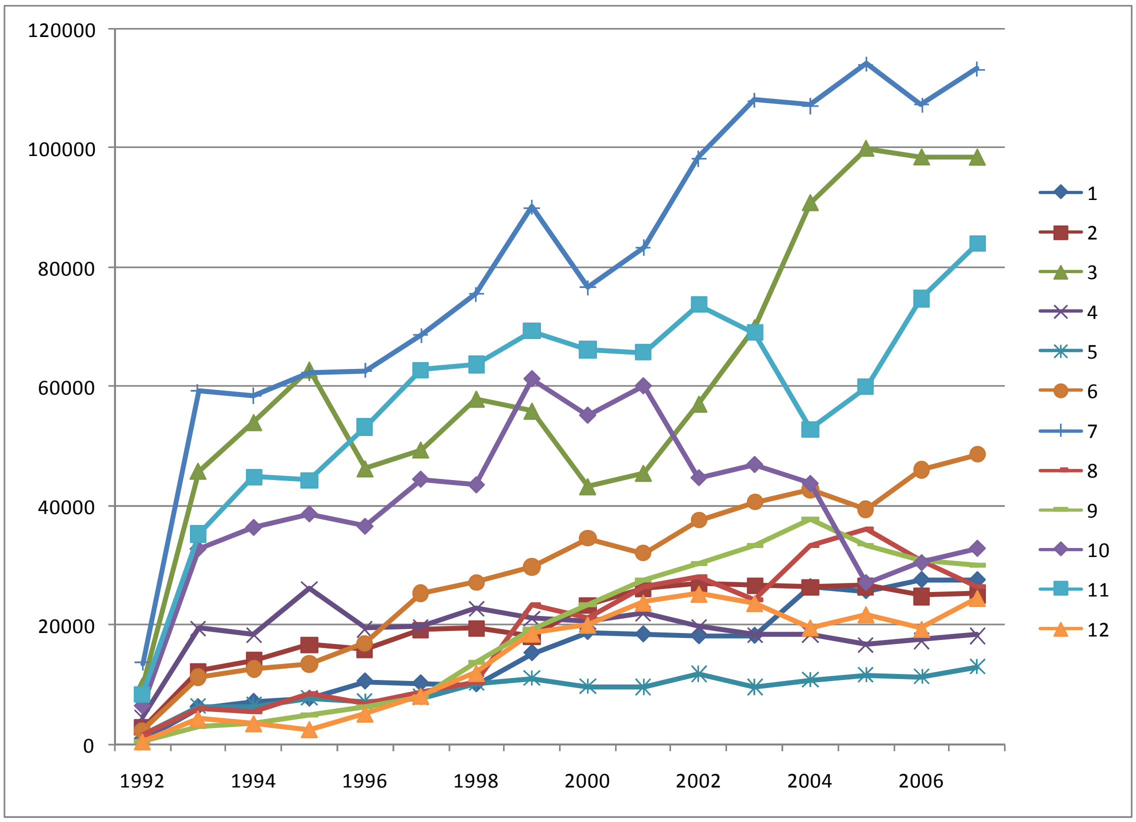


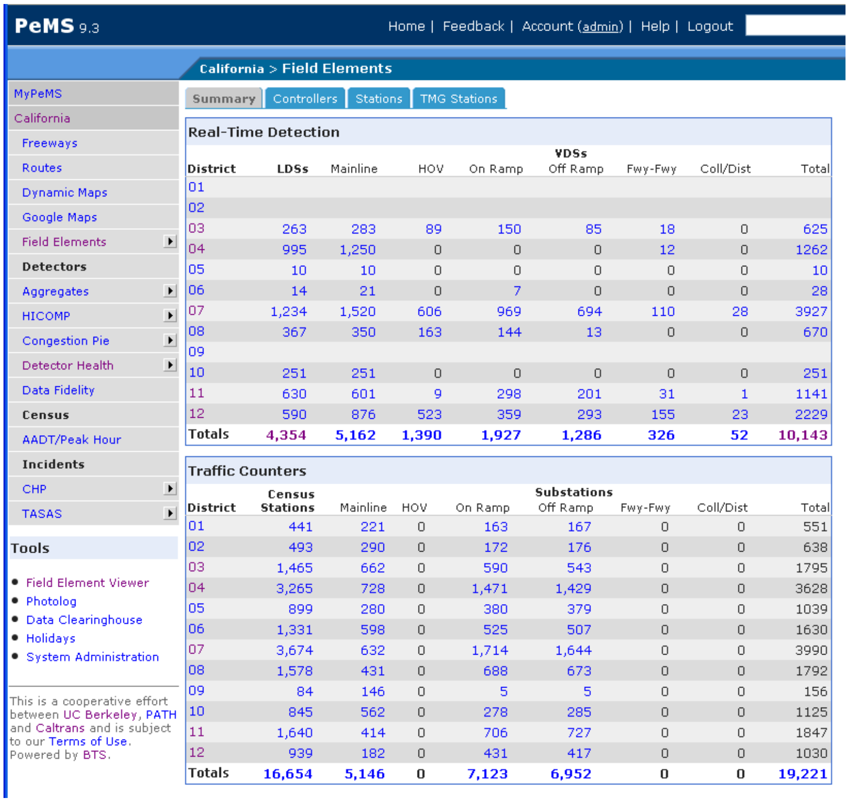
Figure 1. Number of detector days of mainline Census data collected per year per Caltrans District in PeMS.

Figure 1 is a plot of the number of detector days of Census data in each Caltrans District since 1992 that we have in PeMS. It represents all of the mainline and ramp data that is in TSN. We can see that while most districts are collecting more and more Census data, some aren't.

After we got the data loaded into PeMS, we started to build web displays to find and plot the data. We built two types of tools: navigational tools to help users find the data and reporting tools to display values.

2.1 Navigational Tools

We started with a simple set of tools to catalog the data in each district. The first tool is simply a listing of the count of the number of stations in each district, as seen below.



Real-Time Detection

District	LDSs	VDSs Mainline	HOV	On Ramp	Off Ramp	Fwy-Fwy	Coll/Dist	Total
01								
02								
03	263	283	89	150	85	18	0	625
04	995	1,250	0	0	0	12	0	1262
05	10	10	0	0	0	0	0	10
06	14	21	0	7	0	0	0	28
07	1,234	1,520	606	969	694	110	28	3927
08	367	350	163	144	13	0	0	670
09								
10	251	251	0	0	0	0	0	251
11	630	601	9	298	201	31	1	1141
12	590	876	523	359	293	155	23	2229
Totals	4,354	5,162	1,390	1,927	1,286	326	52	10,143

Traffic Counters

District	Census Stations	Substations Mainline	HOV	On Ramp	Off Ramp	Fwy-Fwy	Coll/Dist	Total
01	441	221	0	163	167	0	0	551
02	493	290	0	172	176	0	0	638
03	1,465	662	0	590	543	0	0	1795
04	3,265	728	0	1,471	1,429	0	0	3628
05	899	280	0	380	379	0	0	1039
06	1,331	598	0	525	507	0	0	1630
07	3,674	632	0	1,714	1,644	0	0	3990
08	1,578	431	0	688	673	0	0	1792
09	84	146	0	5	5	0	0	156
10	845	562	0	278	285	0	0	1125
11	1,640	414	0	706	727	0	0	1847
12	939	182	0	431	417	0	0	1030
Totals	16,654	5,146	0	7,123	6,952	0	0	19,221

Figure 2. Count of the number of field elements of each type in PeMS. The Census stations are in the bottom table.

In the figure above we are showing a screen shot of a new page that we built that shows the number of field elements in each district. The top table is the number of ITS stations and the bottom table is the number of Census stations. We break both tables down by type. In the top table we count the number of controllers (LDSs) and the number of VDS stations by function (mainline, HOV, ramps, etc). The bottom table breaks down the Census stations by mainline, on-ramp and off-ramp. If you look at just the total numbers for the state, you can see that while there are approximately the same number of mainline Census stations as mainline ITS stations, there are many more Census ramp stations. But if you look on a district-by-district basis, you can see that a number of the districts, like D4 and D7, have a higher density of mainline ITS detectors than Census.

We built pages to catalog the Census stations by district. This follows the standard PeMS paradigm of providing lists of ITS stations that users can view as well as export to spreadsheets. Each link in the above tables jumps to one of these page that lists off the stations of that particular type in the particular district.

In addition to providing the standard lists of detectors in each district, we also built a tool to find the detectors on a map, as shown below.

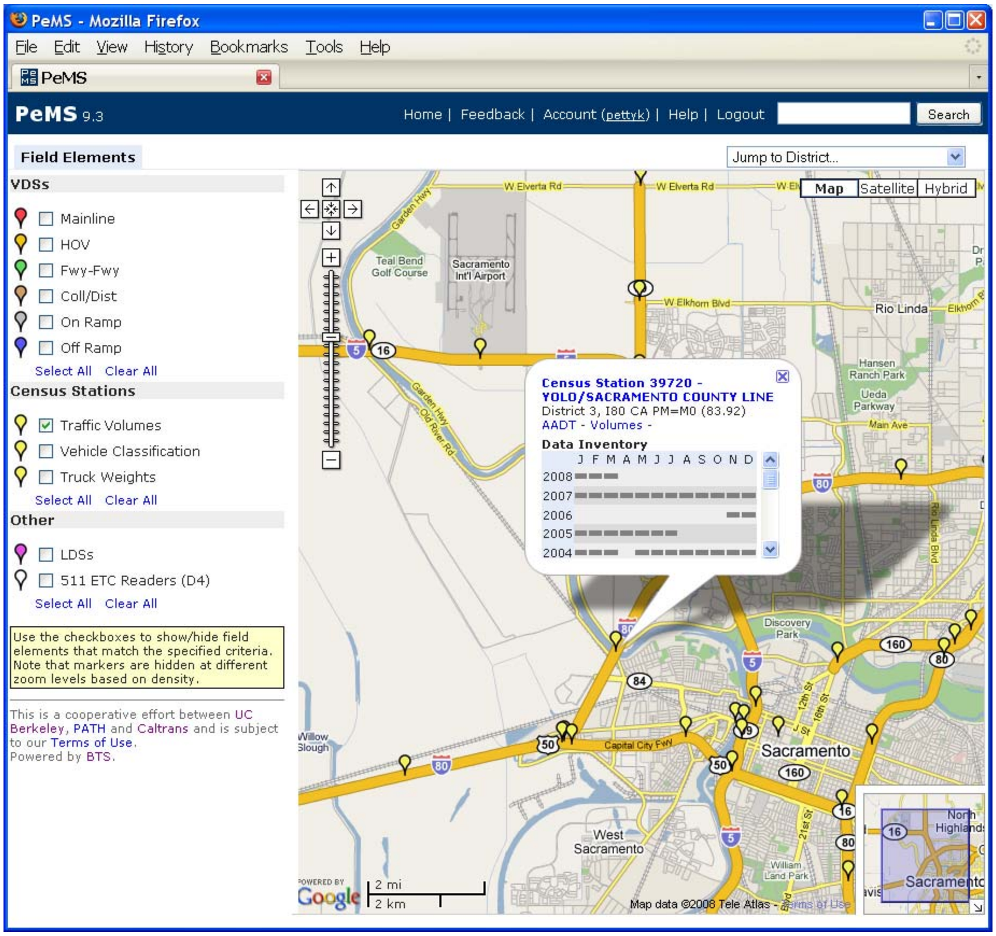


Figure 3. Field Element Viewer showing Census data and data inventory.

In Figure 3 we are showing the field element viewer that we built for PeMS 9. The field element viewer is simply a Google Map mashup where we allow users to turn on and off the location of certain field elements. We extended this past just Census stations to all field elements in PeMS. We currently let users show the following items on the map independently:

- ITS Stations:
  - Mainline VDS stations.
  - HOV VDS stations.
  - Freeway-to-freeway VDS stations.
  - Collector and distributor VDS stations.
  - On-ramp VDS stations.
  - Off-ramp VDS stations.
- Census Stations:
  - Traffic volume stations.
  - Vehicle classification stations.
  - Truck weight stations.

- Other:
    - Controllers (LDSs)
    - 511 ETC tag readers (only in D4).

We automatically limit which ones can be shown on the map as a function of the zoom level. If there are too many labels on the map then it become unresponsive. In the screen shot above you can see that at this zoom level the checkbox for the mainline ITS stations is grayed out.

For the Census stations one of the largest problems is actually finding the data. While the mainline Census stations on urban freeways are typically collected every day, the ramps and rural facilities are only collected periodically, sometimes as infrequently as once every three years. The current navigational paradigm in PeMS doesn't quite work with this type of data because it assumes that there is data all the time. Hence with the ITS stations we can just dump the user into the timeseries plots and they can always find data no matter what time period they are looking at. With the Census stations that is just not true.

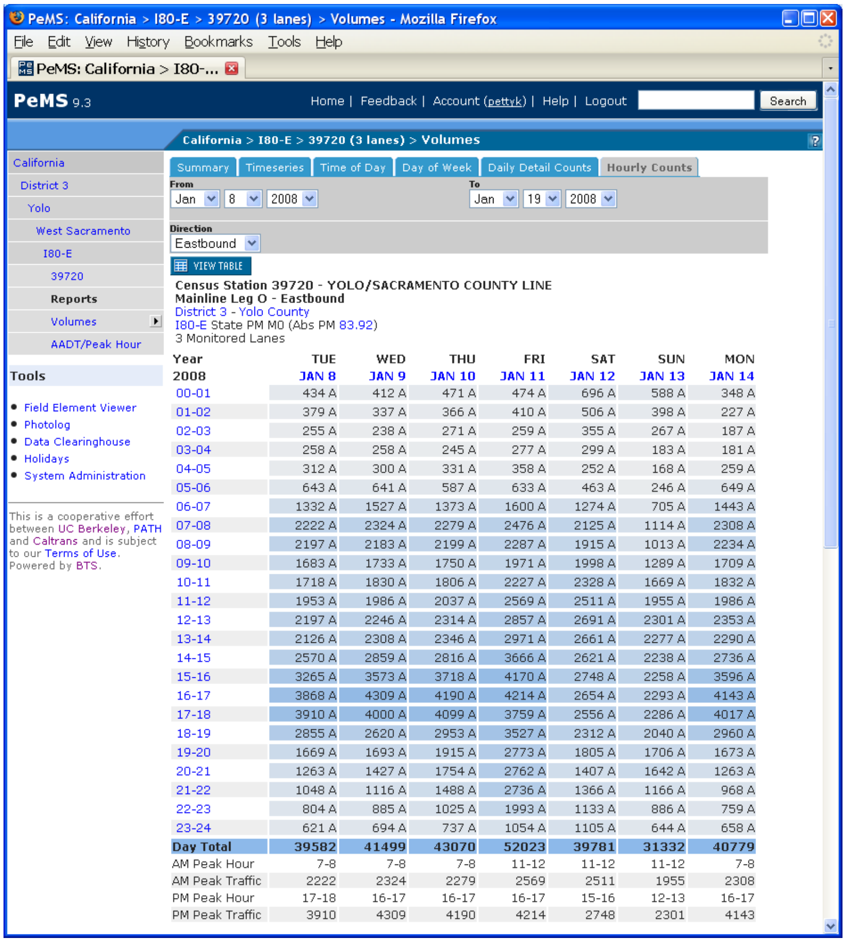
Hence we needed to develop a structure that allows users to actually find the data in time (as opposed to in space, which the Google Maps mashup helps with). What we came up with is a *Data Inventory* widget. You can see an example of the widget in the tooltip popup in Figure 3. In the widget we simply have a table where the rows are all the years and columns are the months of the year. We color the given cell when there is data for this station. You can see in the widget above that there wasn't very much data collected in 2006 for this particular detector in D3. The user can click on the desired cell in the widget to jump from the map directly to a timeseries plot of data for that month. This combination of map and inventory widget provides a nice way to navigate to the detector and the data at the same time.

2.2 Reporting Tools

We built a number of different reporting tools for the Census data. These tools fall into two categories. First, we already have a huge amount of infrastructure in PeMS for display plots and reports of flow for the ITS stations. We adapted these reports so that we can view Census data in addition to ITS data. These include plots of flow versus time with a user-selected granularity of hours, days, weeks and months; plots of flow versus time of day; plots of flow versus day of week; plots of flow down the freeway; etc. Since these are rather standard, we won't discuss them further.

The second set of plots that we build are specific to the Census data. These plots and reports are either an attempt to duplicate an existing Caltrans report or they are something new. We review a few of them here.

One of the existing reports that we wanted to duplicate is the Hourly Count Report, as shown below.

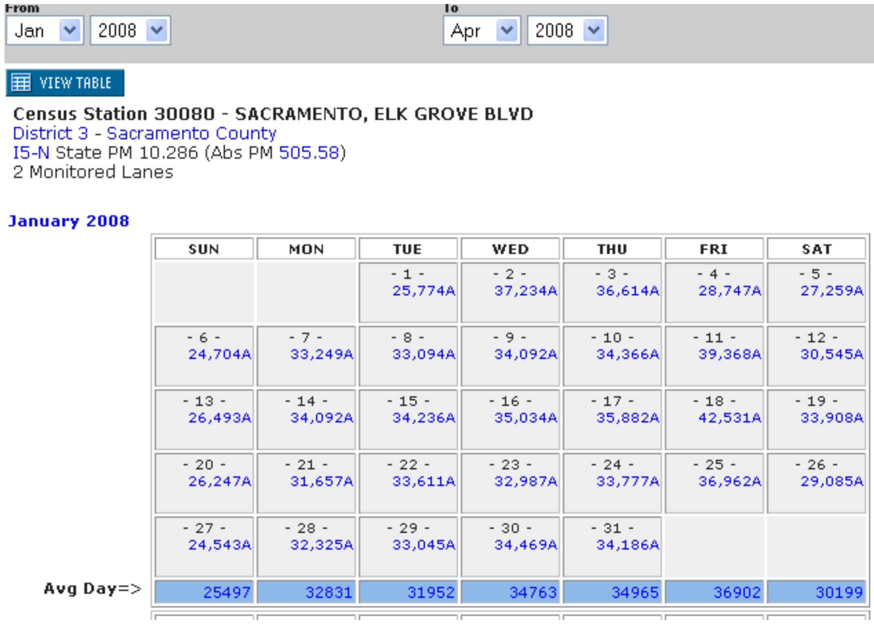


Year	TUE	WED	THU	FRI	SAT	SUN	MON
2008	JAN 8	JAN 9	JAN 10	JAN 11	JAN 12	JAN 13	JAN 14
00-01	434 A	412 A	471 A	474 A	696 A	588 A	348 A
01-02	379 A	337 A	366 A	410 A	506 A	398 A	227 A
02-03	255 A	238 A	271 A	259 A	355 A	267 A	187 A
03-04	258 A	258 A	245 A	277 A	299 A	183 A	181 A
04-05	312 A	300 A	331 A	358 A	252 A	168 A	259 A
05-06	643 A	641 A	587 A	633 A	463 A	246 A	649 A
06-07	1332 A	1527 A	1373 A	1600 A	1274 A	705 A	1443 A
07-08	2222 A	2324 A	2279 A	2476 A	2125 A	1114 A	2308 A
08-09	2197 A	2183 A	2199 A	2287 A	1915 A	1013 A	2234 A
09-10	1683 A	1733 A	1750 A	1971 A	1998 A	1289 A	1709 A
10-11	1718 A	1830 A	1806 A	2227 A	2328 A	1669 A	1832 A
11-12	1953 A	1986 A	2037 A	2569 A	2511 A	1955 A	1986 A
12-13	2197 A	2246 A	2314 A	2857 A	2691 A	2301 A	2353 A
13-14	2126 A	2308 A	2346 A	2971 A	2661 A	2277 A	2290 A
14-15	2570 A	2859 A	2816 A	3666 A	2621 A	2238 A	2736 A
15-16	3265 A	3573 A	3718 A	4170 A	2748 A	2258 A	3596 A
16-17	3868 A	4309 A	4190 A	4214 A	2654 A	2293 A	4143 A
17-18	3910 A	4000 A	4099 A	3759 A	2556 A	2286 A	4017 A
18-19	2855 A	2620 A	2953 A	3527 A	2312 A	2040 A	2960 A
19-20	1669 A	1693 A	1915 A	2773 A	1805 A	1706 A	1673 A
20-21	1263 A	1427 A	1754 A	2762 A	1407 A	1642 A	1263 A
21-22	1048 A	1116 A	1488 A	2736 A	1366 A	1166 A	968 A
22-23	804 A	885 A	1025 A	1993 A	1133 A	886 A	759 A
23-24	621 A	694 A	737 A	1054 A	1105 A	644 A	658 A
Day Total	**39582**	**41499**	**43070**	**52023**	**39781**	**31332**	**40779**
AM Peak Hour	7-8	7-8	7-8	11-12	11-12	11-12	7-8
AM Peak Traffic	2222	2324	2279	2569	2511	1955	2308
PM Peak Hour	17-18	16-17	16-17	16-17	15-16	12-13	16-17
PM Peak Traffic	3910	4309	4190	4214	2748	2301	4143

Figure 4. Hourly count report. Shows flows per hour over each day. Summaries are provided at the bottom of the columns.

The hourly count report shows the reported flow per hour for a number of days in a specific layout, as shown in Figure 4. There are summary columns below each day that indicate the peak hour traffic for each shift. This report is already generated by the TSN Census viewing application. We extended this report a bit by coloring the cells according to the flow and providing links for each day and hour that jump to plots of the data over time.

Another report that we wanted to duplicate is the calendar view of the data, as shown below.



	SUN	MON	TUE	WED	THU	FRI	SAT
			- 1 - 25,774A	- 2 - 37,234A	- 3 - 36,614A	- 4 - 28,747A	- 5 - 27,259A
	- 6 - 24,704A	- 7 - 33,249A	- 8 - 33,094A	- 9 - 34,092A	- 10 - 34,366A	- 11 - 39,368A	- 12 - 30,545A
	- 13 - 26,493A	- 14 - 34,092A	- 15 - 34,236A	- 16 - 35,034A	- 17 - 35,882A	- 18 - 42,531A	- 19 - 33,908A
	- 20 - 26,247A	- 21 - 31,657A	- 22 - 33,611A	- 23 - 32,987A	- 24 - 33,777A	- 25 - 36,962A	- 26 - 29,085A
	- 27 - 24,543A	- 28 - 32,325A	- 29 - 33,045A	- 30 - 34,469A	- 31 - 34,186A		
Avg Day=>	25497	32831	31952	34763	34965	36902	30199

Figure 5. Calendar view of the Census data. This duplicates an existing Census report.

The calendar view shows the total flow per day in a calendar format with averages by day of week at the bottom. We extended this report with the standard PeMS paradigm of providing links to allow the user to drill in to the data.

In terms of new reports that we developed, one of them is the yearly AADT report for a single station, as shown below.

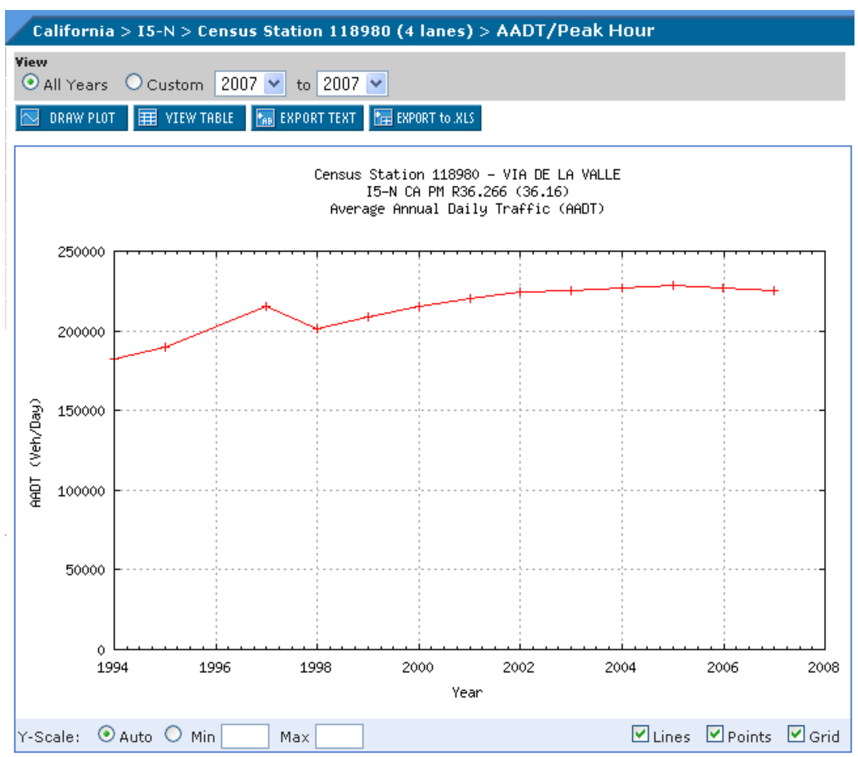


Figure 6. Plot of Caltrans-computed AADT for a Census station for all years.

The above plot is showing the Caltrans-computed AADT for this census station in D11 on I5-N. When we transferred the hourly Census values we also transferred the AADT values for each station as computed by Caltrans. That's what we're plotting here. The table view of this report shows the Caltrans "Peak Hour" report with various engineering quantities.

In summary, we built the following web pages with the Census data:

1. Navigational:
    a. Field element tables that show the number of stations in each district by type.
    b. Field element viewer that shows a map with the location of all the census stations.
    c. Lists of stations in a geographical region by type.
2. For a single station:
    a. Census station configuration page. Shows map, detailed freeway information, substations, and legs.
    b. Summary report showing distribution of flow by peak period.
    c. Standard timeseries plot of flow at hourly, daily, weekly and monthly levels of aggregation.
    d. Time of day plot of flow showing hourly flows versus hour of day.
    e. Day of week plot of flow showing daily flows per day of week.
    f. Daily detail count table. Mimics existing Caltrans report that shows the total flow per day in a calendar view.
    g. Hourly counts table. Mimics existing Caltrans report that shows the hourly flows during each day.
    h. AADT plot for a single station. Shows the AADT for each year for all years for which we have data.
    i. AADT/Peak Hour Table for a single station. Mimics existing Caltrans report that shows, for each year, the AADT, the 1-way peak hour volume, and the K, D and KD factors.
3. For a stretch of freeway:
    a. The list of Census stations down the freeway.
    b. The AADT/Peak Hour report (same content as for a single station).
    c. The Spatial AADT report that shows the AADT for each station down the freeway.

2.3 AADT for ITS Sensors

In addition to incorporating the Census data so that we can display AADT values, we also started to compute AADT for all of the ITS-based detectors. The AADT itself is supposed to represent the average daily traffic at a particular spot. If you had a complete set of data (no missing data) for the entire year then you could simply average the daily total flows and get the correct answer. Unfortunately, we rarely have a complete data set for a year. It's quite common for the data feed to go out for either a short period of time, say 5 minutes, or longer periods of time, like days. We have to properly compute AADT in light of these holes. Doing so involves recognizing that a hole might occur on a weekend in May versus a weekday in December. Since there are seasonal and day-of-week patterns in traffic flow, we have to use special methods for computing the yearly averages that can handle failures like this.

The literature has developed a total of eight different ways to compute AADT for ITS stations[3]. We briefly list them here:

1. **Arithmetic Mean**: Average of all days.
2. **ASTM Standard 1442**: Compute MADW (84), then MADT (12), then average them. Allows for missing weekdays.
3. **Conventional AASHTO Procedures**: Compute MADW (84), then AADW (7), then average them.
4. **Provisional AASHTO Procedures**: Compute MAHW (168), then MADW (84), then AADW (7), then average.
5. **Sum of 24 Annual Average Hourly Traffic Volumes**: Compute hourly average over whole year, then sum them.
6. **Modified ASTM Standard**: Same as ASTM Standard but allow for missing MADT values.
7. **Modified Conventional AASHTO**: Same as AASHTO but 1 of 12 MADW, and/or 1 of 7 AADW values may be missing.
8. **Modified Provisional AASHTO**: Same as Provisional AASHTO but 1 of 12 MADW, and/or 1 of 7 AADW values may be missing.

We implemented a new data processing step to compute AADT for each ITS station in PeMS for each of the methods above. For these computations we toss any data which is imputed. As a result, some of the methods can't be computed with the remaining data for certain stations.

The computation for AADT is done over an entire year. For Caltrans the AADT values are from October 1st of one year through September 30th the next. We do the computation starting for every month, instead of just for October. This means that we have an AADT value starting from October 1st, November 1st, etc. An example of this is given below:

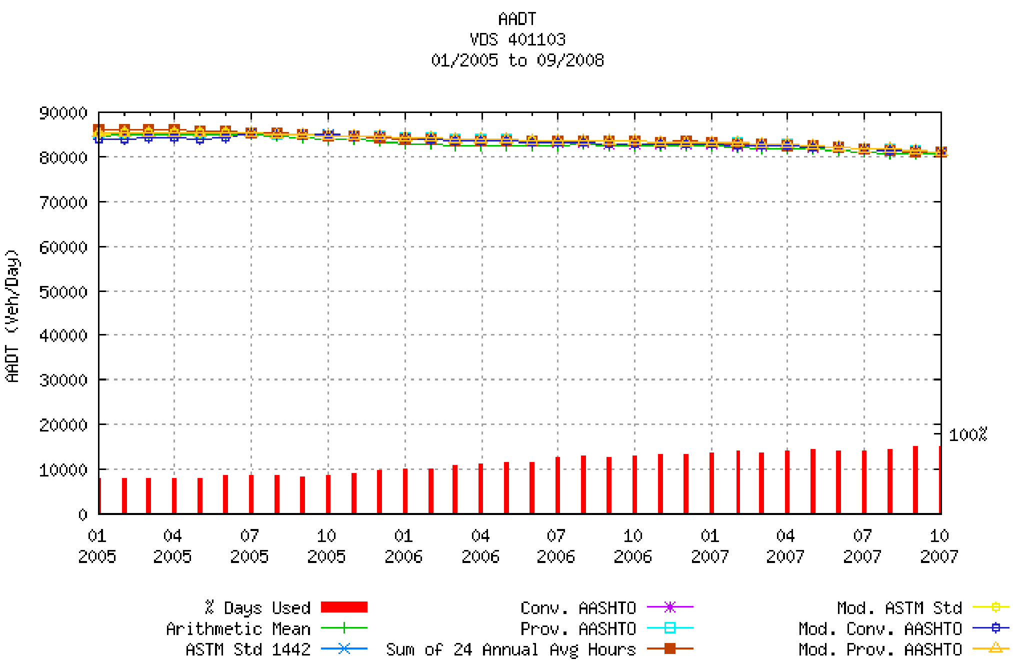


Figure 7. AADT for an ITS sensor showing the monthly values.

---

[3] "*Using Incomplete Archived ITS Data to Calculate Annual Average Traffic Statistics*," Shawn Turner, Eun Sug Park. TRB 2008. Washington, D.C.

In the figure above, we are showing the AADT values starting for each month from January 2005 through October 2007. The first data point covers January 2005 through December 2005. The last data point covers October 2007 through September 2008. Hence the last data point is the latest sample of AADT that we can compute (we grabbed this plot in October 2008). On the plot we are showing the results of all eight methods when they can be computed. Since for some months we can't implement some of the methods due to missing data samples, those values are missing. The red bar plotted along the bottom of the plot is showing the percentage of days out of 365 that were completely observed. We can see in this plot that the AADT has been declining slightly from around 85,000 veh/day to around 80,000 veh/day over the course of about two years. This type of plot is useful when trying to examine how VMT has dropped in response to fuel prices.

2.4 Comparing ITS Sensors with Census Sensors

Although this wasn't a formal part of this project, we did some preliminary investigation into whether the ITS-based values match the Census-based values. In a number of cases the physical sensors in the ground are used by both the TMC folks to extract real-time data as well as the Census folks to collect counts. They are simply connected to two controllers in the same cabinet that use different communication lines back to the District offices. The real-time data for ITS purposes is transported back over dedicated lines, either telephone lines or fiber, that are always connected. The Census controllers are connected to dial-up phone lines that are polled periodically by the Census staff in the district. In other cases the detectors are close enough that they should be reporting values that are very similar.

The point of our investigation was to see if the flow values reported by the Census stations matched what was reported by the ITS stations. We started off looking at the yearly values of AADT as computed from the ITS stations and computed by the Census Team. Unfortunately, they were nowhere close. We then backed up and started looking at the daily values of flow for individual stations.

We dive into a specific pair of stations on I5-N in D12 to explain in detail what we've found.

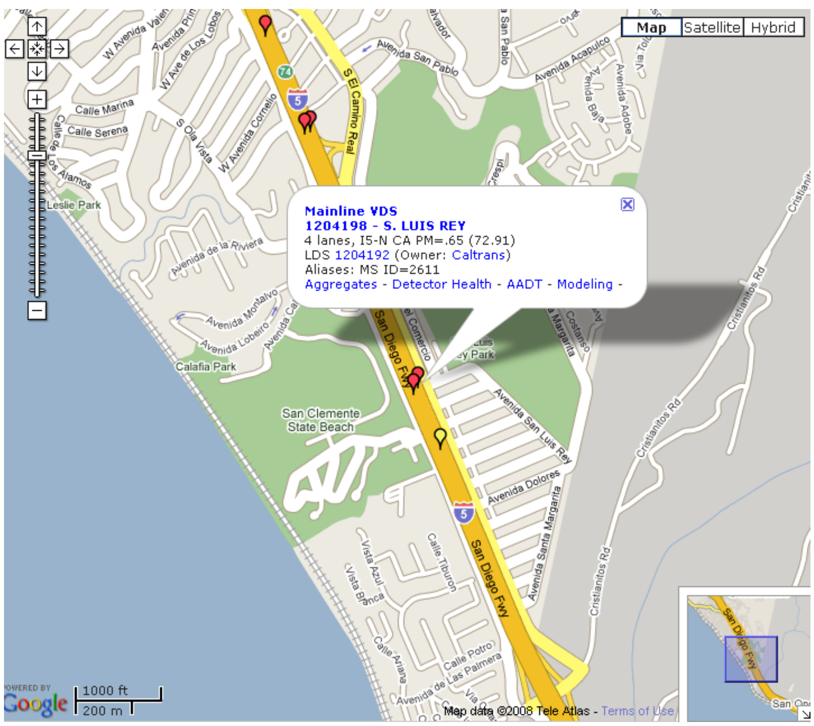


Figure 8. Southern part of D12 by San Onofre State Beach.

We selected a VDS station, 1204198, and a Census station, 124010, that are very close to each other without any intervening on- or off-ramps, as shown in Figure 8. They were also working for the entire time period that we studied (or we tossed any days of data that were imputed). We examined the daily flow from 11/1/07 – 12/31/07. The results are given below.

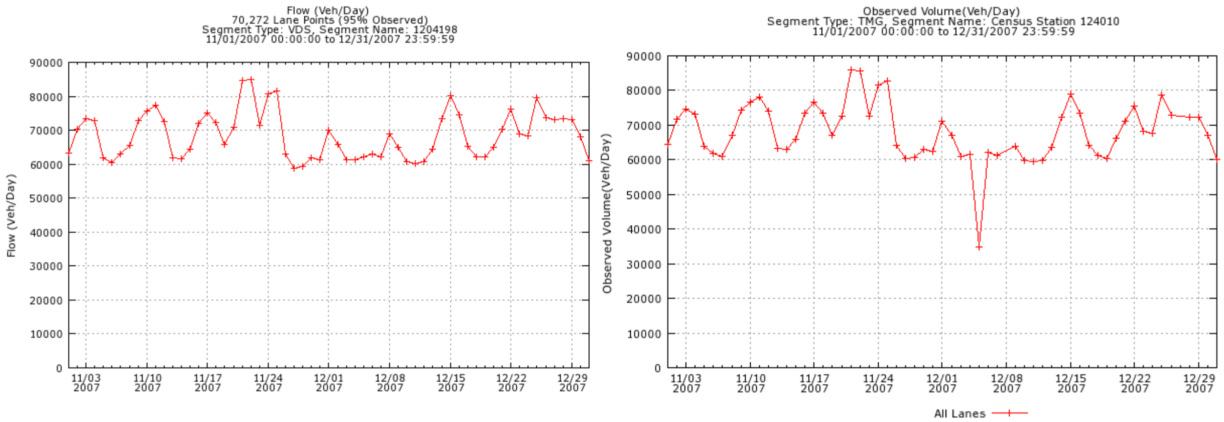


Figure 9. Daily flow for VDS station, left, and Census station, right, for the same time period.

In Figure 9 we are showing the daily flows for the two stations. Besides the anomalous point during the first part of the November 2007 for the Census station, they flows look very similar. We plotted them against each other in a scatter plot on the left below.

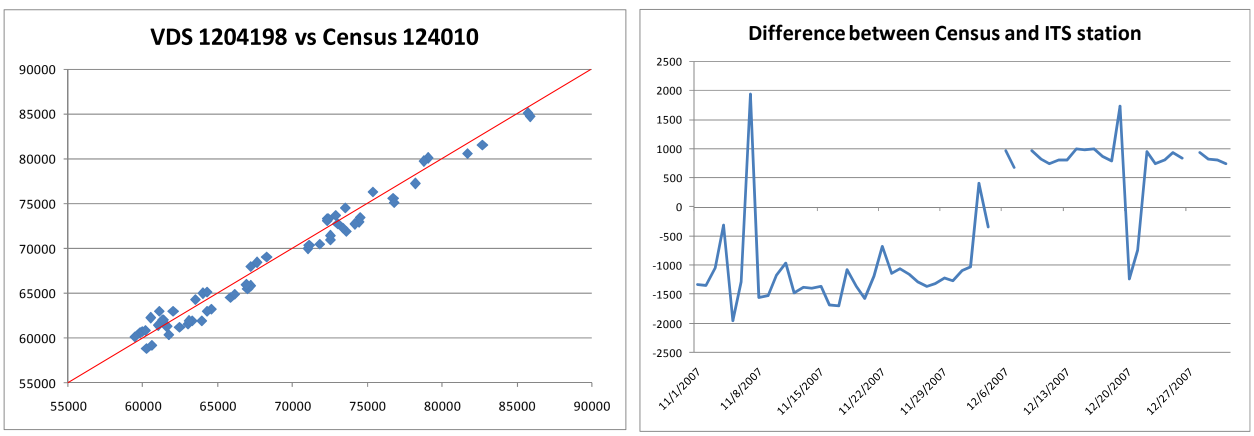


Figure 10. Comparing the daily flows of the two stations against each other. On the left is a scatter plot, on the right we plot the difference over time.

We can clearly see that something is amiss. The points are highly correlated with each other, which is good. But they seem to bunch into two bands around the red line, which is y=x. When we plot the difference in the daily flows over time we see the issue. Around 12/10/07 there was a change. Before that the ITS daily flows were consistently higher than the Census flows. After that they are consistently lower. It's unclear what caused this change. We compared the ITS station with its upstream neighbor to see if the shift was in this single ITS station. There was no noticeable change like on the right side of Figure 10. So either the ITS stations shifted together or the Census station shifted by itself.

We looked at a huge number of other stations across the state to see if we could find any pair of stations that matched up with each other. We typically saw two types of results, as illustrated below.

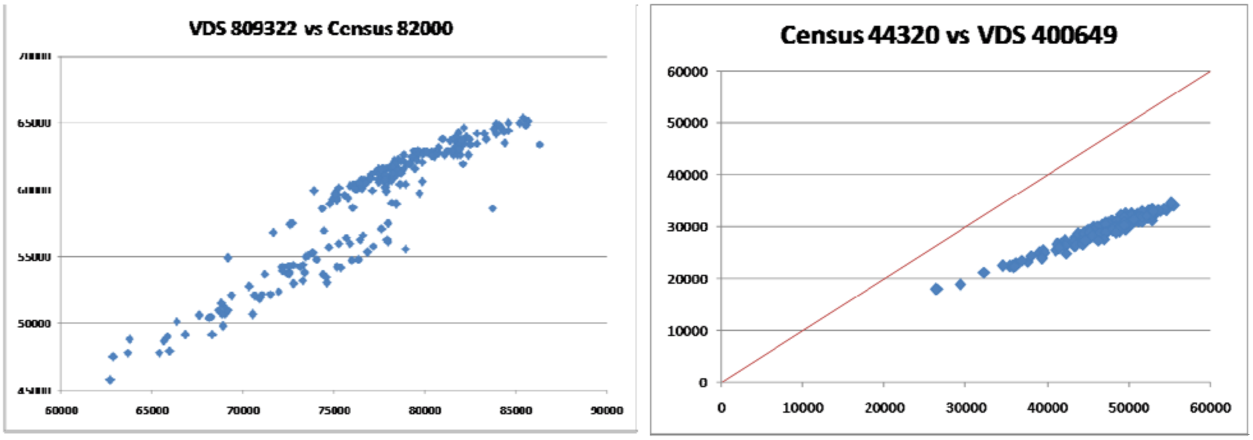


Figure 11. Two more examples comparing ITS and Census stations. The left is in D8 and the right is in D4.

We either saw a shift, like the figure on the left, which was based on a pair of detectors in D8. Or we saw a simple bias, like the figure on the right, which was based on a pair of detectors in D4. In no cases across the state did we see a match between the ITS data and the Census data. This was unfortunate, of course. It deserves further investigation that is outside the scope of this project.

3.0 Task 3: Investigation of Fidelity of 3rd Party Detectors

Up until now we've treated all of the detectors in PeMS the same. We assumed that they sent the same measurements at the same frequency in the same units. We subsequently applied the same thresholds and performed the same aggregations. This model has benefits in terms of simplicity of code and speed of code execution. But it also has severe limitations. Specifically, PeMS was not able to utilize speed values that are reported by some sensors. In addition, we had to perform the same diagnostics and roll up the data the same for every sensor, even when the sensors are different.

Tasks 3, 5 and 7 involved changing this underlying model in PeMS so that we can start to treat individual detectors differently. This required us to change some of the fundamental infrastructure of PeMS. Specifically, the underlying schema that we use to store the detectors in PeMS had to be extended in order to store the additional attributes that are specific to a sensor. These attributes are then used by the diagnostic, aggregation and display code to perform specific actions upon individual sensors.

The first change that we implemented with this new infrastructure involved adding the detector owner to each station. The data that PeMS collects comes from the District TMC. Some of the sensors feeding data to the TMCs are owned by Caltrans but others are owned by 3rd parties that sell the data to Caltrans (or exchange it for other services). We wanted to attach the owner to each detector so that we could facilitate the investigation of the fidelity of detectors by owner. We implemented this feature in the underlying data structures as well as in the diagnostic web pages.

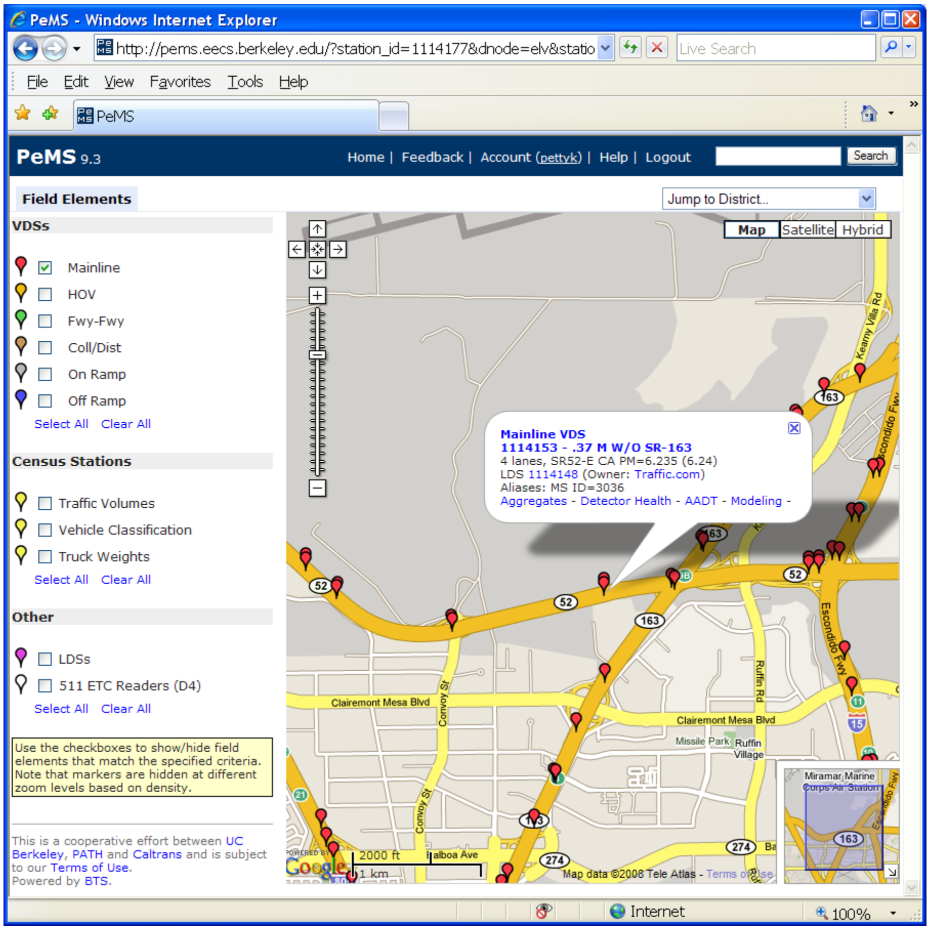
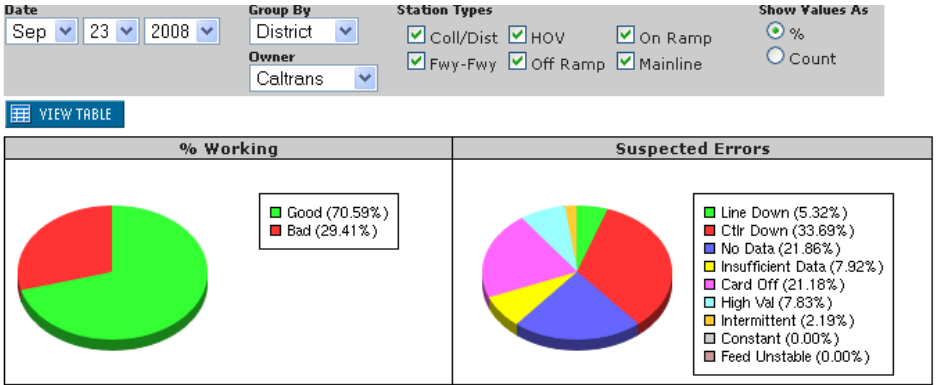


Figure 12. Field element viewer with popup showing the owner of the station, in this case Traffic.com.

We now place the owner of the stations in a number of locations on the web pages. In Figure 12 above we've placed the owner in the little popup that shows up for each detector. We can see on this station in D12 on SR52 that the owner is Traffic.com. In addition, we place the owner on the VDS and controller configuration pages, as you can see below in Figure 20.

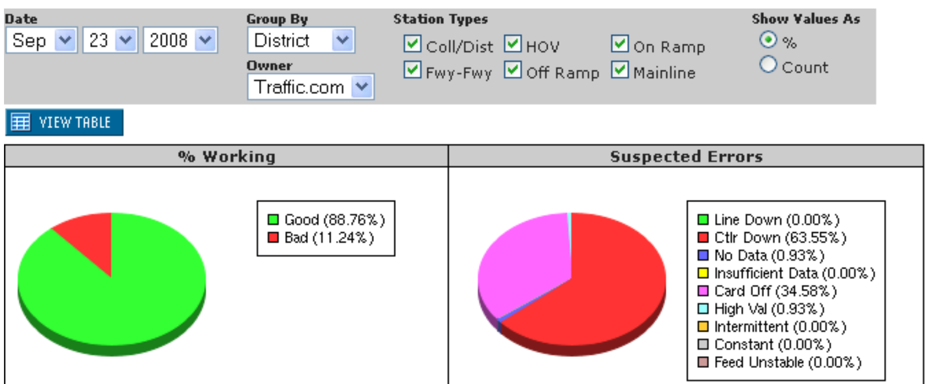
In addition to just showing the owners of the detectors, we also let users filter a number of reports based on the owner. We do this for navigation pages where we list out the detectors in a district or on a freeway. And we also do this extensively in the diagnostic section of PeMS. Specifically, we allow users to see all of the diagnostic reports by owner. For example, below we took a screen shot of the state-wide diagnostic page in PeMS from September 23rd, 2008 for both the Caltrans-owned and 3rd party-owned detectors.



Status by District

District	# Det	% Good	% Bad	Suspected Error: Line Down	Ctlr Down	No Data	Insufficient Data	Card Off	High Val	Intermittent	Constant	Feed Unstable
3	1442	77.8	22.2	0.0	8.5	0.0	3.3	7.4	2.3	0.6	0.0	0.0
4	4720	68.8	31.2	0.0	0.0	18.5	1.1	8.0	3.1	0.6	0.0	0.0
5	23	26.1	73.9	0.0	73.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	78	59.0	41.0	0.0	29.5	0.0	7.7	1.3	2.6	0.0	0.0	0.0
7	8916	60.4	39.6	2.6	15.4	6.1	4.0	7.7	2.8	0.8	0.0	0.0
8	1192	73.0	27.0	6.2	10.8	0.2	5.3	3.6	0.5	0.4	0.0	0.0
10	646	97.1	2.9	0.0	1.1	0.0	1.4	0.2	0.3	0.0	0.0	0.0
11	2921	84.9	15.1	0.5	6.0	0.6	0.7	5.1	2.0	0.2	0.0	0.0
12	5037	76.4	23.6	1.3	12.4	3.2	0.5	3.9	1.5	0.8	0.0	0.0
Totals	**24,975**	**70.6**	**29.4**	**1.6**	**9.9**	**6.4**	**2.3**	**6.2**	**2.3**	**0.6**	**0.0**	**0.0**

Figure 13. State-wide diagnostic page for the Caltrans-owned sensors.



Status by District

District	# Det	% Good	% Bad	Suspected Error: Line Down	Ctlr Down	No Data	Insufficient Data	Card Off	High Val	Intermittent	Constant	Feed Unstable
8	457	87.3	12.7	0.0	6.8	0.0	0.0	5.7	0.2	0.0	0.0	0.0
11	495	90.1	9.9	0.0	7.5	0.2	0.0	2.2	0.0	0.0	0.0	0.0
Totals	**952**	**88.8**	**11.2**	**0.0**	**7.1**	**0.1**	**0.0**	**3.9**	**0.1**	**0.0**	**0.0**	**0.0**

Figure 14. State-wide diagnostic page for the Traffic.com-owned sensors.

We can see that on this particular day the Caltrans-owned sensors are reporting about 70% good, while the Traffic.com sensors are reporting 88% good. Traffic.com only has sensors in D8 and D11 that are reporting to PeMS. In addition to the main diagnostic page, users can filter all of the individual diagnostic pages for detector owner.

There are 3rd party sensors in D4 (owned by Traffic.com and Speedinfo) but we've been unable to come to an agreement with the district on incorporating those sensors into PeMS.

4.0 Task 4: Investigate FasTrak Data Fusion Methodology

MTC has blanketed the Bay Area with tag readers that observe the FasTrak tags in vehicles as they go by. The tag readers in the Bay Area can be shown in PeMS on the field element viewer, which we show below.

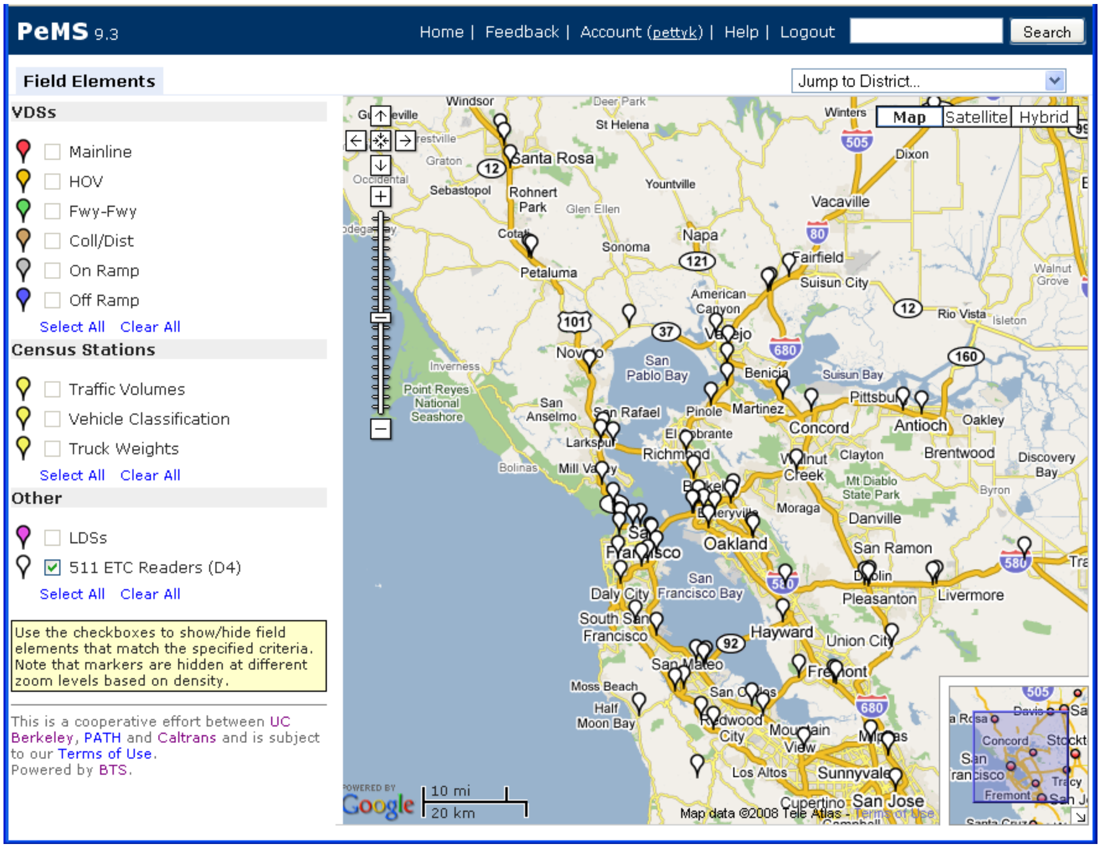


Figure 15.The FasTrak tag readers in the Bay Area as seen in the field element viewer in PeMS.

PeMS has been receiving the FasTrak data from MTC for about two years. The raw data are the samples from each tag reader of when an individual tag went by the reader, or the tuple of [timestamp, reader ID, tag ID]. This can be chained together in order to determine the path that the user took during the entire day. After a day the users tag ID is rotated and further associations are lost. But during the single day, we have the origin and destination of each of the tags that went through the system.

In a paper by Kwon and Varaiya[4] it was demonstrated that it's possible to estimate the OD matrix of the entire population by combining the FasTrak-based OD and the measured flows at each detector that's adjacent to each tag reader. We've implemented a version of that algorithm in PeMS in order to track the OD matrix over time.

The algorithm works in the following high-level steps:

1. Trip summary generation.
    a. We compile the raw data (timestamp, reader ID, tag ID) into vehicle trajectories (the list of tag readers and when they passed them) over the course of a day.

[4] "Real-Time Estimation of Origin-Destination (O-D) Matrices with Partial Trajectories from Electronic Toll Collection Tag Data," by Kwon, J. and Varaiya, P. (2005). Transportation Research Record no. 1923, Transportation Research Board, pp. 119-126.

   b. We cut each trajectory into trips where the consecutive timestamps are less than an hour apart. A trip is a sequence of tag reader IDs.

   c. We count the number of each trip for each hour of each day.

   d. After this the raw data can be tossed and the trip counts are stored.

2. Detection rate computation.

   a. We assume a penetration rate of 40% for all tag readers (the penetration rate and detection rate are confounded – we need to assume one).

   b. We estimate the detection rate at each tag reader by comparing the volume from adjacent loops with the total flow measured by the tag reader.

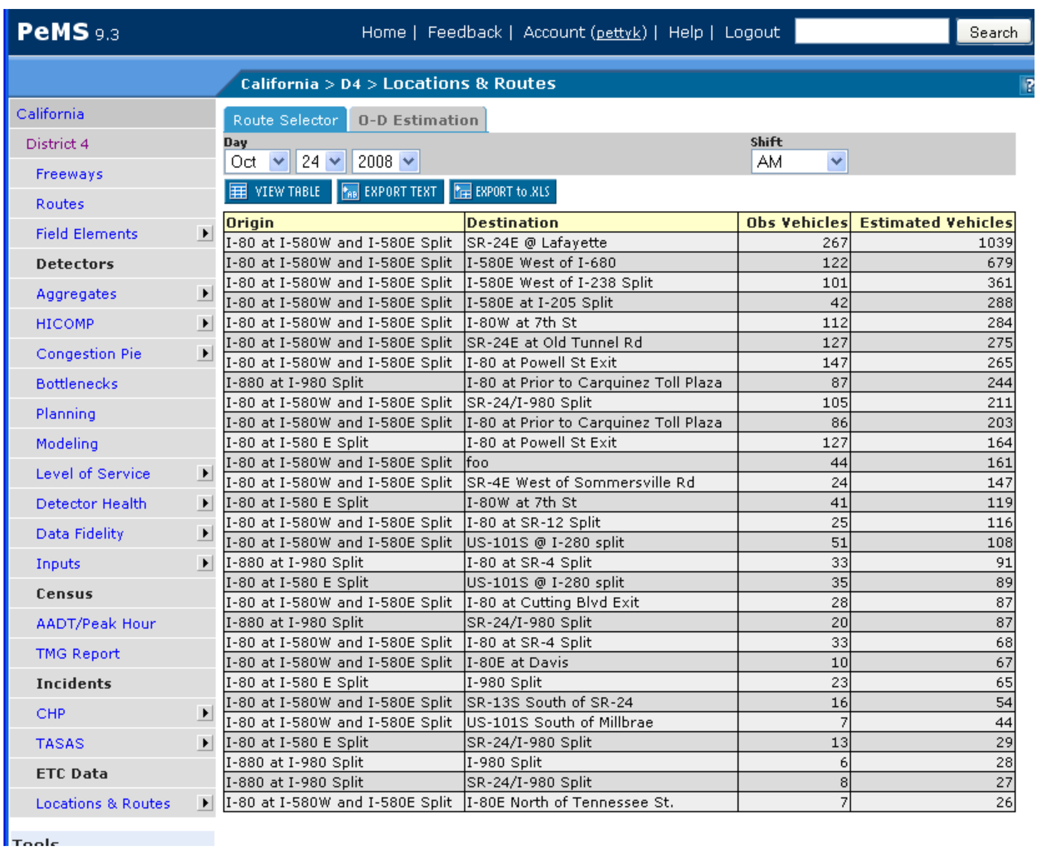
3. OD matrix calculation

   a. We use Bayesian statistics to estimate the likelihood of each OD pair given the observations of trip frequencies and detection rates.

   b. We then store the records by: day, shift (AM, midday, PM), OD, observed volume, estimated volume.

We've implemented this algorithm for PeMS 9.0. The algorithm runs in real-time due to the constraint from MTC that we purge the raw data samples within 24 hours.

Below we have a screen shot showing the results of the OD estimation algorithm.



Origin	Destination	Obs Vehicles	Estimated Vehicles
I-80 at I-580W and I-580E Split	SR-24E @ Lafayette	267	1039
I-80 at I-580W and I-580E Split	I-580E West of I-680	122	679
I-80 at I-580W and I-580E Split	I-580E West of I-238 Split	101	361
I-80 at I-580W and I-580E Split	I-580E at I-205 Split	42	288
I-80 at I-580W and I-580E Split	I-80W at 7th St	112	284
I-80 at I-580W and I-580E Split	SR-24E at Old Tunnel Rd	127	275
I-80 at I-580W and I-580E Split	I-80 at Powell St Exit	147	265
I-880 at I-980 Split	I-80 at Prior to Carquinez Toll Plaza	87	244
I-80 at I-580W and I-580E Split	SR-24/I-980 Split	105	211
I-80 at I-580W and I-580E Split	I-80 at Prior to Carquinez Toll Plaza	86	203
I-80 at I-580 E Split	I-80 at Powell St Exit	127	164
I-80 at I-580W and I-580E Split	foo	44	161
I-80 at I-580W and I-580E Split	SR-4E West of Sommersville Rd	24	147
I-80 at I-580 E Split	I-80W at 7th St	41	119
I-80 at I-580W and I-580E Split	I-80 at SR-12 Split	25	116
I-80 at I-580W and I-580E Split	US-101S @ I-280 split	51	108
I-880 at I-980 Split	I-80 at SR-4 Split	33	91
I-80 at I-580 E Split	US-101S @ I-280 split	35	89
I-80 at I-580W and I-580E Split	I-80 at Cutting Blvd Exit	28	87
I-880 at I-980 Split	SR-24/I-980 Split	20	87
I-80 at I-580W and I-580E Split	I-80 at SR-4 Split	33	68
I-80 at I-580W and I-580E Split	I-80E at Davis	10	67
I-80 at I-580 E Split	I-980 Split	23	65
I-80 at I-580W and I-580E Split	SR-13S South of SR-24	16	54
I-80 at I-580W and I-580E Split	US-101S South of Millbrae	7	44
I-80 at I-580 E Split	SR-24/I-980 Split	13	29
I-880 at I-980 Split	I-980 Split	6	28
I-880 at I-980 Split	SR-24/I-980 Split	8	27
I-80 at I-580W and I-580E Split	I-80E North of Tennessee St.	7	26

Figure 16. Screenshot of OD estimation results.

This is showing the results in tabular form. We simply show the OD pairs with the highest estimated number of vehicles for the selected shift.

It would be nice to show the results in a graphical form as well. For example, it would be nice to show a map with the tag readers on it that illustrates the number of vehicles passing between any two points. This would be similar to the mockup we've included below.

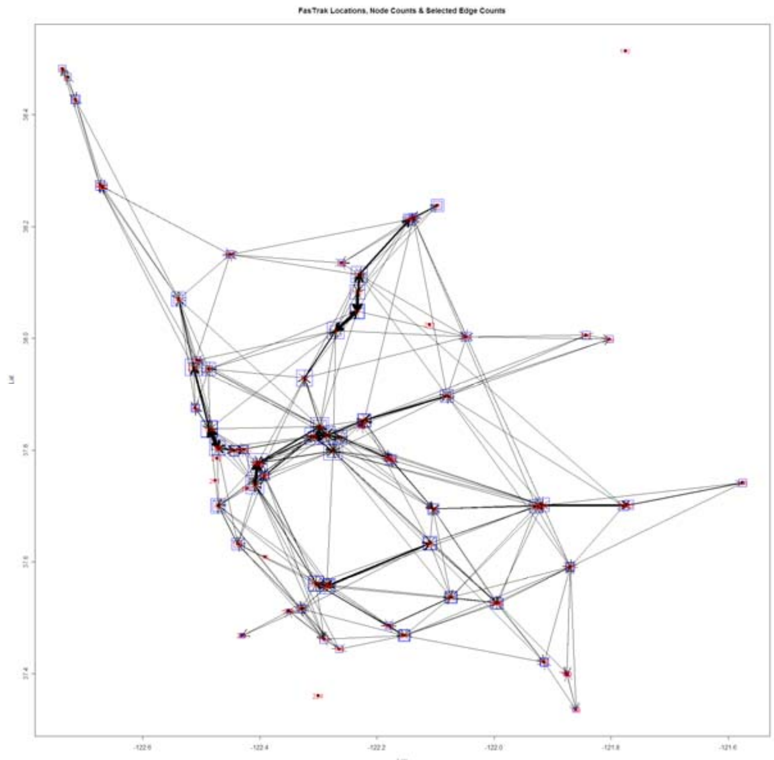


Figure 17. FasTrak OD results on a map. Thickness of line is a function of the number of vehicles.

This figure shows, for a given time period, the flows between individual points with the thickness of the line being a function of the number of vehicles making that trip. The graphical representation is a future project.

5.0 Task 5: Investigate Using Field Speed Measurements

This task involved extending PeMS so that we can use the reported speeds from the detectors. As we described in Section 3.0, this involved changing the underlying schema in PeMS that holds the detector configuration. We needed to indicate in the schema which detectors are reporting speed and for which detectors we still need to compute speed. We implemented this feature in the schema, in the grinding code and in the web configuration management pages and code.

Not all Districts can actually send speed values. The binary format used to send data to PeMS in Districts 3, 6, 7, 8, 11 and 12 currently doesn't have the ability to send additional information other than flow and occupancy per detector. Although there are plans afoot to upgrade to a new version of that protocol, they currently can't use this feature. The two districts that can use this feature are Districts 4 and 10. District 10 has started using this feature.

In order to utilize this feature, PeMS needs to be told which detectors are going to be reporting speed. District 10 provides the configuration of detectors to PeMS via a spreadsheet. They modify the spreadsheet locally and when they are done they upload it to PeMS via a web page. Since we are in charge of the configuration (it's just a spreadsheet) we are able to change the spreadsheet easily. We added an extra column to the station sheet in the spreadsheet, a section of which is shown below.

station_id	type	physical _lanes	county _id	freeway _id	dir	active	controller _id	state_post mile	name	Spd	Diag
1020110	ML	3	77	99	S	0	1020100	19.11	SB 99 S/O On Ramp from EB Fremont St/Rte 26 East	0	Urban
1020210	ML	1	109	108	E	0	1020200	R0.00	EB 108 Sonora, JCT 49	0	Urban
1020310	ML	1	109	108	W	0	1020300	R0.00	WB 108 Sonora, JCT 49	0	Urban
1020410	ML	3	77	5	N	1	1020400	R16.623	NB 5 N/O Louise Ave UC	1	Urban
1020510	ML	3	77	5	N	1	1020500	R17.213	NB 5 S/O Lathrop Rd	1	Urban
1020610	ML	3	77	5	N	1	1020600	R18.023	NB 5 N/O Lathrop Rd On Ramp	1	Urban
1020710	ML	3	77	5	N	1	1020700	R18.906	NB 5 N/O Lathrop Rd On Ramp	1	Urban
1020810	ML	3	77	5	N	1	1020800	R19.504	NB 5 S/O Roth Rd UC	1	Urban

Table 1. Example of the spreadsheet configuration used in D10. The extra column to indicate that the sensor reports speed is on the end.

We added the extra column called "Spd". A 1 indicates that the sensor reports speed and that we should use it. Otherwise we should compute speed as we normally do.

District 10 has already changed a number of their sensors from computed speeds to measured speeds. Hence we are using the reported speeds in PeMS. While we don't know directly if the speeds are better, we can still investigate them qualitatively. In the figures below we are showing a scatter plot of 5-min values of speed versus flow for the same sensor over a day from two different time periods.

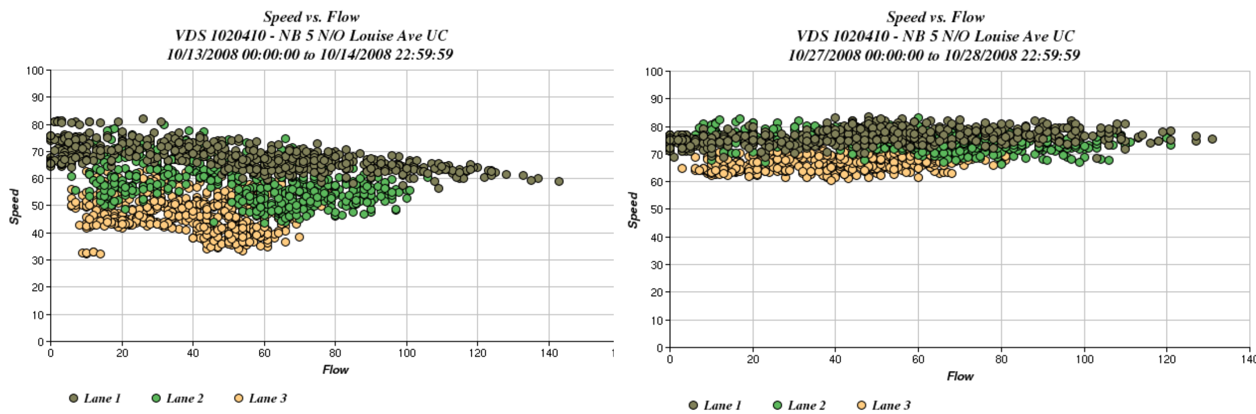


Figure 18. Scatter plots of speed versus flow comparing computed versus measured speeds in D10.

In the figure on the left we are showing data from the time period when PeMS was computing the speeds based on a g-factor. In the figure on the right we are showing data from the time period after we switched to using the reported speeds. The qualitative observation that we can make is that the figure on the right looks "better" than the one on the left – the individual lanes maintain a flat speed over a wide range of flows that are well below capacity. This is what is typically observed. The points on the left are widely scattered at the lower ranges and don't seem to be flat at all over a wide range of flows.

6.0 Task 6: Investigation of User Responses to New Photolog Format

This task involved incorporating the new photolog data that has been collected recently. The old photolog data consisted of movies taken down the freeway. The new photolog data consists of high-resolution images taken every 1/100th of a mile facing forward and sideways.

The new photolog data set is massive. It takes up about 1.2TB of disk space. This far exceeded the capacity of the PeMS web server machine at UCB. Hence we purchased a new machine from Sun. The new machine is an X4500 server that has 24TB of raw disk space. We organized the raw space into protected RAID volumes and placed the new photolog images there. As a side note, we're using the extra space provided by this machine to hold the backups of all the raw data in PeMS, as well as the WIM data that we've started collecting for PeMS 10.

Once the data was in PeMS we needed to display it to the users. The old paradigm that we used was to have little movie icons on the map that users can click on to play the movie. We decided to update this to a Google Map mashup, a screen shot of which is shown below.

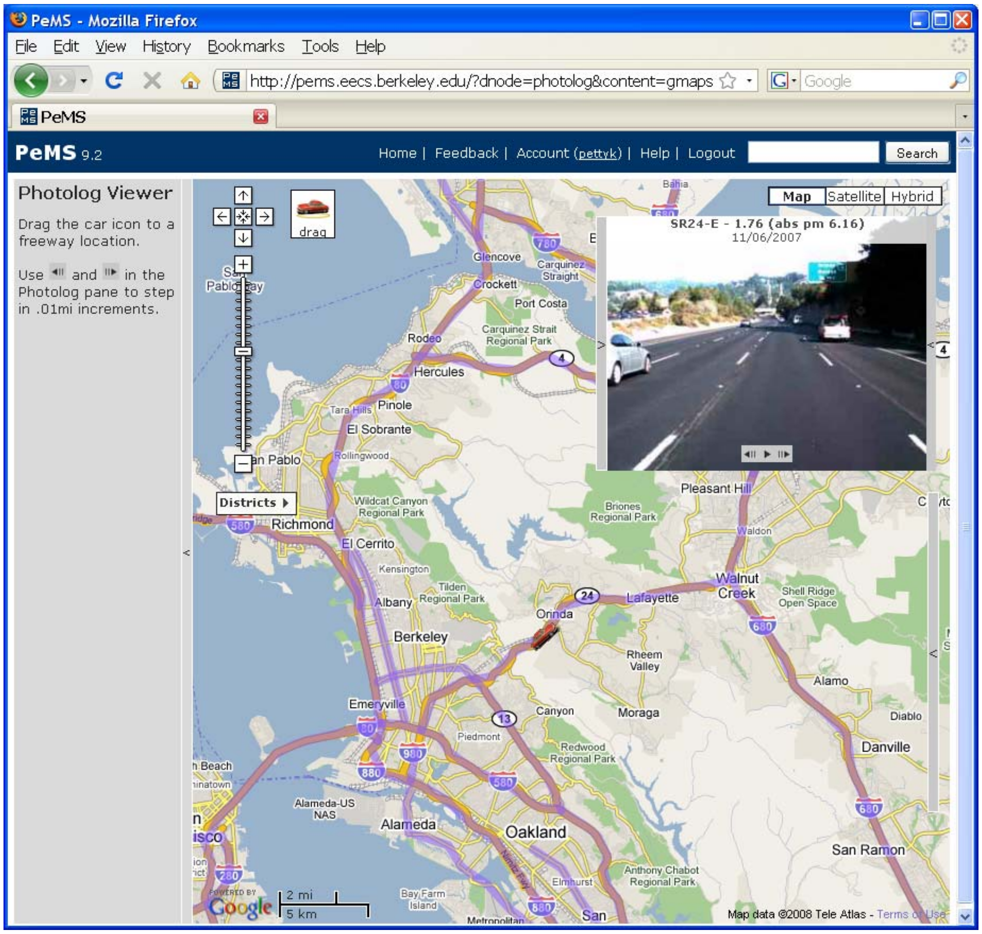


Figure 19. Example of new photolog viewer.

The new mashup starts by highlighting the roads on which we have photolog data (inspired by Google Street Views). In Figure 19 these are the roads that are colored purple. In order to see the photolog image, the user drags the little image of the car over to someplace close to the road that they want to view. The car snaps to the road and displays the image from that location in a popup in the upper right-hand corner, as shown. The user can then step down the freeway to the next picture by using the little controls that show up on top of the picture. In those controls there is a *play* button that causes the system to fetch the next image every ½ a second, making it look like we're playing a movie. Users can also double click on the small image and get the full-resolution image that's very large. Finally, below the photolog image we have a pane that is currently collapsed that holds freeway geometric information from TSN. Users can click on the little arrow to open the pane.

The photolog images present an interesting management problem. The photolog images are taken every three years. It was our original understanding that this meant that they cover the entire state every three years and that we'll simply get a new dump of the entire state that we can copy into place. It turns out that they don't do that. Instead, they are continuously filming roads around the state on a three-year rotation (which should have been obvious). As a result, users might want to know when the picture they are looking at was taken. We added the date the picture was taken on the top of the frame of each image. In addition, we've provisioned the programs so that we can copy in updated photolog images over time.

7.0 Task 7: Research Appropriate Diagnostic Routines for Rural Facilities

Previously PeMS could only apply one set of diagnostic thresholds to all of the detectors in PeMS. The diagnostic thresholds were determined a while ago by looking at a large amount of data and picking sensible thresholds. Since PeMS started in the large urban districts, the thresholds were all "tuned" to traffic in those regions. When we started folding in the rural districts, like Districts 5, 6 and 10, we noticed that there were detectors that were being marked by the diagnostic routes as bad when the raw data itself looked good. It turns out that the thresholds that we developed in urban regions were causing false alarms in rural regions. Specifically, in urban regions we required that there not be too many raw samples with the occupancy equal to zero. If there were more than 59% of the samples with occupancy equal to zero we would label the detector as bad. But in rural locations, where there isn't very much traffic, this doesn't work.

We defined in PeMS the concept of a threshold set. This is a set of parameters that are used by the diagnostic routines to determine if the data from a detector is bad. We defined two threshold sets: one urban, which is just the original defaults, and one rural, with one parameter changed. The two sets and their thresholds are given below.

Category	Test	Rural	Urban
Individual data sample thresholds			
	High Flow Threshold (only used for ramps)	20	20
	High Occupancy Threshold (only for mainline)	0.7	0.7
Detector count thresholds			
	Repeat Occupancy count (absolute count)	50	50
	Percentage of samples where Flow = 0, Occ > 0	2%	2%
	Percentage of samples with High Occupancy (mainline)/Flow (ramps)	20%	20%
	Percentage of samples with Occ = 0 and Flow > 0	50%	50%
	Percentage of samples with Occupancy = 0	75%	59%

Table 2. The two threshold sets defined in PeMS.

Once the threshold sets were defined, we had to have a way to assign them to each station. We extended the detector configuration schema so that every detector can now have its own threshold set, or point to a common one. To change the threshold set for the entire district we simply modify a configuration parameter in PeMS and the change is applied. To change the threshold set on a station by station basis there has to be a way for the district to indicate that in the configuration that they supply to us. In a similar manner to using the speeds on a detector-by-detector basis, this can only be done easily with certain districts – districts that already supply this information in their configuration or whose configuration can be changed to supply it. Specifically, this can be done for Districts 4, 6 and 10.

We implemented this change for District 10. For them to indicate which station should use which threshold set we simply extended the spreadsheet that they use to supply their configuration information to PeMS, as we discussed in Section 5.0. Specifically, Table 1 above shows the new column on the end where the user can specify which threshold set to use. District 10 has done this for a number of their sensors. We have a screen shot below of a configuration page for one of the sensors.

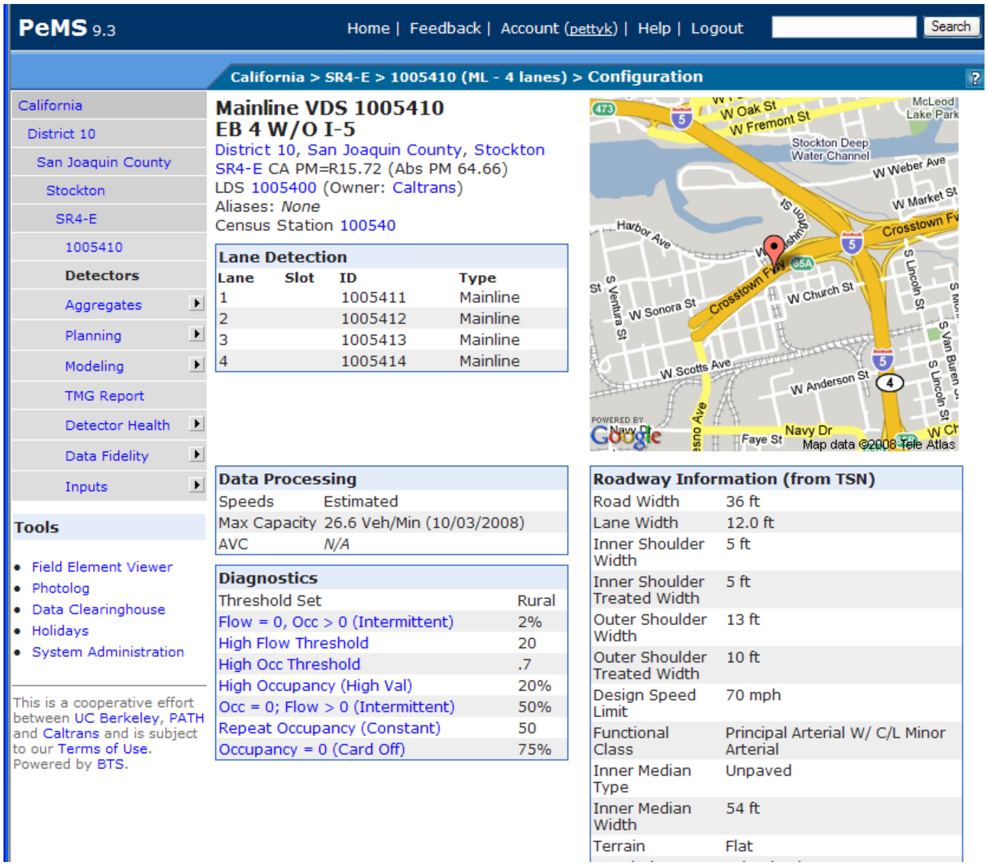


*Figure 20. Station configuration screen showing diagnostic set used for detectors. This particular one is using the **rural** set.*

You can see that we've added the information box labeled **Diagnostics** and inside that we're showing what thresholds we're using, which in this case is the Rural threshold set. If the user hovers on a line in that box a tooltip pops up that explains the test in more detail.

8.0 Conclusion

The collection of work items with this project reflect a number of operational tasks and feature requests that came up during the course of the year. By quickly implementing these we are able to keep PeMS up to the standards expected by Caltrans.

9.0 Acknowledgements

The research summarized here is the joint effort of the PeMS Development Group, especially Alexander Skabardonis and Pravin Varaiya of U.C. Berkeley; Jaimyoung Kwon of CSU East Bay; and the development team at BTS. We have benefited greatly from the advice, interest and guidance of Nick Compin, Carlos Ruiz, Fred Dial and John Wolf of Caltrans, Division of Traffic Operations, and Tarek Hatata of System Metrics Group.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of or policy of the California Department of Transportation. This report does not constitute a standard, specification or regulation.