Probabilistic Analysis of a Large-Scale Urban Traffic Sensor Data Set

Jon Hutchins *
Dept. of Computer Science
University of California, Irvine
CA 96297-3435
johutchi@uci.edu

Alexander Ihler
Dept. of Computer Science
University of California, Irvine
CA 96297-3435
ihler@ics.uci.edu

Padhraic Smyth
Dept. of Computer Science
University of California, Irvine
CA 96297-3435
smyth@ics.uci.edu

ABSTRACT

Real-world sensor time series are often significantly noisier and more difficult to work with than the relatively clean data sets that tend to be used as the basis for experiments in many research papers. In this paper we report on a large case-study involving statistical data mining of over 300 million measurements from 1700 freeway traffic sensors over a period of seven months in Southern California. We discuss the challenges posed by the wide variety of different sensor failures and anomalies present in the data. The volume and complexity of the data precludes the use of manual visualization or simple thresholding techniques to identify these anomalies. We describe the application of probabilistic modeling and unsupervised learning techniques to this data set and illustrate how these approaches can successfully detect underlying systematic patterns even in the presence of substantial noise and missing data.

Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: Statistical Models; I.2.6 [**Artificial Intelligence**]: Learning—*graphical models*

General Terms

probabilistic modeling, traffic model, large-scale analysis, case study

Keywords

loop sensors, MMPP, traffic, Poisson

1. INTRODUCTION

Large-scale sensor instrumentation is now common in a variety of applications including environmental monitoring, industrial automation, surveillance and security. As one example, the California Department of Transportation (Caltrans) maintains an extensive network of over 20,000 inductive loop sensors on California freeways [1, 7]. Every 30 seconds each

---

*contact information: (949) 232-7405

of these traffic sensors reports a count of the number of vehicles that passed over the sensor and the percentage of time the sensor was covered by a vehicle, measurements known as the flow and occupancy respectively. The data are continuously archived, providing a potentially rich source from which to extract information about urban transportation patterns, traffic flow, accidents, and human behavior in general.

Large-scale loop sensor data of this form are well known to transportation researchers, but have resisted systematic analysis due to the significant challenges of dealing with noisy real-world sensor data at this scale. Bickel et al. [1] outline some of the difficulties in a recent survey paper:

> ...loop data are often missing or invalid...a loop detector can fail in various ways even when it reports values...Even under normal conditions, the measurements from loop detectors are noisy...

> Bad and missing samples present problems for any algorithm that uses the data for analysis...we need to detect when data are bad and discard them

> A systematic and principled algorithm [for detecting faulty sensors] is hard to develop mainly due to the size and complexity of the problem. An ideal model needs to work well with thousands of detectors, all with potentially unknown types of malfunction.

> Even constructing a training set is not trivial since there is so much data to examine and it is not always possible to be absolutely sure if the data are correct even after careful visual inspection.

Similar issues arise in many large real-world sensor systems. In particular, the presence of "bad" sensor data is a persistent problem—sensors are often in uncontrolled and relatively hostile environments, subject to a variety of unknown and unpredictable natural and human-induced changes. Research papers on sensor data mining and analysis often pay insufficient attention to these types of issues; for example, our previous work [4, 5] did not address sensor failures directly. However, if research techniques and algorithms for sensor data mining are to be adapted and used for real-world problems it is essential that they can handle the challenges of such data in a robust manner.
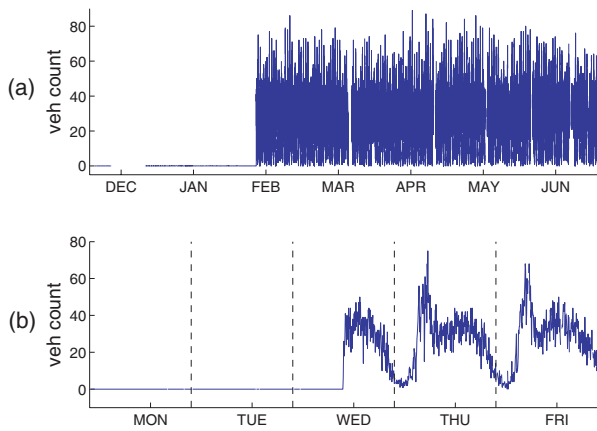
Figure 1: (a) A sensor that is stuck at zero for almost two months. (b) Five days of measurements at the end of the period of sensor failure, after which a typical pattern of low evening activity and higher activity at morning and afternoon rush hour begins to appear.
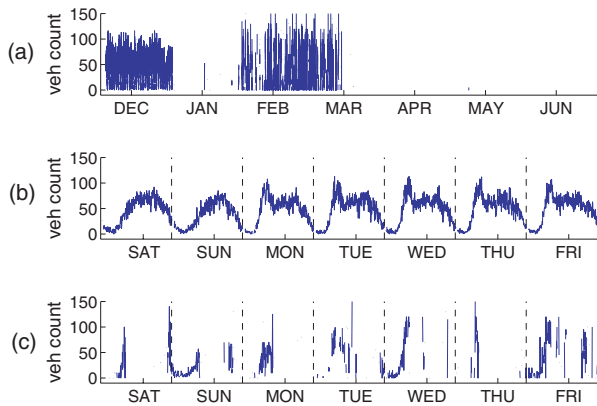
Figure 2: (a) A sensor with normal (periodic) initial behavior, followed by large periods of missing data and suspicious measurements. (b) A week at the beginning of the study showing the periodic behavior typical of traffic. (c) A week in February. Other than the missing data, these values may not appear that unusual. However, they are not consistent with the much clearer pattern seen in the first two months. The presence of unusually large spikes of traffic, particularly late at night, also make these measurements suspicious.

In this paper we present a case study of applying probabilistic sensor modeling algorithms to a data set with 2263 loop sensors involving over 100 million measurements, recorded over seven months in Southern California. The sensor modeling algorithms are based on unsupervised learning techniques that simultaneously learn the regular patterns of human behavior from data as well as the occurrence of unusual events, as described in our previous work [4, 5].

The seven months of time-series data from the 2263 loop sensors contain a wide variety of anomalous behavior including "stuck at zero" failures, missing data, suspiciously high readings, and more. Figure 1 shows a sensor with a "stuck at zero" failure, and Figure 2 shows an example of a sensor with extended periods both of missing data and of suspicious measurements. In this paper we focus specifically on the challenges involved in working with large numbers of sensors having diverse characteristics. Removing bad data via visual inspection is not feasible given the number of sensors and measurements, notwithstanding the fact that it can be non-trivial for a human to visually distinguish good data from bad. In Figure 2, for example, the series of measurements between January and March might plausibly pass for daily traffic variations if we did not know the normal conditions. Figure 2 also illustrates why simple thresholding techniques are generally inadequate, due to the large variety in patterns of anomalous sensor behavior.

We begin by illustrating the results of a probabilistic model that does not include any explicit mechanism for handling sensor failures. As a result, the unsupervised learning algorithms fail to learn a pattern of normal behavior for a large number of sensors. We introduce a relatively simple mechanism into the model to account for sensor failures, resulting in a significant increase in the number of sensors where a true signal can be reliably detected, as well as improved automatic identification of sensors that are so inconsistent as to be unmodelable. The remainder of the paper illustrates how the inferences made by the fault-tolerant model can be used for a variety of analyses, clearly distinguishing (a) the predictable hourly, daily, and weekly rhythms of human behavior, (b) unusual bursts of event traffic activity (for example, due to sporting events or traffic accidents), and (c) sequences of time when the sensor is faulty. We conclude the paper with a discussion of lessons learned from this case study.

2. LOOP SENSOR DATA

We focus on the flow measurements obtained from each loop sensor, defined as the cumulative count of vehicles that passed over the sensor. The flow is reported and reset every 30 seconds, creating a time series of count data. As shown in Figures 1 and 2, the vehicle count data is a combination of a "true" periodic component (e.g., Figure 2(b)) and a variety of different types of failures and noise.

We collected flow measurements between November 26, 2006 and July 7, 2007 for all of the entrance and exit ramps in Los Angeles and Orange County. The data were downloaded via ftp from the PeMS database [1, 7] maintained by U.C. Berkeley in cooperation with Caltrans. Of the 2263 loop sensors, 566 sensors reported missing (no measurement reported) or zero values for the entire duration of the seven month study. The remaining 1716 sensors reported missing measurements 29% of the time on average. Missing data occurred either when PeMS did not report a measurement due to a faulty detector or a faulty collection system, or when our own system was unable to access PeMS.

Aside from missing measurements and sensor failures, the periodic structure in the data reflecting normal (predictable) driving habits of people can be further masked by periods of unusual activity [5]; including those caused by traffic accidents or large events such as concerts and sporting events.
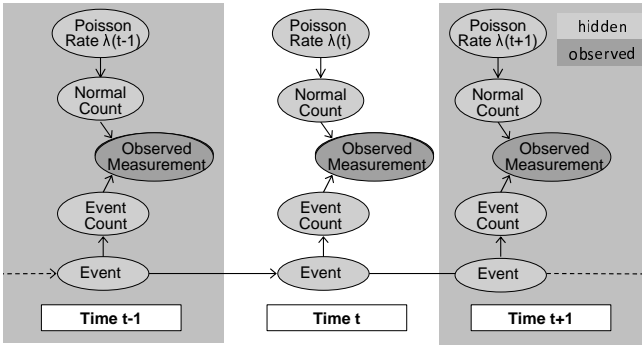
Figure 3: Graphical model of the original approach proposed in [4, 5]. Both the event and rate variables couple the model across time: the Markov event process captures rare, persistent events, while the Poisson rate parameters are linked between similar times (arrows not shown). For example, the rate on a particular Monday during the 3:00 to 3:05pm time slice is linked to all other Mondays at that time.
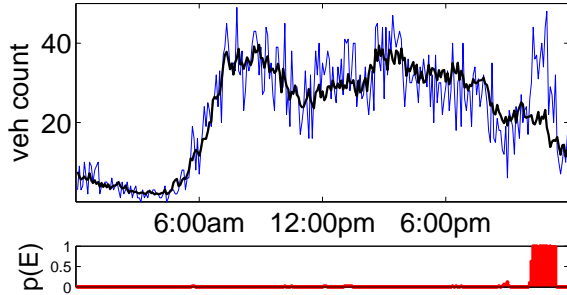
Figure 4: Example inference results with the model from FIgure 3. The blue line shows actual flow measurements for one sensor on one day, while the black line is the model's inferred rate parameters for the normal (predictable) component of the data. The bar plot below shows the estimated probability that an unusual event is taking place.

Noisy measurements, missing observations, unusual event activity, and multiple causes of sensor failure combine to make automated analysis of a large number of sensors quite challenging.

3. ORIGINAL MODEL

In Ihler et al. [5] we presented a general probabilistic model (hereafter referred to as the original model) that learns patterns of human behavior that is hidden in time series of count data. The model was tested on two real-world data sets, and was shown to be significantly more effective than a baseline method at discovering both underlying recurrent patterns of human behavior as well as finding and quantifying periods of unusual activity. Our original model consisting of two components: (a) a time-varying Poisson process that can account for recurrent patterns of behavior, and (b) an additional "bursty" Poisson process, modulated by a Markov process, that accounts for unusual events.

If labeled data are available (i.e. we have prior knowledge of the time periods when unusual events occur), then estimation of this model is straightforward. However, labeled data are difficult to obtain and are likely to be only partially available even in a best-case scenario. Even with close visual inspection it is not always easy to determine whether or not event activity is present. Supervised learning (using labeled data) is even less feasible when applying the model to a group of 1716 sensors.

Instead, the approach we proposed in [4, 5] separates the normal behavior and the unusual event behavior using an unsupervised Markov modulated Poisson process [9, 10]. The graphical model is shown in Figure 3. The normal (predictable) component of the data is modeled using a time-varying Poisson process, and the unusual event activity is modeled separately using a Markov chain. The event variable can be in one of three states: no event, positive event (indicating unusually high activity), or negative event (unusually low activity).

In the model, the Poisson rate parameter defines how the normal, periodic behavior counts are expected to vary, while the Markov chain component allows unusual events to have persistence. If the observed measurement is far from the rate parameter, or if event activity has been predicted in the previous time slice, the probability of an event increases.

Given the model and the observed historical counts, we can infer the unknown parameters of the model (such as the rate parameters of the underlying normal traffic pattern) as well as the values of the hidden states. Note that in addition to the event state variables being connected in time, the rate parameters $\lambda(t)$ are also linked (not shown in Figure 3). This leads to cycles in the graphical model, making exact inference intractable. Fortunately, there are approximation algorithms that are effective in practice. As described in [4, 5], we use a Gibbs sampler [3] for learning the hidden parameters and hidden variables. The algorithm uses standard hidden Markov recursions with a forward inference pass followed by a backwards sampling pass for each iteration of the sampler. The computational complexity of the sampler is linear in the number of time slices, and empirically convergence is quite rapid (see [5] for more details).

Figure 4 shows an example of the results of the inference procedure. The measured vehicle count for this particular day follows the model's inferred time-varying Poisson rate for the normal (predictable) component of the data for most of the day. In the evening, however, the observed measurements deviate significantly. This deviation indicates the presence of unusual event activity and is reflected in the model's estimated event probability (bottom panel). The output of the model also includes information about the magnitude and duration of events.

The event activity in this example looks obvious given the inferred profile of normal behavior; however, simultaneously identifying the normal pattern and unusual event activity hidden within the measurements is non-trivial. In our earlier work [4, 5] we found that the Markov-modulated Poisson process was significantly more accurate at detecting known events than simpler baseline methods such as threshold-type

Event Fraction	Number of Sensors
0 to 10%	912
10 to 20%	386
20 to 50%	265
50 to 100%	153

Table 1: Original model's fit. The study's 1716 sensors are categorized using a measure of the model's ability to find a predictable periodic component in the sensor measurements (if present). The *event fraction* is defined as the fraction of time a sensor's measurements are classified as a positive or negative event. For sensors higher in the table, the model has found a strong periodic component with fewer periods of unusual event activity.
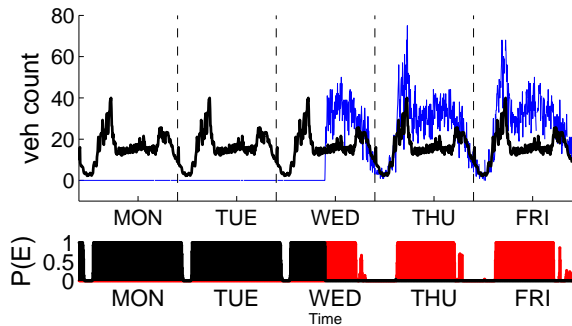
Figure 5: Original model output for the sensor in Figure 1. Shown are the observed measurements (blue) for one week (the same week as Figure 1(b)) along with the model's inferred Poisson rate (black). With a long period stuck at zero, a poor model is inferred for normal behavior in the middle of the day. This is reflected in the event probabilities (bottom), where unusual event activity is predicted for most of each day.

detectors based on Poisson models.

4. SCALE-UP CHALLENGES

After the initial successes described in Section 3, we wanted to test the model on a much larger network of sensors. The model can generally be applied to various types of sensors which record count data, but the loop sensor data set was a particularly appealing choice for our case study. As mentioned earlier, we had access to the measurements of 2263 loop sensors in the Southern California area. We also had additional information about the sensors that could prove useful during analysis, such as geographic location and whether each sensor was on an exit or entrance ramp. In addition, there are many data analysis problems specific to traffic data, including accident detection, dynamic population density estimation, and others. In our work we were motivated by the challenge of extracting useful information from this large data set to provide a basic framework for addressing these questions.

We applied the original model to the data from our seven month study involving 1716 sensors and over 300 million hidden variables. The model was subjected to much greater levels of variability than experienced in our earlier studies. Several weaknesses of the original model were identified as a result.

Table 1 shows one method for judging how well the model fit the data. The table shows the fraction of time that the model inferred unusual event activity for each of the sensors during our seven month study, i.e. the fraction of time slices in which the event variable in Figure 3 was inferred to be in an unusual event state and are thus not explained by the periodic component of the data.

There is reason to be suspicious when the model infers unusual event activity for a large fraction of the time, especially in cases where unusual event activity is more common than normal activity (as in the last row of the table). A review of the sensors where event activity was inferred over 50% of the time revealed some weaknesses of the original model. Some sensors in this category were apparently faulty throughout the study. Another group of sensors recorded non-missing measurements for only a very small fraction of the study, which were not enough to form a good model. However, there were many sensors which appeared to have an underlying periodic behavior pattern that was missed by the original model.

The sensor with the "stuck at zero" failure (Figure 1) is an example of a sensor with a clear periodic pattern that the original model missed. Figure 5 shows the model's attempt to fit the data from this sensor. The model is able to learn early morning and late night behavior, but an inaccurate profile is inferred for normal behavior in the middle of the day. Examples such as this were observed across many other sensors, and in many cases where a poor model was inferred for normal behavior there appeared to be long periods where the sensor was faulty.

We experimented with a number of modifications to the model, including adjusting the priors on the parameters of the Markov process, avoiding poor initialization of the Gibbs sampler which sometimes occurred when extensive periods of failure were present, and dealing with missing data differently. These adjustments improved the performance of the model in some cases. But in many cases (particularly in sensors with extensive periods of sensor failure) inaccurate profiles were still inferred for normal behavior.

5. FAULT-TOLERANT MODEL

It is clear from Section 4 that in order to make the original model more general and robust, sensor failures should be addressed directly instead of bundling them together with unusual event activity. We note that heuristic approaches to sensor fault detection in traffic data have been developed in prior work [2, 6], but these techniques are specific to loop detectors and to certain types of sensor failures. Our focus in this paper is developing an approach that can handle more general types of faults, not only in loop sensor data but also in other sensors that measure count data.

One possible approach to solve these problems would be to modify the model to broaden the definition of "events" to include sensor failures. However, sensor failures and events (as we define them) tend to have quite different characteris-
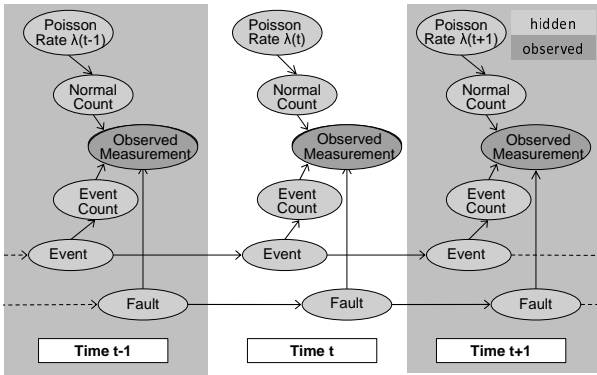
Figure 6: Graphical model of the fault-tolerant method. The fault-tolerant model is very similar to the original model (Figure 3), but with an additional Markov chain added to model sensor failure.

tic signatures. Events tend to persist for a few hours while failures often have a much broader range of temporal duration. Events also tend to be associated with a relatively steady change (positive or negative) in count rates over the duration of the event, while failures can have significant variability in count rates during the duration of the fault. Ultimately, while there is not a crisp boundary between these two types of non-normal measurements, we will show in Sections 6 and 7 that both types are sufficiently different and prominent in the data to merit separate treatment.

When we detect sensor failures visually, we are in essence recognizing extensive periods of time where the periodic structure that we have come to expect is not present. This reasoning is built into a new model, defined by the graphical model in Figure 6. The original model (Figure 3) has been modified to include an additional Markov chain for failure modes. This failure state is a binary variable indicating the presence or absence of a sensor failure.

If the state variable in the event chain is in an event state, the observed measurement is accounted for by both the normal and event count components. If the state variable in the fault chain is in the fault state, however, the observed measurement is treated as if it were missing. This allows our model to ignore the faulty part of the data when inferring the time-varying Poisson (normal) component of the data.

In a Bayesian framework, our belief about the relative duration of events and failures can be encoded into the priors that are put on the transition parameters for the two Markov chains. We expect events to be short, ranging from a few minutes to a couple of hours; in contrast, we expect failures to be relatively lengthy, ranging from days to months.

The modular nature of graphical models makes it possible to extend the original model without starting from scratch. Some modifications were made to the inference calculations when the fault sensor process was added to the original model, but learning and inference proceeds in the same general fashion as before (see [4, 5] for details). From a practical viewpoint, relatively few changes to the software code were needed to extend the model to include a failure state. Details of these changes are given in the Appendix.

Event Fraction	Original Model Number of Sensors	Fault-tolerant Model Number of Sensors
0 to 10%	960	1285
10 to 20%	375	242
20 to 50%	244	117
50 to 100%	137	72

Table 2: Comparison of the fraction of time in the event state for the original and fault-tolerant models. We have excluded missing data and times detected as faulty from the percentage calculation for both models. While there is a slight shift to the upper rows of the table for the original model (compared to Table 1), we see a greater shift for the fault-tolerant model, indicating that it has done a better job of inferring the true periodic structure underlying the data.
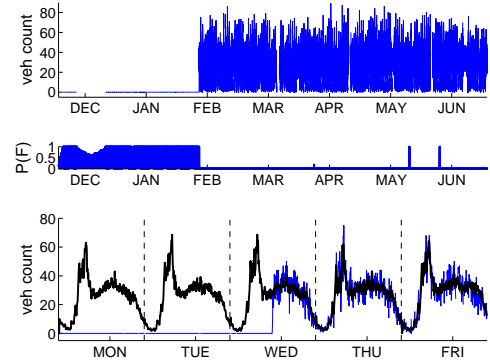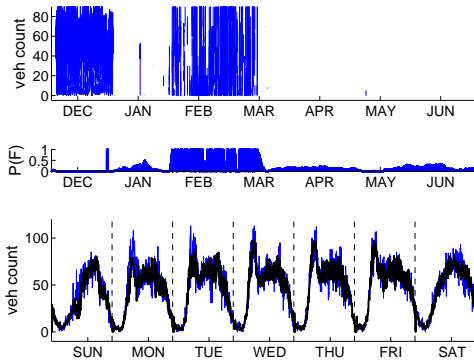

Figure 7: Fault-tolerant model results for the "stuck at zero" example (refer to Figures 1 and 5 and Section 4). Shown are the raw data (top), the model's estimate of the probability of a faulty sensor (center), and the inferred time-varying Poisson rate of the normal component (bottom). The model detects the "stuck at zero" failure, and the model's rate fits the periodic signal in the data that was missed by the original model (Figure 5).

6. EXPERIMENTAL RESULTS

Table 2 gives a sense of the gains made by the fault-tolerant model compared to the original model. We compare the percentage of time spent in an event state for sensors under each model. In order to provide a fair comparison, missing measurements as well as measurements detected as faulty by the fault-tolerant model were removed for *both* models before calculating the event fraction of the remaining measurements predicted by each model. The fault-tolerant model is able to infer a better model of the periodic structure hidden within the flow data, seen as a shift of sensors to the upper rows of the table where unusual activity is detected less frequently and thus more of the data is explained by the time-varying Poisson process. Of the 381 sensors in the >20% range for event fraction with the original model, only 189 remain under the fault-tolerant model, a 50% reduction.

Figure 7 is a plot of the same sensor as in Figures 1 and 5, where the original model was not able to find the normal traffic behavior during the middle of the day. The fault-

Figure 8: Fault-tolerant model results for the corrupted signal example (refer to Figure 2). The corrupted portion of the signal is detected (center), and the model's inferred time-varying Poisson rate (bottom) fits the periodic signal present in the first months of the study.
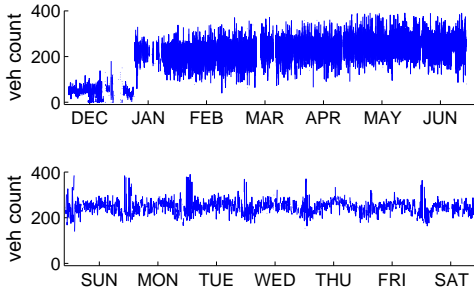
Figure 9: Sensor with a corrupt signal. This sensor appears to be faulty for the entire duration of our study. There is no consistent periodic pattern to the signal, and large spikes often occur in the middle of the night when little traffic is expected.
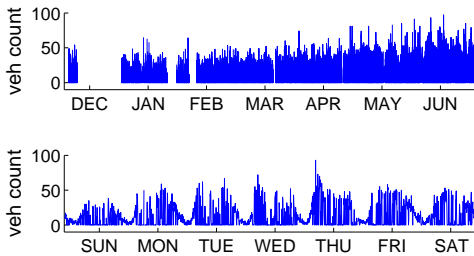
Figure 10: Sensor signal with no consistent periodic component. There may be some periodic structure within a particular week (bottom panel), but there appears to be no consistent week-to-week pattern.

tolerant model was able to detect the "stuck at zero" failure at the beginning of the study and find a much more accurate model of normal behavior.

Figure 8 shows the performance of the fault-tolerant model for a different type of failure. This is the same sensor shown earlier in Figure 2, where the measurements display periodic behavior followed by a signal that appears to be corrupted.

During this questionable period, the measurements are missing more often than not, and unusually large spikes (many 50% higher than the highest vehicle count recorded during the first two months of the study) at unusual times of the day are often observed when the signal returns. The fault-tolerant model can now detect the corrupted signal and also in effect removes the faulty measurements when inferring the time-varying Poisson rate.

In the 50% to 100% row of the table, there are still a number of sensors where the fault-tolerant model is not able to discover a strong periodic pattern. About half of these sensors had large portions of missing data with too few non-missing measurements to form a good model. Others such as seen in Figures 9 and 10, had no consistent periodic structure. Figure 9, is an example of a sensor that appears to be faulty for the duration of our study. The measurements for the sensor in Figure 10, on the other hand, appear to have some structure; morning rush hour with high flow, and low flow in the late evening and early morning as expected. However, the magnitude of the signal seems to alter significantly enough from week to week so that there is no consistent "normal" pattern. Even though the non-zero measurements during the day could perhaps be accurate measurements of flow, the unusual number of measurements of zero flow during the day along with the weekly shifts make the sensor output suspicious.

Before performing our large-scale analysis, we pruned some highly suspicious sensors. With most sensors, the fault-tolerant model makes a decent fit, and can be used to parse the corresponding time-series count data into normal, event, fault, and missing categories, and the results can be used in various analyses. When the model gives a poor fit (Figures 9 and 10 for example), the parsed data can not be trusted, and may cause significant errors in later analysis if included. So, the outputs of such models (and the corresponding sensors) need to be excluded.

We used the information found in Table 2 to prune our sensor list, and limited our evaluation to the 89% of the sensors that predicted less than 20% unusual event activity. The retained sensors sufficiently cover the study area of Los Angeles and Orange County, as seen in Figure 11. Removing sensors with questionable signals visually, without the use of a model, is not feasible. Our model allows us to prune away sensors of which the model can not make any sense in an automated way.

7. LARGE-SCALE ANALYSIS

After pruning the sensor list, 1508 sensor models remain, which together have learned normal, predictable, traffic behavior for approximately 9 million vehicle entrances and exits to and from the freeways of Los Angeles and Orange County. During the seven month study, these models detected over 270,000 events and almost 13,000 periods of sensor failure. Sensors saw unusual event activity approximately once every 30 hours on average, and saw sensor failure once every 26 days on average.

After observing almost 300,000 periods of unusual and faulty activity, the first question we ask is: On what day of the week and at what time of the day is it most common to see
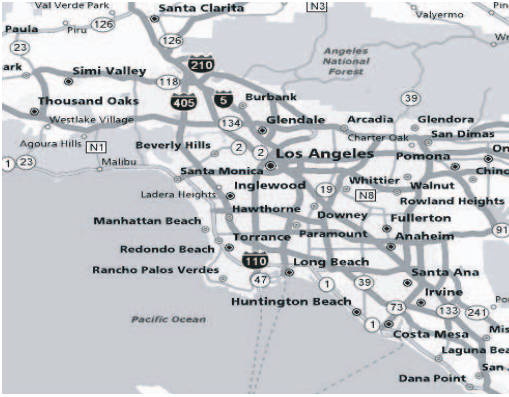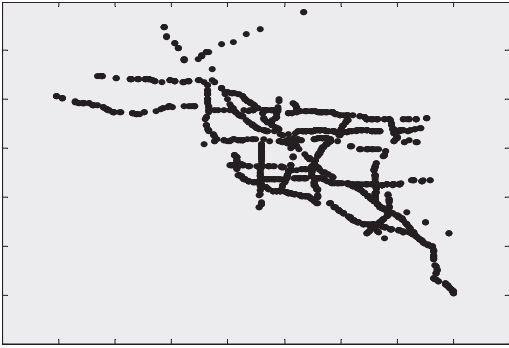
Figure 11: The locations of the sensors used in our large-scale analysis which remain after pruning "suspicious" sensors as described in Section 6 (top), and a road map (bottom) of our study area, Los Angeles and Orange County.
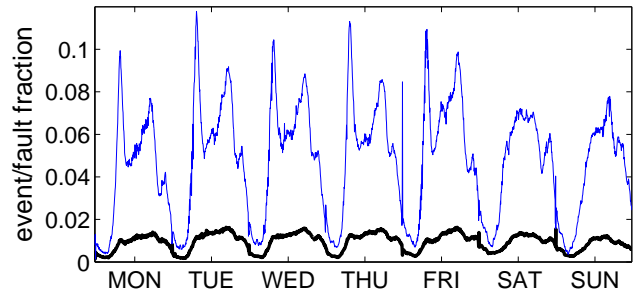

Figure 12: Unusual event frequency and fault frequency. The thin blue line with the greater magnitude shows the fraction of time that events were detected as a function of time of day, while the thick black line shows the fraction of time that faults were detected. Periodic structure is seen in the event frequency.
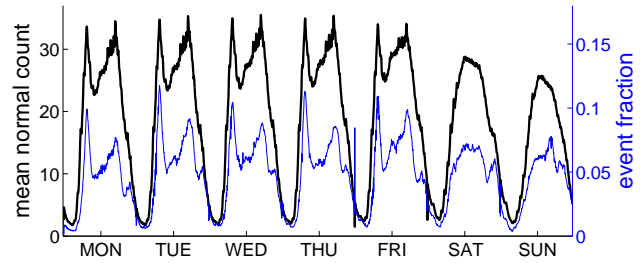

Figure 13: The event fraction (thin blue line, smaller magnitude, right y-axis) is plotted alongside the mean normal vehicle flow profile (i.e., the inferred Poisson rate averaged across the sensors) shown as the thick black line and using the left y-axis. The profiles are similar, with a correlation coefficient of 0.94.

unusual event activity? Figure 12 shows a plot of the frequencies of unusual events and of sensor failures as a function of time of day and day of week. Sensor failures do not appear to have much of a pattern during the day. The troughs at nighttime reflect a limitation of our fault model to detect failures at night when there is little or no traffic. Chen et al. [2] also found it difficult to reliably detect failure events in loop sensor data at night and as a consequence limited fault detection to the time-period between 5am and 10pm.

Of more interest in Figure 12, the frequency of unusual event activity does have a strong pattern that appears proportional to normal traffic patterns. That is, weekdays have spikes in unusual activity that appear to correspond to morning and afternoon rush hour traffic. The event pattern and the normal traffic flow pattern are compared in Figure 13. There is a strong relationship between the two (correlation coefficient 0.94), although there are significant bumps in the event activity each evening, particularly on weekend evenings, that depart from the normal flow pattern.

To explain the shape of the event fraction curve in Figure 13, it is reasonable to consider two types of event activity: events correlated with flow and events independent of flow. Traffic accidents might fall into the correlated event type, because one would expect an accident on the freeway or on an artery close to a ramp to affect traffic patterns more when there is already heavy traffic. Much less of a disruption

is expected if the accident occurs in the middle of the night. Traffic from large sporting events, which often occur in the evening, might fit the second type of event that is not correlated with traffic flow since the extra traffic is not primarily caused by people trying to escape traffic congestion.

Also of note in Figure 13 is that the flow patterns for weekdays look very similar. In Figure 14(a), the inferred time-varying Poisson rate profile for normal activity, averaged across all 1508 sensors, for each week day are plotted on top of each other. This figure shows that the average normal traffic pattern does not vary much between Monday and Friday. Note that in the fault-tolerant model used for the scale up experiments, there is no information-sharing between weekdays, so there is nothing in the model that would influence one weekday to look similar to another. The similarity is not as clear in the raw data (Figure 14(b)).

In Figure 14(a) there is also evidence of bumps occurring at regular intervals, especially in the morning and late afternoon. To investigate if the model was accurately reflecting a true behavior, we plotted the raw flow measurements for
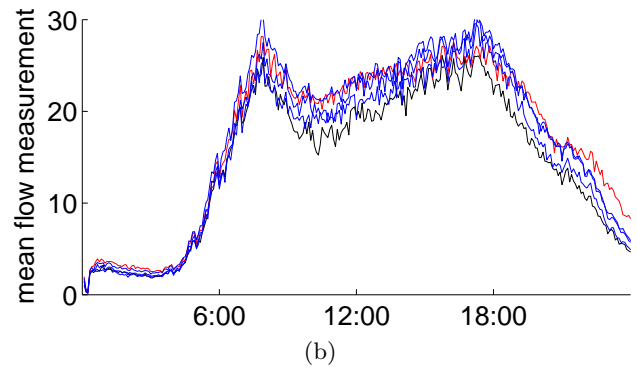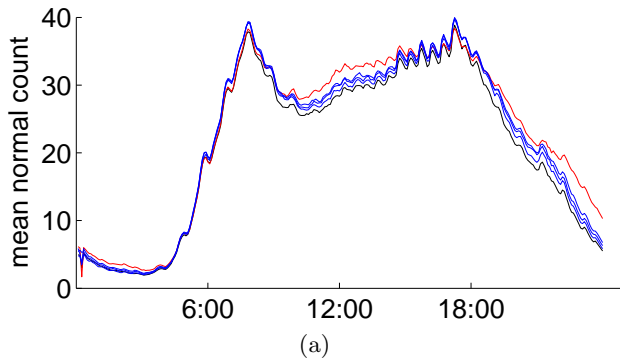
Figure 14: (a) The average Poisson rate (across all sensors) for each weekday, superimposed. Although nothing links different weekdays, their profiles are quite similar, and the oscillation during morning and afternoon rush hour is clearly visible. (b) The mean vehicle flow rates for each weekday (average of raw measurements over all sensors), superimposed. The similarity in patterns is far less clear than in panel (a).
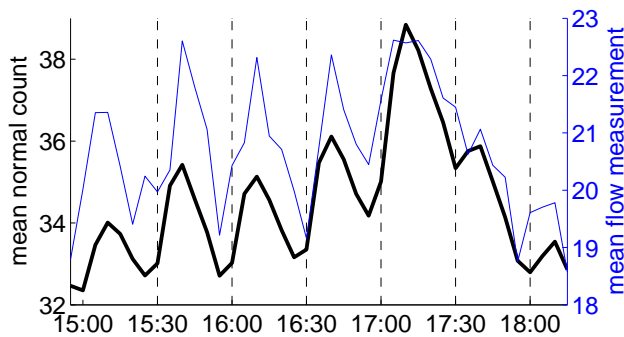

Figure 15: The mean normal vehicle profile (as in Figure 14) shown by the thick black line (using the left y-axis), is plotted against the actual mean flow (light blue line, right y-axis) for Mondays between 3pm and 5pm. The bumps that occur regularly at 30-minute intervals in the model's inferred time-varying Poisson rate are also present in the raw data.
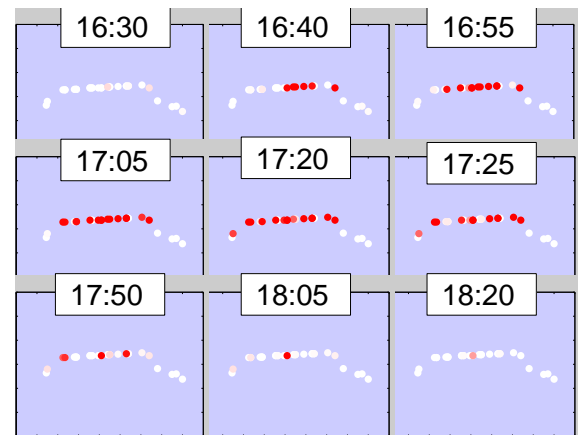

Figure 16: Example of a spatial event that occurs along a stretch of Interstate 10 in Los Angeles. Each circle is a sensor on an exit or entrance ramp. It is light colored when no unusual event activity was inferred by the sensor's model over the past 5 minutes, and is darker as the estimated probability of an unusual event (inferred by the model) increases. The 9 snapshots span a nearly two hour period where unusual activity spreads out spatially then recedes.

each weekday to compare with the model prediction. Figure 15 shows the plot of raw data and model profile for Monday, zoomed in on the afternoon period where the phenomenon is more pronounced. The raw data generally follows the same pattern as the model, confirming that these oscillations are not an artifact of the model. Interestingly, weekend days do not experience this behavior; and when individual ramps were examined, some showed the behavior and some did not. The peaks of the bumps appear regularly at 30 minute intervals. One plausible explanation [8] is that since many businesses are located close to the highway, and people generally report to work and leave work on the half hour and on the hour; the bumps are caused by people getting to work on time and leaving work.

Note that this type of discovery is not easy to make with the raw data. In Figure 14(b), the mean flow profiles the weekdays appear to be potentially different because events and failures corrupt the observed data and mask true patterns of normal behavior. It is not easy to see how similar these daily patterns are, and the half hour bumps in common be-

tween the days (Figure 15) are less likely to be spotted. An important point here is that the model (in Fig 14(a)) has automatically extracted a clear signal of normal behavior, a signal that is buried in the raw data (Fig 14(b)).

Lastly, we present an example of spatial analysis of the model output. Figure 16 shows an example of a "spatial event". The series of plots span a two hour period beginning with a plot of one ramp seeing unusual activity, followed by plots showing a spread of unusual activity detection. At its height, the unusual event activity spans a seven mile stretch of Interstate 10 in Los Angeles, which is followed by a gradual reduction of unusual event activity. One can imagine using information such as this to find the extent of disruption caused by an accident.

8. CONCLUSIONS

We have presented a case study of a large-scale analysis of an urban traffic sensor data set in Southern California. 300 million flow measurements from 1700 loop detectors over a period of seven months were parsed using a probabilistic model into normal activity, unusual event activity, and sensor failure components. The model provides a useful and general framework for systematic analysis of large noisy sensor data sets. In particular, the model was able to provide useful insights about an urban traffic data set that is considered difficult to analyze. Future work could include linking the sensors spatially, and extending the model to detect the spatial and temporal effect of events such as traffic accidents. Other future work could include use of the occupancy values measured by loop sensors in addition to the flow measurements, or making use of census information for dynamic population estimation.

9. REFERENCES

[1] P. Bickel, C. Chen, J. Kwon, J. Rice, E. van Zwet, and P. Varaiya. Measuring traffic. *Statistical Science*, 22(4):581–597, 2007.

[2] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya. Detecting errors and imputing missing data for single-loop surveillance systems. *Transportation Research Record*, 1855(-1):160–167, 2003.

[3] A. Gelman. *Bayesian Data Analysis*. CRC Press, 2004.

[4] A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying Poisson processes. In *ACM Int'l Conf. Knowledge Discovery and Data mining*, pages 207–216, 2006.

[5] A. Ihler, J. Hutchins, and P. Smyth. Learning to detect events with Markov-modulated poisson processes. *TKDD*, 1(3), 2007.

[6] L. N. Jacobson, N. L. Nihan, and J. D. Bender. Detecting erroneous loop detector data in a freeway traffic management system. *Transportation Research Record*, 1287:151–166, 1990.

[7] PeMS. Freeway Performance Measurement System. http://pems.eecs.berkeley.edu/.

[8] W. Recker and J. Marca. Institute of Transportation Studies, UC Irvine, personal communication.

[9] S. Scott. *Bayesian Methods and Extensions for the Two State Markov Modulated Poisson Process*. PhD thesis, Harvard University, 1998.

[10] S. Scott and P. Smyth. The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data. *Bayesian Statistics*, 7:671–680, 2003.

APPENDIX

Using the notation and inference procedure described in our earlier work [5], we explain below the necessary modifications for the fault-tolerant extension of the model.

We use a binary process $f(t)$ to indicate the presence of a failure, i.e., $f(t) = 1$ if there is a sensor failure at time t, and 0 otherwise. We define the probability distribution over $f(t)$ to be Markov in time, with transition probability matrix

$$M_f = \begin{pmatrix} 1 - f_0 & f_0 \\ f_1 & 1 - f_1 \end{pmatrix}.$$

We put Beta distribution priors on f_0 and f_1:

$$f_0 \sim \beta(f; a_0^F, b_0^F) \qquad f_1 \sim \beta(f; a_1^F, b_1^F).$$

In the sampling procedure for the hidden variables given the parameters, the conditional joint distribution of $z(t)$ (the event process) and $f(t)$ is computed using a forward-backward algorithm. In the forward inference pass we compute $p(z(t), f(t)|\{N(t'), t' \le t\})$ using the likelihood functions

$$p(N(t)|z(t), f(t)) =$$
$$\begin{cases} \mathrm{P}(N(t); \lambda(t)) & z(t) = 0, f(t) = 0 \\ \sum_i \mathrm{P}(N(t) - i; \lambda(t))\, \mathrm{NBin}(i) & z(t) = +1, f(t) = 0 \\ \sum_i \mathrm{P}(N(t) + i; \lambda(t))\, \mathrm{NBin}(i) & z(t) = -1, f(t) = 0 \\ \mathrm{U}(N(t); N_{\max}) & \text{otherwise} \end{cases}$$

where $N_{\max}$ is the largest observed flow measurement and $\mathrm{U}(N(t); N_{\max})$ is the uniform distribution over $[0 \dots, N_{\max}]$.

If a failure state is not sampled ($f(t) = 0$), $N_0(t)$ and $N_E(t)$ are sampled as in [5]. However, if a failure state is sampled ($f(t) = 1$), the observed data is treated as missing.

In our previous work [5], $N_0(t)$ and $N_E(t)$ were sampled if the measurement was missing. The fault-tolerant model does not sample $N_0(t)$ and $N_E(t)$ when the data is missing to avoid slow mixing of the Gibbs sampler for sensors with extensive periods of missing data.

By not sampling $N_0(t)$ for missing time slices, the time-varying Poisson rate parameter can no longer be decomposed into day, week, and time-of-day components as in [5]. Instead, a rate parameter is learned for each of the 2016 unique 5-minute time periods of the week. The Poisson rate parameters have prior distributions

$$\lambda_{i,j} \sim \Gamma(\lambda; a_{i,j}^L = 0.05, b_{i,j}^L = 0.01)$$

where i takes on values $\{1, \dots, 7\}$ indicating the day of the week and j indicates the time-of-day interval $\{1, \dots, 288\}$.

We used Dirichlet priors for the rows of the Markov transition matrix for the event process (Z):

$$\begin{pmatrix} a_{00}^Z & a_{01}^Z & a_{02}^Z \\ a_{10}^Z & a_{11}^Z & a_{12}^Z \\ a_{20}^Z & a_{21}^Z & a_{22}^Z \end{pmatrix} = \begin{pmatrix} .999 & .0005 & .0005 \\ .14 & .85 & .01 \\ .14 & .01 & .85 \end{pmatrix} \times 10^6$$

The Beta parameters for the transition matrix of the fault process (F) were:

$$\begin{pmatrix} a_0^F & b_0^F \\ a_1^F & b_0^F \end{pmatrix} = \begin{pmatrix} .00005 & .99995 \\ .0005 & .9995 \end{pmatrix} \times 10^6$$

Strong priors are used for the Markov transition parameters in MMPPs [9] to prevent the model from trying to explain normal sensor noise with the Markov component. The priors above ensure reasonable frequencies and durations for inferred events and faults.