Detecting Changes in Accident Rates using a Hierarchical Bayesian Approach: An Application to the I-710 and the implementation of the PierPASS Program

Ankoor Bhagat, Corresponding Author
Data Scientist
mPulse Mobile
Los Angeles, CA 91436
Phone: 949-331-2374
Email: ankooris@gmail.com

Jean-Daniel Saphores
Professor, Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Irvine
4028 Anteater Instruction and Research Building Irvine, CA 92697-3600
Phone: 949-824-7334
Email: saphores@uci.edu

R. Jayakrishnan
Professor, Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Irvine
4055 Anteater Instruction and Research Building Irvine, CA 92697-3600
Phone: 949-824-8385
Email: rjayakri@uci.edu

Word Count
Abstract: 244
Text: 5,101
6 tables and figures (250 words each): 1,500 words equivalent
Total: 6,845

Submitted to TRB for presentation at the 96th annual meeting on January 8-12, 2017 in Washington, D.C.

ABSTRACT

Road accidents involving heavy duty trucks have long been of concern but detecting the impact on accidents of various policies is typically challenging. The objective of this study is to understand if there was a change in accident rates on a busy freight corridor (the I-710 freeway in Los Angeles County, California) connecting the Ports of Los Angeles and Long Beach to nearby intermodal rail, trans-loading facilities, and warehouses after the implementation of the PierPASS program on July 23, 2005. We analyzed 2,043 accidents that occurred in 2005 on the I-710 freeway; approximately 27.8 percent of these accidents involved trucks. We estimated a three stage hierarchical Bayesian change point model with MCMC developed by Carlin *et al*. (1992) to evaluate whether the implementation of the PierPASS program resulted in a change of accident rates on the I-710. After successfully verifying and validating our model on the dataset used by Carlin *et al*. (1992), and Raftery and Akman (1986), we analyzed road accidents on the I-710 for 2005 filtered by four (approximately 6 mile) segments and four time periods corresponding to different traffic regimes. We generated the probability distribution of the difference between accident rate parameters and built 95% High Density Intervals. Results indicate that there was no significant change in accident rate in 2005 following the implementation of PierPASS. To our knowledge, this is the first time that a three stage hierarchical Bayesian change point model with MCMC was applied to a transportation problem.

KEYWORDS: truck accidents; hierarchical Bayesian analysis; PierPASS program.

INTRODUCTION

Trade is critical to our prosperity, but the transportation of goods especially via drayage trucks generates a number of external costs, including accidents, air pollution, greenhouse gases, noise, vibrations, and uncovered costs associated with the provision, operation and maintenance of public facilities [1, 2]. These external costs are particularly of concern around the Ports of Los Angeles and Long Beach (also known as the San Pedro Bay Ports, or SPBP) in Southern California, which handled ~37% of the container traffic in the United States in 2013 [3]. A quantification of these external costs is important to inform policies related to road infrastructure mitigation, expansion and rehabilitation.

Following widespread concerns about the health impacts of diesel emissions around the SPBP complex [6], measures to address congestion and air pollution from drayage trucks serving the San Pedro Bay Ports (Ports of Los Angeles and Long Beach) were considered by both the Ports and the California Legislature. A number of measures targeted drayage trucks because drayage operations are limited by terminal operating practices, which typically occur on weekdays during normal business hours, whereas ships are typically serviced continuously. These measures include the Clean Trucks Program (which progressively banned older trucks from serving the SPBP complex, established a concession program, and facilitated the replacement of old trucks with low-emission vehicles) and the PierPASS program, which extended gate operations and promoted off-peak deliveries of containers. Although the environmental impacts of these programs has received some attention [5, 7, 10], little is known about the impacts on accidents of the PierPASS program.

Understanding the impact of PierPASS on accidents is critical because accidents represent the largest externality of transportation: in the U.S. the total value of societal harm (economic impacts and valuation for lost quality-of-life) from motor vehicle crashes in 2010 was estimated to be $836 Billion [43]. In the case of PierPASS, it could be argued that spreading out drayage operations could decrease accidents by removing trucks from the I-710 during heavily congested hours. However, this gain could be offset by a higher rate of accidents at night due to decreased visibility and truck driver fatigue. In this context, the purpose of this paper is to analyze accidents on the I-710 in order to understand the net impact on accidents of the PierPASS program. More specifically, we attempt to answer the following two questions:

1. Did the PierPASS program affect the rate of accidents on the I-710?
2. If PierPASS impacted the rate of accidents on the I-710, when did this change occur?

To answer these questions, we estimate a hierarchical Bayesian model for change point detection, following an approach developed by Carlin *et al.* [39]. Although we focus here on accidents linked to drayage operations, this method is widely applicable to detect changes in accidents in other contexts. To our knowledge, this is the first time that a three stage hierarchical Bayesian change point model with MCMC was applied to a transportation problem.

After reviewing selected papers, we introduce our data and our methodology. We then discuss our results before summarizing our conclusions and proposing avenues for future work.

BACKGROUND AND LITERATURE REVIEW

Policies designed to shift freight delivery to off-peak periods have been considered in many urban contexts but they remain somewhat controversial. Several recent studies have highlighted some environmental and health drawbacks of off-peak logistics operations [7, 8, 9, 10].

Off-Peak Logistics Operations

The idea of shifting freight deliveries to off-peak hours has been around for a long time: around 45 B.C., for example, the Roman Emperor Julius Cesar allowed commercial deliveries only during evening hours in order to reduce urban congestion [11]. In recent years shifting logistics operations to nighttime has grabbed the attention of policy makers, because this would reduce freight truck trips during peak hours, thus decreasing traffic congestion, air pollution, and possibly reducing the number of truck accidents. Nighttime off-peak policies are being implemented around the world now to reduce traffic congestion and vehicular emissions [12]. A review of green logistics schemes implemented in various European cities suggests that the main environmental drawback of nighttime delivery is noise pollution from freight vehicles and lifting equipment [13].

In California, off-peak delivery policies were promoted by Assembly Bill (AB) 2650 in 2000, which encouraged port terminals to adopt gate appointments and off-peak operating hours to reduce idling truck queues at gates [14,15]. In Southern California the PierPASS Off-Peak program at the Ports of Los Angeles and Long Beach (also known as the San Pedro Bay Ports, SPBP) was launched on July 23, 2005 to reduce congestion in and around the ports by charging a traffic mitigation fee (TMF) from freight truck movement during peak periods. The traffic mitigation fee charged by the PierPASS off-peak program is $50/TEU (Twenty-foot equivalent unit) and $100 for each container larger than 20 ft [6]. Before the PierPASS program, approximately 17 to 21 percent of port traffic moved during off-peak hours; this percentage jumped to 45 percent after the inception of PierPASS in 2005 [6]. The SPBP established five new shifts per week and a TMF for freight movement during peak hours, which were defined as Monday through Friday from 3 am to 6 pm. No TMF are collected for trucks moving freight during off-peak hours, i.e. after 6 pm [16].

Truck Accidents

Trucks represent a large proportion of vehicles on the I-710 freeway as they connect the SPBP complex to nearby intermodal rail, trans-loading and warehousing facilities. Heavy truck traffic in our study area (see Panel A of Figure 1) generates additional accidents and contributes to air pollution.

In their 2011 paper, Lee *et al.* [17] analyzed 7 years of mid-block crashes in an urban arterial road in Canada using log linear models. They found that mid-block crashes are more likely to occur in road sections with access points (>20%) and high truck percentage (>20%).

Nighttime driving can increase the number of accidents, if visibility is substantially reduced compared to daytime, and if drivers are tired. An older study by Blower *et al.* [18] analyzed accident data for truck operations in Michigan using log-linear techniques to model accident rates of heavy truck-tractors. For single tractors pulling single trailers on rural freeways, they found that the risk of accident at night is 1.4 times greater compared to daytime.

More recently, Golob and Regan [19] analyzed two years of accident data for six Orange County (California) freeways. They calibrated a logistic regression model to estimate truck involvement in crashes using explanatory variables such as average annual daily traffic (AADT), truck traffic as a percentage of AADT, time of day, and weather conditions. They found that a higher percentage of crashes involve trucks between midnight and 3:00 AM.

Fatalities Caused by Truck Accidents

In 2008, according to the Analysis Division of the Federal Motor Carrier Safety Administration (FMCSA), there were 3,733 large truck fatal crashes (2499 crashes between 6 AM and 6 PM and 1,234 crashes between 6 PM and 6 AM) that resulted in 4,229 fatalities in the U.S. [20]. Most of these crashes occurred on weekdays. These statistics show that accidents involving large trucks represent a serious safety problem particularly with respect to fatalities.

In a 2005 paper, Khorashadi *et al.* [21] estimated a multinomial logit model to analyze four years of California accident data. They found that multi-unit trucks and large trucks tend to have more severe accidents than single unit trucks. It confirmed an older study by Chang and Mannering [22] who analyzed accident data from the state of Washington using multinomial logit models to predict the probability of injury severity; that study found that the involvement of a large truck in an accident significantly increases injury severity. More generally, Forkenbrock and Hanley [23] found that conditions like darkness, snow, slush on pavement, the involvement of three or more vehicles, moderate traffic volume, higher speeds, and the involvement of multiple-unit trucks are more likely to result in fatal crashes.

A number of studies have analyzed the monetary cost of accidents involving trucks (see [24] and references therein). These accidents can be very expensive: according to an FMCSA report [24] the average cost of an accident with property damages only is $15,114 while the average cost of a non-fatal injury crash is $195,258. By contrast, the average cost of a crash with one or more fatalities jumps to $3,604,518 (all values in 2005 dollars).

Moreover, accidents occurring during nighttime or during off-peak hours may last longer because incident clearance response units may take longer to respond. Kim *et al.* [25] analyzed Maryland State Highway accident data to identify the variables influencing incident duration. They found that incident duration increases with the number of heavy vehicles involved and that incidents occurring during night time or during off-peak hours last longer than those occurring during day time due to delays in incident clearance operations. In an older study of over 9000 freeway accidents that involved trucks in the Greater Los Angeles, Golob *et al*. [26] found that accidents were more severe and lasted longer when they resulted in overturned trucks. Having separate truck facilities can improve safety; Lord *et al.* suggest separating truck traffic from passenger car traffic improves safety on freeways [27].

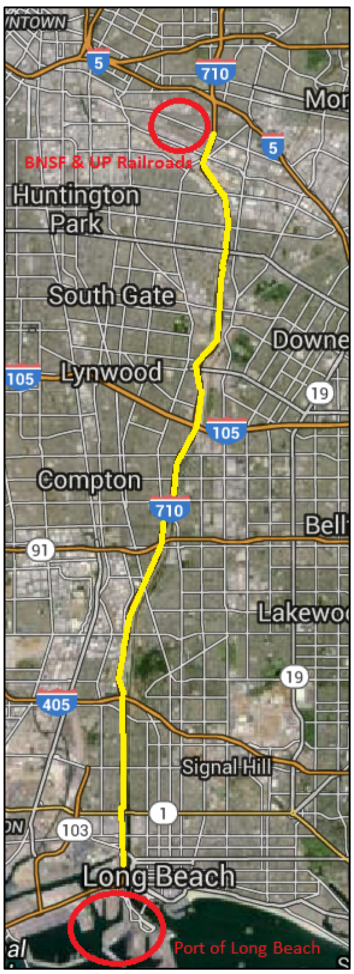
Nighttime Accident Causes

Many nighttime accidents can be attributed to sleepiness. In their study of sleep-related accidents in the U.K., Horne and Reyner [28] found time-of-day effects (i.e., circadian effects) on the timing of accidents with 3 major peak periods associated with driving vulnerability: around 2 AM, 6 AM and 4 PM [28]. In the U.S., Park *et al.* [29] reported that accidents occur primarily at two times of day associated with increased sleepiness – during nighttime (between midnight and 7 AM) and during mid-afternoon (around 3 PM), thus partly confirming the U.K. results of Horne and Reyner [28]. Lenne *et al.* [30] also confirmed that driving performance is affected by time of day and reported a peak in poor driving between 2 AM and 6 AM.

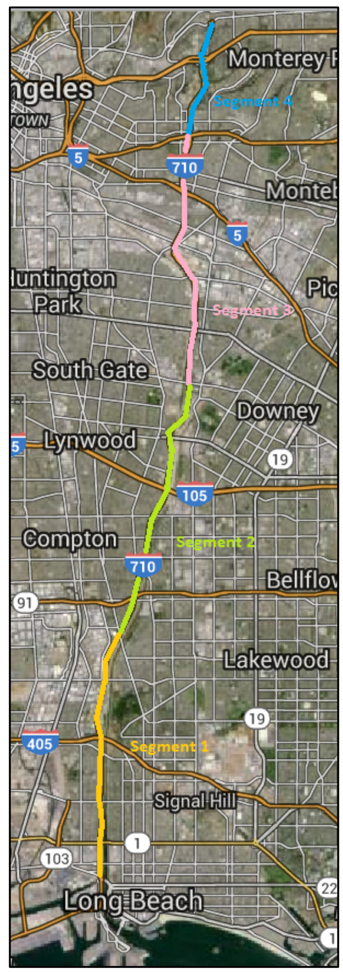
A number of papers have also analyzed the links between truck accidents and geometric design; see for example the two recent papers by Wang et al. [31] and Venkataraman *et al.* [32] and the references therein.

DATA

For this research we obtained accident data for the year 2005 from the Highway Safety Information System (HSIS) database maintained by the University of North Carolina Highway Safety Research Center (HSRC) [33]. California accident data are derived from the California Traffic Accident Surveillance and Analysis System (TASAS) and incorporated in the HSIS system [34]. TASAS has two databases: an accident database and a highway database. For each accident, the accident database has information on location, time, date, severity, primary collision factor, environmental conditions, roadway conditions, type of collision, number of vehicles involved, and information on each party involved. In addition, the highway database contains information on highway segments, intersections, ramps, access control, and AADT [35]. For this research we primarily focused on data from the accident database to estimate change points in the accident rate after the implementation of the PierPASS program.



Panel A Study Area



Panel B Segment Locations

Figure 1. Segment Locations

Our study area (shown in Figure 1) is the I-710 in Southern California. This 24.2 miles freeway connects the San Pedro Bay Ports to nearby intermodal rail and trans-loading facilities and various warehouses. The dataset obtained from HSIS for this study includes the following attributes for each accident that took place in 2005: (1) unique case number; (2) postmile location; (3) date and time; (4) type of accident; (5) number of vehicles involved; (6) involvement of trucks; and (7) accident severity. This database also includes information on weather, road, and ambient light conditions. There were 2,043 accidents reported by the California Highway Patrol (CHP) on the I-710 in 2005, of which 567 (27.8%) involved trucks.

For this analysis, the I-710 freeway was divided into four segments bounded by the following off-ramps:

- Segment-1: West Shoreline Drive to Long Beach Boulevard (postmiles 4.96 to 12);
- Segment-2: Long Beach Boulevard to Firestone Boulevard (postmiles 12 to 18);
- Segment-3: Firestone Boulevard to Pomona Highway (postmiles 18 to 24); and
- Segment-4: Pomona Highway to the end of I-710 (postmiles 24 to 32.69).

Each segment is approximately 6 miles long (see Panel B of Figure 1). The accident data were sorted and filtered by recorded postmiles range for each segment and then from the resulting filtered data accident counts by hour and day of the year were extracted for each segment (i.e. imagine one matrix with 365 rows and 24 columns for each segment). Accidents were further separated between those that occurred before the implementation of PierPASS (July 23, 2005), and those that happened after PierPASS.

An exploratory analysis of the data showed that each day of 2005 saw on average 5.6 accidents on the I-710 of which 1.55 involved trucks. Accidents occurred most often in Segment-3 of our study area. For a detailed exploratory analysis, see Bhagat *et al.* [36].

METHODOLOGY

A conceptually simple approach to find out if the PierPASS program had an impact on accidents on the I-710 is to view accidents as stochastic discrete events and to check if their probability distribution changed following PierPASS. This is the approach (called change point detection) we followed.

This approach dates back to Raftery and Akman [37], who developed a two stage Bayesian model for analyzing a Poisson process with a change point and applied their model to the analysis of British coal-mining disasters data [38]. Carlin *et al.* [39] then developed a three stage hierarchical Bayesian model and tested their model on British coal-mining disasters data [38]. We use a similar approach here.

We rely on a fully Bayesian parametric approach. In the past, the use of Bayesian inference has been impeded by the difficulty of computing marginal posterior distributions of model parameters. Thanks to advances in computer simulation techniques and to more powerful personal computers, marginal posterior densities can now be obtained through Markov Chain Monte Carlo (MCMC) simulations [44].

Hierarchical Bayesian Model

In our proposed framework, accidents are discrete events. Let $Y_1, Y_2, \dots, Y_n$ represent a series of accident counts on the I-710. For our model, we assumed that the number of accidents occurring over a given time period, denoted by Y, follows a Poisson process. The probability mass function of Y is then given by:

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \mathrm{k} = 0, 1, 2, \dots \qquad (1)$$

Following Carlin *et al.* [39], our three stage hierarchical Bayesian model can be formulated as follows. Our model is represented graphically on Figure 2.

Stage-1: Model for accident count data

Let us assume that there was a change point and that the $Y_i's$ are iid. Let us assume the change in accident rate occurs on day τ. Before day τ the accident count follows a Poisson distribution with mean θ. After day τ the accident count follows a Poisson distribution with mean λ.

$$Y_i \sim Poisson(\theta), i = 1, 2, \dots \tau \quad (2)$$
$$Y_i \sim Poisson(\lambda), i = \tau + 1, \dots n \quad (3)$$

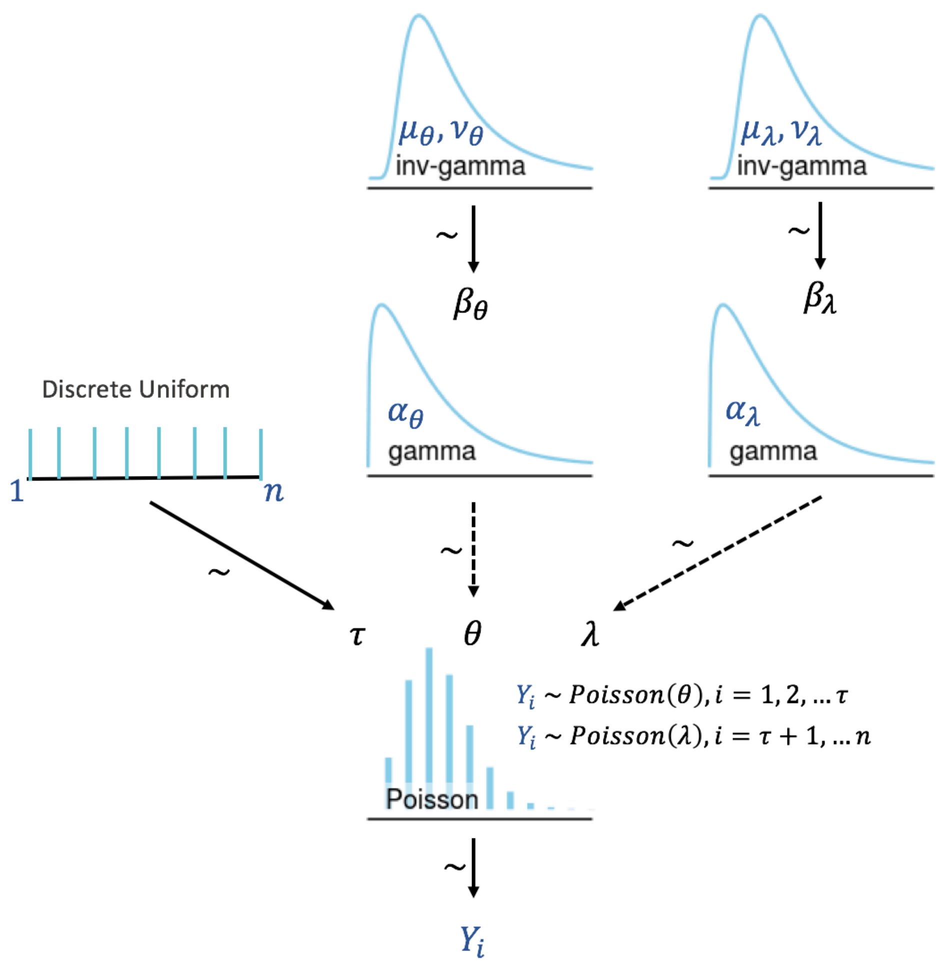


Figure 2. A model of hierarchical dependencies for accident data with change point

Stage-2: Placing independent priors over θ, λ, and τ.

Gamma is a conjugate prior distribution for the parameter of Poisson distribution. The prior distribution for the day of change τ is assumed to be uniformly distributed over the observation period. This suggests that the change point is equally likely to occur on any day during the observation period.

$$\theta \sim Gamma(\alpha_\theta, \beta_\theta) \quad (4)$$
$$\lambda \sim Gamma(\alpha_\lambda, \beta_\lambda) \quad (5)$$
$$\tau \sim DiscreteUniform(1, n), \text{ i.e. } P(\tau = k) = \frac{1}{n}, 1 \le k \le n \quad (6)$$

Stage-3: Placing hyperprior over scale parameters β_θ and β_λ to reduce prior dependency. Also assume β_θ is independent of β_λ.

$$\beta_\theta \sim InverseGamma(\mu_\theta, \nu_\theta) \tag{7}$$

$$\beta_\lambda \sim InverseGamma(\mu_\lambda, \nu_\lambda) \tag{8}$$

Let us assume that $\alpha_\theta, \alpha_\lambda, \mu_\theta, \mu_\lambda, \nu_\theta$ *and* ν_λ are known to us. The unknown variables of interest in the model described above are $\theta, \lambda,$ *and* τ.

Using the hierarchical dependencies (see Figure 2), we can formulate the joint distribution of parameters and data as:

$$P(\theta, \lambda, \tau, \beta_\theta, \beta_\lambda, Y) = P(Y|\theta, \lambda, \tau)P(\tau)P(\theta|\beta_\theta)P(\beta_\theta)P(\lambda|\beta_\lambda)P(\beta_\lambda) \tag{9}$$

Using Bayes theorem, we can find the joint posterior of our parameters given the data from:

$$P(\theta, \lambda, \tau, \beta_\theta, \beta_\lambda|Y) = \frac{P(\theta, \lambda, \tau, \beta_\theta, \beta_\lambda, Y)}{P(Y)} \tag{10}$$

$$P(Y) = \int\int\int\int\int P(Y|\theta, \lambda, \tau)P(\tau)P(\theta|\beta_\theta)P(\beta_\theta)P(\lambda|\beta_\lambda)P(\beta_\lambda)d\theta d\lambda d\tau d\beta_\theta d\beta_\lambda \tag{11}$$

$P(Y)$ is the normalizing constant and its computation involves solving multidimensional integrals. We used Markov Chain Monte Carlo (MCMC) sampling to avoid computing this normalizing constant, so that:

$$P(\theta, \lambda, \tau, \beta_\theta, \beta_\lambda|Y) \propto P(Y|\theta, \lambda, \tau)P(\tau)P(\theta|\beta_\theta)P(\beta_\theta)P(\lambda|\beta_\lambda)P(\beta_\lambda) \tag{12}$$

The likelihood of the unknown parameters given the data, $L(\theta, \lambda, \tau|Y)$, can then be written:

$$L(\theta, \lambda, \tau|Y) \equiv P(Y|\theta, \lambda, \tau) = \prod_{i=1}^{\tau} \frac{\theta^{Y_i} e^{-\theta}}{Y_i!} \prod_{i=\tau+1}^{n} \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!} \tag{13}$$

We also need a way to find priors and hyperpriors. We do this below. Using expressions of the probability mass function of a discrete uniform distribution (for the occurrence of the change point), the probability density functions of the Gamma distribution (for $P(\theta|\beta_\theta)$ and $P(\lambda|\beta_\lambda)$), and of the inverse gamma (for $P(\beta_\theta)$ and $P(\beta_\lambda)$) [44], Equation (12) can be rewritten:

$$\begin{aligned} P(\theta, \lambda, \tau, \beta_\theta, \beta_\lambda|Y) \propto & \prod_{i=1}^{\tau} \frac{\theta^{Y_i} e^{-\theta}}{Y_i!} \prod_{i=\tau+1}^{n} \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!} \times \frac{1}{n} \times \frac{1}{\Gamma(\alpha_\theta)\beta_\theta^{\alpha_\theta}} \theta^{\alpha_\theta - 1} e^{-\theta/\beta_\theta} \\ & \times \frac{\nu_\theta^{\mu_\theta}}{\Gamma(\mu_\theta)} \beta_\theta^{-\mu_\theta - 1} e^{-\nu_\theta/\beta_\theta} \times \frac{1}{\Gamma(\alpha_\lambda)\beta_\lambda^{\alpha_\lambda}} \lambda^{\alpha_\lambda - 1} e^{-\lambda/\beta_\lambda} \times \frac{\nu_\lambda^{\mu_\lambda}}{\Gamma(\mu_\lambda)} \beta_\lambda^{-\mu_\lambda - 1} e^{-\nu_\lambda/\beta_\lambda} \end{aligned} \tag{14}$$

The normalizing constant $P(Y)$ can be computed by integrating Equation (14) which is a formidable analytic problem. We also need to draw samples from the multidimensional posterior distribution $P(\theta, \lambda, \tau, \beta_\theta, \beta_\lambda | Y)$ this is the purpose of using MCMC sampling.

Full conditional distributions of each parameter is obtained by ignoring all terms that are constant with respect to the parameter. Carlin *et al.* [39] derived full conditional distributions for their unknown parameters and used $\alpha_\theta = \alpha_\lambda = 0.5$, $\nu_\theta = \nu_\lambda = 1$, and $\mu_\theta = \mu_\lambda = 0$ to estimate unknown parameters using Gibbs sampling. We use similar values as [39] for $\alpha_\theta, \alpha_\lambda, \mu_\theta, \mu_\lambda, \nu_\theta \; and \; \nu_\lambda$. However $\Gamma(\mu_\theta) = \Gamma(\mu_\lambda) \to \infty$ when $\mu_\theta = \mu_\lambda = 0$, so we are using instead $\mu_\theta = \mu_\lambda = 0.5$ for MCMC sampling.

PyMC3 and MCMC Sampling

PyMC3 (currently a beta software) is an open source Probabilistic Programming framework written in Python. This framework was developed for Bayesian statistical modeling and model fitting that relies on advanced Markov Chain Monte Carlo (MCMC) algorithms [40]. We used PyMC3 to generate MCMC chains for our three stage hierarchical Bayesian model.

In PyMC3, the maximum a posteriori (MAP) estimate for the models unknown parameters is found using numerical optimization methods [40]. MAP estimates are point estimates for the unknown parameters used as initial values during MCMC sampling. To conduct MCMC sampling to generate posterior estimates in PyMC3 we specified three step methods: hyperprior sampling with the No-U-Turn Sampler (NUTS); change point sampling with the Metropolis-Hastings Sampler since it accommodates discrete variables; and rate parameters (θ, λ) sampling using NUTS. NUTS is an extension to Hamiltonian Monte Carlo (HMC) algorithm, it converges faster compared to older MCMC algorithms because it uses the gradient of the log posterior density to identify where regions of higher probability are located [40].

We ran 1,000,000 simulations with a burn-in of 900,000 to discard 90% of initial estimates. Returned samples from MCMC simulation will exhibit autocorrelation partly due to inherent Markovian dependence structure, but thinning can be used to reduce autocorrelation [41]. Thinning of 10 was used to return every 10th sample to reduce autocorrelation. The next section presents the scenarios we considered and our results.

After MCMC sampling we checked if the posterior rate parameters θ and λ are different by generating a distribution of the difference between the posterior rate parameters, i.e. a distribution of $(\lambda - \theta)$ and identified this distribution's credible interval (95% HDI).

A Highest Density Interval (HDI) is a way of summarizing a distribution. The HDI of a posterior distribution indicates which points of this distribution are most credible. The width of the HDI is a way of measuring uncertainty in beliefs. HDI is the set of most probable values of a parameter θ that in total constitute $100 * (1 - \alpha)\%$ of the posterior mass. Mathematically HDI of a distribution is found as [42]: for a given α we find the threshold P^* on the pdf that satisfies:

$$1 - \alpha = \int_{\theta : P(\theta|Y) > P^*} P(\theta|Y) d\theta \tag{20}$$

and then the HDI can be defined as:

$$C_\alpha(Y) = \{\theta : P(\theta|Y) \geq P^*\} \tag{21}$$

Validation

We verified and validated the model by applying it to the British Coal Mining Disaster dataset [38] and comparing the results with the results presented in Raftery and Akman [37], and Carlin et al. [39].

Our results (omitted for brevity) from the analysis of the British Coal Mining dataset are in good agreement with the results presented in Raftery and Akman [37], and Carlin *et al.* [39]: we found a change point around the 40th time period and disaster occurrences decreased after the 40th time period. A look at the posterior distribution of the difference between the rate parameters $\lambda - \theta$ and 95% HDI for their distribution shows that the value 0 does not falls within the 95% HDI, i.e. the value 0 is not among the most credible values of the difference between the parameters, which is negative, so that $\theta > \lambda$. This indicates that there was a significant change in the disaster occurrences after the 40th time period.

Scenarios

After validating our model, we applied it to our I-710 data and ran MCMC simulation for our four I-710 segments and for four time periods (i.e., a total of 16 MCMC sampling). For each segment counts were summed over the following 4 periods to prepare data for the hierarchical Bayesian model:

- Period-1: AM Peak (6 am to 9 am)
- Period-2: Midday (9 am to 3 pm)
- Period-3: PM Peak (3 pm to 7 pm)
- Period-4: Night (7 pm to 6 am)

Each run took approximately 30 minutes on a PC.

RESULTS

Our results are summarized on Figures 3 to 6. These figures show the posterior distribution of difference between accident rate parameters $\lambda - \theta$ for each time period of the four segments on the I-710, and a 95% HDI for each distribution. We can observe that for all time periods the value 0 falls within the 95% HDI, i.e. the value 0 is among the most credible values of the difference between the parameters. This implies that we cannot reject that $\theta = \lambda$ for all four segments time periods, so there was no change in the accident rate after the implementation of the PierPASS program. We note, however, that our conclusions would have been different for the midday and night periods on segments 2 and 3, and also for the morning peak period on segment 2 if we have used a 90% HDI instead of a 95% HDI.

6. CONCLUSIONS AND FUTURE RESEARCH

In this paper a hierarchical Bayesian model for accident change point detection was implemented to understand if there was a change in the accident rate on the I-710 after the implementation of PierPASS program in 2005. The algorithm we implemented was based on the three stage hierarchical Bayesian model developed by Carlin *et al.* [39]. Our model was verified and validated on British Coal Mining Disaster dataset and results were compared with the results obtained by Raftery and Akman [37], and Carlin *et al.* [39]. The model was then applied to 2005 I-710 accident data filtered by four 6 mile segments and four 6-hour time periods. We used PyMC3, a probabilistic programming framework written in Python for Monte Carlo Markov Chain (MCMC) sampling. From the results of MCMC sampling we developed posterior distribution of difference between rate parameters $\lambda - \theta$ (the accident rate parameters before and after the implementation of

PierPASS) for four I-710 segments and for four time periods for each segment. All 95% High Density Intervals include the difference value 0, which suggests that the accident rate parameters are nearly equal implying that there was no change in the accident rate after the implementation of the PierPASS program. Since the model is a count based model, it can be applied to many other areas of transportation. To our knowledge, this is the first time that a three stage hierarchical Bayesian change point model with MCMC was applied to a transportation problem. This approach is of interest because it is powerful, versatile, and it yields results quickly for problems of this size.

In future work, we will investigate if PierPASS resulted in changes in the severity of accidents. As presented in the background and literature section the night time accident risk is much higher than during the day. Therefore, applying the model implemented here to accident count data filtered by severity would be useful to further explore changes in accident rate resulting from PierPASS.

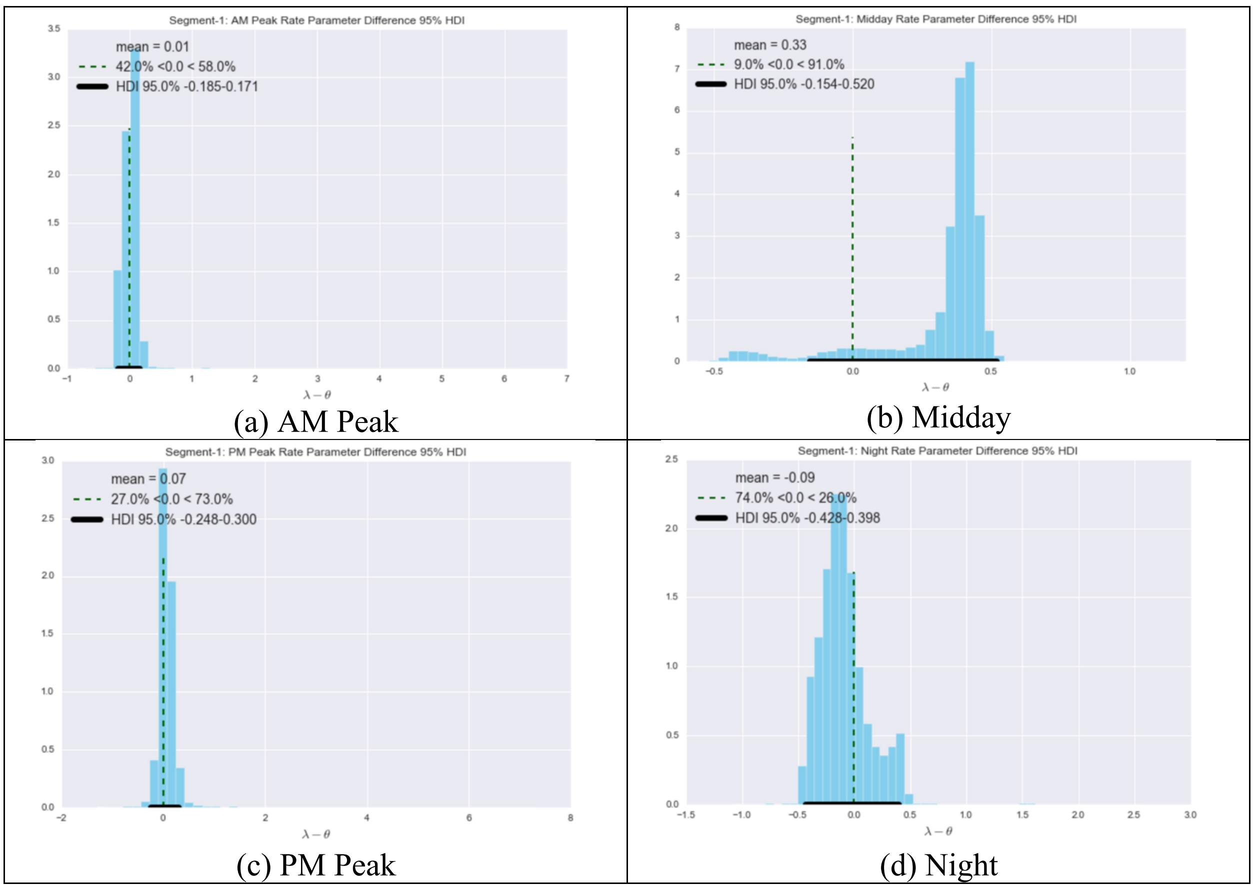


Figure 3. I-710 Segment-1 Accident Rate Parameter Difference 95% HDI

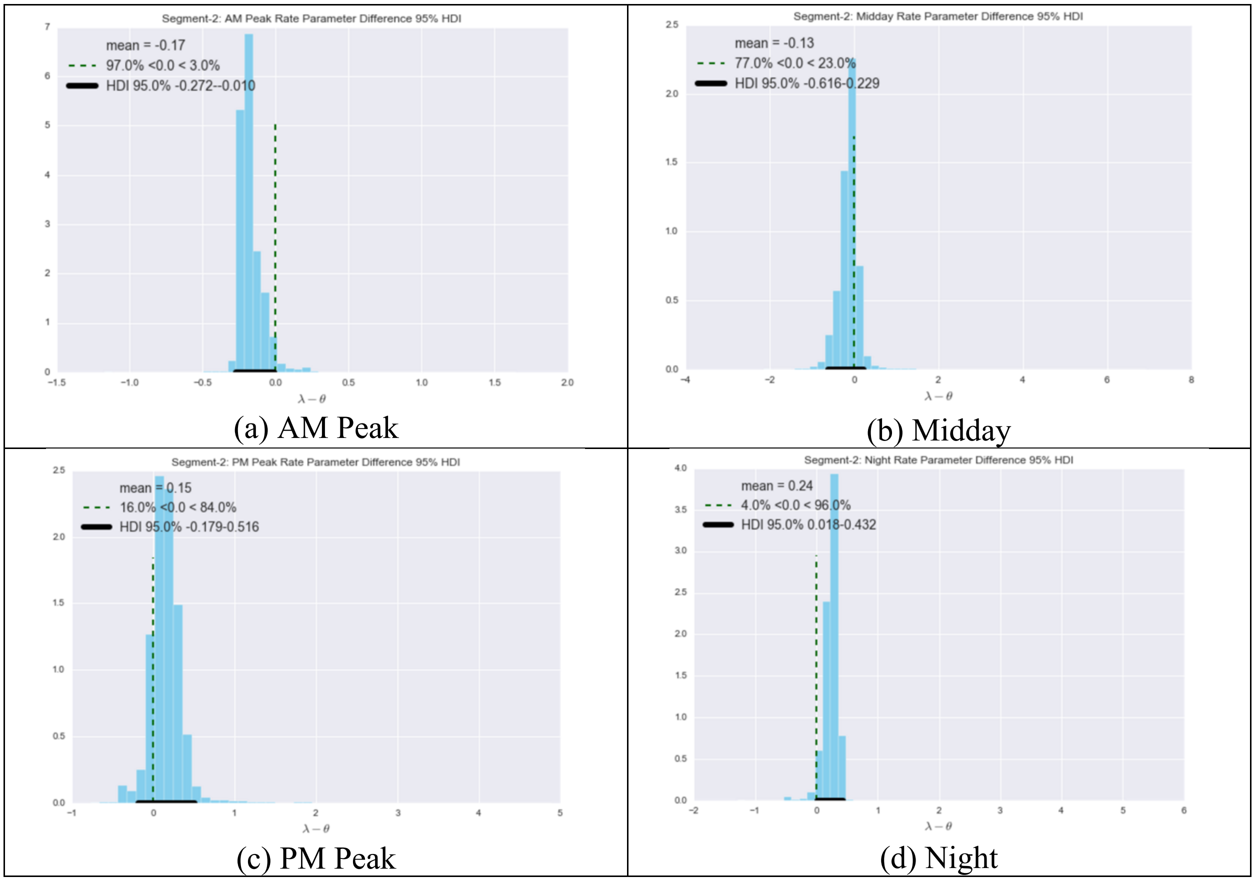
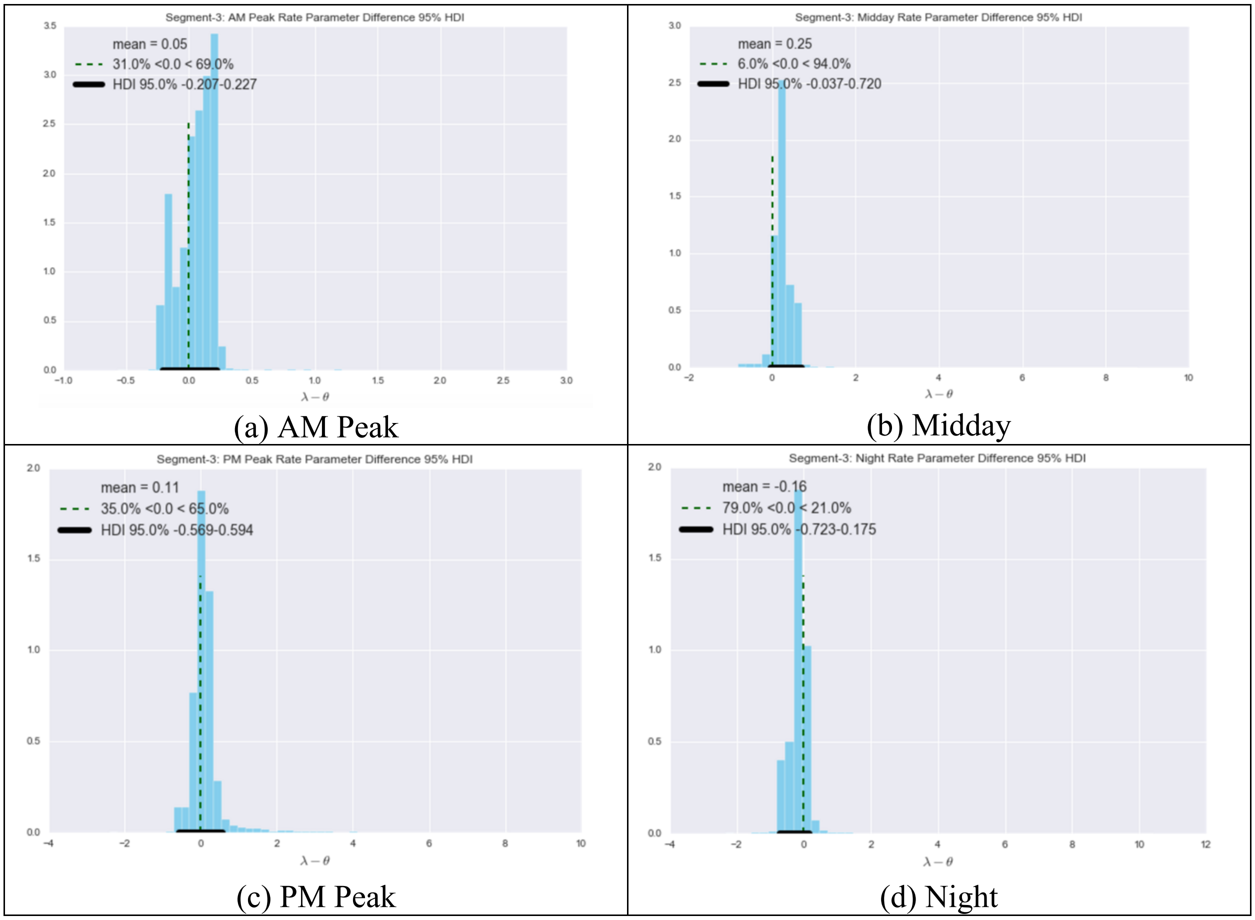


Figure 4. I-710 Segment-2 Accident Rate Parameter Difference 95% HDI



(a) AM Peak (b) Midday

(c) PM Peak (d) Night

Figure 5. I-710 Segment-3 Accident Rate Parameter Difference 95% HDI

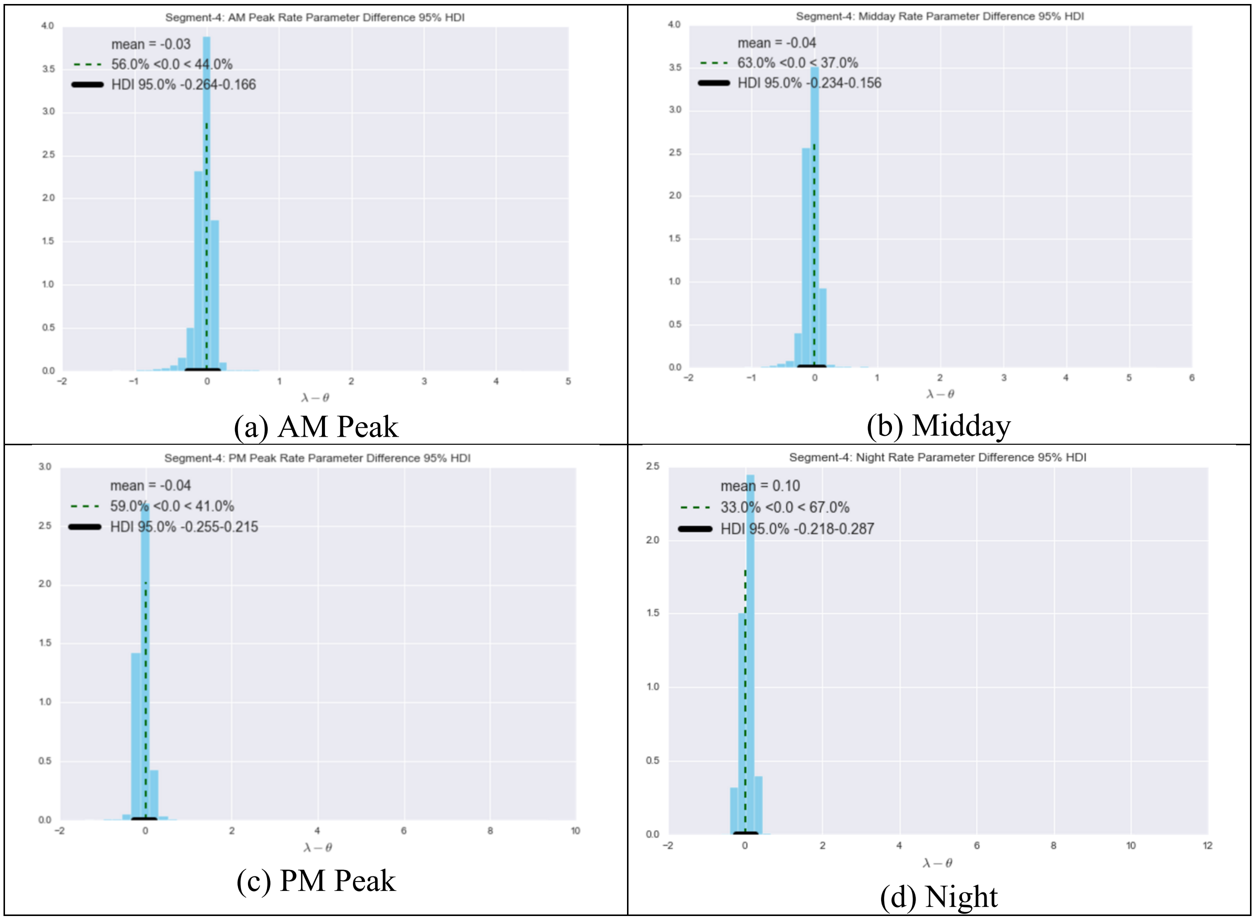


Figure 6. I-710 Segment-4 Accident Rate Parameter Difference 95% HDI

ADDITIONAL INFORMATION

Interactive Data Visualization

Three years (2005-2007) of I-710 accident data can be interactively visualized in the two URLs provided below. D3 JavaScript library was used to create these interactive data visualizations:

1. 2005-2007 I-710 Accidents by Postmile:
   http://ankoorb.github.io/Accidents/index_mile.html
2. 2005-2007 I-710 Accidents by Hour:
   http://ankoorb.github.io/Accidents/index_direction.html

Reproducible Research

The research analysis presented in this paper can be reproduced and verified by reviewers and researchers. Please visit the URL for data and Python codes (Jupyter Notebook):
https://github.com/ankoorb/Bayesian/tree/master/Accident

ACKNOWLEDGEMENTS

We gratefully appreciate and acknowledge the financial support we received from the University of California Transportation Center. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of data presented.

REFERENCES

1. Forkenbrock, D.J. External costs of intercity truck freight transportation. *Transportation Research Part A: Policy and Practice*, Vol. 33, No. 7-8, 1999, pp. 505-526
2. Haling, D. and H. Cohen. Residential Noise Damage Costs Caused by Motor Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1559, No. 1, 1996, pp. 84-93
3. U.S. Department of Transportation, 2014. Freight Facts and Figures 2013 (January). Downloaded on July 30, 2016. Available from http://aapa.files.cms-plus.com/Statistics/Freight%20Facts%20and%20Figures%202013.pdf.
4. Schoch, Deborah, Los Angeles Times Staff Writer (2005). Diesel Fumes From Ports Raising Cancer Risk in Region, Study Says, October 5. Available from http://articles.latimes.com/2005/oct/05/local/me-air4.
5. Lee, G., You, S.I., Ritche S.G., Saphores, J.D., Jayakrishnan, R., Ogunseitan, O. 2012. Assessing air quality and health benefits of Clean Truck Program in the Alameda corridor, CA.
6. PierPASS Review Final Report, July 2008. Prepared by BST Associates. Source: http://pierpass.org/files/offpeak_program/PierPASS_BST_Final_Report_7-9-2008.pdf
7. Sathaye, N., Harley, R., and Madanat, S., 2010. Unintended environmental impacts of nighttime freight logistics activities. Transportation Research Part A (44), pp. 642-659
8. Panis, L.I., and Beckx, C., 2007. Truck driving at night and their effect on local air pollution. Association for European Transport and Contributors, European Transport Conference Proceedings.
9. Hu, S., Fruin, S., Kozawa, K., Mara, S., Paulson, S.E., and Winer, A.M., 2009. A wide area of air pollutant impact downwind of a freeway during pre-sunrise hours. Atmospheric Environment 43, pp. 2541-2549
10. Bhagat, A. (2014). "Expanding Vehicular Microscopic Traffic Simulation for Policy Analysis: An Application to PierPASS in California" *PhD Dissertation*, UC Irvine ITS
11. Holguín-Veras, J. (2008). Necessary conditions for off-hour deliveries and the effectiveness of urban freight road pricing and alternative financial policies in competitive market. Transportation Research Part A 42: 392–413
12. Organization for Economic Co-operation and Development (OECD). 2003. "Delivering the Goods: 21st Century Challenges to Urban Goods Transport."
13. Geroliminis, N. and Daganzo, C., 2005. "A review of green logistics schemes used in cities around the world." U.C. Berkeley Center for Future Urban Transport, A Volvo Center of Excellence.
14. Assembly Bill: 32 (AB:32), Source http://www.arb.ca.gov/cc/ab32/ab32.htm, accessed July 15, 2010.
15. Giuliano, G., and O'Brien, T. (2007). Reducing port-related truck emissions: The terminal gate appointment system at the Ports of Los Angeles and Long Beach. Transportation Research Part D: 460-473.
16. PierPASS (2014). "Off-Peak Information" *About PierPASS.* Source: http://www.pierpass.org/offpeak-information/
17. Lee, C., Xu, X., and Nguyen, V., 2011. Non-intersection related crashes at mid- block in urban divided arterial road with high truck volume. Transportation Research Board, 90th Annual Meeting

18. Blower, D., Campbell, K.L., and Green, P.E., 1993. Accident rates for heavy truck tractors in Michigan. Accident Analysis and Prevention, No. 25, pp. 307-321.
19. Golob, T.F., and Regan, A.C., 2003. Truck involved crashes and traffic levels on urban freeways. Working paper UCI-ITS-TS-WP-03-8.
20. Federal Motor Carrier Safety Administration Analysis Division, 2010. Large truck and bus crash facts 2008. United States Department of Transportation, Washington D.C.
21. Khorashadi, A., Niemeir, D., Shankar, V., and Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis. Accident Analysis and Prevention, No. 37, pp. 910-921.
22. Chang, L.Y., and Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck and non-truck involved accidents. Accident Analysis and Prevention. No. 31, pp. 579-592.
23. Forkenbrock, D.J., and Hanley, P.F., 2003. Fatal crash involvement by multiple- trailer trucks. Transportation Research Part A 37: 419-433.
24. Federal Motor Carrier Safety Administration, 2007. Unit cost of medium and heavy truck crashes. United States Department of Transportation, Washington D.C., Source: http://mcsac.fmcsa.dot.gov/documents/Dec09/UnitCostsTruck%20Crashes2007.pdf
25. Kim, W., Natarajan, S., and Chang, G.L., 2008. Empirical analysis and modeling of freeway incident duration. Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15 2008.
26. Golob, T.F., Recker, W.W., and Leonard, J.D., 1987. An analysis of the severity and incident duration of truck involved freeway accidents. Accident Analysis and Prevention, Vol. 19, No. 5, pp. 375-395.
27. Lord, D., Middleton, D., and Whitacre, J., 2005. Does separating trucks from other traffic improve overall safety? Transportation Research Record 1922: 156- 166.
28. Horne, J.A., and Reyner, L.A., 1995. Sleep related vehicle accidents. BMJ, Volume 310 (6979): 565-7.
29. Pack, A.I., Pack, A.M, Rodgman, E., Cucchiara, A., Dinges, D.F., and Schwab, C.W., 1995. Characteristics of crashes attributed to the driver having fallen asleep. Accident Analysis and Prevention, Vol. 27, No. 6, pp. 769-775.
30. Lenne, M., Triggs, T.J., and Redman, J.R., 1997. Time of day variations in driving performance. Accident Analysis and Prevention, Vol. 29, No. 4, pp. 431-437.
31. Wang, H., K. Jang, C.-Y. Chan, and Z. Shi, 2011. Multi-Level Mixed-Effects Regression Model for Truck Collisions. Paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington D.C.
32. Venkataraman, N.S., G.F. Ulfarsson, V, Shankar, J. Oh, and M. Park, 2011. Modeling the Relationship Between Interstate Crash Occurrence and Geometrics: Exploratory Insights from A Random Parameters Negative Binomial Approach. Paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington D.C.
33. Highway Safety Information System (HSIS). http://www.hsisinfo.org/ (Accessed July 24, 2016)
34. HSIS California Guidebook. http://www.hsisinfo.org/guidebooks/california.cfm#_Toc175365552 (Accessed July 24, 2016)
35. Caltrans Traffic Manual, 1996. Chapter 3: Accident and roadway records. Source: http://www.dot.ca.gov/hq/traffops/signtech/signdel/pdf/TMChapter3.pdf

36. Bhagat, A., Saphores, J.D. and Jayakrishnan, R., 2012. An Analysis of the Timing of Truck and Car Accidents in a busy Freight Corridor. Proceedings of TRB's 91st Annual Meeting. CD-ROM.
37. Raftery, A. and Akman, V. (1986). "Bayesian analysis of a Poisson process with a change-point." *Biometrika*, 73(1), 85-89.
38. Jarrett, R.G. (1979). "A note on the intervals between coal-mining disasters." *Biometrika*, 66(1), 191-193.
39. Carlin, B.P., A.E. Gelfand, A.F.M. Smith (1992). Hierarchical Bayesian Analysis of Changepoint Problems. *Applied Statistics*, Vol. 41, No. 2, pp. 389-405
40. Salvatier J., T.V. Wiecki, C. Fonnesbeck (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* 2:e55 https://doi.org/10.7717/peerj-cs.55
41. Kruschke, J.K., *Doing Bayesian Data Analysis: A Tutorial with R and Bugs*. Academic Press. 2010.
42. Murphy, K.P., *Machine Learning: A Probabilistic Perspective.* The MIT Press, 2012.
43. Blincoe, L, J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. 2015 May. *The economic and societal impact of motor vehicle crashes, 2010. (Revised)*. (Report No. DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration.
44. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. *Bayesian Data Analysis*. CRC Press, 2013.