Development of Duration Models to Determine Rolling Stock Fleet Size

Christopher R. Cherry, University of California, Berkeley

Abstract

Transit agencies are looking for analytical tools to improve their rolling stock maintenance programs and fleet size. If agencies underestimate the number of buses out of service for maintenance at any given time, they cannot adequately cover all scheduled service. If agencies overestimate the number of buses requiring maintenance, they overcapitalize the rolling stock fleet, resulting in an inefficient allocation of scarce funds. This article presents a methodology for using a fully parametric duration model to determine the expected amount of time a vehicle is in service and out of service. Additionally, the transition probabilities are calculated using the hazard data to determine the probability of transitioning from an in-service state to an out-of-service state at a given point in time. An empirical analysis is conducted using data from the San Francisco Municipal Transit Agency and duration models used to identify the transition probabilities from in-service to out-of-service states and the optimum fleet size.

Introduction

Transit operators have a responsibility to ensure full service throughout their systems at all scheduled times. Many times, service is not fully covered for a variety of reasons, including dispatch error, inefficient route planning, driver absenteeism,

and vehicles being out of service. This article focuses on the latter. Controlling and predicting the number of vehicles out of service enables transit operators to effectively cover the entire service area. The time between required maintenance operations and the time it takes to service the bus are stochastic processes. That is, breakdowns and service requirements occur at nonfixed intervals and the time spent out of service varies depending on the type of service and the capacity of the maintenance facility. This article attempts to understand these processes and to make accurate predictions of the number of additional vehicles needed to provide full route coverage as well as have vehicles out of service for maintenance. A duration model is calculated to find the expected time out of service, and from those results, a proportion of extra vehicles is calculated. The parameters are then used to predict the transition probabilities from an operational state to an in-operational state. This article examines literature relevant to modeling approaches for determining transit rolling stock deterioration and maintenance. It also discusses the methodology used to develop the model. An application is presented using San Francisco Municipal Transit Agency bus maintenance data. Finally, conclusions and recommendations are provided.

Literature Review

This article focuses on the development of a statistical model to evaluate transit fleet size to ensure adequate route and schedule coverage as well as include those vehicles that are out of service for maintenance. Specifically, any maintenance or repair that a bus must encounter requires the bus to be out of service, and a replacement bus must take its place to ensure continuous service to customers. Few models have been developed to predict failure and time out of service, and, ultimately, the amount of extra vehicles needed to cover the schedule. There are systems from which comparisons and concepts can be drawn. Namely, absenteeism among transit operators is a similar problem for which transit operators must hire extraboard operators to cover the schedule in the event of unscheduled operator absences. While there are similarities, direct translation of the concepts of extraboard scheduling cannot be transferred to the extra fleet vehicle problem.

A few models have been developed to predict the condition of transit rolling stock or optimize a maintenance schedule. Dutta et al. (1986) developed early simulation models to evaluate scheduling enhancements on the performance of the maintenance system. Haghani et al. (2002) developed a mathematical programming approach to optimize the scheduling of maintenance operations,

given a daily demand for maintenance activities. Schiavone (1997) and Lederer et al. (1990) investigate maintenance performance indicators often used by transit operators. They found that maintenance cost per mile is one of the highest ranking performance indicators among transit agencies. Obeng (1988) regressed maintenance cost per mile on different causal variables. Obeng found that fleet size most significantly increases maintenance cost per mile. These analyses show that minimizing fleet size is essential to decrease maintenance cost per mile, and thus increase the performance of the maintenance system.

A few authors have investigated vehicle reliability. Puntis et al. (1986) qualitatively identified and developed reliability techniques in an attempt to balance the amount of maintenance required and the time to failure of rail cars. The authors identify the need to accurately model reliability to determine how many spare fleet vehicles are needed. The authors specify that reliability is dependent on the duty cycle, maintenance regime, and climatic conditions. Rolling stock condition is one of the primary precursors to the amount of time a bus is receiving scheduled maintenance or is broken down. Karlaftis and Sinha (1997) and Karlaftis (1997) developed ordered probit models that predict the deterioration of a bus using ordinal condition ratings. The authors found that the factors affecting the rate of bus deterioration were age, mileage, capacity, engine type, and climate.

Fukuoka (1987) noted that one of the difficulties when dealing with reliability data is the left truncated and right censored data. Datasets often are missing data for the beginning of the life of the bus, which occurs before the study period. Additionally, the datasets do not record the failure after the study period. This difficulty must be addressed when using reliability data in order to calculate unbiased estimators. The author went on to develop a maximum likelihood estimation model and a nonparametric hazard model to predict failure of rolling stock, both addressing the issues of censoring and truncation.

A parallel system is that of operator absenteeism and extraboard scheduling. Historically, transit agencies have hired extraboard operators based on historical absentee rates (Perry 1983) and dispatcher error rates (Perry and Long 1984). However, this is generally not an optimum solution. MacDorman (1982, 1985) found that transit agencies overstaff their extraboard personnel by more than 7 percent of the optimum assignment. He developed models to balance extraboard and overtime personnel to minimize the cost to the transit agency. Hickman et al. (1988) and Koutsopoulos (1990) later developed strategic and tactical/operational models based on this previous work that could be applied by transit agen-

cies. While there are several similarities that can be drawn from the comparison of extraboard personnel and extra fleet vehicles, there are also several differences that make this comparison fall short. It is clear that historical data on absenteeism and dispatcher error are inaccurate and relying on these measures results in overstaffing. This could be true and much more costly in the event of overcapitalizing a rolling stock fleet, resulting in vehicles being underutilized. Development of prediction models has the potential to optimize this solution. A main difference is labor rules, which is a big driver of open work scheduling and can make extraboard scheduling problematic. Fortunately, rolling stock is not limited by labor rules and can be utilized to the maximum extent possible.

Methodology

To estimate fleet size, a model must be developed to evaluate the number of vehicles out of service at any given time. Managers can then use these results to plan for extra vehicles to provide full service to the system. To this end, a hazard-based duration model (Washington et al. 2003) can be developed to predict the failure of a vehicle and the time spent out of service. Using this duration model, a transit operator can predict the time to failure of an individual bus and aggregate the predictions over the entire fleet to determine the total number of buses out of service at a given time.

The hazard function is the rate at which the failures occur. A fully parametric duration model is used for this analysis. The time between maintenance operations, while the bus is in service intuitively, has a monotonically increasing hazard function. That is, as the length of time from the last failure increases, the probability of failing in the next time step increases. To test this assumption, one can observe the nonparametric hazard function. The nonparametric solution assumes nothing about the functional form of the hazard function. Plotting the nonparametric hazard function of time until maintenance is required confirms the assumption of a monotonically increasing hazard function for the time required until a maintenance operation is performed. Likewise, one might expect a negative duration dependence for time spent out of service, while a maintenance operation is being performed, since many of the longer durations are out of service because of unavailability of parts or qualified labor. The longer the bus is out of service, the lower the likelihood of coming back into service in the next time step. This assumption is confirmed by a visual inspection of the increasing nonparametric

hazard function of time taken to come back into service, given the bus has been out of service for a certain period of time.

A Weibull probability distribution function produces a closed form estimation of the hazard function that accounts for increasing, decreasing, or constant duration dependence. By visual inspection, the Weibull distribution most closely matches the nonparametric solutions of both the interval of time between maintenance operations and the duration of time taken to perform the maintenance operation. The form of this equation is:

$$\lambda(t) = p\lambda^p t^{p-1} \tag{1}$$

where:

λ and p are estimated parameters.

The parameter λ is replaced by $e^{-\beta X}$ for a model with covariate data; however, this model will not include any covariates and thus all unobserved causal variables are included in the λ term. The parameter p is the factor that determines the shape of the PDF. When $p=1$, there is no duration dependence and the hazard rate is constant and follows an exponential distribution. This indicates a Markovian process whereby the probability of changing states is not dependent on the amount of time in that state. When $p>1$, there is positive duration dependence, indicating that the probability of changing states increases as more time is spent in that state. When $0<p<1$, there is negative duration dependence, indicating that the probability of changing states decreases as more time is spent in that state.

Once the parameters are estimated, the expected time until failure can be calculated. The expected duration of the Weibull duration can be calculated by the following equation:

$$E[T] = \lambda^{-p} \tag{2}$$

The survival function is the probability of the time of failure being greater than the time in question. This is important as it indicates the percentage of observations that have a longer duration than the time that a vehicle has already been in service. The functional form of the survival function for a Weibull distribution is:

$$S(t) = P(T>t) = 1-F(t) = \exp[-(\lambda t)^p] \tag{3}$$

The condition of the buses can be identified in one of two states, in service or out of service. A useful tool for a transit operator is to understand the transition probabilities of moving from an in-service state to an out-of-service state within

a given time step. Mishalani et al. (2002) developed a methodology to calculate a discrete time, state-based model for infrastructure deterioration. The transition probability equation is:

$$R\,(t,\,\Delta) = \frac{F(t + \Delta) - F(t)}{S(t)} \tag{4}$$

where:

$F(t)$	equals the cumulative distribution function of the duration random variable T;
$S(t)=1-F(t)$	is the survival function; and
Δ	is the discrete time period.

When utilizing the hazard rate function, the probability of transitioning from the state 1 (in service) to state 0 (out of service) equation becomes:

$$P_{1,0} = R\,(t,\,\Delta) = 1 - \frac{\exp[-\lambda^p(t + \Delta)^{\,p}]}{\exp[-(\lambda t)^{\,p}]} \tag{5}$$

Once the transition probabilities are calculated for the change of state from in service to out of service are calculated, the expected value of the repair time must be calculated. Once the transition probabilities and the expected time out of service have been calculated, the expected number of vehicles out of service at any particular time is calculated. This has very important implications for transit managers as it aids in scheduled maintenance decisions as well as in acquisition of new rolling stock.

Empirical Analysis

Data

The data used in this analysis were supplied by the San Francisco Municipal Transit Agency. The data include maintenance records for two fleets, the diesel engine fleet and the electric engine fleet. Both of these datasets are assumed to have different characteristics between them, but they are assumed to be homogenous

within themselves, even though the buses most certainly have different characteristics (age, routes served, mileage, size, manufacturer, etc.). Unfortunately, these data are not available and cannot be included in the model. The data include the vehicle identification number, date a work order started, date the work order ended, maintenance category, and the specific maintenance operation performed. The discrete time-step interval is one day, since that is the precision of the data. Buses are considered out of service between the time the work order starts and finishes. Buses are considered available for service during the time between the end of a work order and the beginning of the following work order. There are three categories of service: road calls, corrective maintenance, and preventative maintenance. This dataset includes not only corrective maintenance, but also scheduled maintenance, which occurs every 1,000 miles for each bus. Since this model is concerned only with vehicles in service or out of service, it considers all of these types of maintenance operations as failures, even though they have very different characteristics from an operational perspective.

The original dataset consisted of 528 electric buses receiving more than 22,000 maintenance operations between September 12, 2003, and April 30, 2004, as well as 555 diesel engine buses receiving more than 28,000 maintenance operations between August 26, 2003, and April 30, 2004. This totals a fleet of 1,083 buses with more than 50,000 maintenance operations over this nine-month period. In an attempt to remove some of the heterogeneity within the fleets, all buses that were not in full-time service were removed and many of them were retired during the observation interval, according to the maintenance manager of the San Francisco Municipal Transit Agency. Particularly, many of the buses in a certain range of vehicle identification numbers were removed. All obvious data entry errors (e.g., end date is before start date) were removed. Additionally, many of the maintenance operations occurred within the span of a previous operation. That is, they conducted many separate work orders at once or consecutively. Since this analysis is only concerned with the time in and out of service, all observations that were completed at once are modeled as one observation. After all of these adjustments, the final dataset consisted of 3,200 electric bus operations and 7,100 diesel bus operations.

A common issue associated with duration data is that many of the observations can be either right or left censored. Some of the observations used in this analysis are right censored. The data consists of a series of complete observations, with work order start and end dates for each vehicle. The end of the study period was

April 30, so all buses have one censored observation on their last interval in or out of service. All observations start at a different date for each bus, so there is no left censoring. This censoring was accounted for in the statistical model.

Survival and Hazard Model Estimation

The data for each type of bus were modeled separately using LIMDEP econometric software to identify if there are differences between the diesel fleet and the electric fleet. A model was developed for the *interval* of time that the bus was available for service and a separate model was developed for the *duration* of time the bus was out of service. Each of these models is a fully parametric model with a Weibull distribution. Table 1 shows the parameter estimations for p and λ.

Table 1. Duration Model Parameter Estimates

		Interval	*Duration*
Diesel Fleet	**p**	1.22220	0.74144
	p 95% CI	1.2032-1.2412	0.7311-0.7518
	λ	0.10287	0.38997
	λ 95% CI	0.1005-0.1053	0.3713-0.4087
Electric Fleet	**p**	1.11141	0.55970
	p 95% CI	1.0846-1.1382	0.5267-0.5927
	λ	0.09831	0.28094
	λ 95% CI	0.0942-0.1024	0.2330-0.3280

From these parameters one can see that the two populations are different. The λ and p parameters are significantly different from each other, as displayed by them being outside of the other's confidence intervals. For instance, the diesel fleet's $p=1.2220$ is above the confidence interval of the electric fleet's p value. The p values are also significantly different than 1, which indicates that the hazard function is monotonically increasing for the *interval* of time between breakdowns; that is, the longer the vehicle has been in service, the more likely it is to break down. The p value less than one indicates that the hazard function is monotonically decreasing for the *duration* of time being repaired; that is, the longer the bus is in the garage, the likelihood of leaving the garage decreases. This can be expected as longer, more complicated maintenance operations require parts and labor that are not always

available and thus require longer times out of service. These p values also indicate that the distribution is not exponential and the Markovian property of duration independence does not hold for any of the parameters.

These relationships are most easily displayed in the hazard functions for each of the fleets (Figures 1 and 2). From Figure 1, it can be seen that the hazard functions are increasing for the in-service *intervals* of time. From these graphs the instantaneous hazard rate can be determined given that a bus has survived until a certain time. For instance, given that an electric diesel bus has survived 20 days, the probability of failure is 0.15. The graph also indicates that the hazard rate is higher for the diesel fleet than the electric fleet in most cases. Similarly, Figure 2 shows the hazard functions of the time taken to perform a maintenance operation for the electric and diesel fleets. These functions are decreasing, illustrating decreasing duration dependence. The figure also illustrates that the diesel fleet has a higher "hazard" of coming back into service compared to the electric fleet, meaning that it has a higher probability of ending its *duration* of time out of service, given it has been out of service for a period of time.

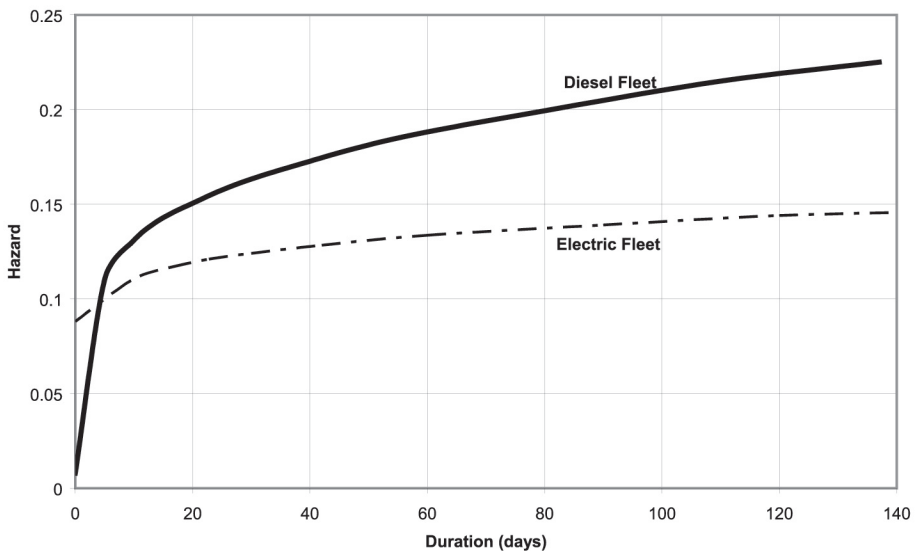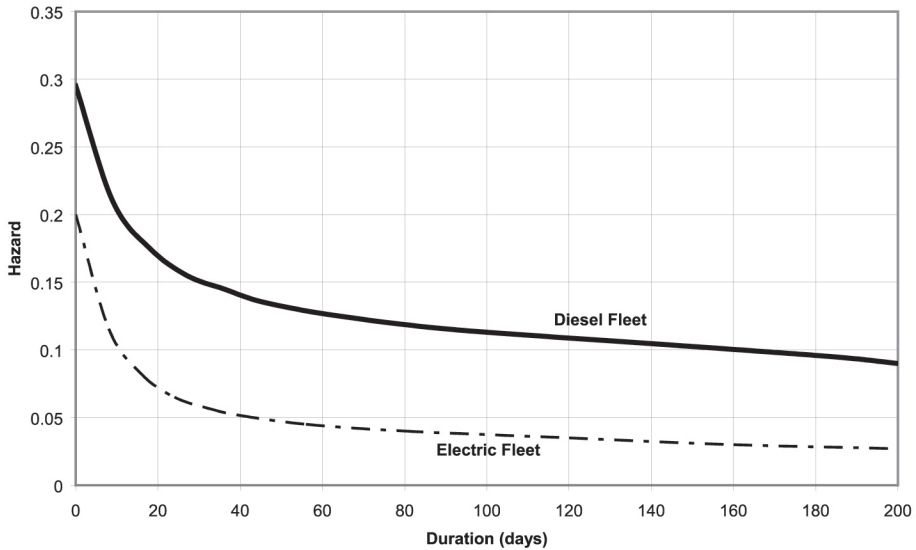## Figure 1. Diesel and Electric Fleet In-Service Hazard Function

Figure 2. Diesel and Electric Fleet Out-of-Service Hazard Function

Transition Probability Estimation

Transition probabilities tell the transit manager the probability that the bus will transition to the worse state (out of service) in the following time step (Δ). This is a very powerful tool as it indicates that the probability does in fact increase as the duration increases. The transition probabilities also give the manager a good approximation of the amount of maintenance operations that are likely to be required in the coming time step. Figure 3 displays the transition probabilities for a bus in each fleet as calculated by Eq. (5). The graph indicates the transition probabilities for transitioning from an in-service to an out-of-service state. It gives these transition probabilities for each fleet for a three-day and a seven-day time horizon. For instance, given an electric bus has been in-service for 55 days, the probability of transitioning to a worse state (out of service) is 32 percent within the next three days and 61 percent within the next seven days. This graph also shows the dependence of the transition probabilities on the time step chosen. The larger the time step, the higher the probability of transitioning to a worse state within the following time step.

Figure 3. Transition Probabilities from In-Service to Out-of-Service State Diesel and Electric Fleet
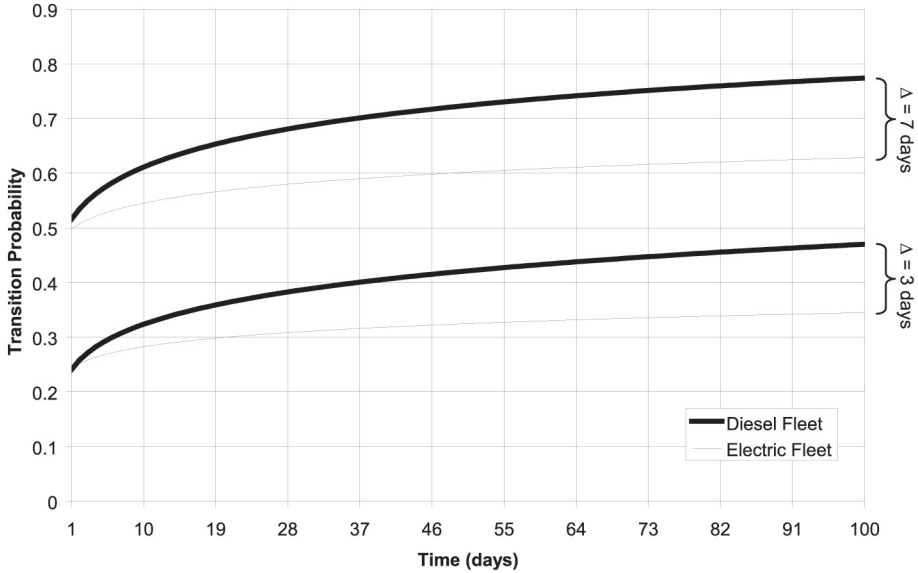
Expected Duration

The expected duration is required when attempting to determine the optimum fleet size. This expected duration is calculated for both the *interval* of in-service time and the *duration* of out-of-service time using Eq. (2). Table 2 shows these calculations for each fleet.

Table 2. Expected Value Estimates

			Interval	*Duration*
Diesel Fleet		p	1.2222	0.74144
		λ	0.10287	0.38997
		E[T]	16.11	2.01
Electric Fleet		p	1.11141	0.5597
		λ	0.09831	0.28094
		E[T]	13.17	2.04

From these data, the expected proportion of time out of service is calculated as the expected *duration* divided by the total time both in and out of service. This can be expressed as:

$$P(out\ of\ service) = \frac{E[T_{dur}]}{E[T_{dur}] + E[T_{int}]} \tag{6}$$

When these proportions are calculated, we find that vehicles are out of service 11.1 percent and 13.4 percent of the time, at any given time for diesel and electric vehicles, respectively. These proportions can easily be extended to a required fleet size. To cover for the expected amount of time a vehicle is out of service, an extra vehicle has to be in service. The required fleet size then becomes the required vehicles in service plus the number of vehicles out of service. The expression is:

$$Fleet\ Size = \frac{Service\ Requirements}{1\text{-}P(out\ of\ service)} \tag{7}$$

For instance, if San Francisco transit planners require 500 diesel buses for service coverage, the required number of buses is 500/(1-0.111)=563 buses. This covers for the expected number of buses, without stochastic fluctuations. Policy questions that must be answered are: What confidence interval is adequate? And, what is the value of full-service coverage? The total number of buses could be higher than this number depending on the variance of durations and the benefit/cost ratio of assuring full service.

Conclusions and Recommendations

This article demonstrated a method of determining rolling stock availability using fully parametric duration models. Two similar datasets were analyzed, an electric and a diesel fleet from the San Francisco Municipal Transit Agency. The data were manipulated to demonstrate that vehicles were either in or out of service. Using LIMDEP statistical software, Weibull distributed hazard and survival functions were developed to identify the instantaneous hazard and survival rates for both the *interval* of time that the bus is in service and the *duration* of time that the bus is out of service. The estimated parameters of the model indicate that the electric and diesel fleets are in fact different and that they should be modeled separately.

This difference could be attributed to the reliability of the vehicles or the different maintenance procedures for the two fleets. Additionally, the magnitude of the estimated parameters indicates that the deterioration, and thus hazard rate of the buses, is monotonically increasing, implying that the longer a bus is in service, the higher the rate of breakdown. Conversely, the longer the bus is in the garage for maintenance, the lower the chances of being released, thus creating a monotonically decreasing relationship. The parameters indicate a significant difference from an exponential distribution that would imply no duration dependence and a Markovian property.

Once the survival and hazard functions were calculated, the discrete time, state-based transition probabilities were calculated as the probability of changing from state 1 (in service) to state 0 (out of service) in the next discrete time step. This allows the planner to identify the probability of a vehicle requiring maintenance the next day, based on the amount of time the vehicle has been in service. This is a potentially powerful tool to use when planning maintenance activities.

Finally, the expected *interval* in service and *duration* out of service are calculated to determine the proportion of time a vehicle is available and thus the fleet size required to cover all scheduled service. This powerful policy tool offers the potential to save a large amount of money for a transit agency considering rolling stock acquisition.

A major improvement of this model would be to include covariate data. The hazard rate is a function of many of the vehicle's variables, including age, mileage, manufacturer, size, route, etc. These variables would certainly have an effect on the time taken to require maintenance and would result in a much more powerful model. These variables would also eliminate the need to account for vehicles that are not in full service or have been taken out of service using the methodology of taking those buses out of the dataset, as the age and mileage parameter would certainly account for those effects. Additionally, this model only analyzes the time in service and the time out of service. An improvement would be to consider the type of service required. Road calls have very different maintenance characteristics than routine maintenance. By differentiating them, policy decisions could be made on maintenance schedules, garage capacity, workforce composition, and support equipment. Additionally, weaknesses in the maintenance schedule could be identified and improved.

Overall, the methodology presented here could be applied to any transit agency to quantify their bus availability and maintenance program. This could potentially save a transit agency millions of dollars through improved rolling stock utilization as well as reduced maintenance and storage facilities. These methods give transit maintenance operators the tools required to make informed decisions to improve their maintenance programs, ultimately improving the quality of service provided to transit customers.

Acknowledgments

References

Dutta, U., T. H. Maze, and A. R. Cook. 1986. Effectiveness of improved repair scheduling in the performance of bus transit maintenance. *Transportation Research Record* 1066.

Fukuoka, Hiroshi. 1987. The analysis method of maintenance data in rolling stock reliability administration-reliability estimation from randomly left truncated and right censored data in restoration-renewal process. *Railway Technical Research Institute* 28 (1).

Haghani, A., and Y. Shafahi. 2002. Bus maintenance systems and maintenance scheduling: Model formulations and solutions. *Transportation Research. Part A: Policy and Practice* 36 (5).

Hickman, Mark D., H. N. Koutsopoulos, and N. H. M. Wilson. 1988. Strategic model for operator work-force planning in the transit industry. *Transportation Research Record* 1165.

Karlaftis, M. G. 1997. A comparison of pattern recognition and probabilistic techniques for capital asset deterioration. *Journal of Public Transportation* 1 (3).

Karlaftis, M.G., and K.C. Sinha. 1997. Modeling approach for transit rolling-stock deterioration prediction. *Journal of Transportation Engineering* 123 (3).

Koutsopoulos, Haris N. 1990. Scheduling of extraboard operators in transit systems. *Transportation Science* 24 (2).

Lederer, B. A., and L. C. MacDorman. 1990. Bus maintenance performance: Findings and direction for research. *Transportation Research Record* 1266.

MacDorman, Littleton C., and J. C. MacDorman. 1982. The transit extraboard: Some opportunities for cost savings. *American Public Transit Association Conference Proceedings*.

MacDorman, Littleton C. 1985. Extraboard management: Procedures and tools. *NCTRP Synthesis of Transit Practice* 5.

Mishalani, Rabi G, and Samer M. Madanat. 2002. Computation of infrastructure transition probabilities using stochastic duration models. *Journal of Infrastructure Systems* (December).

Obeng, K. 1988. Applications of path analysis methodology to transit system maintenance performance. *Transportation* 15 (4).

Perry, James L. 1983. Study of extraboard scheduling, workers' compensation, and operator stress in California public transit. *Report of the Institute of Transportation Studies*. Irvine, California.

Perry, James, and Lyn Long. 1984. Extraboard scheduling, workers' compensation, and operator stress in public transit: Research results and managerial implications. *Transportation Research Record* 1002.

Puntis, R., and D. M. Walley. 1986. The use of reliability techniques on traction and rolling stock. *Railway Engineer* 4.

Schiavone, J. J. 1997. Monitoring bus maintenance performance. *TCRP Synthesis of Transit Practice* 22.

Washington, Simon P., Matthew G. Karlaftis, Fred L. Mannering. 2003. *Statistical and econometric methods for transportation data analysis*. Boca Raton, FL: CRC Press.

About the Author

CHRISTOPHER R. CHERRY (*ccherry@ce.berkeley.edu*) is a Ph.D. student and graduate student researcher in the Department of Civil and Environmental Engineering at the University of California, Berkeley. He holds a master's degree in civil engineering from the University of Arizona. Mr. Cherry's research interests include public transit operations and planning, transit security, transportation policy, and transportation needs of developing countries.