A parametric model of the train delay distribution to improve planning of heavy haul cycle times

M.C. Shih, C.T. Dick & C.P.L. Barkan
University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

ABSTRACT: Efficient heavy haul operations require accurate predictions of terminal arrival times and equipment cycle times. Most analytical and parametric approaches for railway capacity evaluation and operations planning focus on predicting average train delay and not the performance of individual trains and their delay distribution. In practice, train delay varies about this average according to some distribution and the cycle-time performance of certain heavy haul trains may be far from average. Using average values of train delay and cycle time during the creation of a heavy haul operations plan may lead to erroneous conclusions that impact the stability of train operations. To address this shortcoming, this research developed a parametric model for predicting the distributions of train delay on single-track mainlines. The new train delay distribution model is based on a set of indices developed to measure both the amount of traffic and the degree of traffic heterogeneity (differences in train speed and priority) present on the route under study. A quantile regression approach was used to build the model since existing statistical distributions could not adequately represent typical train delay distributions within heterogeneous railway traffic. The developed model can be used to assess the impact of changes in traffic mixture (number of high, medium and low priority trains) and train parameters (speed and priority) on the train delay distribution.

1 INTRODUCTION

The duration of train cycles and arrival times at loading and unloading terminals are highly relevant to the planning of heavy haul operations. Many heavy haul operations use single-track routes that are shared with other types of trains, potentially leading to unplanned train meets that lengthen the cycle time and delay arrival at terminals. The amount of delay may fluctuate with each train run, with very few trains achieving the minimum run time between terminals. A precise estimate of the train delay distribution can be used to predict the reliability of a planned train cycle or terminal arrival time relative to the minimum running time over the route. Better planning-level estimates of the reliability of heavy haul operations can improve both the efficiency and robustness of the system.

Most analytical and parametric approaches to rail traffic performance focus on predicting average train delay (Burdett and Kozan, 2006; Mitra and Tolliver, 2010; Murali et al., 2010) and not the performance of individual trains or their delay distribution. In practice, train delay fluctuates according to some distribution and the performance of an individual train may be far from the average values. Using average values of train delay for planning heavy haul operations without knowledge of the train delay distribution could lead to erroneous conclusions if a system is particularly sensitive to the occurrence of extended train delays.

However, tools currently available to practitioners are not designed to directly estimate the distribution of train delay for a given train without the need for extensive simulation. While detailed simulation models can estimate the distribution of train delay across multiple days of simulated train operations, the models are computationally-intensive. Detailed information on the route infrastructure and signal system is required to develop the simulation model, consuming scarce railway planning resources. Practitioners could be better served by a model that combines the delay distribution output of simulation with the computational efficiency of a parametric approach.

This study develops a parametric model for the distribution of train delay on a single-track line. A quantile regression approach (Koenker and Bassett, 1978; Machado and Mata, 2001; Nielson and Rosholm, 2001) is used to build the parametric model since existing statistical parametric distributions (like Gaussian, Weibull, or Poisson distributions) do not adequately represent the delay distribution of heterogeneous railway traffic.

2   BACKGROUND

Most heavy haul operations that do not operate on dedicated lines involve trains with different length, horsepower, weight, speed and priority characteristics. Some parametric models of average train delay consider variation in train speed and priority. Previous research has proposed factors to better quantify the characteristics of rail traffic with multiple train types (Chen and Harker, 1990; Landex, 2008; Gorman, 2009; Lai et al., 2010; Dingler et al. 2013). The factors proposed by other researchers are either empirical, lack generality or do not have a direct physical meaning that translates to a train delay mechanism. More generalized factors with direct links to the mechanics of train delay may better quantify rail traffic heterogeneity and the causal relationships leading to the distribution of train delay observed on a route.

Additionally, previous parametric models typically do not consider the impact of flexible train operations common in North America. Freight trains in North America do not adhere to a fixed timetable with pre-planned train meets at specific passing sidings (passing loops). Instead, each train operates with a certain amount of schedule flexibility to achieve a desired level of service defined by a maximum allowable train delay. Variation in schedule flexibility and level of service adds another dimension of traffic heterogeneity that needs to be considered by parametric models of train delay.

This study defines schedule flexibility as the departure and trip time flexibility of a single train (Figure 1). Schedule flexibility is a parameter associated with an individual train and different trains on a route can exhibit varying degrees of schedule flexibility depending on their business objectives and level-of-service requirements. The variation of schedule flexibility across all trains operating on a route during a certain dispatching period defines the operating style.

Time-distance diagrams can be used to compare "structured operation" on a fixed timetable (Figure 2a) to the "flexible operation" more common on heavy haul lines in North America (Figure 2b), Under the structured operating style, trains follow predetermined timetables with precise departure times and pre-set meet locations. Under the flexible operating style, the business objectives of heavy haul service usually require dispatchers to dynamically adjust predefined train plans. As a consequence, the departure, and arrival time of trains under flexible operation are ranges instead of points, and the trip time is a band instead of a line. Also, the potential traffic conflict between two trains is a zone instead of a precise point on the structured timetable. These differences need to be considered when evaluating the impact of flexible operation and traffic heterogeneity on train delay.
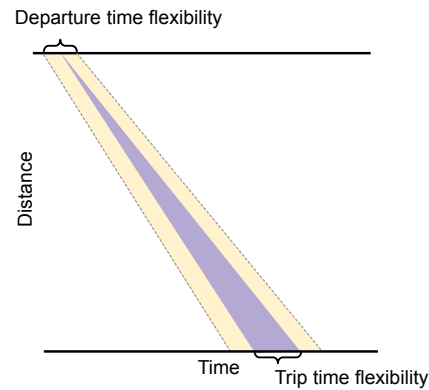

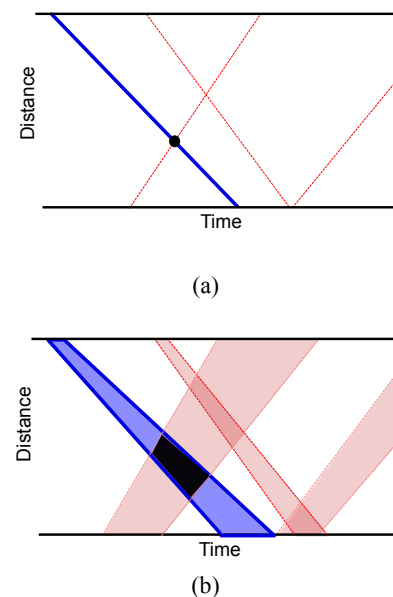Figure 1. Departure time, and trip time flexibility

(a)

(b)

Figure 2. Time and distance diagrams under (a) fixed timetable (b) flexible operation

This study attempts to identify potential parametric model factors to explain the expected variation in train delay based on analysis of traffic conflicts under flexible operations.

3   PARAMETRIC MODEL DEVELOPMENT

The proposed parametric model of the train delay distribution was developed in two stages. In the first stage, appropriate heterogeneity factors were developed based on the concept of traffic conflict analysis. In the second stage, the factors were used to construct a quantile regression model of the train delay distribution. A cross-validation-based process was used to assess the accuracy of the developed quantile regression model.

Train delay is usually related to rail traffic volume through delay-volume curves. However, this ap-

proach only considers the amount of traffic on the route under study and not the degree of traffic heterogeneity. The concept of using train conflicts (also referred to as traffic conflicts) to predict train delay was presented by Gorman (2009). Gorman found that traffic conflicts, represented by the number of meets, passes and overtakes, impact train delay significantly. This study seeks to expand on this idea to determine if the number of traffic conflicts can be used to describe the relationship between the degree of traffic heterogeneity and the distribution of train delay.

As a preliminary study, various heterogeneous railroad traffic scenarios were simulated with Rail Traffic Controller (RTC) simulation software to investigate a potential relationship between the number of traffic conflicts and train delay on a representative North American shared corridor. Under these conditions, the expected number of traffic conflicts (calculated by counting the total traffic conflicts each train could encounter based on a Monte Carlo process) is more closely correlated with train delay than the total traffic volume (Figure 3). This suggests that the number of traffic conflicts captures both the impact of traffic volume and heterogeneity. By capturing additional information about the rail traffic, the number of rail traffic conflicts may be an alternative predictor of the train delay distribution than traffic volume alone.
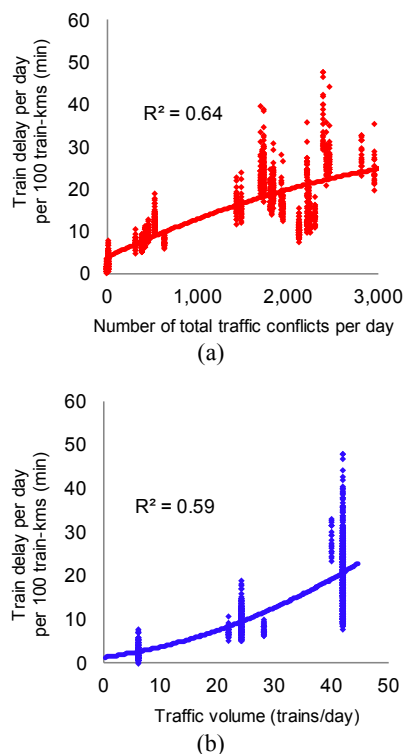

(a)

(b)

Figure 3. Relationship between (a) traffic conflicts, (b) traffic volume, and the average train delay per 100 train-kilometers

Heterogeneity can arise from different combinations of speed variation, priority variation, and operating styles. Based on the concept of traffic conflict analysis, three train delay factors were defined to quantify these attributes of heterogeneous rail traffic:

- Total Conflicts (TC) considers all of the potential conflicts a train may encounter during its trip and that a larger number of traffic conflicts increases the difficulty of the train dispatching task. TC is calculated by examining a set of train departures, creating train paths at the train operating speed and counting the total number of conflicts between train paths. Conflicts between train paths are not resolved.

- Adjusted Train Priority (ATP) quantifies the actual priority of a train within the given traffic mixture on the route. ATP is calculated for a given target train by the summation of inferior conflicts (target train has inferior priority relative to the conflicting train) and half of equal conflicts (target train has equal priority to the conflicting train). The physical interpretation of ATP as a delay mechanic is the number of conflicts where the target train will need to stop and wait for the other conflicting train to pass.

- Inferior Pass (IP) represents the impact of train speed heterogeneity on train conflicts and delay. Variation in speed between trains creates additional delay when one train is required to pass another. IP calculates the expected number of inferior passes (target train has inferior priority to passing train). The physical meaning behind IP is the expected number of passes that will cause the target train to stop or encounter delay.

With the three train delay factors defined, the second step of the study is to apply a quantile regression approach to build the parametric model. Sogin (2013) fit a Weibull distribution to delay data for homogeneous unit train traffic. However, a preliminary test conducted for this study showed that a Weibull distribution, along with other common parametric distributions, did not adequately model the delay distribution of a train in heterogeneous rail traffic. For this reason, a quantile regression approach is used.

Quantile regression is a statistical method used by researchers in the area of macroeconomics to model the interaction between variables and output distributions (Arias et al., 2001; Koenker, 2005). The quantile regression model creates multiple regression lines that each represent a quantile boundary (Figure 4). For example, the 97.5th percentile line

is the best fit such that 97.5 percent of the data points are below the line and 2.5 percent are above the line. In this case, the slope of the curve represents the sensitivity of the 97.5th percentile of train delay to the TC index defined above.
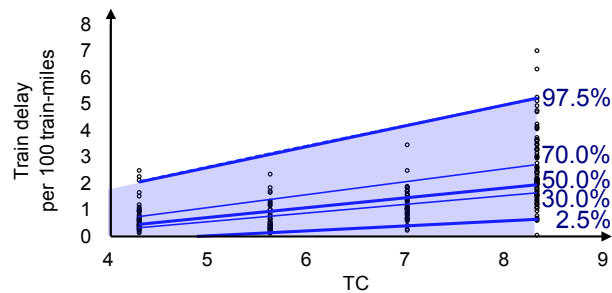

Figure 4. Example of a set of quantile lines

Figure 6. A Monte-Carlo-based process for factor level calculation

Based on the quantile regression technique, the mentioned train delay factors are used as possible variables to consider when building a regression model to predict train delay distributions. The constructed model can be used to analyze the response of the train delay distribution to changes in traffic (as reflected by changes to the three train delay factors).

To develop the quantile regression model, a train plan with associated departure flexibility is required as an input to calculate the values of the train delay factors for each individual train in the train plan (Figure 5).
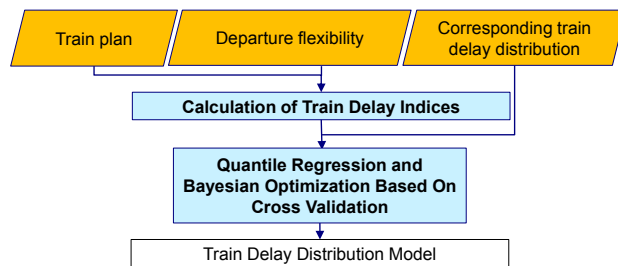

Figure 5. Framework of the development of parametric model

For a given train, each factor is calculated through a Monte-Carlo process that considers the departure time flexibility of each individual train (Figure 6). For each iteration of the process, a set of train paths is selected from within the theoretical train band space defined by the departure time flexibility. From the unresolved train conflicts between these train paths, the corresponding values of the three factors are calculated for each train. The final calculated factor levels for each train are taken as the average values of the factor levels of each train after a certain number of iterations. These values can be used as close approximations of the real factor values.
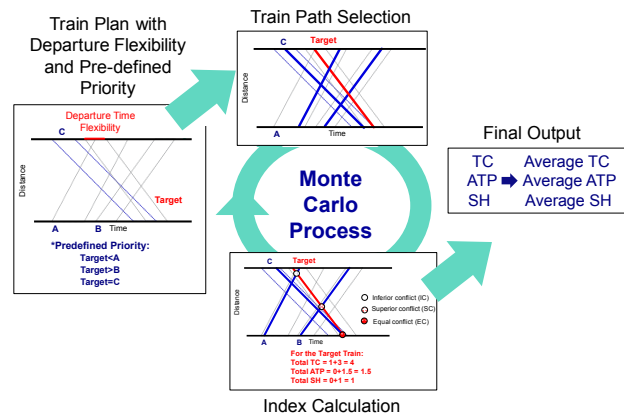
The calculated factor levels for each train and the corresponding distribution of train delays for that train (determined from historical data or simulation of the flexible operations) are then used by the quantile regression approach to construct a Train Delay Distribution Model. The Train Delay Distribution Model inputs are the train delay factor levels associated with a given train path. The model output is the quantile boundaries predicting the distribution of delay for that train.

4   MODEL VALIDATION

The Mean Absolute Deviation (MAD) (Ohashi et al. 2010) is used to assess the performance of the developed train delay distribution model. The function (1) shows the calculation of MAD, where $\hat{Q}_\alpha$ is the estimate of a certain quantile α, and Q_α is the real value of the quantile α.

$$MAD = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{Q}_\alpha - Q_\alpha \right| \qquad (1)$$

For this study, cross validation (Kohavi, 1995) was used to obtain a more robust MAD (Figure 7). The first step in cross validation is to divide the available dataset into k different subsets with equal size. The second step is a looping process where each round a subset is selected as the test data set and other subsets are combined as the training data set. The training data set is used to construct a regression model and the test data set is used to evaluate the performance of the model. Each round of this process generates a MAD for each test set, and the final MAD is the average of all MADs. This final MAD score represents the potential error in predictions of specific quantiles of the train delay distribution made by the constructed model. As

will be demonstrated in the case study, the MAD scores for a model may vary between each quantile of the delay distribution. This indicates that the predictive ability of the model is not consistent across the entire delay distribution.
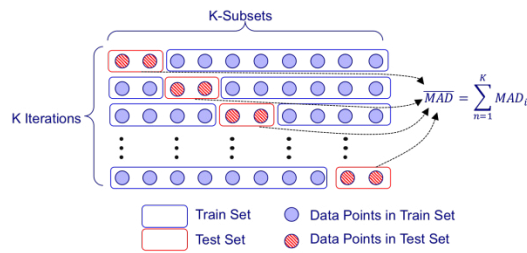

$$\overline{MAD} = \sum_{n=1}^{K} MAD_i$$

Figure 7. Relationship of cross validation and MAD in this study

5 CASE STUDY

To demonstrate the construction and predictive ability of the train delay distribution model, it was applied to a 242-mile single-track route representative of heavy haul operations in North America (Figure 8). The traffic tested on this infrastructure is assumed to have the following characteristics:

- Traffic volume varies from 8 to 26 trains per day
- Three train types with high. medium, and low priority exist within the traffic mixture with the percentage of each train type ranging from 25 to 75 percent
- For each potential combination of traffic volume and mixture, there exist three different patterns of train departures from the terminals at either end of the route
- The departure time flexibility for high, medium, and low train types are 0.5, 1.5, and 3 hours, respectively
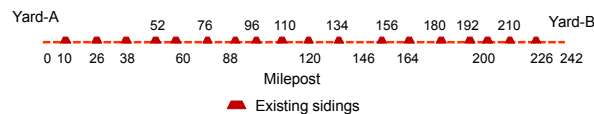

Figure 8. Tested infrastructure layout in RTC

The traffic and infrastructure parameter values above were used to construct an experiment matrix based on the concept of full factorial design. All scenarios in this matrix were simulated with RTC over multiple days of operation to obtain the delay distribution of each train.

The traffic parameters for each scenario in the experiment design were also input to the Monte Carlo process to obtain the factor levels associated with each train. The calculated train delay factors and their associated train delay distributions for each

train were then used to construct the train delay distribution model for this route.

To assess the performance of the case study model, ten subsets were used in the cross validation process. The mean of delay corresponding to each quantile and calculated MADs show the potential prediction error for different quantiles (Table 1). For example, the mean of real delay value corresponding to the 5th percentile is 4.4 minutes. On average, five percent of trains experience less than 4.4 minutes of delay per 100 train-miles. The MAD for the 5th percentile is 1.6 minutes suggesting that the true value of the 5th percentile of train delay falls within the range of 4.4 +/- 1.6 minutes.

For the 5th to 70th quantiles, the MADs range from one to three minutes of train delay. Above the 70th quantile, the MAD increases quickly from three to eight minutes for the 95th quantile. The ability of the model to precisely predict delays corresponding to higher quantiles decreases. This result indicates that the current factors used to construct the model cannot completely explain the extreme train delays observed in the RTC simulation output. Additional research is needed to identify factors that can explain these high-delay cases and incorporate them into the model.

Table 1. MADs from cross validation and the corresponding delays of each quantile

Percentile	Mean of MADs	Mean of delay from raw data
5%	1.6	4.4
10%	1.7	5.9
20%	1.7	8.3
30%	1.8	10.5
40%	1.8	12.6
50%	2.0	14.6
60%	2.3	17.1
70%	2.8	19.8
80%	3.8	23.5
90%	4.8	29.1
95%	8.1	34.5

6 CONCLUSION AND FUTURE STUDY

Equipment cycle and terminal arrival times are important considerations for heavy haul operations planning. Estimating the reliability of these times under different operating scenarios requires knowledge of both the average and distribution of train delay. An accurate prediction of the train delay distribution can improve the efficiency and robustness of heavy haul operation plans. However, previous parametric models used for predicting train delays estimate the average delay and not the entire train delay distribution.

This study proposed a train delay distribution model based on the concept of quantile regression

and newly developed factors that capture the combined impact of speed, priority variation and operating style (with its associated schedule flexibility). The cross validation result in the case study suggests that the model can accurately predict the delay corresponding to different quantiles below the 70th percentile.

To improve the performance of the model at higher percentiles, regularized quantile regression (i.e. Lasso) could potentially be used to replace the standard quantile regression approach implemented in this study. Additionally, conducting root cause analysis of extreme train delays under flexible operation may identify new factors to better characterize heterogeneous train operations. Adding these factors to the model could also improve its performance.

A next step for this research is to test the developed model on other rail corridors and generalize the train delay distribution model. The model developed in the case study was only validated on the same route and under similar traffic conditions used to construct the model. While it can accurately predict the train delay distribution under a wide range of rail traffic conditions on this corridor, it is an open research question to determine how well the model will translate to other single-track routes with different siding spacing. Additional factors describing the route infrastructure could be included in the model to potentially generalize it to any single-track corridor. However, even before this step is taken, this study still provides a general quantile regression framework that can be used by practitioners and researchers to develop their own route-specific versions of the train delay distribution model.

7 ACKNOWLEDGEMENTS

8 REFERENCES

Arias, O., Hallock, K.F., & Sosa-Escudero, W. 2001. Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twin data. *Empirical Economics* 2: 7-40.

Burdett, R.L. & Kozan, E. 2006. Techniques for absolute capacity determination in railways. *Transportation Research Part B: Methodological* 40(8): 616-632.

Chen, B. & Harker, P.T. 1990. Two moments estimation of the delay on single-track rail lines with schedule traffic. *Transportation Science* 24(4): 261-275.

Dingler, M.H., Lai, Y.C., & Barkan, C.P.L. 2013. Mitigating train-type heterogeneity on a single-track line. In *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(2): 140-147.

Gorman, M.F., 2009. Statistical estimation of railroad congestion delay. *Transportation Research Part E: Logistics and Transportation Review* 45(3): 446-456.

Koenker, R. & Bassett, G. 1978. Regression quantiles. *Econometrica* 46: 33-50.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, San Mateo, California: Morgan Kaufmann, 2(12): 1137-1143.

Koenker, R. 2005. *Quantile regression*. Cambridge university press. Cambridge, UK.

Lai, Y.C., Dingler, M., Hsu, C.E., & Chiang P.C. 2010. Optimizing train network routing with heterogeneous traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 2159: 69-76.

Landex, A. 2008. Methods to Estimate Railway Capacity and Passenger Delays. Doctoral Thesis. Technical University of Denmark, Department of Transport. Lyngby, Denmark.

Machado, J.A.F. & Mata, J. 2001. Earning functions in Portugal 1982-1994: evidence from quantile regressions. *Empirical Economics* 26: 115-134.

Mitra, S. & Tolliver, D. 2010. Estimation of railroad capacity using parametric methods. *Journal of the Transportation Research Forum* 49(2): 111-126.

Murali, P., Dessouky, M., Ordonez, F., & Palmer, K. 2010. A delay estimation technique for single and double-track railroads. *Transportation Research Part E: Logistics and Transportation Review* 46: 483-495.

Nielson, H.S. & Rosholm, M. 2001. The public–private sector wage gap in Zambia in the 1990s: a quantile regression approach. *Empirical Economics* 26: 169-182.

Ohashi, O., Torgo, L., & Ribeiro, R.P. 2010. Interval Forecast of Water Quality Parameters. In *Proceedings of the (ECAI) European Conference on Artificial Intelligence 2010*: 283-288.

Sogin, S.L. 2013. Simulations of Mixed Use Rail Corridors: How Infrastructure Affects Interactions Among Train Types. *Master's Thesis*, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering. Urbana, IL, USA.