A Parametric Model of the Train Delay Distribution Based On Traffic Conflicts

Mei-Cheng Shih[a], C. Tyler Dick[a,1], Darkhan Mussanov[a,2], Christopher P.L. Barkan[a,3]

[a] Department of Civil and Environmental Engineering,
University of Illinois at Urbana-Champaign
Newmark Civil Engineering Laboratory, MC-250
205 North Mathews Ave, Urbana, IL 61801-2352
E-mail: mshih2@illinois.edu,
[1] E-mail: ctdick@illinois.edu,
[2] E-mail: mussano2@illinois.edu
[3] E-mail: cbarkan@illinois.edu

Abstract

In North America, most analytical and parametric approaches for railway capacity evaluation and operations planning focus on predicting average traffic delay and not the performance of individual trains and their delay distributions. In practice, a single estimate of average train delay provides the railway planner with no information on the proportion of trains that experience excessive delay. This makes it difficult to develop robust plans for terminal arrival times and equipment cycles, and use of average values of train delay may lead to erroneous conclusions that impact the stability of train operations. Another drawback of existing parametric models for average train delay is that many of the underlying factors do not consider the impact of train operation flexibility on train performance. Different types of underlying factors with clear physical meanings may better quantify the characteristics of rail traffic heterogeneity and, by describing the cause-end-effect relationships between operating factors and train delay, better predict the overall train delay distribution. This study first developed three new features to describe the mechanisms of rail traffic heterogeneity, then these features were used as inputs to construct a parametric model for delay distributions of trains travelling on a specific single-track mainline. The parametric model was developed based on the concept of Lasso Regularized Quantile Regression (LRQR) model. The construction process of the model was demonstrated and an optimization process based on cross-validation and Bayesian Optimization to find the optimal penalty of LRQR was proposed. The outputs from the developed model can help improve the stability and robustness of train operation planning. The case study first showed the accuracy of the model based on the cross validation score, and then displayed an application of the parametric model for train delay management.

Keywords

Rail Capacity Evaluation; Train Delay Distribution; Rail Traffic Heterogeneity; Quantile Regression; Bayesian Optimization

1 Introduction

Most analytical and parametric approaches to railway capacity in North America 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 varies about the average value according to some distribution and the performance of certain trains may be far from average. Using average values of train delay without knowledge of the train-delay performance of the best and worst trains may lead to erroneous conclusions. To address this shortcoming, this study develops a parametric model for the distributions of train delay on a given single-track mainline. A quantile regression approach (Koenker and Bassett, 1978; Machado and Mata, 2001; Nielson and Rosholm, 2001) was used to build the parametric model since existing statistical distributions (like Gaussian, Weibull, or Poisson distributions) do not adequately represent the delay distribution of heterogeneous railway traffic.

Parametric models of average train delay consider variation in train speed and priority. However, they typically do not consider the impact of flexible train operations common in North America. Under flexible operations, individual trains or types of trains may have different schedule and level-of-service requirements, introducing another dimension of rail traffic heterogeneity to the operation that may impact overall train performance. 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; Marković et al., 2015; Jensen, 2015; Jensen et al., 2017). The factors proposed by other researchers are either empirical (such as the ratio of the maximum and minimum train speed in Krueger's study) or lack generality (the features developed by Landex (2008) are only applicable to passenger rail systems under directional operation), and thus do not have a direct physical meaning that translates to a train delay mechanic. For example, the feature representing the ratio of maximum and minimum train speed indicates the degree of speed heterogeneity but does not describe the specific speed heterogeneity mechanics that lead to additional train delay. 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.

To develop appropriate factors for a delay distribution model under flexible train operations, the impact of train operations flexibility needs to be defined first. This study defines *schedule flexibility* as the departure and trip time flexibility of a single train (Figure 1). Thus 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 is defined as *operating style*. Thus operating style is a parameter associated with the overall traffic on a given route and the schedule flexibility of the individual trains that comprise this traffic. There are three operating styles commonly found on the mainlines of the North American rail network (Figure 2). Structured operation contains trains with little schedule flexibility, corresponding to the characteristics of train operation on passenger rail corridors. Oppositely, trains under flexible operation have a higher degree of schedule flexibility. Flexible operations are commonly found on freight lines in North America where the traffic is predominantly trains carrying bulk commodities. There are also scenarios where some trains on a corridor follow structured operation while other trains on the same corridor follow flexible operations. This type of operation is found on shared

corridors in North America where passenger and freight traffic share the same track infrastructure, and is defined as mixed operation in this study.

Different operating styles and their associated schedule flexibility can lead to different types of train conflicts for the same set of train departures. The blue dot, region, and line in Figure 2 shows the potential number and locations of traffic conflict train A is expected to experience under each operating style. This variation in the number, location and type of conflict could have a substantial impact on train performance as measured by train delay. This study attempts to identify potential parametric model factors that relate to the concept of predicting rail traffic conflicts and, by extension, the expected variation in train delay.
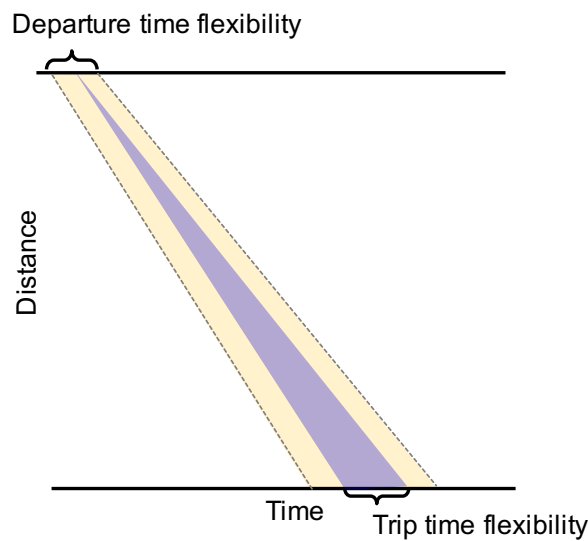
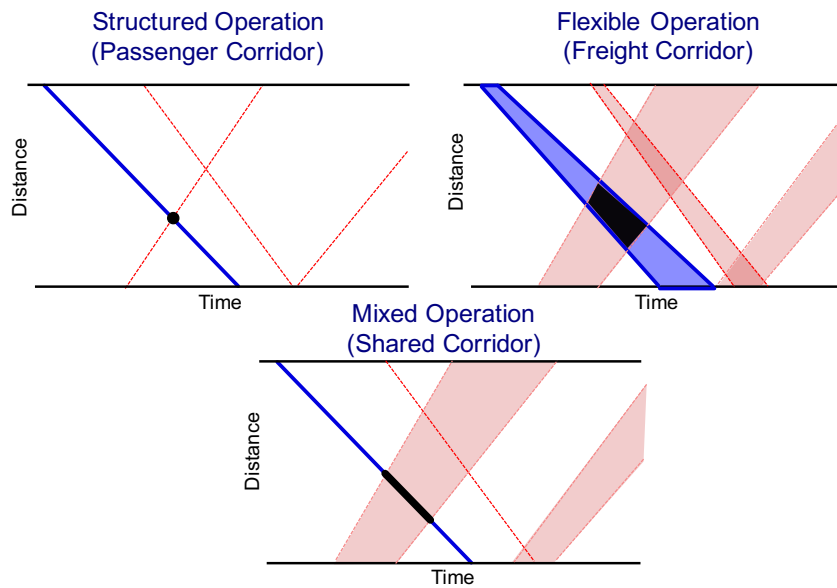

Figure 1: Departure time, and trip time flexibility

Figure 2: Representative operating styles on the North American rail network

The main task of this study is to develop a parametric model for train delay distributions using a new set of train delay features to quantify the degree of rail traffic heterogeneity. The model development can be generalized into two stages. In the first stage, appropriate train delay features were developed to quantify the degree of rail traffic heterogeneity based on the number of potential traffic conflicts encountered by each train. The details of the feature development will be explained in section 2. 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 access the accuracy of the developed quantile regression model. This process will be described in section 3.

2 Development of the New Features

On a given single-track mainline, train delay is usually related to rail traffic volume through delay-volume curves. A shortcoming of this approach is that it only considers the amount of traffic on the route under study and not the degree of traffic heterogeneity. Some previous studies used "train types" as a measure of heterogeneity. Since the requirements for individual trains within a train type can vary according to specific business objectives, general train types do not always correspond to the operational characteristics of an individual train. There is a need for indices to quantify the characteristics of each individual train and not general train types. Some previous studies developed such indices but they usually are too specific to certain operating styles. Also, some of the indices are empirical in that the calculated values showed some correlation to train delay but do not directly depict the delay mechanisms of heterogeneous rail traffic operations.

The concept of using train conflicts (also referred to as traffic conflicts) to better predict train delay was formalized by Gorman (2009). Gorman used historical data from ten single-track freight lines to test the relationship between train running time and various operating and infrastructure factors. Gorman found that traffic conflicts, represented by the number of meets, passes and overtakes, impact train delay significantly. The meets, passes and overtakes encountered on a given rail line are closely related to the degree of traffic heterogeneity. However, Gorman did not connect his factors to traffic heterogeneity. This study seeks 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 delay experienced by trains operating on that route.

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). Also, the number of expected traffic conflicts for a certain traffic volume is a function of both the traffic volume and the traffic mixture encountered at that traffic volume. 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 a better predictor of the train delay distribution than traffic volume alone.
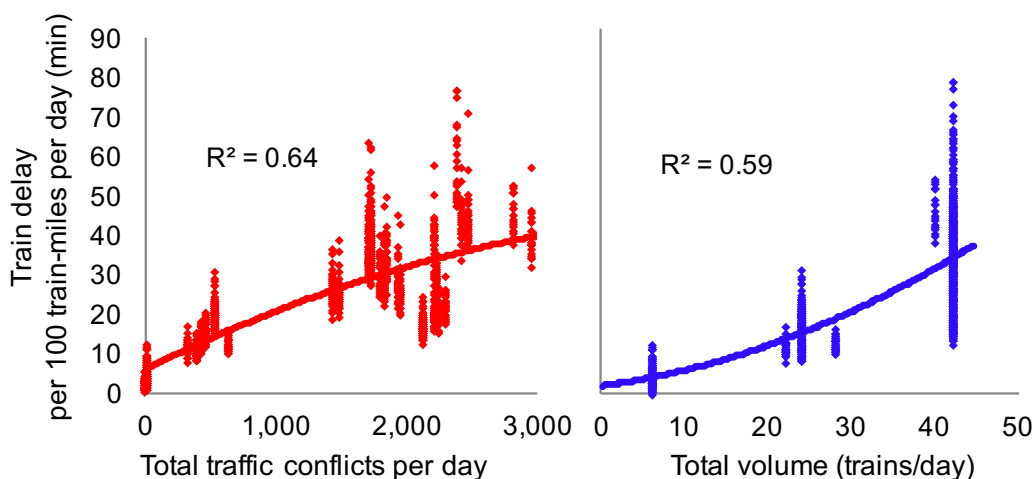
Figure 3: Relationship between traffic conflicts, traffic volume, and
the average train delay per 100 train-miles

As mentioned in the introduction, heterogeneity can arise from different combinations of speed variation, priority variation, and operating styles. Three train delay factors were defined to quantify these attributes of heterogeneous rail traffic and reflect the mechanism of heterogeneous traffic operation:

- Total Conflicts (TC) considers all 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. This feature is also an analogue to traffic volume since higher train volumes usually lead to more train conflicts. TC is calculated by examining a set of train departures and creating train paths at the train operating speed. Conflicts between train paths are not resolved. For a given target train, TC is equal to the total number of conflicts with other train paths regardless of train priority. This same set of train paths is also used to calculate the other two factors described below.

- 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). In past studies, the assigned priority of a train was a static ordinal value based on its train type. The actual priority of a train should be a dynamic value since it varies with the traffic mixture. For example, the actual priority of an intermodal train within traffic composed of 80% inferior trains should be higher than the relative priority of the same train within traffic composed of only 20% inferior trains. 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.

- Speed Heterogeneity (SH) represents the direct impact of differences in train speeds on train conflicts and delay. When SH is high, there is greater diversity in train speed, leading to more frequent passes and meets making up a smaller share of train conflicts. This is in comparison to cases where speed is homogeneous, all train conflicts are meets and SH takes a value of zero. SH calculates the expected number of inferior passes (target train has inferior priority to passing train). The physical meaning behind SH is the expected number of passes that will cause the target train

to stop or encounter delay. Delays for passes are assumed to be the origin of extra delay caused by train speed heterogeneity.

To explain how values for the three features are calculated for each train, three types of traffic conflicts needed to be defined (Figure 4). The three types of conflicts are illustrated from the perspective of the red "target" train by classifying the types of conflicts it encounters according to the relative priority of the other trains. If the target train is inferior to the other train, then that conflict is an inferior conflict (IC). If the target train is superior to the other train, then that conflict is a superior conflict (SC). If the target train priority is equal to the other train, then that conflict is an equal conflict (EC). The concept of IC, SC, and EC zones represent the potential zones in which each type of conflict may occur in time and distance space.
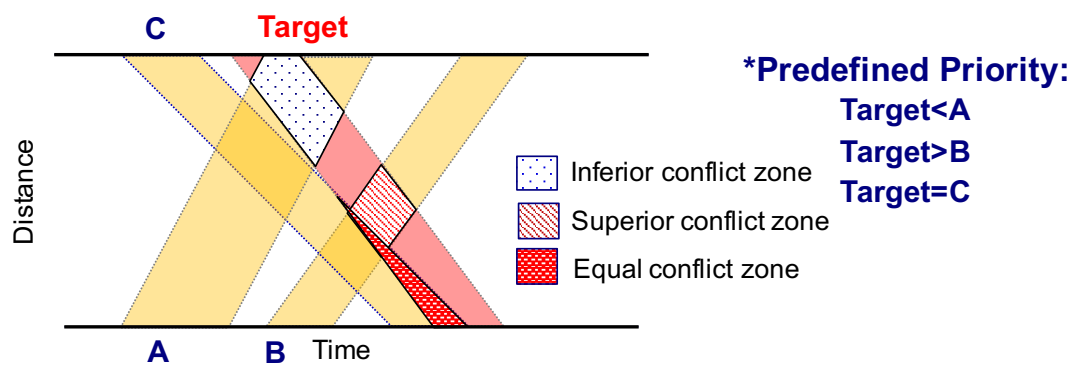

Figure 4: Three types of traffic conflicts used in the calculation of train delay factors

Using these three conflict types, the factors can be calculated for each train by equation (1)~ (3). The IC, SC, and EC in the equation is the number of IC, SC, and EC conflicts a train is expected to encounter during its travelling time across the route segment under study. For the 1/2 equal conflict term in equation (2), it is set based on the assumption that there is a 50% probability the target train will stop and wait at a meet with an opposing train with equal predefined-priority. This value can be adjusted according to the target scenario. For the IC term in equation (3), only the inferior pass conflicts are considered, the inferior meet conflicts are excluded.

$$TC = IC + SC + EC \tag{1}$$
$$ATP = IC + 1/2\,EC \tag{2}$$
$$SH = IC \text{ (Only inferior pass conflcits are considered)} \tag{3}$$

The departure time flexibility may follow different distributions. Different equations for the feature values could be derived from these statistical distributions of departure times for each train assuming they follow distributions with convenient closed-form equations. A more general way to calculate the feature values is thorough a Monte-Carlo process (Figure 5). During each iteration of the Monte Carlo process, a set of train paths was selected from within the set of train bands in time and distance space defined by the departure and trip time flexibility. From the unresolved train conflicts between these train paths, the corresponding values of the three features are calculated for each train. The process is then repeated with the selection of new train paths and recalculation of the feature values for each train (Figure 6). The initialization process requires a pre-

determined train plan with departure flexibility and predefined priority for each train (Figure 6a). The total value of each feature is accumulated through the iterations (Figure 6b, Figure 6c, Figure 6d) and the final output of the process is the projected average values of TC, ATP, and SH for each train (Figure 6d). Since the train paths were generated randomly within each train band based on the stochastic properties of the train operation (i.e. statistic distribution for each departure flexibility), taking the mean of the total projected conflicts is equivalent to calculating the expected number of traffic conflicts along the mainline.
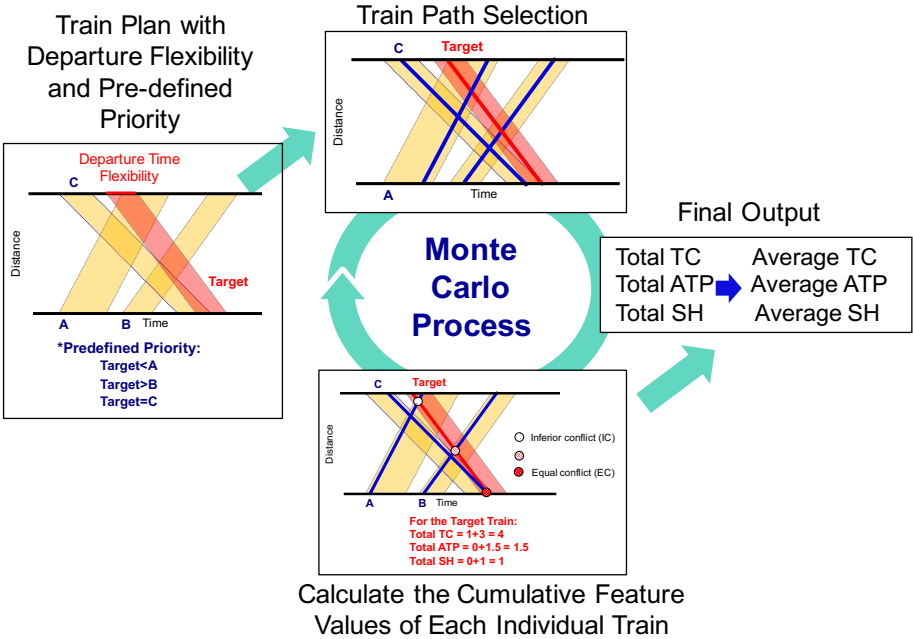

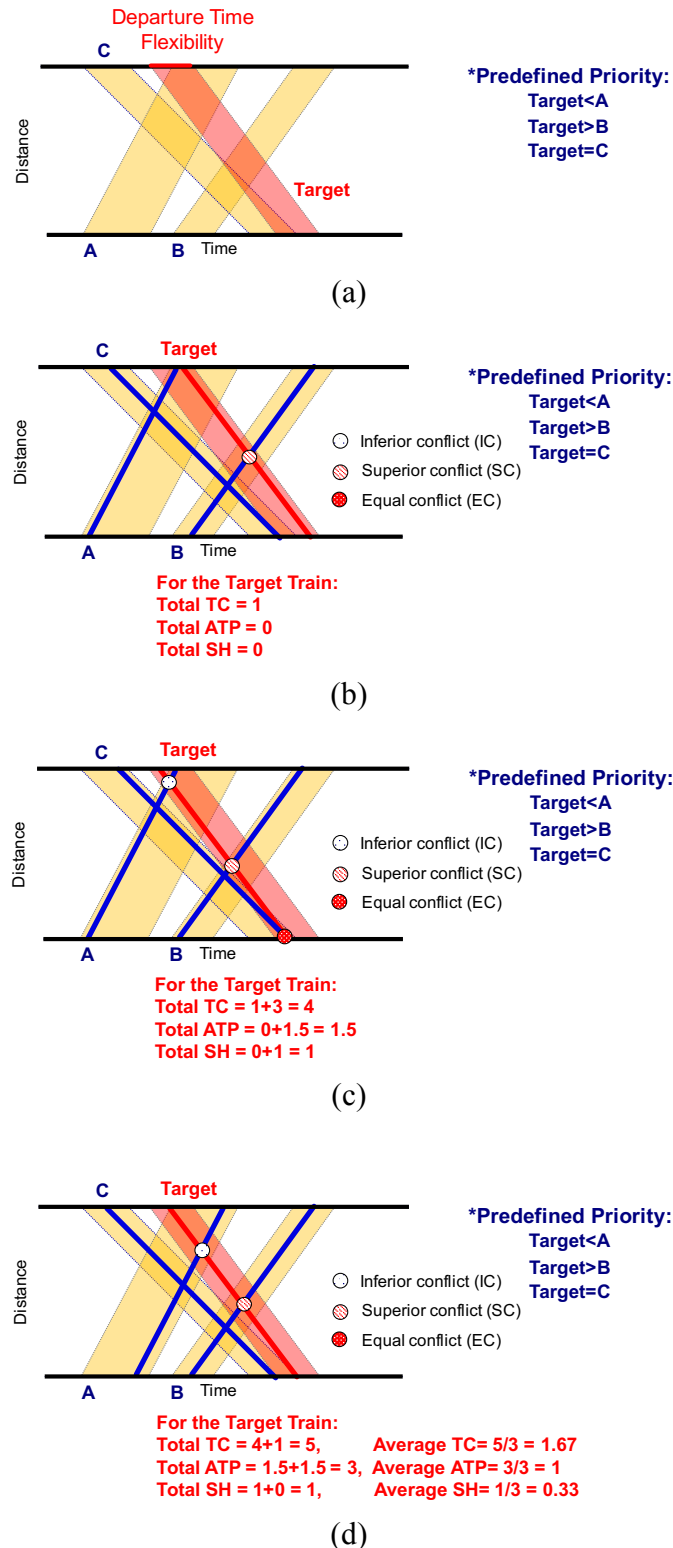Figure 5: A Monte-Carlo-based process for feature value calculation

Figure 6: Example of the Initialization and three iterations of the Monte-Carlo-based process (a) initialization of train plan with departure flexibility and predefined priority (b) first iteration (c) second iteration (d) third iteration

Theoretically, each feature value will converge to a certain value after infinite iterations of the Monte Carlo process. To ensure the computational efficiency, the largest percent change across all feature values between subsequent iterations was used as a threshold to stop the Monte-Carlo process. An arbitrary threshold (0.01%) was selected for the case study.

The output of the Monte-Carlo process is the feature values for each train in the pre-determined train plan. The values for each train along with the corresponding delay distribution for that train are used as inputs to develop the parametric model. When using the constructed parametric model, the features values are the inputs, and the train delay distribution are the outputs. The impact of infrastructure properties were not considered by this study, so for different single-track mainlines with various siding length and spacing properties, different parametric models needed to be constructed separately using, the process outlined in this paper.

3   Development of the Parametric Model

With the three train delay factors defined and the corresponding values calculated, the second step of the study is to apply a quantile regression approach to build the parametric model. The traditional way to model the train delay distribution is to use parametric statistics, like the Gaussian distribution, and regression on the parameters of the selected distribution. Sogin et al. (2013) fitted a Weibull distribution to the delay data from homogeneous unit train traffic and built regression models on the parameters of the distribution. However, a preliminary test conducted for this study showed that a Weibull distribution, along with other common parametric distributions, may 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 field of macroeconomics to model the interaction between variables and distributions (Arias et al., 2001; Koenker, 2005). The quantile regression model creates multiple regression lines that each represent a quantile boundary (Figure 7). An example of the quantile boundary is the 97.5 percentile in Figure 7. It shows the best fit such that approximately 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 feature defined above. It is a first-order linear relationship in the example but it can be polynomial or exponential depends on the terms used. 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).
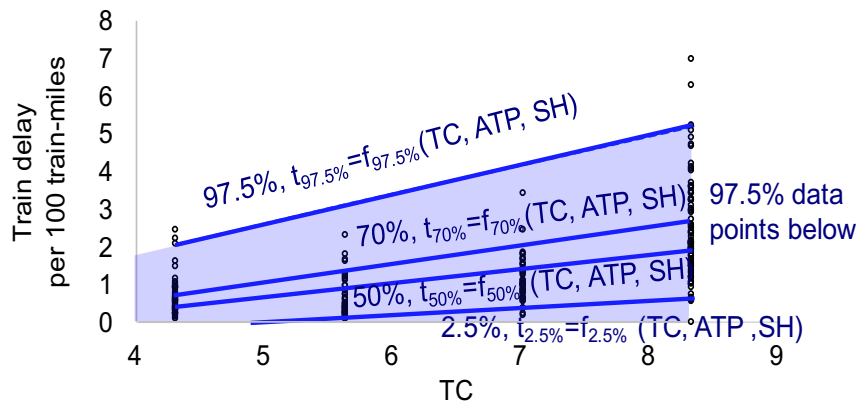
Figure 7: Example of a set of quantile boundaries in the space of TC and train delay

There are some possible basic relationships (e.g. first-order linear, polynomial, exponential, nonlinear, feature interaction) that can be selected to construct the model. Nonlinear quantile regression models like quantile forest or quantile boosting trees are usually more precise and accurate. However, since the features developed have physical meaning, and all the indices should have a positive linear (including first-order linear, polynomial, and feature interaction) relationship to delay distributions, linear quantile models should be the more appropriate choices.

In this study, the first order, second order, and interaction terms of the features were used as inputs for model construction. Since overfitting is a common problem of this type of model, a Lasso (L1-norm) Regularized Quantile Regression (LRQR) approach was used to avoid the problem (Wu and Liu, 2009; Sherwood et al., 2016). Lasso regression approach is a general concept used in all types of regression to prevent overfitting. The Lasso approach applied a penalty to all the regression coefficients to help the model select the best number of coefficients and terms to include while avoiding overfitting (Li et al., 2015; Peng and Wang, 2015). When combined with the quantile regression approach, the Lasso approach can help the quantile models avoid overfitting. Equation (4) is the model fitting function of the LRQR for finding the quantile functions, where y_i are the values of dependent variable, x_i are the values of independent variables, β_o, β are the intercept and coefficients of regression, ρ_τ is the percentile and λ is the penalty. The term $\lambda \|\beta\|_1$ is the L1-norm penalty. Different penalty λ leads to different coefficient values and model accuracy. There is always an optimal value or values that minimize the error of the model, which maximize the accuracy of the model.

$$\min_{\beta_o,\boldsymbol{\beta}} \sum_{i=0}^{n} \rho_\tau(y_i - \beta_o - \beta^T x_i) + \lambda \|\beta\|_1 \tag{4}$$

To find the optimal penalty for a model, two processes are required: a process to robustly assess the model performance under different penalty values, and an assessment process to find the optimal penalty efficiently. For the assessment process, a standard that can be used by the process to quantify performance is also required. The standard used to assess performances is Mean Absolute Deviation (MAD) (Ohashi et al. 2010), and the assessment process is cross validation (Kohavi, 1995). The function (5) 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.

$$MAD = \frac{1}{n} \sum_{i=1}^{n} | \hat{Q}_\alpha - Q_\alpha | \qquad (5)$$

One way to obtain MAD value for a model based on certain penalty is to divide the data into training and testing sets first, construct the model with the training set, and then calculate MAD based on the predicted and real values of the test set. However, the calculated MAD may be subject to a certain training and test partition bias. Cross validation was used to avoid this potential bias (Figure 8). Cross validation generates several MADS based on different training and test partitions, and then takes the average of them to calculate as the final MAD. In more detail, the first step of cross validation is to divide the available dataset into k different subsets with equal or equivalent sizes. The second step is to loop over the k subsets. For each iteration, the kth subset is used as a test set and the other subsets are combined as a training set. A LRQR model is generated based on this training set and a certain penalty. The MAD of each iteration is calculated based on the predictions from the trained model and the real values of the test set. After all iterations are completed, the final MAD is the average of all MADs obtained across the iterations. Since this final MAD is the average of different training and test data partitions, it is more representative compared to MAD obtained by only testing one combination of training and test partition.
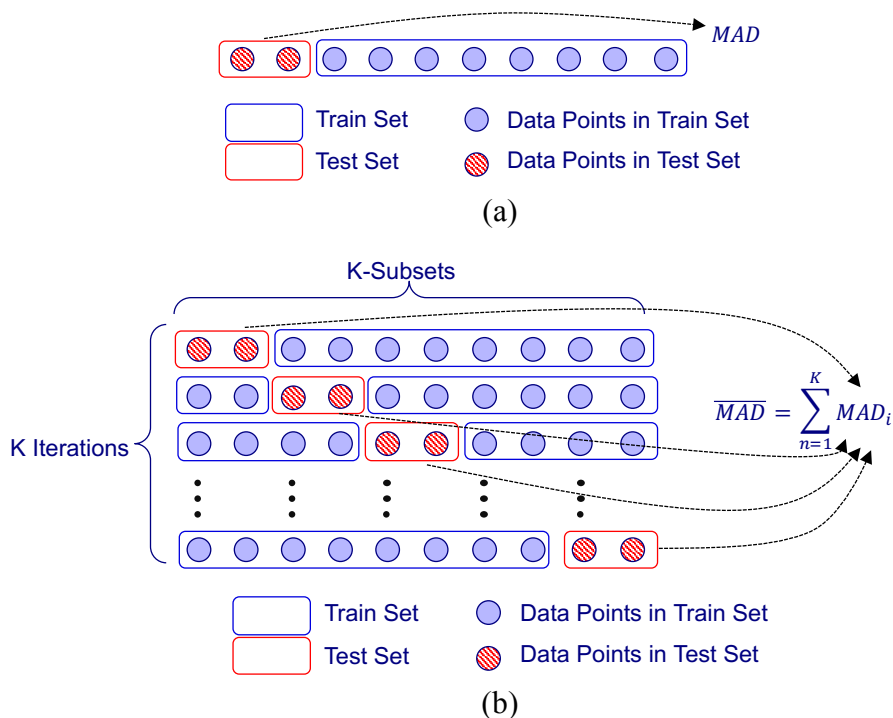

Figure 8: MAD from (a) one train-test partition, (b) cross validation

An algorithm is needed to efficiently determine the optimal LRQR penalty based on the average MADs from cross validations. Three frequently used approaches to finding the optimal penalty are random (Solis and Wets, 1981; Bergstra and Bengio, 2012), manual, and grid search (Bergstra and Bengio, 2012). A Bayesian Optimization-based searching approach (Lizotte, 2008; Snoek et al., 2012) was adopted in this study together with a grid search algorithm since it is more efficient than the other three approaches. In

this combination of searching approaches, grid search is the first step. Grid search sets a potential range of penalty values and then selects several penalty values with evenly divided intervals. These penalty values and their corresponding MADs are used as inputs to the Bayesian Optimization approach. Bayesian Optimization will use the data as initial information to start the search process. The Bayesian Optimization (Mockus, 2012; Snoek et al., 2012; Yen, 2016) is a sequential design strategy for global optimization of black-box functions. Like the other Bayesian methods, it uses the concept of prior and posterior probability. Initially, the search treats the objective function as an unknown (black-box) function and uses the currently available information to place a prior over it (Figure 9). The prior captures the beliefs about the behavior of the function based on the available information. In the example (Figure 9a), the known information is the values of 3 existing data points (p=3). After gathering the function evaluation or evaluations, which is like an additional point (p=4 now), the prior is updated to form the posterior distribution over the objective function, and the posterior distribution is used to construct an acquisition function, which determines the next point or points to explore (Figure 9.b). There are two advantages of using grid search and Bayesian optimization compared to grid or random search only. The first is it searches for the answer in a more efficient way, the second is that Bayesian Optimization tries to approximate the optimal penalty, while grid and random search do not.
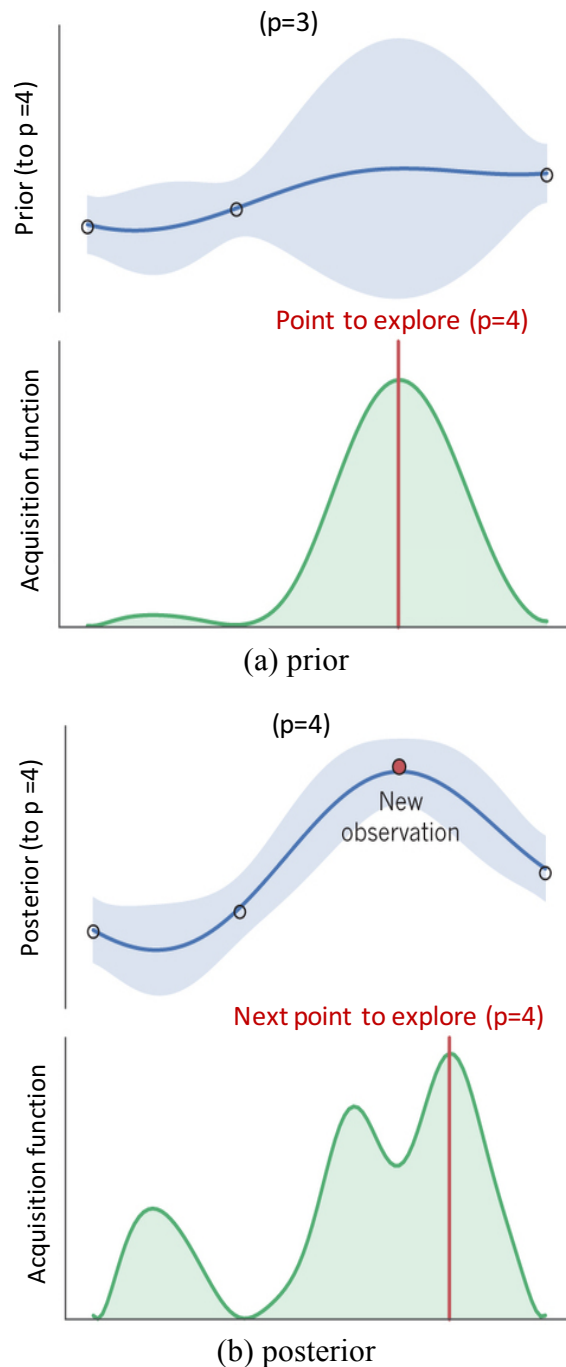
Figure 9: Example of a step of Bayesian Optimization (a) prior (b) posterior
(modified from Ghahramani, 2015)

The general framework for the parametric model construction requires the feature values and delay distribution of each train travelling on the studied single-track mainline as inputs (Figure 10). The output of the framework is a LRQR with the optimal penalty having the lowest average MAD from cross validation based on all available inputs. For initialization, the grid search evaluated a certain number of penalties with fixed intervals and obtained their corresponding MADs. This information was used as initial information

by the Bayesian Optimization to determine the next point to explore. Theoretically, the search of Bayesian Optimization can find the exact optimal penalty after infinite number of iterations but the search is set to stop arbitrarily if the best average MAD does not improve in ten iterations. After Bayesian optimization, an optimal penalty is found and it together with the feature values and delay distributions are used by LRQR to construct the final Train Delay Distribution model.
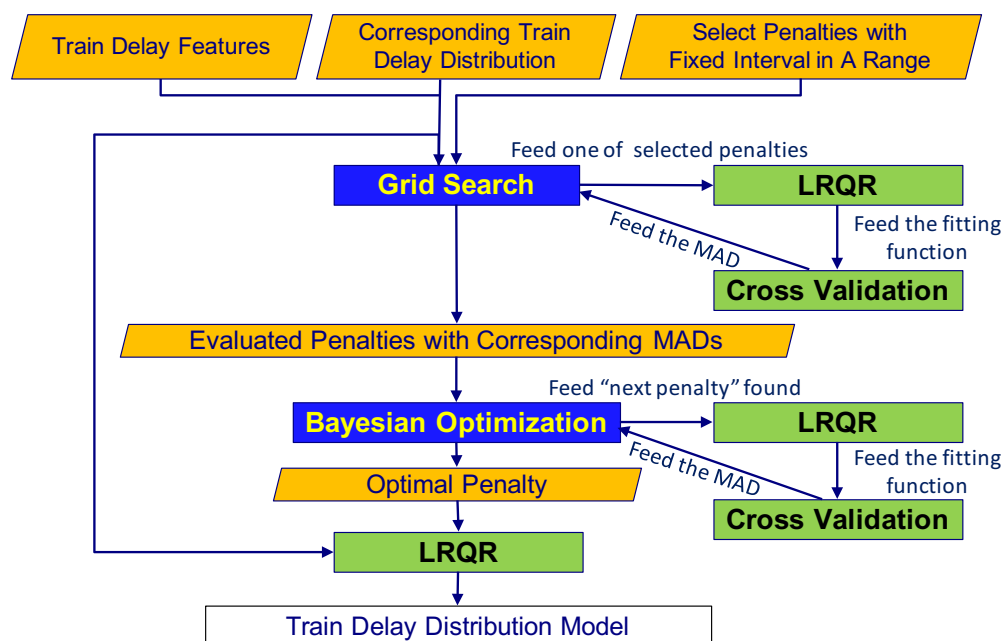

Figure 10: Framework of parametric model construction

4 Case Study

The case study demonstrates the construction of the train delay distribution model. It includes the design of an experiment matrix containing different traffic scenarios used by simulation to generate train delay input data, the calculation of the feature values for each train based on input data, the construction of the model, and the cross validation assessment of the constructed model. The changes of model coefficients and MAD values under different quantiles are displayed to show the characteristics and performance of the developed model.

A hypothetical infrastructure layout was constructed in RTC to obtain delay distribution data (Figure 11). The traffic tested on this infrastructure is assumed to have the following characteristics:

- Traffic volume varies from 8 to 26 trains per day
- Three predefined train priorities (high, medium, and low priority) exist within the traffic mixture with the percentage of each priority ranging from 25 to 75 percent
- Each predefined train priority corresponds to a certain set of characteristics, and the characteristics include train length, train weight, horsepower per ton, maximum travelling speed, and the predefined priority
- For each potential combination of traffic volume and mixture, there exist three different schedules
- The departure flexibilities for high, medium, and low train types are 0.5, 1.5, 3 hours, respectively
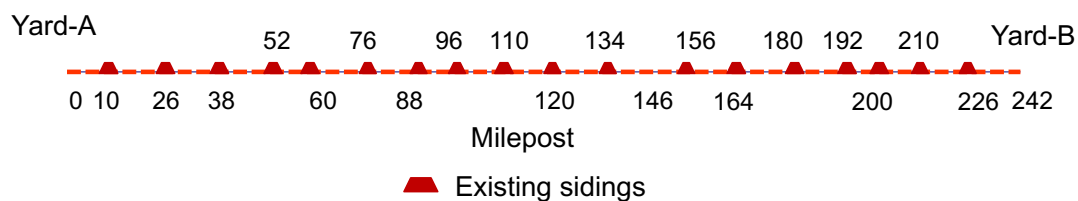
Figure 11: Tested infrastructure layout in RTC

The traffic 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 output from simulation includes traffic parameters for each scenario that were also input to the Monte Carlo process to obtain the feature values associated with each train. The calculated train delay features and their associated train delay distributions were then used as inputs to construct a quantile regression model.

The model construction process was coded in and solved by R (R Core Team, 2015). R package "rqPen" and "rBayesianOptimization" were used for LRQR and Bayesian Optimization. Package "caret" was used for cross validation. The total number of folds in the cross validation is set to 10 for this study. The MADs of different percentile from cross validation show the potential prediction errors. Table 1 shows the MADs, optimal penalties, and the mean of delay values corresponding to each percentile. For percentile 1 to 70, the MAD gradually increases from 1 to 3 min of delay. For percentiles more than 70, the MAD increases relatively rapidly from 8 to 14 mins. This means the features developed cannot predict the long-tail part of train delay distribution precisely. Developing new features which are highly related to the long-tail of train delay distribution may solve this problem. Dispatchers of single-track mainline sometimes create train fleets to facilitate the process of meeting and passing trains. This action may lead to a train or trains with higher speed and priority following a train with relatively slower speed and priority. A train meeting or passing a fleet may need to wait on a passing siding for a longer time than most normal meets with a single train. These mechanisms could be a cause of large train delays. Additional features relating to these processes could improve the ability of the parametric model to predict extreme train delays.

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

Percentile	Optimal Penalty	Mean of MADs	Mean of Delay for each Percentile
1%	4.22E-06	1.48	2.42
5%	1.86E-05	1.63	4.27
10%	3.62E-04	1.71	5.89
20%	3.77E-04	1.72	8.45
30%	3.30E-04	1.77	10.59
40%	4.99E-03	1.87	12.67
50%	2.63E-03	2.09	14.72
60%	4.55E-03	2.69	16.93
70%	2.75E-03	2.91	19.62
80%	3.89E-03	3.93	23.23
90%	1.25E-03	5.94	28.95
95%	7.52E-04	8.16	34.11
99%	3.79E-04	13.77	43.73

5   Conclusion and Future Study

Past parametric models used for train delay prediction generate the average train delay instead of a train delay distribution. The features they use to predict train delay are empirical and only capture the impact of speed and priority variations but not operating styles. This study proposed a train delay distribution model based on the concept of Lasso Regularized Quantile Regression (LRQR) and newly developed features to capture the combined impact of speed, priority variation, and operating style. A Bayesian Optimization approach was also used to determine the optimal penalty of the model under different percentiles. The cross validation result in the case study shows that the model can accurately predict the delay corresponding to different quantile for quantiles up to percentile 70. To improve the performance of the model to predict extreme train delays (exceeding the 70th percentile in the case study for example), new features must be developed based on the observations of the dispatching strategy. However, the constructed model provides predictions with an acceptable range of error for most of the delay distribution on the case study line of interest.

Acknowledgements

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*, vol. 26, pp. 7-40.

Bergstra, J., Bengio, Y., 2012. "Random search for hyper-parameter optimization", *Journal of Machine Learning Research*, vol. 13, pp. 281-305.

Burdett, R.L., and Kozan, E., 2006. "Techniques for absolute capacity determination in railways", *Transportation Research Part B: Methodological*, vol. 40, issue 8, pp. 616-632.

Chen, B., Harker, P.T., 1990. "Two moments estimation of the delay on single-track rail lines with schedule traffic", *Transportation Science*, vol. 24, issue 4, pp. 261-275.

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

Ghahramani, Z., 2015. "Probabilistic machine learning and artificial intelligence", *Nature*, vol. 521, pp. 452-459.

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

Jensen, L.W., Nielsen, O.A., Landex, A., 2015. *Robustness indicators and capacity models for railway networks*, Technical University of Denmark

Jensen, L.W., Landex, A., Nielsen, O. A., Kroon, L.G., Schmidt, M., 2017. "Strategic assessment of capacity consumption in railway networks: Framework and model", *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 126-149.

Koenker, R., Bassett, G., 1978. "Regression quantiles", *Econometrica*, vol. 46, pp. 33-50.

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

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

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*, vol. 2159, pp. 69-76.

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

Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., Li, Y., 2015. "Robust causal dependence mining in big data network and its application to traffic flow predictions", *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 292-307.

Lizotte, D. J., 2008. *Practical bayesian optimization*, University of Alberta.

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

Marković, N., Milinković, S., Tikhonov, K. S., Schonfeld, P., 2015. "Analyzing passenger train arrival delays with support vector regres*sion", Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 251-262.

Mitra, S., Tolliver, D., 2010. "Estimation of railroad capacity using parametric methods", *Journal of the Transportation Research Forum*, vol. 49, issue 2, pp. 111-126.

Mockus, J., 2012. *Bayesian approach to global optimization: theory and applications*, Springer Science & Business Media.

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*, vol. 46, pp. 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*, vol. 26, pp. 169-182.

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

Peng, B., Wang, L., 2015. "An Iterative Coordinate Descent Algorithm for High-Dimensional Nonconvex Penalized Quantile Regression", J*ournal of Computational and Graphical Statistics*, vol. 24, pp. 676-694.

R Core Team. 2015. An Introduction to R. URL https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf. Accessed: 2016-03-14.

Sherwood, B., Maidman, A., Sherwood, M.B., ByteCompile, T.R.U.E., 2016. "Package 'rqPen', version 1.5.1.", Retrieved December 18, 2016, from Retrieved December 18, 2016, from https://cran.r-project.org/web/packages/rqPen/rqPen.pdf

Snoek, J., Larochelle, H., Adams, R.P., 2012. "Practical bayesian optimization of machine learning algorithms", In *Advances in neural information processing systems*, pp. 2951-2959.

Solis, F.J., Wets, R.J.B., 1981. "Minimization by random search techniques", *Mathematics of Operations Research,* vol. 6, issue 1, pp. 19-30.

Wu, Y., Liu, Y., 2009. "Variable selection in quantile regression", *Statistica Sinica*, vol. 19, pp. 801-817.

Yen, Y., 2016. "Package 'rBayesianOptimization', version 1.1.0", Retrieved December 18, 2016, from Retrieved December 18, 2016, from https://cran.r-project.org/web/packages/rBayesianOptimization/ rBayesianOptimization.pdf