2014 IEEE International Conference on Data Mining

Bus Travel Time Predictions Using Additive Models

Matthías Kormáksson, Luciano Barbosa, Marcos R. Vieira, and Bianca Zadrozny
IBM Research – Brazil
Email: {matkorm, lucianoa, mvieira, biancaz}@br.ibm.com

***Abstract*—Many factors can affect the predictability of public bus services such as traffic, weather, day of week, and hour of day. However, the exact nature of such relationships between travel times and predictor variables is, in most situations, not known. In this paper we develop a framework that allows for flexible modeling of bus travel times through the use of Additive Models. The proposed class of models provides a principled statistical framework that is highly flexible in terms of model building. The experimental results demonstrate uniformly superior performance of our best model as compared to previous prediction methods when applied to a very large GPS data set obtained from buses operating in the city of Rio de Janeiro.**

I. Introduction

In this paper we are concerned with the problem of predicting bus travel/arrival times using GPS data from public buses. The main challenge in performing this task arises from the fact that GPS data only provide snapshots of bus locations at predefined (or in some cases irregular) time stamps. The observed GPS coordinates are therefore necessarily irregular in space as signal transmissions are not controlled with respect to bus locations. The difficulty of the problem is further increased when difference between time stamps is large.

The GPS data permit us to study the relationship between bus movements in time and space. However, other factors such as day of week, hour of day, and current traffic conditions may also influence travel times in some systematic way. The exact nature of such relationships between travel times and predictor variables is usually not known. Therefore, these factors need to be incorporated into prediciton algorithms either indirectly through binned analyses or through direct modeling.

We propose to model travel times using Additive Models [1], [2], which provide a principled statistical framework for arrival time predictions. In particular, we model cumulative travel time as a smooth function of route location and further allow this functional relationship to vary smoothly across time. We also construct features that may seemlessly be incorporated into the Additive Model, either as direct main effects or interaction effects in conjunction with other variables.

Previous approaches have used a mixture of statistical and machine learning models for predicting bus travel times. These methods fall broadly into four categories: *Historical Data-Based Models*, [3], *Kalman Filter Models*, [4], *Artificial Neural Network Models*, [5], and *Support Vector Regression Models*, [6]. Due to space limitations we point to [7], [8] for a more comprehensive list (and discussion) of related works.

A recurring problem in previous approaches is that they assume knowledge of travel times between fixed locations in space, in particular bus stops. In the absence of additional information, interpolation is performed to infer these times at the route's bus stops [3]. This is reasonable when difference between time stamps is small, say 20 seconds, but can lead to larger errors when difference is larger, say few minutes. Another problem arises for methods that account for temporal effects (e.g., Kalman filters) due to discretization made in the time dimension. This is again reasonable in the presence of high volumes of data, but may be problematic if data is sparse with irregularities in the time dimension.

The main advantage of Additive Models in this context is their ease of interpretability and flexibility in modeling complex non-linear relationships. Factors that are known (or suspected) to affect traffic may be included in the model as traditional linear features, smooth functional effects, or interactions thereof. Additive Models do not require any discretization or interpolated observations, but rather are capable of handling directly the raw observed data. The only interpolation that applies is made when inferring the departure time from origin. However, a critical feature of our proposed solution is the inclusion of a (corrective) random intercept in the model that attempts to correct for this interpolation step thus redefining time zero for each bus.

To the best of our knowledge our proposed solution is the first method that: **(1)** models bus travel times directly using raw irregular GPS data; **(2)** models spatial and temporal effects through smooth functions thus avoiding any discretization; and **(3)** allows for flexible incorporation of additional traffic related features in a model based manner. The last point is an important one as it implies that our proposed framework may be used as a development framework for building more accurate travel time models through the incorporation of additional (perhaps city dependent) features.

The remainder of the paper is organized as follows: Section II describes the motivating data; background on additive models is provided in Section III; the proposed solution is detailed in Section IV; experimental evaluation is provided in Section V; and Section VI concludes the paper.

II. Preliminaries

A. Motivating Data

The motivating data consist of GPS measurements collected from public buses in the city of Rio de Janeiro, Brazil, during the time period from September 26, 2013 to January 9, 2014. The complete data set contains information about more than 400 bus routes and 9000 buses. Each GPS data point contains information about the position of the bus (longitude, latitude), date and time stamp, bus ID, and route ID. In total there are more than 100 million location entries for the time period of this study. The time between consecutive GPS measurements ranges from anywhere under a minute to over 10

1550-4786/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDM.2014.107

IEEE computer society

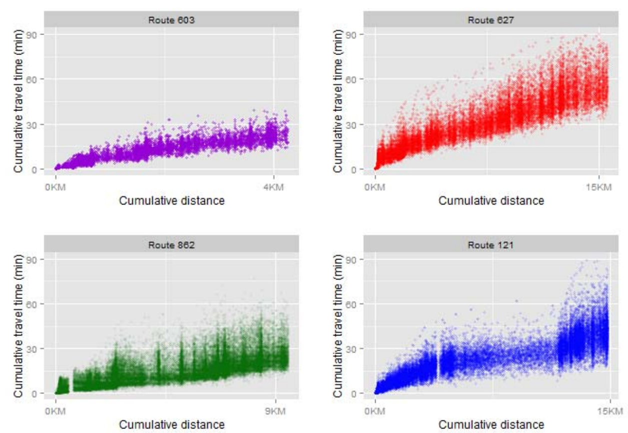

Fig. 1. Cumulative space-time trajectories of the four bus routes analyzed in this paper. Note the distinct x-scales that reflect the different route lengths.

minutes, with an average of ≈ 4 minutes. A sequence of GPS coordinates of a given bus is called a *space-time trajectory* and provides information about bus movement in space and time.

We also had access to GTFS (General Transit Feed Specification[1]) data, which contain general information about the bus routes, such as bus stop locations. In general, each route consists of two trips, one going from origin to destination and the second representing the return trip. The GTFS data contain a complete definition of each such trip as a sequence of latitude/longitude points tracing the streets of the route from origin to destination.

B. Data Normalization

The GPS data in conjunction with the GTFS data provide us with the means to map GPS coordinates onto a 1-dimensional scale measuring distance from origin. For any given bus coordinate we project it onto the closest line segment of the corresponding route and then calculate its distance from origin along the piecewise segments.

By calculating differences between consecutive time stamps we may infer travel times of each bus between its observed locations. However, in order to analyze and compare travel times of buses, running at different hours, we need to normalize the time stamps onto a common cumulative time scale, i.e., we need to define a common time zero. This may be achieved by interpolating all the observed space-time trajectories at a common fixed point in space, e.g., origin, and defining that point as time zero. Space-time trajectories whose GPS coordinates have been mapped onto a cumulative distance scale and whose time stamps have been normalized to a common cumulative time scale are called *cumulative space-time trajectories*.

In Figure 1, we see the cumulative space-time trajectories of all buses (during the specified time period) running on the four routes analyzed in this paper. Note that the only interpolation made is at origin to define the common cumulative time scale. In all other aspects, the scatter plots represent raw measurements observed at irregular spatial locations.

[1] http://developers.google.com/transit/gtfs

C. Mathematical Notation

In general, we may normalize the time stamps at any arbitrary fixed point in space, in particular, at any of the route's bus stops. Let $0 = p_0 < p_1 < \cdots < p_K$ denote the distances of all bus stops of a given route from origin p_0, where K denotes the number of on-route bus stops. *Cumulative space-time trajectories normalized at* p_k consist of cumulative distances $p_k \leq dist_{ijk} \leq p_K$, and corresponding cumulative travel times $T_{ijk} \geq 0$, $j = 1, \ldots, m_{ik}$, where m_{ik} denotes the number of data points for bus i beyond p_k. The distances may either represent interpolated values at prespecified fixed locations (e.g., subsequent bus stops) such as in [3], or raw GPS coordinates as in this paper. Both cumulative distances and cumulative times are defined from p_k onward such that *dist* $= 0$ and $T = 0$ at p_k. The cumulative time scale is inferred by interpolating two consecutive time stamps before and after p_k. We denote by *Traj*(p_k) the set of thus normalized historical cumulative space-time trajectories.

III. Theoretical Background

Additive models [1], [2] are linear models, which allow the linear predictor to not only depend on pure linear terms but also on a sum of unknown smooth functions of predictor variables. In this section we discuss the two classes of additive models that we consider in this paper.

A. Additive Models

The additive models that we consider have the form:

$$y_i = X_{0i}\beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{1i}, x_{2i}) + \varepsilon_i, \quad (1)$$

where y_i and ε_i are the response and error term respectively, $X_{0i}\beta_0$ represents purely linear terms in the model, and f_1, f_2, and f_3 represent smooth functions of the predictors x_1 and x_2. We represent the one-dimensional functions as $f_1(x_1) = \sum_{j=1}^{q_1} \beta_{1j}\phi_j(x_1)$ and $f_2(x_2) = \sum_{k=1}^{q_2} \beta_{2k}\psi_k(x_2)$, where $\phi_j(x_1)$ and $\psi_k(x_2)$ are known (possibly distinct) basis functions. In this paper a tensor product basis [2] is used to represent the bivariate term:

$$f_3(x_1, x_2) = \sum_{j=1}^{q_1}\sum_{k=1}^{q_2} \beta_{3jk}\phi_j(x_1)\psi_k(x_2). \quad (2)$$

Each of the above functions f_i, $i = 1, 2, 3$, may be represented linearly as, $X_i\beta_i$, by appropriately specifying the X_i matrices in terms of the basis functions $\phi_j(\cdot)$, and $\psi_k(\cdot)$. The model terms may then be stacked in the traditional way: $X = [X_0\ X_1\ X_2\ X_3]$, and $\beta = (\beta_0', \beta_1', \beta_2', \beta_3')'$, to obtain the linear model:

$$Y = X\beta + \varepsilon. \quad (3)$$

The above model may be estimated by maximizing the loglikelihood function, $\ell(\beta)$, under a normality assumption on ε. However, in order to control the smoothness of the fit we need to work with the so called penalized loglikelihood

$$\ell_P = \ell(\beta) - \sum_{i=1}^{3} \lambda_i \beta_i' D_i \beta_i, \quad (4)$$

where D_i are matrices of known coefficients and λ_i are function specific smoothness parameters. Estimation of the

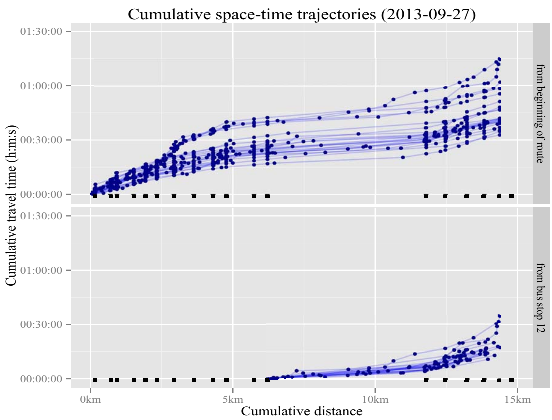

Fig. 2. Cumulative space-time trajectories for route 121 from Sep 27, 2013. Upper panel demonstrates the trajectories from beginning of route and the lower panel from bus stop p_{12} onward. Bus stops are marked by black squares.

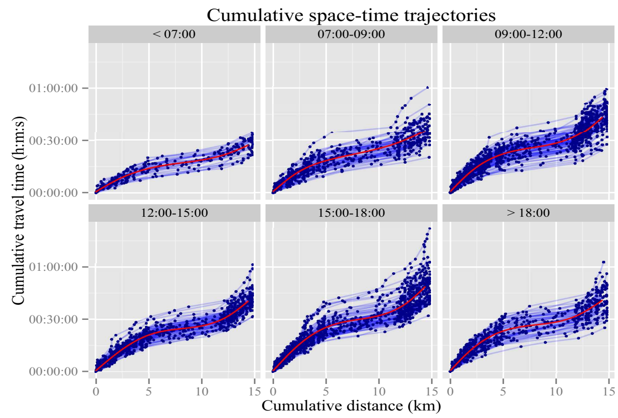

Fig. 3. Cumulative space-time trajectories of route 121, stratified by hour, during Sep 26 - Oct 10, 2013. A smooth mean curve for each time category is depicted in red.

additive model may be performed by maximizing the penalized loglikelihood in (4) and estimating the smoothness parameters through GCV. Once the model has been estimated using training data, one can predict a new response in the usual manner.

B. Additive Model with Random Intercept

In this paper we also consider an additive model with a random intercept

$$y_i = b_{0i} + X_{0i}\beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{1i}, x_{2i}) + \varepsilon_i, \quad (5)$$

where $b_{0i} \sim N(0, \sigma_b^2)$ and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Note that the above model is not overparametrized as the b_{0i} are treated as random and not fixed. This model falls into the general class of Additive Mixed Models [2] (due to the mixed combination of random and fixed model terms) and we note that by specifying the smooth functions as before it may be represented in the matrix form:

$$Y = X\beta + Zb_0 + \varepsilon, \quad (6)$$

where Z is a single column matrix of ones.

The estimation of the above model is not straight forward and since space is limited we point to [2] for full theoretical coverage. However, we note that through a correct likelihood specification an iterative maximization algorithm may be applied to obtain estimates of the random intercept b_{0i} and the parameters β, σ_b^2, and σ_ε^2.

IV. Proposed Solution

In this section we present additive models for analyzing historical cumulative space-time trajectories such as those observed in Figure 2. In the upper and lower panel, respectively, we see examples of historical trajectories, $\mathit{Traj}(p_k)$, that have been normalized at p_k for $k = 0$ (origin) and $k = 12$ (bus stop p_{12}). We note that the cumulative travel time variance beyond bus stop p_{12} is reduced dramatically when normalized at p_{12} as compared to at p_0. Therefore, we propose to train additive models on each of the historical trajectories, $\mathit{Traj}(p_k)$, for bus stops $k = 0, \ldots, K-1$, where $K-1$ corresponds to the second to last bus stop on route. The objective is then to base future travel time predictions of a bus close to bus stop p_k on the corresponding additive model trained on $\mathit{Traj}(p_k)$.

In order to make our presentation more coherent, we model and analyze the bus trajectories of bus route 121 for the first two weeks of our observed time period. We analyze trajectories starting from origin, $\mathit{Traj}(p_0)$ and for ease of notation we omit the k-subindex of Subsection II-C. However, we note that all discussions generalize to trajectories starting from any given bus stop along the route, $\mathit{Traj}(p_k)$. For the first two weeks of our study we observed $n = 385$ trajectories for route 121 with on average $m_i = 13$ measurements per bus ride. Through statistical reasoning, we construct three models whose performances are compared to previous approaches (Section V). All numerical summaries in this section apply to this data set.

A. Basic Additive Model for Travel Times

In Figure 3, we see the cumulative space-time trajectories of all bus rides of route 121 during the specified time period. The trajectories are stratified by hour and a smooth mean curve is fitted through each scatterplot to illustrate travel time trends. We note that morning travel time duration peaks between 9am and noon (morning rush hour). Then a slight reduction in travel times is observed between noon and 3pm, followed by an afternoon rush hour. We also note that there is not only a difference in total travel times across hours, but also in the shapes of the mean curves. This figure inspires the following model of bus travel time, T_{ij}, as function of distance from origin, dist_{ij}, and time of departure, time_i:

Model 1: Basic Additive Model (BAM)

$$T_{ij} = \beta_0 + f_1(\mathit{dist}_{ij}) + f_2(\mathit{time}_i) + f_3(\mathit{dist}_{ij}, \mathit{time}_i) + \varepsilon_{ij},$$

$i = 1, \ldots, n$, and $j = 1, \ldots, m_i$, where β_0, and ε_{ij} represent an overall model mean and error term, respectively. In what follows we assume the random error terms are mean zero and normally distributed. The functions f_1, f_2 and f_3 were represented by cubic regression spline basis functions (see [2] for definition and theory) and tensor product smooths (see

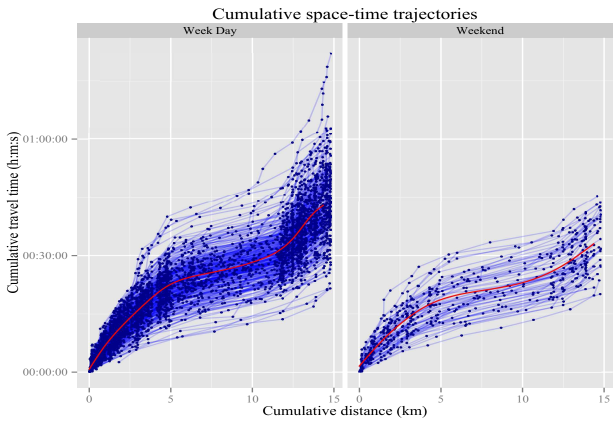

Fig. 4. Cumulative space-time trajectories of route 121 during week days (left) and weekends (right). The dots represent raw measurements and the blue interpolated curves represent each bus trajectory for illustration purposes. Smooth mean curves are depicted in red.

Section III). We placed one spline knot at each bus stop between origin and destination to capture smooth transitions from one station to the next. We placed 5 equally spaced spline knots in the time space, which was large enough to capture the two rush hours trends in the morning and afternoon, respectively. Larger number of time-knots did not seem to affect the fit of the model.

Each of the functional effects f_1, f_2, and f_3 was deemed statistically significant by the F-test (p-values $< 10^{-16}$) and the overall adjusted R^2 of the model was 0.903.

B. Extended Additive Model with Additional Features

To account for differences in mean travel times between weekdays and weekends (see Figure 4) we include in the model the indicator variable $weekend_i$ that determines whether bus ride i occurred on a weekday or on the weekend. However, there is also an evident interaction of the weekend factor with distance from origin since the mean difference between weekday and weekend travel times increases as a function of distance. Therefore, in addition we propose replacing the functional term $f_1(dist_{ij})$ in Model 1 by the interaction term $f_1(dist_{ij}, weekend_i)$. This term in fact generates two different smooths, one for weekday and the other for weekend trajectories.

Another feature that intuitively seems likely to correlate well with travel time of a given bus is the travel time of the last bus in front of it. We therefore define the feature T_{ij}^{last} to be the cumulative travel time at $dist_{ij}$ of the last bus that passed before time of departure of bus i. It is important to point out here that it is unlikely that the last bus will transmit a GPS signal at the same locations $dist_{ij}$ as bus i. Therefore, interpolation of the cumulative space-time trajectory of the last bus is performed at $dist_{ij}$ to construct the feature T_{ij}^{last}.

The following model extends Model 1 to include the features discussed above:

TABLE I. ROUTE DATA SUMMARY

Route	# trajectories	# stops	Length (in km)
603	1,276	15	4
627	1,325	54	15
862	7,882	24	10
121	2,515	18	15

Model 2: Extended Additive Model (EAM)

$$T_{ij} = \beta_0 + \beta_1 \cdot weekend_i + f_1(dist_{ij}, weekend_i) + \beta_2 \cdot T_{ij}^{\text{last}} + f_2(time_i) + f_3(dist_{ij}, time_i) + \varepsilon_{ij}.$$

We fitted the above model to our data set and observed that all effects, including $weekend_i$, the interaction term $f_1(dist_{ij}, weekend_i)$, and the linear predictor T_{ij}^{last} were highly significant (p-values $< 10^{-16}$). The adjusted R^2 increased from the previous model to 0.919.

C. Additive Mixed Model

We recall that in Section II-C we normalized all the space-time trajectories to a common cumulative time scale. Since the actual times of departure are not known, an approximation was made by taking two consecutive time stamps before and after origin and defining $T = 0$ as the interpolated time-stamp at origin. However, note that this introduces an error in the form of a vertical trajectory shift, due to incorrect specification of time of departure.

In order to correct for the misspecification of time of departure, we propose an additive mixed model (see section III-B) that includes a (corrective) random intercept term b_{0i} for each and every bus ride $i = 1, \ldots, n$:

Model 3: Additive Mixed Model (AMM)

$$T_{ij} = \beta_0 + b_{0i} + \beta_1 \cdot weekend_i + f_1(dist_{ij}, weekend_i) + \beta_2 \cdot T_{ij}^{\text{last}} + f_2(time_i) + f_3(dist_{ij}, time_i) + \varepsilon_{ij},$$

where $b_{0i} \sim N(0, \sigma_b^2)$.

We fitted the above model and found that the random intercept term b_{0i} was indeed highly significant (p-value $< 10^{-16}$) and the adjusted R^2 value increased significantly to 0.968. The standard deviation σ_b was estimated to be 3 minutes, which indicates that the estimated (interpolated) time of departure indeed requires adjustment.

V. EXPERIMENTS

A. Experimental Data

We performed a prediction analysis on four bus routes in the city of Rio de Janeiro. These routes are located in distinct regions of the city and further have different lengths, number of bus stops, and frequency of bus rides, see Table I. Further these routes demonstrate distinct traffic patterns as can be seen in Figure 1. These four distinct routes represent a wide range of prediction scenarios we want to cover in our experiments. We have made all the data sets available in [9], as we believe this will stimulate further research in the area.

B. Experimental Setup

The total number of trajectories in each route is presented in Table I. We randomly selected 14 days after November 1st 2013 as our test data. This guaranteed at least 30 days of historical data for each test date. For each bus running on any of these 14 days we performed travel time predictions using three sets of historical data involving all bus rides in the last 10, 20, and 30 days, respectively. This was done to get a sense of whether the size of the historical data set has an influence on the accuracy of the tested models.

Travel time predictions were made for each bus in test set from every bus stop until end of route to reflect the real world problem of predicting bus arrivals from any on-route location onward. More precisely, for each bus stop p_k we recorded for bus i in test set the first observed bus entry $dist_{i1k}$ after p_k. We then made travel time predictions at all remaining (observed) points $dist_{ijk}$, $j = 2, \ldots, m_{ik}$. Since the data at $dist_{ijk}$ represent the raw data whose cumulative travel times T_{ijk} (from bus stop p_k) are known we could thus calculate and analyze prediction errors. In order to get a sense of how error changes as a function of distance from the bus stop beyond which predictions were made we recorded the prediction distances $|dist_{ijk} - p_k|$.

C. Evaluation Measures

To evaluate overall performance of each method for a given bus route we calculated the *mean absolute relative error*, defined as $(1/N)\sum_{ij}|T_{ij} - \hat{T}_{ij}|/T_{ij}$, where N denotes total number of predictions made. We performed a non-parametric paired Wilcoxon test to compare the overall performances between methods.

Since error was greater at later parts of route, we also analyzed the distributions of absolute errors stratified by prediction distances, $|dist_{ijk} - p_k|$. The distance space was binned into one kilometer bins [0,1), [1,2), [2,3), ... etc. Visual comparison of distributions was performed using boxplots and a 95th percentile curve; see Figure 5. Since absolute errors were right skewed for each method we performed a two-sided non-parametric paired Wilcoxon test to compare methods within each distance bin. To account for multiple testing, p-values were recorded for each comparison and then adjusted using the Benjamini-Hochberg method. Statistical significance was determined if adjusted p-values were < 0.05.

D. Implemented Methods

Additive Models: No model selection or parameter tuning was performed during the training. Instead for each and every training set we estimated the exact same three models as defined in Section IV. Once estimation had been performed the estimated model parameters, $\hat{\beta}$, along with a complete set of test features was plugged into the Additive Model formulas to obtain travel time predictions at subsequent route locations. For AMM, in order to estimate the random effect b_{0i} of (5) for a new trajectory i in the test set at least one observed travel time is needed. Since predictions are always made given the current location of the bus, the first observation, T_{i1k}, may be used for that purpose. Implementation was performed using the *mgcv* R package, [2].

TABLE II. MEAN ABSOLUTE RELATIVE ERROR

		Method				
Route	# days	BAM	EAM	AMM	Kernel	SVM
603	10	19.9%	19.7%	**18.4%**	21.3%	64.4%
	20	20.1%	19.8%	**18.5%**	21.3%	64.7%
	30	19.8%	19.6%	**18.3%**	21.3%	64.8%
627	10	16.3%	14.7%	**13.8%**	18.1%	28.8%
	20	15.2%	14.2%	**13.4%**	17.3%	29.1%
	30	15.1%	14.0%	**13.2%**	17.1%	28.8%
862	10	22.1%	19.5%	**18.0%**	23.8%	26.4%
	20	22.5%	19.3%	**18.0%**	23.6%	26.8%
	30	22.2%	19.3%	**17.9%**	23.4%	26.8%
121	10	23.1%	20.9%	**19.2%**	23.9%	41.5%
	20	22.9%	20.7%	**19.1%**	23.6%	41.4%
	30	22.7%	20.3%	**18.9%**	23.4%	41.2%

Support Vector Machine (SVM): Bin et al. [6] used SVM regression to predict the arrival time of the next bus. They divided the bus trajectories in segments and then used as features the travel time of current bus at previous segment and the latest travel time of a previous bus in the next segment to predict the travel time for the next segment. Since we are not only interested in predicting the travel time of the next segment but all subsequent segments until end of route, we added to the training data the latest travel times at all subsequent segments. Similar to [6], we used a linear kernel and the implementation was performed using the R package "e1071".

Kernel Regression: Sinn et al. [3] proposed an instance-based method that uses weighted averages of historical trajectories to make predictions. Trajectories with similar behaviour up to the current bus location are given more weight. Weights are defined by a gaussian kernel: $\exp(-\|x - y\|^2/b)$, where x and y are cumulative space-time trajectories, and b is the bandwidth of the kernel. Similar to [3], we set $b = 1$ in our experiments. For further details, we refer the reader to [3].

E. Experimental Results

In Table II we see the mean absolute relative errors for each method. The first thing to note is that our Additive Models (BAM, EAM, and AMM) outperformed the Kernel Regression and SVM in all scenarios. SVM's overall performance was notably worse than any of the other methods. The main comparisons of interest are thus between the Kernel Regression approach and each one of our Additive Models. The Wilcoxon paired test revealed statistically significant differences between the Kernel Regression and all our proposed three Additive Models, in all scenarios. Further, the Wilcoxon paired test revealed that in all scenarios the AMM outperformed all other methods. Another observation from Table II is that the size of the training data does not seem to affect performance of any of the 5 methods.

To give a more detailed view of the results, we show in Figure 5 boxplots of absolute prediction errors across all training data sets, 10, 20, and 30 days, and further stratified by route and prediction distance. Boxplots are displayed for all methods except for SVM as their performance was greatly inferior for larger distances and only interfered with visualization.

As expected, the error increases with distance from bus stops beyond which the predictions were made. We note that

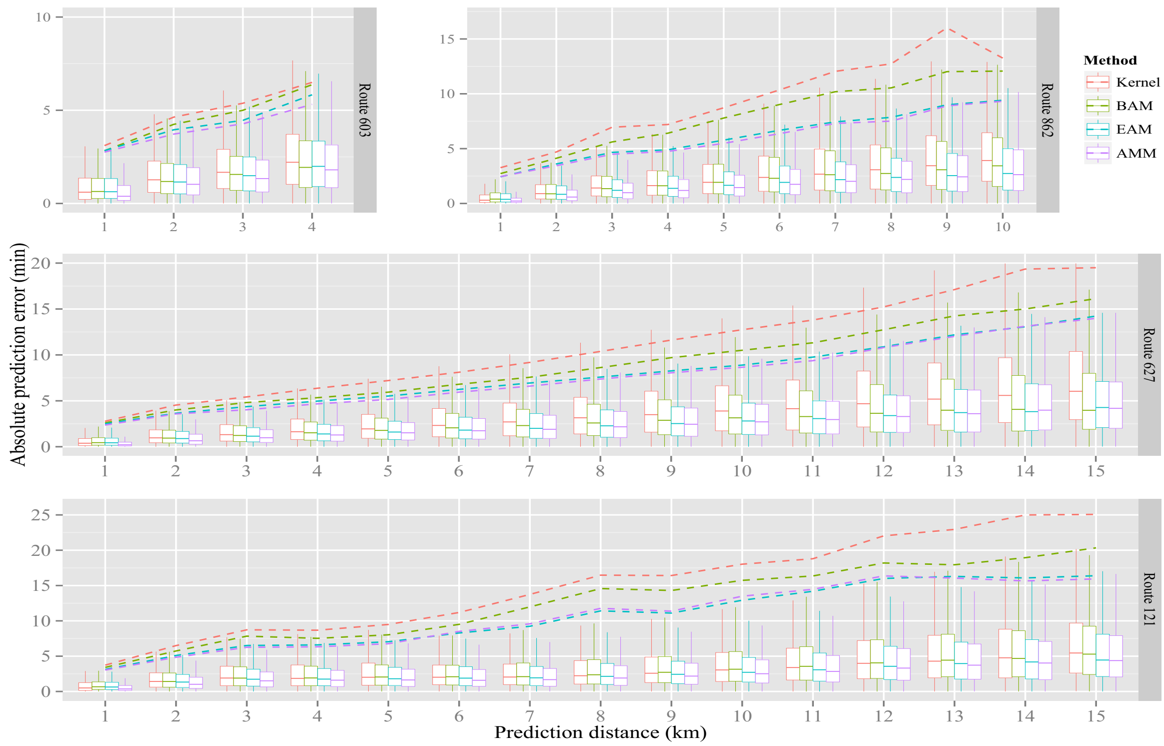

Fig. 5. Boxplots of absolute prediction errors. The dashed lines represent 95th percentiles of corresponding absolute errors.

all distributions were right skewed with several outliers (as defined by the boxplot whiskers) mostly due to heavily delayed buses. These outliers are not displayed as they interfere with visualization and do not reveal any significant trends beyond those seen in the boxplots. However, in order to get a sense of this "outlier effect" we plotted the 95th percentiles (dashed lines) along with the boxplots. These lines give us a sense of "worst case" scenario performance of each method.

Although perhaps not visually striking in all distance bins, the AMM statistically outperformed all methods for all routes and in all distance bins (except for the 14km bin on route 627 where no difference existed between EAM and AMM). In the first distance bin, $[0, 1)$, the Kernel Regression method outperformed both BAM and EAM at all routes except for route 603 (where no statistical difference existed). However, in all other distance bins the two Additive Models statistically outperformed the Kernel Regression. Thus, on the whole, the visualization and stratified analysis confirmed the performance order observed in Table II.

The fact that Kernel Regression outperformed BAM, and EAM in the first distance bin suggests that the Additive Models tend to put more priority on minimizing error in later parts, when it is in fact larger, at the expense of short term predictions. Perhaps this may be fixed by placing more knots at the beginning of the route or through additional features. However, as we discussed before, AMM performed statistically better than all other methods in the first distance bin. This suggests that the random corrective intercept term plays an important roll in rescuing the incorrectly specified cumulative time scales as obtained by interpolation.

VI. Conclusions

In this paper we discussed the problem of predicting travel times of public buses based on GPS data. We proposed Additive Models as a flexible and a statistically principled framework for model building. We modelled cumulative travel time as a sum of linear terms and smooth functions of predictor variables. We showed that by including a random intercept in the model we were able to correct for an interpolation error incurred when normalizing space-time trajectories onto a cumulative time scale. We demonstrated on a large real-world GPS data that our proposed Additive Models achieved superior performance as compared to other existing prediction methods.

References

[1] T. Hastie and R. Tibshirani, *Generalized Additive Models*, Chapman & Hall, 1990.

[2] S. N. Wood, *Generalized Additive Models: an introduction with R*, Chapman & Hall/CRC, 2006.

[3] M. Sinn, J. W. Yoon, F. Calabrese, and E. Bouillet, "Predicting arrival times of buses using real-time GPS measurements," in *ICITS*, 2012.

[4] L. Vanajakshi, S. C. Subramanian, and R. Sivanandan, "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses," *Intelligent Transport Systems*, 2009.

[5] S. Chien, Y. Ding, and C. Wei, "Dynamic bus arrival time prediction with artificial neural networks," *J. of Transportation Engineering*, 2002.

[6] Y. Bin, Y. Zhongzhen, and Y. Baozhen, "Bus arrival time prediction using support vector machines," *J. of Int. Transp. Systems*, 2006.

[7] B. Yu, W. H. K. Lam, and M. L. Tam, "Bus arrival time prediction at bus stop with multiple routes," *Transportation Research Part C*, 2011.

[8] R. Rajbhandari, *Bus Arrival Time Prediction Using Stochastic Time Series and Markov Chains*, New Jersey Institute of Technology, 2005.

[9] Project Webpage, "http://goo.gl/HIuSPX," 2014.