SIMULTANEOUS ESTIMATION OF FLEXIBLE MODELS AND ASSOCIATED HYPERPARAMETERS: AN APPLICATION TO TRAVEL ACTIVITY-DURATION MODELING

Daniel Emaasit
Doctoral Candidate
Transportation Research Center
University of Nevada, Las Vegas
Howard R. Hughes College of Engineering
Department of Civil & Environmental Engineering
4505 Maryland Parkway, PO Box 454007
Las Vegas, NV 89154-4007
emaasit@unlv.nevada.edu

Alexander Paz, Ph.D., P.E.
Associate Professor
Director, Transportation Research Center
University of Nevada, Las Vegas
Howard R. Hughes College of Engineering
Department of Civil & Environmental Engineering
4505 Maryland Parkway, PO Box 454007
Las Vegas, NV 89154-4007
Office: (702) 895-0571
Fax: (702) 895-3936
apaz@unlv.edu

Word Count: 5512 words + 3 figure(s) + 4 table(s) = 7262 words

Submission Date: August 1, 2017
ABSTRACT

Estimation of flexible-statistical models of travel demand involves tuning varying parameters, hyperparameters, manually and iteratively. Proper tuning of hyperparameters results in superior models. However, considerable expertise including technical knowledge of statistics, data mining or machine learning and experience are required to tune hyperparameters and consequently generate appropriate models. Moreover, tuning hyperparameters is prone to subjective error and consequently produces travel demand models that are difficult to reproduce, and extend; and makes the development more an art than a science. There is a need for methods to reduce or eliminate subjectivity during the tuning process. This study proposed a framework to reduce subjectivity during the tuning of hyperparameters required for the estimation of nonparametric models of activity-duration. That is, a flexible-statistical framework, which leverages state-of-the-art innovations in Bayesian optimization (BO), was proposed to estimate Gaussian process models of activity duration and associated hyperparameters. The framework was applied to estimate duration models for five types of out-of-home non-mandatory activity episodes for household individuals in the greater Los Angeles area. Experiments demonstrate that the accuracy of results from the proposed framework are superior than those from the state-of-art tuning process, and are obtained in a fraction of time. The proposed framework could potentially increase the productivity of modelers by reducing time required to tune hyperparameters.
INTRODUCTION

The life cycle of a typical travel demand modeling process requires upgrades to multiple statistical analyses to capture changes in travel behavior patterns. These statistical analyses are tedious and involve an iterative procedure that requires expert knowledge and time to perform several tasks including but not limited to preprocessing data, selecting appropriate attributes, selecting appropriate models, finding model parameters, and critically analyzing results (1). The procedure is governed by multiple assumptions such as functional forms. In general, models that require weak assumptions about the phenomena under study are preferred. Models estimated under weak assumptions are termed in the literature as flexible and the data drives their estimation (2). Flexible models are also known as nonparametric and have demonstrated to have superior predictive performance compared to traditional econometric methods (3) (Emaasit and Paz, unpublished).

Estimation of flexible-statistical models of travel demand (3) involves tuning varying parameters, hyperparameters, manually and iteratively. Hyperparameters enable the estimation of different models for special-purpose applications and datasets. Typically, the values for these hyperparameters are specified by an analyst. An example includes a penalty hyperparameter to control the size of coefficients in a linear ridge regression model. This hyperparameter is different from model parameters, the coefficients in the linear regression model that are estimated by a search algorithm. Considering that there is no general optimizer, all optimization models use hyperparameters that require tuning. Often, optimization algorithms are selected based on how involved is the tuning of their hyperparameters for the solution of a particular problem (4).

Values for the hyperparameters are set by a modeler to initialize and estimate a flexible-statistical model of travel demand (3). Proper tuning of hyperparameters results in superior models (5). However, considerable expertise including technical knowledge of statistics, data mining or machine learning and experience are required to tune hyperparameters and consequently select appropriate models. Tuning hyperparameters is prone to subjective error and consequently produces travel demand models that are difficult to reproduce, and extend; and makes the development more an art than a science (5). Hence, there is a need for methods to reduce or eliminate subjectivity during the tuning process.

Estimation of travel demand models (TDMs) results in costly functions that require significant amounts of computation resources and time. Involving the modeler in the tuning of hyperparameters is computationally too expensive because even a single run may take days or weeks to complete. This is a major limitation in cases with large amounts of data. Using a subset of a large dataset to produce models having superior performance is more computation- and time-efficient than using the entire large dataset to produce models having similar performance. Therefore, there is a need for methods that are data efficient such that only required data points are used to obtain adequate performing models.

Currently, there is increasing use of emerging methods for simultaneous hyperparameter tuning and model selection in scientific fields, including biology (6), natural language processing (7), robotics (8), autonomous vehicles (9) and medicine (10). However, there is very limited research and application in transportation planning. Algorithms that take the role of the modeler or assist his/her task of tuning hyperparameter and selecting appropriate models are important for transportation planning in at least three ways:

- To reduce subjective error in model estimation.
- To increase productivity of modelers by reducing amount of time spent on tuning and model selection. Modelers can focus their time on other tasks, such as model interpreta-
• To enable modelers with minimal technical knowledge to estimate models.

This study proposed an approach to reduce subjectivity during the tuning of hyperparameters required for the estimation of nonparametric models. That is, a flexible-statistical framework was proposed to estimate a model and associated hyperparameters. The framework was applied, as an example, to estimate flexible activity-duration models for out-of-home non-mandatory activity episodes for household individuals in the greater Los Angeles area (similar to Ferdous et al. (11)).

Activity duration modeling has received a considerable amount of research attention (3, 12–19). This is because activity-based modeling (ABM) places emphasis on understanding participation characteristics of individuals including the duration spent performing an activity (1). The amount of time spent at activity locations contributes to dynamic changes in the built environment (18, 19). For instance, locations that experience longer durations of out-of-home non-mandatory activities, such as shopping and recreation, will attract leisure services (such as shopping malls, markets) thereby affecting travel behavior. Several operational frameworks of ABM, implemented in many metropolitan areas in the US, rely on estimates of durations for out-of-home work activities for use in generating and scheduling out-of-home non-mandatory activities (14). Some studies have investigated the factors that influence durations of activity episodes. For instance, Habibi (16) studied factors affecting out-of-home physical activities by children in the city of Dhaka, Bangladesh. Car ownership, tenure type, activity location, and parent-escort arrangements were found to be the most notable predictors of the duration of physical activity. Insights from this study provided important policy implications. The reader is referred to Pinjari et al. (14) for a more detailed exposition of other benefits of activity duration modeling.

The rest of this paper is organized as follows. Section 3, Related Work, reviews literature in the transportation domain that have proposed some comparative methods for hyperparameter tuning and model selection. Section 4, Methodology, presents the details of the proposed methodology including: a methodological context that highlights the current emerging methods for automated hyperparameter tuning and model selection; and methodology formulation which develops two important components of the proposed framework. Section 5, Experiments, presents the experiments performed including a description of the data used, the sample formation process, and empirical analysis for the five different activity episodes. Section 6, Discussion of Results, presents a general discussion of the results as it relates to activity duration modeling and implications to travel planning. Finally, a conclusion that summarizes the important findings from this study are presented in Section 7, Conclusions and Future Work, including potential directions for future research.

RELATED WORK
The current state of the art in transportation has proposed several methods for the estimation of statistical models while reducing the need for inputs from the modeler. For instance, the selection of variables to be considered has received some attention. Recently, Habibi (20) used a non-random hold-out validation framework to select variables to be considered for predicting car type demand. Another study by Ahmed et al. (21) used different search techniques including stepwise, forward, and backward search to identify significant variables associated with crash occurrence. All three search techniques identified the same variables to be significant in their multivariate analysis. Both proposed frameworks only addressed variable selection but did not consider other key aspects such as model selection, parameter estimation, and hyperparameter tuning.

Other studies have focused on search techniques to obtain the best model specification.
For instance, Washington and Wolf (22) compared Ordinary Least Squares (OLS) regression and Hierarchical tree-based regression (HTBR) methods for estimating trip generation models. They demonstrated that HTBR methods automatically search for statistical relationships between response and predictor variables. Kim et al. (23) compared several techniques for calibrating traffic microsimulation flow models including: manual search, gradient-based, simplex-based and artificial intelligence techniques. Their proposed approach required pre-specifying a statistically based objective function for use in the calibration process. Athanasopoulos et al. (24) implemented three time series algorithms, including Forecast Pro, ARIMA and exponential smoothing based algorithms, for forecasting tourism demand. Their methods followed a step-wise algorithm, proposed by Hyndman and Khandakar (25), to traverse the space of models efficiently to obtain a model with the lowest Akaike Information Criterion (AIC) value. Tulic et al. (26) developed a procedure for predicting route travel times and mean speeds in urban networks. Their approach considered features of input data particularly measurement errors, by including a model for heteroscedasticity.

Advanced methods for parameter-tuning have also been proposed. For instance, Abraham and Hunt (27) used a heuristic search method, Levenberg-Marquardt, to tune parameters to improve the goodness of fit of their nested logit model for travel destination and mode choice. This heuristic search method was a much faster approach than an initial manual process. Hollander (28) proposed a multi-objective approach to predict travel times and modes. The approach involved combining several objective functions, rather than using one as is the case of maximum likelihood estimation. The full set of mode choices made by travelers was estimated. Then, the objective functions were used to test the extent to which the estimated choices met travelers’ needs. This was performed repeatedly with different possible sets of parameters. The downhill simplex method was used to search for the combination of parameters that best estimated travel times and modes. The main drawback of this approach is that it is computationally expensive in terms of the number of interactions required to reach convergence, which is not always guaranteed.

To the best of the authors’ knowledge, comprehensive estimation including model selection, hyperparameter tuning and parameter estimation using current emerging techniques (described in the third Section, Methodology) have not been explored for developing TDMs.

**METHODOLOGY**

**Methodology context**

A systematic approach to handle the iterative process of hyperparameter tuning, model selection, and parameter estimation would be to treat the process as an optimization problem (29) in order to search for the best fit between the model and observed data (29). Generally, such optimization can be categorized into two groups, including methods that:

- Search in a finite model space of known standard models, such as logit, probit, and mixed logit; and
- Use a nonparametric approach to search in an infinite model space for the model and parameters that best explain the observed data.

The latter category, to a great extent, lets the data dominate the process of model discovery. This category often results in the discovery of new model specifications not included in the existing literature (30). The methodology proposed in this study is based on the latter category of methods.

Optimization methods that search in an infinite model space seek to find a model that is fundamentally superior to others in terms of predictive performance (31). This includes estimating the corresponding parameters to produce models that best describe the observed data and can
These drawbacks can be addressed by using random search (RS) (33). Instead of searching the entire grid of models and hyperparameters, random search takes random walks over this large grid. This method, which has been shown to be superior to grid search (30), is considered the de facto method for tuning hyperparameters because of its simplicity and efficiency. However, in both grid and random search, explorations are done randomly and blindly. That is, any set of hyperparameters and models are explored independently of the previous one. This results in loss of valuable information that could help the next iteration.

This drawback can be addressed by using sequential model-based methods, such as Sequential Model-Based Global Optimization (SMBO) (5, 30, 34–36). SMBO methods, such as Bayesian optimization (BO) (30), use a surrogate function to approximate the unknown function of the ‘true’ model. Given that this surrogate function is simple to evaluate, it can be optimized to select points that yield the best approximation of the true function while leveraging information gathered from previous iterations. BO has demonstrated success in optimizing millions of hyperparameters in neural networks (9). The proposed approach in this study leverages Bayesian optimization to formulate a framework that simultaneously seeks adequate hyperparameters and selects a flexible activity-duration model.

**Methodology Formulation**

Generally, the proposed methodology consists of two main stages:

- **Stage 1**: A surrogate function for the model that relates activity duration and associated factors is proposed.
- **Stage 2**: The surrogate is optimized iteratively using a few data points until it approaches the ‘true’ model of activity duration.

To formulate the methodology, consider for each individual traveler, \( i \), that \( y_i \) represents the log of activity duration and \( x_i \) is a vector of potential explanatory variables such as trip distance, income levels, and household characteristics. Regression modeling can involve estimating a latent function \( f(x_i) \), which maps a vector of input data, \( x_i \), to output data \( y_i \) for \( i = 1, 2, \ldots, N \), where \( N \) is the total number of travelers. Each of the input vectors \( x_i \) is of dimension \( D \), and \( X \) is a \( N \times D \) matrix with rows \( x_i \). The observations are assumed to satisfy:

\[
y_i = f(x_i) + \varepsilon, \quad \text{where } \varepsilon \sim \mathcal{N}(0, \sigma^2_{\varepsilon})
\]

The noise term, \( \varepsilon \), is assumed to be normally distributed with a zero mean and variance, \( \sigma^2_{\varepsilon} \). Latent function \( f \) represents a hidden underlying travel behavior that produced the observed activity-duration data. The objective of the optimization framework is to find the set of arguments and values, \( x^* \), that maximizes (or minimizes) a continuous function \( f(x) \) over a compact space \( x \).

\[
x^* = \arg\max_{x \in X} f(x)
\]
whether \( f(x) \) is linear, quadratic, sinusoidal or polynomial to some order. Hence, the first stage of
proposed methodology involves using a surrogate function to approximate \( f(x) \).

A probabilistic function sampled from a Gaussian process (GP) prior \( (2, 37, 38) \) was chosen
as the surrogate for the functional relationship between activity duration and associated factors.
The reason for this choice is because a GP prior provides a probability distribution over an infinite
space of possible functions \( (37) \). This surrogate function is fully parameterized by a mean vector
and covariance matrix (kernel), denoted as:

\[
f \sim GP(m, K_{N,N}),
\]

where:

\( f \) = a surrogate function that relates activity duration and associated factors,
\( m \) = a mean vector (mean function) that relates activity duration and associated factors,
\( K_{N,N} \) = a covariance matrix (kernel function) that captures correlation between regressors

Bayes theorem then was used to estimate the posterior distribution over the unknown func-
tion evaluations \( f \) at all data points \( x_i \), as illustrated by:

\[
p(f | y, X) = \frac{p(y | f, X) p(f)}{p(y | X)},
\]

\[
= \frac{p(y | f, X) N(f | m, K_{N,N})}{p(y | X)},
\]

where:

\( p(f | y, X) \) = the posterior distribution of functions that best explain log activity duration,
given the covariates
\( p(y | f, X) \) = the likelihood of log of activity duration, given the functions and covariates
\( p(f) \) = the prior over all possible functions of log of activity duration
\( p(y | X) \) = the data (constant)

This posterior is a Gaussian process composed of a distribution of possible functions that best
explain activity duration patterns. Given that the data are fixed, Eq. (4) was re-formulated as the
unnormalized posterior distribution

\[
p(f | y, X) \propto p(y | f, X) N(f | m, K_{N,N}).
\]

The kernel function, \( K_{N,N} \), was used to encode prior information about the smoothness of
the function that relates activity duration and associated attributes. In this study, it was assumed
that "travelers with similar personal and household characteristics spend more or less similar
time during out-of-home non-mandatory activity episodes". For instance, travelers with similar
characteristics, such as income levels and job types, were assumed to have more similar activity
durations than individuals with different characteristics. This smoothness assumption was encoded
using the Squared Exponential-Automatic Relevance Determination (SE-ARD) kernel \( (39) \). Each
covariance element in this kernel, \( k(x_i, x_j) \), was evaluated as:

\[
k_{se-ard}(x_i, x'_j) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{d=1}^{D} \frac{(x_{i,d} - x_{j,d})^2}{l_d^2}\right),
\]
where $\sigma_f^2$ is the signal or function variance to capture the variance in the estimated functions; and $l_d$ is the characteristic length scale in dimension $d$, to capture the relevance of each covariate to activity duration. The likelihood function of activity duration was expressed as:

$$p(y \mid f, X) = N(f, \sigma^2).$$  

(7)

The hyperparameters in this surrogate model include $\theta = \{\sigma_f^2, l_d, \sigma_\varepsilon\}$. Given the GP prior in Equation (3) and the data likelihood in Equation (7), the posterior of the hyperparameters and parameters of the surrogate were estimated using Bayesian inference as follows:

$$p(f \mid y, X) \propto p(y \mid f, X)N(f \mid m, K_{N,N}).$$  

(8)

The second stage involves optimizing stage 1 to run efficiently. Given that evaluations of the surrogate model $f(x)$ may be expensive due to its complexity and/or use of large dataset, an approach is proposed to select a subset of the large dataset to produce a model of comparable performance. The approach selects a few data points at a time to estimate the surrogate function iteratively. That is, as new data points, $D_{t:T} = \{x_{t:T}, y_{t:T}\}$, are accumulated from each iteration $t$, where $T$ is the total number of iterations, the GP prior is combined with the data likelihood to update the posterior distribution of the surrogate, $p(f \mid y, X)$, using Equation (8). The posterior in iteration $t$ becomes the prior for the next iteration, $t+1$. To select this subset of data points, a secondary utility function, called an acquisition function, is pre-specified. The purpose of this acquisition function, $a(.)$, is to select the next data point to use for re-estimating the surrogate. The acquisition function was designed to make effective and efficient choices of optimal data points by simultaneously considering the following two requirements (30):

- The acquisition function should exploit data points that are close to previous optimal iterations because they may potentially offer improvement.,
- The acquisition function should explore data points where the uncertainty is highest because the optimal point may lie in those areas.

The acquisition function was designed to trade off these two competing requirements to determine points to be explored and exploited with preference given to those points that produce the best improvement of the surrogate function over previous iterations. The acquisition function used in this study is the expected improvement (E.I) (40), expressed as follows: Consider that $f(x_t)$ is the current best (extremal) point evaluation of the surrogate model. That is:

$$x_t = \arg \max_{x \in \mathcal{X}} f(x)$$  

(9)

and $f(x_{t+1})$ is the evaluation of the surrogate function at the next iteration. The magnitude of improvement, $I(x_{t+1})$, over the current best point is given by:

$$I(x_{t+1}) = \max \{0, f(x_{t+1}) - f(x_t)\}$$  

(10)

where $I(x_{t+1})$ is positive when $f(x_{t+1})$ is higher than the $f(x_t)$ and zero otherwise. The likelihood of improvement, $I$, on a normal posterior distribution specified by $\mu(x_{t+1})$, $\sigma(x_{t+1})$ can be computed from the normal density function given by:

$$\frac{1}{\sigma(x_{t+1})} \sqrt{2\pi} \exp \left( -\frac{(\mu(x_{t+1}) - f(x_t) - I)^2}{2\sigma^2(x_{t+1})} \right)$$  

(11)
The expected improvement is the integral over the likelihood of improvement, given by:

\[ E(I) = \int_{I=0}^{I=\infty} \frac{1}{\sigma(x_{t+1})\sqrt{2\pi}} \exp \left( -\frac{(\mu(x_{t+1}) - f(x_t) - I)^2}{2\sigma^2(x_{t+1})} \right) dI \] (12)

\[ E(I) = \sigma(x_{t+1}) \left[ \frac{\mu(x_{t+1}) - f(x_t)}{\sigma(x_{t+1})} \Phi\left( -\frac{\mu(x_{t+1}) - f(x_t)}{\sigma(x_{t+1})} \right) + \phi\left( -\frac{\mu(x_{t+1}) - f(x_t)}{\sigma(x_{t+1})} \right) \right] \] (13)

where \( \Phi(.) \) and \( \phi(.) \) are the probability and cumulative distribution functions of the standard normal distribution, respectively. The analytic solution to \( E(I) \) in Equation (13) is given by:

\[ EI(x_{t+1}) = \begin{cases} \frac{(\mu(x_{t+1}) - f(x_t) - \xi)\Phi(Z) + \sigma(x_{t+1})\phi(Z)}{\sigma(x_{t+1})} & \text{if } \sigma(x_{t+1}) > 0 \\ 0 & \text{if } \sigma(x_{t+1}) = 0 \end{cases} \] (14)

where \( \xi \geq 0 \) is a parameter that controls the trade-off between global search and local optimization (i.e., exploration/exploitation) suggested by Lizotte (41), and

\[ Z = \begin{cases} \frac{\mu(x_{t+1}) - f(x_t) - \xi}{\sigma(x_{t+1})} & \text{if } \sigma(x_{t+1}) > 0 \\ 0 & \text{if } \sigma(x_{t+1}) = 0 \end{cases} \] (15)

For each iteration, E.I is maximized to select a new additional data point, \( x_{t+1} \). This data point is then augmented with previous optimal data points and collectively used to update the surrogate. After updating the surrogate function, the cycle is repeated until convergence or a constraint on the total number of iterations, \( T \), is met. Each new evaluation decreases the distance between the true global maximum and the expected maximum given the model.

The two stages described above (i.e., the surrogate from a Gaussian process prior and the Expected Improvement acquisition function) were combined to produce the proposed method for hyperparameter tuning and selection of an activity-duration model. The proposed framework is summarized in the following algorithm:

- **Step 1. Initialization:**
  - Step 1.1. Select two random data points of activity duration and corresponding attributes \((x_1, y_1), (x_2, y_2)\) from the dataset and define their likelihood, \( p(y \mid f, X) \).
  - Step 1.2. Define a Gaussian process prior, \( p(f \mid X) \), that shows that individuals with similar attributes have similar activity durations.
  - Step 1.3. Combine the prior with the likelihood, using Bayes theorem, to produce the Gaussian process posterior distribution, \( p(f \mid y, X) \), i.e., the surrogate of the activity-duration model.

- **Step 2. Selection of an optimal data point:**
  - Step 2.1. Use the surrogate of the activity-duration model to construct an acquisition function \( EI(x) \).
  - Step 2.2. Maximize \( EI(x) \) to find an optimal data point \( x_{t+1} \) that produces the highest expected improvement of the activity-duration model over the previous data point.

- **Step 3. Evaluation of objective function at selected point:**
Step 3.1. Use the optimal data point $x_t+1$ to re-evaluate the surrogate of the activity-duration model to produce a new Gaussian process posterior.

Step 3.2. Query the new GP posterior to obtain the corresponding value of activity duration $y_{t+1}$.

Step 3.3. Augment the new selected data point to the previous points to obtain updated data set of activity durations and corresponding attributes $(x_1,y_1), (x_2,y_2), (x_{t+1},y_{t+1})$.

- Step 4. Updating probabilistic model of surrogate:
  - Step 4.1. Use the updated data $(x_1,y_1), (x_2,y_2), (x_{t+1},y_{t+1})$ to update the surrogate of the activity-duration model.
  - Step 4.2. Identify extremal value of the activity-duration model, $f_{\text{max}}$.

- Step 5. Stopping criteria:
  - If the prediction error of the surrogate of the activity-duration model has met the desirable error, then the algorithm is terminated. Or if the algorithm has reached the maximum number of possible iterations (due to time or compute constraints), then the algorithm is terminated. Otherwise go back to step 2. That is, if $\text{error} \leq \text{error}_{\text{desirable}}$ or $T = T_{\text{desirable}}$, then stop the optimization algorithm. Otherwise, go to step 2.

Figure 1 illustrates the proposed procedure for performing Bayesian optimization using an integrated acquisition function and a probabilistic model of the objective function.

EXPERIMENTS

Raw data and sample formation

The data used for empirical analysis in this study was drawn from the 2010-2012 California household travel survey (CHTS) collected by the California Department of Transportation (Caltrans), which is available online at the Transportation Secure Data Center (42). This survey consists of socioeconomic and travel behavior information of 42,431 California households.

To obtain the sample for the current analysis, the CHTS data was processed as follows. First, data from the Los Angeles area were filtered. Second, only activities performed on a weekday that was not a holiday were selected, because the focus of the current paper is to study individuals’ activity participation patterns on a typical weekday. Third, only five out-of-home non-mandatory activity types were considered including (similar to Ferdous et al. (11)):

- Personal/family care (including personal care, caring for children in the household, pick-up/drop-off of children/adults, and caring for extended family members),
- Shopping (grocery shopping, purchasing gas/food, and banking, window shopping, cloth, and electronic shopping),
- Meals,
- Physically active recreation (sports, exercise, walking, bicycling, recreational and volunteer activities.), and
- Physically inactive recreation (social, relaxing, movies, and attending religious/cultural/sports events).

Fourth, data regarding travelers’ socioeconomic statuses, trip purposes, household characteristics, and personal characteristics were appended to their activity types. Table 1 provides descriptive statistics of some of the categorical variables. Fifth, the choices for commute mode that were considered include auto driver, auto passenger, non-motorized, bus, and train. Travelers with missing...
FIGURE 1: Proposed algorithm to tune hyperparameters and generate an activity-duration model
data were removed from the sample, resulting in a total sample size of 23,906 observations.

**Table 1**: Summary statistics of some variables in the sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Frequency</th>
<th>Percentage(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>12353</td>
<td>51.7</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>11553</td>
<td>48.3</td>
</tr>
<tr>
<td>Income</td>
<td>Low</td>
<td>9591</td>
<td>40.1</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>9301</td>
<td>38.9</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>5014</td>
<td>21.0</td>
</tr>
<tr>
<td>Education</td>
<td>College</td>
<td>14569</td>
<td>60.9</td>
</tr>
<tr>
<td></td>
<td>Some college</td>
<td>5633</td>
<td>23.6</td>
</tr>
<tr>
<td></td>
<td>No college</td>
<td>3704</td>
<td>15.5</td>
</tr>
<tr>
<td>Mode</td>
<td>Auto-driver</td>
<td>20372</td>
<td>82.7</td>
</tr>
<tr>
<td></td>
<td>Non-motorized</td>
<td>1294</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>Auto-passenger</td>
<td>1054</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>Bus</td>
<td>873</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>Train</td>
<td>313</td>
<td>1.3</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>16627</td>
<td>69.6</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>2393</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>2202</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>Indian</td>
<td>1,382</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>1302</td>
<td>5.4</td>
</tr>
<tr>
<td>Occupation</td>
<td>Management</td>
<td>11670</td>
<td>48.8</td>
</tr>
<tr>
<td></td>
<td>Clerical</td>
<td>5607</td>
<td>23.5</td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>2607</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>2449</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>1573</td>
<td>6.6</td>
</tr>
<tr>
<td>Residence type</td>
<td>House</td>
<td>18920</td>
<td>79.1</td>
</tr>
<tr>
<td></td>
<td>Apartment</td>
<td>4986</td>
<td>20.9</td>
</tr>
<tr>
<td>Home ownership</td>
<td>Own</td>
<td>17755</td>
<td>74.3</td>
</tr>
<tr>
<td></td>
<td>Rent</td>
<td>6151</td>
<td>25.7</td>
</tr>
</tbody>
</table>

Table 2 presents the proportion of each activity type in the final sample. Personal activities were the most frequently undertaken activity type with 51.4% proportion followed by meals (20.2%), shopping (15.6%), physically inactive recreation (7.9%) and physically active recreation (4.8%). Figure 2 presents the density distribution of each activity type in the sample. Generally, most of the duration density of all activity types was concentrated within the 300-minute mark indicating that most out-of-home non-mandatory activities last at-most 300 mins (5 hours). However, personal activities and meals had two distinct peaks in their duration density. For instance, the data suggests that there were two groups of dining behavior; a larger and a small group that
TABLE 2: Descriptive statistics of durations (in mins) by activity type

<table>
<thead>
<tr>
<th>Activity type</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Frequency</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal activities</td>
<td>309.5</td>
<td>269.0</td>
<td>1.0</td>
<td>1,370</td>
<td>12,296</td>
<td>51.4</td>
</tr>
<tr>
<td>Meals</td>
<td>342.1</td>
<td>262.1</td>
<td>1.0</td>
<td>1,249</td>
<td>4,837</td>
<td>20.2</td>
</tr>
<tr>
<td>Shopping</td>
<td>35.2</td>
<td>44.3</td>
<td>1.0</td>
<td>635</td>
<td>3,735</td>
<td>15.6</td>
</tr>
<tr>
<td>Physically inactive recreation</td>
<td>157.3</td>
<td>187.1</td>
<td>1.0</td>
<td>1,069</td>
<td>1,896</td>
<td>7.9</td>
</tr>
<tr>
<td>Physically active recreation</td>
<td>186.4</td>
<td>165.9</td>
<td>1.0</td>
<td>1,254</td>
<td>1,142</td>
<td>4.8</td>
</tr>
</tbody>
</table>

FIGURE 2: Distribution of activity duration by activity type

spent on average 60 and 550 mins (9 hours), respectively. It is typical for dining to last on average one hour. The second-smaller group may constitute long planned dining events such as parties or other types of celebrations. Given that the duration density for shopping is more compact and narrow (i.e. has a small variance), it suggests that shopping is more time-consistent than all others. Almost all shoppers took on average the same amount of time of 35 mins as confirmed in Table 2. These insights on activity duration could be crucial in planning the built environment which in turn influences travel behavior. For instance, parking duration for shopping centers could be optimized to 30 mins per shopper.

Empirical analysis

To tune carefully the GP model parameters using random search, a rigorous and thorough procedure, adopted from Kuhn and Johnson (43), was followed. This procedure, which is considered state of the art in machine learning, was customized in this study as shown in Algorithm 1. For each activity type, the following procedure was followed. A set of hyperparameter values to be evaluated was generated using a random search algorithm (33). The set of hyperparameters to be tuned in the fitted GP model included one signal variance $\sigma_f$ parameter and 33 length-scale parameters $l_d$ for each dimension $d$ of the regressors. One hundred unique combinations of the 34 hyperparameters were generated using random search, resulting in 3,400 tunable hyperparameters.
Algorithm 1 Procedure to carefully tune hyperparameters of a GP model using random search

1: procedure
2: Generate a set of hyperparameters to evaluate, \((\sigma_f, l_d)\), using random search
3: Specify a re-sampling procedure, such as 10-fold cross validation repeated 5 times
4: for each set of \((\sigma_f, l_d)\) do
5: for each of the 5 repetitions of the 10-folds of the data partition do
6: Hold-out 1 random fold of the data
7: Fit the GP model on the remainder of the partitions
8: Use fitted GP model to predict the hold-out samples
9: end for Calculate the RMSE and R-Squared values across hold-out partitions
10: end for Determine the optimal parameter set, \(\{\sigma_{best_f}, l_{best_d}\}\).
11: Fit the final model to all the training data using the optimal parameter set
12: end procedure

The data was partitioned into two with 70% and 30% for training and testing, respectively. The
type of resampling procedure was then specified to involve 10-fold cross validation (CV), each
repeated 5 times (i.e. five separate 10-fold CVs). This is known as repeated k-fold cross valida-
tion. This resulted in 50 training instances of the GP model. For each training instance, 9-folds
of data partition was used for training and 1-fold used as hold-out partition for testing the fitted
GP model. Root mean square error (RMSE) values were estimated for each testing instance as
metrics of model performance. The hyperparameter set which produced the lowest RMSE value
was chosen as the best signal variance, \(\{\sigma_{best_f}, l_{best_d}\}_{RS}\). Given that each of the re-sampled data sets
is independent of the others, the training of the models was performed in parallel so as increase
computational efficiency. This study utilized a linux server with 80 cores and 256GB of RAM
to spread the computation across these cores. Table 3 presents only values for \(\sigma_{best_f}\) by activity
type. Values of 0.012, 0.013, 0.021, 0.015 and 0.023 were estimated for personal activities, meals,
shopping, inactive and active recreation, respectively. Finally, \(\{\sigma_{best_f}, l_{best_d}\}_{RS}\) was used to fit the
GP model to all the training data (70% portion). The final estimated model was tested on the test
dataset (30% portion) to determine the RMSE, MAPE and R-squared values, which are presented
in Table 3.

The proposed BO procedure was then applied to the same data to search for the best 34
hyperparameters \(\{\sigma_{best_f}, l_{best_d}\}_{BO}\). The following two settings were chosen to initialize the BO run
(as previously described in the proposed algorithm in Figure 1). First, two random data points
of activity duration with their corresponding attributes were chosen. Second, the total number
of iterations, \(T_{desirable}\), was chosen to be 100. For each iteration, the corresponding \(\{\sigma_f, l_d\}\) and
RMSE were evaluated.

The hyperparameter set that produced the lowest RMSE was chosen as the optimal/best
set, \(\{\sigma_{best_f}, l_{best_d}\}_{BO}\). Table 3 only shows the \(\sigma_{best_f}\). Values of 0.004, 0.006, 0.013, 0.004 and 0.024
were evaluated for personal activities, meals, shopping, active recreation and inactive recreation,
respectively. These values from BO are considerably different from the random search procedure.
It would have been difficult for an expert to suggest such values given their small scale and that
there is an infinite number of possibilities in the range \(\{0, 0.9\}\). Finally, \(\{\sigma_{best_f}, l_{best_d}\}_{BO}\) was used
to fit the GP model to all the training data. The final estimated model was tested on test data to
evaluate the RMSE, MAPE and R-squared values, which are presented in Table 3.
TABLE 3: A comparison of duration prediction accuracies between random search and Bayesian optimization by activity type.

<table>
<thead>
<tr>
<th>Activity type</th>
<th>Procedure</th>
<th>Best parameter $\sigma_f^{best}$</th>
<th>RMSE (mins)</th>
<th>MAPE (%)</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal activities</td>
<td>Random</td>
<td>0.012</td>
<td>14.26</td>
<td>44.13</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>BO</td>
<td>0.004</td>
<td>4.72</td>
<td>24.39</td>
<td>0.57</td>
</tr>
<tr>
<td>Meals</td>
<td>Random</td>
<td>0.013</td>
<td>12.24</td>
<td>25.33</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>BO</td>
<td>0.006</td>
<td>2.18</td>
<td>13.85</td>
<td>0.59</td>
</tr>
<tr>
<td>Shopping</td>
<td>Random</td>
<td>0.021</td>
<td>12.16</td>
<td>31.45</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>BO</td>
<td>0.013</td>
<td>2.69</td>
<td>12.35</td>
<td>0.49</td>
</tr>
<tr>
<td>Active recreation</td>
<td>Random</td>
<td>0.015</td>
<td>13.11</td>
<td>23.62</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>BO</td>
<td>0.004</td>
<td>2.82</td>
<td>23.31</td>
<td>0.46</td>
</tr>
<tr>
<td>Inactive recreation</td>
<td>Random</td>
<td>0.023</td>
<td>14.24</td>
<td>24.68</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>BO</td>
<td>0.024</td>
<td>3.99</td>
<td>16.99</td>
<td>0.47</td>
</tr>
</tbody>
</table>

The performance metrics for the proposed BO framework were found to be superior to the RS procedure. The RMSE and MAPE from BO were smaller than those from RS across all activity types. For instance, an error of 24.39% was found from BO for personal activities compared to an error of 44.13% from the RS procedure. The R-squared values from BO were higher than those from RS across all activity types. For instance, an R-squared value of 0.57 was found from BO for personal activities compared to an R-squared value of 0.30 from the RS procedure.

Insights related to transportation planning obtained from the final estimated model for personal activities only are discussed in Section 6, Discussion of Results.

9 DISCUSSION OF RESULTS

Another set of empirical results included the characteristic length-scale parameters, $l_d$, for each of the attributes in the final estimated models. For illustration and discussion, only length-scale parameters for personal activities are provided in Table 4. This non-negative parameter is a measure of relevance of an attribute (which could be interpreted as ‘significance’ in frequentist terminology) to predicting activity duration. The larger the $l_d$, the smaller the covariance of the corresponding attribute. Consequently samples from a distribution with a small covariance are more or less similar. This indicates that such attributes do not vary with activity duration. This study retained attributes with $l_d \leq 5$ as those that were relevant to predicting activity durations. In addition, Table 4 includes statistical properties of their corresponding length scale parameters such as a potential scale reduction statistic (Rhat) (44) to show whether estimation reached global convergence. The Rhat value for all parameters were less than 1.05 indicating that all the MCMC chains in the model converged to the global equilibrium (44).

Given that the length scale parameter is a non-negative value, the sign does not have any meaning. Plots of the marginal posterior mean and covariance distributions of each individual attribute in the estimated GP models were generated to investigate the relationship between the attributes and activity duration. For illustration, Figure 3 presents plots for a few selected variables.
TABLE 4: Estimates for the length scale parameters for the attributes

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Conf.low</th>
<th>Conf.high</th>
<th>Rhat</th>
<th>Neff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto driver</td>
<td>0.98</td>
<td>1.01</td>
<td>0.12</td>
<td>2.08</td>
<td>1.00</td>
<td>375</td>
</tr>
<tr>
<td>Auto passenger</td>
<td>0.97</td>
<td>0.90</td>
<td>0.19</td>
<td>2.05</td>
<td>1.00</td>
<td>400</td>
</tr>
<tr>
<td>Non-motorized</td>
<td>1.21</td>
<td>1.09</td>
<td>0.16</td>
<td>2.38</td>
<td>1.01</td>
<td>368</td>
</tr>
<tr>
<td>Trip distance (miles)</td>
<td>0.88</td>
<td>1.00</td>
<td>0.03</td>
<td>2.14</td>
<td>1.01</td>
<td>223</td>
</tr>
<tr>
<td>Vehicle count</td>
<td>1.22</td>
<td>1.04</td>
<td>0.18</td>
<td>2.52</td>
<td>0.99</td>
<td>400</td>
</tr>
<tr>
<td>Operational vehicle count</td>
<td>1.21</td>
<td>1.11</td>
<td>0.20</td>
<td>2.35</td>
<td>0.99</td>
<td>400</td>
</tr>
<tr>
<td>Low income (&lt; $34,999)</td>
<td>0.99</td>
<td>0.91</td>
<td>0.10</td>
<td>2.11</td>
<td>0.99</td>
<td>400</td>
</tr>
<tr>
<td>Medium income ($35,000 - $74,999)</td>
<td>1.22</td>
<td>1.11</td>
<td>0.27</td>
<td>2.67</td>
<td>1.00</td>
<td>400</td>
</tr>
<tr>
<td>Home ownership: Rent</td>
<td>0.86</td>
<td>0.79</td>
<td>0.11</td>
<td>1.78</td>
<td>1.01</td>
<td>400</td>
</tr>
<tr>
<td>Trip count</td>
<td>1.06</td>
<td>1.15</td>
<td>0.14</td>
<td>2.43</td>
<td>1.01</td>
<td>329</td>
</tr>
<tr>
<td>Person trips</td>
<td>1.13</td>
<td>0.70</td>
<td>0.35</td>
<td>2.03</td>
<td>1.01</td>
<td>400</td>
</tr>
<tr>
<td>Activity count</td>
<td>0.67</td>
<td>0.65</td>
<td>0.11</td>
<td>1.26</td>
<td>1.01</td>
<td>400</td>
</tr>
<tr>
<td>No college</td>
<td>1.03</td>
<td>0.93</td>
<td>0.17</td>
<td>2.16</td>
<td>1.02</td>
<td>231</td>
</tr>
<tr>
<td>Some college</td>
<td>1.07</td>
<td>1.02</td>
<td>0.18</td>
<td>2.28</td>
<td>1.01</td>
<td>324</td>
</tr>
<tr>
<td>Black</td>
<td>0.96</td>
<td>0.92</td>
<td>0.16</td>
<td>1.97</td>
<td>1.00</td>
<td>400</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.23</td>
<td>1.47</td>
<td>0.16</td>
<td>2.64</td>
<td>1.00</td>
<td>188</td>
</tr>
<tr>
<td>Indian</td>
<td>0.99</td>
<td>0.91</td>
<td>0.19</td>
<td>2.14</td>
<td>1.00</td>
<td>400</td>
</tr>
<tr>
<td>White</td>
<td>1.02</td>
<td>0.90</td>
<td>0.19</td>
<td>2.13</td>
<td>1.01</td>
<td>400</td>
</tr>
<tr>
<td>Management occupation</td>
<td>0.95</td>
<td>0.82</td>
<td>0.16</td>
<td>2.15</td>
<td>1.00</td>
<td>324</td>
</tr>
<tr>
<td>Manufacturing occupation</td>
<td>1.03</td>
<td>0.97</td>
<td>0.09</td>
<td>2.32</td>
<td>1.00</td>
<td>400</td>
</tr>
<tr>
<td>Social occupation</td>
<td>1.06</td>
<td>1.04</td>
<td>0.13</td>
<td>2.11</td>
<td>1.00</td>
<td>310</td>
</tr>
<tr>
<td>House residence type</td>
<td>1.08</td>
<td>1.17</td>
<td>0.13</td>
<td>2.24</td>
<td>1.00</td>
<td>400</td>
</tr>
</tbody>
</table>

1 The red and blue lines are the posterior mean and covariance functions, respectively. They constitute functions that best explain the observed data. The black dots are samples of test data plotted over the prediction functions. Given that additivity made the offset of each function arbitrary, the critical properties indicated from the graphs were the pattern of variation as well as the relative magnitude of variation across the different covariates.

2 For instance, the marginal posterior distribution illustrated in Figure 3a suggests that individuals who are auto drivers spend longer durations at personal activities than individuals who take the bus (bus was the base category). Figure 3b suggests that individuals who are auto passengers spend shorter durations at personal activities than individuals who take the bus. Considering household- and personal-level factors, the findings revealed that individuals who live in house residences spent longer durations at personal activities than those who lived in apartments, the base category (Figure 3c). Individual who work in the manufacturing occupation spend longer durations at personal activities than those in the clerical business (Figure 3d). Regarding racial groups, the findings in Figure 3e suggest that Black individuals spend shorter durations at personal activities than Asians, the base category. Figure 3f suggests that individuals with some college spend shorter durations at personal activities than those without college education.
Other factors, not shown Figure in 3, that were found to positively affect personal activity duration include vehicle count, operational vehicle count, and medium income. Factors that negatively affected personal activity duration include trip distance, activity counts, trip counts and
person trips. As the distance to the activity location increases, the duration spent there decreases. The number of trips and activity episodes undertaken decrease the amount of time spent at each individual activity location.

CONCLUSIONS AND FUTURE WORK

This paper proposed a Bayesian optimization framework to simultaneously tune hyperparameters and generate flexible duration models for five types of out-of-home non-mandatory activity episodes. The proposed framework addresses the current challenge in travel demand estimation whereby significant time and expertise are required to adjust hyperparameters. Algorithms that take the role of the modeler or assist his/her task of tuning hyperparameters are important because they reduce subjectivity in model estimation, increase productivity of modelers, and enable modelers with minimal technical background of statistics to select appropriate hyperparameters. Experiments demonstrated that results from this framework outperform expert-level performance and are obtained in a fraction of time compared to traditional approaches used in practice. Future work will involve increasing sample sizes and optimizing the algorithm to be efficient with larger datasets.
REFERENCES


