**INTRODUCTION**

**F**UEL consumption models for vehicles are of interest to manufacturers, regulators, and consumers. They are needed across all the phases of the vehicle life-cycle. In this paper, we focus on modeling average fuel consumption for heavy vehicles during the operation and maintenance phase. In general, techniques used to develop models for fuel consumption fall under three main categories:

Physics-based models, which are derived from an in-depth understanding of the physical system. These models describe the dynamics of the components of the vehicle at each time step using detailed mathematical equations.

Machine learning models, which are data-driven and represent an abstract mapping from an input space consisting of a selected set of predictors to an output space that represents the target output, in this case average fuel consumption

Statistical models, which are also data-driven and establish a mapping between the probability distribution of a selected set of predictors and the target outcome

Trade-offs among the above techniques are primarily with respect to cost and accuracy as per the requirements of the intended application.

In this paper, a model that can be easily developed for individual heavy vehicles in a large fleet is proposed. Relying on accurate models of all of the vehicles in a fleet, a fleet manager can optimize the route planning for all of the vehicles based on each unique vehicle predicted fuel consumption thereby ensuring the route assignments are aligned to minimize overall fleet fuel consumption. These types of fleets exist in various sectors including, road transportation of goods [7], public transportation [3], construction trucks [8] and refuse trucks [9]. For each fleet, the methodology must apply and adapt to many different vehicle technologies (including future ones) and configurations without detailed knowledge of the vehicles specific physical characteristics and measurements. These requirements make machine learning the technique of choice when taking into consideration the desired accuracy versus the cost of the development and adaptation of an individualized model for each vehicle in the fleet.

Several previous models for both instantaneous and average fuel consumption have been proposed. Physics-based models are best suited for predicting instantaneous fuel consumption [1], [2] because they can capture the dynamics of the behavior of the system at different time steps. Machine learning models are not able to predict instantaneous fuel consumption [3] with a high level of accuracy because of the difficulty associated with identifying patterns in instantaneous data. However, these models are able to identify and learn trends in average fuel consumption with an adequate level of accuracy.

Previously proposed machine learning models for average fuel consumption use a set of predictors that are collected over a time period to predict the corresponding fuel consumption in terms of either gallons per mile or liters per kilometer. While still focusing on average fuel consumption, our proposed approach differs from that used in previous models because the input space of the predictors is quantized with respect to a fixed distance as opposed to a fixed time period. In the proposed model, all the predictors are aggregated with respect to a fixed window that represents the distance traveled by the vehicle thereby providing a better mapping from the input space to the output space of the model. In contrast, previous machine learning models must not only learn the patterns in the input data but also perform a conversion from the time-based scale of the input domain to the distance-based scale of the output domain (i.e., average fuel consumption). Using the same scale for both the input and output spaces of the model offers several benefits:

Data is collected at a rate that is proportional to its impact on the outcome. When the input space is sampled with respect to time, the amount of data collected from a vehicle at a stop is the same as the amount of data collected when the vehicle is moving.

The predictors in the model are able to capture the impact of both the duty cycle and the environment on the average fuel consumption of the vehicle (e.g., the number of stops in an urban traffic over a given distance).

Data from raw sensors can be aggregated on-board into few predictors with lower storage and transmission bandwidth requirements. Given the increase in computational capabilities of new vehicles, data summarization is best performed on-board near the source of the data.

New technologies such as V2I and dynamic traffic management [10]–[12] can be leveraged for additional fuel efficiency optimization at the level of each specific vehicle, route and time of day.