Preprints, 10th IFAC International Symposium on Advances in Automotive Control Columbus, Ohio, USA, August 28-30, 2022

Route Generation Methodology for Energy Efficiency Evaluation of Connected and Automated Vehicles ⋆

Dennis Kibalama* Giorgio Rizzoni* Matteo Spano**

* *The Ohio State University, Columbus, OH 43212 USA (email: kibalama.3@osu.edu)*
** *Politecnico di Torino, Torino, 10129 Italy*

Abstract: Evaluation of the energy savings potential of Connected and Automated Vehicles (CAVs) technologies necessitates a representative baseline that accounts for the inherent variability due to route, terrain, traffic, traffic lights, etc., in real-world driving conditions. While considerable work has been done in the field of optimal energy management, eco-driving and eco-routing of CAVs, few contributions have addressed the creation of a representative baseline to realistically evaluate the energy savings potential of these technologies. This work proposes a route generation methodology based on leveraging a high-dimension driving dataset to construct diverse subset of synthetic driving trips and synthetic routes for large scale evaluation of energy consumption of CAVs. The generated synthetic routes can then be used to extract real-world routes from open-source mapping platforms, which have similar characteristics as the generated synthetic routes.

Keywords: Autonomous Vehicles, Simulation, General automobile/road-environment strategies.

1. INTRODUCTION

The current trends of vehicle electrification, ever increasing road user safety targets, advances in on-board computation hardware and algorithms, advances in vehicular connectivity and a myriad of other features have led to an increased proliferation of Connected and Automated Vehicle (CAV) technologies with a huge bias on electrified vehicles; with this trend expected to progressively increase as studied by Dingyi et al. (2018). Current state-of-the art methods in the evaluation of vehicle fuel economy rely on running standardized tests over regulatory drive cycles. However, to exhaustively evaluate the energy benefits of CAVs this cannot be performed over regulatory drive cycles since the vehicle velocity is an output of the CAV based on the driving scenario, route, traffic conditions. As such, establishing a representative baseline for evaluating the benefits of CAVs technologies is critical as has been studied in Rajakumar Deshpande et al. (2020) via Monte Carlo Simulation and then transitioned to representative track testing in Rajakumar Deshpande et al. (2021). However, these all assumed that the subset of routes over which these technologies were evaluated could represent real-world driving conditions.

Significant efforts have been dedicated towards generation of synthetic drive cycles for evaluating the energy consumption of vehicles beyond the standard regulatory drive cycles provided by US EPA (2020). Dembski et al. (2002) proposed a procedure to generate synthetic drive cycles manipulating the regulatory missions. In the work of Gong et al. (2010), the authors used a Markov Chain approach to create representative drive cycles for energy evaluation of Plug-in Hybrid Electric Vehicle (PHEV) fleets. Ericsson (2000) formulated a comparison of driving pattern metrics based on driver type, street environment factors and traffic condition factors to capture the variability inherent in urban driving patterns. This can be used to model realistic driving behavior in the development of traffic management systems but does not inherently model the variability that a CAV may encounter with access to route and traffic information. Ericsson (2001) studied 62 driving pattern parameters that were used to empirically quantify the effect of these driving pattern parameters on energy use and emissions. A subset of these parameters are chosen for the proposed work. However to the authors' knowledge, the procedures found in the literature do not accurately assess whether or not the generated drive missions would represent real-world driving conditions.

The aim of the present work is to provide a novel route generation process utilizing a driving data pool to produce representative missions for the evaluation of potential energy savings of CAVs. The data pool was collected during previous works of the authors and comprises both real-world and virtually simulated data. The vast set of route information is statistically analyzed to determine key metrics to later generate synthetic routes. Moreover, a constrained multi-objective eco-driving optimal control problem is formulated to quantify the energy consumption and compare it with the results obtained by virtually simulating a baseline driver. A large scale simulation is performed to evaluate a statistical representation of the potential energy savings over various route scenarios. Finally, the synthetic routes that match the dominant

⋆ This work is supported by the United States Department of Energy, Advanced Research Projects Agency – Energy (award number DE-AR0000794).

energy consumption values per route classification can be used in the route extraction involving real-world trips with similar features.

The work is divided as follows: firstly, the baseline simulator is introduced along with the main equations used to model it; then, the route generation methodology is rigorously described; finally, an example of the potential energy savings that could stem from CAVs is given along with conclusions and possible future works.

2. SIMULATOR - ENHANCED DRIVER MODEL

To model and analyze the energy consumption of the vehicle while accounting for variability due to driving styles and routes, an implementation of the Intelligent Driver Model (IDM) by Kesting et al. (2010) is adopted. To incorporate traffic and stop signs, an extension is made to formulate the Enhanced Driver Model (EDM) by Gupta et al. (2019), and is leveraged for driver behavior modeling in this work. The EDM simulates human driver behavior and predicts the longitudinal speed profile of the vehicle. This deterministic velocity predictor has been tuned with real-world data where the calibration parameter set is as defined in Equation 1, in which a_{max} is the maximum vehicle acceleration, b_{max} is the maximum vehicle deceleration, δ is the acceleration / deceleration exponent, θ_0 is the offset from the speed limit, and c_1 is correlated to braking distance.

$$\Psi_{driver} \triangleq \{a_{max}, b_{max}, \delta, c_1, \theta_0\} \tag{1}$$

The EDM operating modes which include Freeway Driving (FD), Car Following (CF) and approach to a stop location are defined by the following equations:

$$\dot{v}_e = \begin{cases} a_{max}\left(1-\left(\dfrac{v_e}{v_{lim}-\theta_0}\right)^{\delta}\right) & v_e < v_{lim}[FD] \\ -b_{max}\left(1-\left(\dfrac{v_{lim}-\theta_0}{v_e}\right)^{\delta}\right) & v_e \geq v_{lim}[FD] \\ a_{max}\left(1-\left(\dfrac{v_e}{v_l}\right)^{\delta}\right) & v_e < v_l[CF] \\ -b_{max}\left(1-\left(\dfrac{v_l}{v_e}\right)^{\delta}\right) & v_e \geq v_l[CF] \\ -\dfrac{1}{b_{max}}\left(\dfrac{v_e^2}{2s}\right)^2 & v_l = 0 \end{cases} \tag{2}$$

$$s = s_l - s_e - s_{safe} \tag{3}$$

$$s_{brake} = \left(1+\frac{c_1}{\delta}\right)\frac{v_e^2}{2b} \tag{4}$$

where $s_l, s_e, s_{safe}, s_{brake}$ denote lead vehicle position, ego-vehicle position, safe gap and safe braking distance respectively. v_e, v_l, v_{lim} represent the ego-vehicle velocity, lead vehicle velocity and speed limit respectively. Additionally, Rajakumar Deshpande et al. (2020) enhanced the EDM to include a Line of Sight (LoS) formulation to account for driver response when approaching a signalized intersection.

There are multitude of factors that contribute to the energy consumption of the vehicle, namely:

(1) Vehicle characteristics e.g., vehicle dimensions, payload, auxiliary loads, etc.
(2) Route characteristics, Ξ_{route} e.g., v_{lim}^{max}, v_{lim}^{min}, grade α, number of stop signs n_{ss}, number of traffic lights n_{tl}, location of stop signs s_{ss}, traffic light location s_{tl}, etc.
(3) Driving style, Ψ_{driver}
(4) Signal Phase and Timing (SPaT), $\Phi_{SPaT} = f(\Xi_{route})$: including the traffic light sequence for each light along the route. For this work, the traffic lights are assumed to run a fixed time interval traffic light sequence as opposed to an event-driven traffic light sequence.
(5) Traffic: which is embedded in v_l
(6) Other external conditions e.g. weather, road conditions, etc.

For this work, the variability in driving conditions is introduced by varying the route characteristics Ξ_{route}, driving style Ψ_{driver} and the SPaT Φ_{SPaT}. The EDM generates a vast driving dataset that is used in the route generation process outlined. A large scale simulation over the test matrix of routes, driving style and SPaT results in a dataset X with select features d based on statistical metrics defined in the next section and numerous samples n with:

$$X \in \mathbb{R}^{n\times d},\ 0 < d < \infty, n \gg d \tag{5}$$

3. ROUTE GENERATION METHODOLOGY

The initial route data used for generating the driving data pool has been sampled from around the *Columbus, OH* region with varying route characteristics. To aid in evaluating the energy and mobility benefits of predictive technologies that leverage look-ahead information (e.g. route data, traffic and traffic light information), large-scale Monte Carlo simulations were performed over a subset of these routes. In these simulations, multiple scenarios of the driver aggressiveness and SPaT were varied to determine the statistical variation of fuel consumption and travel time. Although the methodology was sufficient for showcasing the energy savings benefits for a mild HEV vehicle, the criteria for route selection may differ when considering vehicles with larger electrification, such as PHEVs and Battery Electric Vehicles (BEVs). In these cases, further insights into the methodology are required when generating feasible test scenarios mainly due to the larger battery size which implies longer all-electric ranges.

To this aim, the authors have formulated a procedure to systematically generate routes for an exhaustive evaluation of the energy benefits of CAVs (see Figure 1). It is divided into four broad processes: (1) Driving data generation, (2) Machine Learning (ML)-based synthetic route generation, (3) Route energy analysis and (4) Real-world route extraction.

In the Driving data generation process, a random route selector extracts trip characteristics (i.e. speed limits: v_{lim}, stop sign locations: s_{ss}, traffic light locations: s_{tl}, SPaT: ϕ_{SPaT}, etc.) that are fed together with driver parameters Ψ_{edm} into the EDM to populate the large driving data pool. This set can also be augmented with real-world driving data collected using Global Positioning System (GPS) or the On-Board Diagnostics (OBD) port

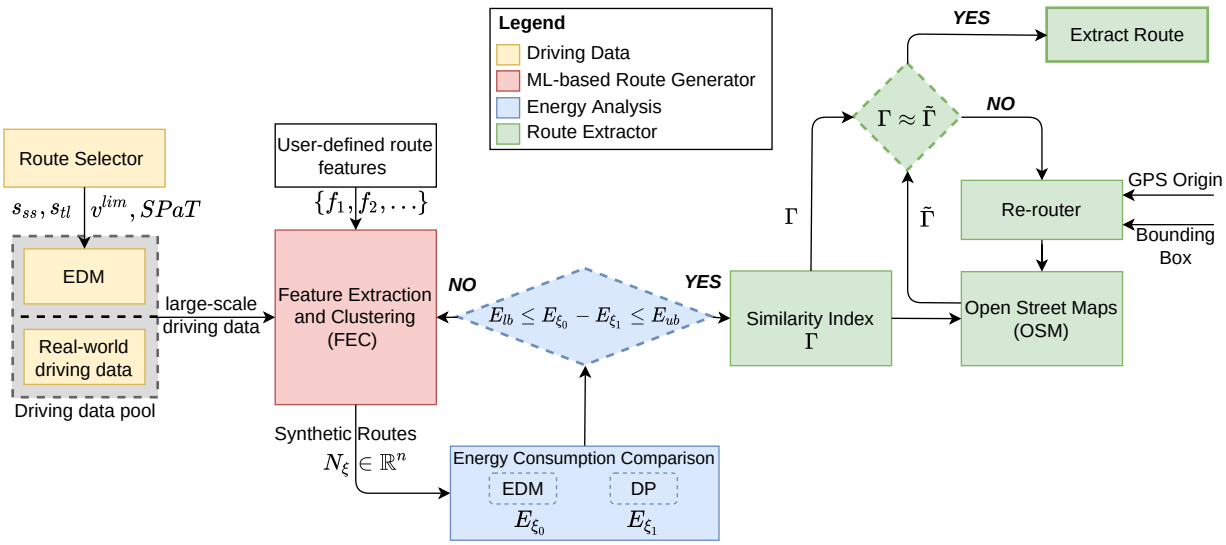

Fig. 1. Overall Route Generation Process

data. In the ML-based synthetic route generation process, user defined route metrics f_i including trip classifications and desired route lengths are fed along with the large-scale driving data obtained in the previous process into a Feature Extraction and Clustering (FEC) subsystem. The FEC generates synthetic routes that correspond to the user-defined route metrics. In the Energy Analysis process, an Energy-based Route Sensitivity (ERS) analysis is performed on the generated synthetic routes N_ξ by comparing the energy requests output from the EDM E_{ξ_0} and the DP E_{ξ_1}. All the synthetic routes $\hat{N}_\xi \in N_\xi$ for which the energy request variations are higher than within a defined bound are fed to the final process. In the Route extraction process, a Similarity Index Γ is computed as a function of route parameters and can be used as a quantifiable pass/fail metric for extracting routes from Open Street Map (OSM), or any other available mapping API that closely match the similarity index. Routes from OSM whose similarity index $\tilde{\Gamma} \approx \Gamma$ can then be extracted as viable routes to exhaustively evaluate energy efficiency benefits for CAVs.

3.1 Route Selector

A subset of 5 routes sampled from the *Columbus, OH* region were used as the basis for the route selector subsystem since the route data for each route was readily available. The corresponding SPaT profiles, extracted using SUMO, generated with the departure time as a random variable were used in the generation of large-scale driving data. Three driver aggressiveness parameter sets Ψ_{driver} (i.e. "aggressive", "normal" and "relaxed") were chosen and calibrated for each route and used to increase variability in the generation of the driving dataset. A Design-of-Experiment (DOE) has been setup where the large-scale driving dataset was generated using the EDM by varying $\{\Psi_{driver}, \phi^k, SPaT^k\}$. The generated results were then stored into a driving data pool that was composed also of real-world data collected through vehicle instrumentation e.g., via GPS or OBD-II odometry information.

3.2 Driving Sequence Analysis

To extract statistical metrics from the driving data, kinematic sequences are extracted from each trip in the driving data pool. A kinematic sequence, κ as formulated by Dembski et al. (2002) and for the context of this paper is defined as the vehicle velocity trajectory between consecutive stops (instances where vehicle is at zero speed). These stops correspond to either stopping at a stop sign or a traffic light, which correlates to idle times for a conventional powertrain. These stop times are appended to the end of each kinematic sequence. For each kinematic sequence, a set of statistical metrics $\in \mathbb{R}^d$ as defined in Table 1 is extracted from the driving data pool to form the feature set. The choice of metrics was motivated from related literature by Ericsson (2000) and Dembski et al. (2002) that characterize driving behavior in urban driving scenarios. Additionally, metrics that are correlated to energy usage while remaining powertrain agnostic are considered.

Table 1. Kinematic sequence statistical metrics

Parameter	Description	μ	σ
t_{κ_i}	Sequence duration, s	81	42
s_{κ_i}	Sequence distance, m	939	778
$t_{\kappa_i}^{stop}$	Stop time, s	14.4	12.7
$t_{\kappa_i}^{cruise}$	Cruise time, s	24.8	27.6
$v_{\kappa_i}^{max}$	Max. sequence velocity, m/s	18.5	4.1
$\overline{v}_{\kappa_i}$	Mean sequence velocity, m/s	10.4	4.4
$\sigma_{v_{\kappa_i}}$	SD sequence velocity, m/s	6.3	1.5
$a_{\kappa_i}^{max}$	Max. sequence acceleration, m/s^2	2.4	0.33
$b_{\kappa_i}^{max}$	Max. sequence deceleration, m/s^2	-2.1	0.12
$\overline{a}_{\kappa_i}$	Mean sequence acceleration, m/s^2	0.8	0.53
$\overline{b}_{\kappa_i}$	Mean sequence deceleration, m/s^2	-1.04	0.18
$\sigma_{a_{\kappa_i}}$	SD sequence acceleration, m/s^2	0.65	0.18
$\sigma_{b_{\kappa_i}}$	SD sequence deceleration, m/s^2	0.313	0.19
$t_{a_{\kappa_i}}$	Sequence acceleration time, s	21	14.7
$t_{b_{\kappa_i}}$	Sequence deceleration time, s	20.7	10.85
$\overline{v}_{\kappa_i}^{lim}$	Mean speed limit, m/s	-	-
$\Delta_{v_{\kappa_i}}^{max,lim,+}$	Max. speed limit increase, m/s	-	-
$\Delta_{v_{\kappa_i}}^{min,lim,}$	Min. speed limit increase, m/s	-	-

Trip Classification A driving trip refers to a complete set of kinematic sequences up to the destination. Trips are classified into three (3) broad categories derived from the sections of the WLTP drive cycle. The classification is based on the average trip velocity, $\overline{v}_{trip}$ in *mph* as shown in Equation 6 which captures variability due to speed limits and route markers.

$$\mathbb{Y}_{trip} = \begin{cases} 1 \implies \text{urban} & \overline{v}_{trip} < 20 \\ 2 \implies \text{mixed} & 20 \leq \overline{v}_{trip} \leq 30 \\ 3 \implies \text{highway} & \overline{v}_{trip} > 30 \end{cases} \quad (6)$$

Statistical Analysis of Driving Sequences For each trip, with variation in $\Xi_{route}, \Psi_{driver}, \Phi_{SPaT}$, the kinematic sequence metrics described in Table 1 are extracted and appended to a matrix $X \in \mathbb{R}^{n \times d}$. This results in a high dimension dataset with varying mean and variance for the various metrics. It is not immediately obvious which statistical metrics are dominant as well as the correlation between them.

To reduce the dimension of the dataset, Principal Component Analysis (PCA) introduced by Pearson (1901) is applied. From this, it is learnt that the 99% of the ratio of explained variance in the kinematic sequences is captured by 3 principal components, with the dominant statistical metrics being $t_\kappa, t_\kappa^{cruise}, t_{a_\kappa}, t_{b_\kappa}, \overline{v}_\kappa$.

Clustering of Driving Sequences An agglomerative hierarchical clustering approach is used to cluster the data recreated via PCA into groups. An agglomerative clustering approach was adopted to aid in making a correlation between the generated clusters and the trip classification of each kinematic sequence. By analyzing the dendrogram for different numbers of cluster looking, the data is clustered into three distinct groups that can be correlated to the original trip classification via a probability matrix as shown in Table 2. Figure 2 shows the three clusters visualized with three of the dominant statistical metrics: $t_{a_\kappa}, t_\kappa, t_{b_\kappa}$, confirming the process accuracy.

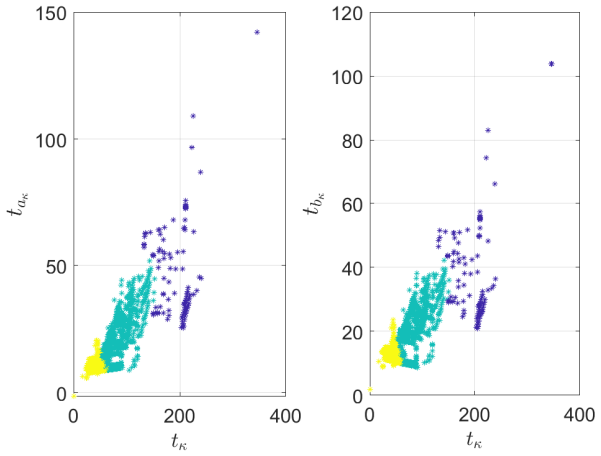

Fig. 2. Visualization of clusters generated from data for dominant metrics $t_{a_\kappa}, t_\kappa, t_{b_\kappa}$.

3.3 Synthetic Route Generator

After computing the probability matrix, the process for developing a robust synthetic route generator has been studied. The objective is to create new routes exploiting the large pool of data collected and then clustered. In general, the synthetic route generator is an iterative process where given different user requirements, the tool can output a large batch of missions complying with them. The flowchart describing the procedure adopted in the synthetic route generation is illustrated in Fig. 3. The first step is to choose the trip length and classification (i.e. urban, mixed or highway) that can be considered the minimum inputs to provide. The process starts by dividing the range $[0, 1]$ into n-parts (where n is the number of clusters detected, i.e. three) each with a length depending on the results of the probability matrix. Then, a random number n_{rdm} from a uniform distribution is generated and its value informs which cluster to pick the kinematic sequence κ from (based on the probability ranges $\mathbb{P}_{C1}$, $\mathbb{P}_{C2}$, and $\mathbb{P}_{C3}$). If the criteria are met, the kinematic sequence κ_i is appended to the synthetic route N_ξ and the whole process is re-iterated until the desired trip length ($\pm$ 2% tolerance) is met. Once the number of synthetic routes generated is sufficient enough, further checks can be performed to output missions that precisely resemble the initial requirements. This additional step is required to avoid routes that have repetitive κ_i or where the mean speed does not comply with the driving scenario requirements.

Table 2. Probability of occurence of clusters in trip classification

	Cluster 1	Cluster 2	Cluster 3
$\mathbb{Y} = 1$ (urban)	0	0.2383	0.7617
$\mathbb{Y} = 2$ (mixed)	0.0972	0.516	0.3868
$\mathbb{Y} = 3$ (highway)	0.2906	0.5554	0.154

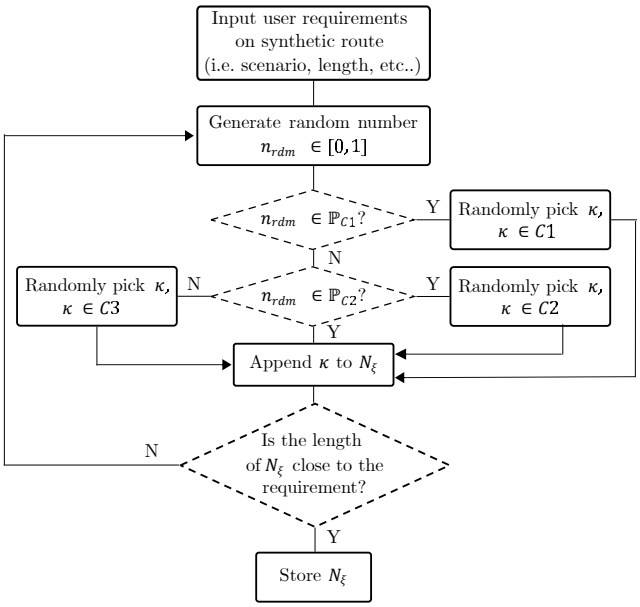

Fig. 3. Flowchart of the synthetic route generation procedure

3.4 Synthetic Route Energy Analysis

The objective of this study is to quantify the potential energy savings of CAVs over the generated synthetic routes. To this aim, the power requested at the wheels has been

assessed using the EDM by performing a full-factorial simulation over the aggressiveness parameter set Ψ_{edm} to exhaustively explore the trade-off of energy consumption and travel time. Then, the EDM power request output, resulting net energy, and travel time over the synthetic routes has been compared to the results obtained using a globally optimal controller algorithm, Dynamic Programming (DP) first studied by Bellman and Lee (1984). This algorithm has been implemented in *Matlab* and the *DynaProg* toolbox provided by Miretti et al. (2021) has been used, due to the relatively fast computation compared to other DP solvers in the literature. In general, a DP-based algorithm provides the solution to an Optimal Control Problem (OCP) by running a backward simulation and computing at each step the cost function for every discretized control value at each discretized state value. Then, the optimal trajectory is found in the forward simulation, i.e. the trajectory associated with the minimum cost capable of simultaneously complying with the set constraints. The vehicle dynamics equations used in the OCP formulation are detailed in Equation 7.

$$\begin{aligned} \omega_{wheel}(t) &= v(t) \cdot r_{wheel} \\ F_{road}(t) &= \big(A + B \cdot v(t) + C \cdot v^2(t)\big) \\ T_{wheel}(t) &= (F_{road}(t) + m_{veh} \cdot a(t)) \cdot r_{wheel} \\ v(t+1) &= v(t) + a(t) \cdot \Delta t \end{aligned} \tag{7}$$

Where ω_{wheel} is the rotational speed of the wheels; v and r_{wheel} are the velocity of the vehicle and the tire radius, respectively; A, B, and C are the road load coefficients used to compute the resistive road force F_{road}; T_{wheel} is the resistive torque at the wheels; m_{veh} and a are the vehicle mass and its acceleration; lastly, Δt is the timestep. The parameters in Equation 7 have been discretized via Euler forward method and then reformulated to express them in distance domain due to the nature of the exogenous inputs (i.e. speed limits and stop locations are expressed as a function of the distance). Moreover as seen from the vehicle dynamics equations, the only vehicle data needed are mass, tire radius and road load coefficients that have been extracted from US EPA (2022), making the implementation powertrain agnostic. The vehicle selected for this analysis is a 2021 Chrysler Pacifica Hybrid PHEV.

The control and state vectors along with the cost function for the OCP formulation are detailed in Equation 8. The optimal trajectory achieves the minimization of a trade-off between energy at the wheels and travel time according to Equation 9:

$$u = [a_s] \in \mathbb{R}, \qquad x = \begin{bmatrix} v_s \\ a_{s-1} \end{bmatrix} \in \mathbb{R}^2$$

$$J_s(x_s) = \sum_{i=1}^{N-1} \Big(\gamma \frac{P_{i,wheels}}{P_{norm}} + (1-\gamma)\Big)\Delta t_i + \mu_{jerk}\big|\Delta a_i\big| \tag{8}$$

$$\begin{aligned} \min_{u} \quad & J_s(x_s) \\ \text{subject to} \quad & x_{i+1} = x_i + f(x,u) \cdot \Delta t \\ & v_{min}(s_i) \le v_s(s_i) \le v_{max}(s_i) \\ & a_{min} \le a_s(s_i) \le a_{max} \\ & v_s(s_i = s_{stop}) = 0 \end{aligned} \tag{9}$$

The stage cost J_s is made of two terms mainly, the ratio between the instantaneous power requested at the wheels and a normalization factor P_{norm}, and the timestep. Even though the focus of this study is on the energy analysis, the second term has been crucial to avoid any trivial solution (e.g. null speed at all steps). Furthermore, the acceleration variation in the stage Δa_i has been included in the running cost to penalize rapid changes in the acceleration. Two (2) different weights have been formulated, namely γ and μ_{jerk}. $\gamma \in [0,1]$ is used to study trade-off between energy minimization and travel time; μ_{jerk} is a fixed weight to balance the derivative of the acceleration. Physical constraints on acceleration and velocity have been set on lower and upper boundaries based on the analysis of regulatory drive cycles and route speed limits. The acceleration lies in the range $[-2, 2]\, m/s^2$, whereas the velocity is bounded to have values greater than or equal to zero but lower than the speed limits. To ensure that the vehicle comes to a stop at the desired stop locations, a null velocity has been imposed in the speed limit constraints at every stop.

4. ENERGY ANALYSIS RESULTS

In this section, the energy comparison between the baseline EDM and the DP-based optimal controller is found. To provide an exhaustive analysis, five different synthetic routes have been generated using the methodology described above for a mixed driving scenario. For each route, a multitude of EDM parameters Ψ_{driver} and DP weights γ have been tested to ensure the robustness of the proposed procedure. In Fig. 4 the velocity and power demanded at the wheels graphs are illustrated for a 15 km synthetic route in mixed scenario. In the upper sub-plot, the red line corresponds to the vehicle velocity determined by the DP-based algorithm while the blue indicates the velocity profile prescribed by EDM, while the grey line indicate the speed limits with the stops corresponding to the points where $v_s = 0$. As can be seen by simultaneously analyzing speed and power, the DP controlled vehicle resulted in a lower energy request throughout the route. This mainly stems from less aggressive accelerations and decelerations along with avoiding abrupt accelerations for short periods of time. In the first part of the trip, the DP-based

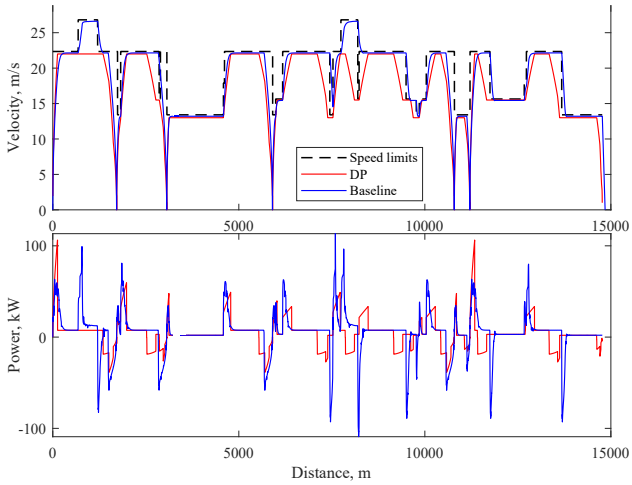

Fig. 4. Velocity and wheels power comparison between the EDM and DP controllers in a mixed driving scenario

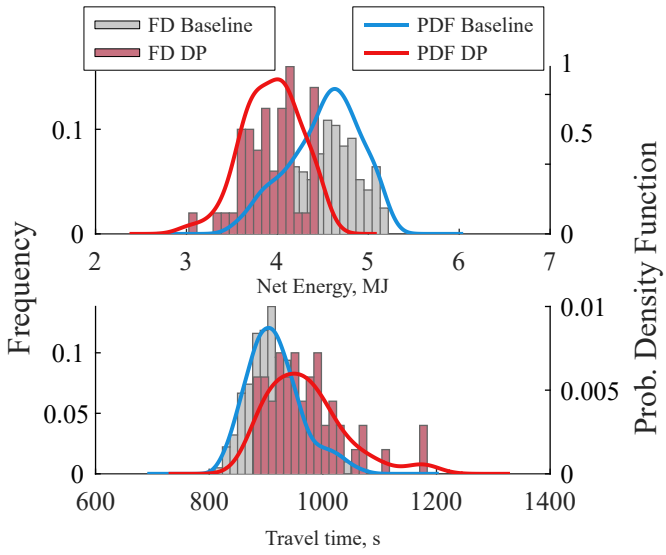

Fig. 5. Frequency Distribution and Probability Density Function comparison for the net energy (graph above) and travel time (graph below) in a *mixed* scenario with different drivers parameters

controller does not accelerate up to the new speed limit avoiding an increase in power request. In this synthetic route illustrated, the baseline EDM net energy is equal to 4.26 MJ whereas the DP resulted in 3.43 MJ with a variation in net energy demand of 24.2% with respect to the baseline. With regards to the travel time, the EDM takes 884 s to run the 15 km whereas the DP-based controller spends 984 s resulting in an increase of approximately 11%.

To provide a broader and more complete analysis, the probability density function (PDF) for both EDM and DP-based controller has been computed for the net energy request and travel time in 5 synthetic routes generated. The PDF can be used to statistically provide the likelihood of a certain variable to be in a range of values, and Fig. 5 shows both the raw frequency distribution (FD shown as histograms) and the PDF for net energy and travel time in a mixed scenario. As evident from this figure, it is more likely to obtain substantially lower values of net energy in the DP-based controller than the baseline, whose mean is 4.51 MJ with respect to 3.93 MJ obtained by the optimal strategy (i.e. a variation of approximately 14.8%). Concerning the travel time, it is clear that the DP more likely results in higher values with an average of 970 s compared to 914 s obtained by the baseline EDM (roughly 6% difference). This statistics tool appropriately depicts the differences between the EDM and the DP controllers to prove the remarkable improvements arising from an optimal strategy, which in turn could be seen as the improvements due to the implementation of Connected and Automated technologies. For the purpose of this study, the synthetic routes showing a net energy variation that corresponds to the dominant modes from the PDF will be considered for the route extraction process.

5. CONCLUSION AND FUTURE WORK

The evaluation of potential energy savings stemming from the implementation of connected and automated driving cannot be exhaustively performed using conventional driving cycles. The proposed study provides a procedure for the creation of synthetic routes starting from real-world and virtually simulated driving data. This data is converted into kinematic sequences and then clustered accordingly through a feature extraction and analysis tool. Then, a batch of synthetic routes is generated and the power request at the wheels is assessed using the Enhanced Driver Model and a Dynamic Programming based algorithm. The synthetic routes in which the dominant energy consumption are found are considered feasible candidates for the Open Street Map route extraction and to later assess the potential benefits of Connected and Automated Vehicles. Significant variations in net energy request have been demonstrated in the example shown, for five (5) generated synthetic routes of 15 km long depicting a mixed driving scenario. The probability density functions obtained by the raw frequency distributions clearly illustrate a lower net energy request when the DP-based controller drives the vehicle, with a difference in mean values of approximately 14.8%. The EDM is more likely to show lower travel times, with a reduction of roughly 6% with respect to the DP strategy.

Work on the route extraction procedure to find real-world routes that resembles the synthetic routes is already underway. This involves using an open-source mapping resource, namely *OpenStreetMaps (OSM)* to extract these real-world routes by performing a similarity scoring between the synthetic route and the real-world routes.

ACKNOWLEDGEMENTS

Our team acknowledges the support from the United States Department of Energy, Advanced Research Projects Agency – Energy (award number DE-AR0000794) that made this work possible.

REFERENCES

Bellman, R. and Lee, E. (1984). History and development of dynamic programming. *IEEE Control Systems Magazine*, 4(4), 24–28. doi:10.1109/MCS.1984.1104824.

Dembski, N., Guezennec, Y., and Soliman, A. (2002). Analysis and experimental refinement of real-world driving cycles. *SAE Technical Paper*. doi:10.4271/2002-01-0069.

Dingyi, Y., Haiyan, W., and Kaiming, Y. (2018). State-of-the-art and trends of autonomous driving technology. In *2018 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE)*, 1–8. doi:10.1109/TEMS-ISIE.2018.8478449.

Ericsson, E. (2000). Variability in urban driving patterns. *Transportation Research Part D: Transport and Environment*, 5(5), 337–354. doi:https://doi.org/10.1016/S1361-9209(00)00003-1. URL `https://www.sciencedirect.com/science/article/pii/S1361920900000031`.

Ericsson, E. (2001). Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment*, 6(5), 325–345. doi:https://doi.org/10.1016/S1361-9209(01)00003-7. URL `https://www.sciencedirect.com/science/article/pii/S1361920901000037`.

Gong, Q., Midlam-Mohler, S., Marano, V., Rizzoni, G., and Guezennec, Y. (2010). Statistical analysis of phev fleet data. In *2010 IEEE Vehicle Power and Propulsion Conference*, 1–6. doi:10.1109/VPPC.2010.5729224.

Gupta, S., Deshpande, S.R., Tulpule, P., Canova, M., and Rizzoni, G. (2019). An enhanced driver model for evaluating fuel economy on real-world routes. *IFAC-PapersOnLine*, 52(5), 574–579. doi:https://doi.org/10.1016/j.ifacol.2019.09.091.

Kesting, A., Treiber, M., and Helbing, D. (2010). Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928), 4585–4605. doi:10.1098/rsta.2010.0084.

Miretti, F., Misul, D., and Spessa, E. (2021). Dynaprog: Deterministic dynamic programming solver for finite horizon multi-stage decision problems. *SoftwareX*, 14, 100690. doi:https://doi.org/10.1016/j.softx.2021.100690.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. doi:10.1080/14786440109462720.

Rajakumar Deshpande, S., Gupta, S., Kibalama, D., Pivaro, N., and Canova, M. (2020). Benchmarking Fuel Economy of Connected and Automated Vehicles in Real World Driving Conditions via Monte Carlo Simulation. *Dynamic Systems and Control Conference*, Volume 1. doi:10.1115/DSCC2020-3250. V001T10A004.

Rajakumar Deshpande, S., Gupta, S., Kibalama, D., Pivaro, N., Canova, M., Rizzoni, G., Aggoune, K., Olin, P., and Kirwan, J. (2021). In-vehicle test results for advanced propulsion and vehicle system controls using connected and automated vehicle information. *SAE International Journal of Advances and Current Practices in Mobility-V130-99EJ*. doi:https://doi.org/10.4271/2021-01-0430.

US EPA (2020). *Dynamometer Drive Schedule*. `https://www.epa.gov/vehicle-and-fuel-emissions-testing/dynamometer-drive-schedules`, [Accessed: 02/25/2022].

US EPA (2022). *Data on Cars used for Testing Fuel Economy*. `https://www.epa.gov/compliance-and-fuel-economy-data/data-cars-used-testing-fuel-economy`, [Accessed: 02/25/2022].