Eco-Driving Data Imputation

May 5, 2019

Some driving behaviors may have dramatic impacts on the fuel consumption. We are investigating the eco-driving strategy from speed profiles in GPS trajectories. Each profile is a time series recording the real-time speed of a vehicle. Let it be $\mathbf{x} = \{(t_1, s_1), (t_2, s_2), \ldots, (t_n, s_n)\} \in \mathcal{X}$, where t_i is the timeframe and $s_i \in S$ is the accompanied non-temporal state at t_i, and the time duration of the profile $\mathbf{x}$ is $T_{\mathbf{x}} = t_n - t_1 > 0$, which may be in minutes or hours and n is the number of recorded entries. The time series may have varied lengths and are allowed to be unevenly distributed in the time space, i.e. the short time durations in $\{t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}\}$ may be distinct. To be simple, we assume that the time series is in unit second. To measure the fuel usage based on the speed profile, an energy consumption function is defined as $\mathcal{C} : \mathcal{X} \to R^+$. Given a speed profile of any length, it's able to produce a reliable estimation of the *average* energy consumption of the profile.

Given a sample $\mathbf{x} \in \mathcal{X}$ with sparse records, for instance $n = 2$ and $T_{\mathbf{x}} = 1$ minutes. A data imputation technique $\mathcal{A}$ will be deployed to fill in the missed records at a higher resolution τ, say 1 seconds. We evaluate $\mathcal{A}$ based on the fuel consumption

$$\mathcal{M}_{\mathcal{A}} = P\Big\{|\mathcal{C}(\mathcal{A}(\mathbf{x})) - \mathcal{C}(\mathbf{x}^*)| \le \epsilon\Big\}, \tag{1}$$

where $0 < \epsilon \ll 1$, and $\mathbf{x}^*$ is a sample drawn from the GPS data set, $\mathbf{x}$ is a sparser sequence with many missed records. A higher $\mathcal{M}_{\mathcal{A}}$ implies a better imputation technique $\mathcal{A}$.

Two simple imputation techniques are evaluated in the report:

- Straight Line: Let $\mathbf{z} = \mathcal{A}(\mathbf{x})$, and $g = (s_n - s_1)/T_{\mathbf{x}}$ be the slope or the gradient in unit time, then $z_1 = (t_1, s_1), z_2 = (t_1 + \tau, s_1 + g), \ldots, z_m = (t_1 + (m-1)\tau, s_1 + (m-1)g), z_{m+1} = (t_n, s_n)$. Here, it's assumed that $n = 2$ and $T_{\mathbf{x}} = m\tau$.
- Straight Line With Random Perturbation: Based on the above straight line imputation, a random perturbation is imposed over the imputed data, i.e. $z_k = (t_1 + (k-1)\tau, s_1 + (k-1)g + r)$, $k = 2, 3, \ldots, m$, $z_1 = (t_1, s_1)$ and $z_m = (t_n, s_n)$. The random variable $r \sim Q(r)$, where Q could be a uniform distribution $Q(r) \sim U(0,1)$ to be the relative deviation from the straight line imputation, e.g. $r = 0.1$ is less or greater than the imputed value, or it is the probability distribution of deviation from the straight line imputation in the empirical data set, as illustrated in Fig. 1.

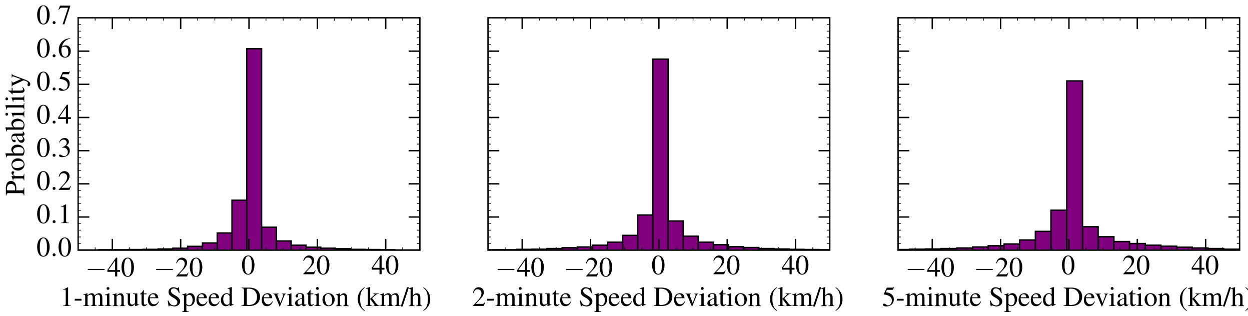

Fig. 1: Probability distribution of speed deviation from linear interpolation of varied time-lengths: 1, 2 and 5 minutes.

There are many different energy consumption functions that can be used. One is defined as

$$\mathcal{C}(s_i) = \mathcal{C}(\{v_i, m_i\}) = \begin{cases} a \times \exp\{b_1 + (c_1 \times \ln v_i)\} v_i, & \ln v_i < 3.75 \text{ and } m_i = 1, \\ a \times \exp\{b_2 + (c_2 \times \ln v_i)\} v_i, & \ln v_i \geq 3.75, \\ 0, & else. \end{cases} \tag{2}$$

where v_i is the instantaneous speed of the truck, $a = 8.80618107 \times 10^{-5}/(3600 \times 1.60934) = 1.519978 \times 10^{-8}$, $b_1 = 8.185202 + 0.002580905/2 = 8.1864924525$, $b_2 = 3.178949 + 0.002580905/2 = 3.1802394525$,$c_1 = -0.480677 - 0.069 = -0.549677$, $c_2 = 0.8072496 - 0.0546 = 0.7526496$. The energy consumption function $\mathcal{C}(\mathbf{x}) = \sum_{i=1}^{n} \mathcal{C}(s_i)$ relies on the speed and the variable *Movement* m_i. To be simple, we simply put all imputed the same value, i.e. setting *Movement* to be 1.

Another energy consumption function used in our simulation comes from the vehicle emission model proposed by Emrah Demir and his colleagues[1], which estimates the tractive power requirements (kw)

$$P_{tract} = (M \times a + M \times g \times \sin\theta + 0.5 C_d \rho A v^2 + M g C_r \cos\theta) v / 1000.$$

where v is the speed (m/s), and M is the weight (kg), with ρ is the air density in kg/m^3 (typically 1.2041), A is the frontal surface area in m^2 (typically between 2.1 and 5.6), and g is the gravitational constant in m/s^2 (typically 9.81). In addition, C_d is the coefficient of aerodynamic drag (typically 0.7), and C_r the coefficient of rolling resistance (typically 0.01). Here, we choose $M = 5000$ kg, $A = 5.6$ and $\theta = 0$. To translate the tractive requirement into engine power requirement, we calculate

$$P = P_{tract}/\eta_{tf}$$

where η_{tf} is the vehicle drive train efficiency (typically 0.4). It's known that 1 gallon of gas can produce energy of about $\alpha = 33.410107$ kwh. Therefore, we can define an alternative energy consumption function

$$\mathcal{C}(s_i) = \frac{(M a_i + 0.5 C_d \rho A v_i^2 + M g C_r) v_i}{3.6 \alpha \eta_{tf} \times 10^6}. \tag{3}$$

It relies on the acceleration rate a_i, which is calculated from the change in speed, as a result we have

$$\mathcal{C}(\mathbf{x}) = \sum_{i=1}^{n-1} \mathcal{C}(s_i).$$

We select a speed profile of one specific route as described in Fig. 2 (Upper), and calculate the related instantaneous (Middle) and cumulative energy consumption (Lower) with the two designed energy consumption functions introduced in the above paragraphs. The two curves indicates the two functions in essential are the same. We also notice that the instantaneous fuel consumption presents similar pattern as the speed profile. To measure the similarity, we calculate the cosine similarity of the two, and find that value is as as high as $\cos(\mathbf{v},\mathbf{c}) \approx 0.99$. In other words, the energy consumption functions heavily rely on the speed.

Two simulation schemes are conduct, one over all possible episodes that take 1, 2 and 5 minutes $T_{\mathbf{x}} = 60s, 120s, 300s$ to run, which has overlapping episodes, while another over all non-overlapping episodes that take 1, 2 and 5 minutes' driving time. Scheme 1 is applying the imputation techniques over a moving sliding window in the driving profile, the size of the sliding window could be 1, 2 and 5 minutes. Once a sliding window has been done, then move to the next one. Scheme 2 split the entire driving history into multiple cells of equal size, each cell is corresponding to a short driving profile of length 1, 2 or 5 minutes.

The simulations are evaluated with the fuel consumption saving due to the imputation techniques relative to the real fuel consumption. The evaluations are reported in Fig. 3 and 4. Because the two fuel consumption functions are essentially the same, hereinafter the fuel consumption is calculated using the first fuel consumption function.

Fig. 3 reports the instantaneous fuel consumption saving following the trajectory generated from the imputation techniques, when compared against with the real fuel consumption of all 12 routes under scheme 1, and Fig. 4 reports the same result under scheme 2. According to the simulation, the two simple imputation techniques have no significant difference when compared against each other, and both give good imputation to the original records. Under scheme 1, let $\epsilon = 0.01$, $\mathcal{M}_{\mathcal{A}} > 0.95$ when $T_{\mathbf{x}} = 60$, $\mathcal{M}_{\mathcal{A}} \approx 0.77$ when $T_{\mathbf{x}} = 120$, and $\mathcal{M}_{\mathcal{A}} \approx 0.51$ when $T_{\mathbf{x}} = 300$ for both techniques. Under scheme 2, let $\epsilon = 0.001$, $\mathcal{M}_{\mathcal{A}} > 0.99$ when $T_{\mathbf{x}} = 60, 120$ or 300 for both techniques. No matter under which simulation scheme, the sparser the samples become, the worse the imputation tend to be, as the gap becomes larger. According to the range the fuel consumption saving, scheme 1 is much greater than scheme 2, because it evaluates all possible short route of specific time length.

The random perturbation imputation makes a perturbation over the straight line imputation. The perturbation could be uniform distributed or collected from the empirical data. To make a comparison, we also report the results based on the empirical distribution (see Fig. 1) in Fig. 5. Overall, the empirical distribution does not provide much information to improve the imputation results w.r.t the fuel consumption.

To have a close examine the performance of the imputation techniques, we illustrate the instantaneous fuel consumption in Fig. 6 of all 12 routes w.r.t the real speed profile, the imputed speed profile with straight line and straight line based random perturbation of 2 minutes length, and the corresponding cumulative fuel consumption of one individual route in Fig. 7.

References

1. Demir, E., Bektaş, T. & Laporte, G. A comparative analysis of several vehicle emission models for road freight transportation. *Transportation Research Part D: Transport and Environment* **16**, 347–357 (2011).

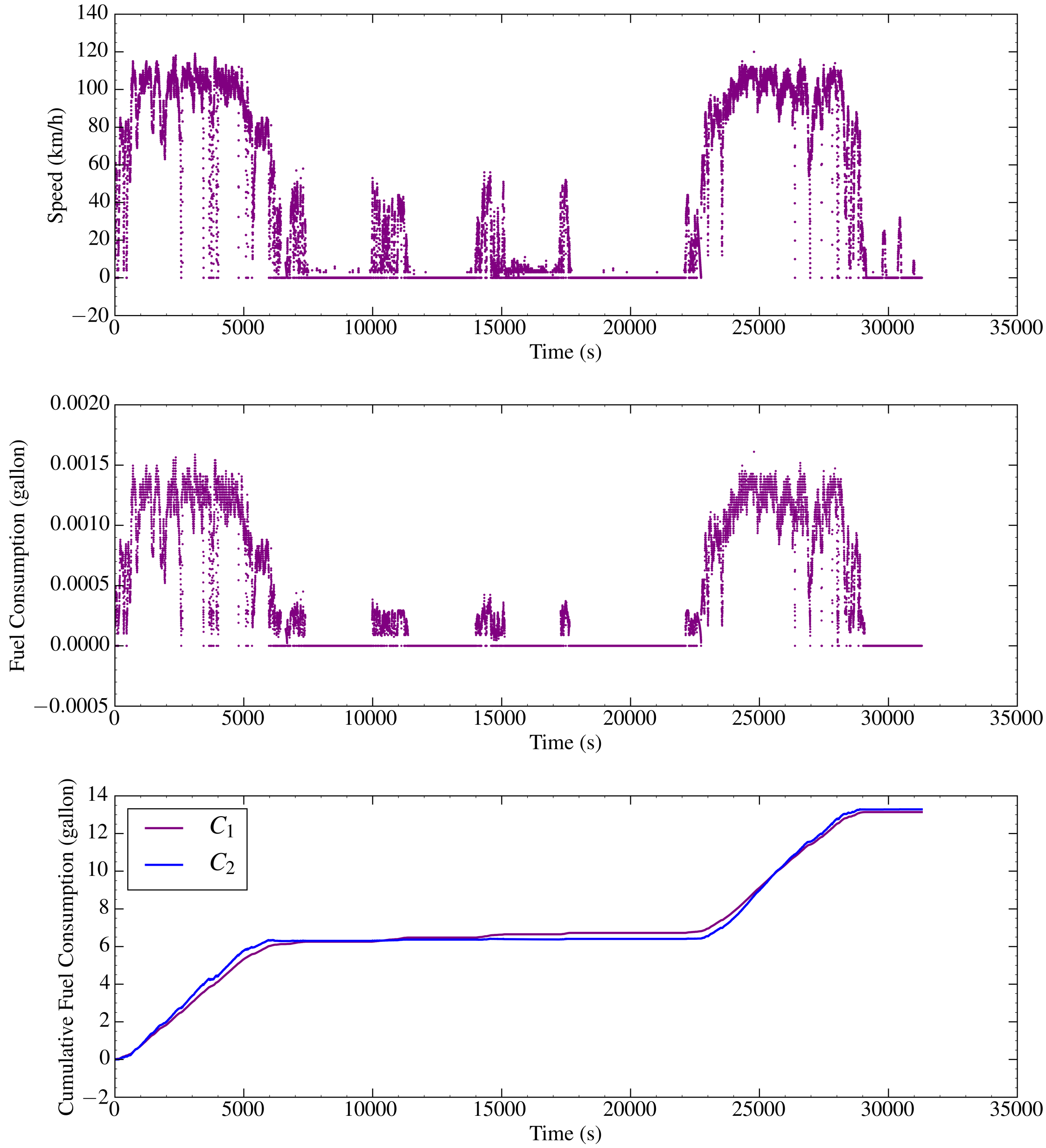

Fig. 2: Speed Profile of a Truck (Upper), Instantaneous Fuel Consumption (Middle) and Its Curve of Cumulative Fuel Consumption (Lower) on One Specific Route

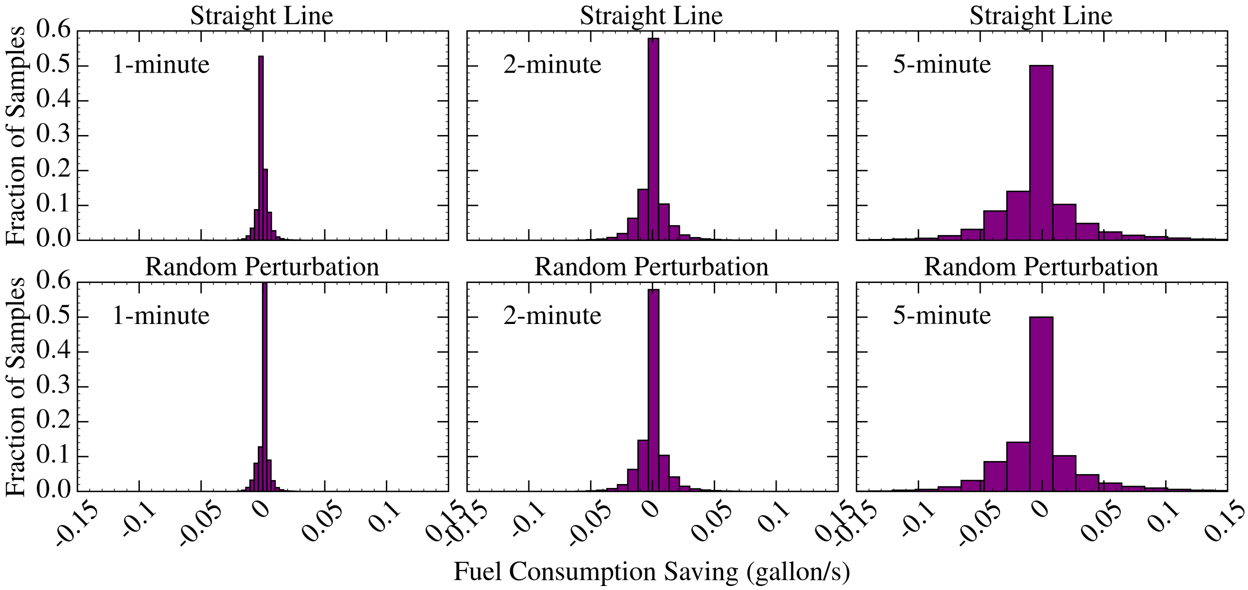

Fig. 3: Instantaneous fuel consumption saving due to the imputation techniques relative to the ground truth consumption of all 12 routes under simulation scheme 1. Each column is corresponding to a time-length when employing the imputation technique (Straight Line in row 1 and Random Perturbation in row 2). Three time-lengths are examined, they are 1-minute, 2-minute and 5-minute.

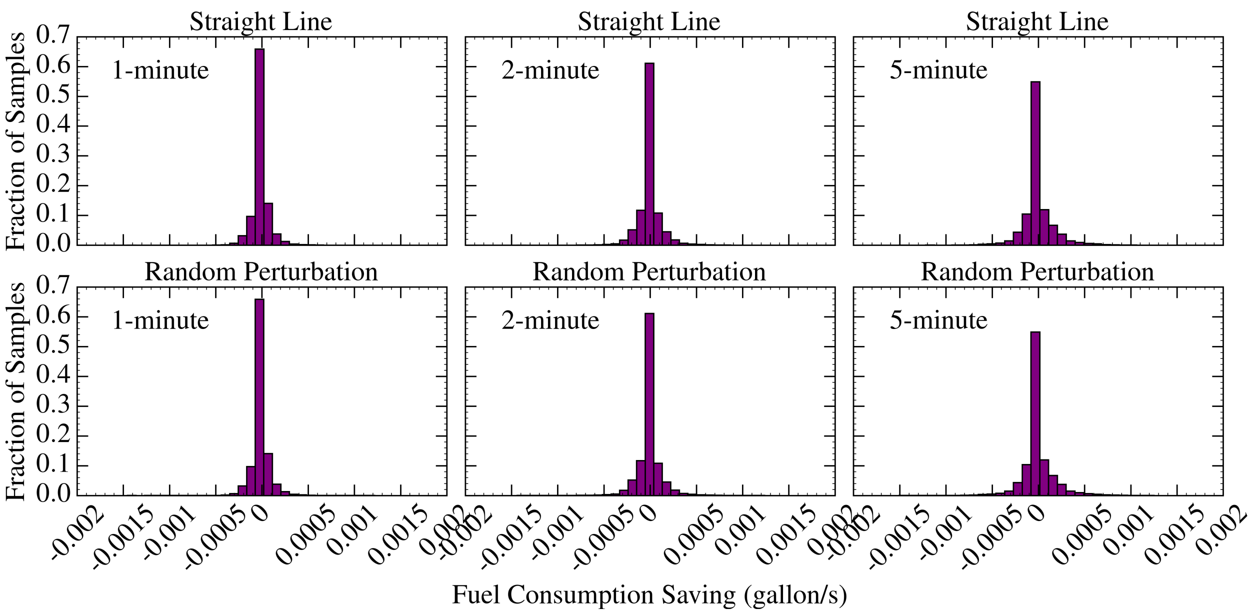

Fig. 4: Instantaneous fuel consumption saving due to the imputation techniques relative to the ground truth consumption of all 12 routes under simulation scheme 2. Each column is corresponding to a time-length when employing the imputation technique (Straight Line in row 1 and Random Perturbation in row 2). Three time-lengths are examined, they are 1-minute, 2-minute and 5-minute.

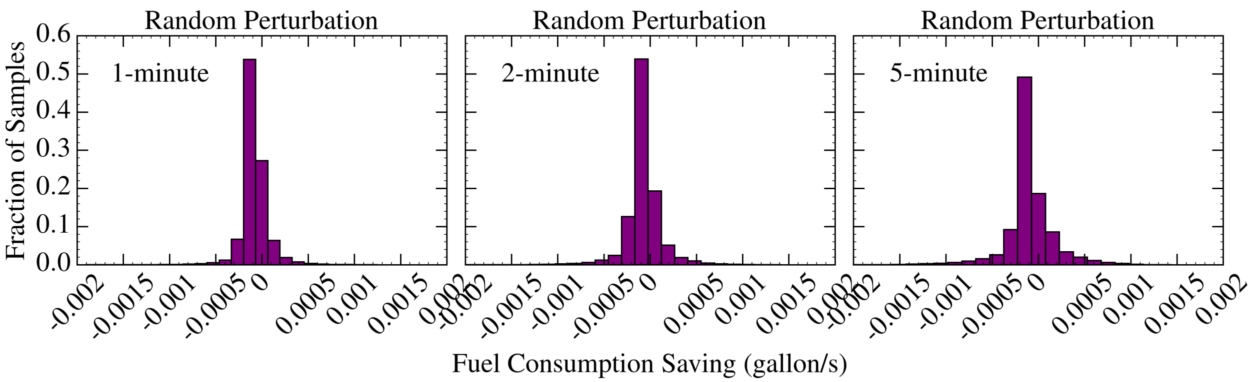

Fig. 5: Instantaneous fuel consumption saving due to the imputation techniques relative to the ground truth consumption of all 12 routes under simulation scheme 2. The perturbations are drawn from the empirical distribution.

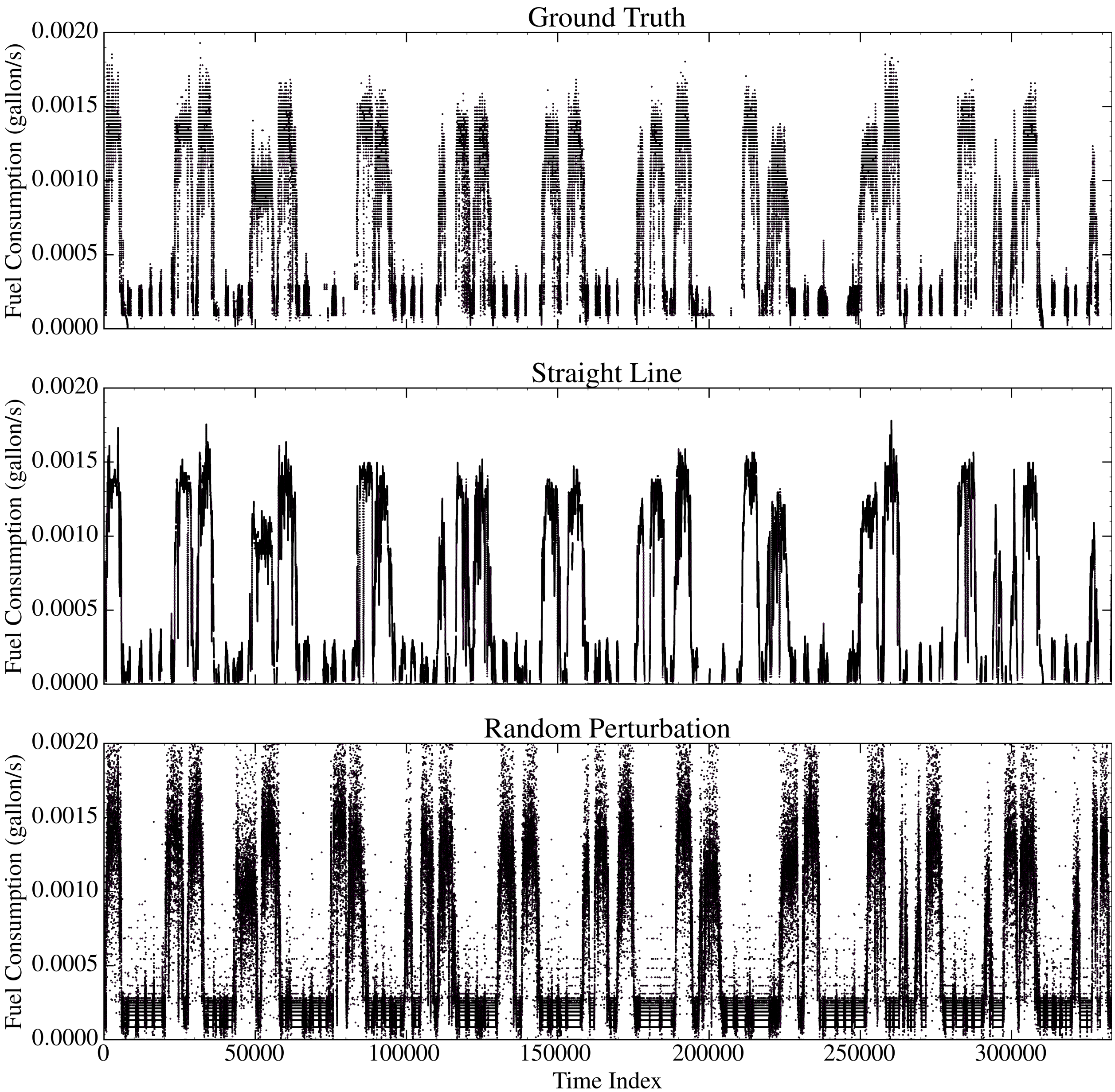

Fig. 6: Instantaneous Fuel Consumption on Ground Truth Speed Profile, Imputed Profile with Straight Linear and Straight Linear based Random Perturbation Imputation Techniques (120s) for All 12 Routes

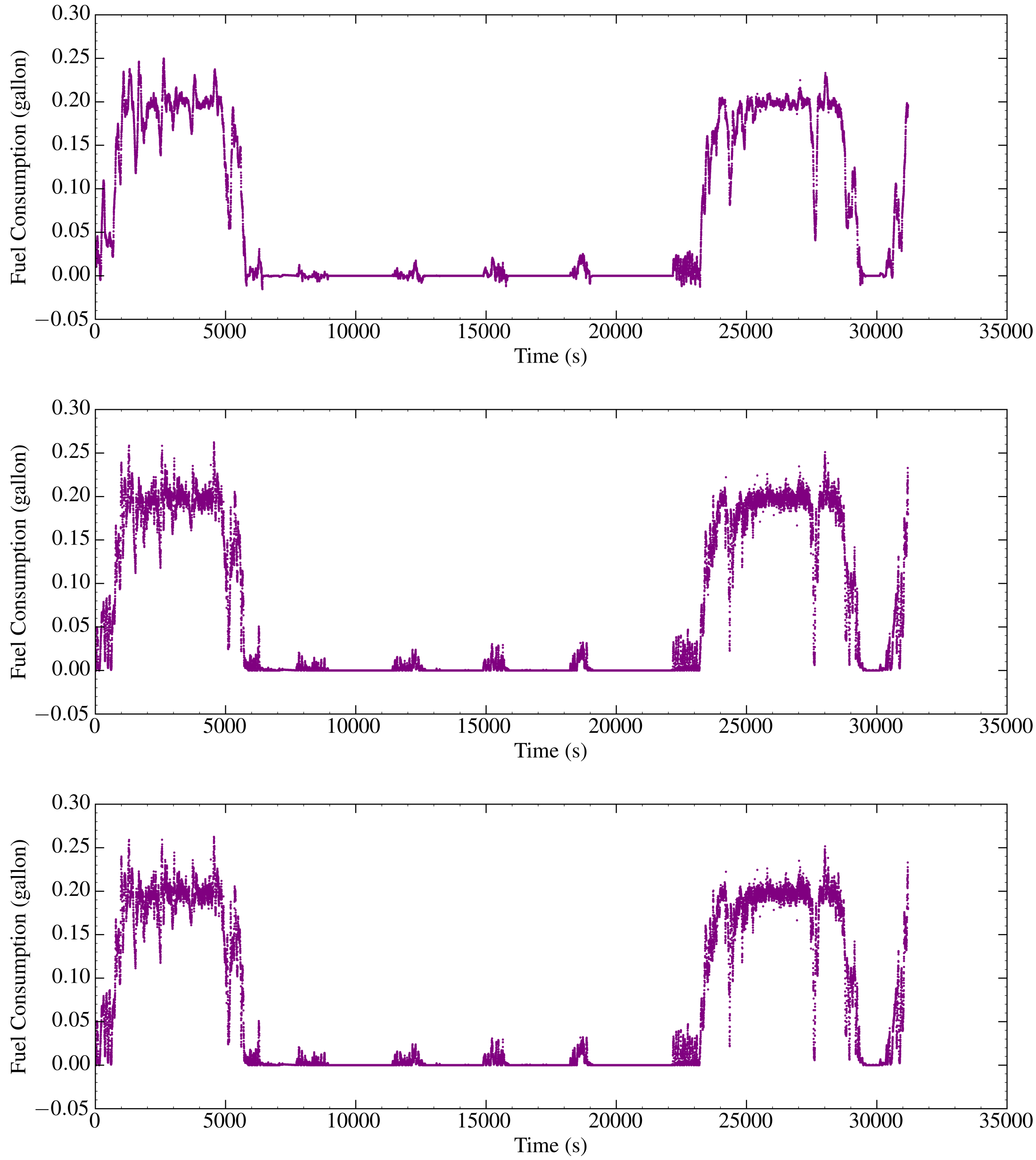

Fig. 7: Cumulative Fuel Consumption of 2-Minute Short Trip, **Upper**: Real Consumption, **Middle**: Consumption for Linear Imputed Profile and **Lower**: Consumption for Random Perturbation Imputed Profile. The x-axis represents the start time index of each short trip.