This is the published version of a paper presented at *IEEE 15th International Conference on Machine Learning and Applications, ICMLA 2016, Anaheim, United States, 18-20 December, 2016.*

N.B. When citing this work, cite the original published paper.

Learning of aggregate features for comparing drivers based on naturalistic data

Iulian Carpatorea, Sławomir Nowaczyk, Thorsteinn Rögnvaldsson, Marcus Elmer, Johan Lodin

Abstract—Fuel used by heavy duty trucks is a major cost for logistics companies, and therefore improvements in this area are highly desired. Many of the factors that influence fuel consumption, such as the road type, vehicle configuration or external environment, are difficult to influence. One of the most under-explored ways to lower the costs is training and incentivizing drivers. However, today it is difficult to measure driver performance in a comprehensive way outside of controlled, experimental setting.

This paper proposes a machine learning methodology for quantifying and qualifying driver performance, with respect to fuel consumption, that is suitable for naturalistic driving situations. The approach is a knowledge-based feature extraction technique, constructing a normalizing fuel consumption value denoted *Fuel under Predefined Conditions* (FPC), which captures the effect of factors that are relevant but are not measured directly.

The FPC, together with information available from truck sensors, is then compared against the actual fuel used on a given road segment, quantifying the effects associated with driver behavior or other variables of interest. We show that raw fuel consumption is a biased measure of driver performance, being heavily influenced by other factors such as high load or adversary weather conditions, and that using FPC leads to more accurate results. In this paper we also show evaluation the proposed method using large-scale, real-world, naturalistic database of heavy-duty vehicle operation.

I. INTRODUCTION

The majority of goods transport within Europe is done on roads using heavy duty vehicles, which means that decreasing the associated fuel consumption is very important. For example, those trucks contribute with a large portion of CO_2 emissions, approximately 20%. In addition, fuel expenses are up to 40% of the operating cost for the truck fleet (cf. Barnes and Langworthy [1]). Various aspects of this issue are being addressed by both the industry and the research community. In this paper we focus on the influence of drivers, who can make up to 30% difference in terms of fuel consumption, according to Nylund [2],

The most common way of expressing vehicle fuel usage is the amount of fuel consumed divided by distance traveled, typically expressed as liters per 100 kilometers. This metric is also often used as a way of comparing the performance of drivers. However it is, in many circumstances, inadequate for identifying areas for improvement and as an incentive mechanism. One example is that for heavy duty trucks the gross weight (including cargo) can be over ten times the weight

Iulian Carpatorea, Sławomir Nowaczyk and Thorsteinn Rögnvaldsson are with Center for Applied Intelligent Systems Research, Halmstad University, Sweden (email: firstname.lastname@hh.se). Marcus Elmer and Johan Lodin are with Volvo Group Trucks Technology, Göteborg, Sweden.

of the tractor. Since the fuel consumption is directly dependent on the total weight, it is clear that its range is very wide, and highly dependent on the specifics of the individual mission. The effect of the driver is not as prominent, and to make it visible, it is necessary to use a different scale.

Much existing literature on comparing driver performance is based on predicting how much fuel would be used given different actions. Bifulco et al. [3] developed a method for calculating instantaneous fuel consumption based on speed, acceleration, gas pedal position and engine air intake. Constantinescu et al. [4] derived 5 categories of aggressiveness for the driver based on a number of predefined driving parameters, such as positive acceleration. Typically, aggressiveness is associated with performance. However, it shares many drawbacks of fuel consumption when used as performance indicator, since it is also missing normalization for drivers operating in different conditions.

The US patent [5] proposes using pre-calculated reference values for selected road segments, capturing both environment variables as well as vehicle operation characteristics. The fuel consumption reference value for a given segment is then modified by fuel consumption modeled based on deviations from the reference values for the other variables. This approach is based on a similar idea to ours, however, we create the reference values dynamically, based on the current measurements only, without the need to assume that most trips on a road segment share similar profiles.

We propose a framework that makes use of expert knowledge, i.e., the physical relations governing the behavior of the system in question. Understanding the vehicle on the road lets us compare the effect of the variable of interest under different conditions. In this work we specifically target driver influence, but the approach can be generalized to other areas. An accurate and precise way of comparing fuel-related driver performance allows for better understanding of when changes in performance occur and how to transfer the best results to others, be it other human drivers or automatic systems. We aim to be able to categorize maneuvers with respect to fuel consumption, provide alternatives to improve bad ones, as well as understand specific skills of a driver, e.g., if they are doing poorly all the time or only in specific situations.

It is clear that the automotive industry is moving towards autonomous driving, where new factors like communication and advanced cruise control strategies also affect the vehicle operation. Such systems also benefit from a deeper understanding of what kind of actions lead to better performance. However, the driver is still the primary and final decision maker in the transportation process.

To summarize, driver performance in naturalistic driving

scenarios is particularly hard to estimate due to the fact that many of the relevant factors are not being measured. To be able to compare, train and incentivize drivers, we need to learn new features that capture their performance even in the absence of complete information. Such features need to incorporate the factors that we have no information about but also be robust enough to facilitate integration of measured variables.

II. DATA

We use two large datasets that have been collected in research and development projects within Volvo Group Trucks Technology (GTT). The first dataset comes from *European Field Operational Test* (EuroFOT) project [6], in which GTT was a partner with the role of testing *Fuel Efficiency Advisor* functionality. The other is an internal Volvo project called *Customer Fuel Follow-up* (CuFF). In both projects, data from multiple trucks have been collected, covering a wide area in Europe and also spanning over a long period of time, offering a variety of both geographic and ambient conditions.

Each truck has an automatic gearbox with 12 gears and a Cruise Control system. The data contains over one hundred signals that are logged from the vehicles' internal Controller Area Network (CAN), as well as additional sensors, at 10 Hz sampling frequency. In conjunction with on-board data that was recorded, we make use of off-line databases that provide map information.

The data recorded is extensive but not complete. Variables measured include ambient air temperature, vehicle speed, distance to vehicle in front, etc. Some variables can be obtained through mulitple sensors, such as vehicle speed which can be obtained using both the GPS location and the wheel-based odometer. Unmeasured variables include wind conditions, tire pressure, pavement characteristics, etc. Other signals, recorded for EuroFOT but not for CuFF database, include distance to vehicle ahead and video data from several cameras.

III. METHODOLOGY

The equation of motion for a vehicle on the road moving through air is:

$$m \cdot a = F_p - F_r - F_d - F_c \qquad (1)$$

where m is the weight of the vehicle, a its acceleration, F_p is the driving force, F_r is the rolling resistance force, F_d is the air drag force and F_c is the climbing force. F_r is dominated by the deformation of tires in contact with the road surface (Clark & Dodge [7]). F_d is dominated by the relative speed of vehicle in relation to the air around it, while F_c is dominated by the slope of the road coupled with the weight of the vehicle. Driver influences the vehicle by increasing or decreasing F_p force, however, they have no real control over the others.

Based on the sensor data available in our databases, it is not possible to calculate all the terms in equation 1. For example, there is no information regarding the wind, neither speed nor direction. Such factors of interest that are missing include both dynamic and static aspects of the vehicle as well as the environment. However, many of those aspects remain approximately constant, or experience minimal changes, across

periods of time, geographical location, or even the whole trip. Examples include tire pressure, gross vehicle weight, engine efficiency, wind parameters or road pavement.

We exploit this fact by means of *Fuel under Predefined Conditions* (FPC) concept, whose purpose is to provide a comparison term which captures some of the unmeasured factors. FPC is calculated on certain trip segments within the trip, based on a predefined set of characteristics. It can later be used on other segments, as long as they share those characteristics, and provides a means for fair comparison of driver performance. The "predefined conditions" need to be selected so that they have a high presence in the data. In our case, which covers mostly highway driving, one example is constant speed, within the 85 to 90 km/h range. An FPC value obtained in such a setting captures a number of vehicle and environment characteristics, while being independent from the driver. Comparing the fuel used on another road segment, for such as when approaching an intersection or overtaking a hill, allows us to highlight the influence of driver performance while diminishing the influence of other factors.

FPC is calculated according to the following equation:

$$FPC(s_0) = \frac{1}{N} \sum_{t \in s_0} fc(t), \qquad (2)$$

where N is the length of the trip segment, and $fc(t)$ is instantaneous fuel consumption measured at time t. Smaller values of N offer more opportunities to calculate FPC, however, due to the high amount of noise in the available sensors, the value is less accurate. On the other hand, there is a limit to how long a FPC segment can be, as trip characteristics can change significantly.

Then we can express driver performance $P(s)$, over any trip segment s, as the ratio between actual fuel consumption and the FPC:

$$P(s) = \frac{\mathrm{FC}(s)}{\mathrm{FPC}}, \qquad (3)$$

where $FC(s)$ is the fuel used on segment s.

The FPC value is calculated on some segment in the trip and then used afterwards on other segments. A particular FPC corresponds to a set of characteristics and when comparing drivers by $P(s)$, it is important to ensure that s shares this set of characteristics with the FPC used.

An illustrative example of the generality of the concept we will use multi-vehicle platooning. It is regarded as fuel saver, reducing air drag for all involved (Alam et al. [8]). However, platooning strategies differ and they are not always beneficial. Question as to when should a vehicle join a platoon directly affect how much fuel is saved or used. Platooning strategies can be defined by i) when a vehicle joins a platoon, ii) what distance should be kept from vehicle ahead and iii) when should a vehicle leave the platoon. We can then compare different platooning strategies using equation 3 where $P(s)$ now represents the performance of some platooning strategy.

The most clear benefit of using FPC is to provide a space where comparisons are meaningful and allow ranking of drivers, taking into account both long- and short-term performance. This can be done based on any set of factors
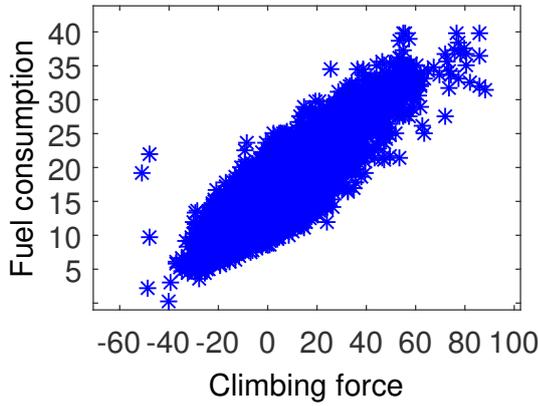
Fig. 1: Relation between road and weight of vehicle vs fuel consumption.

of interest, for example, geographical location. Given several trips that share a road segment, we can calculate FPCs for those trips, and then perform statistical inference on any other variable, such as traffic based on time of day and day of week, or road characteristics like road gradient and surface type.

IV. EXPERIMENTS

In this section we present the results of our experiments with estimating FPC using two different sets of predefined characteristics. We begin by identifying the conditions under which the effect of road gradient, wind, surface type, tire pressure, and so on can be estimated.

A. Base FPC

We approach the problem as a successive approximation, starting with a basic set of characteristics. Since we are looking to quantify driver effect, the first condition has to be that the vehicle speed is constant. Decisions of the driver that we are interested in directly correspond to changes in speed, having constant speed means that there is no driver influence. Therefore, we can later compare other segment s against such an FPC value, and make sure that the effects of driver decisions made there are captured by $P(s)$.

In our study, almost all occasions where constant speed occurs for more than one minute also correspond to having the highest gear. In general transmission setting is a factor that would be important to consider, but in practice we can simply ignore all other situations without losing significant amount of data. Therefore, the base set of characteristics for which we calculate FPC contains two aspects i) *constant speed* and ii) *maximum gear*. We denoted it as C_0. We write FPC_0 when necessary to specify for which set of conditions this FPC was calculated. It is important to note that FPC is calculated over a particular trip segment s_0, however, it is used to compare driver performance at segments $s_{1..n}$.

In long-haul truck operation, C_0 is a common scenario, in particular given the frequent use of cruise control. In addition, it captures the situation when the driver has minimal effect on fuel usage. Differences in fuel consumption are due to other factors, so selecting situations when cruise control is enabled

corresponds to having essentially removed the influence of the driver.

A number of characteristics captured by FPC_0 follow our assumptions about being constant or slow-changing: cargo weight, tire pressure, wind, etc. However, one which does not is the road gradient — and that can have very profound effect of the fuel usage. Therefore, our $P(s_i)$ measure will only be reliable if we can ensure that the road profile of segment s_i is similar to the road profile of segment s_0. In practice, given our setting, this only happens if both road segments are flat[1].

B. Adjustment for road inclination

One of the drawbacks of the approach presented in subsection IV-A is that flat roads, while the most common profile, are still quite rare in practice. Having the road flat and the speed constant for the required duration is overly limiting.

However, the effect of road gradient on fuel consumption can be estimated. As seen in figure 1, a simple linear road gradient model (RGM) is enough. RGM is specific to a vehicle type, however, in our data all vehicles have the same attributes. Having such a model allows for another set of characteristics under which we can calculate FPC.

For FPC_1 therefore we add a third characteristic, making C_1 be i) *constant speed*, ii) *maximum gear* and iii) *road gradient* of 0. It is important to note that we can calculate FPC_1 over segments with any road gradient profile. We take the effect of the climbing force F_c (see Equation 1), estimate the change in fuel consumption using RGM, and the resultant FPC_1 is considered to correspond to a flat road:

$$FPC_1(s_0) = FPC_0(s_0) - RGM(s_0) = FPC_0 - q \cdot \theta \cdot M \quad (4)$$

where q is an empirical constant, θ is the road gradient, and M is the estimated total weight of the vehicle. One of the drawbacks of this methods is that our data is collected from the tractor but not the trailer. This results in inaccurate weight estimations, which translates into higher variation in the FPC_1 values.

It can be seen that FPC_1 is equivalent to FPC_0 when calculated on a road segment s which is "naturally" flat. However, when s is uphill, FPC_1 will be generally smaller than FPC_0, and when s is downhill, it will be larger. Most importantly, though, the opportunities to calculate FPC on a truly flat road are scarce. And as mentioned in the previous section, while we can calculate FPC_0 on non-flat roads, they are not useful for driver comparisons. On the other hand, using the road model and FPC_1, we can calculate $P(s)$ based on:

$$P(s) = \frac{\text{FC}(s) - RGM(s)}{\text{FPC}_1}. \quad (5)$$

---

[1]However, this is not necessarily true in other settings. For example, in the case of city buses that drive in the same area, other road profiles can also be repeated often.
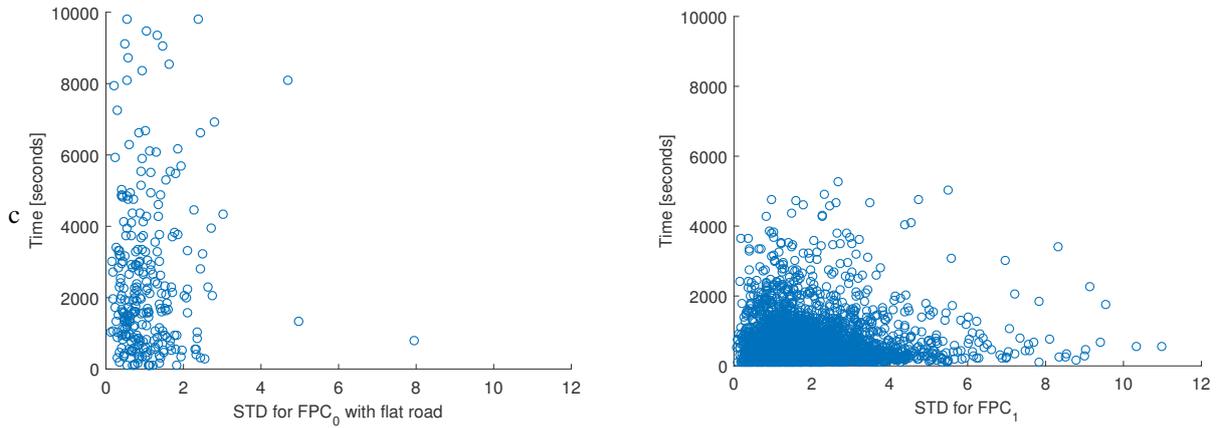
Fig. 2: Standard deviation of FPCs within a trip vs average time difference between consecutive FPCs.

C. FPC Comparison

In section III we have introduced the FPC feature, and in the subsections above we have defined two set of conditions, namely C_0 and C_1. Here we present the comparison between the two, and demonstrate the necessity of introducing FPC_1 despite the extra complications it involves. The main criteria for this comparison will be the standard deviation of FPCs within the same trip, and the time between consecutive FPCs within a trip. Those two measures capture how reliable a support system for driver or fleet operator can be. This data is presented in Figure 2.

We start by looking at *reliability*, i.e., how likely it is that we can even calculate the FPC in a given trip. In those experiments we require a segment of 150 seconds to calculate FPC. For FPC_0 with flat road we have 30% of trips with one FPC segment, and less than 5% of trips with five FPC segments. On the other hand, with FPC_1, we have over 60% of trips with at least three FPC segments and almost 40% of trips ten or more FPC segments.

Another aspect is the time it takes before conditions are met and the system is able to calculate the first FPC. This is represented in figure 3(left) for FPC_0 and figure 3(right) for FPC_1, respectively. Those figures show that one of the main concerns for FPC_0 is the long time from the start of

the trip until we meet the conditions required. It is on average 72 minutes, compared to 33 minutes for FPC_1. This high time required to have an FPC_0 lowers the attractiveness of the method, especially if we consider online driver support applications, as a large part of the support can only be available after the first FPC is calculated.

Next we look at *stability*, expressed as the standard deviation of all FPCs within a single trip. This is used to validate our assumptions regarding how fast the unmeasured aspects of the environment change, as well as our approach to computing FPC. Figure 2(left) shows that FPC_0 with flat road has low and constant standard deviation regardless of how far apart consecutive FPC segments are. This verifies our main expectation regarding slow changes in most of the unmeasured factors, like wind speed or pavement type. On the other hand, Figure 2(right) shows that we have — to some degree — traded quality for quantity. In particular, the average time between consecutive FPC_1 is consistently a lot lower than that for FPC_0. At the same time, FPC_1 has a much higher variability between different FPCs calculated from the same trip. This value is quite a bit higher than expected, which means that the inaccuracies introduced by the road model and imprecise estimation of vehicle weight lead to lower quality.

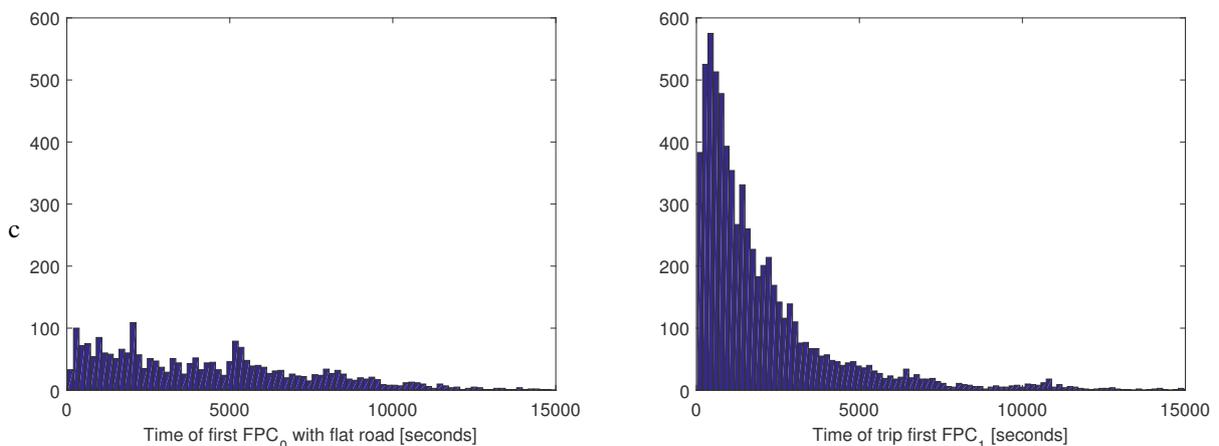Figure 4(left) displays the standard deviation for FPC_0

Fig. 3: How far along [seconds] we are in a given trip before we meet the conditions to calculate FPC for the first time.
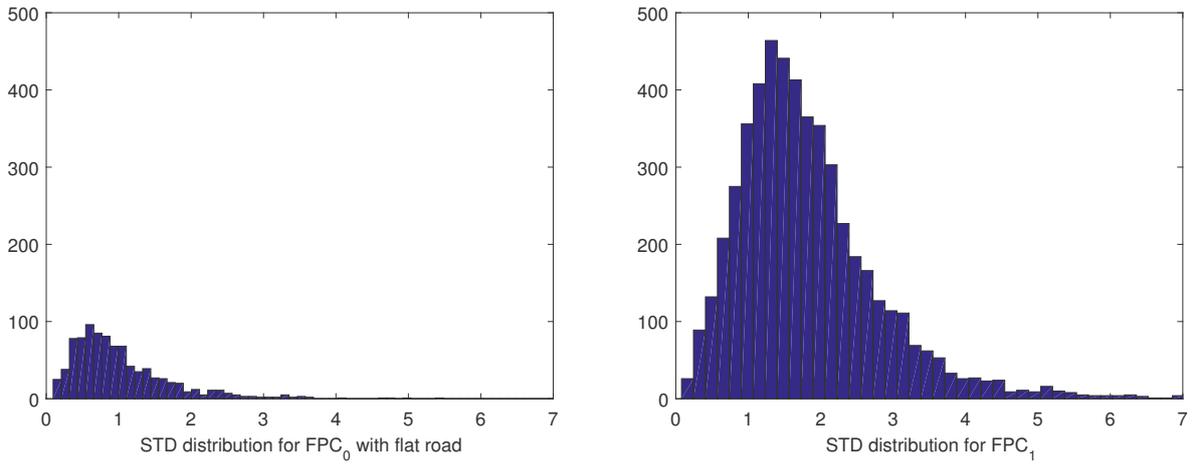
Fig. 4: Standard deviation for FPCs within a single trip, for trips with at least three FPCs found.

with flat road within a trip. We attribute this variability to changes in unmeasured factors, inaccuracies in sensors and noise, and for the FPC to be useful we require the variability to be small. From our experiments we see that 60% of the FPC are within 1 standard deviation, and 91% are within 2 standard deviations, respectively, which is close to a normal distribution and what we would expect as the most dynamic factor, wind, can be modeled with a standard distribution, according to Aksoy et al. [9]. We perform the same analysis for FPC_1. In figure 4(right) we can see that the mean of standard deviation is higher than that of FPC_0, at 1.84 compared to 1.03 [L/100km]. At the same time, the number of FPC_1s available is 13.5 times higher than the number of FPC_0s.

The natural question is then where does the increase in variability comes from: the changing environment conditions, the inaccuracies in the calculations or both. Figure 1 can answer partly, as we employed a linear model to fit that data, so we expect this to introduce some amount of noise. However the tradeoff still improves the usability of the method by increasing the number of missions for which we can calculate FPC, and especially multiple FPCs, as well as provides more up-to-date information.

Even though the time gaps in calculating FPC_0 vary significantly, we can see in figure 2(left) that the variation in FPC_0 is not dependent on how long time has passed since the previous FPC was calculated. The reliability aspect of FPC_1 is illustrated in figure 2(right) and it shows the same thing as for FPC_0, i.e. the quality of FPC, measured through standard deviation, is not dependent on the time between consecutive values.

Finally, we look *timeliness*, i.e., for how long do we need to use an FPC before can replace it with a more up-to-date one. We can see a distribution of this in figure 5a, with mean of 40 minutes for FPC_0 and 15 minutes for FPC_1, respectively. We believe that a reasonable maximums for the FPC to be used for is 30-40 minutes, and after that time we would like it to be replaced with a newer one even though as we see in figure 2 the variability of FPC is not time dependent.

However, the mean of timeliness is not the most interesting metric. It can be heavily dominated by areas where many FPCs can be calculated easily and often. On the other hand, there is very little difference between FPC being replaced by a new one after 1 or 5 minutes — it's the extremes, the tail of the distribution, that is interesting.
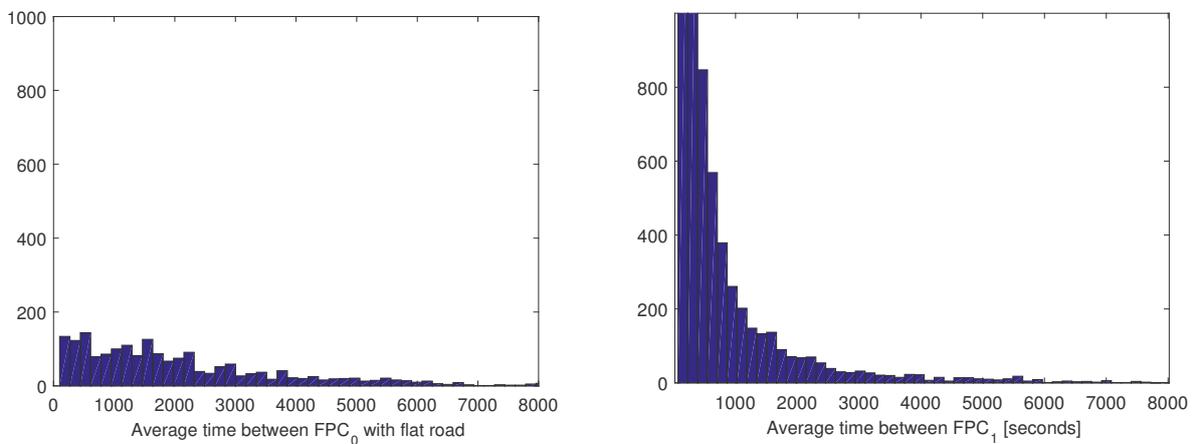

Fig. 5: Average time difference between consecutive FPCs occurring in the same trip. For the last occurrence of FPC in a trip we consider the time difference until the end of the trip.
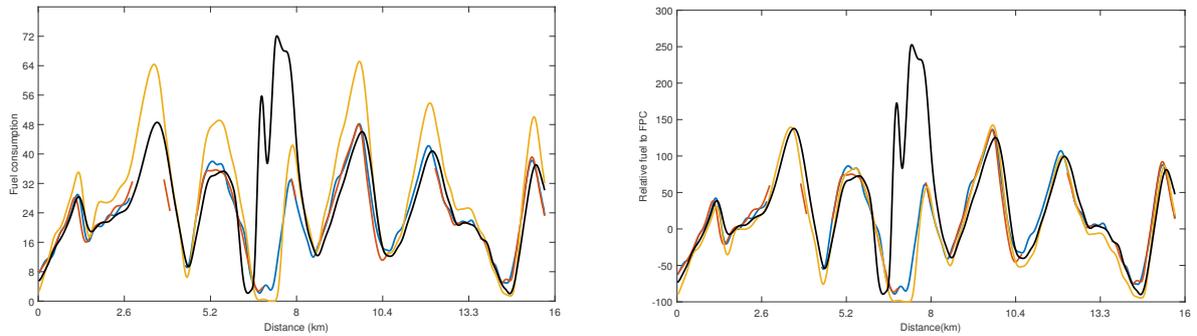
Fig. 6: Measured fuel consumption (left), Normalized fuel by FPC(right).

D. FPC Evaluation

Finally, we look at what we perceive similar and different driver profiles. The most evident situation where driver behavior is similar is when their speeds match on a particular road segment, and the behavior is different when when one driver maintains constant speed while another does not. Initially, we consider the fuel consumption as the indicator for driver performance. To increase the likelihood that the conditions in which the trips occur are the same, we select trips that occur on same road segment and we also select vehicles with similar characteristics, i.e. engine type, vehicle load, etc.

Figure 6(left) shows the fuel consumption profile for 4 trips with the specified characteristics. Three trips have similar both fuel consumption and speed (not shown): they share the same shape for both curves. The fourth, on the other hand, has speed profile that changes in the middle of the depicted segment. The effect of this change can be readily seen in the fuel consumption plot.

We consider the lowest fuel consumption (red line) as the reference and then we have the following performances compared to it: blue 101%, red (reference) 100%, yellow 126% and black 116% respectively. The first three have the similar speed profile. The conclusion based on those numbers would be that "yellow" driver has the worst performance.

However, this conclusions changes when comparing the $P(s)$ values, i.e., fuel consumption normalized by FPC as depicted in figure 6(right). Here we can see that "yellow" driver has performance very similar to "red" and "blue" drivers. The performances given by the average normalized fuel consumption are: blue 113%, red 113%, yellow 109% and black 131%.

To summarize, we have the following order given by measured fuel consumption: red, blue, purple, yellow. The driver performance measure based on the approach proposed in this paper gives us: yellow, red, blue, purple. Given the speed profiles, the second ranking is much more believable.

We can also conclude that impact of other factors is significant, in this case around 26%, which also indicates that better methods for estimating performance in terms of fuel consumptions are needed.

V. CONCLUSIONS AND FUTURE WORK

The paper presents an approach towards quantifying the effects of specific conditions in the presence of factors for which measurements are incomplete or missing, but when general relations between those factors are known. We apply it for a specific case of comparing truck driver performance. In this setting we propose a normalization term, called Fuel under Predefined Conditions.

Using FPC we facilitate the comparison among trips having different characteristics. By arranging the factors responsible for fuel consumption into different categories, we show how to perform better comparison among trips, as well as provide better estimation of driver performance.

We have identified some issues with our current approach, for example, how the inaccuracies that propagate from imprecise sensors affect the values of FPC_1. Improving this aspect of the method is an important task for the future.

Another essential question to consider is when does the driver have actual freedom to make decisions and when external conditions are the limit. Driver decisions are influenced by many factors, such as traffic or delivery deadlines. Fuel consumption is only one of the criteria — so a comprehensive evaluation method should take bigger picture into account.

We aim at extrapolating driver influence as well as the method for assessing freedom of driver when making choices that lead to changes in fuel consumptions. For this, however, additional information is required, e.g. mission parameters such as delivery deadline.

REFERENCES

[1] G. Barnes and P. Langworthy, "The per-mile costs of operating automobiles and trucks," 2003.

[2] N. Nylund, "Fuel savings for heavy-duty vehicles hdenergy. summary report 2003-2005," Project Report VTT, Tech. Rep., 2006.

[3] G. N. Bifulco, F. Galante, L. Pariota, and M. R. Spena, "A linear model for the estimation of fuel consumption and the impact evaluation of advanced driving assistance systems," *Sustainability*, vol. 7, no. 10, pp. 14 326–14 343, 2015.

[4] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *International Journal of Computers Communications & Control*, vol. 5, no. 5, pp. 654–663, 2010.

[5] M. D. Murphy, "Fuel consumption estimation," Jun. 22 1999, US Patent 5,913,917.

[6] WWW, "http://www.eurofot-ip.eu/," accessed: 2016-05-01.

[7] S. K. Clark and R. N. Dodge, "A handbook for the rolling resistance of pneumatic tires," 1979.

[8] A. A. Alam, A. Gattami, and K. H. Johansson, "An experimental study on the fuel reduction potential of heavy duty vehicle platooning," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on.* IEEE, 2010, pp. 306–311.

[9] H. Aksoy, Z. F. Toprak, A. Aytek, and N. E. Ünal, "Stochastic generation of hourly mean wind speed data," *Renewable energy*, vol. 29, no. 14, pp. 2111–2131, 2004.