A stochastic data-based traffic model applied to vehicles energy consumption estimation

Arthur Le Rhun, Frédéric Bonnans, Giovanni De Nunzio, Thomas Leroy, Pierre Martinon

***Abstract*—A new approach to estimate traffic energy consumption via traffic data aggregation in (speed,acceleration) probability distributions is proposed. The aggregation is done on each segment composing the road network. In order to reduce data occupancy, clustering techniques are used to obtain meaningful classes of traffic conditions. Different times of the day with similar speed patterns and traffic behavior are thus grouped together in a single cluster. Different energy consumption models based on the aggregated data are proposed to estimate the energy consumption of the vehicles in the road network. For validation purposes, a microscopic traffic simulator is used to generate the data and compare the estimated energy consumption to the reference one. A thorough sensitivity analysis with respect to the parameters of the proposed method (i.e. number of clusters, size of the distributions support, etc.) is also conducted in simulation. Finally, a real-life scenario using floating car data is analyzed to evaluate the applicability and the robustness of the proposed method.**

***Index Terms*—Traffic modeling, Clustering, Energy Consumption**

I. Introduction

In 2015, according to data from the European Environment Agency, road transportation contributed to 21% of total EU-28 greenhouse gas emissions. In order to meet the long-term emissions reduction target, emissions from transportation need to fall by more than two thirds by 2050 [1]. These emissions are essentially a function of the vehicle propulsion technology and the driving style [2].

Estimating energy consumption of the vehicles is a great challenge in the objective of improving global transportation efficiency, since this information is used in energy management, eco-routing, eco-driving, traffic management, ... Traffic congestion has a major impact on the driving behavior, and thus plays a key role in the level of fuel consumption [3].

Therefore, accurate predictions of vehicles energy consumption must take traffic conditions into account. To perform this objective, faithful modeling of traffic behavior is of primary importance. Energy-oriented modeling approaches can be divided in two main categories.

On the one hand, several mathematical traffic models are available nowadays, see for instance [4]. Such models typically depict the reality either from a macroscopic point of view, based on the road vehicular density [5], or from a microscopic perspective, based on the description of the instantaneous behavior of each vehicle [6]. Both approaches have limitations in providing an accurate energy consumption estimation. Macroscopic models typically provide average traffic speeds to compute energy consumption [7], thus neglecting the impact of speed fluctuations due to congestion. Higher precision of the energy consumption estimation could only be obtained at the expense of a denser discretization of the road network, therefore compromising scalability. Microscopic models could achieve precise energy consumption estimation, but they require a significant calibration and validation effort. Also, the computational burden and the amount of collected data grows rapidly with the size of the network, therefore these models are more suitable for off-line use.

On the other hand, data-based models rely on collected traffic information to estimate traffic behavior and energy consumption. Instantaneous models are able to precisely estimate energy consumption by using large amounts of data, generally the measured driving profile of each vehicle. To tackle this drawback, aggregated models use the average value of the measured speed profiles to compute energy consumption, but they suffer from the same accuracy problems previously discussed for the macroscopic traffic models. Furthermore, the data sparsity and availability is an issue [8]. Other approaches try to solve the problem of the data sparsity by simply classifying road segments by category (e.g. urban, arterial, freeway, etc.), in order to associate each category with a typical energy use. This type of models may lead to inaccuracy in energy consumption estimation, as road segments belonging to the same category may show very different traffic patterns [9].

In this work, a new way to represent traffic behavior on large road networks is proposed. The objective of this model is to accurately depict the effect of traffic conditions on the vehicles energy consumption in each road segment. The key idea is to use a statistical approach based on vehicle speed and acceleration data, measured from real vehicles. In particular, the entire observation time during which speed and acceleration data are collected is subdivided into time-frames. During each time-frame a (speed,acceleration) distribution is generated, and such a distribution is then used as an input for an energy consumption model to estimate the traffic energy consumption on the analyzed road segment during the specific time-frame. Therefore, each road segment is defined by its own collection of (speed,acceleration) distributions.

A. Le Rhun, IFP Energies nouvelles, 1 avenue de Bois-Préau, 92852 Rueil-Malmaison, France, arthur.le-rhun@inria.fr

G. De Nunzio, IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France, giovanni.de-nunzio@ifpen.fr

T. Leroy, IFP Energies nouvelles, 1 avenue de Bois-Préau, 92852 Rueil-Malmaison, France, thomas.leroy@ifpen.fr

F. Bonnans, Inria Saclay and CMAP Ecole Polytechnique, route de Saclay, 91128 Palaiseau, France, frederic.bonnans@inria.fr

P. Martinon, Inria Saclay and CMAP Ecole Polytechnique, route de Saclay, 91128 Palaiseau, France, pierre.martinon@inria.fr

The proposed model also includes the possibility of reducing the dimension of the traffic data and increasing scalability, by applying clustering techniques to the probability distributions of each road segment. For instance, the different distributions representing traffic in one road segment over different hours of the day may be aggregated in clusters modeling only significant traffic conditions (e.g. peak, off-peak, etc.).

The paper is organized as follows. The clustering technique and the proposed energy consumption model are presented in Section II. The traffic data collection and the model validation procedure are discussed in Section III. Using traffic data from the simulator SUMO, Section IV illustrates the method on one road segment, while more complete numerical results on 500 segments are shown in Section V. Finally, section VI presents an application of the method to actual traffic data, as well as a comparison to a standard approach using only average speeds.

II. Proposed Method

A. Road segments

In the following, the road network is assumed to be subdivided into a collection of *segments* which correspond to relatively small portions of road. The segments, usually denoted s, form a set S of size N_S. These portions may typically reflect the topography of the network, so a segment will for instance include a traffic light, a crossroads, a roundabout, or a straight section of road. Segments length can range between a few meters and a few hundreds meters, as illustrated on Fig.1.

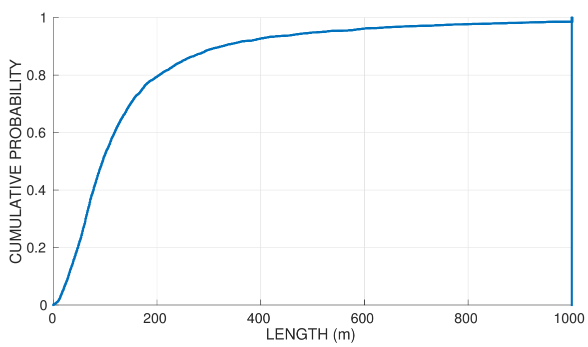


Fig. 1: An illustration of cumulative distribution of the length of road segments in city of Luxembourg as implemented in LuST.

B. (speed,acceleration) probability distribution

The consumption of an engine depends on its operating point, which can be determined by the speed and acceleration of the vehicle. The key point in our method is to assume that an accurate (with respect to energy consumption) description of the traffic can be derived from the probability distributions of measured speeds v and accelerations a for each road segment. Note that these distributions do not retain the temporality of the speed profiles.

We set N_T the number of time-frames for each segment[1]. We denote the family of time-frames $(t_{i,s})_{(1...N_t)\times(1...N_S)}$, and their length $(\Delta T_{i,s})_{(1...N_t)\times(1...N_S)}$. For each segment s, we record the (speed,acceleration) of all the vehicles passing through the segment during all the time-frames $t_{i,s}$.

[1]For the sake of simplicity, we take the same number of time-frames for all the segments

In the following, we work with discrete distributions in the (v, a) space. We denote $\mathbb{V}$ and $\mathbb{A}$ the sets of feasible speed and acceleration. To simplify, these sets are taken identical for all segments and time-frames, thus all (speed,acceleration) discrete probability distributions have the same support in $\mathbb{V} \times \mathbb{A}$. We denote $N_{\mathbb{V}}$ and $N_{\mathbb{A}}$ the discretization size of $\mathbb{V}$ and $\mathbb{A}$. Recalling that N_S and N_T are the number of road segments and time-frames, we obtain a total of $N_S N_T$ discrete distributions of support size $N_{\mathbb{V}} N_{\mathbb{A}}$. Fig.2 shows an example of such a distribution.

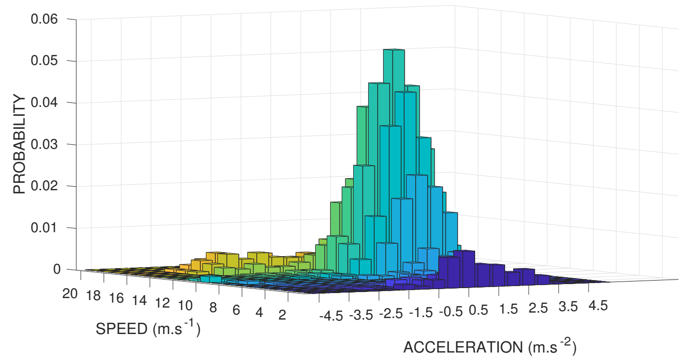


Fig. 2: Example of (speed,acceleration) distribution.

C. K-means clustering with strong patterns

Since the objective is to handle large road networks with several thousands of segments, we try to reduce the information from the probability distributions $\mu_{i,s}$ corresponding to a segment s at the different time-frames $t_{i,s}$. We perform this reduction through clustering techniques, and take the barycenters $\bar{\mu_j}$ of each cluster as representatives of the traffic conditions on all the segment-timeframe pairs $(s, t_{i,s})$ in the cluster.

We use the K-means algorithm [10] to compute the clusters. In the following, k indicates the (fixed) number of clusters. Since the elements to be clustered are probability distributions, we choose to work with the 1-Wasserstein distance [11], [12] whose definition we recall in (1):

$$W_1(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} d \cdot \pi \tag{1}$$

with $\Pi(\mu, \nu)$ being the set of transportation plans defined as the set of nonnegatives matrices of size $(|N_{\mathbb{V}}| \times |N_{\mathbb{A}}|)^2$ with marginals $\pi^\top \mathbf{1} = \mu$ and $\pi \mathbf{1} = \nu$. Since speed and acceleration have comparable magnitude order, we take d_{ij} equal to the Euclidean norm between the corresponding points.. The notion of barycenter has been extended to this distance in [13]. Using this distance based on the optimal transport theory tends to preserve the geometrical aspects of the probability distribution (shapes of the distributions) and then the way that the speeds are spread.

To compensate the sensitivity of the K-means algorithm to the initial guess, we use the strong patterns method [14]. Strong patterns are sub-clusters whose elements always end up in the same cluster regardless of the starting point of the K-means algorithm. In practice, we run a first batch of K-means with random initializations to identify strong patterns. Then a final K-means run is initialized with one element of each k largest strong patterns. We recall the general principle of the K-means in Algorithm 1.

Algorithm 1: K-means algorithm

Input : Distributions to be clustered $\mu_1 \cdots \mu_i$
Output: Clusters barycenters $\bar{\mu_1} \cdots \bar{\mu_k}$
Initialization
for $j \in \{1, 2, ..., k\}$ **do**
  $\bar{\mu_j^0} \leftarrow$ choose randomly in $\mu_1 \cdots \mu_i$
while $\exists j, \bar{\mu_j^k} <> \bar{\mu_j^{k+1}}$ **do**
  for each distribution μ_i **do**
    Find the closest barycenter $\bar{\mu_c^k}$
    Index μ_i by c
  Compute the new barycenter $\bar{\mu_c^{k+1}}$ for each index c

D. Computing energy consumption

The energy consumption computed is the energy at the wheel, neglecting the losses due to the powertrain. The instantaneous power at the wheel is denoted as a general function $P(v, a)$, which can be for instance of the form presented in Eq.5. The ultimate objective of the proposed method is to estimate energy consumption by using only the information extracted from the (v, a) probability distributions μ. More accurately, we seek to obtain the energy consumption from the barycenter $\bar{\mu}$ of the cluster containing μ. In the following we introduce two methods to compute the consumption of a generic vehicle passing through a segment. These will be referred to as "Average Consumption Method" and "Memoryless Sampling Method".

1) Average Consumption method: The first idea is to use the average power $\bar{P}$ (in the probabilistic sense)

$$\bar{P}(\bar{\mu_t}) = \sum_{\mathbb{V}\times\mathbb{A}} \bar{\mu}_t(v, a) P(v, a) \tag{2}$$

where $\bar{\mu_t}$ is the barycenter of the cluster containing the current segment at time t. The barycenter may indeed change if t crosses different time-frames while the vehicle is on the segment. This average power is integrated over the time interval $[t_i, t_f]$ spent by the vehicle on the segment, thus

$$C_{Avg} = \int_{t_i}^{t_f} \bar{P}(\bar{\mu_t}) dt \tag{3}$$

Note that knowledge of the time interval is required in this method, in addition to the (speed,acceleration) distributions. Indeed, using here the average time would give identical consumption for every vehicle. Therefore, we need some more statistically significant time information in order to capture the deviation of the consumption distribution. A resulting drawback of this method is that a faster vehicle has a shorter travel time and thus a lower energy consumption, which may seem unrealistic.

2) Memoryless Sampling method: In the second method the energy consumption is still obtained by integrating the instantaneous power, but we do not use the average power. Instead, we implement the idea that the vehicle must follow the traffic at every time, in a statistical sense. More precisely, its speed and acceleration should follow the probability distribution of the barycenter $\bar{\mu}$ of the cluster for the current pair (s, t). Another difference is that the integration is performed over the segment length L_s instead of travel time.

So the Memoryless Sampling method generates a sequence of (v_n, a_n), independent samples according to the probability distribution $\bar{\mu}$. Setting a time step δt, we use this sequence to integrate both the travelled distance and the instantaneous power. We assume $\delta t = 1s$, in order to have the same order of magnitude as the reaction time of a driver. We stop the generation of (v_n, a_n) when the vehicle reaches the end of the segment[2]. Since the distance will be covered in a finite time, we obtain the finite set of samples $(v_n, a_n)_{n=1,...,n_f}$ and the consumption writes as

$$C_{MSM} = \sum_{n=1}^{n_f} P(v_n, a_n) \delta t \tag{4}$$

III. Validation Approach

A. Traffic Data from simulation

We illustrate our approach with data obtained from the traffic simulator SUMO [15]. The simulation runs for a specific scenario named LuST [16], which models the traffic in the city of Luxembourg during 24 hours, with a time step of one second (1 Hz sampling frequency). This scenario already includes the subdivision of the network in road segments. In LuST scenario, a segment is defined as a continuous lane with no intersection. In the following we group lanes going in the same direction to obtain more data per segment. From the raw data, we extract the segment, speed, and acceleration of each vehicle in the network at all time steps. A Python script imports these values and saves them in two matrices for each time step, indexed by road segment and time-frame.

We also aggregate these records for a fixed time-frame length $\Delta t_{s,i}$ (in practice we use a constant frame length Δt for all segments). This is done both to gather sufficient data on the segments, and to decrease the number of elementary traffic distributions for the clustering phase.

B. Power and Reference energy consumption

We would like to compare the consumption given by the two methods (Average and Memoryless Sampling) with the "reference consumption" from the simulator. Since energy consumption is not an output of SUMO, it needs to be

[2]This event will happen with probability 1, since vehicles never stop indefinitely.

computed. Neglecting the slope effect, the following equation is used to compute the instantaneous power P needed at each time t:

$$P(v(t), a(t)) = (ma(t) + a_2 v^2(t) + a_1 v(t) + a_0)v(t) \quad (5)$$

with m the vehicle mass and a_0, a_1, a_2 defining a polynomial approximation of the road-load force, which is vehicle-dependent. The numerical values used for the simulations are: $m = 1190kg$, $a_0 = 113.5$, $a_1 = 0.774$, $a_1 = 0.4212$, corresponding to a passenger vehicle [17].

The Reference energy consumption is obtained by integrating[3] the instantaneous power while plugging into (5) the recorded speed and acceleration $(\mathbf{v}(t), \mathbf{a}(t))$ of the vehicle:

$$C_{ref} = \int_{t_i}^{t_f} P(\mathbf{v}(t), \mathbf{a}(t))\mathrm{dt} \quad (6)$$

Instead of comparing the consumption of single vehicles, we compute the consumption for all vehicles passing through the segment, as recorded in the traffic simulation. Therefore, we end up with comparing the consumption distributions for the Average and Memoryless Sampling methods to the Reference consumption. In the following, these three consumptions distributions are denoted respectively $\mathbf{C}_{Avg}$, $\mathbf{C}_{MSM}$ and $\mathbf{C}_{Ref}$.

C. Indicators

Since our aim is to compare distributions of energy consumptions, we study several indicators.

- Mean and Standard deviation errors are classical indicators. We compute these relative errors as follows, with $\mathbf{C}_{method}$ denoting either the Average consumption or the Memoryless Sampling consumption:

$$\varepsilon_{mean}(s) = \frac{\overline{\mathbf{C}_{method}(s)} - \overline{\mathbf{C}_{ref}(s)}}{\overline{\mathbf{C}_{ref}(s)}}$$

$$\varepsilon_{\sigma}(s) = \frac{\sigma_{\mathbf{C}_{method}}(s) - \sigma_{\mathbf{C}_{ref}}(s)}{\sigma_{\mathbf{C}_{ref}}(s)}$$

- Kullback-Leibler divergence [18], also called 'relative entropy', is a particular case of φ-divergence. KL divergence can be used to measure distances between two probability distributions P and Q, however it is not a metric (no triangular inequality or symmetry). Another drawback is that it cannot be computed for instance when the probability of the model q is 0 while the probability of the reference p is not.

$$KL(P|Q) = \sum_i p_i \log(\frac{p_i}{q_i}) \quad (7)$$

Instead, we use the Jensen Shannon divergence, which is a symmetrized version of KL divergence, sometimes referred to as 'total divergence to the average' [19]. Note that the square root of the JS divergence is a metric called JS distance [20], [21]. Fig.3 illustrates the JS divergence on a Gaussian with noised parameters.

$$JSD(P|Q) = \frac{1}{2}KL(P|M) + \frac{1}{2}KL(Q|M) \quad (8)$$

with

$$M = \frac{1}{2}P + \frac{1}{2}Q \quad (9)$$

Here, we have $P = \mathbf{C}_{method}$ and $Q = \mathbf{C}_{Ref}$.

[3]In practice integration is done by Euler scheme.

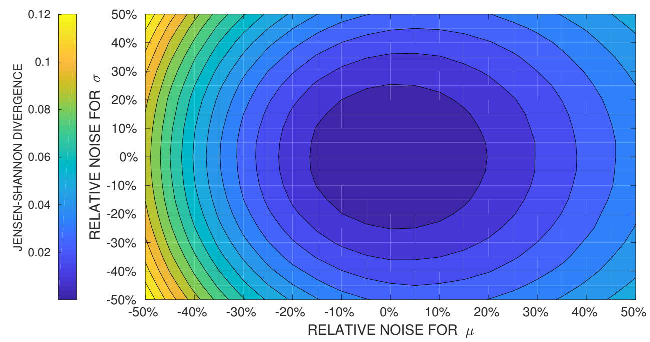


Fig. 3: Jensen Shannon divergence for a noised Gaussian distribution with base parameters $\mu = \sigma = 1$.

IV. Numerical results on one segment

To begin with, we compare the different consumptions on a single road segment. The objects compared are therefore the consumptions distributions of all the vehicles that went through the segment during each time-frame. Results are shown as the cumulative distribution function of the consumptions for all time-frames.

We analyze in particular the influence of the (speed,acceleration) discretization, the clustering, and the choice of time-frame duration. Unless specified otherwise, the distributions are shown for a discretization $N_V = N_A = 10$, a time-frame $\Delta t = 10min$, taking the full set of distributions without clustering.

A. Influence of the (speed,acceleration) discretization

First, Fig.4 shows the consumption distribution for both the Average method and Memoryless Sampling method. We test the discretizations $N_V = N_A = 10, 20$ and 30 and compare to the reference consumption. We see that for the Memoryless Sampling method: i) the general shape of the distribution is similar to the reference and ii) finer (v, a) discretizations give distributions closer to the reference. On the other hand, for the Average method: i) we observe some linearization of the consumption and ii) the effect of discretization is much less significant.

B. Influence of the time-frame

Next we study the effect of the length of the time-frame Δt, which is the time interval over which we aggregate the vehicles data. Longer time-frames may cause some over-averaging and loss of specific traffic information. On the other hand, shorter time-frames may lead to insufficient vehicle data (for statistical relevance), and also increase the number of (v, a) distributions to handle. Fig. 5 shows the energy consumptions obtained for time-frames of $5s$, $1min$ and $10min$. On this segment, for both the Average and the Memoryless Sampling methods, the influence of the time-frame duration seems rather small.

C. Influence of the clustering

To conclude this first batch of results, we examine the information loss due to the clustering stage. The consumptions

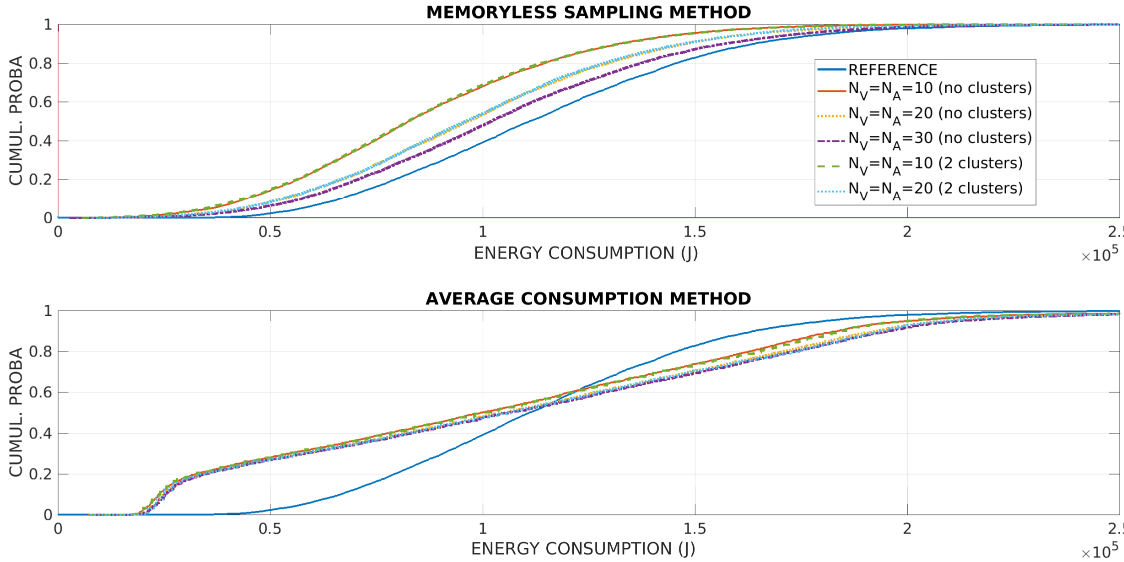


Fig. 4: Cumulated consumption distribution for one segment - effect of (v, a) discretization and clustering.

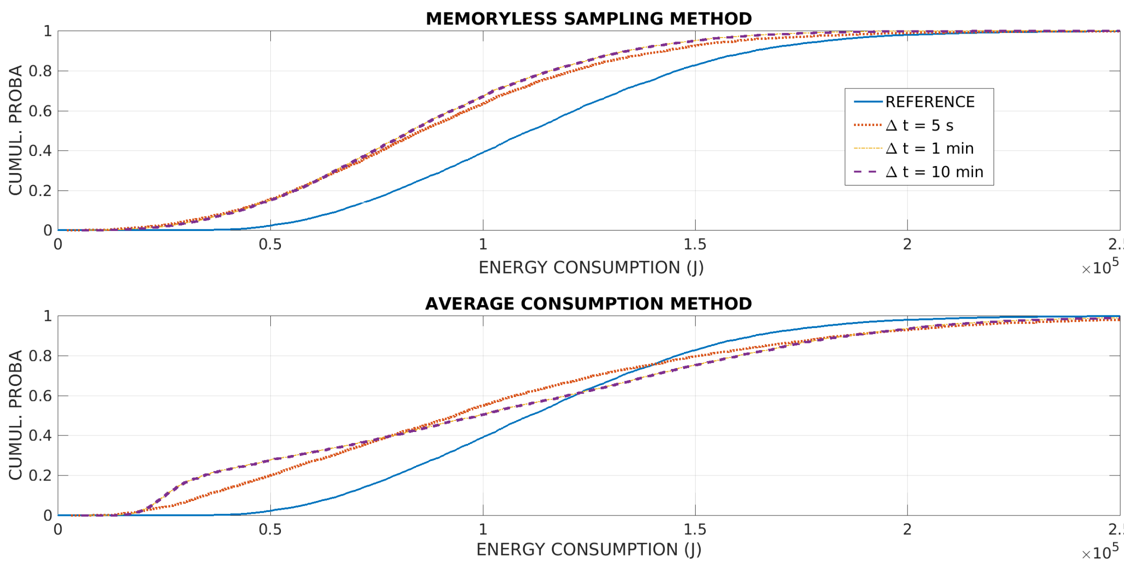


Fig. 5: Cumulated consumption distribution for one segment - effect of the time-frame duration.

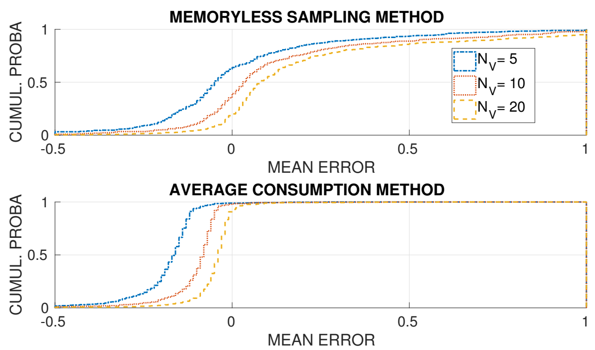
obtained using only the barycenters (green and cyan curves in Fig.4) are almost identical to the ones obtained without the clustering stage. This indicates that the data reduction performed by the clustering comes with a negligible loss of information.

V. Numerical results on 500 segments

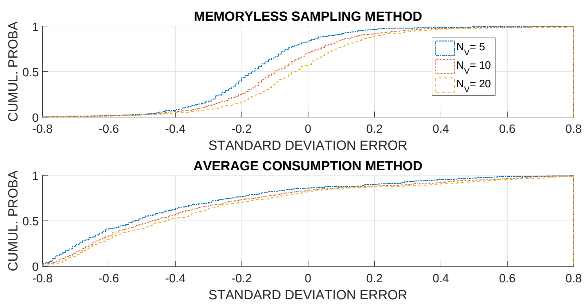
Now, we consider the entire road network of the city of Luxembourg, as implemented in LuST, with $\simeq$ 54000 road segments. We pick a random set of 1% (547) segments, for which we perform the clustering, and compute the energy consumption with the Average and Memoryless methods. For each segment, we use the indicators defined in III-C to compare the computed energy consumptions to the Reference consumption. We discuss the relevance of the models based on the distributions of these indicators on the set of 547 segments. More precisely, we investigate the influence of the discretization of speed and acceleration, and the number of clusters. Unless specified otherwise, the simulations use a speed and acceleration discretizations of 20 steps, a time-frame of 10 minutes, with 2 clusters.

A. Influence of speed discretization

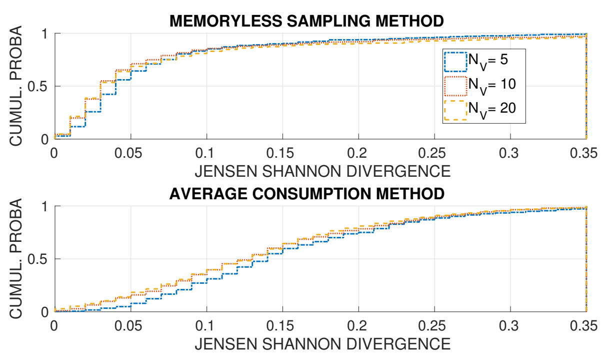
The speed discretization has a direct influence over the barycenters computed by the K-means, since it changes the support of the distribution obtained. We expect a finer discretization to give computed consumptions (Average and Memoryless Sampling methods) closer to the Reference ones. The obvious drawbacks are an increased cost of the barycenter computation and size of the distributions. We test $N_{\mathbb{V}} = 5, 10, 20$ steps for the discretization, the speed interval being $[0, 20]$ in m/s.



(a) Cumulated distribution of relative Mean error.



(b) Cumulated distribution of relative Standard deviation error.



(c) Cumulated distribution of Jensen Shannon divergence.

Fig. 6: Speed discretization.

1) Relative mean and standard deviation errors: We begin with the distribution over the 547 test segments of the mean and standard error (both relative). Fig. 6a and 6b (upper graphs) show the errors between the Reference energy consumption and the Memoryless Sampling method. The mean and standard error both appear to be reasonably well centred around 0. We also observe that finer discretizations of the speed clearly improve the standard error, possibly due to a better reconstruction of the travel times. On the other hand the mean error is shifted towards positive values, and is indeed smaller for 10 steps than 20.

Fig. 6a and 6b (lower graphs) compare the Average method consumptions to the Reference. Here the mean error is almost always negative, and the standard error is also negative for 80% of the segments. This strong unbalance towards the negative indicates that the Average method tends to underestimate the consumption. Increasing the speed discretization reduces the mean error, but does not really improve the standard error. A possible explanation is the fact that the travel time for each vehicle is taken from the simulation, thus the speed discretization has no effect on it.

2) Jensen Shannon Divergence: Fig. 6c shows the distribution over the 547 segments of the Jensen Shannon divergence. Upper graph is for the Memoryless Sampling versus Reference, and lower graph is for Average method versus Reference. We observe that the JS divergence is much smaller overall for the Memoryless Sampling than the Average method, with distributions more concentrated towards zero. For both methods, increasing the speed discretizations from 5 to 10 steps improves the JS divergence, while 20 steps yield very little additional gain.

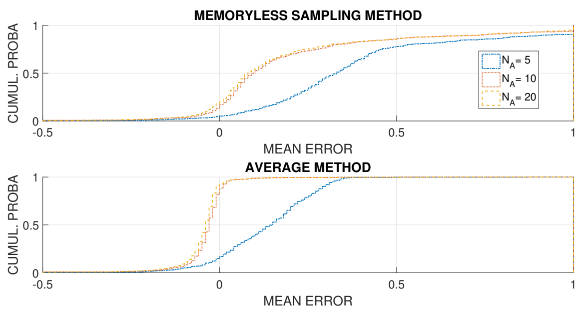
B. Influence of acceleration discretization

Now we study the discretization of the acceleration. As for the speed, this parameter influences the support of the (v, a) distributions, and therefore the K-means clustering. We want to know if finer discretizations of a give more accurate energy consumptions for the Average and Memoryless Sampling methods. We test $N_{\mathbb{A}} = 5, 10, 20$ steps for the discretization, the acceleration interval being $[-4.5, 4.5]$ in ms^{-2}.

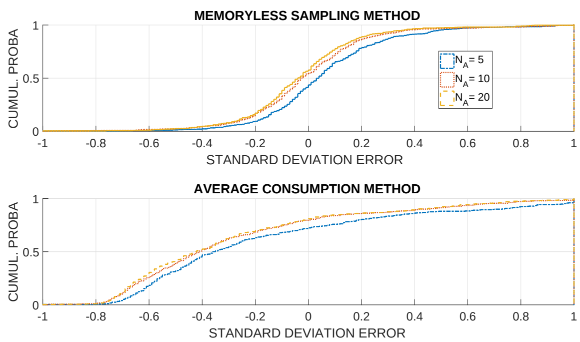
1) Relative mean and standard deviation errors: Fig.7a shows the mean error distribution for the two methods, with different discretizations of a. We observe that for both the Memoryless and Average methods, going from 5 to 10 steps gives a significant improvement, while 20 steps is similar to 10. With a sufficient discretization, the mean error is extremely good for the Average method. The Memoryless method, on the other hand, tends to slightly overestimate the consumption.

Fig.7b shows the standard error distributions. As observed for the speed discretization, error for the Memoryless is well balanced while Average method has mostly negative standard errors. Finer discretizations of a seem to give no improvement for the standard error. This may be due to the fact that acceleration has no influence on the travel time for either method, unlike speed which is used to reconstruct the travel times in the Memoryless method.

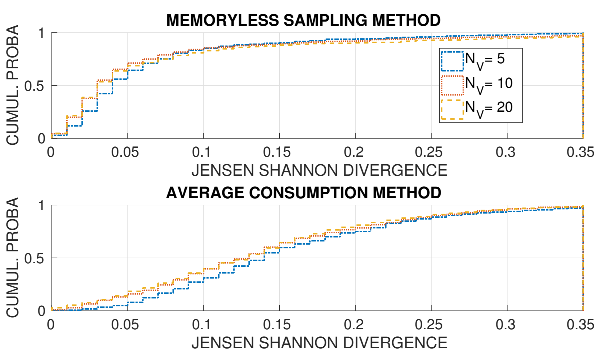
2) Jensen Shannon Divergence: Fig. 7c shows the distribution of the Jensen Shannon divergence when varying the discretization of a. Like in the speed discretization study, we observe that the JS divergence is much smaller overall for the Memoryless Sampling method. Once again, for both methods increasing from 5 to 10 steps improves the indicator, while 20 steps give no additional benefit.



(a) Cumulated distribution of relative Mean error.



(b) Cumulated distribution of relative Standard deviation error.



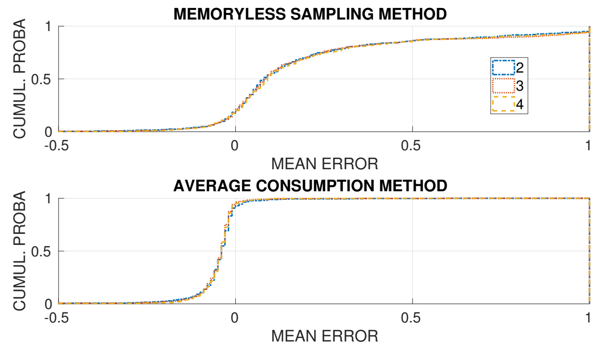
(c) Cumulated distribution of Jensen Shannon divergence.

Fig. 7: Acceleration discretization.

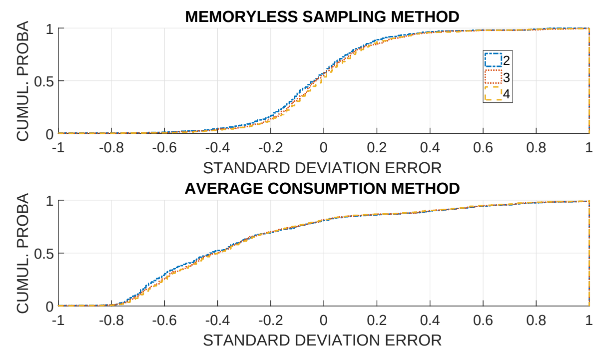
C. Influence of the number of clusters

Our traffic model uses clustering techniques to reduce the size of the traffic data, while retaining the useful information. One would expect some kind of trade-off, where using a larger number of clusters would keep more information at the expense of data size. We will compare the results (in terms of computed energy consumption) for $k = 2, 3$ and 4 clusters per segment. Fig 8a shows the mean error, Fig 8b the standard error and Fig 8c the Jensen Shannon divergence. We observe that increasing the number of clusters to 3 and 4 changes almost nothing, except for a small improvement of the JS divergence. This indicates that for this data set, 2 clusters per segment seem to be enough to retain most of the information in terms of energy consumption.

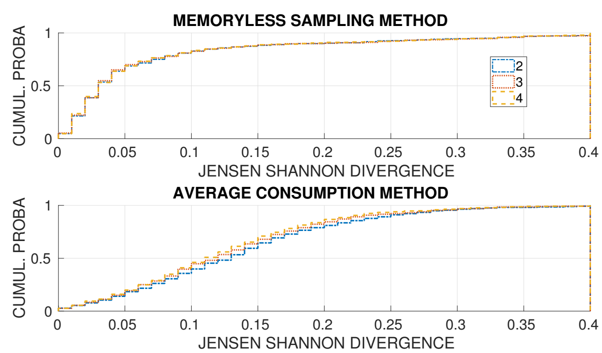
Note: the simulations in this section are for a time-frame $\Delta t = 1h$, due to the increased computational cost for 3 and 4 clusters. It remains to be seen whether a smaller time-frame would benefit more from a higher number of clusters.



(a) Cumulated distribution of relative Mean error.



(b) Cumulated distribution of relative Standard deviation error.



(c) Cumulated distribution of Jensen Shannon divergence.

Fig. 8: Influence of the number of clusters.

D. Summary

Table I summarizes the influence of the speed and acceleration discretizations over the different indicators, for the Memoryless Sampling method and the Average method. Since the time-frame size and number of barycenters seem to have little effect overall, they are not recalled in the table. Also indicated is the overall quality that can be achieved for each indicator.

While the Average method is better to approximate the average consumption, it is superseded by the Memoryless Sampling when it comes to the variability. This leads us to privilege the Memoryless Sampling method to estimate the energy consumption distributions.

TABLE I: + : positive influence, 0 : no influence

Method	Indicator	Speed ($N_{\mathbb{V}}$)	Accel. ($N_{\mathbb{A}}$)	Overall quality
Memoryless	Mean error	0	+	Fair
	Std error	+	+	Good
	JS Divergence	+	+	Good
Average	Mean error	+	+	Very good
	Std error	0	0	Bad
	JS Divergence	+	+	Fair

VI. Results using real data

Thanks to floating car data collected by the smartphone application Geco air [22], we were able to test our method on a real-life scenario. We focused on a portion of the A7 highway near Lyon, France, which is known to be regularly used by commuters. For our analysis, the traffic data collected during the working days of the last two years were aggregated as they were recorded over one day, reasonably assuming that the data share similar traffic patterns. The speed measurements were then divided into 10-minutes time-frames. The Memoryless method is applied to the 1632 collected speed profiles. The discretization of the (speed,acceleration) space is limited at a 10×10 grid to reduce the computation time. One of the main goal of our method is to represent traffic with a small number of (speed,acceleration) distributions, thus we choose to explore the traffic representation with a number of clusters between 1 and 10. In order to evaluate the accuracy of the model, we decided to compare our model to the reference energy consumption distribution computed with the actual speed profiles (Fig.9), and to the energy consumption distribution computed with the average traffic speed, denoted $\bar{v}_{Here}$, obtained from HERE Maps [23].

$$C_{Here} = \frac{P(\bar{v}_{Here}, 0)L}{\bar{v}_{Here}} \tag{10}$$

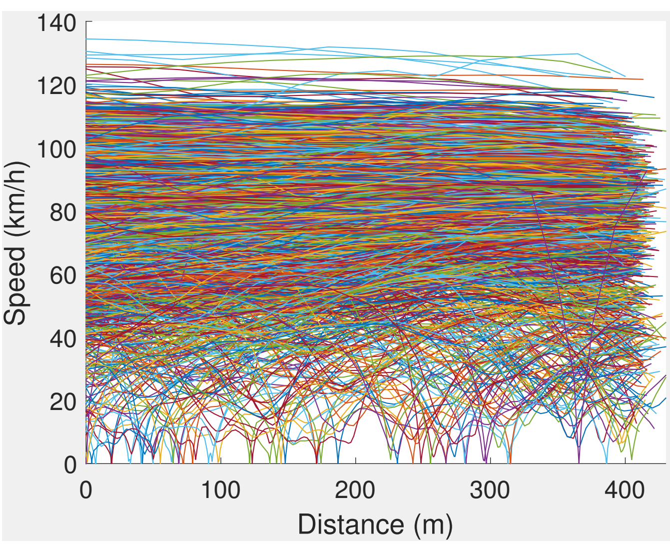


Fig. 9: Speed profiles on the considered highway segment.

A. Influence of the number of clusters

The Memoryless Sampling Method sensitivity to the variation of the number of clusters was analyzed. The number of clusters was chosen in a trade-off between energy consumption estimation accuracy and data occupancy. Fig. 10 shows the absolute mean error and the standard deviation error of the Memoryless method depending on the number of clusters obtained via the K-means algorithm. The performance of the proposed method is compared to the average-speed approach one could use with HERE Maps traffic data, in red in Fig. 10. It is clear that the Memoryless Sampling Method outperforms the HERE average-speed method when it comes to estimate the energy consumption distribution. The mean error remains below 5%, i.e. three times less than the HERE average-speed method. Moreover, the standard deviation tends to decrease when the number of cluster increases. We represent the energy consumptions histograms of the Memoryless method with 4 clusters and of the real driving profiles on the right-hand side of Fig. 11. The histogram of the HERE method is compared to the reference on the left-hand side of the same figure.

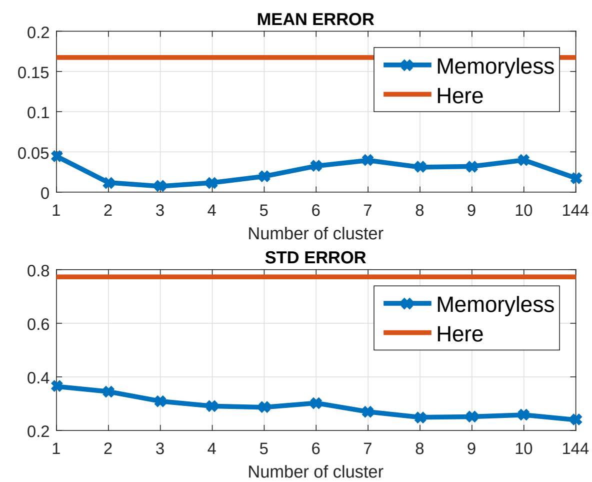


Fig. 10: Absolute mean error and standard deviation for different number of clusters.

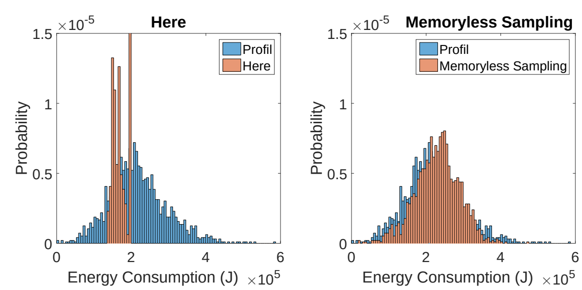


Fig. 11: Histogram of the energy consumption.

B. Interpretation of the clusters

In the following we focus on the 4 clusters case that seems to offer a good trade-off between model size and accuracy. On Fig. 12 we plot the subset of speed profiles belonging to each of the 4 clusters. We observe that these speed profiles do appear rather similar in each cluster. Clusters 1 to 3 correspond to relatively smooth traffic conditions, with little speed variations, 2 being the fastest, followed by 1 then 3. Cluster 4, on the other hand, obviously corresponds to a traffic jam situation, with large variations in speed and frequent drops to null speeds.

Fig. 13 represents the same 4 subsets of profiles for each cluster in the (speed,acceleration) space. This representation confirms that the (speed,acceleration) distributions are quite distinct for each cluster. The level sets on each graph correspond to the distribution of the barycenter of the cluster, and we observe that the barycenters coincide rather well with the profile subsets. Furthermore, we can interpret the clusters in terms of traffic conditions, as summarized on Tab. II.

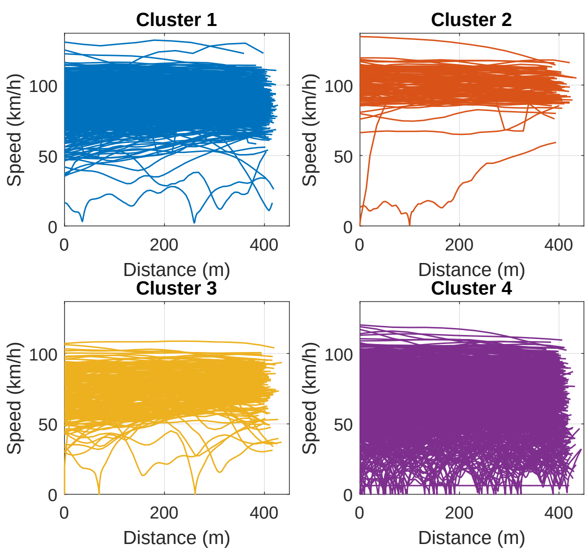


Fig. 12: Speed profiles associated with the clusters.

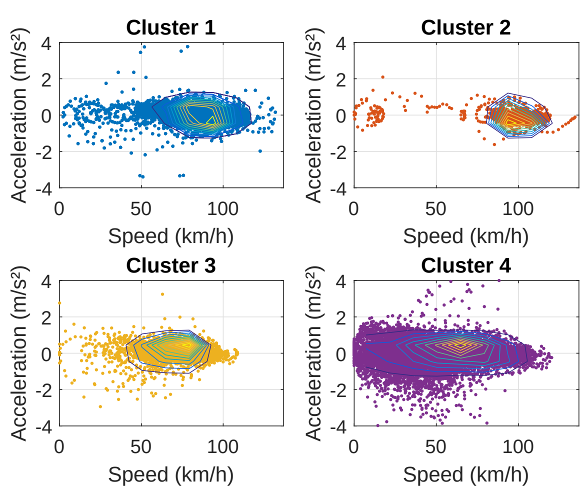


Fig. 13: Real distributions compared to their cluster.

In order to see whether the traffic interpretation is coherent with reality, the clusters associated with each time-frame were plotted in Fig. 14, in which the grey portions correspond to the time-frames with not enough data for the analysis. We can observe that the cluster 3, corresponding to a dense traffic, appears in the morning between 7am and 9:30am. The cluster 4, traffic jam, is present essentially between 4:30pm and 8pm. Cluster 1, normal traffic conditions, during the day and cluster 2, fluid conditions, during the night. Overall, this traffic pattern seems very consistent with reality, exhibiting two peak hours in the morning and evening, typical of commuting behavior.

TABLE II: Traffic Interpretation

Cluster	Mean Speed	Speed Spread	Acceleration	Interpretation
1	Medium	Important	Small	Normal
2	High	Low	Small	Fluid
3	Low	Low	Small	Dense
4	Very low	High	Well spread	Traffic jam

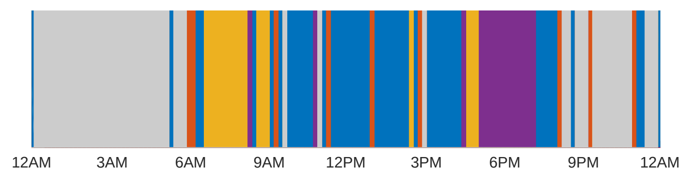


Fig. 14: Traffic clusters according to the time of day.

C. Summary

The analysis conducted on real data confirms that representing traffic by means of (speed,acceleration) distributions is effective not only to estimate the energy consumption distributions in different portions of the road network but also to identify different traffic conditions. Also, clustering proves to be an effective method to reduce data occupancy.

The cluster stage fits its role, allowing us to keep a reasonable data size while retaining most of the useful information from the original set of (v, a) profiles.

VII. Conclusions

In this paper, we have presented a new approach to use traffic data to predict the energy consumption of vehicles. The key point is to consider the (speed,acceleration) data in a statistical sense without the temporal aspect, coupled with a decomposition of the road network into a collection of small segments, based on topological aspects. Numerical experiments carried out with traffic data generated by the traffic simulator SUMO indicate that our approach is able to reconstruct the distribution of the energy consumption over a set of vehicles. We introduce two methods to compute the energy consumption, called Average and Memoryless Sampling methods. Memoryless Sampling method shows better results to estimate the energy consumption distribution.

We also investigate the influence of several parameters such as the (speed,acceleration) discretization, length of time-frame for data aggregation, and number of clusters for the data reduction.

The analysis on real data shows that the Memoryless Sampling method outperforms the common average-speed method. Furthermore, we obtain clusters coherent with real traffic conditions.

Possible future works include a second level of clustering, creating clusters of road segments with close traffic conditions, and variants of the Memoryless Sampling method that would retain some temporality (e.g. Markov).

REFERENCES

[1] "European Strategies - White Paper 2011." [Online]. Available: https://goo.gl/GoSWcf

[2] L. Evans, R. Herman, and T. Lam, "Multivariate Analysis of Traffic Factors Related to Fuel Consumption in Urban Driving," Transportation Science, vol. 10, no. 2, pp. 205–215, 1976. [Online]. Available: https://doi.org/10.1287/trsc.10.2.205

[3] I. D. Vlieger, D. D. Keukeleere, and J. Kretzschmar, "Environmental effects of driving behaviour and congestion related to passenger cars," Atmospheric Environment, vol. 34, no. 27, pp. 4649 – 4655, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S135223100000217X

[4] M. Treiber and A. Kesting, "Traffic flow dynamics," Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg, 2013.

[5] M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, pp. 317–345, 1955.

[6] S. Panwai and H. Dia, "Comparative evaluation of microscopic car-following behavior," IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 3, pp. 314–325, Sept 2005.

[7] S. K. Zegeye, B. De Schutter, J. Hellendoorn, E. A. Breunesse, and A. Hegyi, "Integrated Macroscopic Traffic Flow, Emission, and Fuel Consumption Model for Control Purposes," Transportation Research Part C, vol. 31, pp. 158–171, 2013.

[8] C. Guo, B. Yang, O. Andersen, C. S. Jensen, and K. Torp, "EcoMark 2.0: Empowering Eco-Routing with Vehicular Environmental Models and Actual Vehicle Fuel Consumption Data," Geoinformatica, vol. 19, pp. 567–599, 2015.

[9] E. Ericsson, "Variability in urban driving patterns," Transportation Research Part D: Transport and Environment, vol. 5, no. 5, pp. 337 – 354, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361920900000031

[10] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66). Univ. California Press, Berkeley, Calif., 1967, pp. Vol. I: Statistics, pp. 281–297.

[11] L. N. Vasershtein, "Markov processes over denumerable products of spaces describing large system of automata," Problemy Peredaci Informacii, vol. 5, pp. 64–72, 1969.

[12] R. L. Dobrushin, "Prescribing a System of Random Variables by Conditional Distributions," Theory of Probability & Its Applications, vol. 15, no. 3, pp. 458–486, 1970.

[13] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," SIAM Journal on Mathematical Analysis, vol. 43, no. 2, pp. 904–924, 2011.

[14] D. Lefebvre and M. David, "Dynamic clustering and strong patterns recognition: new tools in automatic classification," Canadian Journal of Earth Sciences, vol. 14, no. 10, pp. 2232–2245, 1977.

[15] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of SUMO - Simulation of Urban MObility," International Journal On Advances in Systems and Measurements, vol. 5, no. 3&4, pp. 128–138, December 2012.

[16] L. Codecá, R. Frank, S. Faye, and T. Engel, "Luxembourg SUMO Traffic (LuST) Scenario: Traffic Demand Evaluation," IEEE Intelligent Transportation Systems Magazine, vol. 9, no. 2, pp. 52–63, 2017.

[17] G. De Nunzio and L. Thibault, "Energy-optimal driving range prediction for electric vehicles," in Intelligent Vehicles Symposium (IV), 2017 IEEE. IEEE, 2017, pp. 1608–1613.

[18] S. Kullback and R. A. Leibler, "On information and sufficiency," Ann. Math. Statistics, vol. 22, pp. 79–86, 1951.

[19] I. Dagan, L. Lee, and F. Pereira, "Similarity-based Methods for Word Sense Disambiguation," in Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, ser. ACL '98. Stroudsburg, PA, USA: Association for Computational Linguistics, 1997, pp. 56–63. [Online]. Available: https://doi.org/10.3115/976909.979625

[20] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," IEEE Transactions on Information Theory, vol. 49, no. 7, pp. 1858–1860, 2003.

[21] F. Österreicher and I. Vajda, "A new class of metric divergences on probability spaces and its applicability in statistics," Annals of the Institute of Statistical Mathematics, vol. 55, no. 3, pp. 639–653, Sep 2003.

[22] "Geco air." [Online]. Available: http://www.gecoair.fr/home_en/

[23] "HERE." [Online]. Available: https://developer.here.com/

Arthur Le Rhun received a M.Sc. degree in engineering from the ENSTA Paristech (France), a M.Sc. degree in Optimization from Université Paris XI both in 2016. He is currently pursuing a Ph.D. in applied mathematics at the Ecole polytechnique and IFP Energies nouvelles. His research theme is dynamic optimization for stochastic systems, and applications to energy management of hybrid vehicles.

Frederic Bonnans Got an engineering degree from the Ecole Centrale Paris, a PhD (docteur ingenieur)from the Université Technique de Compiègne, and the Habilitation from Université Paris IX Dauphine in 1979, 1982 and 1993 resp. He is currently a senior researcher at Inria Saclay and Ecole polytechnique. His research theme is dynamic optimization for deterministic and stochastic systems, and applications to energy management and biological models.

Giovanni De Nunzio received the B.Sc. and M.Sc. degrees in information and automation engineering from the University of L'Aquila, Italy, in 2007 and 2010. He received the doctorate degree in automatic control from the Grenoble Institute of Technology, France, in 2015. He is currently a research engineer at IFP Energies nouvelles, Lyon, France. His research activities focus on modeling, control, and simulation of traffic and transportation systems, vehicle powertrain modeling, advanced driver-assistance systems, energy consumption minimization, and graph theory.

Thomas Leroy graduated from ESSTIN and received the M.Sc. degree in control theory from the University Henri Poincaré in Nancy, France, in 2006. He received the doctorate degree in Control theory and Mathematics from the Ecole des Mines ParisTech, France, in 2010. He is currently a research engineer at IFP Energies nouvelles, Rueil-Malamaison, France. His research activities focus on modeling, control, and simulation of internal combustion engines and hybrid electric powertrains.

Pierre Martinon received his M.Sc. and Ph.D. degrees in applied mathematics from the ENSEEIHT (Toulouse, France) in 2001 and 2005. He is currently a researcher at Inria Saclay and Ecole polytechnique. His research field is optimization for dynamical systems.